Undergraduate Lecture Notes in Physics

Alessandro Bettini

A Course in Classical Physics 4 - Waves and Light

Springer

Undergraduate Lecture Notes in Physics

Undergraduate Lecture Notes in Physics (ULNP) publishes authoritative texts covering topics throughout pure and applied physics. Each title in the series is suitable as a basis for undergraduate instruction, typically containing practice problems, worked examples, chapter summaries, and suggestions for further reading.

ULNP titles must provide at least one of the following:

- An exceptionally clear and concise treatment of a standard undergraduate subject.
- A solid undergraduate-level introduction to a graduate, advanced, or non-standard subject.
- A novel perspective or an unusual approach to teaching a subject.

ULNP especially encourages new, original, and idiosyncratic approaches to physics teaching at the undergraduate level.

The purpose of ULNP is to provide intriguing, absorbing books that will continue to be the reader's preferred reference throughout their academic career.

Series editors

More information about this series at http://www.springer.com/series/8917

Alessandro Bettini

A Course in Classical Physics 4 - Waves and Light

Springer

Alessandro Bettini
Dipartimento di Fisica e Astronomia
Università di Padova
Padua
Italy

The text is partially based on the book “Le onde e la luce”, A. Bettini, Zanichelli, 1993. The author own all rights in the former publications.

ISSN 2192-4791 ISSN 2192-4805 (electronic)
Undergraduate Lecture Notes in Physics
ISBN 978-3-319-48328-3 ISBN 978-3-319-48329-0 (eBook)
DOI 10.1007/978-3-319-48329-0

Library of Congress Control Number: 2016954707

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This is the fourth in a series of four volumes, all written at an elementary calculus level. The complete course covers the most important areas of classical physics, such as mechanics, thermodynamics, statistical mechanics, electromagnetism, waves, and optics. The volumes are the result of a translation, an in-depth revision and an update of the Italian version published by Decibel-Zanichelli. This fourth volume deals with oscillations, waves, and light.

It is assumed that the reader knows differential calculus and the simplest properties of the vector fields (the same as in Volume 3), such as the gradient of a scalar field, the divergence and curl of a vector field, and the basic theorems on the line integral of a gradient, the Gauss divergence theorem and the Stokes curl theorem. We shall also assume that the reader has already learned the basic concepts of mechanics and electromagnetism up to Maxwell's equations, as developed in the first volumes of this course, to which we shall make explicit reference when needed, or equivalent ones.

Oscillations about a stable state of equilibrium are a very common natural phenomenon, present in all sectors of physics ranging from mechanics to electromagnetism and from astrophysics to atomic and nuclear physics. A spider hanging from its gossamer thread, if displaced from the equilibrium position, oscillates back and forth as long as passive resistances do not stop it. A blade of grass pushed by the wind moves periodically up and down. A boat on the surface of a lake oscillates under the action of the waves. Large systems vibrate as well. The earth's atmosphere does so under the periodic action of the moon over a period of about 12 h. The earth itself vibrates for a while when hit by an intense seismic shock, as do extremely small objects. Light itself is produced by the vibrations of atoms, namely the oscillations of the electrons they contain, which, as all accelerating charges do, produce an electromagnetic wave. An electric circuit containing an inductance and a capacitance performs harmonic, electric oscillations that are completely similar to the mechanical oscillations of a pendulum. Electric and the magnetic fields also oscillate in a vacuum, when they are the fields of an electromagnetic wave. And space-time itself vibrates when a gravitational wave crosses it.

Hence, many physical systems exist that can perform oscillations. Each system may obey a different law, the Newtonian law if it is mechanical, the Maxwell's equation if it is electromagnetic, or the quantum equation if it is an atom, but their motions have similar characteristics. In particular, they are very often harmonic motions, in a first, but usually very good, approximation. This is a consequence of the fact that the oscillation takes place in the proximity of a stable equilibrium configuration and the system is attracted to that by a restoring force (or, more generally, by an action) proportional to the displacement from that configuration. As a consequence, the differential equation governing their motions is the same for all of these systems. The first part of this book deals with such small oscillation phenomena.

In the remaining parts, we shall discuss waves. The word immediately brings to mind the motion of the sea. Suppose, then, that we are on a beach and observing the motion of the waves approaching us from the open sea, moving with a defined speed, and crashing on the rocks under our observation point. Each wave is different from the previous one, but some features are always evidently equal for all of them: their speed, the distance between two crests (namely the wavelength), and the period during which a given point of the surface rises and lowers. The show is fascinating and could hold our attention for a long time, but, we may ponder, are there other waves around us? They are not as evident, but knowing a bit about physics, we know that there are indeed. The sound that reaches our ears was caused by the impact of the water on the rocks and the cries of seagulls, both are waves. And so is the sunlight that enlightens and heats us. It is waves that allow us to perceive the image of the sea and of the person standing close to us. Waves run along our nerves from the retina and the eardrum to the brain, and then, there are those that intersect in our brain while we think and feel, although we do not know how.

Wave phenomena, therefore, are present in different physical systems, ranging from mechanical to electrical and from biological to quantum. Like oscillations, the different types of waves have common characteristics. Again, this is due to the fact that the equation governing the different systems is, under many circumstances, exactly the same. Our study will therefore be initially addressed to the general properties of waves, common to their different types. To be concrete, we shall exemplify this phenomenon through two of the most important cases, namely sound and electromagnetic waves. Subsequently, we will focus on the visible electromagnetic waves, which are light. That is to say, we will study optics.

Quantum physics describes the phenomena on the atomic and subatomic levels by associating each particle, atom, electron, nucleus, etc., with a “wave function.” This is a complex function of the coordinates and of time, whose amplitude squared gives the probability of finding the particle at a given point in a given instant. This function behaves exactly like a wave. All of this is outside the scope of these lectures. However, we observe that several surprising aspects of quantum mechanics are entirely similar to completely classical aspects of the physics of oscillations and waves. For example, inverse proportionality relationships, which we shall look at in Chap. 2, between the duration of a signal in time and the width of its frequency spectrum and that between the extension of a wave front and the

width of the angular distribution of the wave vector, strictly correspond in quantum physics to the Heisenberg uncertainty relations. In other words, the uncertainty relations are characteristic of *all wave phenomena*, not only of the quantum examples but of the classical ones as well. Another example is the wave phenomenon of the "frustrated vanishing wave," which we will look at in Chap. 4. It corresponds to the "tunnel effect" in quantum mechanics. Consequently, a deep enough understanding of classical wave phenomena will substantially help the student when he/she tackles quantum physics.

The first two chapters are devoted to the study of oscillations of physical systems whose state is determined by a single coordinate (Chap. 1) and more than one coordinate (Chap. 2). Chapter 1 studies the oscillatory motion of a simple pendulum and similar physical systems, in as much as they are described by the same differential equation. We shall study oscillations in both the presence and absence of damping and the presence and absence of an external periodic solicitation. In the second chapter, we shall study more complicated systems, such as two pendulums connected by a spring. We shall see that while the generic motion of these systems is complicated, there are special motions, called normal modes, which are very simple; namely, they are harmonic oscillations of all parts of the system in phase with one another. We shall then see that even a continuous system like a guitar string has particular motions that are its normal modes. The discussion will lead us to discover an important mathematical tool, namely the harmonic analysis (Fourier transform). This is an instrument that we shall use often in what follows.

In Chap. 3, we define the concept of the wave. We shall then study how a wave can be produced, how it is reflected, and how it can be destroyed (absorbed). We shall deal with electromagnetic waves and sound waves, as particularly important examples. In Chap. 4, we learn that there are different concepts of wave velocity and their relations with energy propagation, in particular the velocity of the phase for a wave of definite wavelength and the group velocity for all types of wave. We shall see that under several physical circumstances, the phase velocity is a function of the wavelength. This is especially true for light waves in material media. We shall examine the consequences and study the physical reason for the phenomenon.

In the subsequent chapters, we will focus solely on light waves, which are electromagnetic waves in the range of wavelength in which they can be perceived by the human eye. These wavelengths are very small, a few tenths of a micrometer, and the frequencies are very high, on the order of hundreds of THz. In Chap. 5, we shall study the phenomena of interference and diffraction of light, which are the characteristics of the wave nature of light. In the sixth chapter, we shall deal with the consequences of the fact that the electric field of light waves can vibrate in different directions, all perpendicular to the propagation direction. These are the polarization phenomena. We shall study the different polarization states of light, how polarization can be produced from non-polarized light, and how the polarization state can be experimentally analyzed.

In the last two chapters, we study the imaging processes. These are the processes that take place in our own eye, as well as in optical instruments, providing us with much of the information we have on the outside world. In Chap. 7, we shall learn,

step-by-step, the process for the formation of images of point-like objects, of objects of two and then of three dimensions. The processes of image formation may be different, depending on whether one uses mirrors, lenses, or interference patterns. Topics covered in the first part of the chapter normally fall under the name of geometrical optics. However, the physics of these phenomena is dealt here with constant attention to the wave nature of light. In the last chapter, we shall see that the image formation process through a lens amounts to a succession of a Fourier transform (from the image to the back focal plane) followed by a Fourier anti-transform (from the back focal plane to the image plane). Finally, using the coherent light of the laser, we shall see how we can produce those actual three-dimensional images known as holograms.

Physics is an experimental science, meaning that it is based on the experimental method, which was developed by Galileo Galilei in the seventeenth century. The process of understanding physical phenomena is not immediate, but rather, it advances through trial and error, in a series of experiments, which might lead, with a bit of fortune and a lot of thought, to the discovery of the governing laws. Induction of the process of physical laws goes back from the observed effects to their causes, and, as such, cannot be purely logical. Once a physical law is found, it is necessary to consider all its possible consequences. This then becomes a deductive process, which is logical and similar to that of mathematics. Each of the consequences of the law, in other words, its predictions, must then be experimentally verified. If only one prediction is found to be false through the experiment, even if thousands of them had been found true, it is enough to prove that the law is false. As R. Feynman wrote on the blackboard at the very beginning of his famous lecture course, "We are not concerned with where a new idea comes from—the sole test of its validity is experiment."

This implies that we can never be completely sure that a law is true; indeed, the number of its possible predictions is limitless, and at any historical moment, not all of them have been controlled. However, this is the price we must pay in choosing the experimental method, which has allowed humankind to advance much further in the last four centuries than in all the preceding millennia.

The path of science is complex, laborious, and highly nonlinear. In its development, errors have been made and hypotheses have been advanced that turned out to be false, but ultimately, laws were discovered. The knowledge of at least a few of the most important aspects of this process is indispensable for developing the mental capabilities necessary for anybody who wishes to contribute to the progress of the natural sciences, whether they pursue applications or teach them. For this reason, we have included brief historical inserts recalling the principal authors and quoting their words describing their discoveries.

Quite often, aspects of oscillation and wave phenomena described in this book can be observed in the nature around us. Some of these observations are mentioned, including pictures when relevant. More photographs and movies can be found on the Web.

Each chapter of the book starts with a brief introduction on the scope that will give the reader a preliminary idea of the arguments he/she will find. There is no

need to fully understand these introductions on the first reading, as all the arguments are fully developed in the subsequent pages.

At the end of each chapter, the reader will find a summary and a number of queries with which to check his/her level of understanding of the chapter's arguments. The difficulty of the queries is variable: Some of them are very simple, some is more complex, and a few are true numerical exercises. However, the book does not contain any sequence of full exercises, owing to the existence of very good textbooks dedicated specifically to that.

Acknowledgments

Thanks go to
Nelson Kenter for the photograph in Fig. 4.12,
Sara Magrin and Roberto Temporin for help with Figs. 5.10, 5.25, and 5.34,
Chris Button Photography for the photograph in Fig. 5.18,
Caterina Braggio for the photograph in Fig. 5.21,
Rik Littlefield, Zerene Systems LLC for the photograph in Fig. 5.29 and
Paul D. Maley for the photo in Fig. 8.14 b.

Padua, Italy Alessandro Bettini

Contents

Symbols

Symbols for the Principal Quantities

d	Absorption distance
$\mathbf{a}$, a_s	Acceleration
T	Amplitude transparency
$\boldsymbol{\alpha}$, α	Angular acceleration
ω	Angular frequency
J	Angular magnification
$\boldsymbol{\omega}$	Angular velocity
C	Capacitance
Z	Characteristic impedance
q, Q	Charge
ρ	Charge density
γ	Contrast
$\mathbf{j}$	Current density
I	Current intensity
Γ	Curve
DOF	Depth of field
D	Diameter
κ	Dielectric constant
k	Elastic constant
$\mathbf{E}$	Electric field
E	Electromotive force (emf)
q_e	Elementary charge
$\mathbf{p}$	Electric dipole moment
$\mathbf{D}$	Electric displacement
χ_e	Electric susceptibility
m_e	Electron mass
U	Energy
w	Energy density (of the field)

Φ	Energy flux
e	Exposure
f	Focal length
$\mathbf{F}$	Force
C	Fourier transform
ν	Frequency
γ	Fringe visibility
$\mathbf{G}$	Gravitational field
g	Gravity acceleration
υ_g	Group velocity
E_L	Illuminance
Z	Impedance
i	Incidence angle
L	Inductance
I	Intensity
E	Irradiance
p	Lens object distance
q	Lens image distance
$\Gamma_{\lim}$	Limit angle (resolution)
B	Luminance
Φ_L	Luminous flux
J_L	Luminous intensity
$\mathbf{H}$	Magnetic auxiliary field
$\boldsymbol{\mu}$	Magnetic dipole moment
Φ, Φ_B	Magnetic flux
μ	Magnetic permeability (absolute)
κ	Magnetic permeability (relative)
χ_m	Magnetic susceptibility
$\mathbf{M}$	Magnetization
m	Magnification
m, M	Mass
$\langle x \rangle$	Mean value, of x
μ	Molar mass
$\mathbf{p}$	Momentum
G_N	Newton constant
NA	Numerical aperture
T	Period
Π	Period in square distance
υ_p	Phase velocity
e_L	Photometric exposure
θ, α	Plane angle
θ, ϕ	Polar angle
ρ, θ, ϕ	Polar coordinates (space)
$\mathbf{P}$	Polarization (density)
$\mathbf{r}$	Position vector

ϕ	Potential (electrostatic and scalar)
U_p	Potential energy
$\mathbf{S}$	Poynting vector
p, P	Pressure
ω_0	Proper angular frequency
Q	Q-factor
J	Radiant emission intensity
I	Radiant intensity
Φ	Radiant power
r, R, ρ	Radius
γ	Ratio of specific heats (gas)
n	Refractive index
r	Refraction angle
R	Resistance (electric)
ρ	Resistivity
γ	Resonance width
L	Self-inductance
Ω	Solid angle
k	Spatial frequency
ρ	Specific rotation constant
η	Spectral luminous efficacy
S, Σ	Surface
t	Time
τ	Time constant
T	Tension (of a rope)
c	Velocity of light (in vacuum)
$\mathbf{v}$,υ	Velocity
$\mathbf{u}_\upsilon$	Unit vector of $\mathbf{v}$
$\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$	Unit vectors of the axes
$\mathbf{n}$	Unit vector normal to a surface
ε_0	Vacuum permittivity
μ_0	Vacuum permeability
$\mathbf{A}$	Vector potential
V	Volume
λ	Wavelength
ψ	Wave function
v_s	Wave number
$\mathbf{k}$	Wave vector
D	Width (slit, aperture)
W	Work

Base Units in the SI

Quantity	Unit	Symbol
Length	metre/meter	m
Mass	kilogram	kg
Time	second	s
Current intensity	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

Decimal Multiples and Submultiples of the Units

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^{9}	giga	G	10^{-12}	pico	p
10^{6}	mega	M	10^{-15}	femto	f
10^{3}	kilo	k	10^{-18}	atto	a
10^{2}	hecto	h	10^{-21}	zepto	z
10	deka	da	10^{-24}	yocto	y

Fundamental Constants

Quantity	Symbol	Value	Uncertainty
Speed of light in vacuum	c	299,792,458 m s^{-1}	Defined
Newton constant	G_N	6.67308(31) × 10^{-11} m^3 kg^{-1} s^{-2}	47 ppm
Avogadro's number	N_A	6.022140857(74) × 10^{23} mol^{-1}	12 ppb
Boltzmann constant	k_B	1.38064852(79) × 10^{-23} J K^{-1}	570 ppb
Vacuum permittivity	$\varepsilon_0 = 1/(c^2\mu_0)$	8.854187817… × 10^{-12} Fm^{-1}	Defined
Vacuum permeability	$\mu_0 = 1/(c^2\varepsilon_0)$	12.566370614… × 10^{-7} NA^{-2}	Defined
Elementary charge	q_e	1.6021766208(98)… × 10^{-19} C	6.1 ppb
Unified atomic mass	u = 1 g/N_A	1.660539040(20) × 10^{-27} kg	12 ppb

(continued)

(continued)

Quantity	Symbol	Value	Uncertainty
Electron mass	m_e	$9.10938356(11) \times 10^{-31}$ kg	12 ppb
Proton mass	m_p	$1.672621898(21) \times 10^{-27}$ kg	12 ppb

Greek Alphabet

alpha	α	A	iota	ι	I	rho	ρ	P
beta	β	B	kappa	κ	K	sigma	σ, ς	Σ
gamma	γ	Γ	lambda	λ	Λ	tau	τ	T
delta	δ	Δ	mu	μ	M	upsilon	υ	Y, ϒ
epsilon	ε	E	nu	ν	N	phi	ϕ, φ	Φ
zeta	ζ	Z	xi	ξ	Ξ	chi	χ	X
eta	η	H	omicron	ο	O	psi	ψ	Ψ
theta	θ, ϑ	Θ	pi	π	Π	omega	ω	Ω

Chapter 1
Oscillations of Systems with One Degree of Freedom

Abstract Oscillations are periodic or quasi-periodic motions, or, more generally, the evolution, of a large number of physical systems, which may be very different from one another. However, the principal characteristics of the oscillatory phenomena are similar. This is because the differential equation describing those systems is the same. In this chapter, we study the oscillations of systems with one degree of freedom, both mechanical and electric. We deal with free, damped and forced oscillations and study the resonance phenomenon.

Many situations exist in which the motion, or, more generally, the evolution, of a system is confined to a limited region. Think, for example, of the oscillations of a pendulum, the vibrations of a drumhead or the string of a guitar, the electric oscillations of a circuit in a radio receiver, the waves flickering in a water glass, etc. In each case, the "oscillatory motion" occurs around a stable equilibrium position of the system. The few examples we just gave, and the many others that come to mind, happen in the most diverse physical systems. However, many characteristics of oscillatory motions are similar and largely independent of the physical nature of the system. The reason for this is that those motions are described by the same differential equation, whose unknown is the function that measures the displacement of the system from equilibrium. This function is different from one system to another; it is the horizontal displacement for the pendulum, the deformation of the drumhead or of the guitar string, the charge of the capacitor of the oscillating circuit, the height of a wave of the water in a glass, etc.

The time dependence of the oscillation of a system is not necessarily a simple mathematical function, but is simple when the restoring force, or, more generally, the restoring agent, is proportional to the displacement from the equilibrium position. The typical, but not unique, case is the elastic force. In practice, this is a case with a good approximation of many systems, provided that the oscillation amplitude is small. Under such conditions, the differential equation of the system is linear, and its solution is a sinusoidal function of time. The motion is said to be harmonic.

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_1

In this chapter, we study the harmonic oscillations of the systems with one degree of freedom, namely the simplest ones. Several elements of the oscillations of mechanical systems were anticipated in Chap. 3 of the 1st volume and of electric circuits in Chap. 7 of the 3rd volume of this course. Here, we repeat and extend the analysis.

A configuration of a system with one degree of freedom is defined by a single coordinate or variable. We shall study, in particular, both mechanical and electric oscillators. The mathematical description is exactly the same. In Sect. 1.1, we deal with the free oscillator under the idealized condition of negligible dissipative agents (viscosity in the mechanical case, resistance in the electric case). In the subsequent section, we take dissipative effects into account. In Sect. 1.3, we study what happens when a periodic external force acts on the oscillator. Its motions are called forced oscillations. In Sects. 1.4 and 1.5, we deal with the important phenomenon of resonance, which happens when the frequency of the external force gets close to the proper frequency of the oscillator.

Energy is stored in any oscillating system. In a free oscillator, energy remains constant if dissipative actions are absent. Contrarily, if they are present, as is always the case in practice, and if they are proportional to velocity, the stored energy decreases exponentially with time.

1.1 Free Harmonic Oscillations

Harmonic oscillations are ubiquitously present in all sectors of physics. The prototype phenomenon is the harmonic motion of a point-like mass. This is a periodic motion around a stable equilibrium position in which the restoring force is proportional and in the opposite direction to the displacement. The oscillating system may be an extended mechanical system, an electric circuit, a molecule, a star, etc., and the restoring agent may be something other than a force, as we shall discuss in this chapter. In this section, we deal with free oscillations, namely those made spontaneously by a system about a stable equilibrium position.

A well-known example is the simple pendulum, which is a material point constrained to move along an arc of a circumference in a vertical plane (see Sect. 2.9 of the 1st volume). Let the pendulum initially be in its natural position of stable equilibrium. Let us now remove it somewhat and let it go with zero initial velocity. We shall see the pendulum moving toward the equilibrium position with increasing speed. The restoring force is a component of the weight. In the initial phases of the motion, the potential energy of the restoring force diminishes, while the kinetic energy increases. The pendulum reaches the equilibrium position with a kinetic energy equal to its initial potential energy. It does not stop there, but rather keeps going through inertia on the other side. Its kinetic energy is now diminishing and its potential energy increasing, until the point at which a position at the same distance from equilibrium as the initial one is reached, and so on. The oscillation continues with an amplitude that would be constant with time if dissipative forces

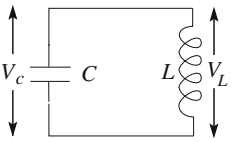


Fig. 1.1 An oscillating circuit

(like the air drag) were not present. The motion is a harmonic motion. Another example is the motion of a material point linked to a spring, as we shall soon discuss quantitatively.

An example of a non-mechanical system is the *oscillating circuit* shown in Fig. 1.1. The plates of a capacitor C are connected to an inductor L through a switch S. Clearly, when the circuit is closed, the stable equilibrium configuration occurs when the charges on the plates are zero. Otherwise, the capacitor would discharge along the circuit. Let us now imagine charging the capacitor with the switch open and then closing it. The current intensity I, which is initially null, will increase while the capacitor discharges. The magnetic field inside the inductor will increase with I. The energy initially stored in the electric field inside the capacitor decreases, while the energy stored in the magnetic field inside the inductor increases. When the capacitor has reached the equilibrium position, namely is discharged, it cannot stop there. Indeed, this would imply a sudden variation of the current. This is opposed by a counter-electromotive force (cemf, for short) that develops in the inductor. The inductance plays the role here of inertia in mechanics. In this case too, the system overtakes the equilibrium position and moves to a configuration symmetrical to the initial one (if the dissipative effects can be neglected). Namely, the current vanishes when the charges on the plates of the capacitor are equal to and opposite of the initial ones.

In this section, we shall consider oscillating systems with a single degree of freedom. Their configuration is determined by a single variable, like the displacement angle for the pendulum or the charge of the capacitor for the oscillating circuit. Let q be such a variable, or coordinate, as we can also call it. We shall neglect dissipative effects and assume that the energy of the system is conserved. Let $U(q)$ be the potential energy of the system (energy of the weight, elastic energy, energy of the capacitor, etc.) and let it have a minimum at $q = q_0$. We shall consider "movements" close to this stable equilibrium state, namely, as they are called, *small oscillations*.

For values of q near enough to q_0, we can develop in series $U(q)$ and stop at the first non-zero term. The first derivative dU/dq is certainly zero in q_0, because this is an extreme, while $d^2U/dq^2 > 0$ there because the extreme is a minimum. Defining $k \equiv (d^2U/dq^2)_{q_0}$, we can write

$$U(q) - U(q_0) = \frac{k}{2}(q - q_0)^2. \tag{1.1}$$

It is convenient to choose the arbitrary additive constant of the potential energy, such as $U(q_0) = 0$, and to define

$$x = q - q_0. \tag{1.2}$$

Namely, x is the displacement from the equilibrium position, or configuration. We then have

$$U(x) = \frac{k}{2}x^2. \tag{1.3}$$

Let us now consider, for the sake of concreteness, a mechanical system, consisting of a material point of mass m. The force corresponding to the potential Eq. (1.3) is

$$F(x) = \frac{dU}{dx} = -kx. \tag{1.4}$$

Namely, the restoring force is *proportional* to the displacement.

To be concrete, let us consider the system shown in Fig. 1.2. A block, of mass m, lies on a horizontal plane, which we assume to be frictionless. You may think, for example, to have a cavity in the block filled with dry ice and a number of small holes in its bottom. A layer of CO_2 gas will develop between the bottom of the block and the support plane, reducing the friction to very small values. A spring is connected to the block at one end and to a fixed point on the other. We assume we are in the range of validity of Hook's law. The restoring force and its potential energy are then given by Eqs. (1.4) and (1.3), respectively, with k equal to the elastic constant of the spring, also called the spring constant.

The equation of motion is

$$-kx(t) = m\frac{d^2x}{dt^2}$$

which we write in the canonical form

$$\frac{d^2x}{dt^2} + \frac{k}{m}x(t) = 0. \tag{1.5}$$

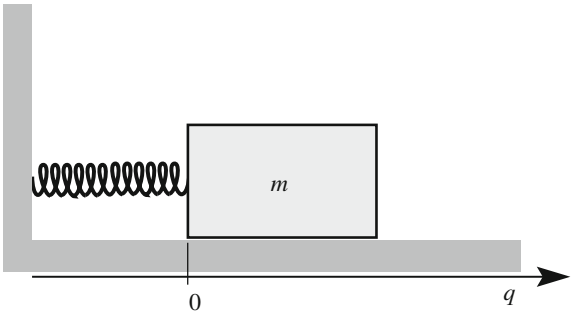


Fig. 1.2 A mechanical oscillator

This is the differential equation of the harmonic oscillator. We now introduce the positive quantity

$$\omega_0^2 = k/m. \tag{1.6}$$

This has a very important dynamical meaning. ω_0^2 is *the restoring force per unit displacement and per unit mass*. It depends on the characteristics of the system. We can then write Eq. (1.5) as

$$\frac{d^2x}{dt^2} + \omega_0^2 x(t) = 0. \tag{1.7}$$

We observe that the differential equation is *linear* as a consequence of having stopped the development in series of U at the first non-zero term (also called the leading term). Consequently, the approximation holds for small values of x only. At larger values of x, terms of order x^3 or larger in the development of U in Eq. (1.3), corresponding to order x^2 or larger in the development of F in Eq. (1.4), may become relevant. The equation of motion is no longer linear under the latter conditions. We shall see important consequences of the linearity subsequently.

The general solution of Eq. (1.7), as determined by calculus, is

$$x(t) = a\,\cos\omega_0 t + b\,\sin\omega_0 t, \tag{1.8}$$

where the constants a and b must be determined from the initial conditions of the motion. They are two in number because the differential equation is of the second order.

The general solution can also be expressed in the, often more convenient, form

$$x(t) = A\,\cos(\omega_0 t + \phi), \tag{1.9}$$

where the constants to be determined from the initial conditions are now A and ϕ.

To find the relations between two pairs of constants, we note that

$$A\,\cos(\omega_0 t + \phi) = A\,\cos\phi\,\cos\omega_0 t - A\,\sin\phi\,\sin\omega_0 t.$$

Hence, it is

$$a = A\,\cos\phi, \quad b = -A\,\sin\phi \tag{1.10}$$

and reciprocally

$$A = \sqrt{a^2 + b^2}, \quad \phi = -\arctan(b/a). \tag{1.11}$$

We now introduce the terms that are used when dealing with this type of motion. To do that in a general way, consider the expression (with a generic ω)

$$x(t) = A\ \cos(\omega t + \phi). \tag{1.12}$$

The motion is not only periodic, but, specifically, its time dependence is given by a circular function. Such motions are said to be *harmonic*. A is called the *oscillation amplitude*, the argument of the cosine, $\omega t + \phi$, is called the *phase* (or *instantaneous phase* in the case of ambiguity) and the constant ϕ is called the *initial phase* (indeed, it is the value of the phase at $t = 0$). The quantity ω, which has the physical dimensions of the inverse of time, is called the *angular frequency* and also the *pulsation*. Its kinematic physical meaning is *the rate of the variation of the phase with time* and, we will notice, is independent of the initial conditions of the motion. In the specific case we have considered above, the harmonic motion is the spontaneous motion of the system (in Sects. 1.3 and 1.4, we shall study motions under the action of external forces) and the angular frequency, ω_0, as in Eq. (1.9), is called the *proper* angular frequency.

The harmonic motion is periodic, with *period*

$$T = 2\pi/\omega. \tag{1.13}$$

The number of oscillations per unit time is called the *frequency*, ν. Obviously, it is linked to the period and to the angular frequency by

$$\nu = \frac{1}{T} = \frac{\omega}{2\pi}. \tag{1.14}$$

The period is measured in seconds, the frequency in hertz (1 Hz = 1 s^{-1}), and the angular frequency in rad s^{-1} or simply in s^{-1}. The unit is named after Heinrich Rudolf Hertz (1857–1899).

The harmonic motion can be viewed from another point of view. Consider a circular disc and a small ball attached to a point of its rim. The disc can rotate on a horizontal plane around a vertical axis at its center. Suppose the disc is rotating with a constant angular velocity ω. If we look at the ball from above, in the direction of the axis, we see a circular motion, but if we look horizontally, with our eye on the plane of the rotation, we see the ball oscillating back and forth periodically. Indeed, the motion is not only periodic, it is harmonic, as we will now show.

Figure 1.3 shows the material point P moving along a circumference of radius A with constant angular velocity ω. We call ϕ the angle between the position vector at $t = 0$ and the x-axis. The co-ordinates of P at the generic time t are

$$x(t) = A\ \cos(\omega t + \phi), \quad y(t) = A\ \sin(\omega t + \phi).$$

The projection of the motion on the axes, in particular on x, is harmonic.

The conclusion leads to the simple graphical representation of the harmonic phenomena shown in Fig. 1.4. To represent a harmonic motion of amplitude A, angular frequency ω and initial phase ϕ, we take a fixed reference axis x and a vector $\mathbf{A}$, of magnitude A, rotating around its origin in the plane of the figure at the constant angular frequency ω and forming, with the x axis, the angle ϕ at $t = 0$. The projection of $\mathbf{A}$ on the reference x axis is our harmonic motion.

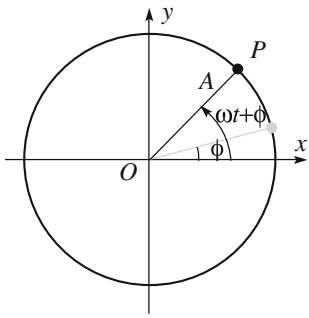


Fig. 1.3 A point P moving in a circular uniform motion

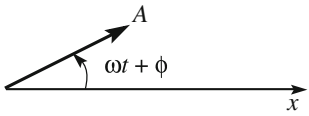


Fig. 1.4 Vector diagram for the harmonic motion

This is shown in Fig. 1.4. The same can be represented analytically by considering Eq. (1.12) as the real part of the complex function

$$z(t) = \left(A\mathrm{e}^{i\phi}\right)\mathrm{e}^{i\omega t} = \alpha \mathrm{e}^{i\omega t}, \tag{1.15}$$

namely as

$$x(t) = \mathrm{Re}[z(t)] = \mathrm{Re}\left(\alpha \mathrm{e}^{i\omega t}\right), \tag{1.16}$$

where the two real integration constants have been included in the complex one

$$\alpha = A\mathrm{e}^{i\phi}, \tag{1.17}$$

which is called the *complex amplitude*. Its modulus is the real amplitude, its argument is the initial phase. The complex notation is often quite simple for the following reasons: (a) operations are simpler with exponentials than they are with trigonometry, (b) differentiating an exponential, one obtains another exponential, (c) the most usual operations (addition, subtraction, multiplication by a constant factor, differentiation and integration) commute with taking the real part; consequently, the latter operation may be done at the end. Notice, however, that the same is not true for non-linear operations like multiplication or exponentiation.

Let us now consider the velocity. The derivative of Eq. (1.12) gives us

$$\frac{dx}{dt} = -A\,\omega\,\sin(\omega t + \phi) = A\omega\,\cos\left(\omega t + \phi + \frac{\pi}{2}\right). \tag{1.18}$$

and, equivalently, the derivative of Eq. (1.15)

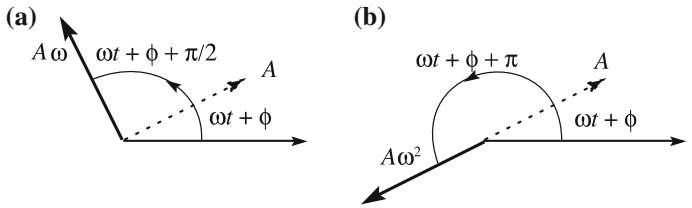


Fig. 1.5 Vector diagram for harmonic motion. **a** velocity, **b** acceleration

$$\frac{dz}{dt} = i\omega\alpha e^{i\omega t} = \omega\alpha e^{i(\omega t + \pi/2)}, \tag{1.19}$$

where we recalled that $i = e^{i\pi/2}$.

In both forms, we see that the velocity varies in a harmonic way as well, with a phase that is forward of $\pi/2$ radians to the displacement. This is shown in Fig. 1.5a.

Differentiating once more, we have the acceleration

$$\frac{d^2x}{dt^2} = -A\omega^2 \cos(\omega t + \phi) = A\omega^2 \cos(\omega t + \phi + \pi). \tag{1.20}$$

or, in complex notation,

$$\frac{d^2z}{dt^2} = -\omega^2\alpha e^{i\omega t} = \omega^2\alpha e^{i(\omega t + \pi)}, \tag{1.21}$$

The acceleration is proportional to the displacement with the negative proportionality constant $-\omega^2$. We can say that its phase is at π radians to the displacement, or in *phase opposition* with it.

In the system we are considering, the total energy, meaning the sum of both the kinetic and potential ones, is conserved. Let us check that. The total energy is

$$U_{\text{tot}} = U_k(t) + U_p(t) = \frac{1}{2}m\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}kx^2 = \frac{1}{2}m\left[\left(\frac{dx}{dt}\right)^2 + \omega_0^2 x^2\right]$$

where $x(t)$ is given by Eq. (1.12), and we obtain

$$U_{\text{tot}} = \frac{1}{2}m\omega_0^2 A^2\left[\sin^2(\omega_0 t + \phi) + \cos^2(\omega_0 t + \phi)\right] = \frac{1}{2}m\omega_0^2 A^2. \tag{1.22}$$

We see that neither the kinetic energy nor the potential energy are constant in time, but rather, they vary as $\sin^2(\omega_0 t + \phi)$ and $\cos^2(\omega_0 t + \phi)$, respectively, but their sum, the total energy, is, as we expected, constant. Also notice that the kinetic, potential and total energies are all proportional to the square of the amplitude on one side and to the square of the angular frequency on the other.

The mean value of a quantity in a given time interval is the integral of that quantity on that interval, divided by the interval. One immediately calculates that the mean values of both functions $\cos^2$ and $\sin^2$ over a period are equal to 1/2. Consequently, the mean values of both potential and kinetic energy over a period are one half of the total energy.

$$\langle U_k \rangle = \langle U_p \rangle = \frac{1}{4} m\omega_0^2 A^2 = \frac{1}{2} U_{\text{tot}}. \tag{1.23}$$

Completely similar arguments hold for the above-considered oscillating circuit in Fig. 1.1. We choose, as a coordinate defining the status of the system, the charge $Q(t)$ of the capacitor. Its time derivative is the current intensity $I(t)$. Let V_C be the emf of the capacitor and $V_L = -LdI/dt$ the emf between the extremes of the inductor. These extremes are directly connected to the plates of the capacitor, and we must have $V_C = V_L$. Namely, it is

$$-L\frac{dI}{dt} = \frac{Q}{C},$$

which, considering that $I = dQ/dt$, we can write as

$$\frac{d^2Q}{dt^2} + \frac{1}{LC} Q(t) = 0, \tag{1.24}$$

which is equal to Eq. (1.7) with the proper angular frequency

$$\omega_0 = \frac{1}{\sqrt{LC}}. \tag{1.25}$$

The solution to the equation is, consequently,

$$Q(t) = Q_0 \cos(\omega_0 t + \phi),$$

where Q_0 and ϕ are integration constants defined by the initial conditions.

We can check that energy is conserved in this case as well. In this case, the total energy is the sum of the energy of the condenser, which is the energy of the electric field, and the energy of the inductor, which is the energy of the magnetic field. Indeed, we have

$$\begin{aligned} U &= \frac{1}{2} LI^2 + \frac{1}{2}\frac{Q^2}{C} = \frac{L}{2} Q_0^2 \omega_0^2 \sin^2(\omega_0 t + \phi) + \frac{1}{2C} Q_0^2 \cos^2(\omega_0 t + \phi) \\ &= \frac{L}{2} Q_0^2 \omega_0^2 \left[\sin^2(\omega_0 t + \phi) + \cos^2(\omega_0 t + \phi)\right] = \frac{L}{2} Q_0^2 \omega_0^2. \end{aligned}$$

1.2 Damped Oscillations

In the previous section, we neglected the dissipative forces, or effects, which, however, are always present. The dissipative forces on a mechanical system may be independent of velocity, as is the case with friction, or dependent on it, as is the viscous drag. The dependence on velocity of the viscous drag may be quite complicated, as we saw in Chap. 1 of the 2nd volume of this course. We shall limit the discussion to the cases in which the resistive force, or more generally action, is proportional to the velocity, or to the rate of change of the coordinate. Consider, for example, a pendulum oscillating in air. The drag force is an increasing function of the velocity. For small velocities, we can expand its expression in series. The term of zero order is zero because the viscous force is zero for zero velocity. We stop the expansion at the first non-zero term, which is proportional and opposite to the velocity. Similarly, in an oscillating circuit, a resistivity is always present. The potential drop across a resistor is proportional to the current intensity, which is the time derivative of the charge of the capacitor. Under these conditions, the (macroscopic) energy of the system is not conserved and the free oscillations are damped.

Let us take back the mechanical oscillator of Fig. 1.2 and include a drag force. In Fig. 1.6, we have schematized the drag with a damper, in which a piston linked to the block of mass m moves in a cylinder full of a gas. Changing the pressure of the gas, we can change the resistive force. Under the assumptions we have made, the viscous drag is proportional to the velocity in magnitude and opposite to it in direction. It is convenient to write the resistive force in the form

$$F_r = -m\gamma \frac{dx}{dt} \tag{1.26}$$

where γ is a constant, representing the drag per unit velocity and per unit mass. We neglect the friction between the support plane and the block. The second Newton law gives us

$$m\frac{d^2x}{dt^2} = -m\gamma \frac{dx}{dt} - kx, \tag{1.27}$$

which we write, dividing by m and taking all the terms to the left-hand side, in the canonical form

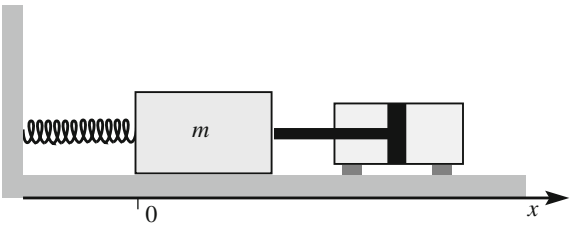


Fig. 1.6 A mechanical damped oscillator

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \omega_0^2 x = 0, \tag{1.28}$$

where ω_0^2 is, as always, the restoring force per unit displacement and per unit mass

$$\omega_0^2 = k/m. \tag{1.29}$$

Notice that both constants ω_0 and γ have the dimension of the inverse of a time. The inverse of γ, namely

$$\tau = 1/\gamma \tag{1.30}$$

is the time that characterizes the damping, as we shall now see.

The solution to the differential Eq. (1.28) is given by calculus. The rule for finding it is as follows. First, we write the algebraic equation obtained by substituting powers of the variable equal to the degree of the derivative into the differential equation. In our case, it is

$$r^2 + \gamma r + \omega_0^2 = 0. \tag{1.31}$$

Then, we solve it. The two roots are

$$r_{1,2} = -\frac{\gamma}{2} \pm \sqrt{\left(\frac{\gamma}{2}\right)^2 - \omega_0^2}. \tag{1.32}$$

The general solution to the differential equation is

$$x(t) = C_1 \mathrm{e}^{r_1 t} + C_2 \mathrm{e}^{r_2 t} \tag{1.33}$$

where C_1 and C_2 are integration constants that must be determined from the initial conditions.

In this volume, we deal with oscillations, and consequently, we shall consider only the case of small damping in which $\gamma/2 < \omega_0$ and the two roots are real and different. Equation (1.33) can then be written as

$$x = C_1 \mathrm{e}^{-(\gamma/2)t + i\omega_1 t} + C_2 \mathrm{e}^{-(\gamma/2)t - i\omega_1 t}$$

where

$$\omega_1 = \sqrt{\omega_0^2 - \left(\frac{\gamma}{2}\right)^2}. \tag{1.34}$$

We can now choose two different integration constants as $a = C_1 + C_2$ and $b = i(C_1 - C_2)$ and obtain the solution in the form

$$x(t) = \mathrm{e}^{-(\gamma/2)t}(a\,\cos\omega_1 t + b\,\sin\omega_1 t). \tag{1.35}$$

For damping tending to zero ($\gamma \to 0$), the equation of motion becomes Eq. (1.8), as we expected. Equation (1.35) can be written in a form analogous to Eq. (1.9),

$$x(t) = A\mathrm{e}^{-(\gamma/2)t}(\cos\omega_1 t + \phi) = A\mathrm{e}^{-t/(2\tau)}(\cos\omega_1 t + \phi), \tag{1.36}$$

where the integration constants are now A and ϕ.

In complex notation, Eq. (1.36) becomes

$$x(t) = \mathrm{Re}\left[A\mathrm{e}^{i\phi}\mathrm{e}^{i(\omega_1 + i\gamma/2)t}\right]. \tag{1.37}$$

The motion is an oscillation similar to the harmonic motion, with an amplitude, $A\mathrm{e}^{-t/(2\tau)}$, which is not constant but rather decreases exponentially in time with a decay time of 2τ. The oscillations are damped. Note that the oscillation frequency is modified by damping, namely it is reduced. As Eq. (1.34) shows, it is $\omega_1 < \omega_0$. This is the consequence of the fact that the drag force tends to reduce the velocity. Note that for very weak damping ($\gamma \ll \omega_0$), ω_0 differs from ω_1 by infinitesimals of the second order of γ/ω_0.

A weakly damped motion, namely with $\gamma \ll \omega_1$, is shown in Fig. 1.7. The oscillation amplitudes diminish gradually over a time long compared to the period. As a matter of fact, rigorously speaking, the motion is not periodic, because the displacement after every oscillation is somewhat smaller than that before it. However, if the damping is small, we can still identify a period

$$T = 2\pi/\omega_1. \tag{1.38}$$

Under a weak damping condition, namely if $\gamma \ll \omega_1$ or, equivalently, $\tau \gg T$, the decay time is much longer than the period. The oscillation amplitude does not vary too much in a period. Consequently, if we want to calculate the mean energy $\langle U \rangle$ in a period, we can consider the factor $e^{-(\gamma/2)t}$ to be constant and take it out of the integral. We get

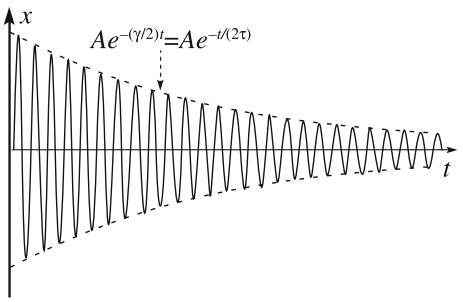


Fig. 1.7 A damped oscillation

$$\langle U\rangle = \langle U_0\rangle e^{-\gamma t}, \tag{1.39}$$

Where $\langle U_0\rangle$ is its initial value. The constant γ has two physical meanings. On one side, as we already saw, γ is *the drag force per unit velocity and unit mass*, while on the other, it is *the fraction of the average energy stored in the oscillator* ($\langle U\rangle$) *lost per unit time*. Indeed, the latter quantity is

$$-\frac{d\langle U\rangle}{dt}\frac{1}{\langle U\rangle} = \gamma. \tag{1.40}$$

Its reciprocal $1/\gamma$, which is the characteristic time τ of Eq. (1.30), is called the *decay time* of the oscillator. This is the time interval in which the energy in the oscillator decreases by a factor of $1/e$.

Question Q 1.1. Consider a damped oscillator with $\omega_0 = 10^3$ s^{-1} and $\gamma = 10$ s^{-1}. How much is ω_1? □

In the analysis of the previous section of an oscillating circuit, we neglected the dissipative effects. These, however, are always present. We now represent them schematically with a resistor in series, which is R in Fig. 1.8.

The equation of the circuit is

$$L\frac{dI}{dt} + RI(t) + \frac{1}{C}Q(t) = 0,$$

which we can write in the form

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{1}{LC}Q(t) = 0. \tag{1.41}$$

This equation is identical to Eq. (1.28), with Q in place of x, $1/\sqrt{LC}$ for the proper angular frequency ω_0 and R/L for the damping constant γ. The condition of small damping is then $R/2 < \sqrt{L/C}$. The corresponding solution is the quasi-harmonic oscillation with exponentially decreasing amplitude with time, namely

$$Q(t) = Q_0 \mathrm{e}^{-\frac{R}{2L}t}(\cos\omega_1 t + \phi), \tag{1.42}$$

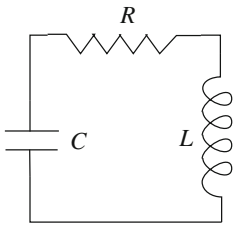


Fig. 1.8 A damped oscillating circuit

where

$$\omega_1 = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}, \tag{1.43}$$

An observation on the exponential function. The amplitude of a damped oscillation in Eq. (1.36) and the energy of the damped oscillator, Eq. (1.39), are examples of physical quantities decreasing exponentially over time. This behavior is often encountered in physics. We make here a simple but important observation. Consider the function

$$f(t) \propto f_0 \mathrm{e}^{-t/\tau}$$

and the ratio between its two values in two different instants t_1 and t_2 ($t_1 < t_2$). We immediately see that this ratio depends only on the interval $t_2 - t_1$ and not separately on the two times or the constant (the initial value) f_0. Indeed,

$$f(t_2)/f(t_2) = e^{-(t_2 - t_1)/\tau}. \tag{1.44}$$

In particular, the function diminishes by a factor of $1/e$ in every time interval $t_2 - t_1 = \tau$ and not only in the initial one.

1.3 Forced Oscillations

Consider again the damped oscillator of the previous section and apply to the body a force in the direction of the x-axis that oscillates as a circular function of time with angular frequency ω and amplitude F_0. The component of the force on the x-axis (its magnitude or its opposite, depending on the direction relative to x) is given by

$$F(t) = F_0 \cos(\omega t), \tag{1.45}$$

where we have chosen the origin of times as the instant in which the force is zero; its initial phase is then null. The second Newton law gives us

$$m\frac{d^2x}{dt^2} = F_0 \cos \omega t - \gamma m \frac{dx}{dt} - kx, \tag{1.46}$$

which we write in the form

$$\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x = \frac{F_0}{m}\cos\omega t. \tag{1.47}$$

The left-hand side of this equation is equal to the left-hand side of the damped oscillation Eq. (1.28). But the right-hand side, which is zero for the latter, is now proportional to the external force. Equation (1.47) is a non-homogeneous differential equation and Eq. (1.28) is its *associated* homogeneous differential equation. A mathematical theorem states that the general solution to the former is the sum of the general solution to the associated homogeneous equation, which we already know, and any particular solution to the non-homogeneous one. The easiest way to find the latter is to consider the same equation of the complex variable $z(t) = x(t) + iy(t)$. We search for a solution to the differential equation, of which (1.47) is the real part, namely

$$\frac{d^2z}{dt^2} + \gamma\frac{dz}{dt} + \omega_0^2 z(t) = \frac{F_0}{m}e^{i\omega t}. \tag{1.48}$$

If we think that the motion of the system, after a sufficiently long period during which the force has acted, should be an oscillation at the frequency of the force, we can search for a solution of the type

$$z(t) = z_0 e^{i\omega t}. \tag{1.49}$$

Let us try this in (1.48). We have

$$-\omega^2 z_0 e^{i\omega t} + i\gamma\omega z_0 e^{i\omega t} + \omega_0^2 z_0 e^{i\omega t} = \frac{F_0}{m}e^{i\omega t},$$

which must be satisfied in every instant of time. And so it is, because all the terms depend on time by the same factor. Hence, Eq. (1.49) is a solution, provided that

$$-\omega^2 z_0 + i\gamma\omega z_0 + \omega_0^2 z_0 = \frac{F_0}{m}, \tag{1.50}$$

which is an algebraic equation. The unknown, which is the parameter we must find to obtain the solution, is the complex quantity z_0. This is immediately found to be

$$z_0 = \frac{F_0/m}{\omega_0^2 - \omega^2 + i\gamma\omega}. \tag{1.51}$$

We see that the solution is completely determined by the characteristics of the oscillator, ω_0 and γ, and of the applied force, F_0 and ω. It does not depend on the initial conditions.

The particular solution to Eq. (1.48) is then

$$z(t) = \frac{F_0/m}{\omega_0^2 - \omega^2 + i\gamma\omega} e^{i\omega t}. \tag{1.52}$$

It is convenient to write z_0 in terms of its modulus B and its argument $-\delta$

$$z_0 = Be^{-i\delta}. \tag{1.53}$$

We recall that the modulus of a ratio of two complex numbers is the ratio of the modulus of the nominator and the modulus of the denominator, hence we have

$$B = \frac{F_0/m}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}}. \tag{1.54}$$

We recall that the argument of a ratio is the difference between the arguments of the nominator (which is 0, in this case) and the argument of the denominator. We are interested in its opposite, which is

$$\delta = \arctan \frac{\gamma\omega}{\omega_0^2 - \omega^2}. \tag{1.55}$$

The particular solution to Eq. (1.48) is then

$$z(t) = Be^{i(\omega t - \delta)} \tag{1.56}$$

and, taking the real part, the particular solution to Eq. (1.47) is

$$x(t) = B\cos(\omega t - \delta). \tag{1.57}$$

Finally, the general solution to Eq. (1.47) is

$$x(t) = Ae^{-(\gamma/2)t}(\cos\omega_1 t + \phi) + B\cos(\omega t - \delta), \tag{1.58}$$

where ω_1 is given by Eq. (1.34). Let us now discuss the motion we have found. It is the sum of two terms. The first one represents a damped oscillation at the angular frequency ω_1 that is proper for the oscillator. The constants A and ϕ depend on the conditions under which the motion started and appear in the first term alone. The second term depends on the applied force. The motion is quite complicated under these conditions. However, the amplitude of the first term decreases by a factor of e in every time interval $2/\gamma$. After a few such intervals, the first term has practically disappeared. Once this transient phase has gone, the regime of the motion is *stationary*. The *stationary oscillation* is described by the particular solution in Eq. (1.57), which is called a *stationary solution*. We rewrite it as

$$x_s(t) = B\cos(\omega t - \delta). \tag{1.59}$$

We repeat that the stationary motion is a harmonic oscillation at the angular frequency of the force, not at the proper frequency of the oscillator. However, both the amplitude B and the phase δ, which is not the initial phase but rather the phase delay of the displacement x relative to the instantaneous phase of the force, do depend on the characteristics of both the oscillator and the force, as in Eqs. (1.54) and (1.55). We shall come back to that in the next section. Before doing that, we must introduce a few more interesting physical quantities.

Absorptive amplitude and *elastic amplitude*. The stationary solution, Eq. (1.59), can be written, in an equivalent form, as the sum of a term in phase with the external force (which we call *elastic*) and a term in quadrature, namely with a 90° phase difference (which we call *absorptive*), namely as

$$x_s(t) = A_{el}\cos\omega t + A_{abs}\sin\omega t, \tag{1.60}$$

where

$$\begin{aligned} A_{el} &= B\cos\delta = \frac{F_0}{m}\frac{\omega_0^2 - \omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \\ A_{abs} &= B\sin\delta = \frac{F_0}{\gamma\omega m}\frac{\gamma^2\omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}. \end{aligned} \tag{1.61}$$

The reason for the use of the adjective "absorptive" is understood by calculating the mean power absorbed by the oscillator from the work of the external force, as we shall immediately do. The reason for the use of "elastic" is that this adjective often means "without absorption" in physics. Let us now express the instantaneous power, which is

$$P(t) = F(t)\frac{dx_s}{dt} = \omega F_0\cos\omega t[-A_{el}\sin\omega t + A_{abs}\cos\omega t].$$

We calculate the mean absorbed power in a period $\langle P\rangle$, remembering that the mean over a period of $\sin^2$ and $\cos^2$ is 1/2 and that of sin times cos is zero. Hence, we have

$$\langle P\rangle = \frac{1}{2}F_0\omega A_{abs} = \frac{F_0^2}{2m}\frac{\gamma\omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}. \tag{1.62}$$

We see that the mean absorbed power is proportional to the absorptive amplitude. This happens because the average power delivered by the force is different from zero only if the force has the same phase as the velocity, namely if it is in quadrature with the displacement.

Another important quantity is the *energy stored* in the oscillator. This is

$$U(t) = \frac{1}{2}\omega_0^2 m x_s^2(t) + \frac{1}{2}m\left(\frac{dx_s}{dt}\right)^2 = \frac{1}{2}mB^2\left[\omega_0^2\cos^2(\omega t - \delta) + \omega^2\sin^2(\omega t - \delta)\right].$$

We see that the stored energy is not constant, but rather it varies periodically over time. In a free oscillator, the mean kinetic and potential energies are equal in every instant, while in a forced one, they are different, in general, being equal only when $\omega = \omega_0$, namely at resonance. This is a consequence of the fact that the power $P(t)$ delivered by the external force is not exactly balanced in each instant of time by the power dissipated by the drag force. The instant-by-instant balance exists only in resonance. Indeed, in resonance, the elastic amplitude vanishes identically, and not only at its average value.

Let us finally express the mean value over a period of the stored energy, which is

$$\langle U\rangle = \frac{1}{2}mB^2\frac{\omega_0^2 + \omega^2}{2} = \frac{F_0^2}{2m}\frac{\left(\omega_0^2 + \omega^2\right)/2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}. \tag{1.63}$$

1.4 Resonance Curves

We now observe that the expressions of the mean absorbed power in Eq. (1.62), the mean stored energy in Eq. (1.63), the absorptive amplitude in Eq. (1.61) and the square of the oscillation amplitude in Eq. (1.54) have the same denominator and different numerators. The most interesting phenomena are when the angular frequency of the external force ω is near the proper frequency of the oscillator ω_0. Here, the denominator is very small, and consequently, the four mentioned quintiles exhibit very similar, if not identical, behaviors. Contrastingly, their behaviors are quite different far from ω_0. The behavior of the elastic amplitude is very different. It vanishes in resonance rather than having a maximum. We shall now discuss the four quantities and subsequently the elastic amplitude.

The dependence on ω of the four quantities is conveniently studied in terms of the dimensionless function

$$R(\omega) = \frac{\gamma^2\omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}, \tag{1.64}$$

which we shall call the *response function* of the oscillator. As Fig. 1.9 shows, the function is a bell-shaped curve with its maximum in $\omega = \omega_0$, symmetric about the maximum. This is a *resonance curve*.

Note that we have defined the function in order to have its maximum be equal to one. Contrastingly, as we shall soon see, the maxima of the curves representing the physical quantities depend on the damping parameter γ.

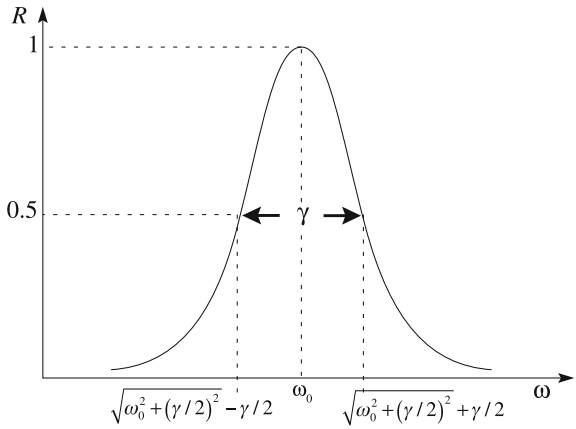


Fig. 1.9 The resonance curve

The resonance curve is narrower for smaller damping. As a matter of fact, the abscissas at which the function is equal to 1/2 are $\sqrt{\omega_0^2+(\gamma/2)^2}\pm\gamma/2$. Consequently, the full width at half maximum (FWHM, for short), which we call $\Delta\omega_{\text{ris}}$, is equal to γ, namely

$$\Delta\omega_{\text{ris}}=\gamma. \tag{1.65}$$

Question Q 1.2. Calculate the values of ω at which the resonance curve is one half of its maximum. □

Let us now consider each of the four resonating quantities. The mean absorbed power is

$$\langle P\rangle=\frac{F_0^2}{2m\gamma}R(\omega). \tag{1.66}$$

Hence, its frequency dependence is exactly the same as for the response function, times, however, a constant factor depending on the damping γ. The power is a maximum when the angular frequency of the force is equal to the proper frequency of the oscillator ($\omega=\omega_0$). The maximum is higher the smaller the damping due to the factor $1/\gamma$ multiplying $R(\omega)$. Hence, the maximum of $\langle P\rangle$ increases without limits when γ decreases. The FWHM is exactly $\Delta\omega_{\text{ris}}$.

The oscillation amplitude squared is

$$B^2=\left(\frac{F_0}{m\gamma\omega}\right)^2R(\omega) \tag{1.67}$$

and is shown in Fig. 1.10.

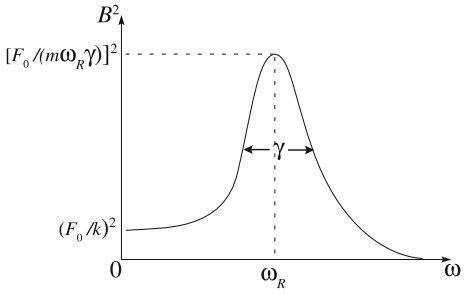


Fig. 1.10 Frequency dependence of the oscillation amplitude square

The frequency dependence differs from that of $R(\omega)$ by the factor $1/\omega^2$. Its effects are not very large near resonance, if the damping is small. The maximum of B^2 is at a value smaller than ω_0, namely

$$\omega_R = \sqrt{\omega_0^2 - \gamma^2/2}. \tag{1.68}$$

This value is, at any rate, very close to ω_0^2 if $\gamma/\omega_0 << 1$.

QUESTION Q 1.3. Consider a damped oscillator with $\omega_0 = 10^3$ s^{-1} and $\gamma = 10$ s^{-1}. How much is ω_R? □

The function of Eq. (1.67) contains the factor $1/\gamma^2$ that makes B^2 increase indefinitely when γ decreases. The width of the curve is approximately, but not exactly, $\Delta\omega_{\text{ris}}$.

Let us now think for a moment about the same oscillator when it is free. The decay time of the energy stored in the oscillator τ is equal to $1/\gamma$ for Eq. (1.30), but is equal to $1/\Delta\omega_{\text{ris}}$ as well, for Eq. (1.65). We reach the conclusion that *the width in angular frequency of the resonance curve for forced oscillations is inversely proportional to the decay time of the free oscillations*, namely that

$$\Delta\omega_{ris} \cdot \tau = 1. \tag{1.69}$$

This relation is extremely important, both in classical and quantum physics. We shall use it in the subsequent chapters.

Arguments similar to those we made for the mean power and the oscillation amplitude square hold for the mean stored energy and the absorptive amplitude.

The *Lorentzian curve*, named after Hendrik Antoon Lorentz (The Netherlands, 1853–1928), is a useful simplification of the expression of the resonance curve, valid if the damping is sufficiently small and in the frequency region close to the resonance. The factor that varies most rapidly in the function $R(\omega)$ is its denominator.

We write the denominator as $\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2 = (\omega_0 - \omega)^2(\omega_0 + \omega)^2 + \gamma^2\omega^2$ and note that, on the right-hand side, the term varying most rapidly is $(\omega_0 - \omega)^2$. Near to ω_0, and if $\gamma/\omega_0 \ll 1$, we can simplify the function by putting

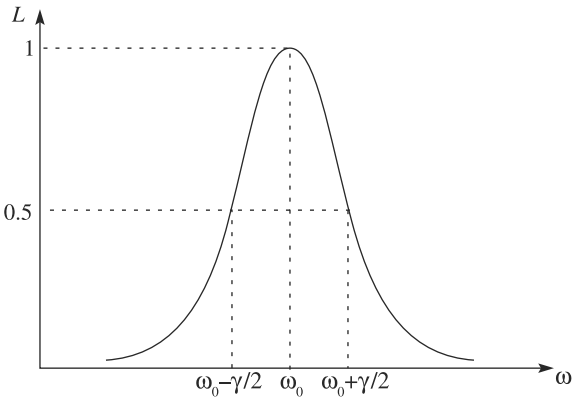


Fig. 1.11 The Lorentzian curve

ω_0 in place of ω in all the terms in the numerator and denominator except in $(\omega_0 - \omega)^2$.

In this approximation, the resonance curve $R(\omega)$ becomes

$$L(\omega) = \frac{(\gamma/2)^2}{(\omega_0 - \omega)^2 + (\gamma/2)^2}, \tag{1.70}$$

which is shown in Fig. 1.11. The curve has its maximum at ω_0 and FWHM equal to γ. Indeed, one can easily calculate that $L(\omega \pm \gamma/2) = 1/2$. The curve very often appears in atomic physics with the name Lorentzian and in nuclear and sub-nuclear physics with the name of the Breit-Wigner curve.

Let us now study the phase delay δ between the displacement x and the applied force F, as given by Eq. (1.55). This function is shown in Fig. 1.12, for two different values of the damping coefficient γ, one smaller (continuous curve), one

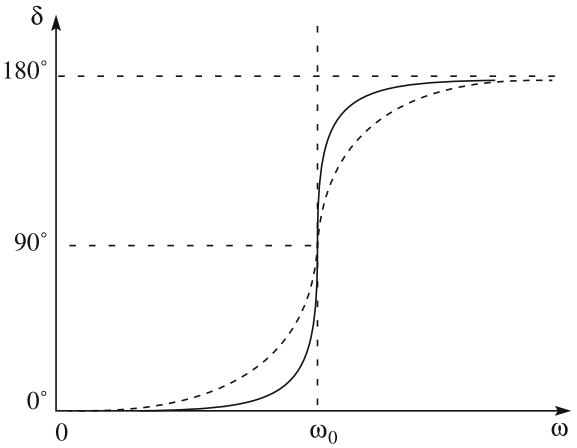


Fig. 1.12 The phase delay of the displacement relative to the external force

larger (dotted curve). Equation (1.55) tells us that when $\omega \ll \omega_0$, the delay of the displacement is very small, namely $\delta \cong 0$ (see Fig. 1.12). This is easily understood: at low frequencies, the accelerations of the oscillator are very small and the main effect of the external force is to balance the restoring force $-kx$. Consequently, the force is in phase with x.

Contrastingly, when $\omega \gg \omega_0$, the phase delay of the displacement is $\delta \cong \pi$. This means that when the spring pulls to the right, the oscillator mass is on the left of the equilibrium position and vice versa. Again, the interpretation is not difficult. At high frequencies, the accelerations are large. Hence, the term is $-md^2x/dt^2$ (due to inertia, we can say) and dominates over the elastic force term. The external force mainly needs to balance that term. Consequently, it is in phase with acceleration, which is in phase opposition with the displacement.

The above arguments help us to understand the behavior of B^2, and of the amplitude B, far from the resonance. For $\omega \to 0$, Eq. (1.54) gives us $B \to F_0/k$. This means that, at small frequencies of the applied force, the oscillation amplitude is (almost) independent of the frequency. It is independent of m and γ as well. It depends only on the spring constant k (strength of the force apart). For very high frequencies, on the other side, $B \cong F_0/m\omega^2$. Hence, the amplitude, which vanishes for $\omega \to \infty$, depends on the mass but not on the spring constant. As a matter of fact, if the spring was not present, the change would be small, its force being very small compared to $-md^2x/dt^2$.

Coming back to the phase delay, we note that the transition between the region of small values $\delta \cong 0$ and values $\delta \cong \pi$ takes place around the resonance in an interval on the order of γ. The smaller the damping, the sharper the transition. At resonance, the displacement is in quadrature ($\delta = \pi/2$) with the applied force. Considering that the velocity is in quadrature with the displacement, one sees that, at resonance, the force is in phase with the velocity. This is why the absorbed power is a maximum here.

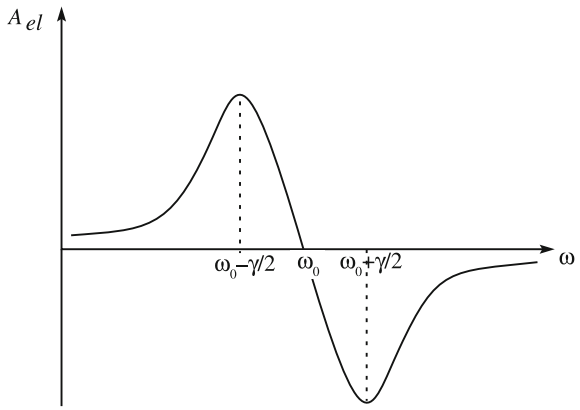


Fig. 1.13 The elastic amplitude

We still have to analyze the behavior of the elastic amplitude. This is shown in Fig. 1.13. The amplitude goes through zero in resonance due to the factor $\omega_0^2 - \omega^2$. It has a maximum and a minimum at the abscissas $\omega_0 \pm \gamma/2$ symmetrically about the resonance. Their values $\pm F_0/(2m\omega_0\gamma)$ are equal and opposite, and are one half of the maximum of the absorptive amplitude. The elastic amplitude, which is considerably smaller than the absorptive amplitude in the region of the resonance, becomes the dominant term far from it. We shall see the importance of both amplitudes in the study of the refractive index of light in Chap. 4.

1.5 Resonance in Nature and in Technology

Resonance phenomena are ubiquitous in physics and have an enormous number of technological applications. Indeed, every system oscillates harmonically, or almost so, when abandoned close to a state of stable equilibrium. These oscillations have a definite frequency, depending on the structure of the system. If an external force, with a sinusoidal dependence on time with angular frequency ω, acts on the system, the motion of the system evolves toward a stationary regime in which it oscillates at the frequency of the external force. The amplitude of the stationary oscillations is a maximum when ω is close to the proper angular frequency ω_0 of the system. The maximum is narrower the smaller the damping γ. To be precise, the systems with one degree of freedom, which we have discussed in this chapter, have a single proper frequency. Systems with several degrees of freedom have several proper frequencies, as we shall study in the next chapter.

In this section, we shall discuss a few examples of resonance. Before doing that, let us explain the origin of the name.

"*Resonance*" comes from the Latin verb "*resonare*", meaning to resound or to sound together. As we are dealing with sound, let us consider the following experiment. We tune two strings of the same musical instrument to the same note; plucking one of them, we now hear the note. If we stop the string with a finger, we still hear the same note, at a volume as loud as when we first heard the sound of the first string. Why? Because the second string, which has not been touched, has entered into oscillations, being "in resonance" with the one that was touched.

The first string is the source of a sound wave (see Sect. 3.4) that has the same frequency as its vibrations. The wave propagates in space, hitting, in particular, the untouched string and exerting a periodic force on it. The force is weak, but it is at the exact proper frequency of that string. Consequently, the efficiency of energy transfer from the sound to the string is a maximum. Let us now repeat the observations after having detuned the second string a bit relative to the first. We still hear the untouched string vibrating, if the detuning was not too great, but with a much smaller amplitude. As a matter of fact, this amplitude decreases with increased detuning, namely for increasing differences between ω and ω_0.

Galileo Galilei, who was a good musician, describes the phenomenon beautifully in his "*Dialogues and mathematical demonstrations concerning two new sciences*" (English translation by Henry Crew and Alfonso de Salvio, McMillan 1914), in which Salviati, who represents Galilei himself, says that

> to explain the wonderful phenomenon of the strings of the cittern [*a musical instrument*] or of the spinet [*another musical instrument*], namely, the fact that a vibrating string will set another string in motion and cause it to sound not only when the latter is in unison but even when it differs from the former by an octave [*namely with proper frequency twice or one half*] or a fifth [*frequency ratio 2/3*] [*because a string has a series of proper frequencies, as we shall see in the next chapter*]. A string which has been struck begins to vibrate and continues the motion as long as one hears the sound; these vibrations cause the immediately surrounding air to vibrate and quiver; then these ripples in the air expand far into space and strike not only all the strings of the same instrument but even those of neighboring instruments. Since that string which is tuned to unison with the one plucked is capable of vibrating with the same frequency, it acquires, at the first impulse, a slight oscillation; after receiving two, three, twenty, or more impulses, delivered at proper intervals, it finally accumulates a vibratory motion equal to that of the plucked string, as is clearly shown by equality of amplitude in their vibrations. This undulation expands through the air and sets into vibration not only strings, but also any other body which happens to have the same period as that of the plucked string. Accordingly, if we attach to the side of an instrument small pieces of bristle or other flexible bodies, we shall observe that, when a spinet is sounded, only those pieces respond that have the same period as the string which has been struck; the remaining pieces do not vibrate in response to this string, nor do the former pieces respond to any other tone.

Resonance phenomena exist in mechanical systems, including the oceans and astronomical bodies, in electromagnetism, in acoustics, and in molecules, atoms, nuclei and subnuclear particles. The phenomenon is employed across a wide spectrum of technologies, ranging from telecommunications and information technology to medical applications, like electron spin resonance and nuclear magnetic resonance, to laboratory instrumentation, as in the resonant cavities used to accelerate particle beams, etc.

As an initial example, let us go back the resonant circuit of Sect. 1.2, where we discussed its free damped oscillations. To obtain a forced oscillating circuit now,

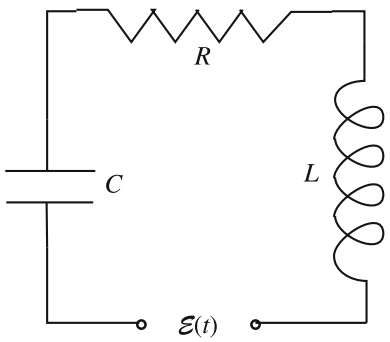


Fig. 1.14 A forced resonating circuit

we insert an emf sinusoidal generator, as in Fig. 1.14. Let $E(t) = E_0 \cos \omega t$ be its emf and let its frequency be adjustable as we wish.

We find the differential equation governing the circuit by imposing that the sum of the electromotive forces along the circuit be zero, namely

$$L\frac{dI}{dt} + RI(t) + \frac{1}{C}Q(t) = E(t),$$

which we can write in the form

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{1}{LC}Q(t) = \frac{E_0}{L}\cos \omega t.$$

The equation is formally identical to Eq. (1.48), valid for the mechanical oscillator. We can then state that, in this case as well, a stationary solution exists that is a harmonic (sinusoidal) oscillation at the angular frequency ω of the external source. The closer ω is to the proper oscillation angular frequency of the circuit $\omega_0 = 1/\sqrt{LC}$, the larger the oscillation amplitude. This is negligibly different from the position of the maximum for very small damping. The width of the maximum is inversely proportional to the damping coefficient, namely R/L, while its height increases with decreasing R/L.

Resonating circuits of adjustable proper frequency are used selectively to detect and amplify electromagnetic signals of definite frequency. For example, the basic structure of a radio receiver includes a tunable resonating circuit. The receiver is always under the action of electromagnetic waves produced by a large number of natural and artificial sources. However, each radio station transmits on a certain characteristic frequency. To select a particular station, one changes the proper frequency of the receiver $\omega_0 = 1/\sqrt{LC}$ acting (with a knob) on a variable capacitor to make it equal to the frequency ω of the desired station (in practice, the electronics are more sophisticated). When $\omega_0 = \omega$, the (electric) oscillations of the receiver have a large amplitude and are detected. The signals from other sources, which are present, are out of resonance and do not excite oscillations of detectable amplitude, provided their frequency differs from ω_0 by about R/L. The R/L parameter, namely the width of the resonance curve, controls the selectivity of the receiver.

Resonances must be carefully considered in the design of mechanical systems, especially if they have parts in rapid rotation, like motors, pumps, turbines, etc., but also for civil structures, like bridges, harbors or tall buildings that may receive periodic solicitations. Consider, for example, the rotating part of a motor. Engineers accurately design the rotor so as to have the rotation axis be as coincident as possible with a principal axis and the center of mass on the axis as accurate as possible. However, even a small imperfection might be enough to induce destructive oscillations, especially for high rotation speeds and small damping. These effects must be accurately calculated in the design phases. A turbine or an ultracentrifuge, for example, might have to cross a number of resonances during the acceleration phase. The design must be such so as to guarantee that the process

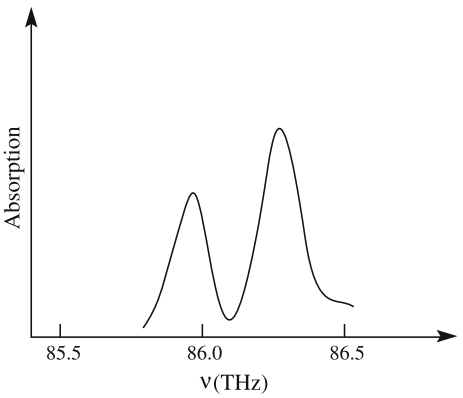


Fig. 1.15 Absorption probability for HCl molecules versus frequency

takes place without damage. This represents a full chapter of engineering, called the mechanics of vibrations.

The resonance phenomenon is also present in the molecular oscillators, at quite high frequencies, on the order of 10^{13} Hz (10 THz). These are the frequencies of the electromagnetic waves in the infrared. Imagine conducting the following experiment. We radiate a container with transparent walls containing an HCl gas with an infrared radiation, of which we can vary the frequency, and we measure the intensity of the radiation transmitted by the gas in correspondence. Taking the ratio between the transmitted and the incident intensities, we have the absorption probability of the gas as a function of frequency.

We obtain Fig. 1.15, in which, we must note, the abscissa is frequency, rather than angular frequency. It is a resonant curve, because in resonance, much more energy is transferred from the radiation to the molecular oscillators than at other frequencies. However, two peaks, not one, are observed. The reason for this is that Chlorine has two isotopes, ^{35}Cl and ^{37}Cl of atomic masses 35 and 37, respectively. The two proper frequencies squared, ω_0^2, are different, as the forces are equal and the masses different, in the two cases. To be thorough, in the spectrum, several doublets like the one in Fig. 1.15 are present. This is because quantum oscillators have several proper oscillation frequencies, rather than a single one.

To see the orders of magnitude, let us try to evaluate the "spring constant" k of a molecular oscillator. Let us consider, for example, a metal. If we press a block of that metal, it reacts with an elastic force. This force, which we can measure, because it is macroscopic, is ultimately a result of the fact that we are changing the distance between molecules. Let us consider the simple shape of a cylinder of section S and length l and apply a force in the direction of the axis. The spring constant is $k = ES/l$, where E is the Young modulus. A typical value for metals is $E \approx 2 \times 10^{11}$ N/m^2. At the molecular scale, we can evaluate the geometric factor S/l to be on the order of a molecular diameter, namely a few hundred picometers. If we take $S/l = 4 \times 10^{-10}$ m, we get $k \approx 80$ N/m. As for the mass, it should be the reduced mass of the system, but the latter is practically equal to the mass of the hydrogen atom, because that of Cl is

about 35 times larger. We then take $m_{\mathrm{H}} = 1.67 \times 10^{-27}$ kg. With these numbers, we obtain $\omega_0 = \sqrt{k/m_H} \cong 2 \times 10^{14}$ Hz. This angular frequency corresponds to $v_0 \cong 200/6.28 \cong 30$ THz, which is the correct order of magnitude.

Let us now consider a gas of monoatomic molecules. We can think of an atom as being a spherical distribution of negative charge (the electrons) with a positive equal and opposite point charge in the center. The centers of the negative and positive charges coincide. The system is globally neutral. As we have studied in the 3rd volume of this course, under the action of an electric field, the atom acquires an electric dipole moment proportional to the electric field intensity. We can think of the centers of the positive and negative charges as displacing one relative to the other, the former in the direction of the field, the latter in the opposite direction. The force resulting from the applied electric field is balanced at equilibrium by an internal restoring force, which is proportional to the deformation (in a first approximation). If the electric field varies periodically in time, the atom behaves like a forced oscillator, a fact that we shall consider often in subsequent sections. The resonance phenomenon appears in this case as well. The proper oscillation frequencies (they are several) are larger than the molecular frequencies considered above, typically by one or two orders of magnitude. If we conduct an experiment with a gas of atoms similar to the above-considered with molecules, we observe similar resonance maxima, but at much higher frequencies. We shall discuss these phenomena in Sect. 4.8, where we shall see that the elastic amplitude determines the frequency dependence of the refractive index. The curious shape of the curve of the elastic amplitude explains why, in a transparent medium, the speed of blue light is smaller than that of red light. This is the dispersion phenomenon of light, which is the physical explanation for the rainbow, to be discussed in Sect. 4.4 (and much more). For this reason, the curve of the elastic amplitude is also called a dispersion curve.

1.6 Superposition Principle

All the differential equations we have encountered in this chapter have had the common characteristic of being *linear* in their unknown that is the function $x(t)$. There are no terms proportional to powers of the function, like $x^2(t)$, or of its derivatives, like $(dx/dt)^2$, or products, like $x(t)dx/dt$, etc. An important aspect is that the solutions to linear equations obey the superposition principle. Let us demonstrate this, for the sake of being concrete, on Eq. (1.47), with a generally valid argument. We shall write this equation

$$m\left(\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x\right) = F(t). \tag{1.71}$$

as

$$m\left(\frac{d^2}{dt^2}+\gamma\frac{d}{dt}+\omega_0^2\right)x(t)=F(t).$$

We then formally define operator O, which acts on the function x as

$$O\equiv m\left(\frac{d^2}{dt^2}+\gamma\frac{d}{dt}+\omega_0^2\right). \tag{1.72}$$

In this notation, the differential equation becomes

$$O(x)=F(t). \tag{1.73}$$

This is obviously the exact same equation, written, we say, in operational form. We say that O is a linear operator. It is so because it satisfies the following properties:

$$O(x+y)=O(x)+O(y) \tag{1.74}$$

and, if a is an arbitrary constant,

$$O(ax)=aO(x). \tag{1.75}$$

The linear systems, namely the systems ruled by a linear differential equation, obey the superposition principle, which can be expressed with the following statements.

If the equation is homogeneous, $O(x) = 0$, any linear combination of solutions is a solution.

Indeed, let $x_1(t)$ and $x_2(t)$ be two solutions. Then, $O(x_1)=0$ and $O(x_2)=0$ and also $O(ax_1+bx_2)=O(ax_1)+O(bx_2)=aO(x_1)+bO(x_2)=0$.

If the equation is non-homogeneous, and if $x_1(t)$ is a solution when the known term is $F_1(t)$ and $x_2(t)$ is a solution when the known term is $F_2(t)$, then $x_1(t) + x_2(t)$ is a solution when the known term is $F_1(t) + F_2(t)$.

Indeed, $O(x_1+x_2)=O(x_1)+O(x_2)=F_1+F_2$.

Note that non-linear differential equations are usually extremely difficult, or even impossible, to solve analytically. Even if the natural systems are never exactly linear, we can often use a linear approximation. This is what we did in this chapter and shall do in the subsequent ones. Bear in mind, however, that the solutions will be valid within a certain degree of approximation.

Let us go back to Eq. (1.71). In Sect. 1.3, we found the solutions to this differential equation for a particular time dependence of the force, namely $F(t) = F_0 \cos \omega t$. Well, the equation being linear, we can apply the superposition principle and immediately find the solution for any $F(t)$ that can be expressed as a linear combination of cosine functions, namely in the form

$$F(t) = \sum_n A_n \cos(\omega_n t + \phi_n) \tag{1.76}$$

or even as

$$F(t) = \int A(\omega) \cos(\omega t + \phi(\omega)) d\omega. \tag{1.77}$$

We shall see in the next chapter that practically all the functions relevant in physics can be expressed as the sum of a series, as in Eq. (1.76), when they are periodic, or as an integral, as in Eq. (1.77). These statements are proved by theorems credited to Joseph Fourier (France, 1768–1830), and the above expressions are called a Fourier series and a Fourier transform, respectively.

In conclusion, the problem of the motion of a linear oscillator subject to an applied force, with arbitrary dependence on time, is solvable by expressing the force as a linear combination of sinusoidal forces of different frequencies. One shall find the motion of the system separately for each of these forces and then take their linear combination to have the motion under the given force.

Summary

In this chapter, we studied the following points.

1. A mechanical system performs harmonic (sinusoidal) oscillations around a stable equilibrium position if the restoring force is proportional to the displacement.
2. The angular frequency of the free oscillation is a property of the oscillator and is equal to the square root of the restoring force per unit mass and unit displacement. Contrastingly, the oscillation amplitude and initial phase depend on the initial conditions of the motion.
3. Velocity and acceleration of a harmonic motion are sinusoidal functions of time as well. They have the same frequency as the displacement, but different phases.
4. The oscillation frequency of the oscillations about a potential minimum is determined by the second space derivative of the potential.
5. If energy is conserved, the mean values of the potential and kinetic energies of an oscillator are equal (and equal to one half of the total energy).
6. An LC circuit is governed by the same differential equation as the mechanic oscillator. Consequently, its electrical oscillations are analogous.
8. A weakly damped oscillator performs quasi-harmonic oscillations of exponentially decreasing amplitude at a frequency slightly less than those in the absence of damping.
9. The damping parameter γ is the relative loss of stored energy per unit time and is equal to the drag force per unit speed and unit mass. This is a characteristic of the system. The time constant with which the stored energy decreases is equal to $1/\gamma$.

10. The stationary oscillation of a forced oscillator has the frequency of the force.
11. The displacement has nearly the same phase as the force at frequencies much smaller than the proper frequency, and is in phase opposition at a frequency much larger than that. At resonance, the force is in phase with the velocity.
12. In a weakly damped oscillation, several kinematical quantities have a maximum at or close to ω_0.
13. The width of the resonance curve of a forced oscillator is inversely proportional to the decay constant of the damped oscillations of the same oscillator when free.
14. An RLC circuit fed by a sinusoidal emf generator obeys the same differential equation as a mechanical forced oscillator. It oscillates electrically as a mechanical oscillator does mechanically.
15. The motion of a linear system under the action of several forces can be found by adding its motions under the action of each force separately.

Problems

1.1. Consider the oscillator in Fig. 1.2 with $m = 0.3$ kg and $k = 30$ N/m. Calculate the angular frequency, frequency and period.

1.2. Calculate amplitude, initial phase and complex amplitude of the oscillation given by $x = (10\,\text{mm})\cos\omega_0 t + (15\,\text{mm})\sin\omega_0 t$.

1.3. Show that the amplitude of a damped oscillator is halved in a time equal to $1.39/\gamma$. How much does the energy vary during this time?

1.4. The proper angular frequency of an oscillator is $\omega_0 = 300$ rad/s and it is $\omega_0/\gamma = 50$. Compare the values of ω_0, of the free oscillations angular frequency ω_1, and of the angular frequency at which the amplitude square is at a maximum.

1.5. We want to assemble a mechanical oscillator similar to that in Fig. 1.2. We have with us a mass and two identical springs. We connect to one side of the mass separately: (a) one spring, (b) two springs in series, (c) two springs in parallel, and then, (d) one spring on each side of the mass. What are the values of the proper frequencies in the different cases, relative to case (a)?

1.6. Consider a forced oscillator in stationary oscillation. Show that the mean energy over a period is mainly potential when $\omega \ll \omega_0$, mainly kinetic when $\omega \gg \omega_0$, and exactly half and half at $\omega = \omega_0$.

1.7. Consider a forced oscillator in stationary oscillation at $\omega = \omega_0$. Show that the mean absorbed power over a period is γ times the mean stored energy.

1.8. We want to assemble an oscillating circuit similar to that in Fig. 1.1, having two identical capacitors and two identical inductors for our use. We separately make circuits with: (a) a capacitor and an inductor, as in Fig. 1.1, (b) two capacitors in series and an inductor, (c) two capacitors in parallel and an inductor, (d) two inductors in series and a capacitor. Find the oscillation frequencies of cases (b), (c) and (d) relative to case (a).

1.9. We know the amplitudes of the displacement and of the velocity oscillations of a harmonic oscillator. What can we say about the angular frequency?

1.10. A damped oscillator performs 100 complete oscillations in 100 s. In the same time, its amplitude decreases by a factor of 2.718. How much is the damping constant γ? How much is the relative decrease in energy in a period $-\Delta\langle U\rangle/\langle U\rangle$?

1.11. Does the amplitude of the stationary oscillations of a mechanical forced oscillator at frequencies much smaller than resonance depend on the amplitude of the applied force? On the mass? On the spring constant? Answer the same questions at oscillation frequencies much larger than resonance.

1.12. After having hit a damped oscillator with a stroke, we observe its oscillations. We measure an amplitude decay time of 100 s. Find the FWHM of the resonance curve in angular frequency of the same oscillator when it is forced

1.13. The amplitude of the stationary oscillation of an oscillator forced by an initial force, with sinusoidal dependence on time, is 20 mm. When separately forced by a second sinusoidal force of the same period, the amplitude of the stationary oscillation is 40 mm. When both forces act simultaneously, the amplitude is 30 mm. What is the phase difference between the forces?

Chapter 2
Oscillations of Systems with Several Degrees of Freedom

Abstract In the first part of this chapter, we study oscillating systems with two (and subsequently with n) degrees of freedom. We learn the existence of particular motions, the normal modes, in which all parts of the system oscillate together harmonically. The number of modes, and the number of resonances, is equal to the number of degrees of freedom. We then study the modes of a vibrating sting. In the second part of the chapter, we study Fourier analysis, in regard to both periodic and non-periodic functions and for functions of both time and space.

The oscillators we studied in the previous chapter had one degree of freedom, namely their state was defined by a single variable. In the first part of this chapter, we study oscillating systems with two and then more degrees of freedom. We shall deal, as an example, with the system of two pendulums linked together by a spring. In general, the motions of the system are not harmonic oscillations, but can be quite complicated. We shall find, however, that special, very simple motions exist in which both pendulums, or, generally speaking, both parts of the system, oscillate in a harmonic motion with the same frequency and in the same initial phase. These stationary motions, called the normal modes, are two in number for a system with two degrees of freedom; indeed, in general, the number of modes is the same as the number of degrees of freedom. The oscillation frequencies of the modes, called proper frequencies, are characteristic of the system.

Subsequently, in Sect. 2.3, we shall study the vibrations of a continuous system through the example of an elastic string with fixed extremes. We shall see that normal modes also exist for continuous systems. In these motions, all the points of the system vibrate harmonically in phase with the same frequency. The number of modes is infinite, but numerable. Once again, the frequencies of the succession of modes are characteristic of the system. For an elastic string, the proper frequencies form the arithmetic succession, which is the succession of the natural numbers. The lowest frequency is called the fundamental, while those subsequent are its harmonics. The name is a consequence of the fact that a sound composed of simple sounds is pleasant if the frequencies of the components are in the ratio of small natural numbers.

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_2

The systems we study are linear, namely the differential equation ruling their motion is linear and the superposition principle holds. A consequence of this is that every motion of the system, as complicated as it can be, can always be expressed as a linear combination of its normal modes. The functions (which are cosines for an elastic string) giving the time dependence of the modes constitute, as we say, a complete ortho-normal system of functions. Every function representing a motion of the system can be developed as a linear combination of these functions. This property allows for important simplifications in the study of the vibrations.

In the second part of the chapter, we study the harmonic (Fourier) analysis. This is the chapter of mathematics that originates from the physical phenomena we have just mentioned. We shall give only the mathematical statements without rigor and without proof. We are interested in the way in which harmonic analysis is very useful in physics. We shall see in Sects. 2.4 and 2.5 how a periodic function of time can be expressed in its Fourier series, which is a linear combination of an infinite sequence of harmonic functions, whose frequencies are the integer multiples of the lowest of them.

In Sect. 2.6, we shall extend the result to non-periodic functions. For them, an integral over the angular frequency, called the Fourier transform, takes the place of the Fourier series. We shall discuss two relevant examples.

Finally, in Sect. 2.7, we shall consider the Fourier analysis for a function of a space coordinate, rather than one of time. The function might be, for example, the gray level of a picture. The problems are substantially equal from the mathematical point of view, but their physical aspects are different. We shall need these concepts when studying optical phenomena.

2.1 Free Oscillators with Several Degrees of Freedom

After having studied the oscillations of systems with one degree of freedom in the previous chapter, we study here oscillations of systems with a discrete number of degrees of freedom. The motions, or more generally the evolution, which we shall discuss will always be about a stable equilibrium configuration under the action of a restoring force, or, more generally, of an agent, of magnitude proportional to the displacement from equilibrium. In other words, we shall deal with *linear systems*.

We start with the free oscillations of systems with two degrees of freedom in the absence of dissipative agents. To be specific, we consider a system of two *coupled oscillators*, as shown in Fig. 2.1. We shall subsequently generalize our findings. The coupled oscillators in the example of Fig. 2.1 are two identical pendulums, a and b, having length l and mass m, joined by a spring, having a rest length equal to the distance between the equilibrium positions of the pendulums. In order to have the oscillators coupled only weakly, we must choose the spring constant k as being small enough to have an elastic force substantially smaller than the weight.

We shall consider motions of the pendulums in the direction joining their equilibrium positions. Let x_a and x_b be the coordinates of the pendulums *each*

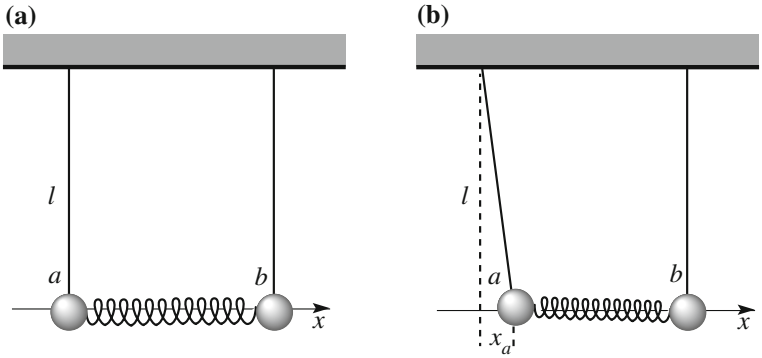


Fig. 2.1 Two coupled pendulums. **a** Equilibrium condition. **b** Initial condition

measured from their respective equilibrium position. We move pendulum a to the distance A from equilibrium, keeping b in its equilibrium position and letting both go with zero speeds. The initial conditions of the motions are the positions and velocities of the two pendulums, namely $x_a(0) = A$, $x_b(0) = 0$, $dx_a/dt(0) = 0$, $dx_b/dt(0) = 0$ (Fig. 2.1b).

We observe the following phenomenon. Initially, a oscillates with an amplitude equal to A. In so doing, it exerts on b a periodic force through the spring. The force is quite feeble (our having chosen to make k small) but it is at the resonance frequency of b (the pendulums being equal). As a consequence, b starts oscillating. The oscillations of b grow in amplitude with time, while those of a decrease, up to the point when a stops, or almost does, for a moment. The amplitude of b is now equal, or almost so, to A. The configuration is like that of the initial one, with the roles exchanged. The amplitude of b starts decreasing with that of a increasing until the system is back in its initial configuration, and so on. In this motion, energy goes back and forth from one pendulum to the other. This motion would continue forever in the absence of dissipative forces.

The motion we just described is more complex than the harmonic motion of a single pendulum. However, a system with two degrees of freedom can perform harmonic motions. More precisely, it is always possible to choose the *initial conditions* in such a way that *all the parts of the system* (namely the two pendulums, in this case) *perform a harmonic motion*. As a matter of fact, two different motions of this type exist. The initial conditions are shown in Fig. 2.2.

The first motion is obtained by taking both pendulums out of equilibrium by the same distance, say A, and letting them go at the same instant with zero speed. Namely, the initial conditions are $x_a(0) = A$, $x_b(0) = A$, $dx_a/dt(0) = 0$, $dx_b/dt(0) = 0$. Clearly, under these conditions, the spring is not deformed, namely its length initially and always is the rest length. Consequently, the spring does not exert any force (we assume its weight to be negligible) and each pendulum oscillates as if it was free (not coupled). Namely, we have

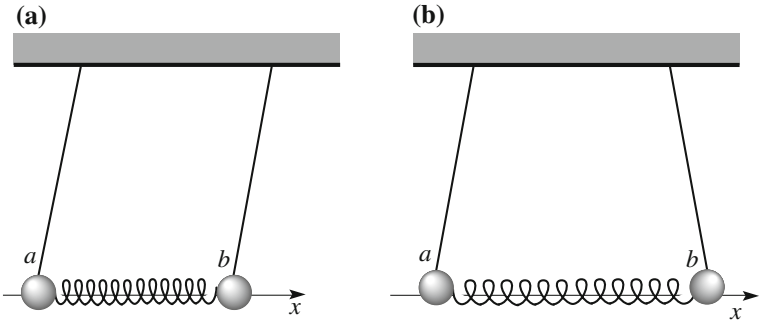


Fig. 2.2 The normal modes of two coupled pendulums

$$\begin{aligned} x_a(t) &= A\cos(\omega_1 t + \phi) \\ x_b(t) &= A\cos(\omega_1 t + \phi), \end{aligned} \tag{2.1}$$

where

$$\omega_1^2 = g/l,$$

which, we must remember, is the restoring force per unit displacement and per unit mass. Note that the amplitude A is arbitrary, as long as we do not stray from the condition of having the restoring force proportional to the displacement. What does matter for having a harmonic motion is the value of the *ratio* between the initial amplitudes of the two pendulums, which must be equal to 1. In Eq. (2.1), we have explicitly written the initial phase ϕ in the arguments of the cosines to be complete. In this case, it is $\phi = 0$. As a matter of fact, the initial phase is arbitrary too, but *must be the same* for both pendulums.

The second way to obtain harmonic oscillations for both pendulums (Fig. 2.2b) is to take them out of equilibrium by the same distance in opposite directions and let them go with zero initial speeds. The initial conditions are $x_a(0) = A$, $x_b(0) = -A$, $dx_a/dt(0) = 0$, $dx_b/dt(0) = 0$. In this case, the spring acts. The forces that it exerts on the pendulums, internal to the system, are equal and opposite of one another in any instant. The center of mass, which was initially at rest, remains at rest. Consequently, in every moment, we have $x_a(t) = -x_b(t)$.

Let us start analyzing the motion of a. Three forces act on the pendulum. Two forces are exactly the same as for a non-coupled pendulum, namely the weight and the tension of the suspension wire. Together, they give a contribution proportional to the displacement equal to $-x_a g/l$. The third force is due to the spring. Taking signs into account, the stretch of the spring is $x_a - x_b$. Consequently, the force is $-k(x_a - x_b) = -2kx_a$. We see that, in the present motion, it is proportional to the displacement. In conclusion, the resultant force on a is a restoring force proportional to its displacement from equilibrium. This is the condition necessary for

harmonic motion. The square of its angular frequency, which we call ω_2, is, as always, the restoring force per unit displacement per unit mass, namely

$$\omega_2^2 = \frac{g}{l} + 2\frac{k}{m}.$$

The situation for b is completely analogous. It moves under the action of its weight, the tension and the spring. The force of the latter is equal to and opposite of that of a and we can write it in terms of x_b only as $k(x_a - x_b) = -2kx_b$. The restoring force is the same as that for a when the displacement is the same. Consequently, being that the masses are also equal, b oscillates in harmonic motion with the same angular frequency ω_2. The motions of the two pendulums are harmonic motions *with the same angular frequency* and in phase opposition, or, to put it more properly, with *the same phase* and equal and opposite amplitudes, namely

$$\begin{aligned} x_a(t) &= A\cos(\omega_2 t + \phi) \\ x_b(t) &= -A\cos(\omega_2 t + \phi), \end{aligned} \tag{2.2}$$

In this case too, the initial amplitude is arbitrary. What matters is the *ratio* between the initial amplitudes, which must be equal to -1. Also, one of the initial phases is arbitrary, just as before. What matters is that the two initial phases must be equal (both $\phi = 0$, in this case).

The motions we described, namely those given by Eqs. (2.1) and (2.2), are called *normal modes* of the system. In general, the motion of a freely oscillating system with several degrees of freedom is said to be a normal mode when *all the parts of the system move in harmonic motion with the same frequency and the same phase.* "With the same phase" means that all the parts pass through their equilibrium positions in the same instant. We can also say that the normal modes are the *stationary motions* of the system, namely the motions whose characteristics are constant over time. Contrastingly, the characteristics of a generic motion, such as the one considered at the beginning of the section, evolve over time.

The system we considered is particularly simple and symmetric, the two pendulums being equal to one another. Its symmetry substantially allowed us to find the normal modes through intuition. To solve the problem in general, we need a bit of mathematics. We shall now show that the number of normal modes (or simply modes) of a system is equal to the number of its degrees of freedom. For linear systems, such as the ones we will consider, the superposition principle holds. Consequently, a linear combination of its normal modes is a possible motion of the system as well. In addition, we shall now show that *every motion* of the system can be expressed as a linear combination of its modes.

Let us start with the formal treatment of the example just discussed. The differential equations of its motion are

$$\begin{aligned}\frac{d^2x_a}{dt^2}+\omega_1^2x_a&=-\frac{k}{m}(x_a-x_b)\\ \frac{d^2x_b}{dt^2}+\omega_1^2x_b&=-\frac{k}{m}(x_b-x_a),\end{aligned}\tag{2.3}$$

where $\omega_1^2=g/l$ is a constant. We re-write the system in the form

$$\begin{aligned}\frac{d^2x_a}{dt^2}&=-\left(\omega_1^2+\frac{k}{m}\right)x_a+\frac{k}{m}x_b\\ \frac{d^2x_b}{dt^2}&=+\frac{k}{m}x_a-\left(\omega_1^2+\frac{k}{m}\right)x_b.\end{aligned}$$

This form can be made general. Indeed, we define a linear, not damped, freely oscillating system with two degree of freedom as a system obeying the following system of differential equations

$$\begin{aligned}\frac{d^2x_a}{dt^2}&=-a_{11}x_a-a_{12}x_b\\ \frac{d^2x_b}{dt^2}&=-a_{21}x_a-a_{22}x_b,\end{aligned}\tag{2.4}$$

where the coefficients a_{ij} are constants characteristic of the system (in the example, they are combinations of the masses, the elastic constant of the spring and the gravity acceleration), which are independent of the initial conditions.

Let us search for normal modes, namely motions in which all the parts of the system oscillate with the same frequency and in the same initial phase. The question is: does any value of ω exist such that the functions of time

$$\begin{aligned}x_a(t)&=A\cos(\omega t+\phi)\\ x_b(t)&=B\cos(\omega t+\phi),\end{aligned}\tag{2.5}$$

are solutions to the system in Eq. (2.4)?

To find an answer, we substitute these functions in Eq. (2.4), obtaining

$$\begin{aligned}\left(a_{11}-\omega^2\right)x_a+a_{12}x_b&=0\\ a_{21}x_a+\left(a_{22}-\omega^2\right)x_b&=0.\end{aligned}\tag{2.6}$$

This is a homogeneous algebraic system. The condition for obtaining non-trivial solutions is that the determinant must be zero, namely that

$$\begin{vmatrix}a_{11}-\omega^2, & a_{12}\\ a_{21}, & a_{22}-\omega^2\end{vmatrix}=0.\tag{2.7}$$

This is a second-degree algebraic equation in the unknown ω^2. The equation is so important that it has a name, the *secular equation*. The equation has two roots (corresponding to the two degrees of freedom), say ω_1 and ω_2, which are called *proper angular frequencies* of the system. For each solution, there is a normal mode, say mode 1 and mode 2. Once the secular equation is satisfied, Eq. (2.6) give, for each mode, the *ratios* between the oscillation amplitudes of x_a and x_b.

For mode 1, we have

$$\left(\frac{x_b}{x_a}\right)_{\text{mode 1}} = \frac{B_1}{A_1} = \frac{\omega_1^2 - a_{11}}{a_{12}}, \tag{2.8}$$

where A_1 and B_1 are the amplitudes of the harmonic motions of x_a and x_b, respectively. The equations of the motions, for mode 1, are then

$$\begin{aligned} x_a(t) &= A_1 \cos(\omega_1 t + \phi_1) \\ x_b(t) &= B_1 \cos(\omega_1 t + \phi_1). \end{aligned} \tag{2.9}$$

Analogously, for mode 2, Eq. (2.6) give us

$$\left(\frac{x_b}{x_a}\right)_{\text{mode 2}} = \frac{B_2}{A_2} = \frac{\omega_2^2 - a_{11}}{a_{12}}. \tag{2.10}$$

where A_1 and B_1 are the amplitudes of the harmonic motions of x_a and x_b, respectively. The equations of motions for mode 2 are

$$\begin{aligned} x_a(t) &= A_2 \cos(\omega_2 t + \phi_2) \\ x_b(t) &= B_2 \cos(\omega_2 t + \phi_2). \end{aligned} \tag{2.11}$$

Note that the physical characteristics of the system, which are encoded in the a_{ij} constants, determine the proper frequency and the ratio between the amplitudes (called the *mode shape*) for each mode. Contrastingly, the values of the amplitude and the initial phase are determined by the initial conditions of the motion.

The most general solution to a system of two linear differential equations is given by a linear combination of two independent solutions. It is then evident that the most general solution of the system in Eq. (2.4) is

$$\begin{aligned} x_a(t) &= A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2) \\ x_b(t) &= B_1 \cos(\omega_1 t + \phi_1) + B_2 \cos(\omega_2 t + \phi_2). \end{aligned} \tag{2.12}$$

The initial conditions determine four quantities. As a matter of fact, only four of the six constants (A_1, A_2, B_1, B_2, ϕ_1 and ϕ_2) in Eq. (2.12) are independent. Indeed, Eqs. (2.8) and (2.10) determine the mode shapes, namely B_1/A and B_2/A_2, independently of the initial conditions, and we can rewrite the general solution in the form

$$\begin{aligned} x_a(t) &= A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2) \\ x_b(t) &= \left(\frac{B_1}{A_1}\right) A_1 \cos(\omega_1 t + \phi_1) + \left(\frac{B_2}{A_2}\right) A_2 \cos(\omega_2 t + \phi_2). \end{aligned} \tag{2.13}$$

The four constants A_1, A_2, ϕ_1 and ϕ_2 are determined by the initial conditions ($x_a(0)$, $x_b(0)$, $dx_a/dt(0)$ and $dx_b/dt(0)$).

A generic motion of a system with two degrees of freedom may be quite complicated. The motions of its parts may not be harmonic, but all these motions can be expressed as linear combinations of two simple harmonic motions.

Coming back to our initial example, let us find the combination of normal modes that expresses it. The shapes of the two modes of the system are given by the ratios $B_1/A_1 = 1$ and $B_2/A_2 = -1$. Thus, we write

$$\begin{aligned} x_a(t) &= A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2) \\ x_b(t) &= A_1 \cos(\omega_1 t + \phi_1) - A_2 \cos(\omega_2 t + \phi_2). \end{aligned}$$

We must find A_1, A_2, ϕ_1 and ϕ_2 in order to have the initial conditions $x_a(0) = A$, $x_b(0) = 0$, $dx_a/dt(0) = 0$ and $dx_b/dt(0) = 0$ satisfied. The solution is clearly $A_1 = A_2 = A/2$, $\phi_1 = \phi_2 = 0$. The equations of the motions are then

$$\begin{aligned} x_a(t) &= \frac{A}{2}(\cos\omega_1 t + \cos\omega_2 t) \\ x_b(t) &= \frac{A}{2}(\cos\omega_1 t - \cos\omega_2 t), \end{aligned} \tag{2.14}$$

which can also be written in the more transparent form

$$\begin{aligned} x_a(t) &= A\cos\left(\frac{\omega_2 - \omega_1}{2}t\right)\cos\left(\frac{\omega_2 + \omega_1}{2}t\right) \\ x_b(t) &= A\sin\left(\frac{\omega_2 - \omega_1}{2}t\right)\sin\left(\frac{\omega_2 + \omega_1}{2}t\right). \end{aligned} \tag{2.15}$$

The two pendulums are weakly coupled, namely we have

$$2\frac{k}{m} \ll \frac{g}{l}. \tag{2.16}$$

The difference between ω_1 and ω_2 is small compared to their values. The motion of each pendulum can be thought of as an almost harmonic motion at the mean angular frequency $(\omega_2 + \omega_1)/2$ with an amplitude $A\cos[(\omega_2 - \omega_1)t/2]$, which is not constant in time, but varies periodically, as a sine function, at low frequency (half the difference between the proper frequencies), as shown in Fig. 2.3. When the amplitude of a is large, that of b is small, and vice versa.

The energy of each pendulum is proportional to the square of its amplitude. The total energy of the system is proportional to the sum

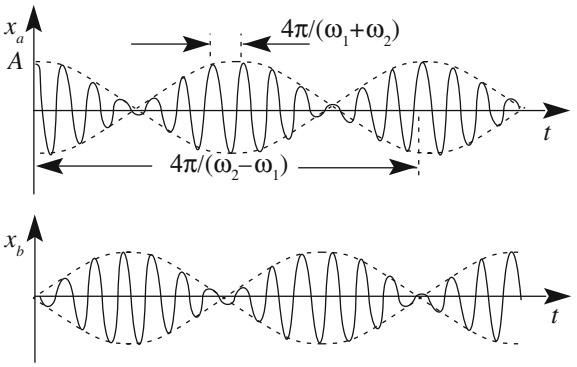


Fig. 2.3 The motions of two equal, weakly-coupled pendulums starting from the initial conditions $x_a(0) = A$, $x_b(0) = 0$, $dx_a/dt(0) = 0$ and $dx_b/dt(0) = 0$

$$A^2\left[\cos^2\left(\frac{\omega_2-\omega_1}{2}t\right)+\sin^2\left(\frac{\omega_2-\omega_1}{2}t\right)\right]=A^2,$$

which is constant, as expected.

Let us again use the example of the two equal, weakly-coupled pendulums to introduce the concept of *normal coordinates*. Let us substitute in the system of Eq. (2.3) the two linear combinations of x_a and x_b, which, remember, are the displacement for each pendulum of its equilibrium position

$$\begin{aligned} x_1(t) &= x_a(t)+x_b(t) \\ x_2(t) &= x_a(t)-x_b(t). \end{aligned} \tag{2.17}$$

Let us add and subtract the two equations in Eq. (2.3). We obtain

$$\begin{aligned} \frac{d^2x_1}{dt^2}+\omega_1^2x_1(t)&=0 \\ \frac{d^2x_2}{dt^2}+\omega_2^2x_2(t)&=0. \end{aligned} \tag{2.18}$$

The equations are now independent of one another. The coordinates enjoying such a property, like x_1 and x_2 in the example, are called *normal coordinates*. Note that each normal coordinate corresponds to one of the modes. Indeed, for mode 1, in which $x_a(t) = x_b(t)$, it is $x_2(t) = 0$ identically. We see that only x_1 is excited. Similarly, in mode 2, only x_2 is non-zero. Note also that the equation of each normal coordinate is the harmonic motion equation.

Analogous considerations hold for any linear, freely oscillating system with two degrees of freedom, neglecting dissipative forces. However, the simple expressions of the normal coordinates and normal modes we have found hold only in the simple and symmetric example we have considered. They are not even valid for a system similar to that shown in Fig. 2.1, but with two pendulums of different lengths, for example, with a shorter than b. The system is still simple, but not symmetric. The

configurations analogous to those shown in Fig. 2.2 are not the normal modes. In addition, if we observe the evolution of the system starting from an initial state analogous to that shown in Fig. 2.1a, we still see that, initially, the oscillation amplitude of a decreases over time and that of b increases. However, the amplitude of a does not go down to zero, but rather reaches a non-zero minimum and then goes back up again. In other words, energy is never transferred completely from the pendulum that is initially moving to the one that starts from rest.

The problem of finding the normal modes in general must be treated with the proper mathematics. We shall not do this here.

There exist freely oscillating systems with two degrees of freedom of very different nature. Figure 2.4 shows some examples, three mechanical systems and one electric. Very important cases of systems of two degrees of freedom exist in molecules, atoms, nuclei and elementary particles as well. Just to quote a few examples, we mention the ammonia molecule (and the MASER based on it), the hydrogen molecule, a number of dye molecules (and the colors they produce), the electrons in a magnetic field (and the electron-spin resonance) and the oscillation phenomena of elementary particles. These systems are properly described by quantum, rather than classical, physics. However, a proper classical oscillation frequency corresponds in quantum physics to the energy, or to the mass, of a state. Consequently, the analogy with classical physics is strong and can help us in intuiting an understanding of these quantum phenomena.

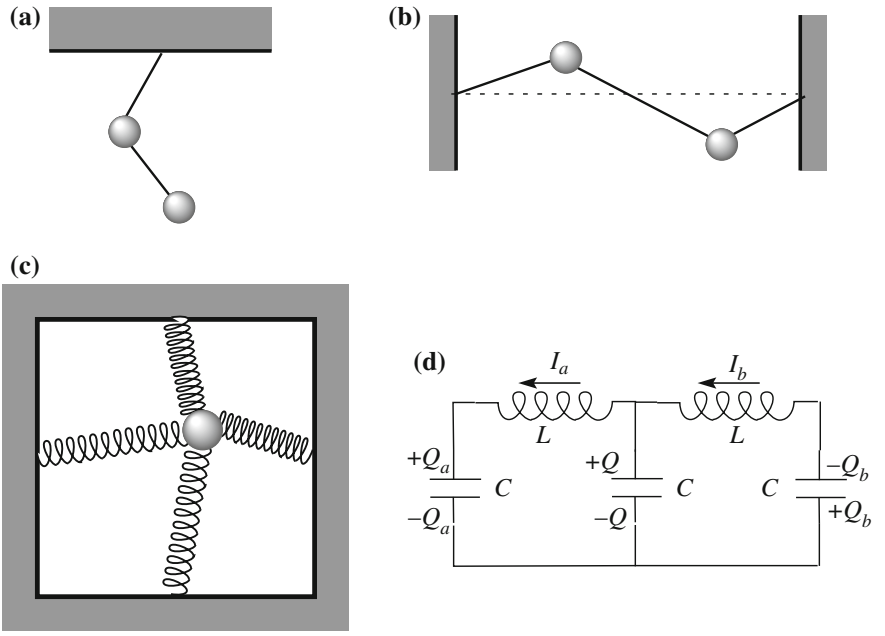


Fig. 2.4 Examples of systems with two degrees of freedom. **a** Double pendulum. **b** Two-mass loaded string, **c** Two-dimensional oscillator. **d** Capacitance coupled oscillating system

Let us consider, as another example, the two oscillating circuits in Fig. 2.4d. They have a capacitor in common and are consequently said to be capacity-coupled (an alternative is to have an inductor in common; the reader can analyze this as an exercise).

With the positive signs for charges and currents shown in the figure, the differential equations of the system are

$$\begin{aligned} -L\frac{dI_a}{dt} &= \frac{Q_a(t)}{C} - \frac{Q(t)}{C} \\ -L\frac{dI_b}{dt} &= \frac{Q_b(t)}{C} + \frac{Q(t)}{C}. \end{aligned}$$

With these sign conventions, we have $dQ_a/dt = +I_a$ and $dQ_b/dt = +I_b$, meaning an I_a running in the positive direction positively charges the capacitor a, and the same goes for I_b and capacitor b. Considering that there are instants in which all the charges on the capacitors are zero and that charge is conserved, we conclude that $Q(t) + Q_a(t) + Q_b(t) = 0$. We can then write the above differential equations as

$$\begin{aligned} \frac{d^2Q_a}{dt^2} &= -\frac{1}{LC}(2Q_a - Q_b) \\ \frac{d^2Q_b}{dt^2} &= -\frac{1}{LC}(2Q_b - Q_a). \end{aligned} \tag{2.19}$$

These equations are also equal to those of the two coupled pendulums with $1/LC$ in place of ω_1^2 and $1/LC$ in place of k/m. We can conclude that the proper frequencies of the two modes are $1/\sqrt{LC}$ and $\sqrt{3}/\sqrt{LC}$.

Question Q 2.1. Analyze an inductance-coupled oscillating circuit, namely a circuit similar to that shown in Fig. 2.4, with capacitances where there are inductances and inductances where there are capacitances. □

That which we have discussed can be easily generalized to systems of n degrees of freedom. We call the coordinate measuring the displacement from its own equilibrium position of each part of the system with $x_1, x_2, \ldots, x_n$, respectively. The differential equations of the motion of the system, neglecting dissipative forces, are

$$\begin{aligned} \tfrac{d^2x_a}{dt^2} &= -a_{11}x_1 - a_{12}x_2 - \ldots - a_{1n}x_n \\ \tfrac{d^2x_2}{dt^2} &= -a_{21}x_1 - a_{22}x_2 - \ldots - a_{2n}x_n. \\ &\ldots \\ \tfrac{d^2x_n}{dt^2} &= -a_{n1}x_1 - a_{n2}x_2 - \ldots - a_{nn}x_n \end{aligned} \tag{2.20}$$

We look for solutions that are normal modes, namely motions in which all the coordinates move in harmonic motion with the same frequency and with the same initial phase. In other words, we look to see if we can find one or more values of ω such that $x_1 = A^{(1)}\cos(\omega t + \phi), \ldots, x_n = A^{(n)}\cos(\omega t + \phi)$ is a solution. We substitute these functions in Eq. (2.20) and obtain the algebraic system

$$\begin{aligned}
&(a_{11} - \omega^2)x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\
&a_{21}x_1 + (a_{22} - \omega^2)x_2 + \ldots + a_{2n}x_n = 0 \\
&\ldots \\
&a_{n1}x_1 + a_{n2}x_2 + \ldots + (a_{nn} - \omega^2)x_n = 0.
\end{aligned} \tag{2.21}$$

In this case too, we must impose that the determinant be zero, in order to obtain non-trivial solutions. We have an algebraic equation of n degrees in ω^2. Its n solutions, say $\omega_1^2, \omega_2^2, \ldots \omega_n^2$, are the proper angular frequencies of the n normal modes. While it can be shown that all the solutions are positive, and consequently physically meaningful, it can happen, depending on the system, that the values of some of them coincide. In this case, the corresponding normal modes are said to be *degenerate*.

The system in Eq. (2.21) gives the ratios between the n amplitudes and one that is arbitrary, namely the mode shape. The initial conditions determine the other constants.

2.2 Forced Oscillators with Several Degrees of Freedom

We shall now consider a linear oscillating system with two degrees of freedom, forced by an external force and in the presence of a drag force proportional to the velocity. We shall start with the example of the two equal pendulums from the previous section, applying a force to one of them, namely pendulum a, as shown in Fig. 2.5. We consider an external force sinusoidally dependent on time, say $F(t) = F_0 \cos \omega t$.

The differential equations of motions of the system are

$$\begin{aligned}
&\frac{d^2x_a}{dt^2} + \gamma\frac{dx_a}{dt} + \frac{g}{l}x_a + \frac{k}{m}(x_a - x_b) = \frac{F_0}{m}\cos\omega t \\
&\frac{d^2x_b}{dt^2} + \gamma\frac{dx_b}{dt} + \frac{g}{l}x_b + \frac{k}{m}(x_b - x_a) = 0.
\end{aligned} \tag{2.22}$$

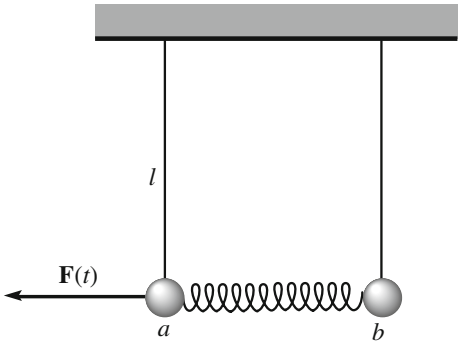


Fig. 2.5 Two equal forced coupled pendulums

We know that the two normal modes of the system when it is free and in the absence of drag, namely if $\gamma = F_0 = 0$, are

$$\begin{array}{ll} \text{mode 1} & x_a(t) = x_b(t); \quad \omega_1^2 = \frac{g}{l} \\ \text{mode 2} & x_a(t) = -x_b(t); \quad \omega_2^2 = \frac{g}{l} + 2\frac{k}{m}. \end{array}$$

Let us try and see if the normal coordinates that worked for the free system, namely

$$\begin{aligned} x_1(t) &= x_a(t) + x_b(t) \\ x_2(t) &= x_a(t) - x_b(t) \end{aligned}$$

still work now. Let us add and subtract the two sides of Eq. (2.22), obtaining

$$\begin{aligned} \frac{d^2x_1}{dt^2} + \gamma\frac{dx_1}{dt} + \frac{g}{l}x_1 &= \frac{F_0}{m}\cos\omega t \\ \frac{d^2x_2}{dt^2} + \gamma\frac{dx_2}{dt} + \left(\frac{g}{l} + \frac{2k}{m}\right)x_2 &= \frac{F_0}{m}\cos\omega t. \end{aligned} \tag{2.23}$$

We find that we have been lucky; the two equations are independent of one another. Consequently, x_1 and x_2 are the normal coordinates for the present system as well.

We see that each of the equations in Eq. (2.23) is the differential equation of a damped and forced oscillation. The normal coordinate x_1 behaves like the coordinate of such an oscillator with proper square angular frequency $\omega_1^2 = g/l$ and damping γ forced by the force $F(t) = F_0 \cos wt$. The normal coordinate x_2 behaves similarly with proper square angular frequency $\omega_2^2 = g/l + 2k/m$. The two oscillations are independent. Each of them behaves like a one degree of freedom oscillator. Remember, however, that x_1 and x_2 are not physical displacements, but rather combinations of these.

The system has two resonances, one for each of its proper frequencies. If the frequency of the external force is close to one of the resonance frequencies, the system, after the transient phase has finished, reaches its steady regime. It moves in the normal mode corresponding to that proper frequency. The other normal coordinate is zero.

The resonances of the two modes, in general, not only have different frequencies but also different widths. This is not the case in the simple example we have just discussed, but suppose, for example, that the spring joining the pendulums dissipates a certain amount of energy when it is stretched back and forth. In this case, the energy loss rate will be larger for mode 2 (in which the spring is stretched) than for mode 1 (in which the length of the spring does not vary). As a consequence, the width of the second resonance will be larger than that of the first.

The discussion we developed around a simple example is valid in general, provided that the difference between the proper frequencies is substantially larger

than the widths of both of them. In these cases, normal motions exist, even if finding the normal coordinates is not as simple as in the above example. This is not the case in the presence of damping if the resonances are too close to one another.

All the arguments can be extended to systems with n degrees of freedom.

2.3 Transverse Oscillations of a String

Up to now, we have considered systems with a discrete number of degrees of freedom. We shall now discuss the normal modes of a system with a continuous, infinite number of degrees of freedom. We shall analyze the important case of an elastic string with fixed extremes. More precisely, our string is an ideal one, namely it is perfectly elastic (its tension is proportional to the stretching), perfectly flexible (it does not oppose to folding) and is homogenous (its linear mass density ρ is uniform). The extremes are fixed at a distance L. At equilibrium, the tension of the rope, T_0, is independent of the position. The motions we shall study are the small oscillations.

Let x, y, z be a reference frame, with the z-axis on the equilibrium position of the string and the origin in its left extreme, as in Fig. 2.6. We identify each element of the string by the z coordinate of the *equilibrium position* of the element. The position of that element when the string is moving at the instant in time t is a vector function of z and t, which we call $\boldsymbol{\psi}(z, t)$. This vector has a component along the string, corresponding to the *longitudinal oscillation*, and two components normal to the string, corresponding to the *transverse oscillations*. We shall limit the discussion here to the transverse oscillations. The z-component of $\boldsymbol{\psi}(z, t)$ is identically zero.

In general, in a transverse oscillation, the direction of $\boldsymbol{\psi}(z, t)$ on the xy plane is different both for different z and for different times. A transverse oscillation is said to be *plane polarized* or, alternatively, *linearly polarized* if the direction of $\boldsymbol{\psi}(z, t)$ is independent of both z and t. In this case, if we take a picture at any instant in time, we see the string shape as a plane curve, always in the same plane. We shall limit the discussion, for the sake of simplicity, to a linearly polarized motion, in which the displacement is represented by a single function $\psi(z, t)$, rather than by a vector.

Let us consider a small segment of the string, of length at rest Δz. Its mass is $\Delta m = \rho \Delta z$. In the generic non-equilibrium configuration, the element we are considering is removed from the z-axis, as shown in Fig. 2.7. The displacements of both extremes are normal to the z-axis, because we assumed a purely transverse

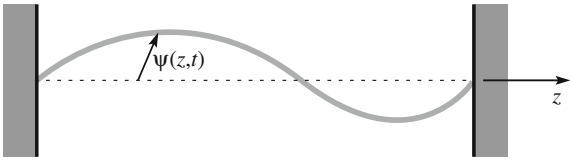


Fig. 2.6 A configuration of an elastic string with fixed extremes

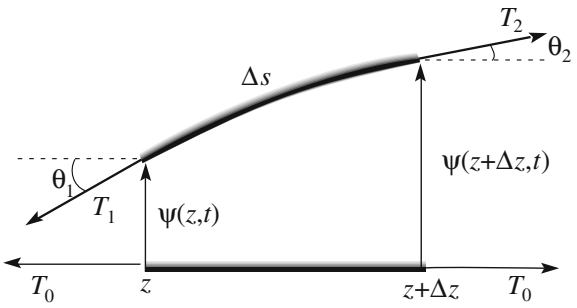


Fig. 2.7 A segment of the elastic string in a generic position

oscillation and they are, in general, different from one another. In general, the element is not straight, namely the angles with the z-axis at the two extremes may be different ($\theta_1 \neq \theta_2$ in Fig. 2.7). The component of the resultant force in the direction opposite to that of the displacement is

$$F(t) = T_2 \sin\theta_2 - T_1 \sin\theta_1.$$

Two important simplifications are possible if the displacements from equilibrium of all the elements of the string are small. First, we can approximate the tangent of the angles with the angle or with its sine. Second, we can use the Pythagorean theorem to find an approximate relation between the length Δs of the element and its length at rest Δz. The theorem gives us

$$\Delta s^2 = \Delta z^2 + [\psi(z+\Delta z,t) - \psi(z,t)]^2 = \Delta z^2 + \left(\frac{\partial \psi}{\partial z}\Delta z\right)^2,$$

namely

$$\Delta s = \left[1 + \left(\frac{\partial \psi}{\partial z}\right)^2\right]^{1/2} \Delta z$$

Now, $\partial\psi/\partial z = \tan\theta$ is small, say infinitesimal, and we can expand the square root on the right-hand side of this expression, stopping at the first term. We have

$$\Delta s = \left[1 + \frac{1}{2}\left(\frac{\partial \psi}{\partial z}\right)^2\right] \Delta z.$$

This expression shows, finally, that the difference between Δs and Δz is infinitesimal of the second order. Under these conditions, the tensions T_1 and T_2 at the two extremes differ from the tension at rest T_0 by infinitesimals of the second order, which we neglect. Under these approximations, for the restoring force, we can write

$$F(t) = T_0 \sin\theta_2 - T_0 \sin\theta_1 = T_0 \left(\frac{\partial\psi}{\partial z}\right)_2 - T_0 \left(\frac{\partial\psi}{\partial z}\right)_1 = T_0 \frac{\partial^2\psi}{\partial z^2}\Delta z.$$

The acceleration of the element is $\partial^2\psi/\partial t^2$ and the second Newton law gives us

$$T_0 \frac{\partial^2\psi}{\partial z^2}\Delta z = \rho \Delta z \frac{\partial^2\psi}{\partial t^2}.$$

Finally, simplifying out Δz, we obtain

$$\frac{\partial^2\psi}{\partial t^2} - \upsilon^2 \frac{\partial^2\psi}{\partial z^2} = 0, \tag{2.24}$$

where we have introduced the constant

$$\upsilon = \sqrt{T_0/\rho}, \tag{2.25}$$

which has the physical dimension of a velocity, as it is easy to check.

Equation (2.24) is a very famous partial differential equation known as the *vibrating string equation* and, more commonly, the *wave equation*, for reasons that will become clear in the subsequent chapter. From now on, through the entire book, we shall discuss the properties of its solutions in a number of sectors of physics.

Let us now search for the *normal modes* of the system. Namely, we look for solutions in which all the parts of the system move sinusoidally with the same frequency and with the same phase, meaning that both ω and ϕ are independent of z. The solution should have the form

$$\psi(z,t) = A(z)\cos(\omega t + \phi). \tag{2.26}$$

Let us substitute this in Eq. (2.24). We obtain

$$\frac{d^2A(z)}{dz^2} = -\frac{\omega^2}{\upsilon^2}A(z). \tag{2.27}$$

The solution to this differential equation $A(z)$ gives the *shape of the mode*. Different modes have different shapes due to the factor ω^2 in the equation, which differs from one mode to another.

Equation (2.27) is formally equal to the equation of the harmonic oscillator, with the space coordinate z at the place of the time t. Hence, we know how to solve it. In the present case, the most useful form is

$$A(z) = A\sin kz + B\cos kz, \tag{2.28}$$

where A and B are the integration constant that we shall soon find and

$$k^2 = \frac{\omega^2}{v^2}. \tag{2.29}$$

The quantity k is called the *wave number*, which is inversely proportional to the *wave length* λ as

$$k = \frac{2\pi}{\lambda}. \tag{2.30}$$

The function $A(z)$ is the oscillation amplitude of the element at z, namely the shape of the mode. We see that it is a sinusoidal function of z. The wavelength is the period in length, which is equivalent to the period in time T in a periodic function of time. The unit of the wavelength is the meter. Similarly, the wave number k is the equivalent in length of the angular frequency ω ($\omega = 2\pi/T$) in time. Consequently, it is also called *spatial frequency*. Its measurement units are the inverse of a meter (m^{-1}). Note that we shall use the same symbol k for spatial frequency as for the spring constant, but this should not generate confusion.

Note also that ω and k are not at all independent. Rather, when one of the two is known, the other is known as well. Indeed, for every system, a relation between ω and k exists, called the *dispersion relation* . The dispersion relation of the ideal elastic string we are discussing is given by Eq. (2.29). Other systems, in general, have more complicated dispersion relations. The dispersion relation is independent of the boundary conditions.

Let us go back to the normal modes we have found, putting together Eqs. (2.26) and (2.28). The solution we have found is

$$\psi(z,t) = (A\sin kz + B\cos kz)\cos(\omega t + \phi). \tag{2.31}$$

We now determine the integration constant A and B by imposing the boundary conditions. These are the equivalent of the initial conditions we have always used when working in the time domain. The boundary conditions are that the extremes do not move, namely $\psi(0, t) = 0$ and $\psi(L, t) = 0$ for any t. The first of these gives $B = 0$ and we can consequently rewrite Eq. (2.31) as

$$\psi(z,t) = A\sin(kz)\cos(\omega t + \phi) = A\sin\left(\frac{2\pi}{\lambda}z\right)\cos(\omega t + \phi). \tag{2.32}$$

The second condition, and $\psi(L, t) = 0$, imposes that $\sin(2\pi L/\lambda) = 0$, trivial solution $A = 0$ apart. The unique quantity we can adjust is the wavelength λ. This means that the modes exist only for definite wavelengths. There is an infinite sequence of them, namely

$$\lambda_1 = 2L,\ \lambda_2 = \lambda_1/2,\ \lambda_3 = \lambda_1/3,\ \lambda_4 = \lambda_1/4, \ldots. \tag{2.33}$$

Understanding the reason for that is easy if we recognize that the wavelengths of the modes are those that have one half of the string length as an integer multiple.

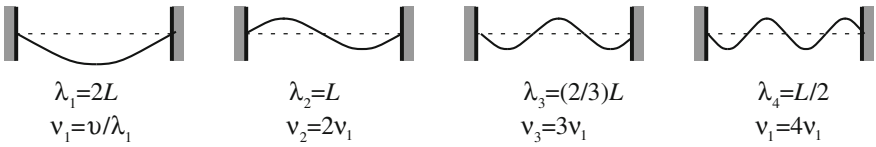


Fig. 2.8 The first four oscillation modes of an elastic string with fixed extremes

Namely, an integer number of half wavelengths must fit exactly between the extremes, as shown in Fig. 2.8.

Summarizing, when the string vibrates in one of its normal modes, each of its elements oscillates in a sinusoidal motion, all of them at the same frequency and in the same phase. There are instants in which the string is straight, passing through its equilibrium configuration. The vibration amplitude is different in the different positions. There are points that are always at rest, which are called the *nodes*, and points that vibrate with maximum amplitude, called the *anti-nodes*. Two contiguous nodes and two contiguous anti-nodes are separated by half a wavelength.

The dispersion relation in Eq. (2.29) can be written in terms of the wavelength λ and of the frequency $\nu = 1/T$ as

$$\nu\lambda = \upsilon. \tag{2.34}$$

It follows that each mode vibrates with a definite frequency. The sequence (or progression) of these frequencies is

$$\nu_1 = \upsilon/\lambda_1,\ \nu_2 = 2\nu_1,\ \nu_3 = 3\nu_1,\ \nu_4 = 4\nu_1, \ldots \tag{2.35}$$

We see that the proper frequencies are the integer multiples of the smallest one, called the *fundamental frequency*. This is a sequence (or progression) well known from mathematics courses, called the harmonic sequence. The frequencies above the fundamental are called harmonics. The reason for these names is in the physical phenomenon we are discussing. Indeed, Pythagoras of Samos (Greece, −570 to −495) discovered early on that two vibrating strings of a musical instrument give a pleasant sound, we say a harmonic sound, if their lengths are multiples of one another, and consequently if the fundamentals are multiples of one another as well. The sound is also pleasant if the lengths are in the ratio of two small integer numbers, when some of the harmonics of the two strings coincide.

Note that what we have just observed is a consequence of the linear relation between ω and k. Namely, the vibration frequency is directly proportional to the wave number, hence inversely proportional to the wavelength ($\nu = \upsilon/\lambda$, with υ being a constant). We also note that the proportionality constant

$$\upsilon = \sqrt{T_0/\rho}. \tag{2.36}$$

is directly proportional to the square root of the tension (namely of the restoring force) and inversely to the square root of the density (namely the inertia). In other words, for a given length of the string, the vibration frequency is higher if the tension is larger and if the density is smaller. As always, the proper frequency square is proportional to the restoring force per unit mass.

As we already stated, a dispersion relation exists for every system, however, such a relation is often more complicated than Eq. (2.29) or Eq. (2.34) $\lambda\nu = \omega/k = \upsilon$ = constant. Piano strings, for example, are not perfectly flexible. They present a small degree of stiffness, which opposes flexion. Being that flexion is larger for larger curvatures, namely for shorter wavelengths or larger wavenumbers k, the contribution of stiffness to the restoring force is larger for larger wave numbers. Let us develop the unknown dispersion relation $\omega^2(k^2)$ in series of powers of k^2 and stop at the first term. This is proportional $(k^2)^2$ by a certain constant α. We write

$$\omega^2 = \frac{T_0}{\rho} k^2 + \alpha k^4.$$

We do not really know the constant α, but, remembering that ω^2 is the restoring force per unit displacement and per unit mass, we know that it is positive for the argument given above.

The boundary conditions for our piano string are the same as those we discussed for an ideal string, and consequently the shapes of the modes are the same. Their wavelengths are still $\lambda_1 = 2L$, $\lambda_2 = \lambda_1/2$, etc., but now the proper frequencies of the modes are not exactly integer multiples of the fundamental. Being that $\alpha > 0$, the frequencies of the higher modes are a bit higher than those of the harmonic sequence. The sound is still (or even more) pleasant if the differences are not too large.

We have seen in Sects. 2.1 and 2.2 that every motion of a linear system with n degrees of freedom (where n is an integer number) can be expressed as a linear combination of its n normal modes. The vibrating string we are now studying has a continuous infinite number of degrees of freedom. As we have seen, its normal modes are numerable infinite. Well, even in this case, it can be shown (although we shall not do that) that every possible motion of the string (with fixed extremes) can be expressed as a linear combination of its normal modes. This means that, given an arbitrary motion represented by the function $\psi(z, t)$, we can find two infinite sequences of numbers, $F_0, F_1, F_2\ldots$ (amplitudes) and $\phi_1, \phi_2,\ldots$ (initial phases) such that

$$\psi(z,t) = \sum_{m=1}^{\infty} F_m \cos(\omega_m t + \phi_m) \sin k_m z, \tag{2.37}$$

where ω_m are the proper angular frequencies and $k_m = \omega_m/\upsilon$.

The normal modes of the vibrating string that are solutions to the wave equation are also called *stationary waves*.

Let us finally consider the energy of a vibrating string. Every element of the string has kinetic and potential energy in every moment. As we have just seen, every motion can be expressed as a superposition of the modes. Consider now the energy the string would have if it were vibrating in the generic mode m with the same amplitude F_m appearing in that superposition. Let us call it the *energy of the mode*. Well, it is easy to show (although we shall not do that) that the energy of the string in the considered motion is equal to the sum of the energies of the modes of the superposition taken separately.

2.4 The Harmonic Analysis

We have already stated that the functions that we usually encounter in the description of physical phenomena can be expressed as linear combinations of sines or cosines. The mathematical process for finding such combinations is called the *harmonic analysis* or *Fourier analysis* of the function under examination. This represents an important chapter in mathematics initiated in modern terms by Joseph Fourier (France; 1768–1830) in 1809 (after important contributions by several predecessors). In both this section and the one that follows it, we shall present the elements of the harmonic analysis that we shall need in the subsequent study, having in mind physics rather than mathematical rigor. We shall not provide the mathematical demonstrations of the statements. However, we shall discuss several examples of interest for physics.

We preliminarily note that harmonic analysis is extremely useful, both for functions of time, such as the displacement of a point of a string or the evolution of a force, and for functions of the position in space, such as the surface of the sea with its waves at a certain instant or the level of gray in a photograph.

This section has an introductory character, intended to provide a sense of the issue as it applies to physics. As we are dealing with an analysis that is called harmonic, let us start from harmony, namely from music and musical tones.

To be specific, let us consider a guitar string tuned to an A at 440 Hz. We assume the string to be perfectly flexible. Consequently, the frequencies of its normal nodes are in harmonic sequence, namely

$$\nu_1 = 440\,\text{Hz},\ \nu_2 = 880\,\text{Hz},\ \nu_3 = 1320\,\text{Hz},\ldots, \tag{2.38}$$

If we pluck the string, we move it out of equilibrium. Initially, the string is not in a normal mode. The disturbance propagates in both directions, reaches the extremes, is reflected there, and then comes back, and so on. After a short moment of transition, the motion of the string becomes stationary. The stationary vibration continues for a duration that is very long compared to the period of the fundamental, which is $1/\nu_1 = 22.7$ ms.

In its vibrations, the string produces periodic variations of the air pressure that propagates in space. This is the sound wave we shall be studying in Sect. 3.4. The

sound wave, in turn, sets our eardrum into vibration, which is a motion completely similar to the vibrations of the points of the string; and thus we perceive the sound.

Let $f(t)$ be the displacement from equilibrium of the eardrum at the instant t. If we were to measure it, we would obtain something like that represented in Fig. 2.9.

As a comparison, let us think about recording the motion of our eardrum when we perceive noise, for example, the clapping of our hands. We would find something similar to Fig. 2.10. Comparing the two cases, we see that a musical sound corresponds to a *periodic* function of time, and a noise, a non-periodic one. The function shown by Fig. 2.9 repeats identically when the time is incremented by a well-defined quantity T, which is the *period* of the function. Mathematically expressed, the function has the property that $f(t + T) = f(t)$ for every t. This is clearly an idealization. Indeed, rigorously speaking, a periodic function should exist for an infinitely long time from minus infinite in the past to plus infinite in the future. No physical phenomenon is represented by a truly periodic function. However, if the duration of the phenomenon is long compared to the period, we can consider the function as being approximately periodic.

The periodic function $f(t)$ representing the sound of the string in general is not, however, a simple sinusoidal function. This is because the string does not vibrate in a normal mode, as a consequence of its initial state not having had the shape of any mode. If the motion was the first normal motion (the fundamental), the vibration of the eardrum (and of any point on the string) would be a cosine of period $T = 1/\nu_1 = 22.7$ ms. The second normal motion is a cosine function as well, with period $T/2$. This means that the function repeats itself every $T/2$ s. But it repeats itself every T seconds as well. Similarly, the third mode repeats itself every $T/3$ s,

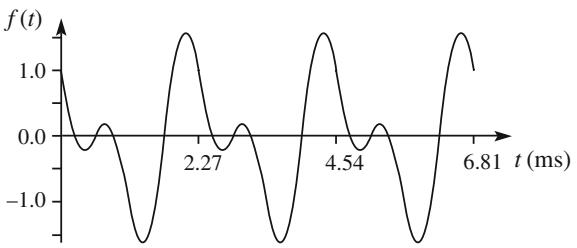


Fig. 2.9 A musical sound

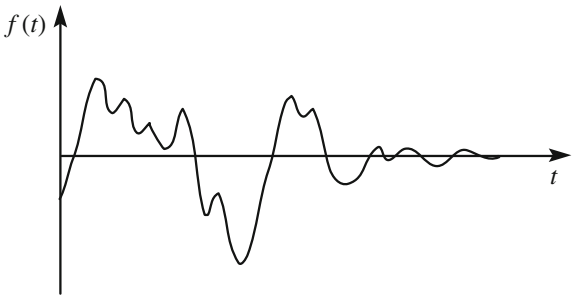


Fig. 2.10 A noise

and hence also every T seconds, etc. In conclusion, all the functions cos $\omega_0 t$, cos $2\omega_0 t$, cos $3\omega_0 t$, … (and sin $\omega_0 t$, sin $2\omega_0 t$, sin $3\omega_0 t$, … as well) are periodic functions with period T. Hence, every linear combination of these functions is periodic with period T as well.

Hence, the motion of our eardrum, which is a linear combination of harmonic motions as given by Eq. (2.38), is a periodic motion at the frequency $T = 1/\nu_1$. The term corresponding to each mode enters into the linear combination with a certain amplitude (we shall call F_m the amplitude of the mode at the frequency ν_m) and with a certain initial phase (which we shall indicate with ϕ_m). Both quantities depend on the initial configuration of the string. If the angular frequency of the fundamental is $\omega_0 = 2\pi/T$, the general motion of the eardrum can be expressed as

$$f(t) = F_1 \cos(\omega_0 t + \phi_1) + F_2 \cos(2\omega_0 t + \phi_2) + F_3 \cos(3\omega_0 t + \phi_3) + \ldots, \quad (2.39)$$

We obtained the example shown in Fig. 2.9 by adding together only two terms, namely as

$$f(t) = \cos(2\pi \cdot 440 \cdot t + 0) + 0.8 \cdot \cos(2\pi \cdot 880 \cdot t + \pi/2),$$

with t in seconds. Figure 2.11 graphically represents the two components and the resulting combination. Note, in particular, how the two components start, at $t = 0$, at a different point of their period. This is because the initial phases are different: one is 0 and the other is $\pi/2$.

We shall now use this simple example to discuss some general features.

Let us first consider doubling the amplitudes of both components. The result is simply twice what we had. The new function has the same shape as the old one. Hence, generally speaking, the shape of the resulting function depends on the ratios between the component amplitudes, and not on their absolute value.

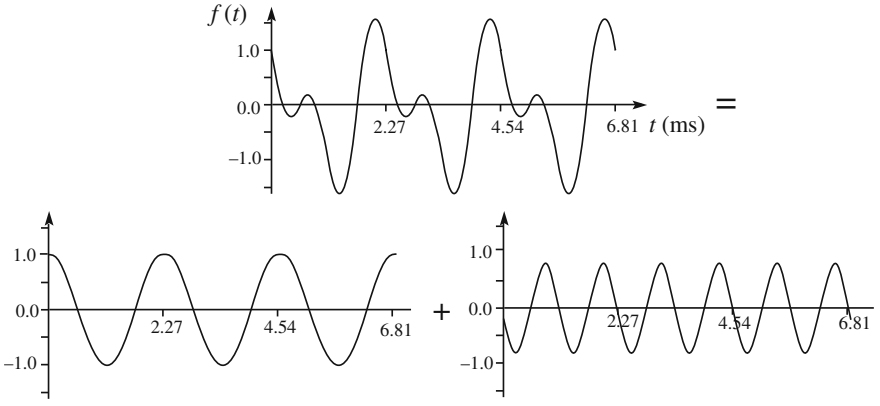


Fig. 2.11 The components of the function in Fig. 2.9 having periods T and $T/2$

Let us now change the initial phases by adding the same quantity to each of them, for example, $\pi/2$. We have the function

$$f(t) = \cos(2\pi \cdot 440 \cdot t + \pi) + 0.8 \cdot \cos(2\pi \cdot 880 \cdot t + 3\pi/2),$$

The result is shown as a dotted curve in Fig. 2.12. The new function has the same shape and magnitude as the old one. It is simply translated forward in time by half a period.

Let us now change the difference between the initial phases. Let us have, for example, $\pi/4$ in the second term instead of $\pi/2$ without changing the argument of the first, obtaining

$$f(t) = F_1 \cos(2\pi \cdot 440 \cdot t) + 0.8 \cdot F_1 \cos(2\pi \cdot 880 \cdot t + \pi/4).$$

Figure 2.13b shows the result, while Fig. 2.13a shows the original function for comparison. Now, the shape has changed. Generalizing the result, the shape of the combination depends on the differences between the initial phases, but is independent of their absolute values.

All the functions we have considered in the above examples have mean values over a period equal to zero. Indeed, this is the case for any function representing the displacement from a fixed equilibrium position in an oscillation about it. More generally, the mean value of a function might be different from zero. Let us consider, for example, the function

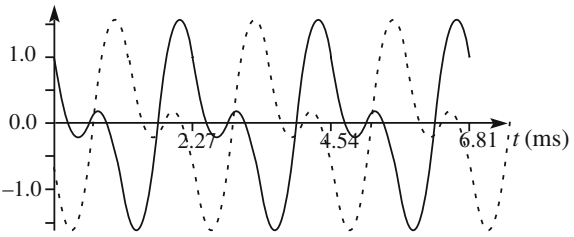


Fig. 2.12 Changing by the same amount both phases of the components

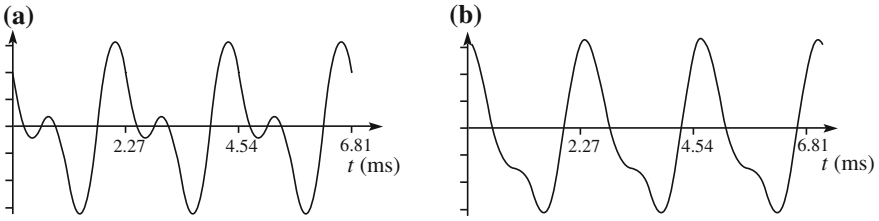


Fig. 2.13 Changing the phase difference between the components

$$f(t) = F_0 + \cos(2\pi \cdot 440 \cdot t) + 0.8 \cdot \cos(2\pi \cdot 880 \cdot t + \pi/2).$$

Being that the mean values over a period of all the cosines are zero, the mean value of f is F_0. Figure 2.14 shows this function as a dotted line compared with the original one, which is a continuous line. Clearly, the shapes of both are equal. We have only a vertical shift up or down, depending on the sign of F_0.

We have concluded the observations for our simple example. They have a general character. We are ready to state the Fourier theorem, which is valid for a very large class of periodic function, as precisely defined by mathematics. We shall not demonstrate the theorem, but shall simply state that practically all the functions encountered in physics obey the theorem. We state that, given any "reasonable" function of time $f(t)$ periodic of period T, one can always find two infinite sequences of real numbers F_0, F_1, F_2, … and ϕ_1, ϕ_2, …, which, with $\omega_0 = 2\pi/T$, are such that

$$f(t) = F_0 + F_1\cos(\omega_0 t + \phi_1) + F_2\cos(2\omega_0 t + \phi_2) + \ldots, \tag{2.40}$$

Let us go back to our example of the musical tone. In general, the motion of our eardrum can be expressed as the series in Eq. (2.40) (with $F_0 = 0$). The term in $\nu_1 = \omega_0/2\pi$ is the fundamental, the subsequent ones are the harmonics. Each of them is the same note in a higher octave. Two notes differ by an octave when the frequency of the second is twice the frequency of the first. The amplitudes F_m define the relative importance of the subsequent harmonics. Our ear is capable of appreciating the relative weights of the harmonics, namely the F_m. In other words, our ear performs a Fourier analysis. On the other hand, we are not sensitive to the phases.

The proportions of the fundamental and of the different harmonics determine what is called the *timbre* (and also *color*) of the sound. A sound is said to be *pure* if it contains the fundamental alone, *rich* if, on the contrary, several harmonics are important. The same note, the A we have been considering, for example, is different if it is played by a piano, an oboe, a violin or another instrument, because the different instruments produce harmonics in different proportions. Similarly, the timbre distinguishes the same note sung once as a-a-a, and then again as o-o-o.

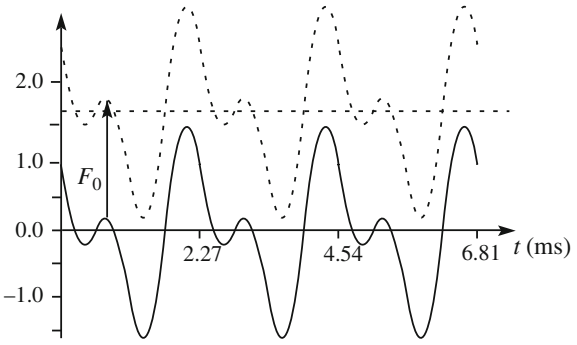


Fig. 2.14 Including a constant term

We recall now that, at the end of Sect. 2.3, we stated that every motion of the string can be expressed as a linear combination of its normal modes writing Eq. (2.37). Let us now consider an arbitrary point of the string, for example, the point at $z = z^*$, and let us indicate its displacement from equilibrium with $g(t)$. This function of time is given by Eq. (2.37), which is a function of z and t, valuated for $z = z^*$, namely

$$g(t) = \psi(z^*, t) = \sum_{m=1}^{\infty} [F_m sink_m z^*] \cos(m\omega_0 t + \phi_m).$$

This expression is just Eq. (2.40) with $F_m sink_m z^*$ in place of F_m, which is just another way to write the constants. This consideration shows that the Fourier series is also the series of the normal modes of a physical system, the flexible string.

The development in Eq. (2.40) can be written in an equivalent form, which will be useful and that we will find immediately. We start from the trigonometric identity $\cos(\omega t + \phi) = \cos\phi \cdot \cos\omega t - \sin\phi \cdot \sin\omega t$. Considering that $\cos\ \phi$ and $\sin\ \phi$ are two constants, we can absorb them into the amplitudes of the terms and express the function as the sum of a linear combination of sines $(\sin m\omega_0 t)$ and one of cosines $(\cos m\omega_0 t)$ with initial phase zero. The equivalent form of Eq. (2.40) is

$$f(t) = A_0 + A_1 \cos\omega_0 t + A_2 \cos 2\omega_0 t + \ldots + B_1 \sin\omega_0 t + B_2 \sin 2\omega_0 t + \ldots, \tag{2.41}$$

2.5 Harmonic Analysis of a Periodic Phenomena

In this section, we generalize the conclusions we have reached, giving, without any demonstration, the proper mathematical expressions for the Fourier series. We shall see how to calculate the coefficients in three equivalent, but all useful, expressions.

Consider the *periodic* function of time $f(t)$ with period T, and, correspondingly, angular frequency $\omega_0 = 2\pi/T$. As we know, all the functions $\cos\omega_0 t$, $\cos 2\omega_0 t$,…, $\cos m\omega_0 t$… and $\sin\omega_0 t$, $\sin 2\omega_0 t$,…, $\sin m\omega_0 t$…are periodic with period T. In addition, these functions constitute a complete set of normal orthogonal functions over the interval 2π. This means that the functions enjoy the following properties:

$$\begin{aligned}
&\frac{1}{\pi}\int_0^{2\pi} \cos mx \cdot \cos nx \cdot dx = \begin{cases} 0 & \text{for } m \neq n \\ 1 & \text{for } m = n \end{cases} \\
&\frac{1}{\pi}\int_0^{2\pi} \sin mx \cdot \sin nx \cdot dx = \begin{cases} 0 & \text{for } m \neq n \\ 1 & \text{for } m = n \end{cases} \\
&\frac{1}{\pi}\int_0^{2\pi} \cos mx \cdot \sin nx \cdot dx = 0 \quad \text{for any } m \text{ and } n.
\end{aligned} \tag{2.42}$$

In other words, the properties are as follows. Firstly, the integral over a period of the product of two *different* functions is zero. Such functions are said to be orthogonal in analogy to the fact that the scalar product of two orthogonal vectors is zero. Here, we have the integration in place of the scalar product. Secondly, the integral of the square of each function is equal to 1. Such functions are said to be normal. This has been obtained by including the "normalization factor" $1/\pi$.

The Fourier theorem states (as we already discussed in Sect. 2.4) that, given any "reasonable" function of time $f(t)$ periodic of period T, it is always possible to find two infinite successions of real numbers A_0, A_1, A_2, … and B_1, B_2, … such that

$$f(t) = A_0 + \sum_{m=1}^{\infty} A_m \cos(m\omega_0 t) + \sum_{m=1}^{\infty} B_m \sin(m\omega_0 t), \tag{2.43}$$

where $\omega_0 = 2\pi/T$. The constant term A_0 is the mean value of the function in *any* time interval T, namely it is

$$A_0 = \langle f \rangle = \frac{1}{T} \int_{\tau}^{\tau+T} f(t)dt \tag{2.44}$$

Note that the cosine is an even function of its argument (its values in opposite values of the argument are equal). Consequently, in the development of an even function $f(t)$ of t, (namely such that $f(-t) = f(t)$), $B_m = 0$ for all m. Similarly, in the development of an odd function ($f(-t) = -f(t)$), $A_m = 0$ for all m. In the general case, expression (2.43) shows the even and odd parts of the function separately.

The ortho-normality property of the set of sine and cosine functions allows us to find the expressions of the coefficients of the series immediately. To find the generic A_n, we multiply both sides of Eq. (2.43) by $\cos n\omega_0 t$, obtaining

$$f(t)\cos(n\omega_0 t) = A_0 \cos(n\omega_0 t) + \sum_{m=1}^{\infty} A_m \cos(m\omega_0 t)\cos(n\omega_0 t) + \sum_{m=1}^{\infty} B_m \sin(m\omega_0 t)\cos(n\omega_0 t)$$

and integrate over a period, say from τ to $\tau + T$. For the orthogonality property, the only integral different from zero on the right-hand side is the term in the first sum with $m = n$, so that we have

$$\int_{\tau}^{\tau+T} f(t)\cos(n\omega_0 t)dt = A_n \int_{\tau}^{\tau+T} \cos^2(n\omega_0 t)dt = A_n \frac{T}{2},$$

and hence, finally,

$$A_n = \frac{2}{T}\int_{\tau}^{\tau+T} f(t)\cos(n\omega_0 t)dt.$$

Similarly, we obtain B_n by multiplying by sin $n\omega_0 t$ and integrating over a period. In conclusion, the coefficients of the Fourier series of the periodic function $f(t)$ are given by

$$A_n = \frac{2}{T}\int_{\tau}^{\tau+T} f(t)\cos(n\omega_0 t)dt; \quad B_n = \frac{2}{T}\int_{\tau}^{\tau+T} f(t)\sin(n\omega_0 t)dt \tag{2.45}$$

The second equivalent expression of the Fourier series is

$$f(t) = F_0 + \sum_{m=1}^{\infty} F_m \cos(m\omega_0 t + \phi_m), \tag{2.46}$$

where the coefficients are now the F_m (which are non-negative) and the $_m$. We immediately state these quantities in terms of the A_m and B_m as

$$F_0 = A_0,\ F_m = \sqrt{A_m^2 + B_m^2},\ \phi_m = -\arctan\frac{B_m}{A_m}. \tag{2.47}$$

The third equivalent form of the series, which we shall use very often, is obtained from Eq. (2.46), using the identity

$$F_m \cos(m\omega_0 t + \phi_m) = \frac{F_m}{2}e^{im\phi}e^{im\omega_0 t} + \frac{F_m}{2}e^{-im\phi}e^{-im\omega_0 t}$$

and defining the coefficients as

$$C_0 = F_0,\ C_m = \frac{F_m}{2}e^{i\phi_m},\ C_{-m} = \frac{F_m}{2}e^{-i\phi_m}. \tag{2.48}$$

Equation (2.46) becomes

$$f(t) = \sum_{m=-\infty}^{\infty} C_m e^{im\omega_0 t}, \tag{2.49}$$

Note that the sum now runs on all the integer numbers, not just the positive ones, and that the coefficients are complex numbers. Note also that the two coefficients corresponding to the opposite values of the index are the complex conjugates of one another. As immediately seen from their definitions, we have

$$C_{-m} = C_m^*. \tag{2.50}$$

We may also immediately verify that the complex coefficients C_m are given by

$$C_m = \frac{1}{T}\int_{\tau}^{\tau+T} f(t)\mathrm{e}^{-in\omega_0 t}dt. \tag{2.51}$$

We shall call the coefficients F_m the *Fourier amplitudes* and ϕ_m the *Fourier phases* of the Fourier series. They are, respectively, the modulus and the argument of the *complex Fourier amplitudes* C_m.

Finding the coefficient of the Fourier series of a given function is called Fourier analysis or harmonic analysis.

The sequence of Fourier amplitudes F_m of a time-dependent phenomenon is called the *amplitude spectrum* or simply the spectrum of the phenomenon. The periodic phenomena we are considering have a discrete spectrum. An example is shown in Fig. 2.15. We can appreciate the importance of the concept of an amplitude spectrum if we remember that the energy of the oscillating system is equal to the sum of the energies of its normal modes, namely of the energies of its Fourier components. The latter are proportional to the squares of the amplitudes F_m and do not depend on the phases. The energies of the Fourier components are completely determined by the amplitude spectrum of the function.

Let us now consider an example, which is important for our study in the subsequent chapters. Consider the function of time shown in Fig. 2.16a. It is a periodic sequence, with period T, of equal rectangular pulses of height L and length Δt ($\Delta t < T$). We assume the function to be exactly periodic, namely that the sequence of pulses extends through infinite time. Clearly, this is an idealized condition. Real situations will approach it the longer the duration of the sequence is compared to the period.

Taking advantage of the symmetry of the problem, we choose the origin of time in the center of a pulse. The function is an even one. The angular frequency is $\omega_0 = 2\pi/T$. We want to find the Fourier coefficients. The choice of the integration

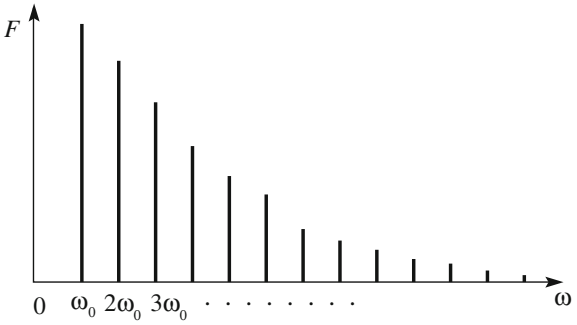


Fig. 2.15 Example of a spectrum of a periodic phenomenon

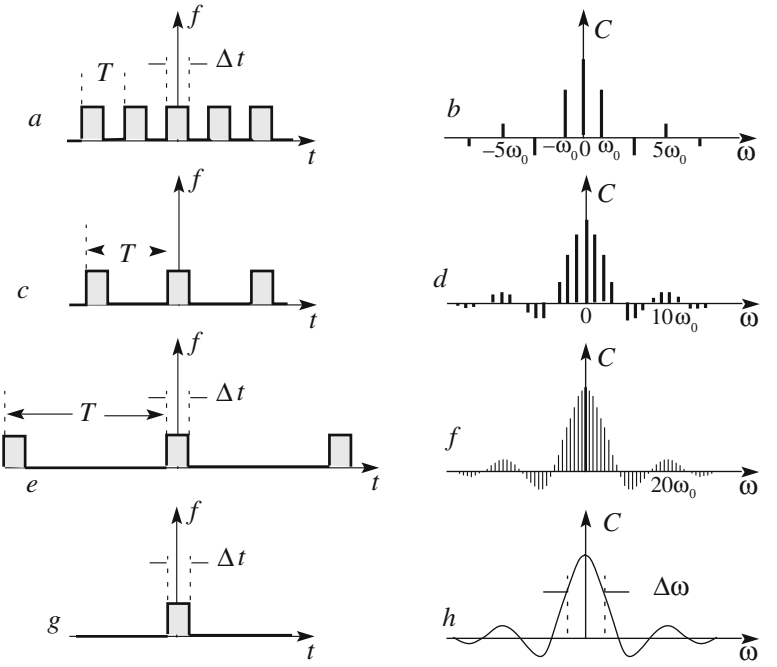


Fig. 2.16 An example of a periodic function and of its spectrum, for different values of the period

period being arbitrary, we choose the easier one, which is cantered on zero, namely $-T/2 \leq t \leq T/2$. In this interval, our function is different from 0 only in the interval $-\Delta t/2 \leq t \leq +\Delta t/2$, in which it is $f(t) = L$. The complex amplitudes are given by

$$C_m = \frac{1}{T}\int_{-\Delta T/2}^{\Delta T/2} L\mathrm{e}^{-im\omega_0 t}dt = -\frac{L}{T}\frac{\mathrm{e}^{-im\omega_0\Delta t/2} - \mathrm{e}^{+im\omega_0\Delta t/2}}{im\omega_0} = \frac{L\Delta t}{T}\frac{\sin(m\omega_0\Delta t/2)}{m\omega_0\Delta t/2}.$$

Let us analyze the result. The mth coefficient of the series is the product of a constant times a function. The constant is the product of the height of the pulse (L) and the ratio between its length (Δt) and the period (T). The function is $(\sin x)/x$, a function that we shall often encounter. To analyze its behavior, we start by observing that the quantity $m\omega_0$ in its argument is the mth angular frequency in the series. Let us call it $\omega_m = m\omega_0$. We then write

$$C_m = \frac{L\Delta t}{T}\frac{\sin(\omega_m\Delta t/2)}{\omega_m\Delta t/2}. \tag{2.52}$$

We note that, as a consequence of $f(t)$ being an even function, the coefficients are real and, for $m \neq 0$, $C_m = C_{-m}$. The coefficients are shown in Fig. 2.16b. We shall use the other parts of the figure in the next section. We recall that these coefficients are one half of the Fourier amplitudes F_m.

2.6 Harmonic Analysis of a Non-periodic Phenomena

The Fourier analysis can be extended to non-periodic functions. These functions can represent phenomena that are "periodic" only for a certain duration or that are not periodic at all, like the noise represented in Fig. 2.2.

Let us consider a function $f(t)$, which is different from zero within a definite time interval Δt (as are all the functions describing physical phenomena). Let us consider another function of time, which is equal to $f(t)$ during Δt and that repeats itself periodically with the same form in an arbitrary period $T > \Delta t$. Being that this new function is periodic, we can calculate its Fourier coefficients, write down its Fourier series and then look to see if we can find its limit for T going to the infinite.

Let us consider a very simple, non-periodic function, namely a rectangular pulse of height L and duration Δt. The corresponding periodic auxiliary function is represented in Fig. 2.16a. We have already calculated its Fourier coefficients. Let T now grow, keeping Δt fixed. Figure 2.16a, c, e represent the result for $T = 2\Delta t$, $T = 4\Delta t$ and $T = 8\Delta t$, respectively. Note that when T increases, the fundamental angular frequency $\omega_0 = 2\pi/T$ decreases in an inverse proportion. The abscissa of the diagram is the angular frequency ω. The m-th Fourier amplitude C_m is at the abscissa $\omega_m = m\omega_0$. Consequently, when ω_0 decreases, the amplitudes C_m get closer to one another, while Eq. (2.52) continues to hold, describing their envelope. In other words, when T varies, the values of the ω_ms vary as well, namely the positions on the abscissa at which the amplitudes are evaluated vary, but the dependence of the amplitudes on ω does not change. This is shown in Fig. 2.16b, d, f. It is then convenient to think of the Fourier spectrum, namely of the amplitudes C_m, as the values of a continuous function $C(\omega)$ evaluated in ω_m, namely as $C_m = C(\omega_m)$. In the limit $T \to \infty$ (Fig. 2.16g), the spectrum becomes a continuum function, namely the domain of the function $C(\omega)$ becomes defined on the entire real axis ω and not only at the discrete values ω_m (Fig. 2.16h).

In conclusion, for a non-periodic function, in place of an infinite discrete sequence of Fourier amplitudes, we have a continuous function of the angular frequency. Correspondingly, in place of the Fourier series, we have an integral. The integral is called the *Fourier transform*. We shall now give, without demonstration, the three equivalent expressions of the Fourier transform (analogous to the three expressions of the Fourier series in the previous section).

Let us start with the analogy of Eq. (2.43). In place of the discrete sequences of A_m and B_m, we now have two functions of the continuous variable ω, which we call $A(\omega)$ and $B(\omega)$, respectively. In place of the sums, we have integrals on ω on the entire domain that is $\omega \geq 0$,

$$f(t) = \int_0^\infty A(\omega) \cos \omega t \, d\omega + \int_0^\infty B(\omega) \sin \omega t \, d\omega. \tag{2.53}$$

The expressions for the two functions ω, $A(\omega)$ and $B(\omega)$ (analogous to Eq. (2.45)), are

$$A(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \cos \omega t \, dt, \quad B(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \sin \omega t \, dt. \tag{2.54}$$

Note that the domain is on the entire axis of time, from $-\infty$ to $+\infty$.

The second equivalent form, analogous to Eq. (2.46) in the periodic case, is

$$f(t) = \int_0^\infty F(\omega) \cos(\omega t + \phi(\omega)) d\omega. \tag{2.55}$$

The functions $F(\omega)$ and (ω) are given in terms of the $A(\omega)$ and $B(\omega)$ by the expressions

$$F(\omega) = \sqrt{A^2(\omega) + B^2(\omega)}, \quad \phi(\omega) = -\arctan \frac{B(\omega)}{A(\omega)}. \tag{2.56}$$

The third expression, analogous to Eq. (2.49), is

$$f(t) = \int_{-\infty}^{+\infty} C(\omega) \mathrm{e}^{i\omega t} d\omega. \tag{2.57}$$

The function $C(\omega)$ is complex and so equivalent to two real functions. Its expression, analogous to Eq. (2.51), is

$$C(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-i\omega t} dt. \tag{2.58}$$

The continuous complex function of the angular frequency ω, $C(\omega)$, is called the *Fourier transform* of the function of time $f(t)$. Twice its absolute value, namely $2|C(\omega)|$, is the *frequency spectrum*, which is a continuous rather than a discrete function, as in the case of the periodic functions. The function $f(t)$ itself, as given by Eq. (2.57), is called the *Fourier antitransform* of the function $C(\omega)$.

Note that the Fourier integral in the third form of Eq. (2.57) extends on both the positive and negative values of the angular frequency. This was already the case for periodic functions in the sum of Eq. (2.49). As we know, the angular frequency is

inversely proportional to the period, and is consequently a physically positive quantity. The negative values of appear as a consequence of the mathematical rearrangements we have made, but do not have a direct physical meaning. However, we shall see in the next section that the corresponding quantity for functions of space, which is the spatial frequency, does have a physical meaning both for positive and negative values.

Let us now go back now to the function in Fig. 2.16g, which is not only simple, but very important as well. Its Fourier transform is found using Eq. (2.58). The calculation is simple and completely analogous to what we did for the C_m. We shall not develop it, but rather go directly to the result, which is

$$C(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t) \mathrm{e}^{-i\omega t} dt = \frac{L}{2\pi} \int_{-\Delta t/2}^{+\Delta t/2} \mathrm{e}^{-i\omega t} dt = \frac{L\Delta t}{2\pi} \frac{\sin(\omega \Delta t/2)}{\omega \Delta t/2}. \tag{2.59}$$

Looking at Fig. 2.16h, we see that this function has its absolute maximum at $\omega = 0$. Two minima, where the function is null, are located symmetrically on the sides of the maximum. Going further, we encounter a succession of maxima and minima. The heights of the maxima decrease monotonically. We note that the most important components of the spectrum are located at low frequencies.

This is, indeed, a general characteristic of all the functions of time that have a beginning and an end, namely a finite duration. The most important part of their spectrum is situated in a region of lower frequencies, which we call *bandwidth*. The concept of bandwidth has a certain degree of arbitrariness, but the main argument for defining it is as follows. Consider a certain function of time $f(t)$ and calculate its Fourier transform $C(\omega)$. Let us then reasonably define a bandwidth $\Delta\omega$ and consider the function $C'(\omega)$, which is equal to $C(\omega)$ inside the bandwidth and 0 outside it. Let us antitransform $C'(\omega)$, obtaining, say, the function of time $f'(t)$. The result will not be exactly equal to $f(t)$, but the differences may be small enough for our purposes. If this is not the case, we need to define a somewhat wider bandwidth.

Coming back to the case under discussion, we now define the bandwidth $\Delta\omega$ as one half of the interval between the first two zeroes of $C(\omega)$, which is about the full width of the peak at half maximum (FWHM). The latter are at the values of ω for which $\omega \Delta t = \pm\pi$. We see that the bandwidth $\Delta\omega$ and the duration Δt are linked by the expression

$$\Delta\omega \Delta t = 2\pi. \tag{2.60}$$

This is an extremely important equation. Indeed, it is a particular case of a theorem of general validity, which we call the bandwidth theorem. We shall not demonstrate the theorem, but just give its statement, which is: *the bandwidth $\Delta\omega$ of the Fourier spectrum of a function limited in time to a duration Δt is inversely proportional to that duration.* The result is general, even if the specific coefficient (2π, in our example) depends both on the shape of the function of time and by our

specific definitions of the bandwidth and of the duration (which is somewhat arbitrary as well for functions that are not rectangular pulses). Equation (2.60) has very relevant consequences for optics, as we shall see in Sect. 5.3. Even more important are the consequences for quantum physics, where its equivalent is one of the fundamental laws, namely the uncertainty relation between energy and time.

In conclusion, the spectrum of a non-periodic function of time is a continuous function of the angular frequency; the spectrum of a periodic function of time is a function defined for discrete values of the frequency alone. In particular, the spectrum of a sinusoidal function of time has one component only, at the frequency of the sine.

As we shall see in the subsequent chapters, light is an electromagnetic wave. When the wave has a sinusoidal dependence on time, light has a definite color to our eyes. Contrastingly, in white light, there are waves of different frequency, continuously distributed over a wide range. As a consequence, an electromagnetic wave with sinusoidal dependence on time is said to be *monochromatic*, meaning a single color in Greek. By extensions, all the sinusoidal functions of time are often called monochromatic as well.

We shall now discuss two important examples.

Damped oscillation. Elastic oscillations in the presence of resistive forces proportional to velocity are quite common phenomena in different fields of physics. We have studied their equations in Sect. 1.2. We would now like to work with an acoustic oscillator vibrating on a single mode. The tuning forks used to tune the musical instruments have this property. They are U-shaped metal objects properly shaped to the purpose. Let ω_1 be the proper angular frequency of our tuning fork. We excite its vibrations by hitting one of its prongs with a small wooden hammer. Let us consider the displacement from equilibrium of one of its points as a function of time $\psi(t)$ and let ψ_0 be the initial displacement. We assume the air drag to be proportional to the velocity as

$$F_r = -m\gamma \frac{d\psi}{dt}.$$

As we know, the equation of motion for $t > 0$, namely after the oscillation has started, is

$$\psi(t) = \psi_0 \mathrm{e}^{-\frac{\gamma}{2}t} \cos \omega_1 t = \psi_0 \mathrm{e}^{-\frac{t}{2\tau}} \cos \omega_1 t, \tag{2.61}$$

where we have taken, as in Sect. 1.2.,

$$\tau = 1/\gamma$$

We recall that the time τ is the time interval in which the energy of the oscillator (which is proportional to the square of the vibration amplitude) decreases to $1/e$ of the initial value. The oscillation is not exactly sinusoidal, at angular frequency ω_1, as a consequence of damping. Consequently, its spectrum contains components at

angular frequencies different from ω_1. The spectrum is obtained by performing the integral in Eq. (2.58) with $f(t) = \psi(t)$. The limits of the integral are 0 and $+\infty$ because $\psi(t) = 0$ for $t < 0$. We also recall that $\omega_1^2 = \omega_0^2 - (\gamma/2)^2$, where ω_0 is the proper frequency in the absence of damping. Performing the Fourier transform integral and taking the square of the result, one finds for the square of the frequency spectrum the expression

$$F^2(\omega) = \frac{1}{(2\pi)^2} \frac{4\omega^2 + \gamma^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}. \tag{2.62}$$

We see that the denominator is the same as that of the resonance curves. If the damping is small, as is often the case, namely if it is $\gamma \ll \omega$, we can neglect the second term in the numerator and write

$$F^2(\omega) = \frac{1}{\pi^2} \frac{\omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}, \tag{2.63}$$

which, constant apart, is exactly the response function, or the resonance curve, of the oscillator in Eq. (1.64). The function $F(\omega)$ is shown in Fig. 2.17. We can then state that *the square of the Fourier transform of a damped oscillation is proportional to the response curve $R(\omega)$ of the resonance of the same oscillator when it is forced.*

Rigorously speaking, the spectrum extends over an infinite frequency range. However, the important contributions are those that are not very different from ω_0 within a certain bandwidth. We here define the bandwidth to be the full width at half maximum (FWHM) of the resonance curve $\Delta\omega_F$. Under this definition, recalling what we stated in Sects. 1.2 and 1.3, we can conclude that:

(a) the bandwidth of the Fourier transform of a weakly damped oscillator is equal to the width of the same oscillator when forced by a periodic external force;
(b) both widths are inversely proportional to the decay time τ of the free oscillations, namely the time in which the stored energy decreases to a value equal to $1/e$ of the initial value.

Namely, we have

$$\Delta\omega_{\text{ris}} = \Delta\omega_F = 1/\tau. \tag{2.64}$$

The bandwidth theorem requires defining the duration Δt of the phenomenon. Rigorously speaking, the duration would be infinite, but, in practice, the vibration is finished after several τ. Somewhat arbitrarily taking $\Delta t = 2\pi\tau$, we can write Eq. (2.64) as

$$\Delta\omega_F \Delta t = 2\pi. \tag{2.65}$$

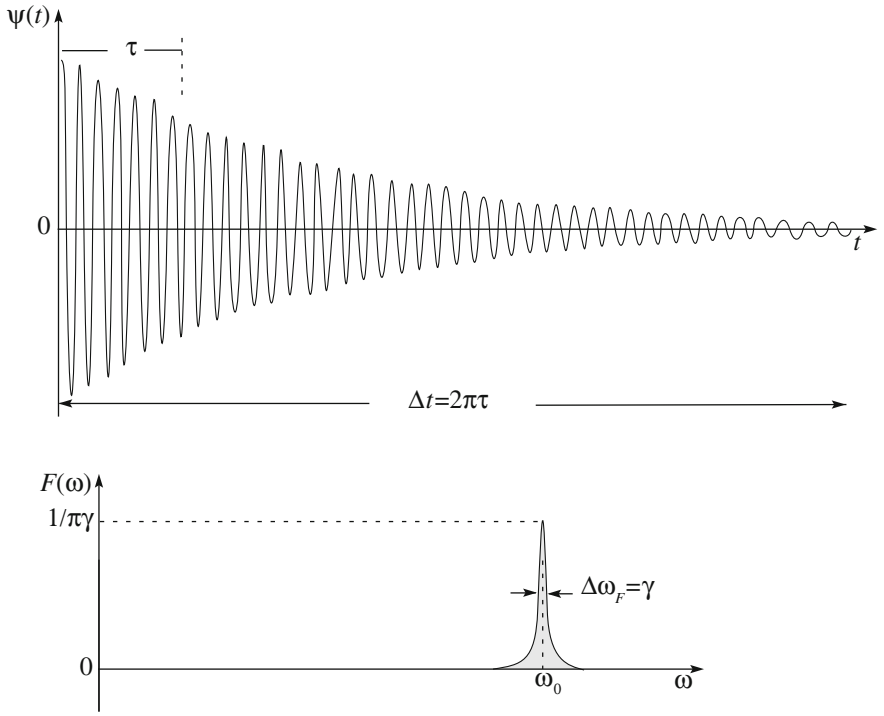


Fig. 2.17 A weakly damped oscillation and its Fourier transform

Even if the oscillators encountered in nature and technology are often damped to some degree, their damping is often small and their behavior approaches that of an ideal oscillator. The *quality factor*, or simply the *Q-factor*, is defined as the ratio between the resonance angular frequency and the FWHM of the resonance curve. The same quantity can be expressed in the following equivalent forms:

$$Q \equiv \frac{\omega_0}{\Delta\omega_F} = \frac{\omega_0}{\gamma} = 2\pi\frac{\tau}{T}. \tag{2.66}$$

Q-factors of mechanical oscillators are typically between several hundreds and several thousands (as in the case of a piano string), and up to millions or more for electronic oscillators.

The just found expressions have important experimental consequences. Indeed, while it is usually easy to measure the decay time of a macroscopic oscillator, for example, making a film of its vibrations, the same cannot be done for an atomic, or subatomic, oscillator. We can, however, enclose a number of atoms, say a gas, in a container and excite them, for example, with an electric discharge that force the ions, a few of which are always present, to violently collide with atoms. The electron cloud of each excited atom will then oscillate with a frequency and decay

time that are characteristic of the atomic species under study (as a matter of fact, the atom has several normal modes, each with a frequency and a decay time). The excited atoms emit the energy they have acquired as electromagnetic radiation. The decay time τ, called the lifetime of the excited atomic state, is an important quantity in atomic physics, but is usually too short to be directly measurable. However, if we measure the Fourier spectrum of the intensity of the emitted radiation, we obtain a resonance curve, whose width $\Delta\omega_F$ we can measure. We then have the lifetime from the relation $\tau = 1/\Delta\omega_F$.

Question Q 2.2. Calculate the Fourier transform of Eq. (2.59). □

Question Q 2.3. You hit the prong of a tuning fork of 440 Hz and hear a sound lasting about 30″. How much is its γ? How much is its Q-factor? □

Question Q 2.4. Consider an excited atom having a resonance angular frequency of 2×10^{16} s^{-1} and a lifetime of 2 ns. How much is the Q-factor? □

Question Q 2.5. The Q-factor of a harmonic oscillator of frequency 850 Hz is 7000. What is the time in which the amplitude reduces by a factor $1/e$? How many oscillations happen in this time? What is the difference between two consequent oscillation amplitudes relative to the amplitude itself? □

Question Q 2.6. With reference to Eq. (1.40), prove that the Q-factor is equal to 2π times $\langle U\rangle/(d\langle U\rangle/dt)$. □

Beats. A beat is a sound resulting from the interference between two sounds of slightly different frequencies. We perceive it as a periodic variation in volume whose rate of change is the difference between the two frequencies, say ω_1 and ω_2. The sound can be easily heard using two equal tuning forks, and altering the frequency of one of them by fixing a small weight to one of its prongs. If the weight is near the bottom of the prong, the change in frequency is quite small. If we now hit both forks, we hear the beat. If we fix the weight a little higher and repeat the experiment, we hear the volume of the sound periodically varying at a higher frequency. If we further increase the difference, we reach a limit in which the system of ear-plus-brain perceives the two sounds as separate. The limit depends on the person, usually being at about $\Delta\omega/\omega = 6$ %. The ear of a musician is substantially more sensitive.

Let us analyze the phenomenon. Assume, for simplicity, that the two vibrations have equal initial amplitudes and phases. The two displacements from equilibrium are then

$$\psi_1(t) = A\cos\omega_1 t, \quad \psi_2(t) = A\cos\omega_2 t.$$

The motion of our eardrum is proportional to the sum of these two functions, namely it is given by

$$\psi = \psi_1 + \psi_2 = A\cos\omega_1 t + A\cos\omega_2 t = 2A\cos\left(\frac{\omega_1 - \omega_2}{2}t\right)\cos\left(\frac{\omega_1 + \omega_2}{2}t\right).$$

As we stated, the beat happens when the difference between the two frequencies $\Delta\omega = \omega_1 - \omega_2$ is small in absolute value. Under these conditions, the average of

the two, namely $\omega_0 = (\omega_1 + \omega_2)/2$, is very close to each of them and we can write within a good approximation

$$\psi(t) = 2A\cos\left(\frac{\Delta\omega}{2}t\right)\cos(\omega_0 t). \tag{2.67}$$

We can think of this expression as describing an almost harmonic motion taking place at the mean frequency ω_0, whose amplitude varies slowly (and harmonically as well) in time with angular frequency $\Delta\omega/2$. The function is shown in Fig. 2.18.

Note that the perceived angular frequency of the modulation is $\Delta\omega$ and not $\Delta\omega/2$. The reason for this is that the ear is sensitive to the *intensity* of sound. The intensity is proportional to the square of the amplitude, namely to $4A^2\cos^2(\Delta\omega t/2) = 2A^2(1 + \cos\Delta\omega t)$. The constant term on the right-hand side is irrelevant; the angular frequency of the varying term is $\Delta\omega$.

Let us now consider the more general case in which the amplitudes of the two motions are different. Let us call them A_1 and A_2, and represent the motions as rotating vectors (Fig. 2.19). The vector sum of the two represents the resultant motion. If the oscillations have the same angular frequency ω, then all vectors rotate as a rigid structure with angular velocity ω. As a consequence, the magnitude of the resultant is constant over time and its x component makes a harmonic motion. Contrastingly, if the two frequencies are a little different, the angle between the vectors A_1 and A_2 slowly varies, and consequently the magnitude of their resultant slowly varies as well. From the figure, we see that the magnitude of the resultant is

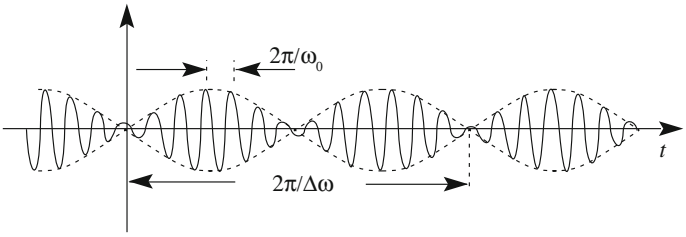


Fig. 2.18 The amplitude of a beat sound

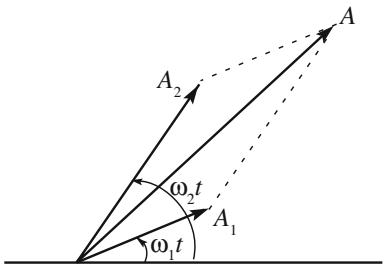


Fig. 2.19 Two rotating vectors and their resultant

$$A = \left[A_1^2 + A_2^2 + 2A_1A_2\cos(\omega_1 - \omega_2)t\right]^{1/2}. \tag{2.68}$$

2.7 Harmonic Analysis in Space

In this section, we shall consider the Fourier analysis as a function of space rather than of time. A function of space is, in general, a function of three variables, which are the coordinates. Such is the case with, for example, the temperature of a fluid. Examples of functions of two spatial coordinates exist as well, like the height of the waves on the surface of a lake or the sea and the gray levels of a photograph. For the sake of simplicity, we shall limit the discussion to functions of one space variable only. Such is the case with, for example, the configuration of a rope at a certain instant in time. Under these conditions, the mathematics of the spatial Fourier analysis is exactly the same as in the time domain we have considered in the previous sections.

Let us start by considering a periodic function of the coordinate x, which we call $f(x)$. As in the case of time, no physical system is described by a strictly periodic function of space, because no system exists of infinite dimensions. However, approximately periodic spatial structures are often encountered.

The period in space is the wavelength λ. The quantity corresponding to the frequency is the wave number, namely the number of wavelengths in one meter

$$v_s = 1/\lambda \tag{2.69}$$

and that, corresponding to the angular frequency, is the spatial frequency

$$k = 2\pi/\lambda = 2\pi v_s. \tag{2.70}$$

Let $f(x)$ now be a periodic function of x of period λ, and let $k_0 = 2\pi/\lambda$ be its fundamental spatial frequency. Clearly, we can express the Fourier series of $f(x)$ in any of the three forms we have seen for a function of time. We shall write down only the third one, which is, in complete analogy with Eq. (2.49),

$$f(x) = \sum_{m=-\infty}^{\infty} C_m \mathrm{e}^{imk_0x}. \tag{2.71}$$

The complex coefficients are given by the integral over a period, analogous to Eq. (2.51), as

$$C_m = \frac{1}{T}\int_{\xi}^{\xi+\lambda} f(x)\mathrm{e}^{-imk_0x}dx, \tag{2.72}$$

starting from any ξ.

If the function $f(x)$ is not periodic, then instead of a Fourier series, we have a Fourier integral, or a spatial Fourier transform. Analogous to Eq. (2.57), the transformation is

$$f(x) = \int_{-\infty}^{+\infty} C(k)\mathrm{e}^{ikx}dk. \tag{2.73}$$

The function $C(k)$, analogous to Eq. (2.58), is given by the Fourier spatial antitransform

$$C(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-ikx}dx. \tag{2.74}$$

We shall now come back to the gray level of a (black and white) picture, as an example of a space function. These are, in general, functions of two variables, say the coordinates x and y, but, for the sake of simplicity, we shall limit the discussion to functions of one coordinate only, say on x.

In the time domain, the Fourier analysis is the simplest for a sine function of time. The same is obviously true in the space domain as well. If the physical quantity described by the function is a gray level of a picture, it cannot have physically negative values. Let us consider the simplest case, namely

$$f(x) = \frac{1}{2} + \frac{1}{2}\cos kx. \tag{2.75}$$

The two constants on the right-hand side might have different values. We have chosen them so as to have the function vary between 0 and 1. The gray level and the function in Eq. (2.75) are shown in Fig. 2.20. Note that the gray level does not depend on y. This is obviously an idealization.

As shown in the figure, the wavelength is the distance between two homologous points, for example, between two consecutive maxima or minima. The wave number is the number of periods in one meter. The space frequency k is the wave number times 2π. The larger the spatial frequency, the closer the clear and dark bands are to one another. This structure is called sine grating in optics.

In the following example, we consider an infinite succession of completely black rectangular bands of width D repeating with a period in x equal to λ. Such successions are called Fraunhofer gratings in optics. This situation, as shown in Fig. 2.21, is again idealized, because neither the succession nor the length of the bars in y can be infinite. Being that the function is periodic, it admits a Fourier series, which is given by Eq. (2.71) with $k_0 = 2\pi/\lambda$. This function of x is identical to the function of time we considered in Sect. 2.5. Consequently, we already know the

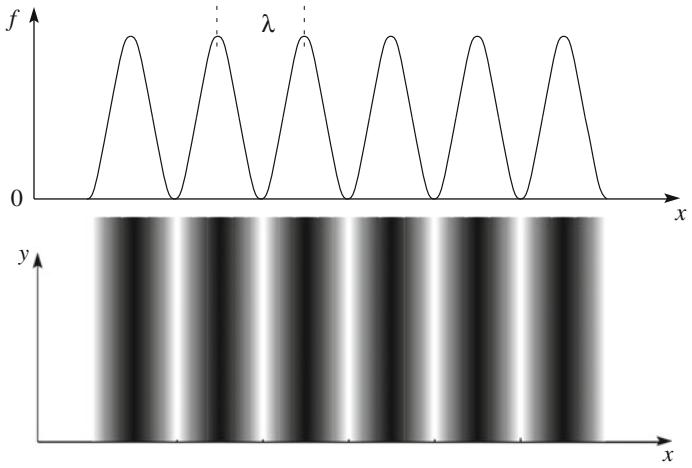
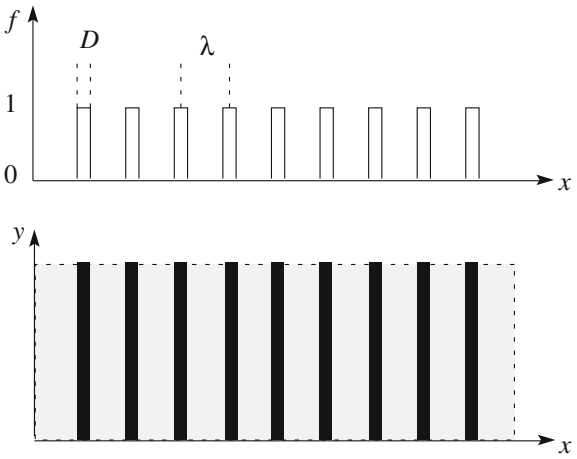


Fig. 2.20 A sinusoidal function of one space coordinate

Fig. 2.21 A periodic spatial grating



coefficients of the series, which are those of Eq. (2.52), with the obvious changes in the variables, namely

$$C_m = \frac{D}{\lambda} \frac{\sin(k_m D/2)}{k_m D/2}. \tag{2.76}$$

Let us now move to a non-periodic function of x, namely a single black band of width D, as shown in Fig. 2.22. Being that the function is not periodic, we must consider the integral of Eq. (2.73). In this case too, this is the spatial analogy of the rectangular function of time that we considered in Sect. 2.6. Consequently, we

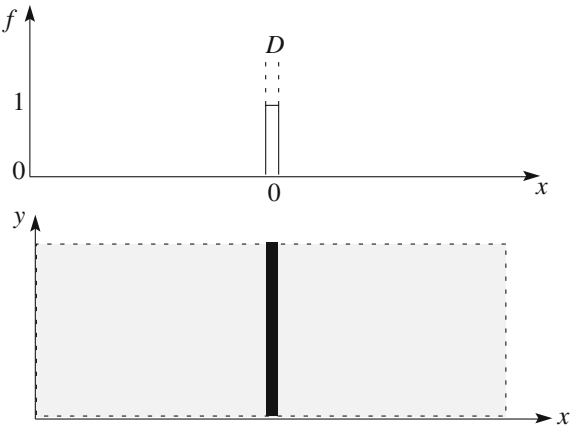


Fig. 2.22 A single black band

already know the Fourier transform in Eq. (2.74) of our function, which we obtain from Eq. (2.59). Changing the variables as needed, we obtain

$$C(k) = \frac{D}{2\pi} \frac{\sin(Dk/2)}{Dk/2}. \tag{2.77}$$

We shall return to this equation when we study the diffraction phenomenon of a grating in optics in Sect. 5.9.

The spatial Fourier transform in Eq. (2.78) is shown in Fig. 2.23, which is obviously the analogy in space of Fig. 2.16h in the time domain. The abscissa is now the spatial frequency k in place of the angular frequency ω.

Clearly, the bandwidth theorem holds in the space domain, as it does in the time domain. In this example, the function $f(x)$ is different from zero in a limited interval, which is $\Delta x = D$. The most important part of the Fourier transform has a certain width Δk, which we define similarly to what we did for the angular frequency (see Fig. 2.23) The relation between them is

$$\Delta k \cdot \Delta x = 2\pi. \tag{2.78}$$

Namely, the narrower the space function, the larger the frequency interval needed to represent it in the Fourier transform. This relation has important

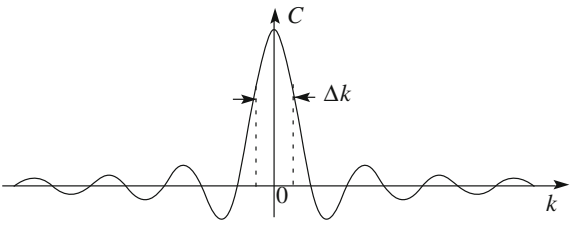


Fig. 2.23 The spatial Fourier transform of the function in Fig. 2.22

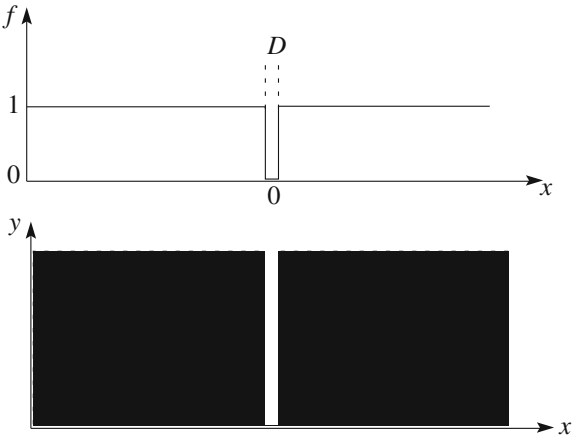


Fig. 2.24 A slit

consequences in optics, which we shall discuss in Chap. 5. Even more important are the consequences for quantum physics, where its equivalent is one of the fundamental laws, namely the uncertainty relation between momentum and position.

Let us finally note that Eq. (2.78) is, a sign apart, the Fourier transform of the "gray level" function shown in Fig. 2.24, which is the same as Fig. 2.22 with inverted blacks and whites. But we can also think of Fig. 2.24 as a screen that completely absorbs an incident light beam, except for the part in the white band, which can go through. This can be obtained, for example, by opening a slit in an absorbing screen. We shall discuss this phenomenon in Chap. 5, where we shall see, in particular, that the light intensity beyond the slit depends on the coordinates as the square of the function shown in Fig. 2.23. What we shall learn to be the (Fraunhofer) diffraction pattern of the slit is the physical materialization of the mathematical concept of the spatial Fourier transform.

Summary

The most important concepts studied in this chapter are the following:

1. Two coupled harmonic oscillators can move in two particular motions, the normal modes, in which all the parts of the system oscillate with the same frequency and in the same initial phase. The oscillation frequencies of the modes are the proper frequencies of the system. The proper frequencies and the shapes of the modes depend on the structure of the system, but not on the initial conditions.
2. Particular coordinates can be found, called the normal coordinates, in which the equations of motions are decoupled.
3. The equations of motion of two forced coupled oscillators are independent when written in normal coordinates. Two resonances exist at the two proper frequencies.

4. Coupled oscillating electrical and mechanical systems obey the same differential equations and behave in analogous manners.
5. The above properties can be generalized to systems with n degrees of freedom.
6. An oscillating string with fixed extremes is a system with an infinite continuous number of degrees of freedom. Its motion is described by an important partial differential equation.
7. The normal modes of the vibrating elastic string have the following characteristics: (a) all the points of the string oscillate with the same amplitude and in the same phase, (b) each mode has its own (proper) oscillation frequency, dependent on the structure of the system and independent of the initial state, (c) the shape of the modes are sine functions, whose period is the wave length.
8. The relation between the angular frequency and the wave number of the modes is the dispersion relation. The dispersion relation is linear for a perfectly elastic string. In this case, the proper frequencies are in the arithmetic succession.
9. A periodic function, both of time and of a space coordinate, can be expressed as a sum of cosine functions and in other equivalent forms.
10. A non-periodic function can be expressed as an integral, performing a Fourier transform. The Fourier transform of a function of time (space coordinate) is a function of the angular frequency (space frequency).
11. The bandwidth of the Fourier transform of a function of time of limited duration is inversely proportional to that duration. The bandwidth of the Fourier transform of a function of a space coordinate extending throughout a limited interval is inversely proportional to that interval.
12. The width of the resonance curve of a weakly damped forced oscillator is equal to the width of the Fourier transform of the displacement as a function of time in the free oscillations of the same oscillator. Both widths are inversely proportional to the decay time of the free oscillations.

Problems

2.1 Consider two coupled pendulums, such as those discussed in Sect. 2.1. Suppose that the squares of their proper angular frequencies differ by 4 s^{-2}. If we change the spring with another one having a spring constant 10 times smaller, how much is the new difference?

2.2 Consider the two coupled pendulums again. Does the proper frequency of the lower frequency mode depend on the spring constant? Does it depend on the masses? Answer the same questions for the higher frequency mode.

2.3 A string of a viola is 0.5 m long and is tuned 440 Hz. You can play it at 550 Hz, pressing it with your finger to make it shorter. How much shorter should we make it?

2.4 A string of an instrument should be tuned to 440 Hz, but instead is at 435 Hz. How should we change its tension to tune it?

2.5 The width of a forced oscillator is $\Delta\omega_{ris} = 35\ s^{-1}$. We let it freely oscillate and we measure its displacement as a function of time. We then Fourier transform the function we have found. What is the bandwidth of this transform?

2.6 Consider the motion $\psi(t) = (10\,\text{mm})\cos\left[(6.28\text{s}^{-1})t + 32^\circ\right] + (15\,\text{mm})\sin\left[(6.21\text{s}^{-1})t - 72^\circ\right]$. Calculate the mean frequency and the beat frequency

2.7 The bandwidth of the Fourier transform of a function of time is 120 s^{-1}. How much is the duration of the function?

Chapter 3
Waves

Abstract In this chapter, we introduce the concept of the progressive wave, first in a one-dimensional medium, then in 3D space. We focus on sound and electromagnetic waves. We discuss the sources of electromagnetic waves, which are ultimately accelerating charges, presenting an expression of the electric field they generate, in an approximation valid at large distances from the source. This is a very useful expression that we shall often use in subsequent sections. We discuss the quadratic-law detectors and the concept of intensity for sound and electromagnetic waves. Finally, we study how the frequency and wavelength of a wave depends on the motion of the source and of the detector (the Doppler effect).

In the previous chapter, we found that the motions of a perfectly elastic string obey a partial differential equation, which is called a wave equation. We then considered the oscillations of a string of limited length and with defined boundary conditions, in particular, with fixed extremes. Under these conditions, the oscillation modes of the string are called stationary waves. In this chapter, we begin our study of the properly-named waves, namely those that advance, or propagate, in a medium, called the progressive waves. Waves are a very common natural phenomenon, appearing in a number of different forms. They are the big waves in a stormy sea, the tiny ripples on a pond, the sounds propagating through the air, the pulses in a nerve. There are seismic waves, light waves, electromagnetic television waves, etc.

When a wave propagates in a medium, there is no matter traveling along with it. For example, in a sound wave, the different points of the medium oscillate back and forth around fixed positions, while in the waves of the sea, the water particles describe closed trajectories about their equilibrium positions, and so on. What propagates with a wave is the disturbance of the medium. The medium exists only as a support for the wave, and is not even necessary. Indeed, no material medium carries electromagnetic waves. They even travel in a vacuum.

What we have called disturbance means that some physical quantity characteristic of the medium (or of the electromagnetic field) assumes values different from those at equilibrium. For example, this quantity may be the pressure of the gas in a

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_3

sound wave, the intensity of the electric field in an electromagnetic wave, the displacement from the equilibrium of a water element, etc. Propagation of the disturbance means that the change from equilibrium affects points at distances increasing with time. Indeed, waves propagate with a definite velocity. Namely, the distance traveled by the disturbance in a certain time is proportional to that time. A wave is not a simple object, like a material point. Consequently, a precise definition of the wave velocity is not as straight forward as for a material point. As a matter of fact, we shall define not one, but several different wave velocities: the velocity of a pulse, the phase velocity and the group velocity.

The velocities of different types of wave span many orders of magnitude. For example, the waves on the surface of water travel on the order of several meters per second, the sound in the atmosphere at about 300 m/s, and light and gravitational waves, the fastest of all, at about 300,000 km/s.

Propagation of a disturbance implies propagation of information. Indeed, waves are the principal vehicle for information. Just think of the sound, of the electromagnetic waves of TV and cell phones, of the light, of the electric waves in cables.

In addition, waves carry energy. For example, the light from the sun brings the energy necessary for life to earth. Its intensity is about 1 kW/m^2. A fraction of this energy is absorbed by the vegetation, the trees, and us when we burn wood or coal. We can also use it to produce electric power with photovoltaic cells. TV stations radiate power on the order of several dozens of kilowatts. Sound waves have relatively small intensities, for example, on the order of l0 nW/m^2, while a seismic wave can transport a huge quantity of energy. In addition, waves carry linear and angular momentum, but we shall not deal with these aspects in this book.

We shall always consider situations in which the displacements from equilibrium are relatively small. Under these conditions, the systems behave approximately linearly (exactly so for electromagnetic waves). This means that the differential equation ruling the system is linear and that the superposition principle holds. Consider, for example, the waves on the surface of water. If their amplitude is small, the system is linear. To control the validity of the superposition principle, we just have to toss two stones into a quiet pond and observe the waves. Looking carefully, you understand that there are two systems of expanding circles, each centered at the points where the stones touched the water. The waves cross one another without disturbing each other. As another example, think of the feeble light of a star that we see in the night. Before reaching us, it has crossed the intense light of the sun without being altered.

Examples of non-linearity are everywhere as well. They appear whenever the displacements from equilibrium are large. Examples are the sound waves from an explosion, the big waves in the ocean that make surfers happy and the light from a powerful LASER. In this course, we shall consider only linear conditions.

In the first three sections, we introduce the concept of a progressive wave in a one-dimensional medium, which will be an elastic string. We study how a wave reflects at an extreme of the string. We then go to the progressive waves in space, namely in three dimensions, with a focus on the important cases of sound and electromagnetic waves. In Sect. 3.7, we discuss the sources of electromagnetic

waves, which are ultimately accelerating charges. We present and discuss an expression of the electric field they generate, according to R. Feynman, which is an approximation valid at large distances from the source and for source velocities much smaller than light. This is a very useful expression that we shall use often in the subsequent sections. Then, we discuss a category of frequently used detectors, namely the quadratic-law detectors, whose response is proportional to the square of the disturbance. In Sects. 3.9 and 3.10, we deal with wave intensity and the energy transport of sound and electromagnetic waves. In Sect. 3.11, we discuss the equation and properties of a very common transmission line, namely the coaxial cable.

Finally, in Sect. 3.12, we study how the frequency and wavelength of a wave depend on the motion of the source and the motion of the detector. This is the Doppler effect, which will be analyzed through the examples of sound and electromagnetic waves. This will show us the profound differences between the two cases.

3.1 Progressive Waves

In Sect. 2.3, we studied the steady oscillations of an elastic string. We have seen that the partial differential equation ruling its motions is the equation

$$\frac{\partial^2 \psi}{\partial t^2} - v^2 \frac{\partial^2 \psi}{\partial z^2} = 0, \tag{3.1}$$

where $\psi(z, t)$ is the wave function, which is the displacement from equilibrium at time t of that point of the string that is located at z when in equilibrium and

$$v^2 = T_0/\rho. \tag{3.2}$$

Here, T_0 is the tension and ρ is the linear density of the string. Equation (3.1) is often simply called the *wave equation*.

We also recall that in Sect. 2.3, we found those particularly important motions of the elastic string with fixed extremes, which are the normal modes. As a matter of fact, normal modes, or stationary waves as they are also called, also exist when both extremes are free to move or when one is fixed and the other is free. The important feature that is necessary for a stationary wave is the existence of boundary conditions. The system must have boundaries, which are the extremes in the case of the string.

In this section, we shall study the opposite case, namely the oscillations of an open system, for example, an infinitely extended string. Suppose we apply a perturbation at a point of our infinitely long string. The perturbation immediately propagates along the string, in both directions, in the form of waves that move farther and farther as time goes by. If boundaries were present, these waves, reaching an extreme, would reflect and come back, interfering with the progressive waves. The resulting, initially quite complicated situation will evolve towards a

stationary oscillation. If the string is open, we need to consider only the primary waves generated by the applied force. As another example, consider a sound source, a trumpet, for example. Boundaries of the medium in which the sound waves propagate are any walls, the floor and the ceiling. In an open medium, these boundaries are absent or, at least, very far away. Another example is a wave on the surface of an infinitely extended water surface.

In practice, infinitely extended media do not exist. However, what really concerns us here is the absence of reflections. This condition can also be realized in a non-infinite medium. For example, in a theater, sound reflections are avoided by placing heavy tents in the proper locations, and at the borders of a pond, waves may be absorbed by thick vegetation living in the water.

Let us consider, for simplicity, a one-dimensional medium, our elastic and flexible string, and let us limit the discussion to completely transversal and linearly polarized motions, as we did previously in Sect. 2.3. This system is ruled by the differential equation in Eq. (3.1), but we shall not need it. Indeed, we shall now define the concept of the progressive wave, or simply wave, independently of the differential equation ruling the system. Namely, we want to establish the characteristics that the wave function $\psi(z, t)$ should have in order to be considered a progressive wave.

Let us suppose that we take a shot of the moving string at a certain instant t, finding the continuous curve shown in Fig. 3.1. Let us take another shot after a time interval Δt, namely at the instant $t + \Delta t$. If the movement is that of a progressive wave, we must find the same curve translated by a certain distance Δz, as shown dotted in the figure. In addition, this translation must be proportional to Δt if the entire wave has to move with a definite speed. Let us call it w.

Let us formally analyze what we have just stated verbally. We want the function ψ to have the same value for t and z when we substitute $t + \Delta t$ for t and $z + \Delta z$ for z at the same time, namely it should be $\psi(z + \Delta z, t + \Delta t) = \psi(z, t)$, where Δz and Δt are proportional to one another, namely it should be $\Delta z = w\Delta t$. This can only happen if ψ does not depend on z independently of t; rather, it should be of the type

$$\psi(z, t) = f(z - wt). \tag{3.3}$$

Indeed, for such a function, we have

$$\psi(z + \Delta z, t + \Delta t) = f(z + \Delta z - wt - w\Delta t) = f(z - wt) = \psi(z, t).$$

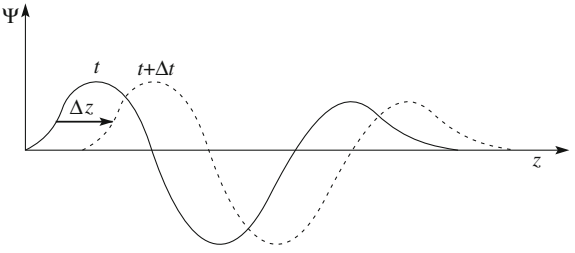


Fig. 3.1 Snapshot of a wave at time t (*continuous*) and $t + \Delta t$ (*dotted*) with $\Delta z/\Delta t = w$

Note that $f(z - wt)$ is a function of one variable only.

We have checked that $\psi(z, t)$ should obey Eq. (3.3) to represent a progressive wave advancing without deformations. Under these conditions, $\psi(z, t)$ is also a solution to the wave equation in Eq. (3.1) for whatever function f. Let us check through direct substitution. Calling f'' the second derivative of f, we have $\partial^2 f/\partial t^2 = w^2 f''$ and $\partial^2 f/\partial z^2 = f''$, and, by substitution in Eq. (3.1), we obtain $w^2 f'' - v^2 f'' = 0$. This equation is satisfied by any function f, provided that

$$w = \pm v. \tag{3.4}$$

First of all, this means that the physical meaning of the constant v in the wave equation is that it is the wave velocity. Second, the result we have found shows that there are two possibilities, with velocities equal in absolute value but opposite in sign. We can say that, beyond solutions of the type $f(z - vt)$ from which we started, solutions of the type, say, $g(z + vt)$, where g is also an arbitrary function, exist as well. The former waves advance in the positive direction of z (progressive wave), the latter move in its negative direction (regressive wave).

One can show that the most general solution to Eq. (3.1) can be expressed as the sum of two properly chosen solutions, one being progressive and one regressive, namely as

$$\psi(z, t) = f(z - vt) + g(z + vt). \tag{3.5}$$

The latter is neither a progressive nor a regressive wave, in general. We can obviously also write it as

$$\psi(z, t) = f(t - z/v) + g(t + z/v).$$

Let us go back to considerations independent of whether or not the differential equation is Eq. (3.1), and consider the relevant case of a progressive wave being a circular function. We call it a *monochromatic wave* or also a *harmonic wave*. Its wave function can be written as

$$\psi(z, t) = A \cos \omega(t - z/v). \tag{3.6}$$

We see that this equation describes a cosine (or sine) function moving in time with constant velocity v. Taking into account that ω and k are linked by the relation $k = \omega/v$, we can write the equation as

$$\psi(z, t) = A \cos(\omega t - kz). \tag{3.7}$$

Remembering that the phase is the argument of the cosine, we immediately see that ω represents the phase advance per unit time and k the phase advance per unit length in the direction of motion of the wave.

We recall that, in a stationary wave at a given instant, all the points on the string have the same oscillation phase. Contrastingly, in a progressive wave, the phase of the points of the string increases linearly with z. More precisely, considering that the phase is the function

$$\phi(z,t) = \omega t - kz \tag{3.8}$$

let us ask ourselves what the velocity is at which the phase advances, which is the *phase velocity*, and which we indicate with v_p. This is the velocity of any point of the wave, for example, a maximum. Namely, let ϕ be the phase at the instant t of the point at z and let dz be the increment of z such that, at the instant $t + dt$, the wave function of the point at $z + dz$ has the same phase ϕ. The situation is shown in Fig. 3.2. Then, the phase velocity is dz/dt. In other words, the phase velocity is $(dz/dt)_{\phi=\text{constant}}$. Consequently, we find the phase velocity imposing on the total differential of the phase to be zero, namely

$$d\phi = \frac{\partial \phi}{\partial t} dt + \frac{\partial \phi}{\partial z} dz = \omega dt - k dz = 0,$$

which gives us

$$\left(\frac{\partial z}{\partial t}\right)_{\phi=\text{const}} = \frac{\omega}{k} = v_p. \tag{3.9}$$

This result is the phase velocity for whatever the differential equation ruling the system may be, because we did not make any assumptions about it. We repeat that one can only speak of phase velocity if a phase exists. Being that the phase is the argument of a circular function, phase velocity makes sense only for monochromatic waves. It does not make sense for waves that have different shapes.

If the equation ruling the system is Eq. (3.1), then the phase velocity is also

$$v_p = \frac{\omega}{k} = v. \tag{3.10}$$

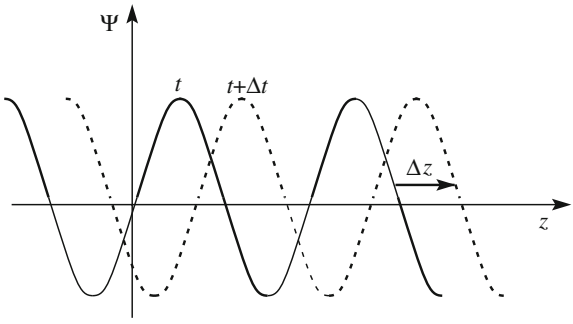


Fig. 3.2 Snapshot of a sine (*harmonic*) wave at time t (*continuous*) and $t + \Delta t$ (*dotted*) with $\Delta z/\Delta t = v$

We see that the phase velocity, in this case, is equal to the constant v in the wave equation.

Let us consider a progressive harmonic wave advancing in an arbitrary medium, not necessarily obeying Eq. (3.1). It might be, for example, an imperfectly flexible and stiff string. Two possible situations exist, namely the phase velocity may or may not depend on the wave number. The medium is said to be non-dispersive in the second case, dispersive in the first.

Being that, in any case, the phase velocity is the ratio between angular frequency ω and wave number k, the non-dispersive case is when this ratio is a constant, namely when the dispersion relation is Eq. (2.29). This equation can also be written in the form

$$\frac{\omega}{k} = v = \text{constant}. \tag{3.11}$$

We see that systems described by the wave equation in Eq. (3.1) are not dispersive. Examples of such systems, beyond that of the elastic string, are the sound waves in air, the light in a vacuum, etc. In all these cases, the phase velocity is independent of k and equal to the constant v, appearing, squared, in the wave equation. Clearly, the expression of v in terms of the physical characteristics of the medium is different in the different cases.

When the dispersion relation is not given by Eq. (3.11), as is often the case, the medium is dispersive. We gave an example of that in Sect. 2.3 with a piano string that is not perfectly flexible. In a dispersive medium, the harmonic waves with different wavenumbers k (or, as we can also say it, with different wavelengths) travel at different speeds. Note that in dispersive cases, the wave equation cannot be Eq. (3.1).

Even if the harmonic waves move in a dispersive medium at different speeds for different wavenumbers, they still propagate, keeping their shape unaltered. Contrastingly, non-harmonic waves change their shape as they propagate in a dispersive medium. The very concept of the progressive wave needs to be extended in comparison to the definition we gave in the previous section. As a matter of fact, the concept cannot be defined very precisely. In practice, however, the dependence of the phase velocity on the wavenumber is usually quite modest. For example, a white light pulse traveling in a dispersive medium like water does not separate into its different colors while it propagates. Contrastingly, the longer ocean waves produced by a storm in the open sea reach shore before the shorter ones.

In this chapter, we shall study the main properties of non-dispersive waves. In the subsequent one, we shall discuss a few examples of waves in dispersive media, in particular, light in a material medium. We shall see how, even if the differential equation of the system is unknown, the knowledge of the dispersion relation will be enough to understand the physical processes.

3.2 Production of a Progressive Wave

In this section, we study how to produce a progressive wave. Concretely, we consider an elastic string, but our arguments have a general character. We consider purely transversal and linearly polarized motions. Let T_0 be the tension and ρ the linear mass density of the string. We assume the string to be semi-infinite and take a z-axis along its position at rest, with the origin in its extreme. We can inject a progressive wave into the string acting on an extreme by moving it up and down. This action requires applying a force on the extreme. Let us consider using the device shown in Fig. 3.3. If $\psi(z, t)$ is the wave function, the displacement of the extreme at time t is $\psi(0, t)$ and the slope of the string in the origin, which is the direction of the tension there, is $(\partial\psi/\partial t)_{z=0}$. Hence, the component of the tension in the direction of the motion of the extreme is $T_0(\partial\psi/\partial t)_{z=0}$. The force to be applied must be equal to and opposite of this. Indeed, being that the mass of an element of the string is infinitesimally small, the resultant force must be zero so as not to have an infinite acceleration.

Now, $\psi(z, t)$ represents a progressive wave. Hence, it must be of the type $\psi(z, t) = f(z - \upsilon t)$, for which $\upsilon = \sqrt{T_0/\rho}$ is the wave velocity.

This condition implies a relation between the partial derivatives of $\psi(z, t)$. In particular, at the extreme, we must have

$$\left(\frac{\partial\psi}{\partial z}\right)_{z=0} = -\frac{1}{\upsilon}\left(\frac{\partial\psi}{\partial t}\right)_{z=0}, \tag{3.12}$$

which must be satisfied at any instant in time. The externally applied force must then be

$$F_e(t) = -T_0\left(\frac{\partial\psi}{\partial z}\right)_{z=0} = \frac{T_0}{\upsilon}\left(\frac{\partial\psi}{\partial t}\right)_{z=0} = \sqrt{T_0\rho}\left(\frac{\partial\psi}{\partial t}\right)_{z=0}.$$

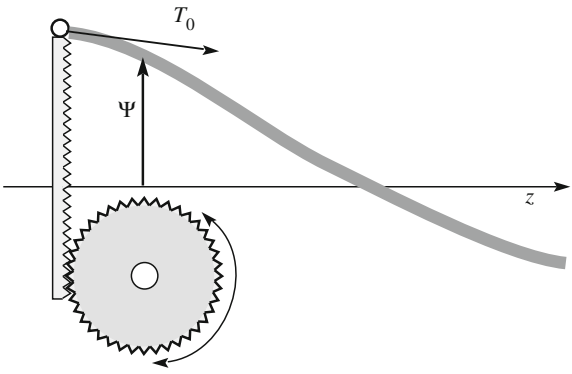


Fig. 3.3 Injecting a progressive wave into an elastic string

We see that the applied force is proportional to the velocity of the application point, namely of the extreme. Let us call it w_0. The proportionality constant

$$Z = \sqrt{T_0 \rho} \tag{3.13}$$

is a characteristic of the medium, the string in this case, and is called the *characteristic impedance* of the medium. We see that the two physical characteristics, tension and linear mass density, determine the two quantities, wave velocity $v = \sqrt{T_0/\rho}$ and characteristic impedance $Z = \sqrt{T_0 \rho}$, that completely characterize the waves' propagation.

Coming back to the external force, as we noticed, it is proportional to the speed of its application point, namely it is

$$F_e = Z w_0. \tag{3.14}$$

This force acts against the resistance of the string that is equal to $-Zw_0$. Forces proportional to velocity and opposite to it absorb energy from outside the system. Indeed, when we generate the wave, the elements of a section of the string of increasing length acquire both potential and kinetic energy. This energy must be injected into the system by an external agent, which is the force in this example.

3.3 Reflection of a Wave

Let us now assume the string to have a second extreme and look at what happens when it is reached by the front of the wave we have generated. The phenomenon is the reflection. The type of reflection depends on how the extreme is constrained.

Let us start by assuming the extreme to be fixed to a body of very large mass, which is immovable. Let us choose the origin of the coordinate z in this extreme and let the string lay on the negative z-axis. Let us consider the progressive wave $f(z - vt)$ approaching the extreme moving along the z-axis. When the wave reaches the extreme, the function f is, in general, different from zero, at least for some value t. Under these conditions, the extreme would move. As this does not happen, the wave function must have, in its general form, the expression $\psi = f(z - vt) + g(z + vt)$, with the second term also being present. Better still, this term is fully determined by the condition $\psi(0, t) \equiv 0$, namely that $g(vt) = -f(-vt)$). We can say that g is the opposite of f as evaluated at the opposite value of the argument. In conclusion, we have

$$\psi(z, t) = f(z - vt) - f(-z - vt). \tag{3.15}$$

We can think of the phenomenon visually, as shown in Fig. 3.4. We imagine having an imaginary string on the positive z-axis and a regressive wave moving toward the origin. This wave is f inverted and reflected, as stated above. In the physical region $z < 0$, namely the place where the real string is located, the motion

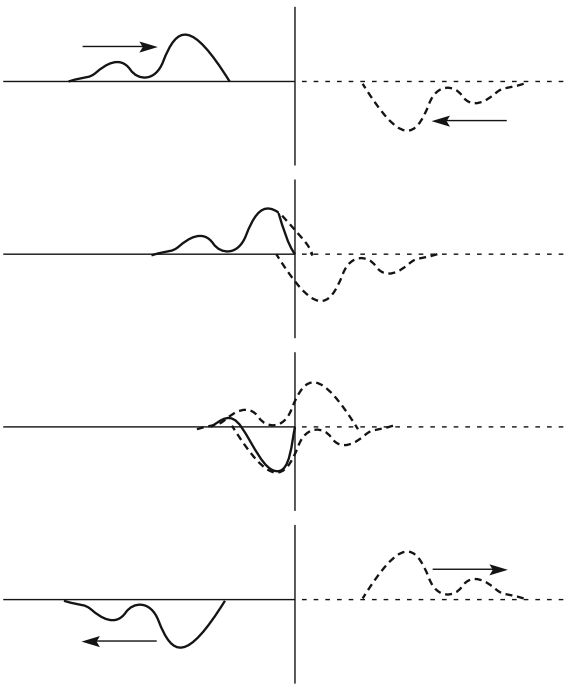


Fig. 3.4 Reflection of a pulse wave

is the sum of the two waves, the real one and the imaginary one. They propagate in opposite directions, in such a way as to cancel one another out at $z = 0$ at every instant in time.

Let us now look at what happens when the incoming wave is sinusoidal, namely when it is

$$f(z - vt) = A \cos(\omega t - kz).$$

Recalling that the cosine is an even function, the reflected wave is

$$-f(z - vt) = -A \cos(\omega t + kz),$$

which we can also write as

$$-f(z - vt) = A \cos(\omega t + kz + \pi).$$

We can then state that the reflected wave is a regressive wave that we can consider to be the incident wave coming back after having had a phase jump up by π in reflection. We shall see this situation repeating for several types of waves.

When the incident and reflected waves are present, the resultant displacement function is

$$\psi(z,t) = A[\cos(\omega t - kz) - \cos(\omega t + kz)] = 2A \sin kz \sin \omega t.$$

We see that all the points of the string, at whatever z, oscillate harmonically with the same angular frequency ω and in the same initial phase. The combination of incident and reflected waves is a stationary wave.

As a matter of fact, quite similar conclusions are reached if the extreme is free rather than fixed. To have the extreme be free and the string under tension, we can attach a small ring to the extreme and have it be free to move up and down on a vertical post fixed to an immovable body. Under these conditions, and neglecting friction between ring and post, the vertical component of the tension at the extreme is zero, and consequently the boundary condition is $(\partial\psi/\partial z)_{z=0} = 0$. With arguments very similar to those for the fixed extreme, it is easy to find that a wave pulse reflects without inversion and a sinusoidal wave reflects without changing its phase. In the latter case, the combination of an incoming and a reflected wave is still a stationary wave.

Question Q 3.1. Prove the above statements. □

We now make an observation that will be useful in the subsequent sections. In the case we considered of a fixed extreme, the string was connected to an object of very large, ideally infinite, mass that could not move. Let us now consider the case in which the farther extreme of our string is connected to the first extreme of a second string of mass density larger than the first one. Consider the propagation of harmonic waves. The tensions of the two strings are equal. Consequently, the wave speed on the second string is smaller than that on the first one. We can think of the condition of the fixed extreme as being the limit of this one when the density of the second string goes to infinity. In the present case, we state without proving that when the wave on the first string reaches the junction, it partially reflects back and partially continues on the second string, with the corresponding velocity. The transmitted wave, as it is called, does not have any phase change at the joint, while the reflected wave has a change of phase of π.

Let us now assume the density of the second string to be smaller than that of the first. We can think of the free extreme as being the limit of this situation when the density of the second string vanishes. Now, the wave speed on the second string is larger than that on the first. In this case too, a progressive harmonic wave at the junction is partially transmitted and partially reflected. Neither the transmitted nor the reflected wave has a change of phase at the junction. We shall come back to this in Sects. 4.6 and 5.4.

Let us now consider the following question: what are the conditions to impose to the far end of the string to avoid any reflection? Consider a progressive wave traveling on an elastic string stretched along the negative z semi axis incoming from $-\infty$. Let the extreme of the string be at $z = 0$. If we want to destroy the wave at the extreme we must act inversely as we did to generate it. As a matter of fact we should apply to the extreme a force equal to the velocity of the extreme due to the wave times the opposite of the characteristic impedance $-Z$. Indeed, suppose for a moment that our string would continue beyond $z = 0$. Clearly there would be no

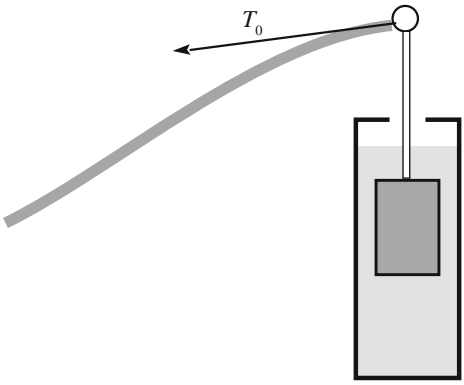


Fig. 3.5 Principle of termination on characteristic impedance

reflection under these conditions because there would be no extreme at $z = 0$. Well, the string section at $z > 0$ would exert on the element at $z = 0$ just that force.

Figure 3.5 schematically shows a possible device for achieving this scope. Attached to the extreme, we have a piston of negligible mass immersed in a fluid, which generates a viscous force, namely a force proportional to and opposite of the velocity of the extreme. The impedance delivered by the "absorber" must be exactly equal to the characteristic impedance of the string. Under these conditions, the extreme moves as if the string extends indefinitely beyond it. The energy transported by the wave is absorbed at the just rate by the absorber. Under these conditions, we talk of matched termination or of termination in the characteristic impedance.

In the cases discussed above of the junction between two strings of different densities, the two characteristic impedances are different as well, the junction is not matched and reflections exist.

Similar situations are encountered for all the transmission lines, in particular, for electric pulses traveling through cables. In this case, reflections are also generated at the end of a cable or at the junction between two cables if the termination or the junction are not matched. We shall discuss this in Sect. 3.12.

3.4 Sound Waves

We shall now consider a first example of a wave in three dimensions, rather than in one, as we have done up to now. A very important example is a sound wave in a gas. A sound wave is produced, for example, by a vibrating string or membrane. The movement, harmonic or not, displaces a part of the gas in which the source is immersed, causing a change in the gas's density and, as a consequence, a change in the pressure. These differences from equilibrium values of density and pressure, in turn, generate changes in density and pressure a little farther from the vibrating

object that is the source of sound. This cyclic succession of events generates a wave that propagates in space, departing from the source.

Let us fix our attention on a particular small volume of gas. When the wave goes through, the volume oscillates back and forth about its equilibrium position. If, in particular, the wave is monochromatic, the motion of the volume is harmonic. Let ψ be the displacement of the small volume from equilibrium. Note that the displacement is now longitudinal relative to the propagation direction of the wave. Sound is a longitudinal wave. Pay attention to the fact that we are considering the motion of a small gas volume and not of the single molecules it contains. Indeed, the gas molecules always move, including when no wave is present. We are interested in the macroscopic ordered motion, which is the average motion of the molecules in the volume. The latter is very small on a macroscopic scale, but still large enough to contain an enormous number of molecules. Let us now quantitatively see what we have stated in words.

We are considering the wave propagation in a continuum medium. This means that the wavelengths must always be much larger than the average distance between molecules, which, as we have seen in the 2nd volume of this course, in a gas at STP, is on the order of several nanometers. Being that the wavelength of the sound waves is on the order of centimeters or larger, the condition is certainly satisfied. A second condition we shall assume is that the variations from equilibrium of pressure and density are small relative to the unperturbed values. Indeed, this is the case for the sound waves, in which these variations are on the order of per mille or less (see Sect. 3.10). It is not the case, however, for the "bang" of a supersonic airplane or for the waves generated by a blast in a mine.

Let us proceed under these assumptions, assuming, for the sake of simplicity, all the relevant quantities, namely position, velocity, pressure and density, to be functions of one space coordinate only (rather than three), which we shall call z. In this case, we speak of a *plane wave*, because all the points on a given plane normal to the propagation direction are equivalent (having the same displacement, velocity, density and pressure).

We first express the fact that the movement of a small volume of gas generates a change in density. Let $\psi(z,\ t)$ be the longitudinal displacement (namely in the direction z) of the gas particle that **at equilibrium** is located at z. Consider a volume of section S enclosed between two planes perpendicular to the propagation direction located at z and $z + dz$ when at rest. The mass of air in the volume Sdz at equilibrium is $\rho_0 Sdz$, where ρ_0 is the equilibrium density. In the presence of the perturbation, the volume changes, because its side at $z + dz$ displaces by $\psi(z + dz, t)$ (if it is positive, the volume increases) and its side at z displaces by $\psi(z,\ t)$ (if it is positive, the volume decreases). The volume consequently becomes $Sdz + S\psi(z + dz, t) - S\psi(z, t) = Sdz + S(\partial\psi/\partial z)dz$. Considering that mass is conserved, we then have $\rho_0 Sdz = \rho S(1 + \partial\psi/\partial z)dz$, where ρ is the new density.

Let us call $\delta\rho$ the difference between the actual and the equilibrium density, namely $\delta\rho = \rho - \rho_0$. For small values of $\partial\psi/\partial z$, we can approximate the above

expression solved for ρ, namely $\rho = \rho_0(1+\partial\psi/\partial z)^{-1}$, as $\rho = \rho_0(1-\partial\psi/\partial z)$. We then obtain

$$\delta\rho = -\rho_0 \frac{\partial\psi}{\partial z}. \tag{3.16}$$

The second process is the change in the pressure due to the change in density. Indeed, in every gas, ideal or real, definite relations exist between pressure and density, say $p = p(\rho)$. Let p_0 be the pressure at equilibrium and $\delta p = p - p_0$ the difference of pressure in the presence of the wave. Being that this is small, we can write $\delta p = (\partial p/\partial\rho)_{\rho_0}\delta\rho$ and, calling $\alpha \equiv (\partial p/\partial\rho)_{\rho_0}$, also

$$\delta p = \alpha\delta\rho, \tag{3.17}$$

which tells us that the variation in pressure is proportional to the variation in density.

Changes in pressure act on the gas, causing variations in its motion. This is clearly the second Newton law. The mass of gas in the volume we are considering is $\rho_0 Sdz$ and its acceleration is $\partial^2\psi/\partial t^2$.

The acting force is the difference between the pressure forces at z and $z + dz$, namely $S[p(z,t) - p(z+dz,t)] = -\frac{\partial p}{\partial z}Sdz = -\frac{\partial\delta p}{\partial z}Sdz$. Hence, the second Newton law gives us $\rho_0 Sdz\frac{\partial^2\psi}{\partial t^2} = -\frac{\partial\delta p}{\partial t}Sdz$. Simplifying, we have

$$\rho_0\frac{\partial^2\psi}{\partial t^2} = -\frac{\partial\delta p}{\partial t}.$$

We now use Eq. (3.17) to eliminate δp and write

$$\rho_0\frac{\partial^2\psi}{\partial t^2} = -\alpha\frac{\partial\delta\rho}{\partial t}$$

and, finally, using Eq. (3.16) to eliminate $\delta\rho$, we obtain

$$\frac{\partial^2\psi}{\partial t^2} - \alpha\frac{\partial^2\psi}{\partial z^2} = 0. \tag{3.18}$$

This is the wave equation we know. The propagation velocity is

$$v = \sqrt{\alpha} = \sqrt{\left(\frac{\partial p}{\partial\rho}\right)_{\rho_0}}. \tag{3.19}$$

We must now find how the pressure depends on the density, namely the function $p(\rho)$. For that, we need two assumptions, one on the nature of the gas (its state equation) and one on the type of thermodynamic process happening during its

compression and expansion. As for the state equation, we shall consider the gas to be ideal, which is a good approximation under normal conditions. To establish which is the type of thermodynamic process, consider two adjacent regions of maximum and minimum pressure due to the wave passage. They are also regions of maximum and minimum temperature. They are separated by half a wavelength. You might think that heat would flow from the high to the low temperature region, but this is not so (do not worry; Newton himself made this mistake). Indeed, heat is transferred through collisions between faster and slower molecules. However, the time needed for this process is much longer than the time available, which is less than half a period, because the latter is the time interval in which the roles of hot and cold regions are interchanged. Recall that the wavelength is much larger than the mean path between collisions. We further observe that we are considering small relative variations of the thermodynamic variables. Under these conditions, the process is reversible.

In conclusion, the process in the gas is a reversible adiabatic transformation of an ideal gas, whose equation is $pV^\gamma = \text{const}$ (where γ is the ratio of the specific heats). Considering that ρ is inversely proportional to V, we can write the equation as $p = \text{const}\rho^\gamma$. Differentiating this expression, we obtain

$$\frac{\partial p}{\partial \rho} = \frac{\gamma p}{\rho}.$$

Coming back to the velocity in Eq. (3.19), we write

$$v^2 = \gamma p_0/\rho_0 = \gamma(Vp_0)/(V\rho_0).$$

We use the gas equation $Vp_0 = nRT$, in which n is the number of moles, and express the mass of the gas as $\rho_0 V = n\mu$, where μ is the molar mass. Finally, we obtain

$$v = \sqrt{\frac{\gamma RT}{\mu}}. \tag{3.20}$$

This is the speed of sound in air and, more generally, in a gas. Note how the speed of sound depends on the temperature and the nature of the gas, but is independent of its pressure and its density.

Let us compare the result with the experiment. As we shall see in Sect. 4.3, the velocity of sound in air at STP is 332 m/s. On the other hand, Eq. (3.20) can also be written as

$$v = \sqrt{\frac{\gamma p_0}{\rho_0}}. \tag{3.21}$$

For air, γ = 1.4 and the molar mass (being an average) is 29 g/mol.

At atmospheric pressure $p_0 = 1.01 \times 10^5$ Pa, the density is $\rho_0 = (29 \times 10^3 \text{ kg/mol})/(22.4 \text{ l/mol}) = 1.29 \text{ kg/m}^3$ and we have

$$v = \sqrt{\frac{1.01 \times 10^5 \times 1.4}{1.29}} = 332 \text{ ms}^{-1},$$

in accordance with the measured value.

Question Q 3.2. The Amundsen-Scott scientific station is located at the South Pole at 2835 m above sea level. During winter, the outside temperature may reach − 70 °C. What would be the speed of sound? □

Question Q 3.3. What would be the speed of sound in He at STP (molar mass is 4 g/mol)? □

3.5 Plane Harmonic Plane Waves in Space

In this section, we shall more precisely define the concept of a wave in three dimensions. In the preceding section, we considered a plane wave. This concept can be easily defined for a harmonic wave, namely when the phase of the wave is defined. In this case, the wave surface is defined as the locus of points having the same phase at a given instant. Clearly, the surface geometry of a wave can be anything. For example, a bell ringing in air produces a wave whose surface at a large distance from the bell is a sphere. A portion of the same wave at a still larger distance may be thought to be plane. We shall consider here the simplest geometry, namely a plane wave and the wave equation for which it is a solution. We shall not analyze different wave shapes, but simply call the attention of the reader to the fact that small portions of even complicated shapes can often be locally approximated with a plane surface, to which the concepts that we shall now develop can be applied.

Let us consider a progressive plane harmonic wave. We call z' the direction in which it propagates. Being a plane wave, its wave function depends on time and on z' but not on the other two coordinates, which we call x' and y'. In this reference frame, and with a proper choice of the initial phase, we can write the wave function as

$$\psi(z', t) = A \cos(\omega t - kz').$$

This expression is identical to that in one dimension due to the particular choice of the reference frame. In general, an arbitrary reference frame, say x, y, z, is rotated relative to x', y', z' by certain angles about the origin (we can forget the translations that only imply an irrelevant change of the initial phase).

To have the equation of our wave in a generic (x, y, z) reference frame, we must express in this frame the product that is kz' in the (x', y', z') frame. Let $\mathbf{r}' = (x', y', z')$ and $\mathbf{r} = (x, y, z)$ be the position vectors of a generic point on the wave in the two

frames. Let us now introduce the vector **k** having magnitude k and the same direction and sense as those of the propagation of the wave. The components of **k** in the two reference frames are (0, 0, k) and (k_x, k_y, k_z), respectively. We now see that kz' is no less than the scalar product of **k** and **r**′ in our original reference frame. Remembering that the scalar product is invariant under rotations, we can write $kz' = \mathbf{k} \cdot \mathbf{r}' = \mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z$. In conclusion, a progressive harmonic plane wave is represented by the equation

$$\psi(x, y, z, t) = A \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) \tag{3.22}$$

or, in complex notation, by

$$\psi(x, y, z, t) = A e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}. \tag{3.23}$$

We have stated that **k** is a vector, namely it is the wave vector. However, as is well known, not all the ordered triplets of numbers are the components of a vector. To be so, they must properly transform under rotations of the reference axes. Let us show that **k** is indeed a vector by looking at its physical meaning. The meaning of the absolute value of **k** is to be the change of phase (in radiants) per unit displacement in the propagation direction of the wave. Consequently, the meaning of one of its components, k_x, for example, must be the change of phase per unit displacement in the x direction. Let θ be the angle of **k** with the x-axis. Now, k is inversely proportional to the wavelength λ, which is the distance between, for example, two consecutive maxima, as shown in Fig. 3.6. From the figure, we see that the distance between two consecutive maxima measured in the x direction, which we might call λ_x, is $\lambda_x = \lambda/\cos\theta$, which is larger than λ. Hence, k_x, which is

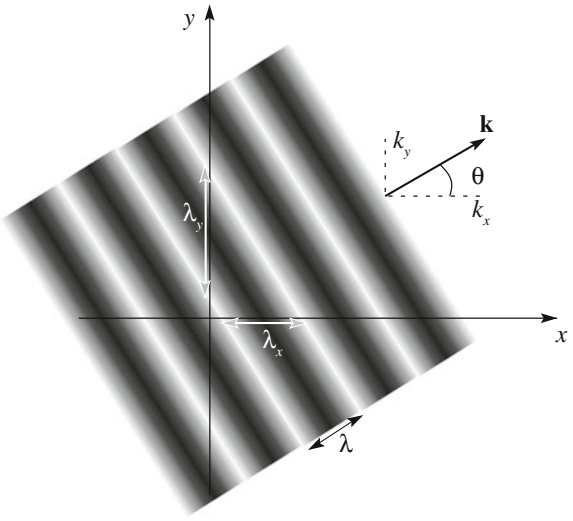


Fig. 3.6 A plane wave, the wavelength and the wave vector and their projections on the axes

inversely proportional to λ_x, is $k_x = k \cos\theta$. Hence, **k** is a vector. Contrastingly, the triplet of numbers $(\lambda_x, \lambda_y, \lambda_z)$ is not a vector. There is no vector wavelength.

One can immediately verify that a progressive harmonic plane wave satisfies the relations

$$\frac{\partial^2\psi}{\partial t^2} = -\omega^2\psi,\ \frac{\partial^2\psi}{\partial x^2} = -k_x^2\psi,\ \frac{\partial^2\psi}{\partial y^2} = -k_y^2\psi,\ \frac{\partial^2\psi}{\partial z^2} = -k_z^2\psi.$$

We can find the wave equation for which the wave is a solution only if we know the dispersion relation of the system. In three dimensions, this is a relation between the angular frequency ω and the magnitude of the wave vector **k**. For the non-dispersive media, the relation is the one we know, in which we must reinterpret k^2 as the square of the wave vector. We have

$$\omega^2 = v^2k^2 = v^2\left(k_x^2 + k_y^2 + k_z^2\right). \tag{3.24}$$

Hence, the differential equation is

$$\frac{\partial^2\psi}{\partial t^2} - v^2\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) = 0, \tag{3.25}$$

which we can write in a compact notation as

$$\frac{\partial^2\psi}{\partial t^2} - v^2\nabla^2\psi = 0. \tag{3.26}$$

This is called the wave equation in three dimensions. In the following section, we shall study several waves, which are solutions to these equations in different physical situations. The plane harmonic waves are particular solutions of Eq. (3.26). Other solutions may be different, both in their shape (spherical, cylindrical, any) and in their dependence on time (not necessarily sinusoidal).

3.6 Electromagnetic Waves

We have already seen, in Chap. 10 of the 3rd volume of this course, how the Maxwell equations, which are the partial differential equations governing all electric and magnetic phenomena, imply the existence of electromagnetic waves. In this solution to the Maxwell equations, both electric and magnetic fields in a vacuum obey the wave equation with a propagation velocity that is equal to the speed of light. We shall repeat here the demonstration for the sake of self-containment and briefly recall how Maxwell experimentally verified the predictions of the theory he had developed.

Let us first recall that James Clerk Maxwell (Scotland UK, 1831–1879) worked on the developments that finally led to his famous equations for ten years between 1855 and 1865, when he published the complete theory in the paper "On the dynamical theory of the electromagnetic field".

The wave equations we obtained in the previous sections were consequences of the Newton laws applied to a mechanical medium. Now that there is no material medium, the underlying equations are different, but the wave equation is the same (even better, it is now exact rather then an approximation). This is a consequence of the fact that, in the case of the electromagnetic field, a cyclic mechanism exists similar to that which we discussed for sound. This process of cyclically chained causes and effects generates a disturbance that, detached from the source that gave it its origin, propagates in space with a well-defined speed.

Let us first describe this mechanism qualitatively. Suppose we have a point charge initially at rest in the origin of our reference frame. It produces a static electric field (Coulomb's law) and no magnetic field. Let us now assume that the charge accelerates in a certain direction for a brief time interval Δt and then continues in a uniform motion. In the immediate surroundings of the charge, we now observe the fields generated by a charge in motion. In particular, the magnetic field is now different from zero. This means that, in the brief time interval, the magnetic field has varied from zero to its final value. However, for the Maxwell equations, a variation over time of the **B** field produces a curl of the **E** field, and consequently also an **E** field, which, in turn, also varies with time. Additionally, the Maxwell equations tell us that an electric field varying over time generates a curl of **B**, and hence a **B**, which varies with time,…., and so on. Initially, we had the fields of a charge at rest in the entire space. Now, they are those of a charge in motion, but only within a sphere, whose radius is growing over time with the speed of light. Outside the sphere, the fields are still those of the charge at rest. We can say that the news that the charge had begun its motion did have not time to reach the points outside the sphere. At the surface of the sphere, a region $c\Delta t$ thick separates the two types of field. In this region, we have the fields of an accelerating charge connecting the internal to the external regions. Figure 3.7 schematically shows the electric field lines. The advancing spherical shell is the electromagnetic wave.

It is important to distinguish two "actors" in the process described. The first actor is the charge that produces the wave with its *acceleration*. Indeed, the sources of electromagnetic waves are, in any case, accelerating electric charges. A charge at rest or moving with constant velocity does not produce a wave. The second actor is the wave, which, once it has been born, has, so to speak, a proper life, propagating independently of the source. If, for example, the source charge were to be annihilated by an equal and opposite charge, the wave would continue its travel.

Let us now analyze how the Maxwell equations quantitatively describe the interplay between electric and magnetic fields, as we did in Chap. 10 of the 3rd volume. We shall consider what happens in a vacuum, namely in the absence of charges and currents. Under these conditions, the Maxwell equations are

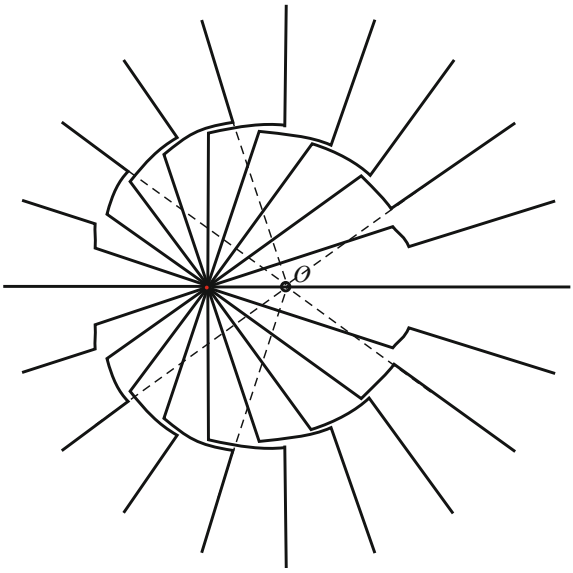


Fig. 3.7 The electric field lines for a charge that is at rest in O, accelerates during a brief time interval and then moves at constant speed to the *left*. The *dotted lines* guide the eye to the initial position

$$\nabla \cdot \mathbf{E} = 0, \tag{3.27}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{3.28}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{3.29}$$

and

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \tag{3.30}$$

Let us take the curl of both sides of Eq. (3.28), obtaining

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial (\nabla \times \mathbf{B})}{\partial t}.$$

We now use the vector identity

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$$

and express $\nabla \times \mathbf{B}$ on the right-hand side with Eq. (3.30). We also use Eq. (3.27) for $\nabla \cdot \mathbf{E}$, obtaining

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\mu_0 \varepsilon_0} \nabla^2 \mathbf{E} = 0. \tag{3.31}$$

A similar argument, starting from Eq. (3.30), leads to the same equation for the magnetic field, namely to

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} - \frac{1}{\mu_0 \varepsilon_0} \nabla^2 \mathbf{B} = 0, \tag{3.32}$$

We have thus found that both electric and magnetic fields in a vacuum obey the wave equation. Namely, the Maxwell equations foresee the existence of electromagnetic waves, propagating in a vacuum with speed

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}. \tag{3.33}$$

The two constants on the right-hand side are measured with electrostatic and magnetostatic experiments respectively. We recall that the differential equations ruling the electric and magnetic fields under time-independent conditions are

$$\nabla \cdot \mathbf{E} = \rho / \varepsilon_0, \; \nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \tag{3.34}$$

Knowing these values, Maxwell not only predicted the electromagnetic waves but also found that their velocity should have been exactly equal, within the experimental errors, to the speed of light. Maxwell concluded that light must be an electromagnetic phenomenon, unifying electromagnetism and optics, two chapters of physics that had been completely separate before him.

However, in 1865, the experimental values of both sides of Eq. (3.33) were known with a rather limited accuracy. We shall discuss the measurements of the velocity of light in Chap. 4. Here, we will just mention that, in 1865, the results of two laboratory measurements were available. In 1849, Hippolyte Fizeau (France, 1819–1896) had measured the value $c = 3.15 \times 10^8$ m/s. In 1862, in a more accurate experiment, Léon Foucault (France, 1819–1868) had obtained $c = 2.98 \times 10^8$ m/s. The two results differ by about 5 %. The quantity on the right-hand side, called the ratio of units at the time, had been determined by Rudolf Kohlrausch (Germany, 1809–1858) and Wilhelm Weber (Germany, 1804–1891) in 1856. Those authors had measured the potential difference of a capacitor of known capacitance, thereby establishing the charge electrostatically. The capacitor was then discharged through a ballistic galvanometer, measuring the same charge as current intensity integrated over time. The result was $(\varepsilon_0 \mu_0)^{-1/2} = 3.11 \times 10^8$ m/s. As Maxwell puts it, “the only use made of light in the experiment was to see the instruments”.

While the values were not in disagreement, their experimental uncertainties were large, and Maxwell thought he had to improve the situation. He developed an ingenious and elegant experiment, which we described in the 3rd volume. We shall not repeat the description here, but will only recall that it was an absolute measurement of the ratio between the force between two current-carrying coils and between the plates of a capacitor charged through a voltage generated by the same current running on a standard resistor. The outcome, presented to the Royal Society of London in 1868 under the title "On a direct comparison of electrostatic with electromagnetic force", was $(\varepsilon_0\mu_0)^{-1/2} = 2.88 \times 10^8$ m/s.

Indeed, precise measurements both of the ratio of units and of the speed of light are quite difficult. After the publication of the Maxwell theory, work to increase the accuracy of both started worldwide. In 1878, in the third edition of his "A treatise on electricity and magnetism", James Clerk Maxwell wrote

> It is manifest that the velocity of light and the ratio of the units are quantities of the same order of magnitude. Neither of them can be said to be determined as yet with such degree of accuracy as to enable us to assert that the one is greater than the other. It is to be hoped that, by further experiments, the relation between the magnitudes of the two quantities may be more accurately determined.
>
> In the mean time our theory, which asserts that these two quantities are equal, and assigns a physical reason for this equality, is not contradicted by the comparison of these results such as they are.

By the next year, the year of the Maxwell's death, the equality of the two quantities had been established with 1 % accuracy. William Ayrton (UK, 1847–1908) and John Perry (UK, 1850–1920) had measured the ratio of units as $(\mu_0\varepsilon_0)^{-1/2} = 2.96 \pm 0.03 \times 10^8$ m/s and Albert Michelson (USA, 1852–1931) the velocity of light (in air) as $c = 2.99864 \pm 0.00051$.

We shall now discuss the fundamental properties of the electromagnetic waves in a vacuum. The wave equation has an enormous number of different solutions. Let us consider the simplest one, which is the plane monochromatic progressive wave. In complex notation, we can write for the electric and magnetic fields

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})}, \quad \mathbf{B} = \mathbf{B}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi)}, \tag{3.35}$$

where the "amplitudes" $\mathbf{E}_0$ and $\mathbf{B}_0$ are now two vectors. We have chosen the origin of the time in order to have the initial phase of the electric field be equal to zero.

Maxwell equations tell us much more about the properties of the electromagnetic waves in a vacuum. Electromagnetic waves are transversal, namely the vectors $\mathbf{E}$ and $\mathbf{B}$ are both perpendicular to the propagation direction. In addition, they are perpendicular to one another. The sense of the wave propagation is the direction of $\mathbf{E} \times \mathbf{B}$. Finally, the ratio of the magnitudes of $\mathbf{E}$ and $\mathbf{B}$ is equal to the velocity c.

We demonstrate these statements, which are true in general, in the particular case of a plane monochromatic wave. Differentiating Eq. (3.35), we immediately get

$$\frac{\partial \mathbf{E}}{\partial t} = i\omega \mathbf{E}, \quad \frac{\partial \mathbf{B}}{\partial t} = i\omega \mathbf{B}, \tag{3.36}$$

$$\nabla \cdot \mathbf{E} = -i\mathbf{k} \cdot \mathbf{E}, \quad \nabla \cdot \mathbf{B} = -i\mathbf{k} \cdot \mathbf{B}, \tag{3.37}$$

and

$$\nabla \cdot \mathbf{E} = -i\mathbf{k} \times \mathbf{E}, \quad \nabla \cdot \mathbf{B} = -i\mathbf{k} \times \mathbf{B}. \tag{3.38}$$

The Maxwell equations for the divergences in a vacuum in Eqs. (3.27) and (3.29) give us, for the monochromatic plane wave,

$$\mathbf{k} \cdot \mathbf{E} = 0, \quad \mathbf{k} \cdot \mathbf{B} = 0, \tag{3.39}$$

namely, both fields are perpendicular to the propagation direction of the wave, which is the direction of the wave vector **k**.

The Maxwell equations for the curls in a vacuum in Eqs. (3.28) and (3.30) give us, for the monochromatic plane wave,

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}, \, \mathbf{k} \times \mathbf{B} = -\varepsilon_0 \mu_0 \omega \mathbf{E} = -c^{-2} \omega \mathbf{E}. \tag{3.40}$$

These equations imply that **E** and **B** are mutually perpendicular and that **k** has the same sense as $\mathbf{E} \times \mathbf{B}$. Remembering that $\omega/k = c$ and taking into account that **E** is perpendicular to **k**, we immediately find that the ratio of the magnitudes of the fields is

$$B = E/c. \tag{3.41}$$

In conclusion, the electric and magnetic fields in an electromagnetic wave are tightly connected. Once we know one of them, we know the other as well. In the study of electromagnetic waves, it is consequently sufficient to consider one field only, calculating the other one when necessary. In the following section, we shall deal with the electric field, chosen because it is usually the one giving the most directly observable effects.

We also note that the particular wave in Eq. (3.35) that we have considered is not only monochromatic and plane but also plane polarized. Namely, the directions of the fields do not depend on time or position. In general, the directions of **E** and **B** vary from point to point and from instant to instant, while remaining perpendicular to one another and to the propagation direction.

3.7 The Discovery of Electromagnetic Waves

With his beautiful theory, Maxwell had unified (meaning he had described them under the same set of equations) two previously separated fields, namely electromagnetism and optics. However, the existence of electromagnetic waves, produced by electric charges moving in a circuit as predicted by the theory, had to be experimentally verified. Starting immediately after 1865, this difficult experimental problem was attacked by a number of scientists for a period stretching over twenty years, finally being solved by Heinrich Hertz (Germany, 1857–1894) in 1887.

The basic prediction of the theory that needed to be checked experimentally was whether electric charges and currents do, indeed, generate waves in which the electric and magnetic forces, using the language of the time, propagate with the speed of light. We might think, for example, of quickly discharging a capacitor and measuring the delay of a possible electric effect on an electrometer at a distance. Or we might rapidly excite a magnet and measure the delay with which the needle of a compass reacts, again at a certain distance. We note that the distance between the source where the wave is produced and the position where we attempt to detect it should be sufficiently large. As a matter of fact, the electric field is generated by two causes: electric charges and variations in time of the magnetic field. Similarly, the magnetic field is the result of the currents in the conductors and the variations in time of the electric field. The contributions resulting from the charges and the currents are those dominant near the source, while the contributions due to the rate of change of the other field, which are the ones in which we are interested, become dominant at distances substantially larger than the main wavelengths of the game. The wavelengths generated by a conductor are on the order of the dimensions of the conductor itself. In conclusion, we should place our detectors at, say, ten meters from a one meter or so long source. As a matter of fact, Michael Faraday (UK, 1791–1867) had already attempted this type of experiment, without finding any effect. This was before the development of the Maxwell theory, and thus he did not know what delay to expect.

As opposed to Faraday, Hertz knew that it was a minuscule delay, only 30 ns at 10 m. The main problem in using a phenomenon like the discharge of a capacitor or the excitation of a magnet is that it is too slow, namely its duration greatly exceeds 30 ns. In an address delivered at a conference in Heidelberg in 1889, Hertz explained this point as follows (from the English translation of 1889 by D.E. Jones):

> Such a small fraction of time we cannot directly measure or even perceive. It is still more unfortunate that there are no adequate means at our disposal for indicating with sufficient sharpness the beginning and the end of such a short interval. If we wish to measure a length correctly of the tenth part of a millimeter it would be absurd to indicate the beginning of it with a broad chalk line. If we wish to measure a time correctly to the thousands part of a second (*which is still about 30 000 longer than 30 ns*) it would be absurd to indicate its beginning by the stroke of a big clock.

Hertz carefully studied the discharge of capacitors and conductors, finding durations on the order of tens of microseconds and, hence, too long for his purpose. He did, however, discover that, under suitable conditions, the discharge is not a continuous process, but rather

> it consists of a large number of oscillations in opposite senses which follow each other exactly at equal intervals.

Hertz was able to discharge conductors and excite oscillations having periods between 10 and 1 ns.

> We now have indicators for which 30 ns is not too short *provided that* we can get only two or three of such sharply-defined indicators.

In practice, Hertz employed a Ruhmkorff induction coil, a device that had been patented in 1851 by Daniel Ruhmkorff (Germany, 1803–1877) and was in common use at the time. Hertz modified the coil as described below. The coil, schematically shown in Fig. 3.8, is essentially a transformer composed of two cylindrical coaxial solenoids wrapped around an iron core. One of them, the primary circuit, has a limited number, say N_1, of turns of a thick wire, while the secondary has a large number of turns, N_2, of a thin wire. The primary is connected to the poles of a battery, providing a continuous voltage through a switch. The switch is periodically closed and opened by a small motor, about one hundred times per second. In this way, a current variable over time is generated in the primary, which electromagnetically induces an electromotive force in the secondary. The amplitude of the oscillating emf in the secondary is N_2/N_1 times larger than that in the primary. This transformation ratio, as it is called, can be as high as several hundred. In the Ruhmkorff coil, the secondary is open. The emf between its terminals (a and a' in Fig. 3.8) can be of several kilovolts.

Hertz joined two straight strands of copper, each 75 cm long with a 5 mm diameter, to the terminals and two spherical conductors of radius $R = 15$ cm at their far extremes (B and B' in the figure). The two near extremes of the wires were terminated by two small conducting spheres at a distance of a few millimeters, constituting a spinterometer.

Hertz could adjust the distance s between the spheres of the spinterometer with a micrometer in order that a spark would occur when the emf between the small spheres reached a certain value.

The operation of the device is as follows. The switch of the primary having been closed, the emf between the terminals of the secondary increases with time. In

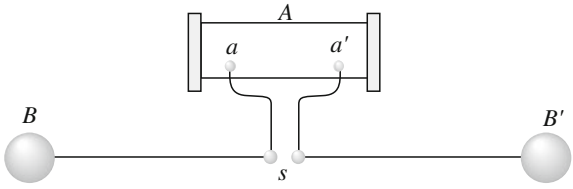


Fig. 3.8 The Hertz device for generating electromagnetic waves. A is the Ruhmkorff coil

particular, the two larger conductors B and B' charge up with opposite charges, like a condenser. When the emf reaches the preset value, the spark occurs. The two conductors B and B' discharge through the copper wires, whose resistance is small, on the order of the ohm. As a matter of fact, the "circuit" made of B, B', the wires and the spark behave like an oscillating circuit and the discharge oscillates as in the above sentences by Hertz. The circuit is not in a stationary regime. The current intensity is different in its different sections. In addition, the circuit is not even closed, similarly to the one we considered in Sect. 10.2 of the 3rd volume. Note that the secondary circuit, which is in parallel with the spark, offers a completely negligible contribution during the discharge, due to its much higher impedance at the high frequencies of the oscillating discharge. However, it will be active in the next phase, when the two large spheres will be charged again.

We need to know the period of the oscillating discharge. This is clearly $T = 2\pi\sqrt{LC}$, where L is the inductance and C is the capacitance of the oscillating circuit. The capacitance is that of the two spherical conductors, namely $C = 4\pi\varepsilon_0 R$. With $R = 15$ cm, we have $C = 17$ pF. For the calculation of the inductance, Hertz used an approximate formula by Neumann, finding $L = 0.22$ μH. The calculated value of the period is then $T = 12$ ns. Hertz had a source of sufficiently sharp indicators.

> But these would be of little use to us if we were not in a position to actually perceive their action up to the distance under consideration, namely about ten meters. This can be done by very simple means. Just at the spot where we wish to detect the force we place a conductor, say a straight wire, which is interrupted in the middle by a small air-gap. The rapidly alternating (*electromotive*) force sets the electricity of the conductor in motion, and gives rise to a spark in the gap. The method had to be found by experience, for no amount of thought could well have enabled one to predict it would work satisfactorily. For the sparks are microscopically short, scarcely a hundredth of millimeter long; they only last about one millionth of a second. It almost seems absurd and impossible that they should be visible; but in a perfectly dark room they *are* (*evidence by Hertz*) visible to an eye which has well rested in the dark.

Once the method had been found, Hertz worked on it, optimizing the dimensions and the geometry of the detecting wire. He found its detector behaving as an acoustic resonator, namely a high sensitivity could be achieved by tuning the detector to the same frequency as the wave, which was the frequency of its generator.

Then, Hertz designed the experimental conditions so as to obtain stationary, rather than progressive, waves in his laboratory. He located the source at a fixed point and observed the intensity of the sparks in its detector in different locations in the laboratory. He found, as expected, positions of maximum intensity alternating with positions of minimum intensity of the sparks. By measuring the distance between two consecutive minima, he determined the wavelength. He had calculated the theoretical period, as we have seen, and he calculated the wave velocity (wavelength divided by the period). The result was the speed of light, within the experimental uncertainties.

He further investigated the effect of rotating the detecting wire in a given position. The intensity of the spark was a maximum if the direction was perpendicular to the direction joining the observation point to the source, while the sparks completely disappeared when the wire was pointing to the source. In agreement with the Maxwell theory, the wave is transversal.

Hertz had proved experimentally that the Maxwell theory is true. Not only had the unification of electromagnetism and optics been completed, but new phenomena of physics had been foreseen and were waiting to be explored and exploited: the phenomena of electromagnetic waves of wavelengths both larger than those of light, namely on the order of millimeters and centimeters up to kilometers, and shorter than those of light in the ultraviolet and beyond.

Hertz himself continued his study of the electromagnetic waves that he had learnt to generate, which had wavelengths on the order of meters. Locating the source in the focus of a parabolic mirror, he produced a parallel beam. He showed that the beam propagated in a straight line, that it was producing a shadow when encountering obstacles, and that it reflected and refracted in exactly the same manner as the light. In addition, he showed that the polarization properties (see Chap. 6) were also equal to those of light. Hertz was now able to perform all the classical experiments of optics with "hertzian waves".

Fourteen years after the first successful experiment by Hertz, in December 1901, Guglielmo Marconi (Italy, 1874–1937) was able to exploit, for the first time, the "hertzian waves" in a transatlantic radio transmission from Poldhu, Cornwall, England to St John's, Newfoundland, Canada, over a distance of 3500 km.

3.8 Sources and Detectors of Electromagnetic Waves

The sources of electromagnetic waves are, in any case, accelerating electric charges. A charge at rest generates only a time-independent field, but no wave. For the relativity principle, a charge in uniform rectilinear motion cannot generate a wave either. In every physical source of electromagnetic waves, an atom, a lamp, a radio station or a star, accelerating charges are always present. Often, their motion is an oscillation.

We shall pay attention here to the electric field produced by a charge in an arbitrary motion (remember that, in a wave, electric and magnetic fields are connected). From the 3rd volume, we know the electric field produced by a charge at rest, which is given by Coulomb's law. From this, we can obtain the electric field of a point charge in uniform rectilinear motion through a Lorentz transformation, but we shall not need it. The expression of the electric field of a point charge in arbitrarily accelerated motion obtained by solving the Maxwell equation is rather complicated, but we shall not need it in its complete form.

Figure 3.9 shows the trajectory of the point charge q. Let $\mathbf{E}(\mathbf{r}, t)$ be the electric field of the charge at a generic point P at a distance r from the charge at the instant t. As shown in the figure, $\mathbf{r}$ is the vector from the charge position at the instant t to

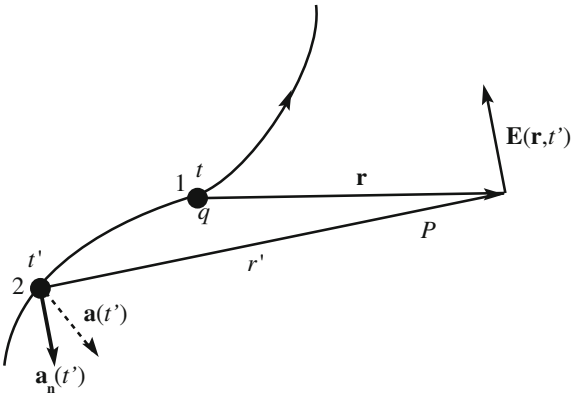


Fig. 3.9 The electric field of a charge moving in an arbitrary motion

the point P. Well, by solving the Maxwell equations, one finds that the field in P does not depend on r but rather on the position the charge had in a preceding instant t'. This instant is precedent by the time necessary for the information to reach the point P, moving at the speed of light c. Namely, if r' was the distance at the instant t', it is $t' = t - r'/c$. The time t' is called *retarded time*. To better understand this, suppose the charge emits light pulses while moving. An observer in P looking in the instant t does not see the position occupied by the charge at the instant t but the one it occupied at $t' = t - r'/c$.

Hence, the field of a point charge in an arbitrary motion is a function of the distance at the retarded time r'. As a matter of fact, in the study of electromagnetic radiation, we are always at distances that are large compared to the wavelength of the radiation under study. The complete expression of the electric field contains terms decreasing with increasing distance r' as $1/r'$, terms decreasing as $1/r'^2$ and terms decreasing as $1/r'^3$. Only terms proportional to $1/r'$ are sizeable at large distances, while the others rapidly become negligible.

We call the part of the electric field that decreases as $1/r'$ with the distance the *radiation field*. This is the only term relevant for our discussion. This is the term that causes the light of objects very far away, like stars, to be visible to our eyes and our instruments, namely so their field will still be large enough to excite the cells of our retina or our sensors. If the field was decreasing as $1/r'^2$, as the Coulomb field of a charge at rest would, we would not see any stars; our sky would be dark both at night and during the day.

A further simplification is possible when the motion of the charge develops in a limited space, which is much smaller than the distance to the observation point. In this case, the distance does not vary too much with time and we can make its value r' at the retarded time equal to its actual value r. In addition, we shall assume the velocity of the charge to be much smaller than the speed of c. All these assumptions are well satisfied in the situations we shall discuss, in which we shall deal with the field generated by electrons inside atoms or by the conduction electrons oscillating

in an aerial at a large distance from it. Under these conditions, the expression of the electric field of a point charge, found by R. Feynman, is

$$\mathbf{E}_{\text{rad}} = -\frac{q}{4\pi\varepsilon_0}\frac{\mathbf{a}_n(t - r/c)}{rc^2}, \tag{3.42}$$

where r is the present distance between the charge and the observation point P and the vector $\mathbf{a}_n$ is the component of the acceleration of the charge in the direction normal to $\mathbf{r}$, namely the projection of the acceleration on the plane normal to the direction in which the charge is seen from P. We call this direction the line of sight. We shall use Eq. (3.42) very often in the subsequent sections.

As already stated, all the electromagnetic waves in a vacuum are described by the same equation, are generated by oscillating charges, propagate with the same velocity, are, in any case transversal, etc., independently of their wavelengths. However, their wavelengths span an enormous number of orders of magnitude, and consequently they can give origin to very different phenomena, depending on the wavelength. Different are their physical sources and the instruments required to detect them are different.

For example, FM radio, TV broadcasts and cellphones use wavelengths from 100 mm to 10 m (corresponding to frequencies in the range from 30 MHz to 3 GHz). Radars employ shorter wavelengths, say from 1 mm to 100 mm (3–300 GHz). These waves exist in nature and can be generated artificially as well. They can be detected using an aerial, which can be schematically thought of as a metallic bar. The electric field of the incoming wave forces the conduction electrons into oscillations, generating an emf in the bar, which is amplified by an electronic circuit. In this way, we can follow the evolution of the magnitude of the wave field as a function of time. Inversely, we can generate an oscillating electromotive force with an electronic circuit and excite the aerial with it. The aerial will transmit an electromagnetic wave.

Electromagnetic waves are visible, namely they are light, in the wavelength range of 0.38 μm < λ < 0.78 μm, where the color blue has the shortest length, the color red the largest. The corresponding frequency band is roughly from 400 to 750 THz. The band is quite narrow and is centered around the maximum on the solar spectrum. The evolution of our eyes has determined this to be visible. Examples of artificial sources are common lamps and LASERs. The latter produce a qualitatively different type of light, as we shall see. Several detectors are available, ranging from photographic emulsions to photodiodes to photomultipliers.

The infrared band is at frequencies immediately smaller than the smallest visible one, ranging from several dozen THz to 750 THz (in wavelength from dozens of μm to 0.78 μm). They are produced by hot bodies and by our MASERs. We can detect them, for example, with bolometers, which are blackened foils, or other absorbing bodies that, when exposed to the radiation, heat up so that their temperature of equilibrium with the environment can be measured.

The ultraviolet band is on the other side of visibility, down to wavelengths of a few dozen nanometers. At still higher frequencies, we have X-rays and then gamma-rays.

This very limited discussion was meant simply to familiarize the reader with the most common terminology and will not be pursued any further. However, the following very important issue must be still mentioned. The signal induced by an electromagnetic wave can be detected and amplified by an electronic circuit, provided the circuit is fast enough to follow the evolution in time of the electric field of the wave. This can be done only up to a certain maximum frequency. The exact value of this limit changes with the progress of technology, but we can think of it as being in the range of the GHz. Similar high frequency limits exist for the sources.

Let us consider the sources first. In the infrared, in the range of visibility and at higher frequencies, the macroscopic sources, like a lamp or a star, contain an enormous number of atoms. A fraction of these atoms, once excited, are the microscopic sources of radiation. The important point is that each atom oscillates independently of all the other ones. Even when the oscillations are at the same frequency, their initial phases are chaotically distributed. Each single oscillation lasts for a specific total time, which is typically 10^{-8} s. This duration is much longer than the oscillation period (say 10^{-14} s) and much shorter than the macroscopic times. As a consequence, even if the emitted wave is practically monochromatic, its instantaneous phase varies in a chaotic way, completely outside our control. These sources are called thermic sources.

This is not the case with the MASERs and the LASERs. In these systems, we are capable of having all the atoms oscillate in phase with one another in a process called stimulated emission. MASER and LASER stand, respectively, for Microwave (in the infrared) and Light Amplification by Stimulated Emission of Radiation. The nature of the coherent radiation they produce is completely difference from that of the thermal sources. We shall come back to this in Sect. 4.8.

Let us now discuss how we can detect an electromagnetic wave. As we shall see in the next section, any progressive wave, and the electromagnetic ones in particular, carry energy with them. We can detect a wave only if our detector, or sense organ, absorbs a sufficient amount of energy. To proceed, we shall now introduce a few physical quantities, with definitions that are valid for every type of progressive wave.

Let $d\Sigma$ be an infinitesimal surface normal to the direction of propagation of the energy of the wave. Energy propagates in the direction of the wave vector $\mathbf{k}$ for a harmonic wave (for which it can be defined) in an isotropic medium, but not under every circumstance. We shall study the case of non-isotropic media in Chap. 6. Let dU be the energy of the wave crossing $d\Sigma$ in the infinitesimal time interval dt. We define as elementary the *energy flux* $d\Phi$ through the surface element $d\Sigma$ normal to the propagation direction of the energy crossing $d\Sigma$ per unit time, namely

$$d\Phi \equiv \frac{dU}{dt} = Id\Sigma, \tag{3.43}$$

where, on the right-hand side, I is the *wave intensity*, which is defined as the energy flux through the normal surface $d\Sigma$ per unit time, namely

$$I = \frac{dU}{d\Sigma dt}. \tag{3.44}$$

More generally, we obtain the energy flux through a surface element dS of arbitrary orientation considering the projection of dS onto the normal in the direction of propagation. If $\mathbf{n}$ is the unit vector normal to dS and $\mathbf{u}$ the unit vector of the direction of propagation, we have

$$d\Phi \equiv \frac{dU}{dt} = I\mathbf{u} \cdot \mathbf{n}dS. \tag{3.45}$$

The *energy flux* through a finite surface S is obtained through integration. Namely, it is

$$\Phi = \int_S d\Phi = \int_S I\mathbf{u} \cdot \mathbf{n}dS. \tag{3.46}$$

The physical dimensions of the energy flux are those of an energy divided by a time, namely they are the same as power. The flux is measured in watt (W). The energy intensity is the flux per unit surface and is measured in watt per square meter (Wm^{-2}).

Let us now consider the detectors of electromagnetic waves. They have different characteristics depending on the frequency interval in which they work, but have common characteristics as well. In any case, a detector can respond, giving its signal, only if it absorbs a sufficient amount of energy. As a consequence, every detector has a sensitive surface of a non-zero area, which is crossed by the energy carried by the wave. The detector will absorb a fraction, up to 100 %, of this power. After a certain time interval, the total absorbed energy will be sufficient for detection. The important point here is that any detector must integrate the incoming intensity over a non-zero surface and over a non-zero time interval. The specific process leading to detection is different from one detector type to another. For example, it is a physico-chemical process in the sensitive cells of our retina and in a photographic emulsion, it is a temperature increase in a bolometer, it is the transfer of electrons from one energy state to another in a photo-camera, etc. However, whatever the process may be, it needs a minimum energy to happen.

Note that all these detectors are not directly sensitive to the amplitude of the wave, but rather to its square, because the energy of a wave is proportional to the square of the amplitude. For this reason, they are called *square-law detectors*. In addition, as we have just seen, any square-law detector measures the *average*

energy flux over a certain area and over a certain time interval. For electromagnetic radiation at frequencies above the many GHz scale, the integration times of our detectors are much longer than the period of oscillation. Namely, the detectors measure the average intensity over a time interval containing a large number of periods. This number being very large, we can well approximate it in our calculations with an integer number, say n, and consider that the average over n periods is equal to the average over one period. We then define as the *average intensity* of a harmonic wave the average value over a period (or over a time interval much longer than a period) of the energy flux it carries through a unit surface perpendicular to the energy propagation direction. Note that the term intensity alone is often used in the sense of average intensity. Note also that the definitions we are given are valid in general, not only for electromagnetic waves.

3.9 Impedance of Free Space

In Sect. 3.3, we saw how to terminate the characteristic impedance on a string carrying an elastic wave. The termination was such so as to absorb the exact energy carried by the wave. The same problem exists for every type of progressive wave. Particularly important is the case of electromagnetic waves in a vacuum, which we will now analyze.

Let us consider a plane progressive harmonic wave propagating in the positive z direction, linearly polarized with the electric field in the x direction, say $\mathbf{E} = (E_x, 0, 0)$. Consequently, the magnetic field is in the y direction, namely $\mathbf{B} = (0, B_y, 0)$. In our wave, the magnitudes of the fields are related, and we can write

$$cB_y = E_x. \tag{3.47}$$

Let us now suppose that the wave advances, in the z direction as noted, in the semi-space of the negative z and that we want absorb it into the plane $z = 0$. We do that by placing a conducting surface on that plane having a surface resistivity Z of the value we shall now find.

The electric field E_x of the wave produces a current on this surface in the x direction, having a surface density, which we call K_x, given by Ohm's law. We have

$$K_x = E_x/Z = cB_y/Z. \tag{3.48}$$

The energy flux of the electric field is transferred to the charge carriers and dissipated, transformed into thermal energy, via the Joule effect. As a result, the electric field of the wave vanishes on the surface.

On the other hand, as we learned in the 3rd volume (Sect. 10.8), the component of the magnetic field parallel to a current-carrying surface has a discontinuity in

crossing that surface equal (in the present case) to K_x/μ_0. We take advantage of this to impose a zero magnetic field beyond the surface. The condition for that is

$$K_x = B_y/\mu_0. \tag{3.49}$$

We see that we can satisfy the conditions in both Eqs. (3.48) and (3.49), by choosing the surface resistivity to be such that $Z/c = \mu_0$. Remembering that $c = (\varepsilon_0\mu_0)^{-1/2}$, the condition is

$$Z = \mu_0 c = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 376.730\ldots \quad \Omega. \tag{3.50}$$

This fundamental quantity is called the *impedance of the free space* or *impedance of the empty space*.

Note that the two fundamental constants, the electromagnetism ε_0 and μ_0, determine the two fundamental constants of electromagnetic waves in a vacuum, velocity and impedance. Note also that in the SI units system, the values of c and μ_0 are given by definition (of the meter and the ampere, respectively). Consequently, the value of Z is given by definition as well. In Eq. (3.50), only the first few digits are shown.

3.10 Intensity of the Sound Waves

In this section, we shall consider the energy carried by sound waves. A sound wave is basically an elastic wave. The simplest elastic wave is the wave on an elastic string, as we considered in Sect. 3.1. Consider, for example, a pulse, namely a wave of limited length, traveling along the string. The elements of the string affected by the pulse at a certain time, being displaced from their equilibrium position, have some potential (elastic) energy, and, being in motion, some kinetic energy. Once the pulse is gone, both their potential and kinetic energies have returned to zero. The string is a continuous medium, and consequently we must talk of energy density, which is the energy per unit length in this one-dimensional case.

Let us consider a progressive harmonic plane sound wave moving in the positive z direction in a homogenous and isotropic medium. Under these conditions, the propagation velocity and direction of energy are the same as that of the wave. As we shall see subsequently, special care must be taken with the definition of "velocity" in the case of dispersive waves. In this case, the wave velocity of a harmonic wave depends on frequency, and the velocity of the pulses needs to be carefully considered. In addition, the velocity of the energy flow may be different from the velocity of the wave. For the non-dispersive waves we are considering now, contrastingly, all the velocities that can be defined in relation to the wave have the same value.

Let $\psi(z,t) = f(z - \upsilon t)$ be the wave function of our sound wave, namely the displacement of a fluid element from equilibrium. In a progressive wave, there is a relation between the time and space partial derivative, as in Eq. (3.12). In this case, we have

$$\frac{\partial \psi}{\partial z} = -\frac{1}{\upsilon}\frac{\partial \psi}{\partial t}. \tag{3.51}$$

We now consider the force exerted by the gas laying to the left of the coordinate z to the gas on its right through the generic section S normal to z. If the pressure is p, the force is pS. Let us consider the volume of gas having section S that, *at equilibrium*, is located between z and $z + a$, where a is a small distance. This volume is $V_0 = Sa$. When the wave is present (remember that it is longitudinal), the volume varies by the quantity δV given by the equation

$$\delta V = [\psi(z+a,t) - \psi(z,t)]S = \frac{\partial \psi}{\partial t} Sa.$$

We shall need the relative change in volume, which is

$$\frac{\delta V}{V_0} = \frac{\partial \psi}{\partial z}.$$

As we know that we are dealing with adiabatic transformation of an ideal gas, we obtain the corresponding relative change in pressure by differentiating the adiabatic equation $pV^\gamma = \text{const}$, obtaining

$$\frac{\delta p}{p_0} = -\gamma \frac{\delta V}{V_0}.$$

We can then write

$$\delta p = -p_0 \gamma \frac{\partial \psi}{\partial z} = -\upsilon^2 \rho_0 \frac{\partial \psi}{\partial z}, \tag{3.52}$$

where ρ_0 is the density in absence of the wave and $\upsilon^2 = \gamma p_0/\rho_0$ is the wave velocity squared. The work done on the unit surface per unit time by the pressure force, which is positive on one side of the volume and negative on the other, is the pressure difference δp times the velocity of the gas elements due to the wave. The latter is $\partial\psi/\partial t$. But this work per unit area and unit time is just the energy flux per unit area, which is the intensity. We can then write

$$I(t) = -\upsilon^2 \rho_0 \frac{\partial \psi}{\partial z}\frac{\partial \psi}{\partial t}$$

and, using Eq. (3.51),

$$I(t) = \upsilon\rho_0 \left(\frac{\partial\psi}{\partial t}\right)^2. \tag{3.53}$$

Considering now a harmonic wave, namely the wave function

$$\psi(z,t) = A\,\cos(\omega t - kz) \tag{3.54}$$

the instantaneous intensity is

$$I(t) = \upsilon\rho_0 A^2\omega^2 \sin^2(\omega t - kz). \tag{3.55}$$

We see that the instantaneous intensity varies periodically over time as a circular function squared. We obtain the average intensity by taking the average on a period and recalling that such an average is equal to ½ for a squared circular function. We obtain

$$\langle I(t)\rangle = \frac{1}{2}\upsilon\rho_0 A^2\omega^2. \tag{3.56}$$

We now recall that, in Eq. (1.22), we found that the energy of a harmonic oscillator of mass m of amplitude A and angular frequency ω is $(1/2)mA^2\omega^2$. In the present case, thus, the energy per unit volume is $(1/2)\rho_0 A^2\omega^2$, and Eq. (3.56) tells us that the energy flux is the energy density times the wave velocity. This apparently obvious conclusion, namely that energy propagates with the wave velocity, is true in the present case, but there are cases in which it is not so.

Note that, as anticipated, the wave intensity is proportional to the square of the oscillation amplitude.

The sound intensity units are the watt per square meter ($\mathrm{Wm^{-2}}$).

The frequency range in which a sound wave is audible depends on the person. Its assumed standard values are between 20 and 20,000 Hz (between 17 m and 17 mm wavelengths). At higher frequencies, one talks of ultrasounds. In any case, instruments capable of following the wave oscillations in time can be easily built. For example, a common microphone produces an emf that is proportional to the value of the pressure varying with time, which can be amplified, registered, etc. Often, however, what matters is the average intensity, meaning a square-law detector can be used.

Our ear is not only a quadratic detector, meaning that the system integrates over times much longer than the period of the sound wave, but it has a non-linear response as well. The response of our auditory system is logarithmic. In this way, we are sensitive to sounds over an enormous range of intensity. This feature, called the *dynamic range*, extends, for a normal ear, over 13 orders of magnitude, from 10^{-12} W/m^2 (you need very good hearing for that) to 10 $\mathrm{Wm^{-2}}$ (a level producing a painful sound).

The unit for the sound intensity on a logarithmic scale is called the *decibel* (dB). The decibel is one tenth of a *bel*. Rigorously speaking, these units measure the

logarithm of the ratio between two intensities. Namely, the intensity I_2 is higher then the intensity I_1 by K bel when $K = \log(I_2/I_1)$. Clearly, the intensity I_2 is higher then the intensity I_1 by K decibel when

$$K = 10 \log(I_2/I_1). \tag{3.57}$$

Note that the ratio between two sound intensities of 1 dB is equal to 1.26 and is barely perceivable by the human ear. For this reason, in practice, the decibel is used in acoustics, rather than the bel. In addition, the decibel is often used for absolute values of the intensity as well. This is done by assuming the minimum audible intensity as a standard reference, namely $I_1 = 10^{-12}$ W/m^2, which is then equal to 0 db by definition. For example, a whisper at one meter might have an intensity of 10–15 db, while the engine of an airplane, even at a few meters, would be 120 db.

We shall give further examples after having considered the oscillation amplitudes of the displacement of our air drum and of the pressure acting on it. Considering a harmonic wave, we obtain from Eq. (3.52), taking Eq. (3.51) into account, that the relation between the maximum pressure, say $p_{\max}$, and the maximum oscillation velocity, say $u_{\max} = (\partial\psi/\partial t)_{\max} = A\omega$, is $p_{\max} = \rho_0 \upsilon u_{\max} = \rho_0 \upsilon A\omega$. From Eq. (3.56), the average intensity is $\langle I(t)\rangle = \upsilon\rho_0 u_{\max}^2/2$, and we can write the relation between intensity and maximum oscillation pressure as

$$p_{\max} = \sqrt{2\rho_0 \upsilon}\sqrt{\langle I\rangle}. \tag{3.58}$$

Let us consider the relevant example of air at STP with $\rho_0 = 1.3$ kg/m^3 and $\upsilon = 332$ m/s. Equation (3.58) gives us $p_{\max} = 29.3\sqrt{I}$. Then, the maximum pressure corresponding to a painful sound of 10 W/m^2 is $p_{\max} \approx 90$ Pa. This is a very small pressure, on the order of one thousandth of the atmospheric pressure.

Let us see how much the corresponding maximum displacement of the eardrum is. From Eq. (3.56), we have that the oscillation amplitude is

$$A = \frac{1}{\omega}\sqrt{\frac{2\langle I\rangle}{\rho_0 \upsilon}}. \tag{3.59}$$

For example, at a typical frequency of 1 kHz, namely for $\omega = 2\pi \times 10^3$ rad/s, we have

$$A = \frac{1}{6.28 \times 10^3}\sqrt{\frac{2}{1.3 \times 3.3 \times 10^2}}\sqrt{\langle I\rangle} = 1.1 \times 10^{-5}\sqrt{\langle I\rangle} \text{ m}.$$

Hence, for a painful sound of 10 W/m^2 intensity at 1 kHz frequency, the oscillation amplitude of our eardrum is $A = 35$ μm. If the intensity is at the audible threshold, the amplitude of the eardrum motion is $A = 0.01$ nm, which is much smaller than an atomic radius. Table 3.1 shows approximate values of sound intensity, both in W/m^2 and in decibel and maximum pressure excursions for several typical sounds.

Table 3.1 Shows approximate values of sound intensity, both in W/m^2 and in decibel and maximum pressure excursions for several typical sounds

	p (Pa)	I (W/m^2)	K (db)
Audibility threshold	3×10^{-5}	10^{-12}	0
Drizzle	3×10^{-4}	10^{-10}	20
Conversation at 3 m	3×10^{-3}	10^{-8}	40
Orchestra	0.3	10^{-4}	80
Airplane at 5 m	30	1	120

3.11 Intensity of Electromagnetic Waves

In this section, we shall see that the speed of electromagnetic waves in a vacuum is also the propagation velocity of the energy of the electromagnetic field. We recall that, in Chap. 10 of the 3rd volume, we found that the energy density in an electromagnetic field is given by

$$w = \frac{\varepsilon_0}{2}E^2 + \frac{1}{2\mu_0}B^2 = \frac{\varepsilon_0}{2}\left(E^2 + c^2B^2\right) \tag{3.60}$$

and that the electromagnetic energy flux, namely the energy through the surface unit normal to the propagation direction in a second, is given by the Poynting vector

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B}. \tag{3.61}$$

Consider now a progressive monochromatic (harmonic) plane electromagnetic wave in a vacuum. We immediately note that **S** has the same positive direction as the wave vector **k** (namely as $\mathbf{E} \times \mathbf{B}$). To find its magnitude, we recall that, in our wave, **E** and **B** are perpendicular to one another and that their magnitudes are related by $B = (1/c)E$. The energy flow per unit section and unit time is then $S = E^2/(c\mu_0) = \varepsilon_0 c E^2$. The intensity is the average of S over a period, namely

$$I = \langle S \rangle = \varepsilon_0 c \langle E^2 \rangle.$$

The average stored energy is immediately obtained from Eq. (3.60) as

$$\langle w \rangle = \varepsilon_0 \langle E^2 \rangle$$

and we finally have that

$$I = c\langle w \rangle.$$

The (average) intensity is equal to the average energy density times the wave velocity.

Consider, as an example of energy carried by an electromagnetic wave, the energy we receive from the sun. Above the atmosphere, its intensity is about

$S = 1.3$ kW/m^2. The atmosphere absorbs and reflects a fraction of this flux, depending on the weather conditions, on the latitude, etc. On average, half of the flux reaches the soil. Typically, on a surface of 100 m^2, you can have an incident power of about 50 kW on a clear sunny day at intermediate latitudes. If you want to know the usable power, you must multiply this by the efficiency of your device (say 10–20 %).

Let us now find the magnitudes of the electric and magnetic fields of the wave. With $S = 1.3$ kW/m^2, we have $\langle E^2 \rangle = \langle S \rangle / (\varepsilon_0 c) = 4.9 \times 10^5$ V^2 m^{-2} that is $\sqrt{\langle E^2 \rangle} = 700$ V/m, a quite small value. For the magnetic field, we have $\sqrt{\langle B^2 \rangle} = \sqrt{\langle E^2 \rangle}/c = 2.3$ μT, which is a very small value. For comparison, think about the fact that the earth's magnetic field is an order of magnitude larger.

3.12 Electromagnetic Waves in a Coaxial Cable

Coaxial cables, or coax for short, are very frequently used as transmission lines for electromagnetic signals when a protection from external electromagnetic interference is requested. They consist of an inner copper wire (1 mm or so in diameter) called the core, surrounded by a cylindrical insulating sheath, a second cylindrical conductor outside the sheath, and finally, a second external insulating and protecting layer, called the jacket. The electric pulse to be transmitted is applied to the core electrode, while the external electrode is grounded. A function of the external conductor is to act as a shield against any electromagnetic interference from the environment. We shall now find the two most important characteristics of the cable: the wave velocity and the characteristic impedance.

Let R_1 and R_2 be the radiuses of the internal and external conductor, respectively, ε the dielectric constant of the insulator, which we assume to be a normal dielectric, and C_u and L_u the capacitance and the inductance per unit length, respectively. The latter are given by the expressions (see volume 3, Sects. 2.9 and 7.9)

$$C_u = \frac{2\pi\kappa\varepsilon_0}{\ln(R_2/R_1)}, \quad L_u = \frac{\mu_0}{2\pi}\ln(R_2/R_1), \tag{3.62}$$

where we have assumed, as is always the case in practice, the magnetic permeability of the insulator to be equal to that of a vacuum. We call the attention of the reader to the fact that the dielectric constant κ is, rigorously speaking, a function of the frequency of the electric field under dynamic conditions. This issue is discussed in Chap. 10 of the 3rd volume of this course, to which we refer. We only state here that, in the majority of cases, the frequencies at which the coaxial cables are employed are not too high and we can consider the dielectric constant to have its frequency independent electrostatic value.

Let us start our analysis by approximating the cable as a series of discrete elements, each of length Δx. Each element consists of a capacitor $\Delta C = C_u \Delta x$ to ground and an inductance in series $\Delta L = L_u \Delta x$, as shown in Fig. 3.10.

Calling $\phi(x)$ and $\phi(x + \Delta x)$ the electromotive forces between the plates of the capacitors at the positions x and $x + \Delta x$, respectively, the emf between the terminals of the inductor ΔL is $\phi(x + \Delta x) - \phi(x)$. We can then write

$$\phi(x+\Delta x) - \phi(x) = -\Delta L \frac{\partial I(x)}{\partial t} = -L_u \Delta x \frac{\partial I(x)}{\partial t}.$$

On the other hand, if $Q(x)$ is the charge on the capacitor ΔC at x, $Q(x)$ varies with time, because the current $I(x)$ acts to increase it and the current $I(x + \Delta x)$ to decrease it. We then have

$$I(x+\Delta x) - I(x) = -\frac{\partial Q(x)}{\partial t} = -\Delta C \frac{\partial \phi(x)}{\partial t} = -C_u \Delta x \frac{\partial \phi(x)}{\partial t}.$$

We divide the two equations by Δx and find

$$\frac{\phi(x+\Delta x) - \phi(x)}{\Delta x} = -L_u \frac{\partial I(x)}{\partial t}, \quad \frac{I(x+\Delta x) - I(x)}{\Delta x} = -C_u \frac{\partial \phi(x)}{\partial t}$$

and, taking the limit for $\Delta x \to 0$, we find

$$\frac{\partial \phi}{\partial x} = -L_u \frac{\partial I}{\partial t}, \quad \frac{\partial I}{\partial x} = -C_u \frac{\partial \phi}{\partial t}.$$

We now differentiate the first equation with respect to x and the second with respect to t and equate the resulting mixed derivative of I, obtaining

$$\frac{\partial^2 \phi}{\partial x^2} - L_u C_u \frac{\partial^2 \phi}{\partial t^2} = 0, \tag{3.63}$$

Similarly, differentiating the first equation with respect to t and the second with respect to x and equating the resulting mixed derivative of ϕ, we obtain

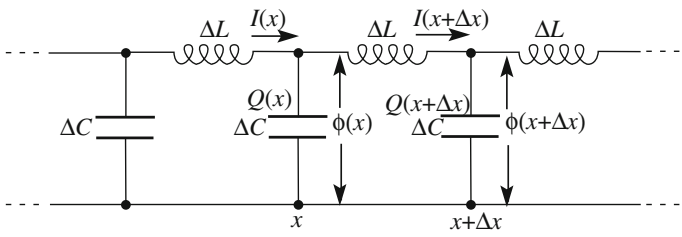


Fig. 3.10 Discrete approximation of a coaxial cable

$$\frac{\partial^2 I}{\partial x^2} - L_u C_u \frac{\partial^2 I}{\partial t^2} = 0. \tag{3.64}$$

We see that both the emf and the current intensity obey the same wave equation. The propagation velocity of the wave is

$$\upsilon = \frac{1}{\sqrt{L_u C_u}} = \frac{1}{\sqrt{\kappa \varepsilon_0 \mu_0}}. \tag{3.65}$$

We see that the wave velocity is independent of the geometry of the cable, namely the radiuses of the conductors. It is on the order of the speed of light in a vacuum, namely a factor $\sqrt{\kappa}$ of it. Plastic materials used as insulators have dielectric constants typically in the range $\kappa = 3 - 9$. In the approximation of neglecting the dependence of the dielectric constant on frequency, the wave in the cable is non-dispersive.

Let us now find the characteristic impedance of a cable. Our argument is similar to that which we developed in Sect. 3.2 for the elastic string. We apply an emf generator between the two conductors at one extreme. Initially, the charge and the current of the cable are zero. When the generator starts acting, it charges a segment of length $dx = \upsilon dt$ in the time interval dt. The capacity of this segment is $dC = C_u dx = C_u \upsilon dt$. Hence, the injected charge is $dQ = \phi\ dC = \phi\ C_u \upsilon dt$. The corresponding current intensity delivered by the generator is

$$I = \frac{dQ}{dt} = \phi C_u \upsilon = \phi \sqrt{C_u / L_u}.$$

Everything proceeds as if the generator would have to provide the current I on a load resistance

$$Z = \sqrt{L_u / C_u}. \tag{3.66}$$

This is the characteristic impedance of the cable.

With exactly the same arguments as those for an elastic string in Sect. 3.2, we can conclude that if the line is interrupted at a point and terminated in its characteristic impedance, physically a resistor of resistance Z, this appears to the wave as if the line would continue indefinitely. There are no reflections, and the energy traveling with the wave is absorbed and degraded to thermal energy by the resistor through the Joule effect.

Contrastingly, reflections at the far extreme exist in any other case. Suppose, for example, that we shorten the two conductors of the cable at the far extreme. The boundary condition is that the emf should be zero at the extreme. This is analogous to the fixed extreme of the string. We can think about satisfying the boundary condition with an imaginary voltage pulse traveling in the opposite direction, inverted relative to the incident pulse. As a result, the reflected voltage pulse is inverted. If the cable is open at the far end, the condition is that the current intensity

should be zero at the extreme. Under these conditions, the reflected current pulse, rather than the voltage one, is inverted.

Equation (3.66) holds for every type of cable. In the case of a coax, we can use Eq. (3.62) to express the capacitance and inductance per unit length in terms of the physical characteristics of the cable. We obtain

$$Z = \frac{1}{2\pi}\sqrt{\frac{\mu_0}{\kappa\varepsilon_0}} ln\left(\frac{R_2}{R_1}\right). \tag{3.67}$$

We see that the characteristic impedance depends both on the dielectric (through its dielectric constant) and on the geometry of the cable (the radiuses of the conductors). The square root factor would be the impedance of the empty space, namely 377 Ω if there were no insulator. In practice, it is a few times smaller. The other factors are, in general, smaller than one.

Question Q 3.4. A common standard for the characteristic impedance of coaxial cables is $Z = 50\ \Omega$. Consider such a cable having a core conductor diameter of 1 mm and a dielectric made of polyethylene, whose dielectric constant is $\kappa = 2.1$. What should the radius of the external conductor be? □

In practice, we can measure both the propagation velocity and the characteristic impedance using an oscilloscope and a pulse generator as follows. We connect the generator and the scope to the same extreme of the cable and leave the far extreme open. Having synchronized the scope with the pulse generator, we observe both the injected pulses and the reflected ones. Measuring the delay of the second and knowing the length of the cable, we have the propagation speed. We can subsequently measure the characteristic impedance by connecting the conductors at the far extreme through a variable resistor, and adjusting its resistance until we no longer observe a reflection.

3.13 Doppler Effect

Let us consider an approximately monochromatic sound source, for example, a tuning fork, in motion relative to an observer, or, better yet, a measurement instrument. Let the tuning fork be on a carriage movable along a straight rail and let us locate a microphone and a recording apparatus near the rail at a certain distance in front of the carriage and similar ones behind it. We first register the sound at both stations with the fork at rest. Then, we repeat our measurements with the carriage moving at constant velocity along the rail. Finally, we measure with the tuning fork at rest, and the registering apparatus in motion on a carriage. The question now is: are the frequencies, the wavelengths and the wave velocities measured in the different cases equal or different?

The effect we are discussing is the Doppler effect, named after Christian Doppler (Austria, 1803–1853), who first observed in 1842 that the tone of a note emitted by

a source in motion heard by an observer on the ground is higher when the source is approaching then when it is moving away.

We shall now analyze the effect, which exists for all wave phenomena. We shall consider two cases: sound waves and electromagnetic waves. In the first case, the wave propagates in a material medium, while in the second, there is no medium. We shall see that, for an electromagnetic wave, the effects depend on the relative motion of the source and detector only, while for the sound, they additionally depend on the velocity of the source and that of the detector relative to the medium. Indeed, in the case of sound, a "privileged" reference frame exists, being the reference in which air is at rest. In the case of the electromagnetic wave, there is no privileged reference, simply because there is no medium supporting the wave.

This important difference may appear to be obvious today, but it was not so before the development of special relativity, when a hypothetical medium, called ether, was assumed to exist to support the propagation of light. A privileged reference frame would have existed under this hypothesis, namely the frame in which the ether is at rest. Under these conditions, the Doppler effect for light would be similar to that for sound. The fundamental experiment in 1887 by Michelson and Morely proved that such a frame does not exist.

Let us now discuss the Doppler effect for sound. We shall limit the discussion to velocities, both of the source and of the detector, both smaller than the sound velocity, which we indicate with c. Clearly, all the velocities in the problem are much smaller than the speed of light, and we can surely consider time intervals and distances between two points independent of the reference frame.

Let S_0 be the reference frame in which the medium, namely the air, is at rest. Consider a source moving with uniform velocity v_S relative to S_0. Let the detector be at rest in S_0, located in front of the moving source along the straight line from the source in the direction of its velocity (the line of sight). The velocity of the wave in the medium (c) is a characteristic of the medium, and consequently is independent of the motion of the source. Let us consider two consecutive ridges of the wave emitted by the source in two instants separated by the time interval τ. In this interval, the source has approached the detector by the distance $v_S\tau$. Consequently, the time taken by the second ridge to reach the detector is shorter than the time taken by the first ridge of $v_S\tau/c$. This means that the time interval between the arrival at the detector of two consecutive ridges is not τ, but $\tau(1 - v_S/c)$. Let ν_o be the frequency of the wave emitted by the source and N the number of periods emitted in τ (not necessarily an integer number). We then have $N = \nu_o\tau$. These N periods reach the detector in the time interval $\tau(1 - v_S/c)$. Consequently, the frequency seen by the detector is

$$\nu = \frac{N}{\tau(1 - v_S/c)} = \frac{\nu_0}{1 - v_S/c}.$$

If the source moves away from the detector, again along the line of sight, the argument is similar and leads to the same result, with a + (plus) in place of

the − (minus). In conclusion, we can write, for the frequencies of the acoustic wave measured by a detector at rest in S_0, the expressions

$$\nu = \frac{\nu_0}{1 - v_S/c} \quad \text{source moving toward the detector.} \tag{3.68}$$

$$\nu = \frac{\nu_0}{1 + v_S/c} \quad \text{source moving away from the detector.} \tag{3.69}$$

Consider now the wavelength and let λ_0 be the wavelength when both source and detector are at rest in S_0, namely $\lambda_0 = c/\nu_0$. When the source moves, the velocity does not change, but the frequency does, as we just saw. Consequently, the wavelength becomes $\lambda = c/\nu$. Using Eqs. (3.68) and (3.69), we can write

$$\lambda = \lambda_0(1 - v_S/c) \quad \text{source moving toward the detector.} \tag{3.70}$$

$$\lambda = \lambda_0(1 + v_S/c) \quad \text{source moving away from the detector.} \tag{3.71}$$

In other words, when the source is in motion, an experiment measuring the wave velocity gives us the same result as when the source is at rest, an interference experiment to measure the wavelength gives us different results, and different results give us an experiment measuring the pitch of the sound.

Let us now consider the case in which the detector moves in a rectilinear uniform motion with velocity v_D relative to the medium. The source is at rest in the medium. The motion is, again, along the line joining source and detector. Consider first the case in which the detector approaches the source. Now, the wave velocity of the wave relative to the detector is not the velocity in the medium, but rather $c + v_D$. The remaining portion of the argument is identical to that of the previous one. We can thus obtain our result by substituting $c + v_D$ in the place of c in the previous one and v_D in place of v_S. We find that the frequency measured by the detector is

$$\nu = \nu_0[1 - v_D/(c + v_D)] = \nu_0(1 + v_D/c).$$

The argument in the case of the detector moving away from the source is similar, and we can write the conclusions as

$$\nu = \nu_0(1 + v_D/c) \quad \text{detector moving toward the source.} \tag{3.72}$$

$$\nu = \nu_0(1 - v_D/c) \quad \text{detector moving away from the source.} \tag{3.73}$$

We see that the expression is different from the case in which the source was moving and the detectors standing. However, ν is larger than ν_0 in this case as well.

Hence, in the case of the detector moving relative to the medium and the source at rest in the medium, the frequency measured by the detector and the wave velocity are different from those measured by a detector at rest. Contrastingly, as easily verified, the wavelengths are equal.

Question Q 3.5. Prove the last statement.

Suppose now that both detector and source move relative to the medium of rectilinear uniform motions along the same line, parallel to the direction of the wave, with velocities v_D and v_S, respectively. The problem of finding the frequency measured by the detector is solved, for example, in the case in which the detector runs after the source in the same direction using, in sequence, Eq. (3.72) first and then Eq. (3.69). One finds

$$\nu = \nu_0 \frac{1 + v_D/c}{1 + v_S/c}. \tag{3.74}$$

Note that when the velocities of the source and detector relative to the medium are equal, namely if $v_D = v_S$, for example, if the source and detector are both on a carriage moving with that velocity, then the measured frequency is the same as that when both are at rest. Contrastingly, if the velocities are different, the effect depends separately on v_D and v_S, and not simply on the relative value v_D–v_S. Consequently, experiments based on the Doppler effect determine the velocities of the source and detector relative to the medium.

We shall not address the general case here in which the line joining the source and detector is at an angle with the propagation velocity different from zero. We only observe that one must consider the projections of the velocities in that direction.

Let us now consider electromagnetic waves. As we know, there is no medium supporting the waves, as they propagate in a vacuum. As a consequence, as opposed to sound waves, there is no privileged reference frame in which a supporting medium would be at rest. All the experiments have shown that only the *relative* motion of the source and detector can be experimentally detected, while the concept of *absolute* motion, say relative to a vacuum, has no physical sense. This was clearly shown by Albert Michelson (USA, 1852–1931) and Edward Morley (USA, 1838–1923) in their famous experiment in 1887, which we discussed in the 1st volume. On that basis, and on that of other experimental results, Henry Poincaré (France, 1854–1912) concluded, in 1900, at the International Congress on Physics in Paris:

> Does our ether really exist? I do not believe that more precise observations will never be able to reveal nothing but relative motions.

We thus expect that equations like those valid for sound from Eqs. (3.68) to (3.74) cannot hold for electromagnetic waves. Otherwise, the Doppler effect would allow us to determine the absolute motion. Let us look at the reason for that, considering the case having the detector at rest in an inertial frame and the source moving with uniform velocity v_S. The argument is equal to that which we developed for the sound, with the important difference that the time intervals cannot be considered invariant. The time interval τ between two consecutive ridges is a *proper time* because it is in the frame in which the source is at rest (which is inertial as well). The frame in which the detector is at rest moves relative to that with velocity v_S. Consequently, the time interval between ridges appears to the detector

dilated by the factor $\left[1-(v_S/c)^2\right]^{-1/2}$, where c is now the speed of light, namely of our wave. Repeating the argument we made for the sound with this important variant, we state that the interval between consecutive ridges at the detector is

$$\tau(1-v_S/c)\left[1-(v_S/c)^2\right]^{-1/2}=\tau[(1-v_S/c)/(1+v_S/c)]^{1/2}.$$

In conclusion, the frequency measured by the detector is

$$\nu=\nu_0\sqrt{\frac{1+v_S/c}{1-v_S/c}} \quad \text{distance between detector and source decreasing.} \tag{3.75}$$

$$\nu=\nu_0\sqrt{\frac{1-v_S/c}{1+v_S/c}} \quad \text{distance between detector and source increasing.} \tag{3.76}$$

The same result is reached in the case of the detector moving and the source at rest. In this case, the contraction of the distances must be considered.

Question Q 3.6. Prove the last statement. □

The Doppler effect is important for several aspects of physics, because it allows for measuring the speed of far away, or, in general, not directly accessible, objects. We must purposefully recall that each species of molecule and atom emits electromagnetic radiation oscillating in a set of proper frequencies, which are characteristic of the species. A record of the light intensity as a function of frequency is called a spectrum. The spectrum of an atomic or molecular species shows a series of peaks, called lines, at the proper frequencies. If the atoms, or molecules, move relative to the detector, all the line patterns in the spectrum shift as a result of to the Doppler effect. Measuring that shift, we can determine the speed of the emitting object relative to our detector or better, the component of that speed on the line of sight.

Consider, for example, a gas in a transparent container. We give energy to the gas molecules in order to excite their vibrations, for example, with an electric discharge. We want to have them completing their oscillation motions without colliding with other molecules. To insure that, we reduce the density to a small enough value. Measuring the spectrum, we observe that the lines are broader than their natural value and that the broadening increases with increasing temperature. The effect, called Doppler broadening, is due to the chaotic thermal motion. Some molecules approach the detector, and the observed frequency increases, some others move away, and the observed frequency decreases. The broadening increases with temperature, because the average kinetic energy is proportional to the absolute temperature.

As a second example, consider a star, or a heavenly body, moving relative to an observer on earth. Its entire spectrum appears to be shifted to lower frequencies (an effect called redshift) if the star is moving away; contrastingly, if it is approaching, the spectrum is "blue-shifted". Measuring these shifts, we can determine the velocity

of the object along the line of sight. In 1929, Edwin Hubble (USA, 1989–1953), measuring the distance and the redshift of several galaxies, discovered that they were moving away with velocities proportional to their distances. This was the discovery of the expansion of the Universe that opened the way for modern cosmology.

Summary

In this chapter, we learned the following important concepts.

1. A pulse of an elastic string travels with non-altered shape with a definite velocity dependent on the string density and tension.
2. The concept of a phase is defined for harmonic waves. The phase varies with time at a given point by ω radiants per second and in a given instant by k radiants per meter.
3. The phase of a monochromatic wave moves with the velocity ω/k.
4. A dispersion relation is the relation between ω and k and is a characteristic of the medium. This is an extremely important relation for all wave phenomena.
5. To inject a progressive wave into an elastic string, one must apply to its input extreme a force proportional to the velocity of that extreme. The proportionality constant is the characteristic impedance of the system
6. An elastic string is like a transmission line of elastic pulses. To avoid reflections at the far extreme, one must terminate the string on its characteristic impedance. The situation is the same for electric cables.
7. The concept of a wave vector for a harmonic wave in space.
8. Sound waves in a gas and their phase velocity.
9. The electromagnetic waves predicted by the Maxwell equations, light in particular, and their characteristics.
10. How waves, sound and electromagnetic, in particular, transport energy.
11. Accelerating charges are the sources of electromagnetic waves. We gave a very useful expression of their electric field at a distance.
12. The square wave detectors are sensitive to the square of the wave amplitude averaged over a period, or a time much longer than the period.
13. The Doppler effect for sound and light. The fundamental differences between the two cases.

Problems

3.1. Consider the harmonic progressive $x = (10\ \text{mm})\cos[(10\ \text{s}^{-1})t + (8\ \text{m}^{-1})k]$ wave and calculate frequency, wavelength and phase velocity.

3.2. Figure 3.11 is a snapshot of a progressive harmonic wave on an elastic string moving to the right. Establish for each of the marked points whether it moves up or down. Does A or B move more quickly? Answer the same questions for the case in which the photo is of a standing wave at maximum deformation.

3.3. In which of the following cases is there a reflection and in which are there not? (a) Wave on an elastic string with free extreme. (b) Wave on an elastic string with fixed extreme. (c) Current pulse on a cable with extreme in short. (d) Current pulse on a cable with extreme open.

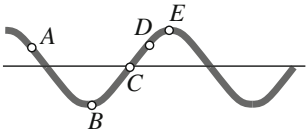


Fig. 3.11 A snapshot of a harmonic wave

3.4. How much does the characteristic impedance of an elastic string vary if; (a) You double the tension, (b) You double the density? How much does the characteristic impedance of a cable vary if; (a) You double the inductance per unit length, (b) You double the capacitance per unit length?
3.5. An observer at a distance of 1 km from a charge oscillating in a harmonic motion measures a certain oscillation amplitude of the radiation electric field. A second observer does a similar measurement at 10 km from the source. What is the ratio of the two amplitudes?
3.6. Two coaxial cables have the same geometry and dielectric constants, one twice the other. What are the ratios of their characteristic impedances and their phase velocities? Is the system dispersive?
3.7. How much do the sound velocities in air differ in Death Valley at 50 °C and on the Mont Blanc at −40 °C?
3.8. What is the electric radiation field at 10 m distance from a charge of 1 °C moving in a rectilinear uniform motion with velocity $0.3c$?
3.9. What is the ratio between the pressure oscillation amplitudes of two sounds differing by 2 dB?
3.10. An electromagnetic monochromatic wave propagates in a dielectric having constant $\kappa = 3$. What is the ratio of the electric and magnetic field amplitudes? What is the phase velocity?
3.11. Consider the monochromatic plane electromagnetic wave in a vacuum $\mathbf{E} = \mathbf{E}_0\cos(\omega t - kx)$ advancing in the x direction. Find the direction, sense and magnitude of the Poynting vector.
3.12. The magnetic field amplitude of an electromagnetic plane monochromatic wave in a vacuum is 10^{-6} T. Determine: (a) The amplitude of the wave electric field, (b) The average wave intensity, (c) The average energy density.
3.13. A car traveling at 72 m/s carries a loudspeaker emitting a note at 1 kHz. The sound waves emitted in the forward direction encounter an obstacle and are reflected back, adding to the primary waves. What is the beat frequency of the resulting signal?
3.14. Superman is traveling at a low altitude along an avenue in Metropolis. Coming to a traffic light, he sees it as green (520 nm) and crosses. A policeman stops him, stating he had crossed on the red (650 nm). Assuming both to be right, find the velocity of Superman.
3.15. A train whistles while passing near an observer standing on the ground. When the train passes from front to back, the observer perceives a change in the tone of the whistle. By how much does the perceived frequency change if the train travels at 90 km/h?

Chapter 4
Dispersion

Abstract In this chapter, we study dispersive waves. Examples of these are the waves on the surface of water and the electromagnetic waves in a transparent medium. Dispersive waves, when monochromatic, propagate with a phase velocity that depends on wavelength, where they change shape while they propagate if they are not monochromatic. We introduce the concepts of group velocity and energy propagation velocity. We see how light velocities have been measured and study the phenomena of reflection, refraction and dispersion. We also discuss the physical origin of the refractive index.

In the previous chapter, we considered non-dispersive waves. If they are monochromatic, their phase velocity is independent of frequency; in any case, monochromatic or not, their shape remains unaltered during propagation. However, a different type of wave exists, the dispersive waves. Examples of these are the waves on the surface of water and the electromagnetic waves in a transparent medium. In these cases, the wave equation is more complicated and might even be unknown. These waves, if monochromatic, propagate with a wavelength-dependent phase velocity, while they change in shape while they propagate if they are not monochromatic. We shall see that, even if the ruling differential equation is not known, the knowledge of the dispersion relation, which links frequency and wave number, is sufficient for our understanding of many phenomena.

We shall see that different concepts of velocity need to be defined for a dispersive wave beyond phase velocity. The most important of these is group velocity, because energy and information propagate with group velocity under almost all circumstances. In Sect. 4.2, we shall consider the measurement of group and phase velocity of light.

In Sect. 4.3, we study the behavior of a light wave at the interface between two transparent media, namely the reflection and refraction phenomena and the Snell law governing the latter. After an interlude during which we will go into the fascinating phenomenon of the rainbow in Sect. 4.4, we shall justify the laws of reflection and refraction from the undulatory point of view in Sect. 4.5. We then

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_4

discuss the dispersion of white light in the different colors in a medium due to the dependence of phase velocity on the wavelength.

In Sect. 4.7, we try to understand the physical origin of the difference between light velocity in a medium and that in a vacuum and of dependence on wavelength. We do that in a low-density medium in order to focus on physics, avoiding mathematical complications. We shall find, in particular, how the elastic and absorptive resonance curves studied in Chap. 1 will now become useful. Finally, in Sect. 4.8, we shall find the wave equation and the dispersion relation of an electromagnetic wave in a dense normal dielectric.

4.1 Propagation in a Dispersive Medium. Wave Velocities

The progressive waves we encounter in physics are never exactly monochromatic (the terms harmonic and sinusoidal are synonyms, as we know), if for no other reason than the fact that they always have a beginning and an end, and therefore a limited duration. Obviously, a "truncated" sine is far more similar to a mathematical sine in as much as the duration is long compared to the period. In other cases, the wave shape is more different from a sine. Think, for example, of the elastic wave propagating on a metal bar hit with a hammer at an extreme or of the sound of a hand clapping propagating in air. We shall now study how non-harmonic waves propagate in a medium. We shall see that several concepts of propagation speed exist depending on the propagating quantity associated with the wave.

As we already mentioned in Chap. 3, we can distinguish the following two cases, depending on the nature of the wave and that of the medium.

(1) *non-dispersive wave*, when the dispersion relation is a proportion, namely when the wave number of a monochromatic wave is directly proportional to its angular frequency. Under these conditions, the phase velocity is the same for all the monochromatic waves, independent of their frequency. In addition, the non-monochromatic waves propagate with invariable shape. The partial differential equation of the system is the wave equation in Eq. (3.1).

(2) *dispersive wave*, when the dispersion relation is not a proportionality relation. Under these circumstances, the phase velocity of the monochromatic waves depends on frequency (and wavelength) and the non-monochromatic waves change their shape when they propagate. This is what happens for the electromagnetic waves in a dielectric (like glasses and water) and for the most evident waves, namely the waves on the surface of water. We shall discuss the latter example later in this section. The differential equation for dispersive waves is not Eq. (3.1). As a matter of fact, we shall not need to know the equation, because the dispersion relation will be sufficient to describe the phenomena.

A pulse or a wave of similar shape is not monochromatic and we cannot rigorously speak of phase velocity. The pulse substantially advances with the group velocity. This is a very important quantity that we will now define. In these circumstances, a "group" is a group of wave crests of finite duration (or length in space).

A non-monochromatic wave has several, in general, infinite, Fourier components. Each of them is monochromatic and has a phase velocity, different from one another. Let us start by considering the simplest case, namely a superposition of two monochromatic waves only (dichromatic wave) of different frequencies, say ω_1 and ω_2 with $\omega_2 > \omega_1$. Let us assume, for simplicity, that the two amplitudes are equal. The resulting wave is

$$\psi(z,t) = A\,\cos(\omega_1 t - k_1 z) + A\,\cos(\omega_2 t - k_2 z). \tag{4.1}$$

The problem is very similar to the case of the beats we discussed in Sect. 2.6. The difference is that the function now depends not only on time but also on a space coordinate. In any case, similarly to what we did then, let us put forth

$$\omega_0 = \frac{\omega_1 + \omega_2}{2}, \quad \omega = \frac{\omega_2 - \omega_1}{2}, \quad k_0 = \frac{k_1 + k_2}{2}, \quad \omega = \frac{k_2 - k_1}{2} \tag{4.2}$$

and re-write Eq. (4.1) in the form

$$\psi(z,t) = f(z,t)\cos(\omega_0 t - k_0 z), \tag{4.3}$$

where

$$f(z,t) = A\,\cos(\omega t - kz). \tag{4.4}$$

To better understand the phenomenon, let us consider ω_1 and ω_2 to be very close to one another, as in the case of the beats. Under these circumstances, it is $\omega_0 \gg \omega$. Then, Eq. (4.3) represents a progressive wave, which is almost harmonic with angular frequency ω_0. Its amplitude $f(z,t)$ is not constant, either in time or in space, but can be thought of as being a progressive wave itself. It is harmonic with angular frequency ω, as Eq. (4.4) tells us. Figure 4.1 shows a snapshot of the wave at a certain instant.

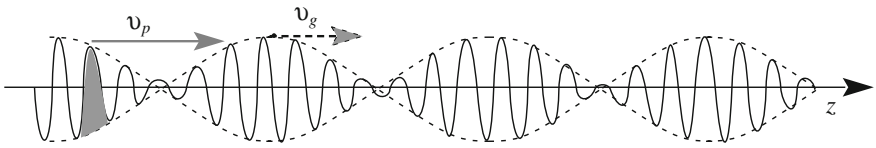


Fig. 4.1 Phase and group velocities in a dichromatic wave

The higher frequency (ω_0) wave, which is called a carrier wave, propagates with a phase velocity given by

$$v_p = \omega_0/k_0. \tag{4.5}$$

This is the velocity of the wave crests of the smaller wavelength, whose amplitudes are modulated by the modulating wave. This is also a sine, with a smaller frequency (ω).

Let us now look at the propagation velocity of the modulation. This is an extremely relevant question. The velocity of the modulation is the propagation velocity of the information. Indeed, all the crests of a non-modulated sine wave invariably have exactly the same shape, and they cannot transport any information. If we want to send a message to another observer, we must alter at least one of these crests. In doing so, we modulate the wave and make its frequency spectrum non-monochromatic.

With reference to Fig. 4.1, we now consider the velocity with which an arbitrary point of the modulating wave propagates. We choose a maximum, namely a point at which $f(z,t) = 2A$. We find its velocity by imposing the argument of $f(z,t)$ to remain constant, namely for the total differential of $f(z,t)$ to be zero, that is, $d(\omega t–kz) = 0$. This is also $\omega dt–kdz = 0$. Hence, the ratio between dz and dt, namely the velocity of the modulation maximum, which we call v_g, must be

$$v_g = \frac{dz}{dt} = \frac{\omega}{k} = \frac{\omega_2 - \omega_1}{k_2 - k_1}.$$

Taking the limit for $\omega_2 - \omega_1$ going to zero, we have

$$v_g = \frac{d\omega}{dk}. \tag{4.6}$$

This expression, which we have found in a particular case, is taken by definition as the *group velocity* under any circumstance. Its physical meaning depends somewhat on both the waveform and the medium. However, it is always true that only in a non-dispersive medium in which the dispersion relation is $\omega = vk$ are group and phase velocities equal to one another.

The example we have just discussed is useful for reaching the concept of group velocity with a simple argument, but it is also unrealistic. Figure 4.2 shows a more realistic case of a "wave packet", in which the term means a group of wave crests. We may think of one of the infinite groups in Fig. 4.1.

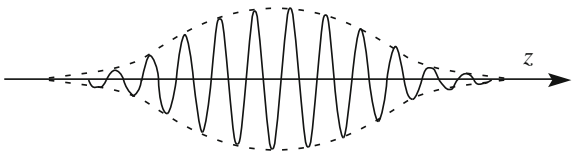


Fig. 4.2 A wave packet, or wave group

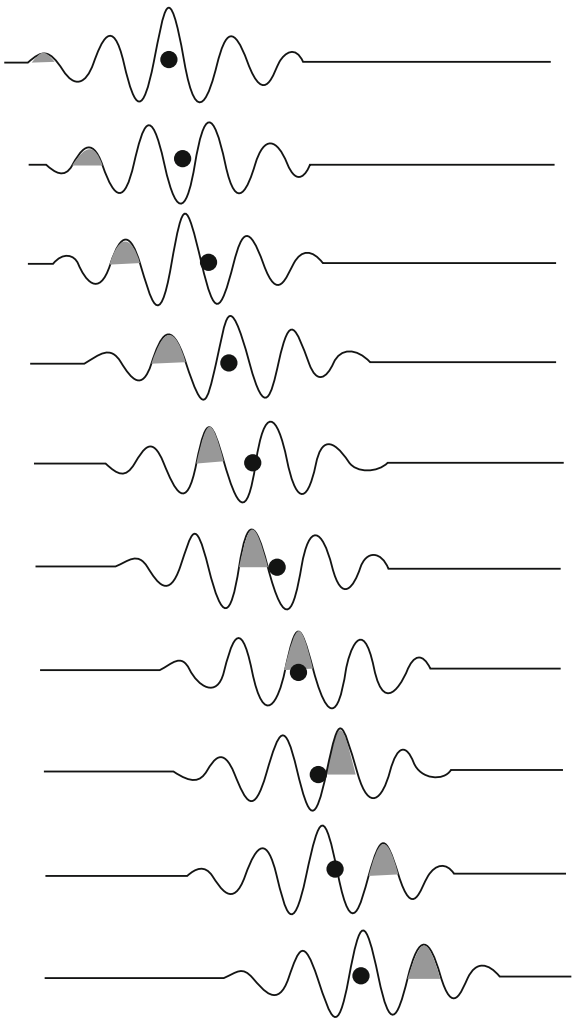


Fig. 4.3 Evolution of a wave packet. The phase velocity is twice the group velocity

In a good approximation, the wave packet (or wave group) is a "sine" of finite length. We can recognize it as having a main wavelength, or, in an equivalent manner, a main wave number, which we call k_0. The amplitude of the "sine" has a maximum in the middle of the packet and gradually vanishes on both sides. Figure 4.3 helps us in understanding how a wave packet propagates in a dispersive medium. The figure shows a series of snapshots taken at regular time intervals. In this example, the phase velocity is twice as large as the group velocity. In the figure, we have marked a certain wave crest with the color gray. It moves forward at the phase velocity. We have also marked, with a black dot, the center of the group. It advances with the group velocity. The reader can appreciate the difference between the velocity of the group and the velocity of the phase of the dominant wavelength

component by following the two marks with his/her eye. Being that the phase velocity is larger than the group velocity, a crest born in the rear of the packet moves forward relative to the group. In doing so, its amplitude grows, reaches a maximum when it is in the middle of the packet, and then decreases until it disappears in the front of the group.

This phenomenon can be observed with a simple, although not easy, experiment on the still surface of a pond. As we shall see at the end of the section, the phase velocity of the water waves is twice as large as the group velocity under these conditions. You can produce a wave group by placing a wooden stick on a still water surface (in a swimming pool when nobody is swimming, for example) and then moving it rhythmically up and down four or five times. Then, try to observe a wave crest appearing at the tail of the packet, overtaking it and disappearing at the front. You shall need to practice somewhat, and work with a friend, because the speed of the waves is quite fast, on the order of a few meters per second.

We now discuss, in general, why a non-harmonic wave changes shape when it propagates in a dispersive medium. Let us think about taking a snapshot of the wave at a certain instant t. We obtain a curve that is a function of the coordinate. What will be its shape after a certain time interval Δt? To answer this question, we start by developing the function we have found in a Fourier integral. If $v_p(k)$ is the phase velocity at the wave number k, the phase of the monochromatic component of that wave number advances by the distance $v_p(k)\Delta t$ in the considered time interval. The dependence on k of v_p implies that the differences between the phases of the Fourier components at $t + \Delta t$ are different than those at t. To find the new shape, we must now anti-transform. As the phase differences have changed, the resulting shape is different from what it was at t.

Two further important concepts are the energy propagation speed and the information propagation speed. Under many circumstances, but not all, the two velocities coincide. This is always the case when waves are detected by a square-law detector. Indeed, as we have already discussed, the signal delivered by such detectors is proportional to the energy flux received, integrated over a time interval that is much longer than the period (or the periods) of the wave. Consequently, the signal is proportional to the average of the squared amplitude.

Under many circumstances, the energy and the information propagation speeds for a wave packet (or a wave group) are both equal to the group velocity. Indeed, we have found the group velocity by considering the propagation velocity of the modulation maximum. Around that point, the largest fraction of the energy transported by the wave is located. Consider an observer A sending a wave packet signal to an observer B. We distinguish two cases, depending on whether the detector in B is a square-law detector or a detector sensitive to the amplitude. In the former case, which is the case at high frequencies, such as those of light, the detector will signal the arrival of the pulse substantially in the instant in which it is reached by the maximum of the modulation. Consequently, the measured velocity is the group velocity.

The situation is different if the detector in B is sensitive to the amplitude. Think, for example, of the packet in Fig. 4.2 as consisting of waves on the surface of water

or of elastic waves in a dispersive medium. The signal in B can then be recorded with a pen connected to a buoy or with a microphone, respectively. As such, the signal will be detected as soon as the amplitude of the incoming wave has risen above the sensitivity threshold of our instrument. The information propagation speed is not necessarily the group velocity under these circumstances.

The information propagation speed may also be different from the group velocity in Eq. (4.6) for a square-law detector if the waveform is different from a wave packet. Take, for example, a "beginning sine", the ideal case of a wave that starts its existence at $t = 0$ and then continues as a sine forever. In this case, a square-law detector in B will not necessarily detect the front of the incoming wave, which is altered by dispersion, as having moved at the group velocity. As a matter of fact, the group velocity of light in a dispersive medium may be larger than the light velocity in a vacuum, as we shall see in the subsequent sections. This is quite an exceptional case, but historically, it raised doubts about a possible violation of the relativity principle. However, it was soon shown that, under those circumstances, the information propagation speed is different from group velocity and is, in any case, smaller than c.

Another similar example is the ideal case of a "finishing sinusoid", in which a sinusoidal wave that always existed suddenly ends at $t = 0$. We shall see later in this section how the speed of the end of the waves, as opposed to the above quoted exception, is equal to the group velocity for the surface waves on water.

The most familiar example of wavy phenomena is undoubtedly the waves on water surfaces. They are strongly dispersive and give us the opportunity to discuss the concepts we have introduced in concrete examples. Note, however, that a mathematical description of the surface waves is not at all simple, and we shall not enter into the demonstrations. Let us look at the principal characteristics.

First of all, the surface waves are neither transversal nor longitudinal. Let us fix our attention on a particle of water on the surface (or immediately below the surface) and on its trajectory when the wave passes through. The motion is neither an oscillation on a line perpendicular to the wave velocity, nor one on a line in the velocity direction. Looking carefully, we see that the water particle trajectory is an ellipses. This is simply a consequence of water being incompressible. If, at a certain point, the water surface should rise because a wave ridge is forming, more water has to move there, in practice, from a point in front, where a gorge is forming.

Secondly, as we anticipated, the surface waves are dispersive. The dispersion relation depends on the restoring forces acting on the water particle when it is out of equilibrium. These are the forces that make water surfaces flat at equilibrium (considering a dimension much smaller than the earth's radius). As a matter of fact, there are two restoring forces: the weight and the surface tension. Surface tension is important for the small ripples having wavelengths on the order of a millimeter (waves in a coffee pot). For larger wavelengths, the weight completely dominates. We shall only discuss the latter conditions here.

Thirdly, effects that are quite often dissipative are present. For example, in shallow water, the wave motion can reach the bottom of the basin, and here the friction with the ground can subtract energy from the wave. This is what happens

when a wave approaches a beach. We see that when the water depth decreases enough, the crest of the wave overturns and the wave breaks. Under these conditions, the system is not linear.

We shall limit our discussion to the gravity waves on the surface of an ideal liquid, which is incompressible (constant density ρ) and inviscous (viscosity $\eta = 0$).

Let h be the depth of the basin, namely the distance between the surface and the bottom. Observing the motion, we can state that the elliptic trajectories of the water particles have decreasing diameters with increasing depth. At the depth of one wavelength, the amplitude of the motion is reduced to 0.2 % of the amplitude on the surface. In addition, it can be shown that the dispersion relation is

$$\omega^2 = kg \tanh(kh), \tag{4.7}$$

where g is the gravity acceleration. Clearly, harmonic waves of different wavelengths have different wave velocities. We draw the attention of the reader to the argument of the hyperbolic tangent function, which is $kh = 2\pi h/\lambda$. It is proportional to the ratio of two lengths, the depth of the basin and the wavelength of the wave. There are two interesting limit cases, in which one of the two lengths is much greater than the other.

When $h \gg \lambda$, we speak of a deep water wave. In this case, considering that $\lim_{h\to\infty} \tanh kh = 1$, we have

$$\omega^2 = kg. \tag{4.8}$$

The phase velocity is then

$$v_p = \frac{\omega}{k} = \sqrt{\frac{g\lambda}{2\pi}}. \tag{4.9}$$

The group velocity is one half of the phase velocity, namely

$$v_g = \frac{d\omega}{dk} = \frac{1}{2} v_p. \tag{4.10}$$

We note that both velocities are independent of the depth of the water. This is because the motion of the water particles has already ceased at a depth on the order of a wavelength, as we already mentioned, and consequently, everything proceeds as if the bottom did not exist.

The phase velocity grows proportionally to the square root of the wavelength, namely longer waves travel faster than shorter ones. Figure 4.4 shows phase and group velocities as being functions of the wavelength. If, for example, a speedboat produces waves in the sea somewhat offshore, an observer on the beach will see the longer waves arriving first. Subsequently, waves of shorter wavelengths, and higher frequencies, will gradually appear.

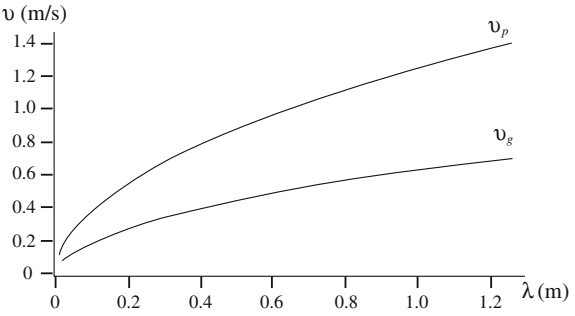


Fig. 4.4 Phase and group velocities of surface waves in deep water versus wavelength

QUESTION Q 4.1. What is the period of a sea wave of 10 m wavelength in deep water? And that of a wave of 2 m wavelength? What are their group and phase velocities? □

Let us now consider the opposite case of shallow water waves. We can now approximate $\tanh(kh) \approx kh$, and Eq. (4.7) becomes

$$\omega^2 = ghk^2. \tag{4.11}$$

Under these conditions, the system is non-dispersive, namely ω is proportional to k. Consequently, the phase and group velocities are equal to one another. They are given by

$$\upsilon_p = \upsilon_p = \sqrt{gh}. \tag{4.12}$$

The velocity increases with the square root of the depth.

The non-dispersive property of shallow water waves explains the disastrous phenomenon of the tsunamis originated by high magnitude quakes having their epicenter in the open sea. The seismic movement of the sea floor propagates to the entire water column up to the surface. The height of the column may be of a few kilometers. The resulting waves generally have small amplitude, only several centimeters in the open sea, but the mass of moving water is huge. The wavelengths are enormous, up to one hundred kilometers, namely so large that even the ocean is shallow water for them. As a consequence, these waves propagate without deformation, transporting all the original energy in a few crests. Their speed in the deep sea is very high, comparable to a sound wave in the atmosphere. Indeed, Eq. (4.12) gives us, for example, for a depth h = 3 km, υ_g = 170 m/s. This corresponds, for a typical wavelength of 170 km, to a period of 1000 s. At this speed, if the epicenter is, say, at 10 km offshore, the waves will reach the seacoast in about 10'. When the wave is close to shore and the sea depth decreases, the height of the submerged liquid column and the speed of the wave decrease as well. In order for energy to be conserved, a large fraction of the kinetic energy transforms into potential energy and the height of the wave increases enormously. When the two or three crests hit the coastline, they sow destruction everywhere they hit.

Let us now schematically look at how we can measure the different velocities of the surface gravity waves in water. This will help us to understand how phase velocity, group velocity and energy (information) velocity are also different quantities when considered from the operational point of view.

Let us start by recalling the definitions. The phase velocity of a harmonic wave is

$$v_p = \frac{\omega}{k}, \tag{4.13}$$

which can be written in terms of the wavelength and the frequency as

$$v_p = \lambda \nu. \tag{4.14}$$

The group velocity is

$$v_g = \frac{d\omega}{dk}. \tag{4.15}$$

Suppose we perform our measurements in a canal of square cross-section, 1 m wide and 1 m deep, and some 20 m long. This is a wave-guide in which we can inject waves moving up and down a wooden septum on the surface at one extreme of the canal. In this way, we produce a wave group of a few crests separated by, say, a dozen centimeters and then study its propagation. We have prepared two reference lines forming a baseline that we have measured. We shall determine the velocities by measuring the time taken by the wave to travel through this base.

Let us start with the phase velocity. We can measure it by two methods.

With the first method, we fix our attention on a particular crest of the packet and measure the time it takes for it to go through the base. The phase velocity is obviously the ratio between the base length and this time.

With the second method, we take a photo of the packet at a certain instant and measure, on the picture, the average distance between crests. This is the average wavelength. We had also arranged a small buoy connected to a recorder that provides us with a record of the height of the surface as a function of time. From this record, we can easily extract the average oscillation frequency. Having measured wavelength and frequency, we multiply them to obtain the phase velocity.

The two methods are operationally different, but, as we can verify, give the same result within the experimental uncertainties.

We now measure the group velocity with a method similar to the first one above. However, we now look at the times of passage at the two lines of the maximum of the packet rather than of a single crest. In this way, we find that the group velocity is one half of the phase velocity.

We can finally measure the energy velocity by measuring the speed of the end of the waves. We inject into the guide a certain number of crests, fix our attention on the moving point at which the wave motion is just finished, and measure the time it takes to go through the base. We thus find the velocity with which energy propagates through the system. Note that, operationally, energy and group velocities are

different concepts. In this case, we experimentally find that energy and group velocities are equal to one another. However, this might not be the case for other systems.

4.2 Measurement of the Speed of Light

The word "light", strictly speaking, means the electromagnetic radiation in the frequency interval in which it can be "seen" by human eyes. In this section, we describe the first historical measurements of the speed of light. They were accomplished with astronomical observations, hence in vacuum, and subsequently in the atmosphere, in which the difference from a vacuum is, however, very small.

Let us start by recalling a few fundamental concepts that should be well known to the reader. The speed of light is extremely fast, or, more properly stated, it is the largest possible velocity. Indeed, as we discussed in the 1st volume of this course, the relativity principle, originally stated by Galileo Galilei (Italy, 1564–1642), holds for all laws of physics. We also recall that it can be shown that, under very general assumptions on the properties of space and time and the validity of the cause and effect principle, only two sets of transformations between the spatial and time coordinates of two inertial reference frames exist, namely the Galilei and the Lorentz transformations. The latter contain a constant, c, which has the physical dimensions of a velocity. The Galilei transformations are the limit of the Lorentz transformations for $c \to \infty$. The quantity c is one of the fundamental constants of physics, is invariant under Lorentz transformations, is the largest possible velocity, and is the velocity of the electromagnetic and gravitational waves in a vacuum. Given its importance, c was measured with precision increasing with time. The relative uncertainty was as small as four parts per billion ($\Delta c/c < 4 \times 10^{-9}$) in 1975, when the Conférence Générale de Poids et Mesures recommended redefining the unit of lengths, the meter, as the distance traveled by light in a vacuum in 1/299 792 458 of a second. Consequently, in the SI system of units, the value of c is fixed by definition to be equal to

$$c = 299\,792\,458 \ \mathrm{m\,s^{-1}}. \tag{4.16}$$

We finally recall that we learned in Sect. 3.6 that the wave equation in Eq. (3.31) is a solution to the Maxwell equations in a vacuum. Consequently, light waves in a vacuum are not dispersive and the phase and group velocities are equal.

Considering the extremely large value of the speed of light, it is not surprising that its propagation has been considered as instantaneous through the millennia of human culture. It was G. Galilei who first thought that light might propagate with a finite velocity and tried an experiment to measure it, as he later wrote in his "Two new sciences" published in 1638. The idea was as follows. Two people hold two lanterns, such that, by interposition of the hand, the one can shut off or admit the light to the vision of the other. They first practice, at a short distance from one

another, a scenario in which one, in the instant he sees the light of his companion, uncovers his own. Having acquired the skill necessary for a response with a "negligible delay", the two experimenters perform the same experiments at night at two locations separated by "two or three miles". Galilei could not measure any delay under such conditions. Indeed, we know now that the delay in observing over a back and forth distance of about 10 km is about 30 μs, far too short a time to be detectable in this experiment.

Much longer delays can be observed on astronomical baselines. Note that in an experiment on earth, one can measure the time spent by light in back and forth travel, as in the Galilei experiment and all those that followed, while an astronomical observation must be based on a light travelling in one direction only, from the source in the sky to the detector on earth. As a consequence, the source must be periodic, namely a clock. Its period should be accurately known and short enough to allow observations over many periods. Indeed, such heavenly clocks had been available, with the discovery by Galilei of the Jupiter satellites. In 1669, Gian Domenico Cassini (Italy-France, 1625–1712) was called to Paris by King Luis XIV to build an advanced astronomical observatory, the Observatoire des Paris. In the following years, Cassini, initially by himself and, from 1672, with his assistant Ole Rømer (Denmark, 1644–1710), performed systematic measurements and calculations on the Jupiter satellites. They measured the intervals between two subsequent occultations by Jupiter of each satellite. For a given satellite, these intervals measured the orbit period. However, the two astronomers observed that the interval between occultations, in particular of the most external one, Io, continuously increased over a period of six months and decreased over the subsequent six months. The maximum delay amounted to about 22 min. They calculated and applied all the necessary corrections, but the effect did not disappear.

In 1675, Cassini concluded that it was very likely that the delay was due to the time "taken by light to reach us". On November 22, 1676, Rømer finally presented to the French Academy his "memoire". The delays and advances were due to the finite velocity of light. Consider the scheme shown in Fig. 4.5 and assume, for a moment, that Jupiter is immobile in the position J in the figure. The occultations of Io are luminous signals leaving Jupiter at equal intervals of time. However, the distance

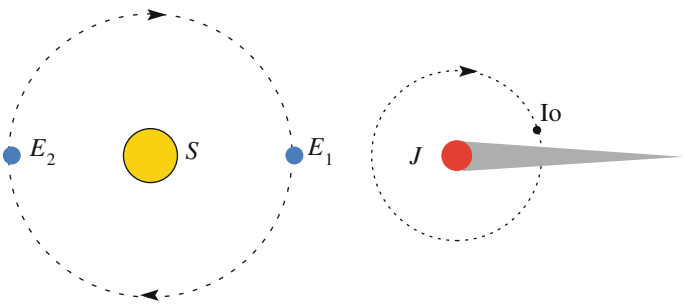


Fig. 4.5 Scheme of Cassini—Rømer argument for the speed of light

they have to cross to reach earth is growing when earth moves from position E_1 to position E_2 in the figure, namely over six months. As a consequence, the intervals between two occultations as seen from earth keep growing during this period, and diminishing during the subsequent half a year. The delays of the first period (and the anticipations of the second) accumulate up to the mentioned total of about 22 min. This is the time spent by light to cross the diameter of the earth's orbit.

The above-simplified argument requires several corrections, due to the motion of Jupiter about the sun (a revolution of about ten years), due to the earth's orbit being an ellipses, and the like. However, the conclusions are correct.

Today, we know that the diameter of the earth's orbit is, in round figures, 300 million kilometers and can calculate a value of c. However, this distance was not known with any precision in the XVII century and Rømer, perfectly aware of that, did not provide a number for the speed of light.

Hippolyte Fizeau (France 1819–1896) was the first scientist to measure the speed of light in a laboratory, in 1849. Even in this case, the light flux was periodically interrupted, as it now is automatically, employing a mechanical system built by the experimenter. Another difference with an astronomic measurement, as we already mentioned, is that, in the laboratory, light travels over a distance and then comes back. Figure 4.6 shows the scheme of the apparatus built by Fizeau. It consists of a cogwheel, W in the figure, and a mirror, M_2, placed 8 km apart. It is essential that the distances between consecutive cogs of the rim of the wheel are exactly equal to one another.

The lens L_1 produces an image of the small light source S in F on the border of the wheel in such a way that light can go through the space between two cogs. The lens L_2 makes the light beam parallel, while L_3 converges the beam on the mirror M_2. The latter reflects the light, which now comes back following the same path as did going in, again passing through the space between two cogs in F. The mirror M_1 is half-silvered and reflects part of the returning light in a direction in which the observer can see it without interfering. The lens L_4 converges the light into the eye of the experimenter.

This is what happens when the wheel is at rest. If we were now to rotate the cogwheel at an increasing speed, we would continue to see the light of the beam coming back when the angular velocity is small. Increasing the angular velocity,

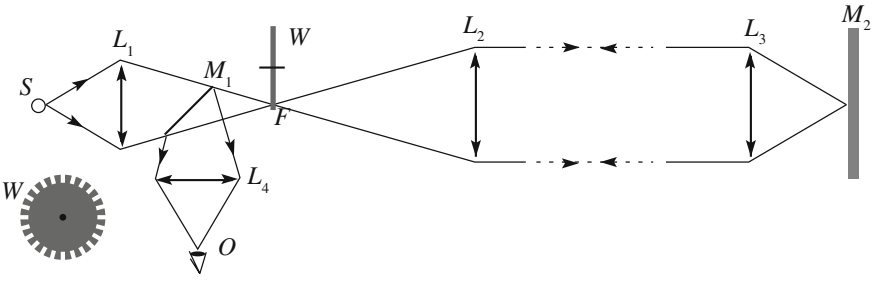


Fig. 4.6 Scheme of the Fizeau cogwheel experiment

Fizeau reached a value for which, in the time spent by light to travel back and forth, a cog had time to move to the position F where light had passed going in. Under these conditions, the reflection back from the mirror was obscured, because the light had struck one of the cogs.

If n is the number of cogs, the first obscuration happens when the number of turns per second v satisfies the condition

$$\frac{1}{2nv} = \frac{2l}{c}, \tag{4.17}$$

where l is the distance between the cogwheel and M_2. Rigorously speaking, Fizeau was measuring the light velocity in air, rather than in a vacuum. However, the difference between the two is extremely small, so small that we can neglect it here. If we now further increase the angular velocity of the cogwheel, we observe a maximum light intensity when the reflected light finds the free space after the cog it had encountered previously. The condition on v is

$$\frac{2}{2nv} = \frac{2l}{c}.$$

We encounter a second eclipse when

$$\frac{3}{2nv} = \frac{2l}{c},$$

and so on. By measuring the spin frequencies at the occultations, we obtain the speed of light c.

In this experiment, the distance was $l = 8633$ m, and the number of cogs was $n = 720$. In one of his first measurements, Fizeau found the first minimum occurring at $v = 12.6$ turns per second and the first maximum at $v = 25.2\ \text{s}^{-1}$. The flash of light had traveled 17,266 m in $1/(25.2 \times 720)$ s, namely at the speed of 3.13×10^8 m/s. Fizeau's final result is $c = 3.15 \times 10^8$ m/s (uncertainty was not quoted).

A few years later, in 1862, Léon Foucault (France, 1819–1868) realized a second method, which is now called the rotating mirror method, one that had been put forward by François Arago (France, 1786–1853) in 1838. Figure 4.7 shows a

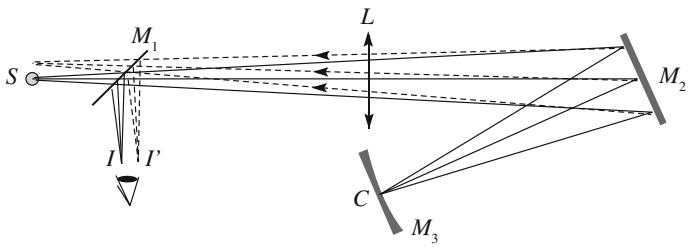


Fig. 4.7 Scheme of the Foucault rotating mirror experiment

scheme of the experiment. The light from the point source S, after having crossed the semitransparent mirror M_1, is focused by the lens L on the point C, after having been reflected by the plane rotating mirror M_2. C is on the surface of the mirror M_3, which is spherical. The reflected light travels back through the path CM_2LM_1, where it is partially reflected to the telescope of the observer. The smart idea is that the curvature center of the spherical mirror M_3 lies on the rotation axis of M_2. In this way, the light reflected by M_3 always arrives at the same point of the rotating mirror, whatever the angle between M_2 and M_3. This would not be possible if M_3 was a plane. In the figure, I is the position of the reflected image of the source when M_2 does not move and I' the image with the mirror in rotation with a certain angular velocity.

If D is the distance from C to M_2, the time taken by the light to go from the rotating mirror to M_3 and back is $\tau = 2D/c$. In this time, the mirror rotated by an angle $\alpha = \tau\omega$. Here, ω is its angular velocity, which was measured with high accuracy. The distance $I\,I'$, which is the result of the experiment, is proportional to this angle. The proportionality constant is known by construction. In 1862, Foucault measured $c = 2.98 \pm 0.005 \times 10^8$ m/s. The precision was considerably improved compared to Fizeau.

The rotating mirror method is apt to be operated on shorter base lines and, consequently, for measuring the speed of light in transparent media. Foucault was able to reduce the distance D down to 4 m, reach a rotation speed of 800 s^{-1}, and measure the speed of light in water.

When measuring the speed of light in a medium, which is always more or less dispersing, we must ask ourselves, what is the speed we are measuring? In principle, this question should even be addressed for air. What Fizeau measured is the speed of the end of the light wave. This is the same logically as what we did in the previous section with the end of the surface waves. We understand that Fizeau measured the group velocity. Similar arguments hold for the measurements of Foucault.

A difference between phase velocity v_p and group velocity v_g for light, which is extremely small in air, was established for the first time by Albert Michelson on carbon disulfide (CS2), which is a highly dispersing, colorless liquid. He found the value $c/v_g = 1.76$ for yellow light pulses, while it was known that $c/v_p = 1.64$.

Let us now see how the phase velocity of light should be measured. As given by Eq. (4.14), we must generate a monochromatic light (at least, approximately so) and measure both wavelength and frequency. For (visible) light, λ is about 0.5 μm, and ν is on the order of 10^{15} Hz. We can easily measure λ taking advantage of interference phenomena. We shall study that in the next chapter. Contrastingly, the measurement of ν is difficult, due to its very large value, but quite possible with modern techniques. Even better, these techniques lead to an extreme precision on the order of a part in 10^{11}. In practice, phase velocity in a medium is obtained indirectly by measuring the refractive index, which is inversely proportional to it, as we shall now see.

4.3 Refraction, Reflection and Dispersion of Light

Electromagnetic waves not only propagate in a vacuum but in material media as well, if the latter are sufficiently transparent. We have seen in Sect. 3.6 that Maxwell equations predict that the electric and magnetic field obey the wave equation in a vacuum [see Eqs. (3.31) and (3.32)]. These equations imply that electromagnetic waves are non-dispersive in a vacuum, a fact that is experimentally verified. Contrastingly, material media are dispersive for electromagnetic waves. We shall find, in Sect. 4.8, the partial differential equations, in place of Eq. (3.31) governing the propagation of the electromagnetic field in dielectric transparent media. We shall now focus on the physics of light. We start with discussion of the phenomena of reflection and refraction at the interface between two different media and the phenomenon of dispersion. We assume the media to be transparent, homogeneous and isotropic.

The refraction phenomenon is governed by a dimensionless quantity characteristic of the medium, which is defined for monochromatic waves and is a function of the wavelength. This is the *refractive index*, which is the ratio between the speed of light in a vacuum and the phase velocity in the medium, namely

$$n(\lambda) = c/v_p(\lambda). \tag{4.18}$$

We immediately observe that measuring the refractive index is much easier than measuring the phase velocity. Let us consider two transparent media separated by a plane surface and a plane light wave traveling in the first medium, crossing the boundary between the two and continuing in the second. Let n_1 and n_2 be the refraction indices of the first and second medium, respectively. We assume the extension of the separation surface to be very large compared to the wavelength of the waves we shall consider. Under these conditions, we can neglect the wave character of light and speak of geometric optics. We also define the trajectory of the *energy* transported by light as a *light ray*. In an isotropic and homogeneous medium for a monochromatic wave, the propagation direction of energy coincides with the direction normal to the wave front, namely of the wave vector $\mathbf{k}$. This is not the case for anisotropic media, as we shall see in Chap. 6. A *light beam* is a set of light rays with a certain cross-section.

Let us start by considering the case in which the first medium is a vacuum (or, in practice, air), which obviously has an index $n_1 = 1$, and let $n_2 = n$ be the index of the material medium. You can think of water or of a glass, for example. Let us consider a monochromatic wave and let ω be its angular frequency and λ_0 its wavelength in a vacuum. The corresponding wave number is $k_0 = 2\pi/\lambda_0$ linked to the angular frequency by the dispersion relation

$$\omega/k_0 = c. \tag{4.19}$$

In the material medium, the angular frequency of the wave is the same as in a vacuum, because, for continuity, the time dependence of the fields must be the same on the two faces of the interface surface (more discussion of this in Sect. 4.5). Let k be the wave number and $v_p(k)$ the phase velocity in the medium. The dispersion relation in the medium is

$$\omega/k = v_p(k) = c/n(k). \tag{4.20}$$

In conclusion, wavelength and wave number in the medium are different than in a vacuum. The relations are

$$k = n(k)k_0, \quad \lambda = \lambda_0/n(k). \tag{4.21}$$

To fix the ideas, the refractive index is about 3/2 for glass and about 4/3 for water. The index of the air is very close to 1, while at STP, it is $1 + 3 \times 10^{-4}$.

As is well known to the reader through common experience, lenses are built by assembling media of different refraction indices limited by plane, spherical or cylindrical surfaces. These issues will be treated in Chap. 7. Here, we discuss the basic phenomena connected to the propagation of light, namely reflection, refraction and dispersion.

Refraction. Consider a beam of parallel light rays that is incident on the interface between two transparent media, as shown in Fig. 4.8. Let n_1 and n_2 be the refraction indices of the first and second media, respectively. We assume the separation surface to be a plane and of dimensions very large compared to the wavelength. The rays in the first medium are called incident rays and the angle they form with the normal to the surface (θ_1 in the figure) is the angle of incidence. The plane of the direction of incidence and normal to the surface is called the *plane of incidence*. The rays in the second medium are called refracted rays and the angle they form with the normal (in the figure) is called the angle of refraction. Historically, it was the Greek Egyptian scientist Claudius Ptolemy (Alexandria in Egypt, 100–170) who first measured the relation between the angle of incidence and the angle of refraction. He made accurate measurements in steps of 10° for the three most important couples of media, namely between air and water, air and glass, and water and glass. In the Xth century, the Arab scientist Ibn Sahal (Bagdad, 940–1000) developed a theory regarding the use of glass lenses to focus light for burning purposes. His work, published in 984 with the title "On burning mirrors and lenses", contains a geometrical expression fully equivalent to the correct analytical expression we use today. The latter was finally found in 1621 by Willebord Snell (The Netherland-1580–1626) and is called Snell's law.

Snell's law tells us that: (a) the refracted ray lies in the plane of incidence, (b) the relation between the angle of incidence and angle of refraction is

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{4.22}$$

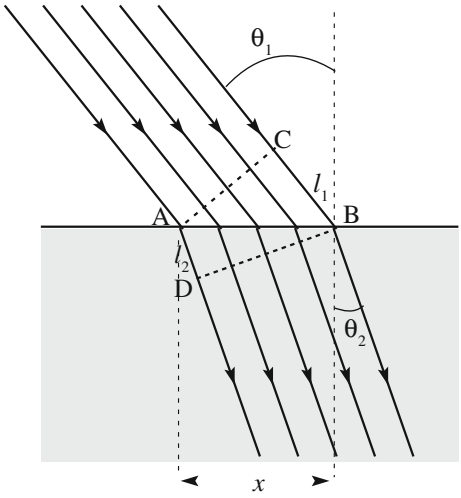


Fig. 4.8 Refraction of a light beam from a less dispersing to a more dispersing media

In Sect. 4.5, we shall give the interpretation of the law based on the wave theory. Here, we give a justification based on simple arguments. Point (a) is an obvious consequence of the symmetry of the problem. Point (b) is illustrated in Fig. 4.8 for the case $n_2 > n_1$. The figure shows, as a dotted line, the incident wave surface in the moment at which the first ray of the beam hits the interface between the media (A in the figure). The points of the wave close to A are the first to enter into the second medium, in which the speed is reduced, and the first to slow down. Other points of the incident wave surface are still in the "faster" medium. For a quantitative analysis, let us consider the triangles ABC and ABD in the figure. Both are rectangular and share the hypotenuse AB. Let x be its length. The two wave surfaces shown in the figure are the last completely in the first medium and the first completely in the second one. Their extremes are separated by distances that we call l_1 and l_2. These distances are given by $l_1 = x \sin\theta_1$ and $l_2 = x \sin\theta_2$. Now, the time taken by the phase to cross the distance l_1 in the first medium with phase velocity v_1 must be equal to the time it takes to cross l_2 in the second medium with phase velocity v_2. Namely, we must have $l_1/v_1 = l_1/v_2$. Finally, we have $n_1 l_1/c = n_2 l_2/c$, which is Snell's law in Eq. (4.22).

Figure 4.8, as already stated, is drawn for the case in which the rays go from a less refracting to a more refracting medium. In this case, as we say, the rays get closer to the normal in the second medium. In the opposite case, from greater to lesser refracting media, the rays are along exactly the same route, moving in the opposite direction. In this case, the rays depart from the normal.

For the majority of the transparent media, like water and glasses, the refractive index of light is a slowly varying, monotonically increasing, function of frequency. In particular, it is a little larger for the blue than for the red. Table 4.1 reports, as an example, the case of a crown glass (as it is called). One sees that the index for blue is about 1 % larger than that for red. This small difference is sufficient to originate the dispersion of light, for example, in a prism, as shown in Fig. 4.9.

Table 4.1 Refractive index of crown glass

	λ vacuum (nm)	ν (THz)	n
Ultraviolet	361	831	1.539
Violet	434	692	1.528
Blue green	486	618	1.523
Yellow	589	510	1.517
Red	656	457	1.514
Dark red	768	391	1.511
Infrared	1200	250	1.505

We call dispersion of light the action made by a prism of glass, and by other transparent dielectrics, on a white light beam, which becomes dispersed in the different colors it contains. Figure 4.9 shows the scheme. Entering into the glass, the components of the beam of different frequencies, which correspond to different colors to our eyes, are refracted through different angles, because their refraction indices are slightly different. At the exit, the rays are refracted again with differences between them that add up to the first ones. The total *deflection angle*, as it is called, is maximum for blue and minimum for red. The dispersion phenomenon of white light into its colors is a consequence of the dependence of the phase velocity on frequency. This is at the origin of the term dispersion relation, a term that has been generalized to all types of wave.

Two important comments are necessary here. Firstly, it is often said that colors are seven in number. This comes from a statement by Newton, who was the first to use a prism to experimentally study the separation of the colors of a sunbeam entering into a dark room from a small hole in a window. He identified and named seven different colors. This number, however, is arbitrary, being purely a matter of definition. Secondly, while it is true that we perceive a monochromatic light per se as a definite color, the opposite is not necessarily so. Indeed, the color perception is a subjective process that involves not only our eye as a detector but also a lot of processing in our retina and our brain. As a consequence, for example, a field on a surface illuminated by light of a certain frequency may be perceived as having different colors depending on the illumination of the nearby fields. We shall not deal with these interesting questions in this course.

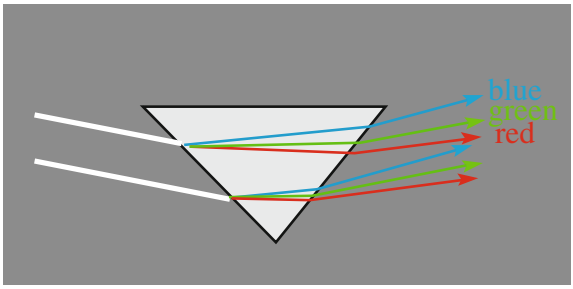


Fig. 4.9 Diffraction of a white beam by a prism

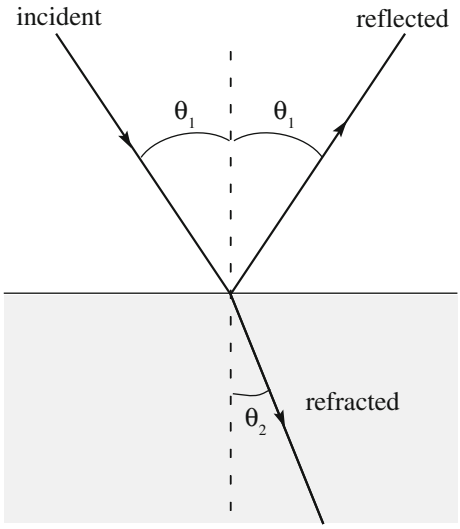


Fig. 4.10 Reflection and refraction from a less refracting to a more refracting medium

Reflection. When a ray is incident on the interface between two media of different indices, not all of its intensity goes to the refracted ray. Part of it gives origin to the reflected ray. The reflected ray propagates in the same medium as the incident ray, and in the incident plane. In addition, the angle of reflection is equal to the angle of incidence. Figure 4.10 shows the geometry in the case of increasing index from the first to the second medium (like from air to water or from air to glass).

The intensity of the incident ray is shared by the refracted and reflected rays in a proportion depending on the angle of incidence. We shall not study this in this book, but only the sharing between the two rays at normal incidence in Sect. 4.6. We shall also see, in Sect. 4.5, that the presence of the reflected radiation, in addition to the refracted one, is a necessary consequence of the wave nature of the phenomenon.

Total refraction or *total internal refraction*. The phenomenon happens in the passage from a more refracting to a less refracting medium, $n_2 < n_1$ (from water to air, for example), as shown in Fig. 4.11.

Snell's law in Eq. (4.22) tells us that, under these conditions, the angle of refraction is larger than the angle of incidence. If we now increase θ_1, θ_2 will increase as well, up to the point of reaching 90°. Under this condition, the refracted ray grazes the surface and cannot go further. In other words, there are no solutions for the refracted ray for larger values of θ_1. Indeed, the condition $\sin\theta_2 \leq 1$ tells us that the refracted ray does not exist for incident angles $\theta_1 \geq \theta_{1,lim}$, for which the limit angle is given by

$$\sin\theta_{1,\mathrm{lim}} = n_2/n_1. \tag{4.23}$$

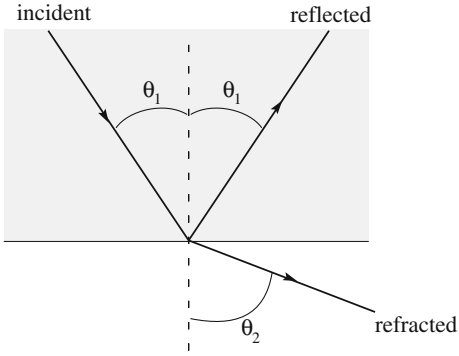


Fig. 4.11 Total internal refraction

These are the conditions of the total internal reflection. You can observe this phenomenon by swimming several meters under the quiet surface of the sea. Looking up, you see the sunlight coming in through a circle, like a large manhole, of a certain radius (depending on your depth). Outside the circle, the water is silvery, like a mirror.

4.4 Rainbow

The rainbow is a fascinating phenomenon, which has always attracted the attention of humanity, and of scientists, in particular. We can observe it when there are water drops in the air, during a rain shower, in the water spray of a waterfall or of a fountain. Figure 4.12 shows an example.

The phenomenon is explained by the dispersion of solar light by spherical water drops, as shown schematically in Fig. 4.13. Let us see how.

The colors of the rainbow are bright, with the red in its external, or higher, part, the violet on the internal, lower, one. Often a second bow can be seen, higher and much fainter than the first one. The arcs are called primary and secondary bows, respectively. The order of the color dispersion is inverted on the secondary bow, compared to the primary one, with red as the lowest, violet the highest. The sky between the two bows is noticeably darker than below the principal bow. This is called Alexander's dark band, from the Greek philosopher Alexander of Aphrodisias (Greece, 2nd century AD), who first described the effect around 200 A.D. Even when the secondary bow cannot be seen, one can easily appreciate how the sky is much more luminous on the lower side than on the upper side of the principal bow.

Humans have been enchanted by these phenomena since ancient times and that enchantment continued even as their explanation gradually became clearer. By the 10th century in Cairo, Ibn al-Haytham (Basra and Cairo, c. 965-c. 1040), better known in the West as Alhazen, was already studying the phenomenon and the dispersion of colors. In 1266 in England, Roger Bacon (UK, 1214–1292) measured

Fig. 4.12 The primary and secondary bows and the dark band. The supernumerary bows are visible under the primary arc. Photograph by Nelson Kenter with permission

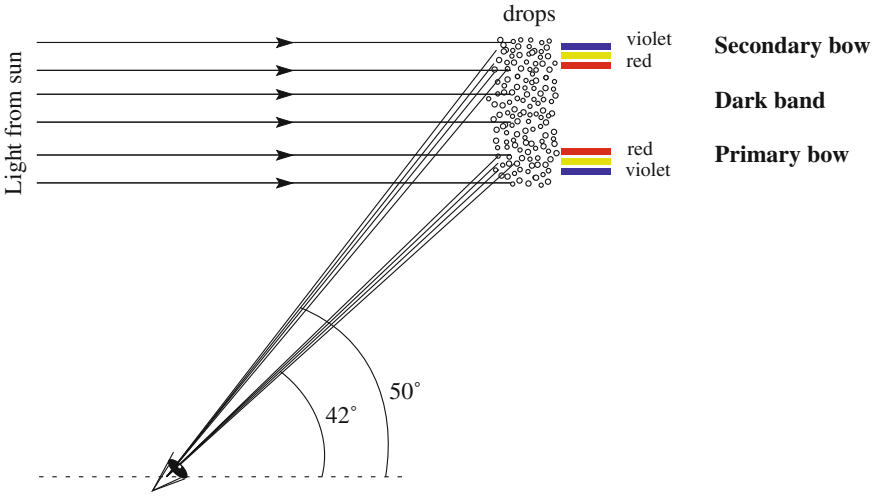


Fig. 4.13 Schematic representation of the primary and secondary bow formation

the angles between the directions of the light incident from the sun and diffused in the bows. He found 138° and 130°, respectively (see Fig. 4.13). At the beginning of the XIV century, Theodoric from Freiberg (Germany, 1250–1310) advanced the hypothesis that the bow could be caused by water drops. To check the idea, he conducted experiments on a spherical glass bowl full of very pure water, serving as a model of the drop. He sent a light ray through the bowl at different distances from the center and studied the trajectory of the ray in the water for each of them. He

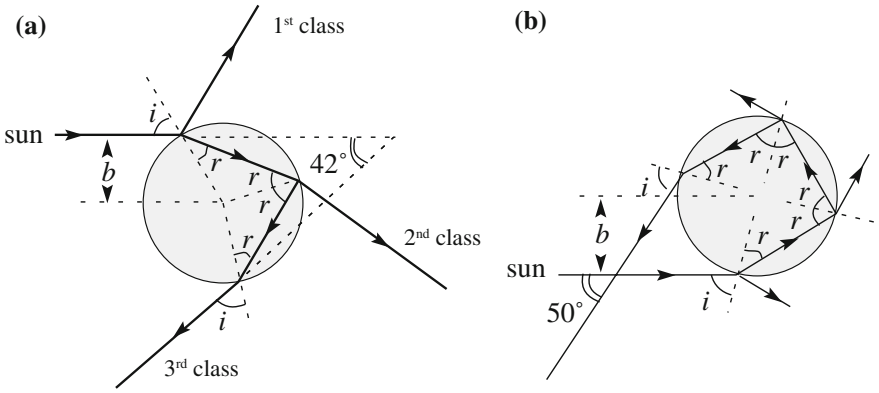


Fig. 4.14 The trajectories of the rays in the drop for **a** the primary bow, **b** the secondary bow

found that the principal bow is due to the rays that enter the drop, are internally reflected once and then exit, while the secondary bow is due to rays that internally reflect twice and then exit (Fig. 4.14). Unknown to him, Ibn al-Haytham had already performed similar experiments four centuries earlier. Three centuries after Theodoric, in 1637, René Descartes (France, 1596–1650) gave a clear interpretation of the phenomenon based on Snell's law (1621).

Let us analyze what happens when a light ray, which we initially consider monochromatic, reaches the surface of a spherical water drop.

We start by noticing that, given the cylindrical symmetry of the problem, the only geometrical parameter is the distance b between the line defined by the incoming ray and the center of the drop. This distance is called the *impact parameter*.

The ray hitting the drop at a given impact parameter is partially reflected, giving origin to what we call a 1st class ray, in part refracted, following Snell's law. When the refracted ray reaches the surface, it, in turn, is partially reflected (inside the drop), partially refracted, leaving the drop as a 2nd class ray. Again, the ray remaining in the drop is partially reflected and partially refracted (3rd class ray). The process continues. The 3rd class rays produce the primary bow, the 4th class rays the secondary bow. Rays of higher classes exist, but their intensities are lower and lower with increasing class order. They produce very faint higher order bows, which can be observed with sensitive photographic techniques.

In conclusion, every incident ray at a certain impact parameter produces a set of outgoing rays, called scattered rays, of different classes. For each class, the scattering angle, which is the angle between the outgoing and incoming rays, varies with the impact parameter. Considering that the sun casts light on the drop at all the impact parameters uniformly, one might think that light should be scattered practically at all angles. What, then, is the reason that we see the bow around a definite angle?

To answer the question, we study how the 3rd class ray scattering angle varies as a function of the impact parameter. When this is zero, the incident angle is also zero, and the 3rd class ray is scattered at 180° after having crossed the drop twice on its diameter. If the impact parameter increases, the scattering angle decreases, but not forever. Indeed, it reaches a minimum and then increases back. In order to see that, let us consider the scattering angle as a function of the incident angle, which we call i. This is a function of b given by $\sin i = b/R$, where R is the drop radius. Looking at Fig. 4.14, we see that the scattering angle is the sum of the deviation at the first refraction ($i - r$), of the deviation at the following reflection ($\pi - 2r$) and of the deviation at the second refraction ($i - r$). Hence, it is

$$\Delta = i - r + \pi - 2r + i - r = \pi + 2i - 4r$$

We can eliminate r using Snell's law. Calling n the index of water (its value is about $n = 1.333$) and taking 1 as the index of air, we find

$$r = \arcsin\left(\frac{\sin i}{n}\right).$$

We then obtain

$$\Delta = \pi + 2i - 4\arcsin\frac{\sin i}{n} = \pi + 2i - 4\arcsin\frac{b}{nR}.$$

For small incident angles (i.e., small impact parameters), Δ is a decreasing function of i (and of b), as one can easily check approximating the sine with its argument, namely as

$$\Delta \simeq \pi + 2i - 4\frac{i}{n} = \pi + 2\left(1 - \frac{2}{n}\right)i = \pi + 2\left(1 - \frac{2}{n}\right)\frac{b}{R}$$

and noticing that $2(1–2/n) < 0$. However, the scattering angle Δ only initially decreases with increasing i. It subsequently reaches a minimum and then increases. It is simply the existence of the minimum, or better yet, of an extreme, that is at the origin of the rainbow, as we shall now see. We find the minimum by putting the derivative as equal to zero and solving the equation

$$\frac{d\Delta}{di} = 2 - \frac{4}{\sqrt{1 - \frac{\sin^2 i}{n^2}}}\frac{\cos i}{n} = 0.$$

We get $4\cos^2 i = n^2 - \sin^2 i$, and hence, $\sin^2 i = (4 - n^2)/3$, which gives us $i = 59.4°$.

The angle of refraction is given by $\sin r = \sin i / n = 0.65$, that is $r = 40.2°$.

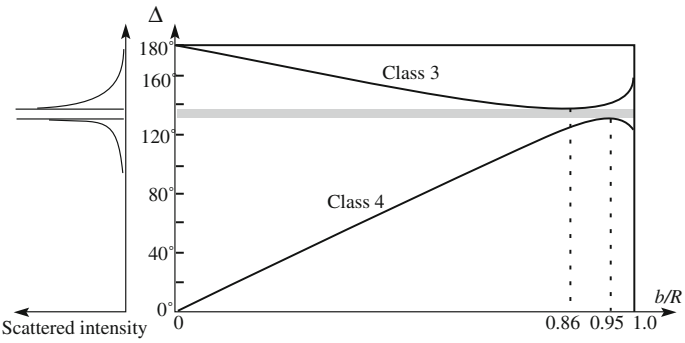


Fig. 4.15 Scattering angle as a function of the impact parameter to radius ratio for 3rd class and 4th class rays and scattered light intensity versus scattering angle

The minimum deflection angle is thus $\Delta_{\min} = 180° + 118.8° - 160.9° = 137.9°$. It is reached at the impact parameter value of about 7/8 (0.86) of the radius.

Figure 4.15 shows the behavior of the scattering angle (also called the deflection angle) as a function of the impact parameter.

The primary rainbow is observed at the minimum angle. To understand why, consider that the sunlight illuminates the drop uniformly at all the impact parameters. The light incident at a certain impact parameter is scattered at a certain angle, as we computed. At angles close to the minimum, where the curve varies very slowly, several impact parameter values contribute. Consequently, there is a sharp intensity maximum (see Fig. 4.15 on the left). In reality, the light is not monochromatic, as we thought up until now. We must consider that the refractive index is a (slowly) increasing function of the frequency. Consequently, the various monochromatic components of the white light are deflected more and more for larger and larger frequencies (the blue is deflected more than the red) and the intensity peaks appear in slightly different directions for the different colors, with blue being lower than red.

The reason for the bow shape is as follows. The rays of certain color are deflected at a certain angle, i.e., from a certain angle with the incident direction. Consequently, they lie on a cone, the axis of which is the incident direction. We observe a section of that cone, and hence, see an arc in the sky.

We have thus found that the dispersion of the light in its monochromatic components is at the origin of the rainbow. The study of the dispersion is originally credited to Isaac Newton (UK, 1643–1727), who published his "Opticks" treatise in 1704. Starting from his measurements of the refractive index at different wavelengths, Newton calculated the rainbow angle to be 137°58′ for the red and 139°43′ for the violet, in agreement with the measured values.

The explanation for the secondary bow is similar, with the difference being due to the 4th class rays. From Fig. 4.14b, we have

$$\Delta = i - r + \pi - 2r + \pi - 2r + i - r = 2\pi + 2i - 6r.$$

At the zero impact parameter (i.e., for $i = 0$), the ray is scattered at 0° after having crossed the drop three times on its diameter. When b increases, the deflection Δ increases, reaches a maximum and then decreases. It is easily found that the maximum deviation is reached at $i = 72°$, corresponding to $r = 45.5°$. The maximum deflection angle is consequently $\Delta_{max} = 360° + 144° - 270°$, or, modulo 360°, $\Delta_{max} = 130°$.

Question Q 4.2. Compare the rainbow angle (of the primary bow) for fresh water ($n = 1.333$) and seawater ($n = 1.340$). (N. B. indicative values; they depend on temperature) □

Question Q 4.3. Calculate the maximum deflection angle for 5th class rays. Is there an extreme for the 5th class rays? Is it a maximum or a minimum? Calculate it. □

Notice in Fig. 4.15b that an observer, being located lower than the sun relative to the drop, sees the scattered light hitting the drop below its axis. As a result, the blue appears higher than the red, as can be observed in that figure.

We also notice that no light is scattered between 130° and 138° in rays of 3rd and 4th classes. In addition, as we mentioned, the higher class rays are extremely faint. This explains Alexander's dark band.

Notice finally that the scattering angles at a given impact parameter are, for each class, independent of the radius of the drop. Water drops of different diameters, such as those within rain, contribute to the rainbow in the same manner. Also, the geometry of the scattering is the same for the small raindrops, and the water spheres of Ibn al-Haytham, Theodoric and Descartes.

We have discussed the showiest characteristics of the rainbow here. However, the physics of the phenomenon is much richer than that. We shall now give some hints.

The light of the rainbow is polarized "by scattering" , a phenomenon we shall discuss in Sect. 6.5. It can be easily observed by looking at the rainbow through polarizing sunglasses and turning them by 90° (see Sect. 6.4).

One can sometimes observe a few colored bands below the principal arc, in general rose and green, called supernumerary bows. They appear when the diameters of the raindrops are fairly equal to one another and are clearly visible in Fig. 4.12. The supernumerary bows are due to an interference phenomenon (we shall study the interference in Chap. 5).

In the geometric optic approximation we have taken, Alexander's dark band should be completely dark. But it is not so due to the diffraction phenomenon (see Chap. 5).

The left part of Fig. 4.15 can be interpreted as a diagram showing the probability of light being scattered as a function of the scattering angle (separately for the 3rd and 4th classes). Similarly, scattering experiments are powerful tools for the study of quantum systems like atoms, nuclei and elementary particles.

4.5 Wave Interpretation of Reflection and Refraction

We shall now show how the wave nature of light explains reflection and refraction, as first done by Christiaan Huygens (The Netherland, 1629–1695) in 1690. We shall see how, in addition, the explanation foresees new phenomena that are not explained by the geometrical optics. Let us consider a plane monochromatic wave incident on the plane surface separating two homogeneous and isotropic transparent media. We choose a reference system with the origin and the y-axis on both the interface and the plane of incidence, the z-axis normal to the interface in the direction of the semi-space of the incident wave and the x-axis on the interface to complete the frame. Let $\mathbf{k}_i$ be the wave vector of the incident wave and ω_i its angular frequency. With our choice of axes, $\mathbf{k}_i$ belongs to the y,z plane.

The electric incident field is

$$\mathbf{E}_i(\mathbf{r},t) = \mathbf{E}_{0i}\mathrm{e}^{i(\omega_i t - \mathbf{k}_i\cdot\mathbf{r})}. \tag{4.24}$$

We use the footers r and t for the reflected and transmitted (namely refracted) waves. Their fields are then $\mathbf{E}_r$ and $\mathbf{E}_t$ and are given by

$$\mathbf{E}_r(\mathbf{r},t) = \mathbf{E}_{0r}\mathrm{e}^{i(\omega_r t - \mathbf{k}_r\cdot\mathbf{r})}. \tag{4.25}$$

and

$$\mathbf{E}_t(\mathbf{r},t) = \mathbf{E}_{0t}\mathrm{e}^{i(\omega_t t - \mathbf{k}_t\cdot\mathbf{r})}. \tag{4.26}$$

Notice that we have represented the angular frequencies and the wave vectors in the two media with different symbols, to avoid making any a priori hypothesis on the relationships between them. We want to find these relations. We know that the tangential component of the total electric field parallel to the interface surface is continuous. We are sure of this statement despite not knowing whether the surface is charged or not. If it is charged, the discontinuity is in the normal component, not in the tangential ones. We then impose the tangential field components (which we indicate with the superscript $\|$) on the two sides of the interface (the plane $z = 0$) so as to be equal to one another, namely

$$\mathbf{E}_{0i}^{\|}\mathrm{e}^{i\left(\omega_i t - k_{ix}x - k_{iy}y\right)} + \mathbf{E}_{0r}^{\|}\mathrm{e}^{i\left(\omega_r t - k_{rx}x - k_{ry}y\right)} = \mathbf{E}_{0t}^{\|}\mathrm{e}^{i\left(\omega_t t - k_{tx}x - k_{ty}y\right)}, \tag{4.27}$$

which must be identically valid. In particular, at $\mathbf{r} = 0$, we have

$$\mathbf{E}_{0i}^{\|}\mathrm{e}^{i\omega_i t} + \mathbf{E}_{0r}^{\|}\mathrm{e}^{i\omega_r t} = \mathbf{E}_{0t}^{\|}\mathrm{e}^{i\omega_t t},$$

which must hold for any t. Consequently, as we had anticipated in the previous section, the three frequencies are equal, and we can represent them with a single symbol, namely with ω (Fig. 4.16).

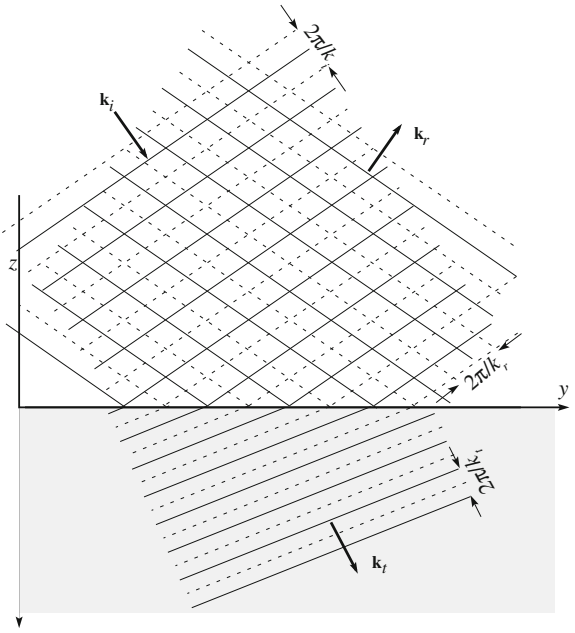


Fig. 4.16 Incident, refracted and reflected waves at the interface between two media

Similarly, at $t = 0$ and $z = 0$, taking into account that $k_{ix} = 0$, the continuity conditions gives us

$$\mathbf{E}_{0i}^{\parallel} e^{i\left(-k_{iy}y\right)} + \mathbf{E}_{0r}^{\parallel} e^{i\left(-k_{rx}x - k_{ry}y\right)} = \mathbf{E}_{0t}^{\parallel} e^{i\left(-k_{tx}x - k_{ty}y\right)},$$

which holds for every x and y. This can happen only if all the exponents are identically equal, namely if

$$k_{iy}y = k_{rx}x + k_{ry}y = k_{tx}x + k_{ty}y. \tag{4.28}$$

In particular, for $y = 0$ and arbitrary (but not zero) x, the conditions become

$$0 = k_{rx}x = k_{tx}x.$$

Hence, in conclusion, it must be

$$k_{rx} = k_{tx} = 0.$$

This tells us that both the reflected and the refracted rays belong to the incident plane.

The continuity equation in Eq. (4.28), now for $x = 0$ and arbitrary (but different from zero) y, becomes

$$k_{iy} = k_{ry} = k_{ty}.$$

We can read these equations by saying that the distance between two consecutive crests on the interface must be the same for all the waves. Let θ_i be the angle of incidence (namely the acute angle between $\mathbf{k}_i$ and the z-axis) and θ_r the angle of reflection (namely the acute angle between $\mathbf{k}_r$ and the z-axis) and θ_t the angle of refraction (namely the acute angle between $\mathbf{k}_t$ and the z-axis). Then, the last relation can be written as

$$|\mathbf{k}_i| \sin\theta_i = |\mathbf{k}_r| \sin\theta_r = |\mathbf{k}_t| \sin\theta_t. \tag{4.29}$$

Being that the incident and reflected waves are in the same medium, the magnitudes of their wave vectors are equal. Then, the first equation tells us that $\theta_r = \theta_i$. This is the law of reflection: the angle of reflection is equal to the angle of incidence.

In the second equation, we take into account the dispersion relation in the two media, which are

$$|\mathbf{k}_i| = \frac{n_1}{c}\omega, \quad |\mathbf{k}_t| = \frac{n_2}{c}\omega \tag{4.30}$$

and immediately obtain

$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

This is Snell's law. We have now shown how all the reflection and refraction laws stem from the wave nature of light. We have learned that they are consequences of the different wave velocities in the two media and of the continuity of the wave field through the interface. Note that the last condition can be satisfied only if all three waves are present. The same conclusions hold for every type of wave.

In the previous section, we saw that, in the passage from a more refrangent to a less refrangent medium, under the condition of total reflection, namely for incidence angles larger than the limit angle, there is no refracted ray. How is it possible to insure the continuity of the wave field if, in the second medium, there is no wave? The answer is that a wave is indeed present in the second medium, near to the interface, even if it does not correspond to any ray of the geometric optics. This is called the vanishing wave. Let us see how it works.

We explicitly write the dispersion relations from Eq. (4.30) as

$$k_i^2 = k_{iz}^2 + k_{iy}^2 = \frac{n_1^2}{c^2}\omega^2, \quad k_t^2 = k_{tz}^2 + k_{ty}^2 = \frac{n_2^2}{c^2}\omega^2. \tag{4.31}$$

Let us find k_{ty} from Eqs. (4.29) and (4.30). We obtain

$$k_{ty} = k_{iy} = |\mathbf{k}_i| \sin\theta_1 = \frac{n_1}{c}\omega \sin\theta_1. \tag{4.32}$$

Let us substitute it in Eq. (4.30) and find

$$k_{tz}^2 = \frac{\omega^2}{c^2}\left(n_2^2 - n_1^2 \sin^2\theta_1\right). \tag{4.33}$$

We see that for angles of incidence larger than the limit angle, it is $k_{tz}^2 < 0$, namely the component of the wave vector of the refracted wave normal to the interface is imaginary. What does this mean? Introducing the real quantity $K_{tz} = ik_{tz}$, we can write the refracted wave as

$$\mathbf{E}_t(\mathbf{r}, t) = \mathbf{E}_0 \mathrm{e}^{-K_{tz}z} \mathrm{e}^{i\left(\omega t - k_{ty}y\right)}. \tag{4.34}$$

This is a progressive wave moving in the y direction along the interface. Its amplitude, $\mathbf{E}_0 \mathrm{e}^{-K_{tz}z}$, is a function of z, namely of the distance from the surface inside the second medium. Its amplitude reduces by a factor of $1/e$ on the distance $1/K_{tz}$, which is on the order of the wavelength. This is the *vanishing wave*, which, we summarize, propagates along the interface on the side of the "forbidden" medium with amplitude that exponentially decreases with increasing distance from the surface. It becomes negligible, "vanishes", at a few wavelengths, namely at a couple of micrometers for light. The presence of the vanishing wave is mandatory, because no wave can disappear sharply and discontinuously. Figure 4.17a schematically shows the ridges (continuous lines) and gorges (dotted lines) of the incident and reflected waves, as well as those of the evanescent wave. In Fig. 4.17b, the reflected wave is not represented to avoid confusion.

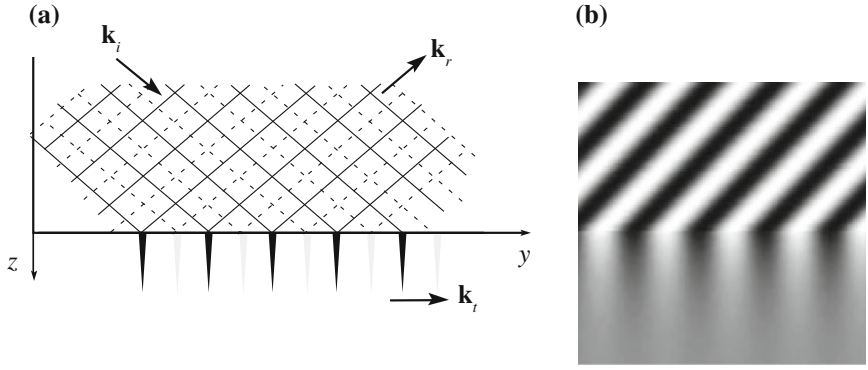


Fig. 4.17 **a** The fronts of the incident, evanescent and reflected waves at the interface between two media under internal total reflection conditions, **b** intensities of the same, omitting reflected wave to avoid confusion

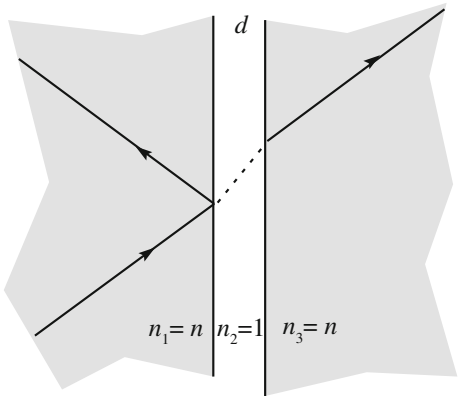


Fig. 4.18 The tunnel effect

The opposite case, namely of $K_{tz} < 0$, while mathematically possible, is not so physically. Indeed, it corresponds to a wave of exponentially increasing amplitude, something that would imply an infinite energy.

Let us check the presence of the vanishing wave as sketched in Fig. 4.18. We use two thin and perfectly flat glass sheets and lay them parallel, with an air gap between them of thickness d on the order of the wavelength. This is not easy to achieve, but suppose we succeeded. We now send a light beam along the first glass sheet at an angle greater than the limit angle for total reflection (at the exit to the gap). Being that d is small enough, the amplitude of the evanescent wave in the gap has not yet vanished completely at the surface of the second glass. Here, the phenomenon opposite to that at the first interface happens. The vanishing wave acts as an incident wave, even if with a reduced amplitude, and produces a refracted wave (or ray) in the second glass. The wave vector of this wave has no imaginary component. We have a "normal" ray. A fraction of the energy coming in from the first glass has been able to cross the forbidden gap. The phenomenon can happen only if the gap width is on the order of the wavelength. This is called, in optics, a frustrated vanishing wave. However, we more often use the term that comes from quantum mechanics, namely the tunnel effect. Indeed, in quantum mechanics, a particle of a given energy can pass to a "valley" on the other side of an energy "hill", namely a region forbidden by energy conservation. The crossing of the forbidden region is exactly the same phenomenon we just considered, as if a tunnel was present under the hill,

An easy way to observe the effect qualitatively is as follows. Fill a glass having a smooth surface with cold water from the refrigerator, wait for the air humidity to condensate on the external surface, keep the glass between your fingers and observe your fingertips looking through the water at an angle larger than the limit angle. The pattern of your fingerprints appears brilliantly silvery. The silvery lines are where the distance of the finger from the glass is larger than a few wavelengths, the pink ones where the vanishing wave is frustrated.

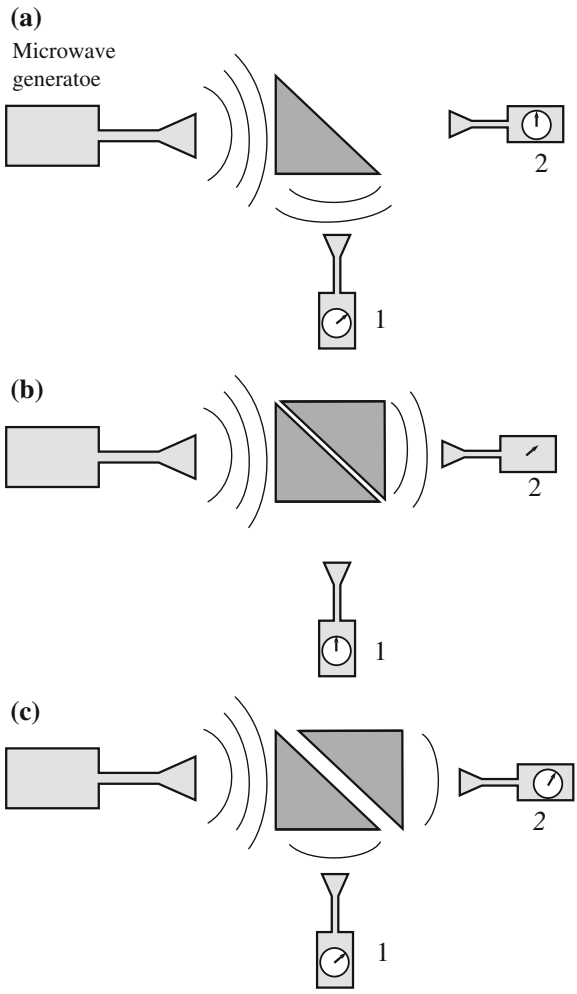


Fig. 4.19 Apparatus to demonstrate the tunnel effect

Experiments are easier if we use a wavelength on the order of a centimeter ($\lambda = 1$ cm corresponds to $\nu = 30$ GHz in a vacuum). Figure 4.19 shows a demonstration apparatus.

We produce two rectangular isosceles prisms of paraffin. This substance, which is opaque at optical frequencies, is quite transparent in the GHz, where its index is about $n = 1.5$. The corresponding limit angle is $\theta_{\text{lim}} = 41°$. Figure 4.19 shows a source of microwaves and two detectors. We shall send the wave beam through one of the prisms and study its internal reflection at 45° where it is total. When the apparatus is as in Fig. 4.19a, detector 1 detects a strong radiation flux, while detector 2 does not detect anything. Indeed, the internal reflection is total. In Fig. 4.19b, we put the second prism in contact with the first along the common hypotenuse. The two blocks are seen by the wave as a single one, any possible gap

between the hypotenuses being much smaller than the wavelength. Detector 2 gives a strong signal, detector 1 does not give any. If we now slowly separate the blocks, gradually opening a gap between them, the signal of detector 1 gradually increases, while that of detector 2 continues to decrease.

4.6 Reflected and Transmitted Amplitudes

As we have seen, at the interface between two media of different indices, the energy of the incident ray is shared between the reflected and the refracted rays. The relative proportions depend on the two indices and the incident angle. We shall only consider the simplest case of normal incidence here (namely $\theta_1 = 0$) in which geometric complications are avoided, while keeping the physics insight. Figure 4.20 shows the situation. The interface is a plane surface, *i, r* and *t* are, respectively, incident, reflected and refracted rays. Let n_1 and n_2 be the refraction indices of the first and second medium, respectively, and let $n_1 < n_2$. We now request the continuity of the relevant components of both the electric and magnetic fields through the surface.

We choose a reference frame with the z-axis in the direction of the incident ray, namely normal to the interface surface, the x in the positive direction of the electric field of the incident wave, and the y-axis in the positive direction of its magnetic field. Let $\mathbf{E}_{0i}$ and $\mathbf{B}_{0i}$ be the amplitudes of the fields of the incident wave. Note that the wave vector $\mathbf{k}_i$, which has the direction and sense of $\mathbf{E}_{0i} \times \mathbf{B}_{0i}$, has the direction and sense of the z-axis as well.

The wave vector $\mathbf{k}_r$ of the reflected wave is directed opposite to $\mathbf{k}_i$, namely as $-z$. Consequently, if $\mathbf{E}_{0r}$ and $\mathbf{B}_{0r}$ are the amplitudes of the fields of the reflected wave, their cross-product $\mathbf{E}_{0r} \times \mathbf{B}_{0r}$ must be directed as $-z$ as well. Hence, one of the two fields must have changed direction relative to the incident waves, the other one not. Let us assume that it is the electric field that changed signs. We then have the situation shown in Fig. 4.21, with $\mathbf{B}_{0r}$ in the y direction, $\mathbf{E}_{0r}$ in the direction opposite to the x axis, and $\mathbf{k}_r$ in the direction opposite to z. Note that the hypothesis

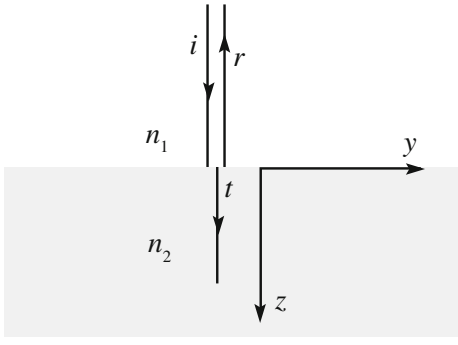


Fig. 4.20 Incident, reflected and refracted rays at normal incidence

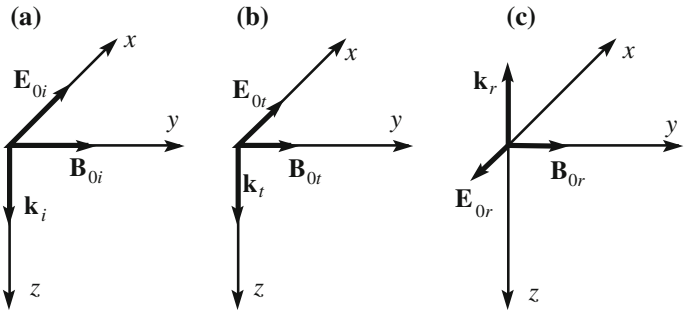


Fig. 4.21 The electric and magnetic fields of the **a** incident wave, **b** refracted wave, **c** reflected wave

we made does not change the substance of the argument. If it were the magnetic field that inverted, we would have the x component of $\mathbf{E}_{0r}$ as positive and the y component of $\mathbf{B}_{0r}$ as negative. In any case, the inversion of amplitude is equivalent to an abrupt change by π of the phase at reflection. We have already encountered this situation.

Let now consider the refracted wave. The direction and sense of the refracted wave, namely those of $\mathbf{E}_{0r} \times \mathbf{B}_{0r}$, are the same as those of the incident wave, as shown in Fig. 4.21b. We impose the continuity of the components parallel to the surface of both fields, by writing

$$E_{0i} - E_{0r} = E_{0t}, \quad B_{0i} + B_{0r} = B_{0t}, \tag{4.35}$$

where we have taken $E_{0r} > 0$ if directed opposite to the x-axis, as shown in Fig. 4.21b. We need to express the second of these equations in terms of the electric field. We already know that, in a vacuum, the ratio of the amplitudes of the electric and magnetic fields is c. We anticipate that, in Sect. 4.8, we shall mention that a similar relation exists in a normal dielectric limited to monochromatic waves, which is the case we are considering. Well, under these conditions, the ratio of the amplitudes of the electric and magnetic fields is the phase velocity in the medium, namely c/n. We can then write

$$B_{0i} = \frac{n_1}{c} E_{0i}, \quad B_{0r} = \frac{n_1}{c} E_{0r}, \quad B_{0t} = \frac{n_2}{c} E_{0t},$$

The Eq. (4.35) relative to the magnetic field becomes $n_1(E_{0i} + E_{0r}) = n_2 E_{0t}$. Together with Eq. (4.35) relative to the electric field, we have now a system of two equations from which we can obtain E_{0r} and E_{0t} as functions of E_{0i}. For convenience, let us introduce the relative refractive index between the media, which is

$$n = n_2/n_1. \tag{4.36}$$

and let us write our system as

$$E_{0i} - E_{0r} = E_{0t}, \quad E_{0i} + E_{0r} = nE_{0t}. \tag{4.37}$$

We solve the system, obtaining

$$E_{0r} = \frac{n-1}{n+1}E_{0i} = \frac{n_2 - n_1}{n_2 + n_1}E_{0i}, \quad E_{0t} = \frac{2}{n+1}E_{0i} = \frac{2n_1}{n_2 + n_1}E_{0i}. \tag{4.38}$$

We now recall that we are considering the case of $n_2 > n_1$, namely $n > 1$. From Eq. (4.38), we see that if the incident field is in the positive direction of x, namely if E_{0i} is positive, E_{0r} is positive as well. Namely, under the convention we have taken, the electric field of the reflected wave has the opposite direction, consistent with our initial assumption. The equations in Eq. (4.38) are valid in the case of $n_2 < n_1$ too. They tell us, in this case, that if E_{0i} is positive, E_{0r} is negative, namely directed in the same sense.

In conclusion, when a monochromatic plane wave is normally incident on the plane interface between two media of different indices, the electric field of the reflected wave is in phase opposition with the field of the incident wave if the index of the second medium is larger than that of the first (i.e., if the phase velocity in the second medium is smaller). In the opposite case, the two fields are in phase. The situation is similar to that of the elastic string that we studied in Sect. 3.3. These conclusions on the relative phases that we have found for normal incidence are, in fact, valid for any incidence angle.

We now define the amplitude reflection coefficient r and the amplitude refraction coefficient t as the ratios between the reflected and refracted amplitudes, respectively, and the incident amplitude. From what we have just found, we can state that the coefficients are given by

$$r = \frac{E_{0r}}{E_{0i}} = \frac{n_2 - n_1}{n_2 + n_1}, \quad t = \frac{E_{0t}}{E_{0i}} = \frac{2n_1}{n_2 + n_1}. \tag{4.39}$$

Often, we are interested in the ratios of the intensities, rather than in the amplitudes. These are called reflection coefficient R and refraction coefficient T. We recall that the intensity of a monochromatic wave, meaning its average intensity, is the average over a period of the energy crossing the unit surface normal to the propagation direction in a second. This is given by the average over a period of the Poynting vector $\mathbf{S} = (1/\mu_0)\mathbf{E} \times \mathbf{B}$. On the other hand, in each of the waves we are considering, the electric and magnetic fields are perpendicular to one another, and both are perpendicular to the wave vector. Their magnitudes are proportional to one another as $B = nE/c$. We can thus establish the following proportionality relations for the intensities: $I_i \propto (n_1/c)E_{0i}^2$, $I_r \propto (n_1/c)E_{0r}^2$ and $I_t \propto (n_t/c)E_{0t}^2$. The reflection and transmission coefficients are then

$$R = \frac{I_r}{I_i} = r^2 \quad , \quad T = \frac{I_t}{I_i} = \frac{n_2}{n_1} t^2 .$$

Finally, we have

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1} \right)^2 , \quad T = \frac{4 n_1 n_2}{(n_2 + n_1)^2} . \tag{4.40}$$

One verifies immediately that $R + T = 1$, as it should be for energy conservation.

Let us finally look into a couple of examples. For the reflection from air ($n_1 \sim 1$) to glass ($n_2 = 1.5$), the reflection coefficient given by Eq. (4.40) is $R = 0.04$ and the same is true for the inverse passage. In the passage from air to water ($n_2 = 1.5$) and vice versa, we have $R = 0.02$.

4.7 Origin of the Refractive Index

We have seen that the phase velocity of light, or the index, in a material medium is different than that in a vacuum, and that it is frequency dependent. We shall try to understand here the physical origin of the refractive index. The effect is due to the electric charges within the molecules and to the polarization induced on these charge distributions by the electric field of the incoming electromagnetic wave. This is the dynamic analogous to the electrostatic polarization of a dielectric. Under the action of an electric field, the charge distribution constituting a molecule stretches, giving origin to an electric dipole. As we studied in the 3rd volume, the electric field "acting" on a given molecule is the sum of the external, applied, field and the fields due to all the sister molecules resulting from their polarization. The resulting problem is quite complicated. In order to avoid unnecessary complexity and to clarify the physical process, we shall limit the discussion to a sparse medium. We shall assume the distances between molecules to be large enough so that their interactions are negligible. This is the case for gases under normal conditions, in a good approximation. Under this assumption, the electric field acting on each molecule is the field of the incoming wave alone. We shall find the dependence of the refractive index on the frequency, namely we shall formulate a theory of dispersion. The theory is valid only under the assumptions we have made, but is general enough to be able to discuss the main physical characteristics of the dispersion phenomena. This approach was first proposed by Richard Feynman, in his beautiful "Lectures on physics", which we shall follow in our discussion.

Let us consider our homogeneous, low-density medium in the form of a plate having faces of infinite area and a small thickness h between them. On the two sides of the plate, there is a vacuum. We choose a reference frame with the origin and the x and y axes on the first face of the plate and the z-axis perpendicular. We consider a

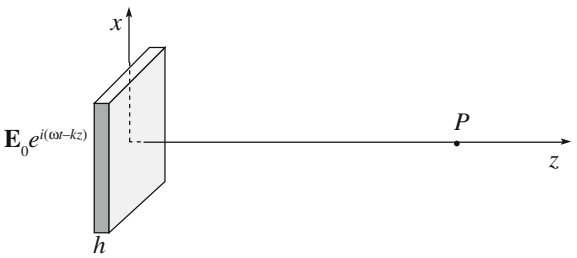


Fig. 4.22 A plane monochromatic electromagnetic wave normally incident on a dielectric plate

progressive, monochromatic electromagnetic plane wave traveling in the positive z direction. In addition, let it be linearly polarized with the electric field in the x-direction, as shown in Fig. 4.22.

From a macroscopic point of view, we know that the wave inside the plate propagates with a phase velocity different than that in a vacuum. From a microscopic point of view, the plate is composed of a large number of charges, inside the molecules. Between the molecules, we have a vacuum. Consequently, the *incident* wave also continues with its velocity c in the medium. Why, then, is the phase velocity different? The reason is obviously to be found in the presence of the charges. The electric field of the incident wave transforms each of the molecules in an electric dipole, whose charges oscillate in time, namely accelerate. Each accelerating charge emits a small spherical electromagnetic wave. The important point is that the phase of each emitted wavelet has a definite relation with the phase of the incoming wave. In other words, all the secondary wavelets are locked in phase with that of the incident. The sum of these wavelets and of the incident wave is what we call the refracted wave, and the reflected wave as well.

We now consider a point P on the z-axis beyond the plate and compare the electric fields there in the presence and in the absence of the plate. Let them be $\mathbf{E}_1$ and $\mathbf{E}_2$, respectively. The incident field on the plate is

$$\mathbf{E}_1(z,t) = \mathbf{E}_0 \mathrm{e}^{-i(\omega t - kz)}, \tag{4.41}$$

which, on the first face of the plate, which is at $z = 0$, is simply

$$\mathbf{E}_1(0,t) = \mathbf{E}_0 \mathrm{e}^{-i\omega t}. \tag{4.42}$$

In the plate, the phase of the wave advances at the speed of c/n, rather than c. Hence, the difference between the time taken by the phase to cross the plate and the time that it would have taken to cross the same distance in a vacuum is $\Delta t = hn/c - h/c = (n-1)h/c$. Hence, compared to the situation in a vacuum, the phase at the second face of the plate lags behind by $\Delta\phi = \omega(n-1)h/c$.

What we have to explain, then, is that the sum of the field of the incident wave and of the fields of the oscillating molecular charges is equal to $\mathbf{E}_1$ delayed in phase by $\Delta\phi$. Namely, we should we should show that the field

$$\mathbf{E}_2 = \mathbf{E}_1 \mathrm{e}^{-i\omega(n-1)h/c} \tag{4.43}$$

can be expressed as the sum

$$\mathbf{E}_2 = \mathbf{E}_1 + \mathbf{E}_c, \tag{4.44}$$

Now, if $h(n–1)$ is small enough, we can approximate Eq. (4.43) as

$$\mathbf{E}_2 = \mathbf{E}_1[1 - i\omega(n-1)h/c] = \mathbf{E}_1 - i\omega(n-1)\frac{h}{c}\mathbf{E}_1.$$

Our thesis is now to show that the field resulting from the charges in the plate is

$$\mathbf{E}_c = -i\omega(n-1)\frac{h}{c}\mathbf{E}_1. \tag{4.45}$$

Under the hypothesis of low density that we have made, the electric field acting on the charges is given by Eq. (4.42). Note that our hypothesis is a realistic one for gases at STP, as proven by considering that, for them, the refractive index is $n \approx 1$ and, in addition, the reflected wave has negligible amplitude. Contrastingly, it is not a good approximation for a condensed medium.

We now look at the effects of the field of Eq. (4.42) on the molecules. We imagine each molecule to be made of a central positive charge in which practically the entire mass is concentrated, surrounded by a cloud of negative charge, made of electrons. Let q_e and m_e be the electron charge and mass. The motion of an electron under the force $\mathbf{F}(t) = q_e\mathbf{E}_1(0,t)$ is correctly described by quantum mechanics. However, we would not be very wrong to consider a model in which the electrons behave as classical matter points of mass m_e subject to an elastic restoring force. Let x then be the displacement of a generic electron from its equilibrium position. The equation of motion of an electron in the plane $z = 0$ can then be written as

$$\frac{d^2x}{dt^2} + \omega_0^2 x = \frac{q_e E_0}{m_e}\mathrm{e}^{i\omega t}. \tag{4.46}$$

We recognize the equation of a forced oscillator of proper angular frequency ω_0. Note that in writing Eq. (4.41) for the incident field, we have implicitly assumed the wave to be present for an infinite time. Consequently, we must look for the stationary solution to Eq. (4.46). In practice, "infinite time" here means a time much longer than the duration of the transient of the oscillator. This is indeed the case under usual conditions. The stationary solution is

$$x(t) = x_0\mathrm{e}^{i\omega t} = \frac{q_e E_0}{m_e\left(\omega_0^2 - \omega^2\right)}\mathrm{e}^{i\omega t}. \tag{4.47}$$

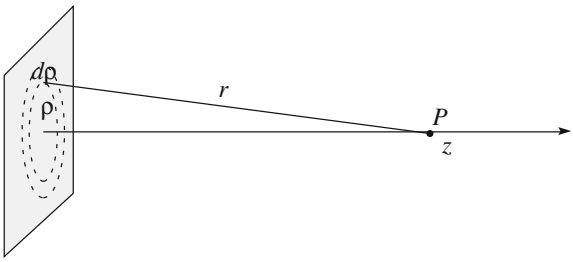


Fig. 4.23 Geometry on the surface of the plate

Consider now an electron in the plane $z = 0$ at a distance ρ from the origin. Let r be its distance from P, as shown in Fig. 4.23. Let us calculate the radiation field of this accelerating electron in P. Considering that the distance r is much larger than the wavelength, we can safely use Eq. (3.42). We need the acceleration at the instant $t - r/c$. This is

$$a\left(t - \frac{r}{c}\right) = -\omega^2 x_0 \mathrm{e}^{i\omega(t-r/c)} = -\frac{\omega^2 q_e E_0}{m_e\left(\omega_0^2 - \omega^2\right)} \mathrm{e}^{i\omega(t-r/c)}.$$

We should now take the component of the acceleration perpendicular to the line of sight, which is the line joining the considered point to P. However, in our case, it is practically $\rho << r$, and we can approximate the normal component with the entire absolute value and write the radiation field as

$$E_e(t) = \frac{q_e}{4\pi\varepsilon_0 c^2} \frac{\omega^2 x_0}{r} \mathrm{e}^{i\omega(t-r/c)}. \tag{4.48}$$

The reader will have noticed that we are not considering the field to be a vector quantity. This is indeed possible because the field has only one component, namely x.

We must now sum up the contributions of all the charges of the plate. Clearly, due to the symmetry of the problem, the contributions of all the charges at the same distance ρ from the z-axis are equal. Let dE_c be the contribution of the circular ring between ρ and $\rho + d\rho$. If n_e is the number of electrons per unit volume, their number in the ring is $n_e h 2\pi\rho\, d\rho$. Their radiation field at P is then

$$dE_c(t) = \frac{q_e}{4\pi\varepsilon_0 c^2} \frac{\omega^2 x_0}{r} \mathrm{e}^{i\omega(t-r/c)} n_e h 2\pi\rho d\rho. \tag{4.49}$$

To have the field in P due to the entire plate, we must now integrate on ρ from 0 to $+\infty$. We write, taking out of the integral all factors independent of ρ (part of the exponential included):

$$E_c(t) = \frac{q_e}{4\pi\varepsilon_0 c^2}\omega^2 x_0 n_e h 2\pi e^{i\omega t}\int_0^\infty \frac{e^{-i\omega r/c}}{r}\rho d\rho.$$

Now, it is $r^2 = \rho^2 + z^2$, and hence, $\rho d\rho = rdr$, z being fixed. We then transform the integral on ρ in an integral on r, changing the limits accordingly. We have

$$\int_0^\infty \frac{e^{-i\omega r/c}}{r}\rho d\rho = \int_z^\infty e^{-i\omega r/c} dr.$$

The integration immediately leads to

$$\int_z^\infty e^{-i\omega r/c} dr = \frac{ic}{\omega}\Big|_z^\infty e^{-i\omega r/c} = \frac{ic}{\omega}\lim_{r\to\infty} e^{-i\omega r/c} - \frac{ic}{\omega}e^{-i\omega z/c}.$$

Now, rigorously speaking, the limit on the right-hand side is indefinite. Indeed, it is the limit of a complex number of unitary magnitude whose argument grows indefinitely. You can think of a unit vector rotating indefinitely in the complex plane. In practice, however, we can think of the limit as being zero. Indeed, the surface of the plate is not infinite, but limited. Consequently, when we get our integration close to the borders, the area of the ring will become smaller than $2\pi\rho\, d\rho$, because part of it will be outside, and shall tend smoothly to zero. In addition, the difference between the normal component of the acceleration and its magnitude shall start to become appreciable sooner or later, once more decreasing the integrand relative to that which we have considered. The rotating vector we have mentioned should then have a magnitude gradually decreasing to zero.

In conclusion, the radiation field in P resulting from the molecular charges in the plate is

$$E_c(t) - i\frac{q_e}{2\varepsilon_0}\omega x_0 \frac{n_e h}{c} e^{i\omega(t-z/c)} = \left(\frac{q_e}{2m_e\varepsilon_0}\right)\left(\frac{n_e}{\omega_0^2 - \omega^2}\right)\left(-\frac{i\omega h}{c}\right)\left[E_0 e^{i\omega(t-z/c)}\right], \tag{4.50}$$

where, on the right-hand side, we have collected the factors in parentheses into four groups. They have different physical meanings that we shall now discuss. The last factor depends on time and on the space coordinate. It tells us that we deal with the field of a progressive wave in the positive direction of the z-axis at the frequency of the incoming wave. This factor is none other than the wave field $\mathbf{E}_1$ that would be the field in P in absence of the plate. We learn then that the field $\mathbf{E}_c$ is a plane progressive wave, completely similar to the incident wave, with amplitude given by the product of the remaining three factors on the right-hand side. We have so shown one aspect of our thesis in Eq. (4.45). Another aspect of this thesis is the factor

$-i\omega h/c$, which is present as well. It remains to be shown that the remaining two factors in Eq. (4.50) can be identified as $(n - 1)$ where n is the index of refraction of the medium. For this to be the case, the factors must be functions of the characteristics of the medium and of the frequency of the wave, but not of the position of P. Indeed, it is so, and we can finally write

$$n = 1 + \frac{q_e^2 n_e}{2m_e\varepsilon_0\left(\omega_0^2 - \omega^2\right)} \tag{4.51}$$

We have found a dispersion formula, namely an expression of the refractive index, which is a quantity determined with macroscopic measurements, in terms of the microscopic characteristics of the medium (electron density, mass and charge), of the proper angular frequency ω_0 of the molecular oscillators and of the angular frequency ω of the light wave. Let us now discuss it.

As a consequence of the assumptions we have made, Eq. (4.51) is valid only for values of n not too different from the unit. Even under these limitations, however, we can ascertain a number of physical aspects.

Let us state now that molecular oscillators are systems with several degrees of freedom and consequently have a number of proper frequencies. To take that into account, we rewrite Eq. (4.51) as

$$n = 1 + \frac{q_e^2}{2m_e\varepsilon_0}\sum_k \frac{n_{e,k}}{\left(\omega_k^2 - \omega^2\right)}, \tag{4.52}$$

This expression is substantially correct, being justified by the fact that it is equal to the result of quantum mechanics (once the completion we shall soon discuss is included). Quantum mechanics predicts the values of the quantities $n_{e,k}$ and ω_k as well. Here, we must leave them as phenomenological parameters.

We start by observing that in the transparent gases, like air, the proper oscillations frequencies ω_k are in the ultraviolet. We know that in the visible, namely $\omega << \omega_k$, the refractive index is larger than 1 (namely the phase velocity is smaller than in a vacuum) and is a slowly increasing function of frequency. This is also true for condensed media like glasses and water, even if in these media, the approximations of our "theory" do not hold. Contrastingly, what we just said is not valid, for example, for colored gases, which have resonances in the visible (see the following).

We still need to add an element to our formula. Indeed, we know that no medium is perfectly transparent. The amplitude of the electromagnetic wave gradually decreases as it propagates in the medium. We easily interpret this fact by considering that the oscillating atomic electrons lose the energy they are emitting. This means that an atomic oscillator, once initially excited, will oscillate for a definite amount of time. We are dealing with damped oscillators. Calling γ_k the damping coefficients, we take this into account by substituting $\omega^2 - \omega_k^2$ in Eq. (4.52), $\omega^2 - \omega_k^2 + i\gamma_k\omega$. Our final dispersion formula is

$$n = 1 + \frac{q_e^2 n_e}{2m_e \varepsilon_0} \sum_k \frac{n_{e,k}}{\left(\omega_k^2 - \omega^2 + i\gamma_k \omega\right)}. \tag{4.53}$$

We have now got a complex diffraction index. What does this mean? We understand this by looking at the field just after the plate, given by Eq. (4.43). Calling n_R and n_I the real and imaginary parts of n, respectively, namely with $n = n_R + in_I$, Eq. (4.43) becomes

$$\mathbf{E}_2 = \left(\mathbf{E}_1 e^{-i\omega(n_R - 1)h/c}\right) e^{-\omega n_I h/c}. \tag{4.54}$$

We see that the factor in parentheses is the field beyond the plate if n had been real (with n_R in place of n). The second factor is an exponential with a *real* exponent. The latter is negative, because quantum calculations show us that n_I is positive under ordinary conditions. Through this factor, the plate acts on the amplitude of the wave rather than on its phase. It expresses the absorption phenomenon. We can say that the amplitude of the wave changes in a thickness h of the medium from, say $\mathbf{E}_{20}$ to $\mathbf{E}_{10}$, which are in the relation

$$\mathbf{E}_{20} = \mathbf{E}_{10} e^{-\omega n_I h/c}. \tag{4.55}$$

We shall come back to this expression soon, after having discussed the real part of the index.

Figure 4.24 shows the real part of Eq. (4.53) in an example with three resonances. What matters here, for each resonance, is the elastic amplitude of the resonance curve, which we discussed with reference to Fig. 1.13. In a frequency interval below a resonance, and far enough from other resonances, the real part of the index is larger than 1 and slowly increases with frequency. The situation is equal to that which we discussed above in the absence of damping. Under these conditions, namely when the index increases with increasing frequencies, we talk of *normal dispersion*, because this is the usual case.

We encounter a new situation near a resonance, in a frequency interval on the order of γ. We see that the index of refraction decreases with increasing frequency. This phenomenon is called *anomalous dispersion*. Just below a resonance, the real

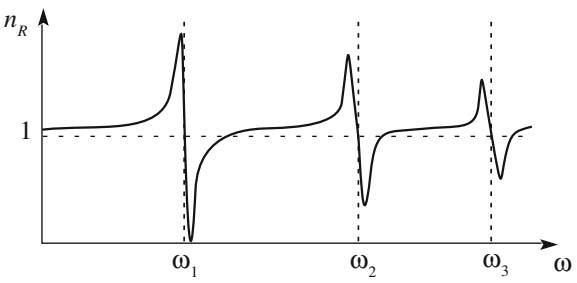


Fig. 4.24 The real part of the refractive index in the region of three resonances

part of the refractive index is $n_R > 1$, while above the resonance, it is $n_R < 1$, namely the phase velocity is larger than c. This is not in contradiction with the relativity principle, because the signals do not travel with the phase velocity.

The reason for the phase velocity being larger in the medium than in a vacuum is simply a consequence of the fact that, at frequencies of the exciting force larger than the resonance frequency, the displacement of an oscillator is in phase opposition to the force. Consequently, for example, a positive charge moves in a sense opposite to the direction of the incoming field. Under these conditions, the contribution of the charge tends to increase that field, causing the phase velocity to increase. The opposite happens below resonance where the field and displacement are in phase. We notice in passing that, in a narrow region just above a resonance, not only is the phase larger than c, but so is the group velocity. However, it has been shown, analyzing the behavior of a "beginning sine" wave, that under these conditions, the signal carrying information does not propagate with the group velocity, but rather with a velocity smaller than c.

Let us now consider the imaginary part of the index. As we already stated, under usual conditions, it is $n_I > 0$ (we shall see an important exception at the end of the section). As shown in Fig. 4.25, the dependence of n_I on frequency is a resonance curve. It is very large at each atomic or molecular resonance, but is negligible at a few widths far from resonances (as a matter of fact, under these conditions, n_I is even smaller, by a significant amount, than what the figure shows). Namely, far from resonances, the refractive index is substantially real. In resonance, as we discussed in Chap. 1, the force, namely the electric field of the wave, is in phase, or almost so, with the velocity, or, more precisely, the deformation rate of the electronic cloud, and consequently, it transfers power to the oscillator with high efficiency. As a consequence, near every resonance, the medium strongly absorbs the incoming radiation. We talk of *absorption bands* (meaning frequency bands). When one of these bands is in the visible, the medium absorbs the corresponding color and the light transmitted by a layer of a certain thickness appears as the complementary color when we look through it.

Let us now take a different point of view and consider a wave advancing in a medium filling the semispace $z > 0$, rather than being a plate. Clearly, we can reinterpret Eq. (4.55) as giving the wave amplitude at the depth of $z = h$ inside the

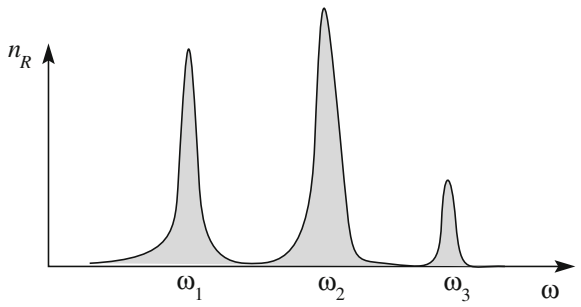


Fig. 4.25 The imaginary part of the refractive index in the region of three resonances

medium. We understand that the amplitude of the field decreases exponentially with increasing distance crossed in the medium. The wave intensity, which is proportional to the square of the amplitude, decreases exponentially as well.

We define as the absorption distance, or alternately the attenuation distance, the distance d on which the wave intensity decreases by a factor of $1/e$. Hence, the amplitude decreases by a factor of $1/e$ along a distance twice as large, namely $2d$. Let a $\mathbf{E}_0(0)$ and $\mathbf{E}_0(h)$ be the field amplitudes at the entrance and at depth h, respectively. The absorption distance is then, by definition, given by the equation

$$\mathbf{E}_0(h) = \mathbf{E}_0(0)\mathrm{e}^{-h/2d}. \tag{4.56}$$

Equation (4.55) gives us the absorption distance in terms of the imaginary part of the index as

$$d = \frac{c}{2\omega n_I} = \frac{\lambda}{4\pi n_I}. \tag{4.57}$$

In nature, the absorption lengths of the transparent gases are on the order of several kilometers. This means, according to Eq. (4.57), that the imaginary part of the index is extremely small. For example, for typical values of $d = 10$ km and $\lambda = 0.5$ μm, we have $n_I = 4 \times 10^{-10}$, which is a really small number.

We see that a positive imaginary part of the refractive index corresponds to a decreasing amplitude of the wave propagating in the medium, namely to an absorption of the wave energy by the medium. Figure 4.26a shows the wave amplitude as a function of the crossed depth. We have drawn the figure for an absorption length as not much larger than the wavelength to make the effect more visible. It is possible to prepare certain classes of materials with the majority of their atoms at an excited energy level. To do that, we have stored, or, as we say, pumped, energy into the medium, which is in a metastable state, like a compressed spring locked in that state but ready to snap. The electric field of the wave propagating in the medium is capable of triggering the de-excitation of the molecules, provided its frequency is equal to the resonance frequency of the molecules. In this case, the amplitude of the wave grows exponentially with increasing distance z crossed in the medium. This situation is shown in Fig. 4.26b. Equation (4.55) holds in this case

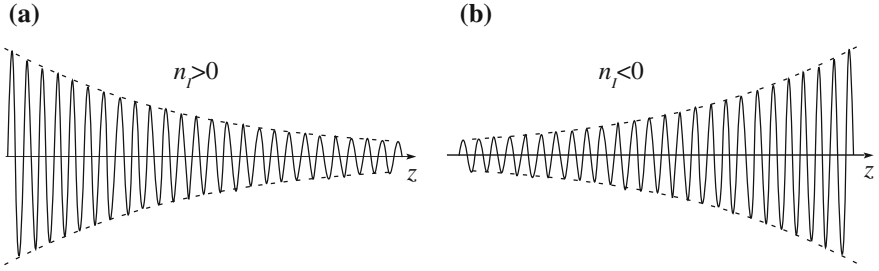


Fig. 4.26 Quasi-monochromatic oscillation with **a** exponentially decreasing, **b** exponentially increasing amplitude

too, now with $n_I < 0$. We talk of negative absorption and a negative imaginary index. This is what happens in a LASER, which stands for Light Amplification by Stimulated Emission of Radiation.

Notice that in a LASER, all the molecules' oscillations are locked in phase, namely they are such so as to remember the phase of the incident wave. Contrastingly, in a usual source of light, called a thermal source, each excited molecule de-excites independently of the others. Each emitted wavelet has a random initial phase. In other words, in a LASER, the memory of the initial phase is conserved for a very long time, while in a thermal source, it is so only for a time comparable with the decay times of the atoms. We shall see some examples in Chap. 8 of how this important feature of the LASER light can be exploited in image formation.

4.8 Electromagnetic Waves in Transparent Dielectric Media

In Sect. 3.6, we found the wave equations of the electromagnetic field in a vacuum starting from the Maxwell equations. These waves are non-dispersive. Their phase velocity, the speed of light in a vacuum, is a universal constant of physics, which is independent of the (inertial) frame. We shall find here the wave equation for the electromagnetic waves in a normal (namely linear, homogenous and isotropic) dielectric medium. We shall find these waves to be dispersive and to obey a different wave equation. We shall find a dispersion relation valid for condensed media.

We shall proceed in a manner similar to that which we used in a vacuum in Sect. 3.6, taking into account, however, the existence of polarization charges and currents. We shall, on the contrary, neglect any magnetic effect, because under the largest fraction of the conditions encountered in practice, the magnetic susceptibility is very small.

We start from the Maxwell equations in the fields **E**, **D** and **B** (See Volume 3, Sect. 10.7), taking into account that the free charge density and the conduction current are zero. The equations are

$$\nabla \cdot \mathbf{D} = 0, \tag{4.58}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{4.59}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{4.60}$$

$$\nabla \times \mathbf{B} = \mu_0 \frac{\partial \mathbf{D}}{\partial t}. \tag{4.61}$$

To these, we must add the relation between **E** and **D**, which is

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}. \tag{4.62}$$

As in Sect. 3.6, we take the curls of both sides of Eq. (4.59), obtaining

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial \nabla \times \mathbf{B}}{\partial t}.$$

We then use the vector identity $\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$ and use Eq. (4.61), obtaining

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu_0 \frac{\partial^2 \mathbf{D}}{\partial t^2}. \tag{4.63}$$

We take the divergence of both sides of Eq. (4.62), taking Eq. (4.58) into account. We find

$$\nabla \cdot \mathbf{E} = -\frac{1}{\varepsilon_0} \nabla \cdot \mathbf{P}.$$

We substitute in Eq. (4.63) and eliminate **D** on the right-hand side using Eq. (4.62), obtaining

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\varepsilon_0 \mu_0} \nabla^2 \mathbf{E} = \frac{1}{\varepsilon_0^2 \mu_0} \nabla(\nabla \cdot \mathbf{P}) - \frac{1}{\varepsilon_0} \frac{\partial^2 \mathbf{P}}{\partial t^2}. \tag{4.64}$$

This is the wave equation we were looking for (a similar equation holds for **B**, but we shall not need it). We see that the left-hand side is the same as for the wave equation in a vacuum, but we now have a non-zero right-hand side depending on the polarization density **P** and on its variations in time and space.

To continue, we need a relation between polarization **P** (the effect) and electric field **E** (the cause). In Volume 3 (Sect. 10. 7), we saw that such a relation does not exist for fields with an arbitrary dependence on time. The physical reason is that polarization induced by a given electric field intensity depends on how fast the field is varying with time. The relation exists, however, when the time dependence is a circular function (namely $\sin\omega t$ or $\cos\omega t$). Under these conditions, we have

$$\mathbf{P} = \varepsilon_0 \chi_e(\omega) \mathbf{E}, \tag{4.65}$$

where the electric susceptibility χ_e is a function of the angular frequency of the field. As we know, under the action of the electric field, the molecules of the medium are stretched and, if they have an intrinsic dipole moment, reoriented. Well, when the frequency of the field increases, the time interval in which the field inverts direction diminishes. The molecules have less time to react. Consequently,

the net effect, namely the polarization, decreases. We must also consider that the effect lags somewhat behind its cause. Consequently, the phase of **P** at a certain instant is not the phase of **E**. This translates into the fact that the susceptivity $\chi_e(\omega)$ is a complex number. Its magnitude gives the "size" of the polarization, and its argument is the phase difference between **E** and **P**. Finally, being that the medium is isotropic, **P** is parallel to **E**.

We are considering monochromatic and plane waves. Let us check if they are solutions of the wave equation in Eq. (4.64). Let the electric field and polarization be

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0 \mathrm{e}^{i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{4.66}$$

and

$$\mathbf{P}(\mathbf{r},t) = \mathbf{P}_0 \mathrm{e}^{i(\omega t - \mathbf{k}\cdot\mathbf{r}+\delta)}, \tag{4.67}$$

where we have taken into account a possible phase difference δ.

We can simplify Eq. (4.64) for monochromatic fields by showing that the divergence of the polarization is zero (as it is in electrostatics). Indeed, we take the divergence of Eq. (4.62), taking Eq. (4.65) into account, and obtain

$$\nabla\cdot\mathbf{D} = \nabla\cdot[1+1/\chi_e(\omega)]\mathbf{P} = [1+1/\chi_e(\omega)]\nabla\cdot\mathbf{P},$$

where, on the right-hand side, we took into account that, the dielectric being homogeneous, the susceptibility does not depend on the coordinates. Equation (4.58) then gives

$$\nabla\cdot\mathbf{P} = 0.$$

The wave equation Eq. (4.64) becomes

$$\frac{\partial^2\mathbf{E}}{\partial t^2} - \frac{1}{\varepsilon_0\mu_0}\nabla^2\mathbf{E} = -\frac{1}{\varepsilon_0}\frac{\partial^2\mathbf{P}}{\partial t^2}. \tag{4.68}$$

Let us now substitute in this equation our tentative solution in Eqs. (4.66) and (4.67). We obtain the equation

$$-\omega^2\mathbf{E}(\mathbf{r},t) + c^2k^2\mathbf{E}(\mathbf{r},t) = \frac{\omega^2}{\varepsilon_0}\mathbf{P}(\mathbf{r},t),$$

We substitute for **P** its expression of Eq. (4.65), obtaining

$$-\omega^2\mathbf{E}(\mathbf{r},t) + c^2k^2\mathbf{E}(\mathbf{r},t) = \omega^2\chi_e(\omega)\mathbf{E}(\mathbf{r},t).$$

We can now simplify out **E**, which appears in all terms, obtaining

$$\omega^2 = \frac{c^2}{1+\chi_e(\omega)}k^2. \tag{4.69}$$

This dispersion relation in terms of the phase velocity is

$$v_p(\omega) = \frac{\omega}{k} = \frac{c}{\sqrt{1+\chi_e(\omega)}} = \frac{c}{\sqrt{\kappa(\omega)}}, \tag{4.70}$$

where, on the right-hand side, κ is the dielectric constant of the medium relative to a vacuum. Obviously, we can write the relation in terms of the refractive index as

$$n(\omega) = \sqrt{\kappa(\omega)}, \tag{4.71}$$

In this form, the relation can be verified experimentally by measuring, for a monochromatic wave of angular frequency ω, the dielectric constant and the refractive index. We determine the dielectric constant by measuring the capacitance of a condenser with the medium under study as a dielectric. This must be done under dynamic, rather than static, conditions, namely polarizing the condenser with an alternate emf at the angular frequency ω. Note that this is feasible up to the frequencies of microwaves, but not at those of light.

In Sect. 4.7, we discussed the relation between the refractive index and the microscopic properties of a low-density medium, such as a gas. The equations discussed in this section are valid for condensed media as well. Even in this case, the polarization is the result of the stretching and reorienting of the molecules under the action of the electric field. This is expressed by Eq. (4.46). However, in a condensed medium, the electric field is not the field of the incoming wave alone, because the fields of the other molecules also contribute. Consequently, Eq. (4.52) does not hold for a condensed medium. The corresponding equation can be obtained in a similar manner to that which we used under static conditions in Sect. 4.8 of the 3rd volume.

In Sect. 4.6, we used the fact that, in a normal dielectric, the ratio between the amplitudes of the electric and magnetic fields of a progressive monochromatic wave is equal to the phase velocity, namely that $E/B = v_p(\omega)$. Here, we simply state that the demonstration is identical to that which we conducted in Sect. 3.6 in a vacuum to show that $E/B = c$, with the only difference being that one must start from Eq. (4.59) or Eq. (4.61) and, obviously, remember that $v_p(\omega) = \omega/k$.

Summary

In this chapter, we have studied several phenomena in which the phase velocity depends on the wavelength. In particular, we learned the following important concepts.

1. In a dispersive medium, a harmonic wave propagates without deformations, but its phase velocity ω/k depends on the wavelength. A non-harmonic wave changes shape while it propagates.
2. The group velocity of a wave packet is given by the derivative $d\omega/dk$ at the dominant wavelength of the group. Group and phase velocity are equal only in the dispersive media.
3. The phase velocity of an electromagnetic wave in a dispersive medium is different from that in a vacuum. The refractive index is the ratio between the two velocities. They both depend on wavelength.
4. The wave equation for an electromagnetic wave in a normal dielectric has a left-hand side equal to the equation in a vacuum, but a different right-hand side
5. The sudden change of the index at the interface between two transparent media is at the origin of the reflection and the refraction.
6. The phenomenon of the total internal reflection and the evanescent wave.
7. The microscopic origin of the refractive index.

Problems

4.1 Find the dependence on frequency of phase and group velocities for a piano string having the dispersion relation $\omega^2 = (T_0/\rho)k^2 + \alpha k^4$.

4.2 Calculate the frequency dependence of the group velocity in the following cases of dispersion relations (α is a constant): (a) $v_p = \alpha$ (example: sound waves in air), (b) $v_p = \alpha\sqrt{\lambda}$ (example: surface gravity waves on a water surface), (c) $v_p = \alpha/\sqrt{\lambda}$ (example: capillary waves on a water surface).

4.3 The curve in Fig. 4.27 represents a dispersion relation. What are the relations of the phase and group velocities for $k = k_0$, with the angles α and β, respectively? Line t is the tangent to the curve in P, line c joins P with the origin of the axes. What should the curve be in order for the two velocities to be equal to one another?

4.4 Calculate the limit angle for total internal reflection between a glass of index $n = 1.65$ and air and between water ($n = 1.33$) and air.

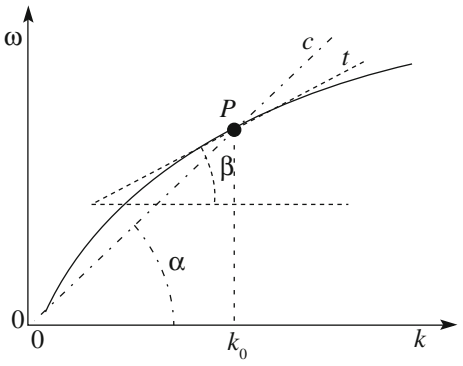


Fig. 4.27 A dispersion relation

4.5 What are the intensity reflection and transmission coefficients for normal incidence at the interface between water (n_1 = 1.33) and glass (n_2 = 1.53)?

4.6 Give an estimate of the distance between the surfaces of two glass blocks in air allowing for an appreciable tunnel effect (say, in order to have an amplitude reduction by 1/10) if the index of the glass is n = 1.5, the vacuum wavelength of light is 0.5 μm and the incidence angle is 45°.

4.7 A lamp on the bottom of a swimming pool 1.5 m deep emits light. An observer outside the water sees a light circle on the surface of 3.40 m diameter. What is the index of water?

4.11 The fields of an electromagnetic wave are $\mathbf{E} = \mathbf{E}_0 \cos(\omega t - kz)$ and $\mathbf{B} = \mathbf{B}_0 \sin(\omega t - kz)$. It reflects on a plane surface normal to the propagation direction at z = 0. Write down its fields.

Chapter 5
Diffraction, Interference, Coherence

Abstract In this chapter, we study different types of interference phenomena. Interference happens when two or more waves having a phase relation with one another fixed in time overlap in a region of space. We study Young's two-slit experiment and then the coherence conditions, namely the conditions that must be satisfied for interference phenomena to be observable. After having introduced diffraction with Grimaldi's discovery, we treat the phenomenon under Fraunhofer conditions in the important cases of the slit, the circular aperture, randomly distributed centers and the diffraction grating. Finally, we study the close relations between the physics of diffraction and the mathematics of the Fourier transform.

In this chapter, we study different types of interference phenomena. Interference happens when two or more waves having a phase relation with one another fixed in time (during the observation) overlap in a region of space. Suppose, for example, you take in your hands two sticks, each terminating with a small sphere like a tennis ball, you lay them on the quiet surface of a pond and you move them rhythmically up and down. You will observe two systems of expanding circular waves. At certain points, the waves are constantly in phase with one another, at others, constantly in phase opposition, etc. In the former positions, the crests of one system arrive at the same time as the crests of the other and similarly for the throats. The resulting oscillation amplitude is large (constructive interference). In other positions, the crests of one system arrive together with the throats of the other, and the resulting oscillation amplitude is small or null (destructive interference).

To be detectable, any interference phenomenon should not change during the observation time. In particular, two interference waves should be at least approximately monochromatic with the same frequency and with a fixed phase difference with one another. These conditions are called coherence conditions (Sect. 5.3). For transverse waves, interference is between oscillations in the same direction.

Interference phenomena happen for waves of every type. Their mathematical descriptions are similar. Contrastingly, the physical characteristics and the ways of producing and detecting the phenomenon vary strongly from one case to another. We shall only discuss the interference of light here. In this case, the coherence conditions

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_5

are often not satisfied in nature. However, when they are, we observe such beautiful phenomena as the color of mother-of-pearl, those of the wings and elytra of certain insects and the coronas about light sources in the presence of mist or ice in air.

Diffraction is similar to interference. It happens every time a wave that was freely advancing meets an obstacle that delimits its front. In such a case, the wave not only continues to move straight forward, but parts of the front of it go around the obstacle, changing the direction of propagation in doing so. We shall also discuss the scattering of light by illuminated objects. The scattered light is the result of the infinite wavelets emitted by the different parts of the object, as excited by the incident wave. One phenomenon explained by the dependence on wavelength of the scattering probability is the blue color of the sky.

We shall limit our mathematical description of interference and diffraction phenomena to the simplest conditions. These, called the Fraunhofer conditions, are implemented by having both the primary light source and the system producing the phenomenon and this system and the observation plane sitting at large distances from each other. The "infinite distance" conditions can be obtained in practice using lenses.

We shall start in Sect. 5.1 by demonstrating simple examples as to how the Huygens-Fresnel principle allows one to construct the propagating wave front beyond an obstacle. In Sect. 5.2, we discuss the conceptually simplest (and historically first) interference experiment, namely the Young two-hole (or two-slit) experiment. In Sect. 5.3, we define the conditions to be implemented in order to perform a successful interference experiment, which are the coherence conditions. In Sect. 5.4, we shall see how exceptionally interference phenomena are observable under unsatisfied coherence conditions. The interference fringes are localized under these conditions.

We then address important cases of diffraction, namely diffraction by a slit, by a circular opening, by randomly distributed centers, and by periodically distributed centers. We shall come, in this way, to the diffraction gratings, which are an important class of optical instrument, and study their most relevant properties.

In Sect. 5.10, we study the close relations between the physics of diffraction and the mathematics of the Fourier transform.

5.1 Huygens-Fresnel Principle

We have already studied the propagation of a light wave in space in some important cases, namely in a vacuum, in a homogeneous medium and through the interface surface between two homogeneous media. These are quite simple situations in which the approximations of geometrical optics can be safely taken. A completely general and rigorous treatment of the wavefront propagation in three dimensions in the presence of limiting obstacles or apertures of arbitrary form, is of considerable mathematical complexity. Simpler methods exist, however, capable of treating the simplest geometries with various degrees of approximation. The most important of these, which we shall use in the following sections, is called the Huygens-Fresnel principle.

In 1690, Christiaan Huygens (The Netherlands, 1629–1695) published the "Treaité de la lumière", in which he was the first to advance a theory of light as a wave phenomenon. With his theory, in particular, he correctly interpreted reflection and refraction along the lines we studied in Chap. 4. Here, we are interested in the recipe proposed by Huygens for a geometrical construction of wavefront propagation. He proposed that every point of a wavefront at a given time becomes the source of a hemispherical wavelet, the sum of these secondary wavelets being the wavefront at any subsequent time. The Huygens principle was extended by Augustin-Jean Fresnel (France, 1788–1827) in 1818 in his "Mémoire sur la diffraction de la lumière", including the interference (a phenomenon that we shall discuss in the next section) between the secondary wavelets. In this way, the principle became an extremely powerful tool, allowing Fresnel to explain all the diffraction phenomena, as we shall see in the subsequent sections. To be complete, we will just mention the further elaboration by Gustav Kirkhoff (Germany, 1824–1887). His diffraction formula provides a precise mathematical prescription for the propagation and diffraction, but we shall not need it.

Proceeding intuitively, let us consider, for example, a wave in a medium such as on the surface of water. We might toss a stone into a calm pond and observe the expanding circular wave pattern. Let us fix our attention on the outermost ring. Beyond it, the water surface is still calm. How will the perturbation reach further regions? Evidently, the oscillations of the different portions of the wavefront we are considering induce movements in the water particles that are immediately in contact with them in the external part of the wavefront. When these particles oscillate, it will induce motions in the still forward water particles, and so on. This implies that what matters for the future motion of the water beyond the considered wavefront is only the motion of the wavefront itself in the instant in time being considered. In other words, the boundary conditions determine the solution in the external space. That is, everything proceeds as if all the points of the wavefront were small secondary sources of waves to the outside.

Let us now establish the Huygens-Fresnel principle. Consider a monochromatic wave, of phase velocity v_p, and suppose we know that the wavefront at the instant t is the circle AB shown in Fig. 5.1, moving in the direction of the arrows. To construct the wavefront at the subsequent instant $t + \Delta t$, we take the following steps. (a) We consider every small element of the front AB as being a source of secondary hemispheric wavelets emitted in the region external to the front. The radiuses of the wavelets are $v_p \Delta t$. Being that all the secondary sources are driven by the same incident waves, the phases of the secondary waves emitted are equal to one another, namely they oscillate coherently. (b) We draw the envelope of the secondary wavelets, obtaining the wavefront at $t + \Delta t$ (A_1B_1 in the figure).

It is easy to see that, if the wavefront AB is spherical at time t, at time $t + \Delta t$, the front is spherical as well, with a radius larger by $v_p \Delta t$. Similarly, if the wavefront AB at t is plane, then A_1B_1 is plane as well, etc.

The Huygens-Fresnel principle finds its most relevant applications in the construction of the wavefront beyond obstacles and openings that block a segment of the wave. Consider, for example, a monochromatic plane wave reaching a screen

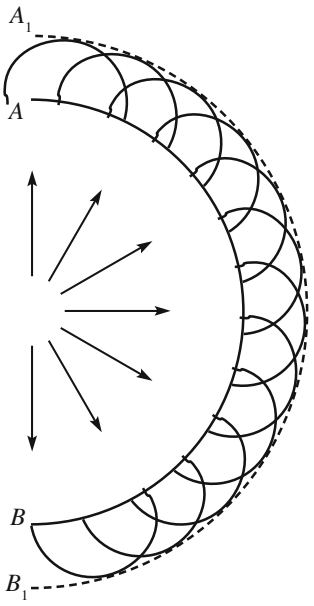


Fig. 5.1 Wavefront construction according to the Huygens-Fresnel principle

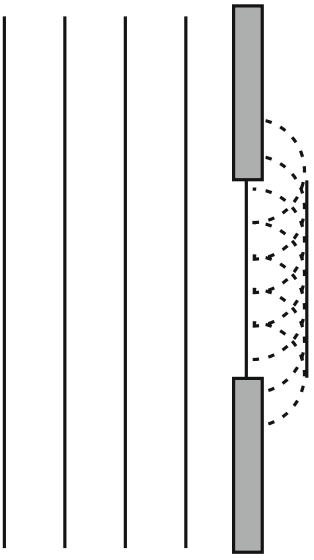


Fig. 5.2 Huygens-Fresnel construction at an obstacle

normal to its propagation direction, in which an opening is present, as shown in Fig. 5.2. We consider the wavefront in the aperture and apply the Huygens-Fresnel construction. We see that the proper envelope of the secondary wavelets corresponds to the light beam in the region of the geometrical light (as opposed to the geometrical shadow).

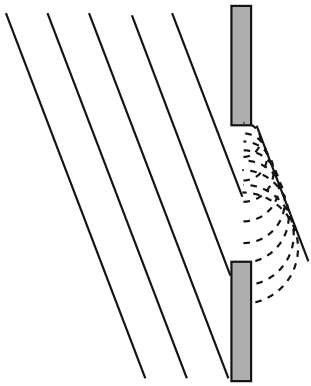


Fig. 5.3 Huygens-Fresnel construction at a slant obstacle

Indeed, on the envelope, all the secondary wavelets add on in phase, enhancing one another's contributions. This produces a wave of large amplitude, as large as the incident one. Note, however, that at both sides of the envelope, more wavelets exist that expand into the region of the geometrical shadow. Their contribution is small, but not negligible. Indeed, it was Fresnel who understood the importance of the wavelets outside the region of geometric optics and calculated their contribution, developing the theory of diffraction, as we shall discuss in the subsequent sections.

We shall also deal with apertures lying on surfaces different from a wavefront. Consider, for example, a monochromatic plane wave coming in slantwise, as in Fig. 5.3. In this case, we must extend the wavefront construction recipe a bit. We consider the surface of the aperture as being the locus of the secondary sources. Even if their phases are not equal to one another, they still have fixed relations between them, determined by the incident wave, as shown in the figure.

Notice also that we constructed the hypothesis that the secondary wavelets are hemispherical, rather than spherical. Indeed, in the latter case, it would have a second front propagating backwards that does not exist. This hypothesis was introduced by Huygens on the basis of physical sense. Later on, Fresnel found it necessary to admit that the amplitude of each secondary wavelet is a maximum in the forward direction and smoothly decreases for increasing angles to the incident direction, vanishing at 90°. This is called the *obliquity factor*.

5.2 Light Interference

Interference and diffraction are two fundamental phenomena characteristic of all waves. As a matter of fact, they are substantially the same phenomenon and the two terms are almost synonymous. In any case, we have a number of contributing sources. If this number is small, we talk of interference; if it is large, or infinite, we talk of diffraction. In this book, we shall focus on light. Being that its wavelengths

are very small, a fraction of a μm, compared with the usual dimensions, and its frequencies very large, on the order of hundreds of THz, its wave characteristics are not easy to observe. Historically, it was only in the XVII century that diffraction was discovered by Francesco Grimaldi (Italy, 1618–1663). His observations were published in the posthumous book *De lumine* in 1665. Here, Grimaldi also introduced the term diffraction, from the Latin verb *frangere*, meaning to break, as he was thinking of light as being an extremely fast moving fluid that breaks when it encounters an obstacle. The title of the first proposition of *De lumine*, translated from Latin, is:

> Light propagates or diffuses not only directly, by reflection and by refraction, but also in a fourth way, namely by diffraction.

The wave nature of light was discovered by Thomas Young (UK, 1773–1829). The report of his first results to the Royal Academy in London in 1803 starts with the sentences

> In making some experiments on the fringes of colours accompanying shadows, I have found so simple and so demonstrative a proof of the general law of the interference of two portions of light (*i.e. a light beam*), which I have already endeavoured to establish, that I think it right to lay before the Royal Society, a short statement of the facts which appear to me decisive. The position on which I mean to insist at present, is simply this, that fringes of colours are produced by the interference of two portions of light; and I think it will not be denied by the most prejudiced, that the assertion is proved by the experiments I am to relate, which may be repeated with great ease whenever the sun shines, and without any other apparatus than at hand of every one.

The first important element for any diffraction or interference experiment is the primary light source, which must be quite small, no wider than a few millimeters, for reasons that will become clear in the next section. The source should also be sufficiently intense. Grimaldi, Young and Fresnel conducted their experiments on sunny days in a completely dark room, but for a pinhole in a window shutter sending in a narrow sunbeam. Young additionally used a mirror to direct the pinhole beam horizontally across the room, in order to work more comfortably.

Question Q 5.1. Considering that the sun power in the visible spectrum on a summer sunny day at intermediate latitudes is on the order of 500 Wm^{-2}, calculate the light intensity of a beam through a pinhole in a window shutter of 2 mm diameter. □

The second fundamental element is to have two (or more) sources with a fixed phase difference with one another. This is done using the light waves originated by a *single source*, namely the pinhole, and having them follow two different paths and subsequently rejoin and interfere. In his first experiments, reported to the Academy in 1803, Young used a 0.8 mm thick paper card to break the beam in two, similarly to what Grimaldi had done with a wire. One part of the beam was passing on the left and one part on the right side of the card. In a later experiment, the famous two-slit experiment, Young used an absorbing screen, normal to the primary beam, in which he had opened two narrow slits very close to one another. For simplicity, we shall begin by considering two pinholes rather than two slits.

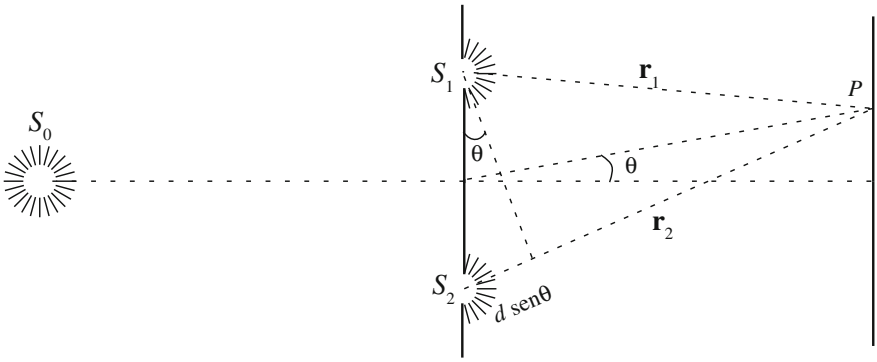


Fig. 5.4 Scheme of Young's two-hole experiment

Figure 5.4 shows the logic scheme of the two-hole experiment. The source S_0 illuminates two small holes S_1 and S_2. We assume, for the moment, that S_0 is point-like and monochromatic. We shall discuss in the next section how and within which limits these hypotheses can be relaxed.

For the Huygens-Fresnel principle, the radiation beyond the screen is equal to that which would occur if, instead of the two holes S_1 and S_2, there were two sources emitting hemispherical waves with a phase relationship determined by the wave illuminating the holes.

If S_0 is now exactly on the axis of the segment S_1S_2, the phases of the waves incident on S_1 and S_2 are equal. Here, "exactly" means that the difference between the distances from S_0 to S_1 and from S_0 to S_2 should be much smaller than a wavelength. Indeed, a difference on the order of 200 nm is enough to have a phase difference between S_1 to S_2 as large as π. Clearly, this is impossible in practice, but fortunately, it is not necessary. Calling ϕ_1 and ϕ_2 the phases in S_1 and S_2, respectively, we find that the time dependence of the electric fields radiated by the two secondary sources are $E_1 = A_1\cos(\omega t + \phi_1)$ and $E_2 = A_2\cos(\omega t + \phi_2)$. The condition for observing the interference phenomenon is not that of ϕ_1 and ϕ_2 being equal, but their difference must be defined and remain constant over a sufficiently long interval of time.

We shall think here of an electric field of the wave always oscillating in the same direction. We can treat it as a single, non-vector quantity.

Let us consider a point P beyond the plane of the holes and let $\mathbf{r}_1$ and $\mathbf{r}_2$ be its position vectors relative to S_1 and S_2, respectively. The light intensity in P is proportional to the average square of the electric field. The electric is the sum of the fields due to S_1 and S_2 in P. Taking into account the proportionality to the inverse distance from the source and the phase delay due to propagation $\mathbf{k} \cdot \mathbf{r}_i$, the field in P is

$$E(P) = \frac{A_1}{r_1}\cos(\omega t - kr_1 + \phi_1) + \frac{A_2}{r_2}\cos(\omega t - kr_2 + \phi_2). \tag{5.1}$$

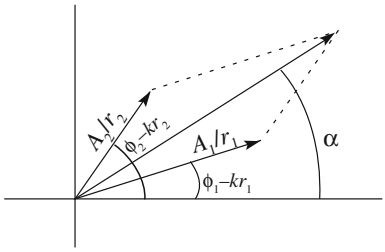


Fig. 5.5 Rotating vectors representation of interference

The easiest way to calculate the average of the square of this expression

$$E^2(P) = \left[\frac{A_1}{r_1}\cos(\omega t - kr_1 + \phi_1) + \frac{A_2}{r_2}\cos(\omega t - kr_2 + \phi_2)\right]^2$$

is to use the rotating vector representation of harmonic oscillations, as shown in Fig. 5.5 at the instant $t = 0$. With time, the two vectors rotate with the same angular velocity ω, and the angle between them, which is the phase difference between the two fields, remains constant. The magnitude of the resultant rotating vector is consequently constant as well. From the geometry of the figure, we immediately have that

$$E^2(P) = \left[\frac{A_1^2}{r_1^2} + \frac{A_2^2}{r_2^2} + 2\frac{A_1}{r_1}\frac{A_2}{r_2}\cos(\phi_2 - \phi_1 - k(r_2 - r_1))\right]\cos^2(\omega t + \alpha),$$

where α is the initial phase of the resultant vector. We do not need to find it, because we just need the average over a period of the only time-dependent factor, namely $\cos^2(\omega t + \alpha)$, which is equal to ½ independently of α. In conclusion, the light intensity in P is proportional to

$$\langle E^2(P)\rangle = \frac{A_1^2}{2r_1^2} + \frac{A_2^2}{2r_2^2} + \frac{A_1}{r_1}\frac{A_2}{r_2}\cos(\phi_2 - \phi_1 - k(r_2 - r_1)).$$

We now look for the physical meaning of the expression we have found. Let us ask us what the intensities would be in P, say I_1 and I_2, if only one of the holes had been open. Clearly, I_1 and I_2 are proportional to $\frac{A_1^2}{2r_1^2}$ and $\frac{A_2^2}{2r_2^2}$, respectively. In conclusion, we can write

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\phi_2 - \phi_1 - k(r_2 - r_1)). \tag{5.2}$$

Let us now discuss the expression we have found. First of all, we note that the total intensity is not simply the sum of the intensities I_1 and I_2 of the two sources, but that the term $2\sqrt{I_1 I_2}\cos(\phi_2 - \phi_1 - k(r_2 - r_1))$ must be added to them. This is called the *interference term*. Indeed, the superposition principle holds for the field, not for the intensity.

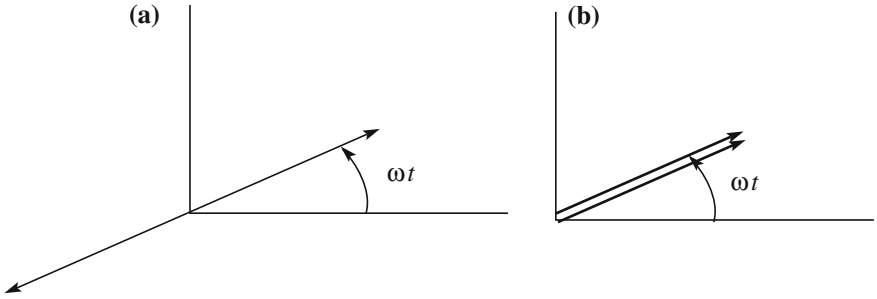


Fig. 5.6 Rotating vector diagrams for equal amplitudes **a** completely destructive interference, **b** completely constructive interference

The interference term may be positive (constructive interference), null or negative (destructive interference), depending on the argument of the cosine in Eq. (5.2). The latter is *the phase difference in P* and depends on two quantities: the initial phase difference $\phi_2 - \phi_1$, which depends on the sources and is independent of the position of P, and the term $k(r_2 - r_1)$, which is independent of the phases of the sources and depends on the position of P relative to them. At the points where the interference is destructive, the intensity is less than the sum of the intensities that would be in the presence of S_1 and S_2 separately. It is a minimum when the argument of the cosine is an odd multiple of π. The two rotating vectors are opposite or in phase opposition. In particular, if we adjust the system to have $I_1 = I_2$, we have complete dark under these conditions. The corresponding rotating vector representation is shown in Fig. 5.6a.

Contrastingly, at the points of constructive interference, the intensity is larger than the sum $I_1 + I_2$, being at a maximum when the argument of the cosine is an integer multiple of 2π, namely when the rotating vectors are in phase with one another. If, in particular, $I_1 = I_2$, the intensity in the maxima is four times larger than it would be in the presence of only one of the sources. The corresponding rotating vectors representation is shown in Fig. 5.6b.

The intensity differences between maxima and minima exist, but are smaller when I_1 and I_2 are different. It may be quite small if $I_1 \gg I_2$ or $I_1 \ll I_2$. Figure 5.7 shows an example.

Let us have a closer look at the maximum interference condition, which is given by

$$\phi_2 - \phi_1 - k(r_2 - r_1) = 2n\pi, \text{ with } n = 0, 1, 2\ldots \tag{5.3}$$

This is the equation of the loci of the maxima in the semi-space beyond the hole's plane. The loci are rotation hyperboloids with foci in S_1 and S_2. To better understand the situation, look at Fig. 5.8, which shows the crests of the circular waves emitted by S_1 and S_2. You can observe a similar pattern on the quiet surface of a pond. To this purpose, take two balls of equal size (tennis balls or similar) and

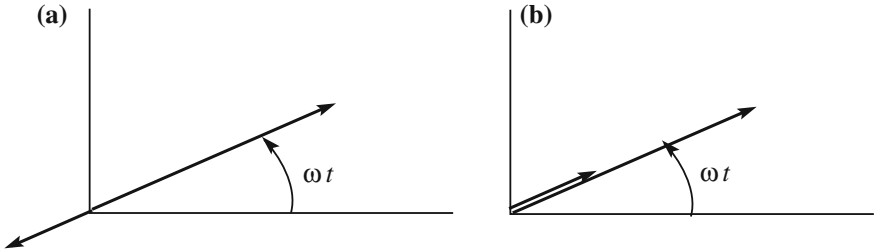


Fig. 5.7 Rotating vector diagrams for different amplitudes **a** completely destructive interference, **b** completely constructive interference

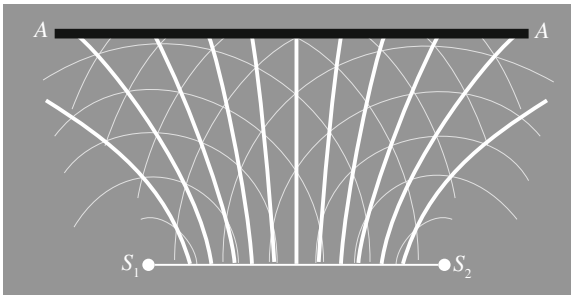


Fig. 5.8 Interference of two circular wave systems

attach each of them to two sticks. With both sticks in your hands, lay the balls on the surface and then move them up and down in phase. The thicker curves in Fig. 5.8 are the loci in which the two wave systems are in phase with one another. There, the oscillation amplitude and intensity have their maxima. With time, the two systems of circular waves expand, but the loci of maximum interference remain stationary. We see that they are hyperbolas. These are the sections of the hyperboloids we have mentioned with the plane of the figure.

In practice, we can see the phenomenon using a white screen parallel to the plane of the holes at a certain distance beyond them, such as the plane *AA* in Fig. 5.8. The screen cuts the hyperboloids, showing a system of luminous hyperbolas with the same foci, called interference fringes, separated by dark bands.

Particularly simple are the Fraunhofer conditions, named after Joseph von Fraunhofer (Germany, 1787–1826), who made many contributions to optics. We shall define these conditions precisely in Sect. 5.6. Under Fraunhofer conditions, the interference pattern is collected at an infinite distance. In practice, we can go a distance of several meters or we can use a convergent lens and locate the screen *AA* in its back focal plane. Figure 5.9 shows the situation. Under these conditions, we have $r_1 \approx r_2$ and, taking, for simplicity, $A_1 = A_2$, $I_1 = I_2$ as well.

We shall now find an expression for the *interference pattern*, namely for the light intensity as a function of the position on the screen. We choose a system of Cartesian coordinates on the screen with the origin O on the axis of the segment

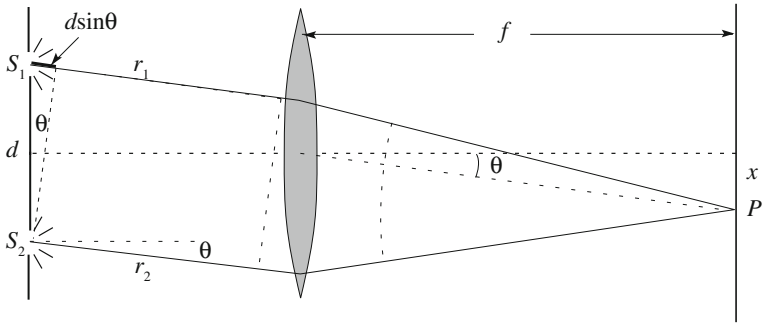


Fig. 5.9 The two-hole experiment under the Fraunhofer conditions

joining S_1 and S_2, the x-axis parallel to this segment and the y-axis perpendicular to x. Let P be a point on the x-axis at the coordinate x. As we shall see in Chap. 7, the radiation arriving in P if the screen is in the focal plane at the focal distance f from the lens is the radiation emitted by S_1 and S_2 at the angle θ, as shown in the figure. In the approximation of small angles, this is approximately $\theta \approx x/f$. Notice that the lens does not introduce any phase difference between the points on a surface normal to the propagation direction.

Consider the dotted plane perpendicular to the propagation direction just before the lens shown in the figure. To reach it, the waves from S_1 must travel a path longer than that from S_2 by $d\sin\,\theta$, where d is the distance between the holes. Consequently, its phase delays relatively by $kd\sin\,\theta$. After that, no further phase difference develops. In conclusion, using $I_1 = I_2 = i$, the light intensity I on the screen is

$$I = 2i[1 + 2i\,\cos(\phi_2 - \phi_1 - kd\,\sin\,\theta)] = 4i\,\cos^2\frac{\phi_2 - \phi_1 - kd\,\sin\,\theta}{2}. \qquad (5.4)$$

We see that, being the phase difference $\phi_2 - \phi_1$ between the two sources fixed, the light intensity on the screen varies as a cosine square as a function of $\sin\,\theta$, and of x approximately as well, being that, for small angles, $\sin\,\theta \approx x/f$. To understand the dependence on y, we consider that the hyperbolic interference fringes will appear almost as straight lines parallel to y about their vertex on a screen at a large distance. In a first approximation, when the screen does not extend by much in the y direction, light intensity is a function of x alone.

Now imagine using two parallel slits in the y direction in place of the two holes. Each pair of elements of the slits will produce interference fringes that differ from one another only by a shift in the y direction, nicely overlapping with one another. In this way, we gain a lot in intensity. This was the two-slit experiment by Young.

The upper panel of Fig. 5.10 represents the light intensity in Eq. (5.4) as a function of the sine of the angle, namely of $\sin\,\theta$. This, for large distances and small angles, is proportional to the coordinate x on the screen, as we have already noticed.

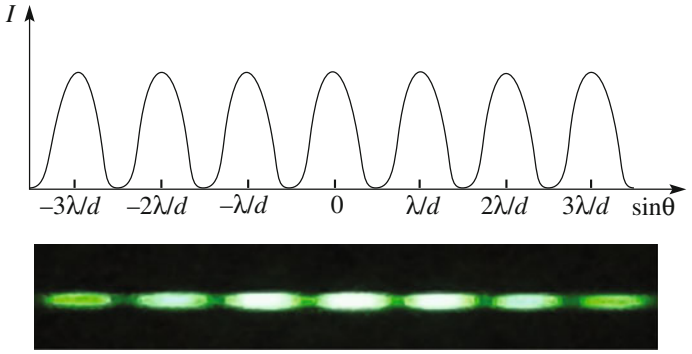


Fig. 5.10 *Upper panel* two-slit interference intensity as in Eq. (5.4). *Lower panel* interference fringes of two-slits of 54 μm width separated by $d = 108$ μm illuminated by laser light with $\lambda = 532$ nm. Photo A. Bettini

Figure 5.10 shows the particular case in which the two sources are in phase ($\phi_2 - \phi_1 = 0$). Under these conditions, the maximum corresponding to the null value of the argument of the cosine, namely, $\phi_2 - \phi_1 - kd \sin\theta = 0$, falls at $x = \sin\theta = 0$. This is called the *central maximum* or, alternatively, the *maximum of order zero*. Notice that the maximum conditions are independent of wavelength for the central maximum.

On the two sides, the light intensity varies periodically as a cosine squared. Observe that the function is quite flat around both the maxima and the minima, while it varies rapidly between a maximum and a minimum. Consequently, the maxima appear as clear fringes separated by dark ones.

The lower panel of Fig. 5.10 is a photo of the two-slit interference pattern obtained with a laser green light, monochromatic at $\lambda = 532$ nm. The width of the slits was 54 μm width and their distance $d = 108$ μm. No lens was used, and the screen was at 8.5 m from the slits. The limited extension of the fringes in the y direction is due to the limited diameter of the laser beam. The reader will notice that the intensity of the maxima decreases on the two sides of the central one. This is due to the finite width of the slits, as we shall explain in Sect. 5.9.

Coming back to the fringe pattern, we obtain its period, namely the distance between two consecutive maxima or minima, remembering that the period of the $\cos^2$ function is π. In the $\sin\theta$ variable, the period is then equal to λ/d and, in x on the lens focal plane (or at the large distance f), is $\lambda f/d$. Note that it is independent of $\phi_2 - \phi_1$. The luminous fringes on the two sides of the central one are found at $x = \pm\lambda f/d$ (first order fringes), $x = \pm 2\lambda f/d$ (second order fringes), and so on.

If $\phi_2 - \phi_1$ has a generic non-zero value, still being fixed by the positions of the slits, the only difference is that the entire fringe pattern rigidly shifts in the x direction. If, for example, it is $\phi_2 - \phi_1 = \pi$, then the central maximum is not at $\theta = x = 0$ but the fringe period is the same as for $\phi_2 - \phi_1 = 0$.

Note that the positions of all the fringes, with the exception of the central one, depend on the wavelength λ. The distance between two contiguous fringes increases when λ increases. If we perform the experiment, as Young did, with white light, we see the central fringe as white, and the lateral ones as composed of different colors, with blue on the internal side and red on the external, because the wavelength of the latter is larger. The width of the colored pattern increases with the order, as understood immediately looking back at Fig. 5.10.

The physical process of observing the interference pattern with our eyes or recording it with an instrument (for example, by taking a photo) requires integrating the light intensity over a certain time interval. This interval depends on the instrument, but is never zero (the exposure time for a photo). We can thus easily understand that the phenomenon is only observable if the phase difference $\phi_2 - \phi_1$ remains constant during the observation time. Otherwise, the interference pattern would move back and forth in the x direction and observation would become impossible. For this reason, if we were to try with two independent, point-like sources, for example, two LEDs (LED means Light Emitting Diode), in place of the two holes S_1 and S_2, we would not observe any interference. Indeed, under these conditions, the phase difference between the two sources varies, taking different values at random over times much shorter than the integration time of our detector. The argument of the $\cos^2$ function in Eq. (5.4) casually takes all possible values and the function averages to ½, independently of the position on the screen. The light intensity on the screen appears uniform, equal to $2i$, namely the sum of the two intensities. We shall come back to this important aspect after having discussed the concept of coherence in the next section.

5.3 Spatial and Temporal Coherence

The two-slit experiment discussed in the previous section is logically a very simple interference experiment. It was historically followed by many more experiments of the same type with increasing levels of accuracy and sophistication up to the present. In any case, a primary light source is necessary. This is the source we called S_0 in the two-slit experiment. We considered it to be point-like and monochromatic. Neither one nor the other of these conditions can be rigorously realized in practice. We shall now analyze the consequences of the fact that sources always have non zero dimensions and are not perfectly monochromatic. More precisely, we shall define the concepts of spatial coherence and temporal coherence of the light source and quantitatively determine the values necessary for a given interference experiment.

Let us start by discussing the geometrical dimensions of the primary source. Our argument will be general, making reference to the Young two-slit experiment for concreteness. If S_0 is extended rather than point-like, we can consider it to be composed of many point sources located at different positions within it. Let D be the diameter of the source and l its mean distance from the holes. For simplicity,

and without compromising the generality of the argument, we assume the center of the source to be on the axis. The phase difference $\phi_2 - \phi_1$ between S_2 and S_1, a result of the light waves coming from the center of S_0, is equal to zero. However, S_1 and S_2 receive light from the other points of the source too, and for each of them, $\phi_2 - \phi_1$ may be somewhat different. Consequently, the phase difference between S_2 and S_1 becomes less and less defined the larger the extension of the source. This results in an increasing confusion of the interference pattern beyond the holes. Indeed, the maxima corresponding to a certain value of $\phi_2 - \phi_1$ may be in the position at which another $\phi_2 - \phi_1$ has an intermediate value or even a minimum. The maxima will not be as intense as in the ideal case and the minima will not be completely dark. To describe the situation quantitatively, we define the concept of *fringe visibility*. Calling $I_{\max}$ and $I_{\min}$ the intensity in the maxima and minima of the interference pattern, respectively, the fringe visibility is defined as

$$\gamma = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \tag{5.5}$$

One sees immediately that the visibility can have values between 1 in the ideal case, in which $I_{\min} = 0$, and 0 when $I_{\max} = I_{\min}$, namely the interference has completely disappeared. Under the latter condition, all the values of $\phi_2 - \phi_1$ between 0 and 2π are present with equal probability. Consequently, the interference term in Eq. (5.4), which is now averaged on all the values of $\phi_2 - \phi_1$, is zero everywhere, and the intensity is simply the sum of the intensities due to the two sources S_1 and S_2.

We understand how the fringe visibility is a measure of the degree of the definition of the phase relationship between the two sources S_2 and S_1 or, in other words, of the phase relationship between the wave field originated by S_0 at the points S_2 and S_1 in the same instant. We express this property by saying that the fringe visibility defines the *degree of spatial coherence* between S_2 and S_1 of the wave emitted by S_0.

Let us now find the largest dimensions that the primary source can have so as to produce an observable interference pattern.

Let us consider two points of S_0, like 1 and 2 in Fig. 5.11, separated in the longitudinal distance along the axis. Both of them are equidistant from the holes, and thus, for both, $\phi_2 - \phi_1$ has the same value (it is zero, but it is the equality that is relevant). We can then extend the source longitudinally as much as we want without changing the phase relation between S_2 and S_1.

We reach the same conclusion extending the source in the direction parallel to the segment joining the holes (namely perpendicularly to the plane of the figure). Indeed, $\phi_2 - \phi_1$ maintains the same value (zero) moving in that direction.

Let us now move in the transverse direction and consider the two farthest points, namely those on the border, like points 3 and 4 in the figure. Their distance is the diameter of the source D. The waves they emit can be considered to be plane when they reach the holes, being that l is always sufficiently large. The angle of the wavefront emitted by 3 with the plane of the holes is approximately $\theta \approx D/(2\,l)$.

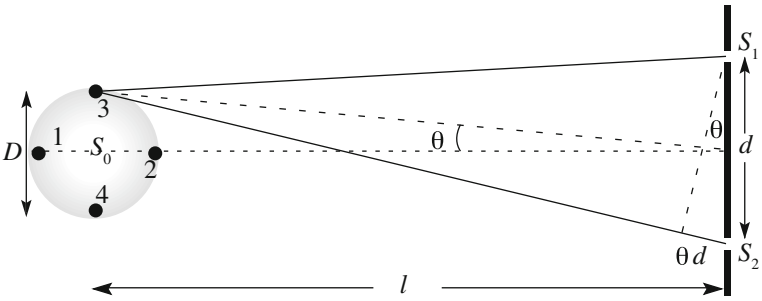


Fig. 5.11 An extended source illuminating a two-hole interferometer

The difference between the paths from 3 to S_2 and to S_1 is thus $d\sin\theta \approx dD/(2\,l)$. The phase difference $\phi_2 - \phi_1$, due to point source 3, is

$$(\phi_2 - \phi_1)_3 = -\frac{kD}{2l} = -\frac{\pi Dd}{\lambda l}.$$

The two secondary sources S_2 and S_1 will then oscillate, when illuminated by point 3, with this constant but not zero phase difference $(\phi_2 - \phi_1)_3$. Similarly, the phase difference between S_2 and S_1, due to point 4 of S_0, say $(\phi_2 - \phi_1)_4$, is

$$(\phi_2 - \phi_1)_4 = +\frac{\pi Dd}{\lambda l}.$$

In the presence of both source 3 and source 4, two interference fringe patterns are simultaneously present on the screen upon which we look for the phenomenon. If D were null, we would have $(\phi_2 - \phi_1)_4 - (\phi_2 - \phi_1)_3 = 0$ and the two fringe patterns would be exactly superposed one over the other; the fringe visibility would be a maximum. Consider now gradually increasing D and observing the fringes on the screen. What do we see? Initially, when D is still small, we practically do not notice any difference. But increasing it further, we see each fringe splitting in two and the distance between the two increasing as D increases. When the positions of the maxima of one interference pattern coincide with those of the minima of the other, the interference disappears. The fringe visibility is zero.

Having discussed the case of two point sources in S_0 separated by the maximum distance D, we now consider that S_0 is composed of a continuous distribution of such points at distances less than or equal to D. Considering that points 3 and 4 give the maximally differing contributions, and that the points of S_0 closer to the axis give contributions that partially sum each other, we can state that D should be smaller than the value for which the contributions of 3 and 4 cancel one another out. If D_{lim} is such a value, this means that for $D = D_{\text{lim}}$, the difference between the extreme phase differences $(\phi_2 - \phi_1)_4 - (\phi_2 - \phi_1)_3$ should be equal to half a period, namely to π. This condition is expressed by the relation

$$D_{\text{lim}} = \frac{l\lambda}{2d}. \tag{5.6}$$

This is the condition on the transverse dimensions of the primary source, not only in the case of the two-slit experiment, but also for every interference experiment. In every interference experiment, there is a set of secondary sources, namely the interference device, receiving light from a primary source. The condition in Eq. (5.6) holds with d being the transverse dimension of the interference device and l its distance from the primary source.

We see that an interference experiment can tolerate larger values of D_{lim} for larger distances of the primary source S_0 to the interference device (the two slits, in the case of the Young experiment). Indeed, the condition for the fringes to be visible depends on the angular diameter of the source, as seen by the interference device. If we call Γ this angle, we have $\Gamma = 2\theta \approx D/l$, and the limit angle is

$$\Gamma_{\text{lim}} = \frac{\lambda}{2d}. \tag{5.7}$$

In conclusion, if the source is seen from the interference device under an angle smaller than Γ_{lim}, the observed interference pattern is indistinguishable from that of a point-like source. We can thus state that the *point-like source* for the considered device is *a source seen under an angle smaller than* Γ_{lim}. Notice also that the conditions we found in Eqs. (5.6) and (5.7) depend on λ. The largest dimensions to be a point sources are smaller for smaller wavelengths.

The general conclusions are as follows. You should never attempt an interference experiment without a design capable of guaranteeing the spatial coherence conditions. In the general case, d is the transverse size of the device that has to produce interference. We can state such a condition in a form easy to remember by heart by rewriting Eq. (5.7) as

$$\frac{l\lambda}{Dd} \leq 2, \tag{5.8}$$

which we read as: *the ratio of the products of the two longitudinal lengths over the two transverse lengths must be less than 2*.

It is worth considering the orders of magnitude. Let us make two holes at the distance $d = 1$ mm and let us operate with light of the typical wavelength of $\lambda = 0.5$ μm. The primary source can be considered point-like if its angular diameter is less than $\Gamma_{\text{lim}} = 0.25$ mrad. If it is located, for example, at $l = 5$ m from the holes, its diameter should be smaller than 1.25 mm. As another example, consider two holes at $d = 1$ cm, and the source at $l = 10$ m from them. Its diameter must be less than a quarter of a millimeter.

Question Q 5.2. In one of his experiments Young worked with a distance between the interfering sources of $d = 2.1$ mm, at a distance from the primary source of $l = 813$ mm. What was the maximum allowed diameter of the primary

source? He writes: "I made a small hole in a window-shutter, and covered it with a piece of thick paper, which I perforated with a fine needle." He observed the fringes on a wall at a distance of $f = 5537$ mm from the interfering sources. What was the distance between the two dark fringes? □

Let us now discuss the consequence of the second limitations, those due to the fact that no source can be perfectly monochromatic. Indeed, the signal emitted by any source has a limited duration in time. Correspondingly, its frequency spectrum has a non-zero bandwidth. This *bandwidth* is inversely proportional to the duration of the signal for the bandwidth theorem. Natural light sources, like the sun and the stars, and the most common artificial ones, like lamps, basically consist of a medium at high temperature. They are collectively called *thermal sources*. A thermal source contains an enormous number of excited atoms that spontaneously decay, oscillating at a number of characteristic frequencies. The frequencies in the visible spectrum are on the order of tenths of PHz. We can select with a filter the light from one of these characteristic oscillations. We can think of it as a sine with a damped amplitude and a total duration, say Δt. This Δt depends on the atomic species, on the phase of matter, on the pressure and temperature of the medium, etc. Its typical values range from picoseconds in a liquid to nanoseconds in a gas. The limited duration has two consequences: (a) the wave has a bandwidth $\Delta\omega$, related to the duration Δt by the bandwidth theorem in Eq. (2.65), namely $\Delta\omega\Delta t = 2\pi$, (b) each of the atoms starts its emission independently of the others, and consequently, the initial phases of their oscillations are randomly distributed. The resulting total wave can be considered to be a chaotic mixture of damped or interrupted sinusoids, having initial phases that are casually distributed. In other words, the memory of the initial phase is maintained only over time intervals of Δt. The duration Δt, within which the information on the initial phase is maintained, is called *coherence time*.

From the operational point of view, the measurement of the coherence time of a source (or of the radiation it produces) is the measurement of the degree of correlation between the electric field emitted by the source in two different instants of time at the same spatial point.

Figure 5.12 shows the concept of an experiment for measuring the temporal coherence. The light produced by the source S is split in two beams by the

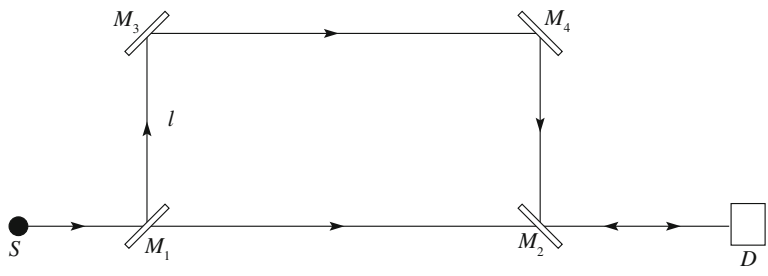


Fig. 5.12 Conceptual configuration for measuring the temporal coherence

semi-reflecting mirror M_1. Semi-reflecting means it reflects half of the incident light, and transmits the other half. Beam number 1 continues in a straight line, while beam number 2 is reflected in sequence by the mirrors M_3 and M_4. The two beams recombine at the semi-reflecting mirror M_2. Call $2l$ the difference between the distances travelled by the two beams. As they have the same phase at M_1, at M_2 they will have a phase difference of $\Delta\phi = 2lk$, where k is the wave number. If such a phase difference is an integer multiple of 2π, the signal at the detector D has maximum intensity, because the crests of one wave arrive at the same time as the crests of the other. On the contrary, if the phase difference is an odd multiple of π, the signal will be minimum, because the crests of one arrive together with the valleys of the other. If we then slowly vary the distance l, the detector will see a series of alternate interference maxima and minima (notice that the experiment is ideal, because the variation of l between a maximum and the next minimum is a quarter of a wavelength, i.e., on the order of a tenth of a micrometer). However, this phenomenon does not go on forever when we increase l; rather, by increasing l, the height of the maxima becomes smaller and smaller, that of the minima larger and larger. Finally, the difference between maxima and minima becomes zero; the interference phenomenon does not exist any more. The reason can be easily understood. When l is small, the two wave trains are overlapped for their entire length, or almost so. For larger values of l, the overlap becomes less and less complete, and consequently the wave trains interfere for only a fraction of their duration. When l is such that the difference between the times of the two paths, which is $2l/c$, is equal to or larger than the coherence time Δt, the succession of maxima and minima ceases to exist.

Figure 5.13 represents schematically the situation approximating a damped sinusoid with a truncated sinusoid of length Δt equal to five periods, for different values of l. In practice, the number of oscillations is much larger, on the order of millions for the light emitted by atoms.

Let us now go back to the interference experiment with two point-like sources. We can now understand why the interference can be observed with two secondary sources, illuminated by the same source, but not with two independent sources. Indeed, let us consider two independent light sources S_1 and S_2. Each of them is made of a large number of excited atoms (they might be two LEDs, for example). In every instant of time, we can think of an atom in S_1 emitting a wave with initial phase ϕ_1, and one atom in S_2 with initial phase ϕ_2. The two emitted waves last for a time on the order of Δt. Within this time interval, the interference between them *does happen*, and on the screen, the luminous and dark fringes are present. Immediately afterwards, however, two other atoms, one in S_1, the other in S_2, would emit light, with a phase difference $\phi_2 - \phi_1$ different from the previous pair. As a consequence, the interference pattern on the screen will move perpendicularly to the fringes.

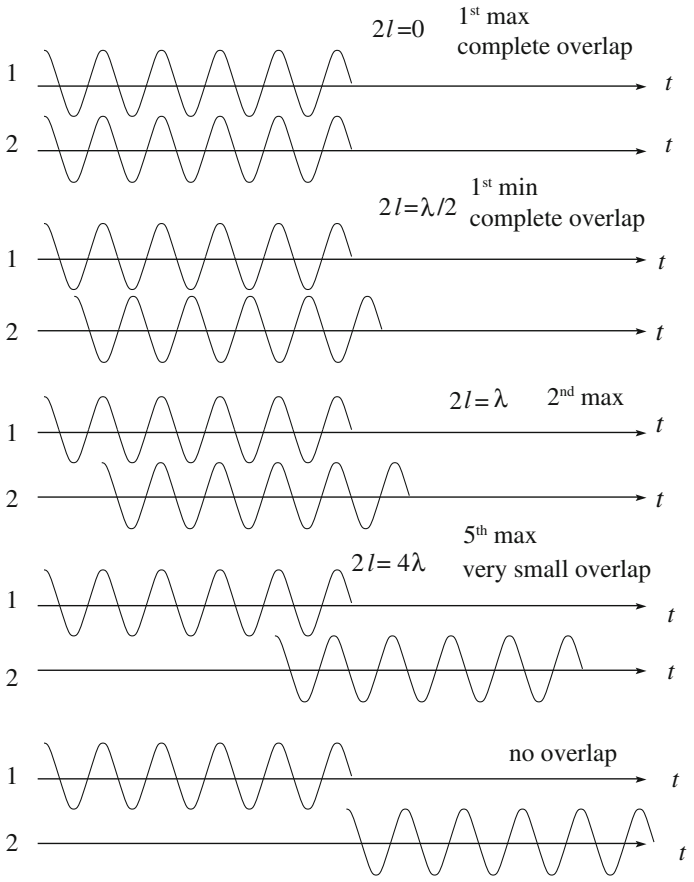


Fig. 5.13 Two waves five wavelengths long for different values of l

The phase difference, and consequently the position of the pattern, remains constant only for a time Δt. If the detector we use to detect the interference integrates over a time that is short compared to Δt, we will observe the phenomenon; contrastingly, if it is much longer, it will produce an average at all the positions of the fringes and the phenomenon will not be observed.

In practice, the coherence times of the common light natural sources are on the order of picoseconds up to nanoseconds, which are much shorter then the integration times of the detectors we usually build. We thus say that independent sources do not interfere. This language, however, is not very precise, because, as we saw, the interference does happen, and whether or not we are able to observe it depends on technology. As a matter of fact, experiments have been conducted in specialized laboratories in which the experimenters were able to observe the interference of two separate light sources.

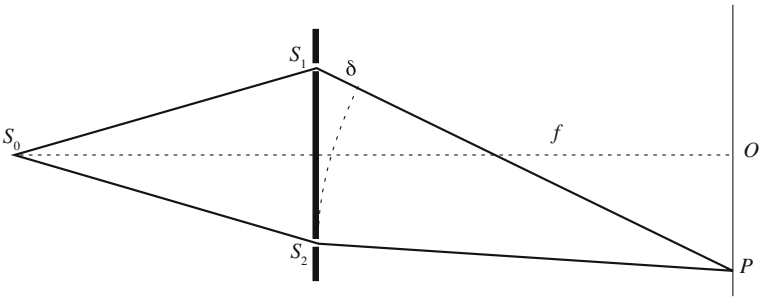


Fig. 5.14 Path difference between two interfering waves in the two-slit experiment

We now go back to the Young holes and see how this experiment allows us, in practice (as opposed to the ideal considered above), to measure the degree of temporal coherence of the primary source S_0 that illuminates the holes. Indeed, its bandwidth determines the number of visible fringes. Let us look at the interference pattern on a screen at the distance f beyond the holes (or in the focal plane of a lens of focal length f). Figure 5.14 shows a scheme of the experiment. We suppose, without compromising the general validity of the argument, that the phases of the two holes are equal, namely that $(\phi_2 - \phi_1) = 0$.

As we have seen, the central fringe is in O, namely the point of the screen at the same distance from the holes, independently of the wavelength. If S_0 is monochromatic with wavelength λ, the fringes of order n are at distances $nf\lambda/d$ from O (on both sides). Notice that the positions depend on λ. If S_0 is not monochromatic, we call ω_0 its mean angular frequency, λ the corresponding wavelength and $\Delta\omega$ the bandwidth (corresponding to an interval $\Delta\lambda$ of wavelengths). We assume $\Delta\omega$ to be small compared to ω_0. We call the ratio $\Delta\omega/\omega_0$ the specific bandwidth. This is small under our hypothesis. The relation between λ and ω_0 is

$$\lambda\omega_0 = 2\pi c. \tag{5.9}$$

Differentiating this expression, we obtain, in absolute values,

$$\left|\frac{\Delta\omega}{\omega_0}\right| = \left|\frac{\Delta\lambda}{\lambda}\right|. \tag{5.10}$$

If the source is not monochromatic, the fringes corresponding to different monochromatic components (colours) fall into different positions, becoming increasingly different with increasing order n. The central fringe is always white, but the other, higher ones show colours (red farther than blue) with separation increasing with n. Increasing n, we come to the point, call $n_{\max}$ the corresponding value of n, at which the clear fringe corresponding to one of the bandwidth limits falls into the same position as the dark fringe of the component at the other bandwidth limit.

If we now take a black and white photo of the screen (not to distinguish the colours, which would effectively reduce the bandwidth), we will see only $n_{\max}$ fringes on each side of the central fringe, namely $2n_{\max} + 1$ fringes in total.

We shall now calculate $n_{\max}$. Consider a generic point P on the screen, as shown in Fig. 5.14. Two light waves meet there, after having travelled the paths S_0S_1P and S_0S_2P, respectively. We see the analogy with the situation in Fig. 5.12. When P moves away from 0, the path difference, call it δ, increases, and we observe maxima or minima depending on whether δ is an even or odd multiple of $\lambda/2$. Between P and O, we have $n = \delta/\lambda$ periods of the fringe system.

Notice now that when δ increases, the difference δ/c between the times taken by the wave on the two paths also increases. If this difference is larger than the coherence time $\Delta t = 2\pi/\Delta\omega$ of S_0, we can no longer observe the interference. This means that, in order for the interference to be observed, δ cannot be larger than the maximum value

$$\delta_{\max} = \frac{c2\pi}{\Delta\omega} = \frac{\lambda\omega_0}{\Delta\omega},$$

where we have used Eq. (5.10). In conclusion, the number of fringes we observe on each side of the white central fringe is

$$n_{\{max} = \frac{\delta_{\max}}{\lambda} = \frac{\omega_0}{\Delta\omega} = \frac{\lambda}{\Delta\lambda}. \tag{5.11}$$

In other words: *the number of observable fringes (on both sides) is equal to the reciprocal of the specific bandwidth.*

If the source S_0 is white, the visible bandwidth is in the wavelengths between 0.38 μm in the violet and 0.78 μm in the red. Hence, $\Delta\lambda = 0.4$ μm and the average wavelength is $\lambda = 0.58$ μm. We should then observe two fringes (multicolor) on each side of the central one only. In practice, one observes four or five of them, for two reasons. First, our eye is not very sensitive to the limits of the spectrum, reducing the effective $\Delta\lambda$ to about 0.2 μm (see Fig. 7.34, which gives the sensitivity of our eye to the different wavelengths); secondly, the perception of the colors helps in distinguishing the fringes even when they are overlapped.

Thermal sources with different degrees of monochromaticity $\lambda/\Delta\lambda$ can be obtained starting from a white source. For example, with a colored filter, we can select a specific bandwidth $\lambda/\Delta\lambda$ on the order of dozens. We can do better selecting a specific spectral line. Indeed, each type of atom or molecule has a series of normal modes of oscillation, each with a certain proper frequency. The spectrum of such molecules contains a series of lines (as they are called because they appear as such when we observe the light decomposed by an interferometer or a prism after having gone through a slit). In condensed matter, molecules are very close together, and the interaction between them does not allow them to oscillate freely, thus the lines are not observable. We can have lines if we operate with a gas at low enough pressures. Even under these conditions, each line is never perfectly monochromatic, but its

bandwidth is at least the one corresponding to the fact that the wave trains emitted by an atom or a molecule have a limited duration (on the order of nanoseconds to fix the ideas). This is called *natural bandwidth* (on the order of the GHz).

There are several effects that produce an increase in the bandwidth of the lines. The principal ones are collisions and thermal motion. A collision during the emission interrupts it, or, in any case, shortens the memory of the initial phase. In a condensed medium, the mean time between collisions is on the order of picoseconds, much shorter than the typical lifetimes. This effect is called collision broadening. We can reduce this effect working at low pressures. As for the second effect, consider that, in their motion, some of the molecules approach the observer, others go away in the opposite direction. The observed frequency depends on the relative velocity between source and observer (Doppler effect, see Sect. 3.13), and the line widens (Doppler broadening).

In conclusion, we can obtain light with a high degree of monochromaticity from thermal sources, using filters, prisms, or systems of slits or interference devices to select a frequency band in which a single line of the source is present, taking care to choose particularly narrow lines (and we shall need to accurately prepare the conditions we have mentioned above). In this way, it is possible to obtain values of $\lambda/\Delta\lambda$ on the order of 1000 with Sodium lamps, 10,000 with Mercury or Helium lamps, 100,000 with Hydrogen lamps and 500,000 with the red line of Cadmium.

LASERs, which are easily available today, belong to another class of light sources, as we already mentioned in Sect. 4.7. LASERs are not thermal sources in which excited atoms de-excite randomly and independently of one another. In a LASER, the de-excitation is by stimulated emission. This is a quantum process that we cannot discuss here. It will be sufficient to know that the entire system of atoms emits energy as a single oscillator in which all the atoms of the medium oscillate with fixed phase relations to one another. Consequently, the degree of monochromaticity of the emitted light is extremely high.

5.4 Interference with Non-coherent Light

The coherence conditions we established in the last section need to be guaranteed to observe an interference phenomenon. For this reason, it is not very common to observe natural interference phenomena, because the natural light sources, like the sun and the atmosphere, are extended and not monochromatic. However, exceptional conditions exist in which interference phenomena are observed in white light and with extended sources. For example, the colors of the elytra of certain beetles and the wings of some butterflies are due to a type of interference phenomenon, as one can understand observing that colors change when we change the angle under which we look at the insect. Similarly, the colors observed on gasoline films on the surface of water are due to interference. In these cases, the light source is the atmosphere, namely a white extended source, and the interference takes place in a thin transparent film. The consequence of the source not being monochromatic is

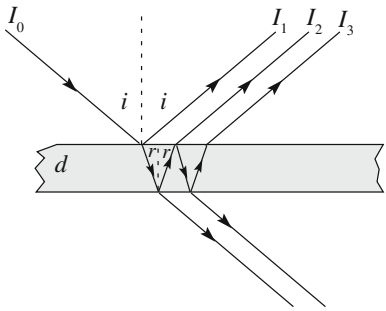


Fig. 5.15 The geometry of the multiple reflections by a parallel face thin film

that the fringes of different colors appear in different positions or under different angles. The consequence of the source being extended is that the fringes are not observable in any arbitrary position in space, but can only be seen in a definite position, to be precise, only on the film. Thus, we talk of *localized fringes*.

We shall now study the interference in a thin film in two geometrically simple cases, namely a film with plane and parallel surfaces and with plane and non-parallel faces (a wedge). The first case is shown in Fig. 5.15

Let us consider a plane wave (hence spatially coherent) incident onto a film of plane and parallel faces at the incident angle *i*. Let d be the width and $n > 1$ the refractive index of the film. The incident ray is partially reflected and partially refracted on the first surface. On the second surface, the refracted ray, in turn, is partially refracted (transmitted ray) and partially reflected back to the first surface, where it is again partially reflected and partially refracted out, and so on. In this way, a set of reflected and a set of transmitted rays is produced. All the rays of one set are mutually coherent, namely the phase differences between the corresponding waves of each frequency have fixed values. Consequently, we have the conditions to observe the interference in both sets of rays.

In practice, the incident angle is usually close to 90° and the reflection coefficient is quite small. Under these conditions, only the very first rays give a perceptible contribution. For example, as we saw in Sect. 4.7, at the air-water interface, the reflection coefficient is about 0.04. Let us consider the reflected rays, namely the case of the examples we have mentioned. Calling I_0 the intensity of the incident ray, the intensity of the first reflected ray is obviously $I_1 = 0.04\ I_0$. The intensity of the second reflected ray, which exits after two refractions and one reflection, is $I_2 = 0.96 \times 0.96 \times 0.04 \times I_0 = 0.037 \times I_0$, which is close to I_1. But the intensity of the third reflected ray is already much smaller, namely $I_3 = (0.96)^2 \times (0.04)^3 \times I_0$, which is 1/625 of I_2. In conclusion, then, if the reflection coefficient is small, the interference phenomenon is practically only due to the first two reflected rays.

On the other hand, no interference is perceptible for the transmitted rays if the reflection coefficient is small, because intensities are all very different from one another. Indeed, under the just-considered conditions, the intensity of the second ray is 1/625 of the first, the intensity of the third 1/625 of the second, and so on.

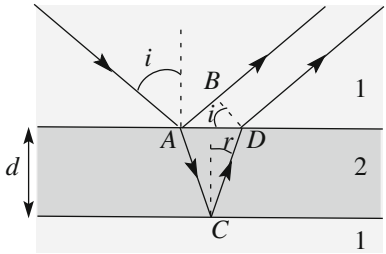


Fig. 5.16 Interference of the first two reflected rays by a thin parallel face film

Consider a monochromatic component of the incident plane wave. Under the conditions we have discussed, the interference is determined by the phase difference between the first and the second reflected waves. The corresponding rays are shown in Fig. 5.16. We assume the first medium to be air, whose index we take to be equal to 1. Let the index of the second medium be $n > 1$.

Consider the wavefront BD (normal to both rays, obviously) immediately outside the film. The paths the two rays must take to reach it are different. The first ray reflects off the first surface and then travels the distance AB through the air. Looking at the figure, we see that this distance is $AB = d \tan r \sin i$. If k is the wave number in air (or a vacuum), the corresponding phase delay is $\Delta\phi_1 = k2d \tan r \sin i$. The length of the path ACD of the second ray is $2AC = 2d/\cos r$. Being that this path is entirely in the medium, in which the wave number is nk, the corresponding phase delay is $\Delta\phi_2 = nk2d/\cos r$. The phase difference between the two rays due to their different paths is, in conclusion,

$$\Delta = \Delta\phi_2 - \Delta\phi_1 = \left(\frac{n2d}{\cos i} - 2d \tan r \cos i\right)k.$$

We can simplify this expression using the Snell law $\sin i = n\sin r$. We obtain

$$\Delta = \left(\frac{2d}{\cos r} - 2d\ \frac{\sin^2 r}{\cos r}\right)nk = nk2d \cos r.$$

However, we have not yet finished, because there is a further contribution to the phase difference between the two reflected waves. We recall that in Sect. 4.7, we saw that, at the interface between a less refracting and more refracting media, the phase of a wave abruptly changes by π, while it does not vary in the opposite case. In our case, we have a change of π only in the reflection off the first face (namely for the first ray), and in conclusion, the phase difference between the two reflected waves is

$$\delta = nk2d \cos r + \pi, \tag{5.12}$$

Let us now see where maxima and minima shall appear. The maxima will do so where the phase difference is an integer multiple of 2π, namely for $\delta = m2\pi$, with

$m = 0, 1, 2,\ldots$. The condition for destructive interference is $\delta = (2\ m + 1)\pi$, where we have the minima. These conditions can be written in terms of the vacuum wavelength λ_0 as

$$\begin{array}{lll} 2dn\cos r = (m+1/2)\lambda_0, & m = 1,2,3,\ldots; & \text{maxima} \\ 2dn\cos r = m\lambda_0, & m = 1,2,3,\ldots; & \text{minima.} \end{array} \tag{5.13}$$

Let us now analyze a few phenomena observed in the thin films, on the basis of the result just found.

Let us first assume the incident angle to be fixed (and the refraction angle consequently to be fixed as well). It often happens that the thickness of the film is not constant, but varies a little, with variations on the order of the wavelength from point to point. Under these conditions, the loci of constructive and destructive interference are the loci of equal thickness. In monochromatic light, we shall see luminous curves alternated with dark ones. In white light, we see iridescent curves of all the colors. These are called equal thickness fringes. This is what we observe on the wings and elytra of insects and on the gasoline films on water surfaces. The colors are beautiful and change with the relative position of the observer.

We must, however, ask ourselves how the assumed conditions of a fixed incident angle are realized. Indeed, this is the condition necessary to observe interference in this case. Let us consider the simplest case of variable thickness, namely a film with plane surfaces forming a (small) angle. Figure 5.17 shows the geometry we are going to discuss. Such a wedge can be easily made using a rectangular frame made of a thin metal wire terminating in a handle. If you dip the frame into soapy water and then raise it smoothly, a soapy film will form inside the frame. If you lay the frame vertically, the soapy film gets thicker in its lower parts, under the action of the weight. The equilibrium configuration is just a sledge, as one can deduce by seeing horizontal equispaced straight interference fringes on the film.

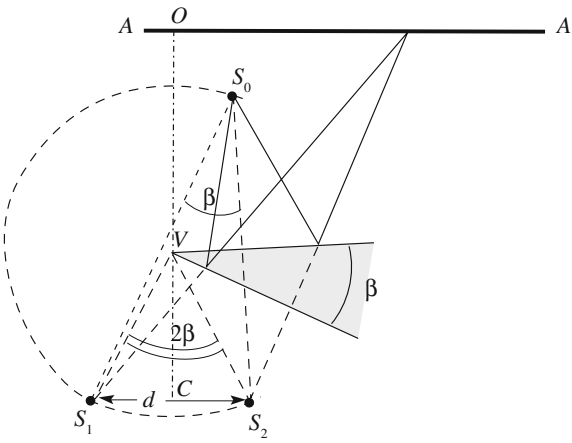


Fig. 5.17 Geometry of the localized fringes

Let us start by assuming the source illuminating the film to be point-like and let us call it S_0. We can think of the two interfering waves as originating from two virtual sources, S_1 and S_2, which are the images of S_0 resulting from the two surfaces of the film acting as mirrors. From this point of view, the interference is completely analogous to the Young two-hole experiment. We conclude that the interference fringes are hyperboloids, having foci in S_1 and S_2. In order to have the same geometry as in the Young experiment, consider now a screen, like the one spanning AA in the figure, parallel to the segment S_1S_2 and to the edge V of the wedge. Let d be the distance between S_1 and S_2, C the center of the segment S_1S_2, O the foot on the screen of the perpendicular from C to the vertex V and l the distance from C to O. Clearly, O is the center of the interference pattern on AA, because it is equidistant from S_1 and S_2.

Let us now analyze how the situation changes if S_0 is extended, which is what happens in practice when S_0 is the sun or the sky. To understand that, suppose keeping S_0 point-like and moving it in a direction parallel to S_1S_2. The images S_1 and S_2 shall move as well, the fringes on the screen AA shall translate perpendicularly to their direction, and their period shall change as well. Indeed, the distance between two contiguous fringes is $p = l\lambda/d$. The distance d between S_1 and S_2 depends on the position of S_0. Let us find out how. If β is the dihedral angle of the sledge, then the angle under which V sees the two secondary sources is 2β. Being that 2β is always very small, calling c the length VC, we can write $d = 2\beta c$. In conclusion, the period of the fringes is

$$p = \lambda l/(2c\beta).$$

This means that the fringes cannot be observed on an arbitrarily-located screen AA when the primary source S_0 is extended. Let us, however, move AA so that it contains the vertex V of the sledge. In practice, we put AA on the sledge. Let us now displace S_0 as we did before. The lateral displacement of the fringe pattern does not happen now, because its center must be in V anyway. In addition, the distance between fringes p does not vary, because it is now $c = l$, which is independent of the position of the point-source S_0. As a consequence, when S_0 is extended, the fringe patterns on the sledge produced by all its points are superimposed on one another. In conclusion, the interference fringes are observable on the sledge, but not in other positions. We say that the fringes are localized on the sledge.

We are now ready to go back and discuss the phenomena of fringes of equal thickness. We shall consider the usual circumstances in which the light incidence on the film is normal or almost so. We can then take $\cos r \approx 1$. According to Eq. (5.13), the loci of the maxima are where the thickness d is such that $2dn = m\lambda_0 + \lambda_0/2$, where, we will recall, λ_0 is the vacuum wavelength, n is the refractive index and m is an integer number. The difference in thickness between two contiguous fringes is then $\lambda_0/2d$. We see that thickness differences on the order of several nanometers can be evaluated.

In some cases, for example, when the film is a sledge, the thickness may be very small at some points. Indeed, the interference is destructive, and we see a dark fringe everywhere the thickness is $d < \lambda_0/4n$. Moving from those regions, the first clear fringe appears at $d = \lambda_0/4n$. Under these conditions, we can not only measure changes in the thickness, but the thickness itself, simply by counting the clear fringes extending out from the first, dark one.

The interference phenomena have a number of applications in optics, allowing us to appreciate thickness differences on the order of one hundredth of a wavelength, namely a few nanometers, as we have already stated. We shall only hint at a few of them here. A way to control the quality of the surfaces of a lens, for example, is to look at the equally thick fringes in a layer of air between the lens itself and a glass reference surface known to be perfectly plane. The surface of the lens under control being spherical, the fringes should be a set of concentric circles. If, on the contrary, a small anomaly were to be present at some point, say a small bump or a small valley, the fringes would show that. Once the irregularity is detected, it can usually be corrected, and the control then done again.

Similar methods are used to accurately measure the dilatation coefficient of materials with varying temperature or pressure and, again, for measuring the refractive index.

Interference methods have been and are of utmost importance in all research fields. Let us simply think of the measurements of the atomic spectra.

5.5 Diffraction

Diffraction exists whenever a propagating wave meets an obstacle in its path limiting its extension by blocking or absorbing part of the front. After the obstacle, the trajectories of each small segment of the front no longer necessarily move in the original direction. This phenomenon is present for every type of wave. The mathematical descriptions are similar, but the physical aspects may be very different from case to case. When the size of the obstacle is on the same order of magnitude as the wavelength, diffraction is most evident. This is the case, for example, for surface waves on the sea. In the area just beyond a strait or the mouth of a harbor, the wave motion does not simply project geometrically forward out of the aperture, but rather invades regions of "geometrical shadow" as well. Figure 5.18 shows, as an example, the diffraction of sea waves through the natural opening of Lulworth Cove in the UK. Similarly, walking down a street, we hear the music produced by a source in a nearby house through an open window above our heads, even when the source of the sound is out of view inside a room. Indeed, the wavelength of the sound waves is on the order of meters, comparable with the objects of everyday life and an approximation with "geometrical acoustics" is not useful.

We shall deal here with the diffraction of light, which gives origin to phenomena that are, even in their largest numbers, not easily observed, or recognized as such, due to the very small wavelength. Indeed, the approximation of geometrical optics

Fig. 5.18 Diffraction of sea waves in the Lulworth Cove, Dorset, UK. Photo by Chris Button Photography

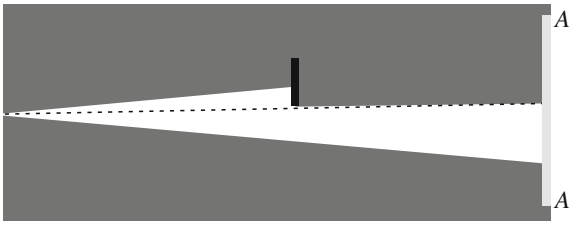


Fig. 5.19 Geometrical shadow

is very often adequate. As we have already mentioned, the diffraction of light was discovered by Francesco Maria Grimaldi in the XVII century. He made his observations in a room that was completely dark but for a small hole that he had opened in the shutter of a window. He placed obstacles of different shapes in the path of the light beam entering the room and looked carefully at their shadows on a white paper, paying attention, in particular, to the transition border between shadow and light. Imagine introducing the simplest obstacle into the path of a light beam, a sheet of paper with a straight border, namely a semi-infinite obstacle. Figure 5.19 shows what we should expect from geometric optics. On the screen, at *AA*, we should observe a white region and a dark region, separated by a sharp line, as long as penumbra situations are avoided.

Let us analyze this phenomenon on the basis of the Huygens-Fresnel principle, as in Fig. 5.20. Consider a wavefront, which we take to be a plane for simplicity's

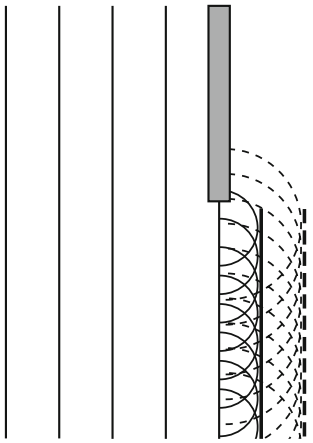


Fig. 5.20 Huygens-Fresnel construction after a semi-infinite obstacle

sake, reaching the obstacle at time t. We geometrically construct the wavefront beyond the obstacle at time $t + \Delta t$, considering a number of elementary secondary sources on the plane of the obstacle and the spherical wavelets they emit, as shown in the figure. The figure shows the wavelets as dotted lines at $t + \Delta t$ as well.

The important point is that the relation between the phases of all the secondary wavelets at a certain time is fixed. Consequently, the infinite hemispherical wavelets interfere where they meet. The resulting wave amplitude is the largest in the geometrical light region, but is not zero in the geometrical shadow. Indeed, the wave gets around the obstacle and widens more and more in the "shadow" as it gets farther from the obstacle. Considering that the phenomenon is due to the interference of the wavelets, we can imagine that colors may be observed both in the geometrical shadow and in the geometrical light, near the border between them. This is exactly what happens, as we shall now see.

Following Grimaldi, we can say that diffraction of light occurs whenever light does not propagates rectilinearly for phenomena different from reflection and refraction. As we have discussed, diffraction is just an interference phenomenon of the infinite secondary Huygens-Fresnel wavelets on a wavefront and differs only for the number of secondary sources from the two-slit Young experiment. In that case, we had two of them, while we now have an infinite number. In conclusion, the terms 'interference' and 'diffraction' are almost synonyms, the difference being in the number of interfering sources.

The diffraction phenomenon, even in its simple form, as in Fig. 5.20, is not easy to observe with natural light for several reasons. First, the hole in the window must be small, no larger than a few millimeters, to guarantee spatial coherence,, and consequently, the beam intensity is not high, only a few milliwatt. Second, the colored fringes are very near to the shadow limit and to one another. For example, in a typical configuration with, say, 3 m distance between the primacy source and the obstacle and 3 m from the obstacle to the observation screen as well, the distance between the first and second minima is about 1.3 mm for blue and 1.0 mm

for red. Contrastingly, the basic diffraction and interference phenomena considered in this book are very easily observed with any laser beam, due to its very high level of both temporal and spatial coherence.

We translate here some of Grimaldi's text on the fringes he observed in the geometrical light region.

> In the illuminated part of the base (*namely the white paper used for observations*) a few starches, or fringes, of colored light spread and separate, so that in each fringe the light is very pure, and sincere, in the middle, while at its extremes is colored, namely blue on the side closer to the shadow and red on the farther extreme. …The first (fringe) is wider than the second and this one is wider than the third (and it never happens that more than three are seen) and have a decreasing intensity of light and colors, in the same order as they get farther from shadow.

Figure 5.21 shows what he should have seen.
He continues describing the clear fringes in the shadow:

> It should also be observed that these colored light fringes appear sometimes in the shadow itself. Their number is sometimes larger, sometime smaller….

As we already mentioned, Grimaldi did not interpret his observations as a proof of the wave nature of light, neither did he perform systematic measurements. Both of these must be credited to Thomas Young, who worked on interference between 1804 and 1807, as we already discussed, and to Augustin Jean Fresnel, who completely clarified the diffraction phenomena with experiments and theoretical interpretation between 1815 and 1827.

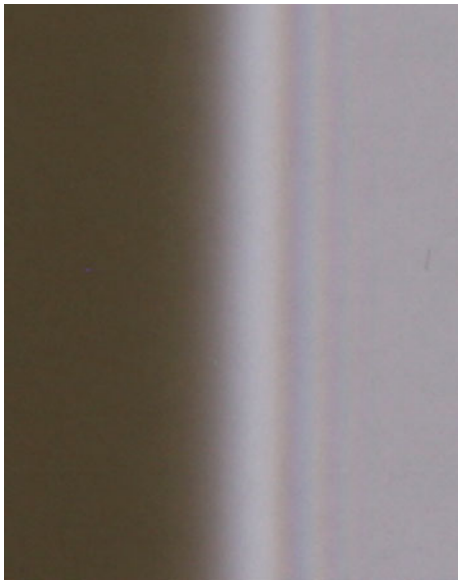


Fig. 5.21 White light diffraction fringes in the geometrical shadow after a semi-infinite obstacle. Photo C. Braggio

We note here that the wave field diffracted by an obstacle depends, in general, on both the distance from the obstacle and the considered direction. Its general description requires rather advanced mathematics, and we shall not deal with it in this book. As we already mentioned, in discussing the two-slit experiment, the situations are simpler in a far field, namely at a large distance beyond the obstacle, where the wave field depends on the directions alone. These are the Fraunhofer conditions, which we will continue to discuss throughout the book (with the exception of Chap. 8). The more general ones, in a near field, are called Fresnel conditions. In the next section, we shall define the Fraunhofer conditions in exact terms.

5.6 Diffraction by a Slit

In this section, we quantitatively analyze the diffraction of light by a slit under the Fraunhofer conditions. This is a simple and practically very important case, and it will give us the opportunity to show the principal characteristics of diffraction. We consider a monochromatic plane wave of infinite extension normally incident on an absorbing screen in which a slit is open. We take a reference frame with origin in the center of the slit, the y-axis in the direction of the slit and the x-axis perpendicular. Let D be the width of the slit (in the x direction) and let be it infinitely extended in the y direction. Notice that, in general, the shape of an opening in a screen depends on two coordinates. The simple case we are considering depends on x alone. The screen is completely transparent for $-D/2 \leq x \leq D/2$, completely absorbing everywhere else.

Fraunhofer conditions are satisfied when the diffraction pattern is observed at an infinite distance or in the back focal plane of a converging lens, in other words, when the propagation directions of the diffracted waves are *parallel* to one another. In practice, the conditions are satisfied, even without using a lens, provided the phenomenon is observed at a large enough distance L from the diffracting element. To understand what "large enough" means, let us consider the arrangement shown in Fig. 5.2, in which the diffraction pattern is viewed on a screen at a "large" distance L from the slit as a function of the coordinate ξ parallel to the x-axis. The Fraunhofer condition is satisfied if the *phases* of the Huygens-Fresnel waves that originated in different parts of the slit are almost equal to one another at all points of the diffraction pattern, like P in the figure. We can consider the phase differences small enough if they are much smaller than 2π. This means that the path difference δL to P from the extremes of the slit, which is the largest path difference, must be much smaller than the wavelength λ. Let θ be the angle under which the center O of the slit is seen from P. This angle being small, we consider it to be infinitesimal, and we approximate at the first order in θ, namely we neglect terms in θ^2 and higher. In other words, we take $\tan\theta \cong \sin\theta \cong \theta$. At the same order, the length of the segment AA in the figure, which is perpendicular to OP, is equal to D.

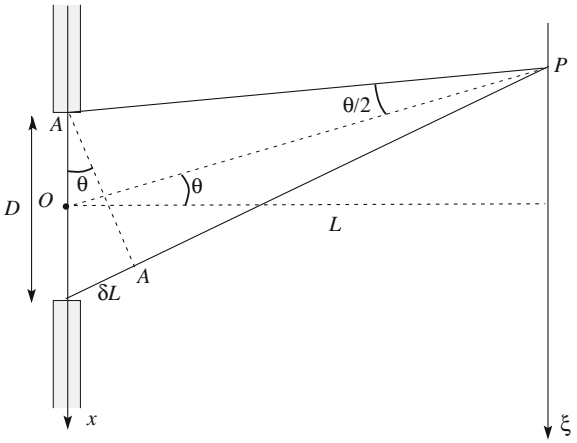


Fig. 5.22 Geometry for the diffraction of a slit at a finite distance L

Looking at Fig. 5.22, we see that $\delta L = D \sin\theta \approx D\theta$ and that $\theta/2 \approx D/(2L)$, and we can write $\delta L \approx D^2/L$. The Fraunhofer condition is then $D^2/L << \lambda$, or

$$L >> D^2/\lambda. \tag{5.14}$$

For example, consider the typical values $D = 1$ mm and $\lambda = 0.5$ μm. The Fraunhofer conditions are satisfied if we look at the diffraction, without using a lens, at a distance L of a few meters. If Eq. (5.14) is not satisfied, we are under Fresnel, or near field, conditions.

Let us now discuss the diffraction by the slit, as shown in Fig. 5.23, in which, having assumed the Fraunhofer conditions to be satisfied, we have drawn the propagation directions of the elements of the diffracted wave as being parallel.

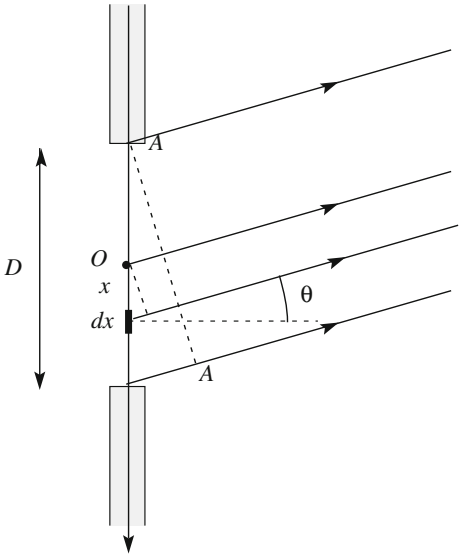


Fig. 5.23 Geometry for the Fraunhofer diffraction of a slit

Let us apply the Huygens-Fresnel principle to the wave surface across the slit. We consider it to be a plane of secondary sources, oscillating with the same phase emitting hemispherical waves in the semi-space beyond the slit. The light intensity at point P at an infinite distance can depend only on the angle relative to the incoming direction (assumed to be perpendicular to the slit), which we call θ, under which the point is seen by the slit. Indeed, at infinite distance, this angle is the same for all the points of the slit.

Our calculation of the diffraction pattern as a function of θ shall proceed with the following steps: (1) divide the wave surface at the slit into infinitesimal elements, which we take to be infinitely extended in the y direction and at a length of dx in the x direction; (2) consider the field emitted by each of these elements in the θ direction; (3) take their sum (integral), (4) take the square of the sum; (5) take the mean value over a period of the result.

The secondary sources emit waves of the same amplitude, say adx, because they have equal areas proportional to dx, and with the same initial phase, because they are all on the same wavefront. The secondary waves, however, will have phases different from one another when they reach the observation point, because they must travel different path lengths to reach it. Now, the distances between any two planes perpendicular to the considered direction are equal. Consequently, the path differences to be considered are between the secondary sources and a plane like AA in the figure.

Taking as a reference the wavelet emitted at $x = 0$, consider the wavelet emitted by the generic element between x and $x + dx$. The path length that the latter must travel is longer by $x \sin\theta$ than that of the former, and consequently, the phase of the latter will lag behind the phase of the former by $kx\sin\theta$. Hence, the field emitted by the element at infinite distance in the direction θ is

$$dE = \mathrm{e}^{i(\omega t - kx\sin\theta)} adx.$$

We obtain the total field, resulting from all the slit elements, by integration on the slit width. We obtain

$$\begin{aligned} E(\theta) &= \mathrm{e}^{i\omega t} \int_{-D/2}^{+D/2} \mathrm{e}^{-ikx\sin\theta} dx = -\frac{a\mathrm{e}^{i\omega t}}{ik\sin\theta}\left(\mathrm{e}^{-ik\frac{D}{2}\sin\theta} - \mathrm{e}^{+ik\frac{D}{2}\sin\theta}\right) \\ &= Da\mathrm{e}^{i\omega t}\left[\frac{\sin\left(k\frac{D}{2}\sin\theta\right)}{k\frac{D}{2}\sin\theta}\right]. \end{aligned}$$

It is useful to introduce the quantity

$$\Phi = k\frac{D}{2}\sin\theta = \frac{\pi D}{\lambda}\sin\theta. \tag{5.15}$$

The physical meaning of Φ is the phase difference at infinite distance in direction θ between the contribution of the center and that of a border of the slit. Namely, it is one half of the maximum phase difference between contributions.

Taking the real part of the above equation, we have

$$E(\theta) = aD\frac{\sin\Phi}{\Phi}\cos(\omega t). \tag{5.16}$$

To obtain the intensity, we now take the mean value over a period of the square of the field, obtaining

$$I(\theta) = \frac{a^2D^2}{2}\frac{\sin^2\Phi}{\Phi^2}.$$

We immediately recognize the physical meaning of the factor $a^2D^2/2$. It is clearly the maximum intensity, which is found in the forward direction, namely for $\Phi = \theta = 0$. Calling the maximum intensity $I_{\max} = a^2D^2/2$, we have, in conclusion,

$$I(\theta) = I_{\max}\frac{\sin^2\Phi}{\Phi^2}. \tag{5.17}$$

In practice, we observe the diffraction pattern on a white screen at a large distance L beyond the slit (or in the focal plane of a converging lens of focal length L). Note that, in practice, D should be on the order of a millimeter or less in a laboratory experiment, in order to have the condition $L \gg D$ satisfied at distances on the order of meters. Let ξ be the coordinate on the observation screen parallel to x and with origin at the point corresponding to $\theta = 0$. The light intensity pattern we observe on the screen is proportional to I, as a function of ξ.

In the majority of the situations, we are interested in small values of θ. Indeed, as we shall now see, the intensity becomes very small at large angles. Under these conditions, we can take approximately

$$\sin\theta \approx \xi/L, \tag{5.18}$$

corresponding to the approximate expression for Φ

$$\Phi \approx \pi\frac{D}{\lambda}\frac{\xi}{L}. \tag{5.19}$$

Figure 5.24a shows the function $\sin^2\Phi/\Phi^2$ as a function of Φ. Being that Φ is proportional to ξ under the approximation of Eq. (5.19), the same curve also represents the intensity as a function of ξ. Figure 5.24b shows the intensity on the screen. Indeed, we see a high central maximum, and sequences of alternated maxima and minima symmetrically on the two sides. Note that in Fig. 5.24b, we have exaggerated the intensities outside the central peak to make them more visible.

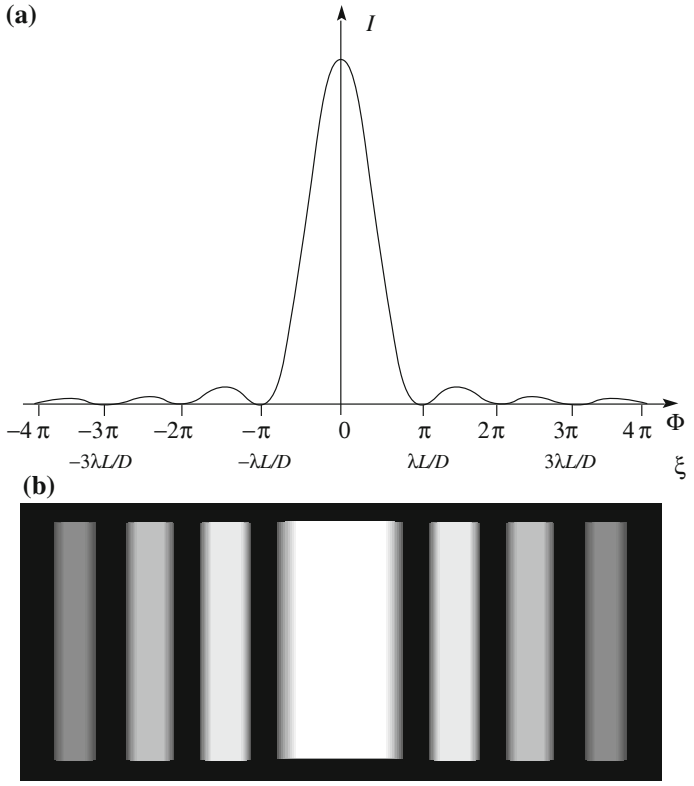


Fig. 5.24 The diffraction figure of a slit under Fraunhofer conditions. **a** The intensity curve, **b** the diffraction fringes (the central fringe is over-exposed to enhance the visibility of the lateral ones)

The positions of the minima, in which the intensity is zero, are easily found, because it must be $\Phi = \pm m\pi$, with an integer of m, at those positions. Hence, in the small angles approximation, the first zero is at

$$\xi_{\min} \approx \frac{L\lambda}{D}. \tag{5.20}$$

As immediately checked, this is also the distance between two contiguous minima.

The lateral maxima of the diffraction pattern, rigorously speaking, are the maxima of the function $\sin^2 \Phi/\Phi^2$. In practice, however, the denominator of this expression varies very slowly compared with the numerator for varying ξ (or, equivalently, for varying θ). We can then safely look for the maxima of $\sin^2 \Phi$. Their positions are for $\Phi \approx (m + 1/2)\pi$, with $m = 1, 2, \ldots$. The intensity in the secondary maxima is much smaller than that in the principal one. For example, for the first one, we have

Fig. 5.25 Photo of the diffraction pattern of a slit under Fraunhofer conditions with monochromatic laser light at $\lambda = 532$ nm. The width of the slit was 50 μm. No lens was used and the screen was at 2.45 m from the slit. Photo by A. Bettini

$$I_1 = I_{\max} \frac{1}{(3\pi/2)^2} \approx 0.04 I_{\max}, \tag{5.21}$$

Indeed, the vast majority of the diffracted light is in the central peak.

Question Q 5.3. Calculate the intensities of the second and third maxima □.

Figure 5.25 is a photograph of the diffraction pattern on a screen at 2.45 m from a 50 μm wide slit with monochromatic light of a laser of $\lambda = 532$ nm wavelength. The central maximum is overexposed to make the secondary maxima more visible.

Let us now discuss the quantity $\xi_{\min}$, which is one half of the lateral extension of the main part of the diffraction pattern. Equation (5.20) contains four linear dimensions, namely $\xi_{\min}$, L, D and λ. The last quantity, namely the wavelength, depends on the radiation we are employing. If we limit our considerations to light and to the orders of magnitude, λ is substantially fixed, on the order of 0.5 μm. Let us then consider how the other lengths play.

If we are in a laboratory, the distance between the slit and the screen shall always be on the order of a meter; let us take the typical value $L = 2$ m. Under these conditions, if we want to be able to easily observe the diffraction pattern, we will need to have $\xi_{\min} = 1$ mm or larger. For that, we need a slit of width $D = 1$ mm or less (see the conditions in Fig. 5.25, for example).

This does not mean that diffraction patterns of wider slits, even much wider ones, are not observable. Indeed, we can observe them by moving to larger distances or, alternatively, by enlarging the diffraction pattern with a system of lenses. We can, for example, observe the diffraction pattern of a slit so wide as to have $D = 5$ cm, by observing at a distance $L = 100$ m, where we have $\xi_{\min} = 1$ mm.

Consider now the following observations that can be done with a slit whose width D can be varied by acting on a micrometric screw. We observe, on a screen at a large enough fixed distance L, how the diffraction pattern varies when we vary D. If the slit is initially quite open, then $\xi_{\min}$ is very small, and we do not see the diffraction pattern. When we gradually close the slit, we observe the diffraction pattern widening, as $\xi_{\min}$ increases. When we have reduced the width to the order of the wavelength, the slit will practically have become a single secondary source of negligible transverse dimensions, because its element radiates in phase one with another in any direction. In practice, all the directions of observation are equivalent. The source radiates light in all directions beyond the screen.

Fig. 5.26 Diffraction between two fingers. Photo by A. Bettini

Up to now, we have considered the diffraction of a monochromatic light. What happens if it is polychromatic? Under these conditions, each monochromatic component produces a pattern on our screen similar to that in Fig. 5.24. The position of the central maximum is independent of the wavelength and is consequently the same for all the components. If light is white, the central fringe is white. Contrastingly, the positions of the first and subsequent minima, and those of the secondary maxima, depend on λ. Consequently, the lateral fringes appear white in their centers, where all colors contribute, and colored on their borders, with red (which has the largest wavelength) at the external part, blue on the internal. This is what Grimaldi described. Notice, finally, that the distance between the red and blue parts of a fringe increases with the order of the fringe itself.

Before concluding the section, let us mention how the effect of diffraction through a slit can be observed without any instrument at all. You merely have to put the fingertips of your thumb and forefinger very close to one another and look through the gap between them at a clear sky. When the fingertips are almost touching, you will see a bridge suddenly appearing to join them, as shown in Fig. 5.26. Carefully adjusting the distance between fingertips, you will see the bridge splitting into a few darker and a few lighter fringes parallel to the tips. The light source being a diffuse one, the fringes are localized.

5.7 Diffraction by a Circular Aperture

When a light wave encounters an opaque screen with an aperture of any shape, diffraction takes place with characteristics similar to those of a slit. The diffraction pattern depends on the shape of the aperture, but the general laws governing the phenomenon are the same. As already mentioned, they were laid down by Augustin Jean Fresnel between 1815 and 1827. In this section, we shall discuss a second case, beyond the slit, namely the circular aperture. Its importance can be appreciated by thinking about the fact that the majority of optical elements, such as lenses, mirrors, diaphragms, filters, etc., are circular, namely they limit the incident wave allowing only a circular part of it to go through.

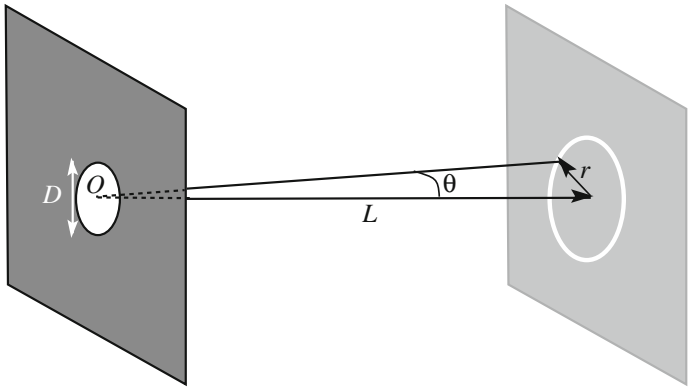


Fig. 5.27 Geometry of the Fraunhofer diffraction from a circular aperture

Consider again a monochromatic plane wave incident in the direction perpendicular to the aperture, as shown in Fig. 5.27. Let D be the diameter of the circle and L the distance from the aperture to the screen on which we look at the diffraction pattern, under Fraunhofer conditions. In this case, the geometry of the problem is cylindrical about an axis perpendicular to the aperture (which is also the incident direction) through its center O. We choose a polar cylindrical reference, as shown in the figure, with origin in the center of the aperture and the z axis as the symmetry axis. In general, the diffraction pattern depends on the distance from the origin and on the polar angle θ, but is independent of the azimuth as a consequence of the symmetry of the problem. Under Fraunhofer conditions, the pattern depends only on θ, namely, on the screen, it is a function of the distance from that axis, which we call r. Let this function be $I(r)$.

The arguments leading to the diffraction pattern are similar to those we developed for the slit in the previous section. The same five steps shall be followed. However, in this case, the integral encountered when summing up the contributions of the secondary sources on the incoming wavefront, namely step 3, cannot be expressed in terms of any simple function like sines and cosines. Indeed, the integral is a higher transcendental function, known as a Bessel function, named after Friederch Bessel (Germany, 1784–1846).

The function $I(r)$ is shown in Fig. 5.28. It is similar to $I(x)$ for the slit in Fig. 5.24a, taking into account the cylindrical geometry. The diffraction pattern features a bright central disc, called the diffraction disc or the Airy disc, containing the largest fraction of the light. It is surrounded by a succession of alternated dark and clear rings. The intensities of the rings are much lesser than that of the central maximum, and decrease with increasing order. The first theoretical description of the phenomenon was given by George Airy (UK, 1801–1892) in 1835. Consequently, the diffraction pattern under Fraunhofer conditions of a circular aperture and its central disc are also called the Airy pattern and the Airy disc, as we just mentioned.

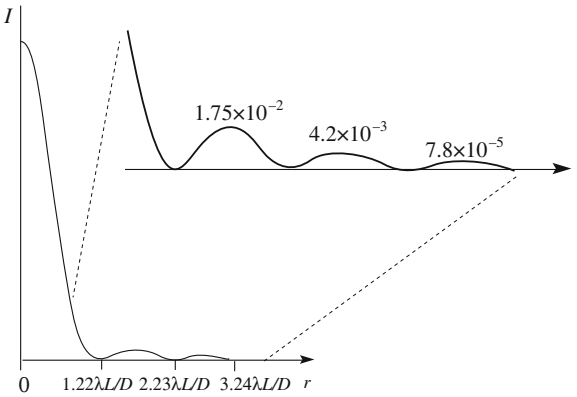


Fig. 5.28 Intensity of diffracted light by a circular aperture under Fraunhofer conditions. In the insert, the first three secondary maxima numbers are their intensities relative to the principal maximum

There are quantitative differences relative to the slit. In particular, the distance from the center of the first dark ring, which is the first minimum, namely the radius of the Airy disc, is not equal to $L\lambda/D$ as for the slit, but to

$$\xi_{\min} \approx 1.22 \frac{L\lambda}{D}. \tag{5.22}$$

The strange-looking factor 1.22 corresponds to the position of the first zero of the Bessel function. Figure 5.28 shows the positions of the second and third minima as well.

We can also say that the Airy disc is seen from the center of the aperture under the radius

$$\theta_{\min} \approx 1.22 \frac{\lambda}{D}. \tag{5.23}$$

The heights of the secondary maxima are extremely small, even smaller than for the slit. Figure 5.28 reports, in the insert, the heights of the first secondary maxima relative to the central one.

Figure 5.29 shows a reproduction of the diffraction pattern of a circular aperture under Fraunhofer conditions. The images are credited to Rik Littlefield, Zerene Systems LLC, obtained with a white LED shining through a pinhole at a distance of about 5 m from the pinhole. In the right panel, the Airy disc has been given the correct exposure. In the left panel, the exposure was twelve times longer, to make the rings of the secondary maxima visible (and super-exposing the central disc). In white light, the central disc is white, while the rings are colored on their borders, with red on the external part, blue on the internal.

Given its importance, the diffraction pattern in monochromatic light by a circular aperture at a large distance is called the fundamental diffraction pattern.

You can observe this diffraction pattern even in white light in the following way. Make a round small hole in a piece of cardboard with the tip of a pin, with a

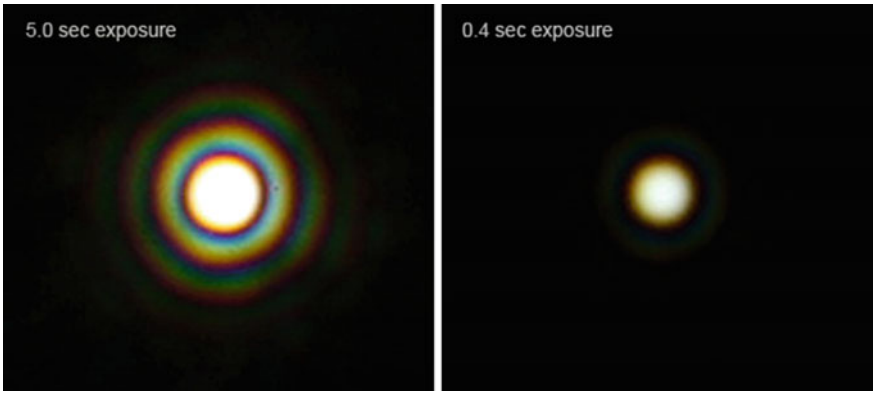


Fig. 5.29 Photo of a Fraunhofer diffraction pattern by a circular aperture. Courtesy of Rik Littlefield, Zerene Systems LLC. In the right panel, the exposure is correct for the Airy disc. In the left panel, the Airy disc is over-exposed to make the ring visible

diameter on the order of 0.5 mm. Look at a light source through the pinhole, positioned right up against your eye (take off your spectacles, if you wear them). The diffraction pattern will form on your retina with clearly distinguishable rings. Indeed, the size of the pattern is on the order of magnitude of the diameter of the central disk, which is, at a typical wavelength $\lambda = 0.5$ μm, $2\theta_{\min} = 2.44\lambda/D = 2.44 \times 5 \times 10^{-7}\text{m}/5 \times 10^{-4}\text{m} = 2.44$ mrad. This is an order of magnitude larger than the minimum angular distance the human eye can resolve (see Sect. 7.11).

When you do the experiment, you will notice that the diffraction pattern appears to be around the light source beyond the pinhole. As a matter of fact, the diffraction pattern is not in a definite place. This perception is due to a psychological mechanism in which the observed figure is located where we, as an observer, believe the source to be located.

5.8 Diffraction by Random Distributed Centers

In the preceding sections, we discussed the interference of light waves coming from two small (secondary) sources, Young's pinholes (or slits) and the diffraction from a single (secondary) source having an extension large enough to necessitate taking into account the interference between the waves coming from its different points. Young's pinholes, as opposed to the slit and the circular aperture of the previous sections, are point-like. What does this mean? Well, it simply means that the waves radiating from all the points of the secondary source are, in any case, in phase with one another independently of the position from which we look at the resultant wave.

In the following discussion, we shall call a single object that diffranges light a *diffraction center*, and also a *scattering center*. It can be an aperture, like the slit and the pinhole we have considered, but may also be a physical object, like a droplet of water. Indeed, we shall now study the case of several diffraction centers (as Young's holes are), whose structure, however, is not necessarily point-like (as was the case with the slit and the circular aperture). We shall limit our considerations to a system of equal centers. For example, we can consider a system of circular apertures of the same diameter D on the same plane, located regularly or irregularly at various distances from one another (although they might be, for example, squares, or stars, etc., equal in size and orientation). Once more, we shall work under Fraunhofer conditions.

The diffraction pattern of the system of circular apertures is determined by the interference between the waves from the different parts of the same aperture and those emitted by all the different apertures. The resultant field in the generic direction θ can be expressed as a sum of one term for each aperture. Each of these is the product of two factors. The first factor, which we call the *form factor* and indicate with $E_f(\theta)$, is due to the diffraction by the aperture. The second factor takes into account the difference between the phases of the waves reaching the observation point from homologous points of the different slits. This factor corresponds to the factor $\cos[\phi_2 - \phi_1 - k(r_2 - r_1)]$ that we encountered in the discussion of Young's experiment.

The form factor $E_f(\theta)$ is equal for all the centers and can be factorized, and we can conclude that the diffraction pattern of a system of equal (including orientation) centers is the product of the form factor of a single center times a factor depending on the relative positions of all the centers, which we shall call the *structure factor*.

The calculation of the structure factor is rather simple in two particular cases, both of practical interest, namely when the centers are randomly distributed and when they are periodically distributed. We shall treat the first case in this section, the second in the following one.

Before starting, we note that, as we already mentioned, the centers may be material objects like molecules, water droplets, powder grains or smoke particles. The apertures we had considered are secondary sources according to the Huygens-Fresnel construction, driven in phase with one another by the incoming wavefront. A water droplet, a molecule, etc., contains a large number of electrons, which are driven into oscillation by the electric field of the incoming wave. The wavelets that these accelerated charges radiate have initial phases determined by the incident wave. These are real, rather than virtual, secondary sources of the Huygens-Fresnel construction, but the phenomena we are considering are similar, being determined by the mutual coherence between the secondary sources.

Let us now consider one of the diffraction centers and let its diameter be D on the order of several wavelengths of the incident monochromatic wave. The center behaves, as we understand, similarly to the pinhole we have discussed, sending light mainly within an angle of about $\pm\lambda/\!/D$ (times a numerical factor of the order of one, depending on the shape of the center, which we can ignore here). If we look at angles larger than λ/D, we will see practically no light. Consider now centers of

larger and larger diameter. Clearly, the forward cone in which the light is diffracted becomes narrower and narrower. A homogeneous transparent medium, such as a clear glass sheet, for example, can be considered to be a single diffracting center. Its diameter is enormous when compared with the wavelength of light. The "diffracted" light practically moves only in the same direction as the incident wave. Indeed, these are the conditions we considered in Sect. 4.7, when we calculated the refractive index (in a sparse medium). We say that an optically homogeneous medium does not scatter light.

Let us then consider a monochromatic plane light wave incident on a system of diffraction centers, equal to one another, that are causally distributed. To fix the ideas, let us think of a slide having a black field on which a number of equal, transparent, circular apertures, or holes, have been registered. The apertures are randomly distributed. Let the average distance between the holes be much larger than the wavelength of the incoming wave. Otherwise, we would be dealing with an optically continuous medium. Such a slide can be easily prepared with a computer code. Under these conditions, the centers act incoherently with one another, and we talk of light scattering rather than of diffraction, but the phenomenon is basically the same.

Let us recall the expression in Eq. (5.2) that we found for two pinholes, which is valid for two narrow slits as well, namely

$$I = I_1 + I_2 + \sqrt{I_1 I_2}\cos[\phi_2 - \phi_1 - k(r_2 - r_1)]. \tag{5.24}$$

If the slits are wide (and equal), I_1 and I_2 are no longer constant, but rather a function of the considered angle θ, as given by Eq. (5.17). In the more general case of equal centers (for example, squares or pinholes), the expression is the same, with $I_1(\theta)$ and $I_2(\theta)$ being the form factor squared of the center being considered. Let us now generalize this result to an arbitrary number N of randomly distributed, equal centers. Calling $I_C(\theta) = I_1(\theta) = I_2(\theta) = \cdots = I_N(\theta)$, we have

$$\begin{aligned} I(\theta) &= I_1(\theta) + I_2(\theta) + \cdots + I_N(\theta) \\ &\quad + 2\sqrt{I_1 I_2}\cos[\phi_2 - \phi_1 - k(r_2 - r_1)] + 2\sqrt{I_1 I_3}\cos[\phi_3 - \phi_1 - k(r_3 - r_1)] + \cdots \\ &= NI_C(\theta) + 2I_C(\theta)\{\cos[\phi_2 - \phi_1 - k(r_2 - r_1)] + \cos[\phi_3 - \phi_1 - k(r_3 - r_1)] + \cdots\}. \end{aligned}$$

Now, the sum of the cosines on the right hand side is zero, because their arguments have all possible values with equal probability, and finally, we have

$$I(\theta) = NI_C(\theta). \tag{5.25}$$

We found that the light-scattering pattern of a system of randomly distributed centers, or particles, is equal to the scattering pattern of one of them multiplied by the number of particles in the system. If we send a monochromatic beam along the slide and we collect the transmitted light with a converging lens, we can observe the diffraction pattern on a screen in the focal plane.

The phenomenon of diffraction (or scattering) by casually distributed centers can sometimes be seen in the form of an iridescent ring, called a corona, appearing around a light source like a street lamp or the moon on a misty night. What we observe is simply the diffraction pattern of a single mist droplet or ice crystal present in the atmosphere, amplified by the factor N. As a matter of fact, the scattering particles have different sizes, in general. Only when the distribution of size is narrow enough does the pattern appear neat.

In general, quantitative measurements of the scattered light intensity as a function of the "scattering" angle are a powerful means for reconstructing the shape of the scattering object. Indeed, this is the method used to study the structures of molecules, atoms and nuclei. A sample containing the objects to be studied is "illuminated" with a beam of sufficiently short wavelength, ultraviolet, X-rays and gamma-rays, respectively, and the intensity of the scattered radiation is measured as a function of the angle.

The scattering of light by incoherent centers explains another everyday phenomenon, namely why the sky is blue. In the highest levels of the atmosphere, at 100 km above sea level, the average distance between molecules is substantially larger than the typical wavelength of light (namely than 0.5 μm). When illuminated by the sunlight, the molecules behave as independent scattering centers (in the visible part of the spectrum we are considering). Consider, for the moment, a monochromatic component of the incoming wave at the angular frequency ω. The molecule behaves as a forced oscillator in its stationary oscillating motion at the same angular frequency ω.

For a given amplitude of the periodic force acting on the oscillator, which is the electric field amplitude of the incoming wave, the amplitude of the forced oscillation depends both on ω and on the proper oscillation angular frequency ω_0 of the oscillator. As we know, neglecting damping, the oscillation amplitude is proportional to $1/(\omega^2 - \omega_0^2)$. Now, the proper angular frequencies ω_0 of the atmospheric gases are in the ultraviolet part of the spectrum. Hence, in the visible ($\omega \ll \omega_0$), the molecules behave as forced oscillators at frequencies much smaller than the proper one. Neglecting ω in comparison to ω_0, the oscillation amplitude is independent of ω (hence, of the color of the incoming wave).

On the other hand, the amplitude of the electric field scattered by an oscillating electron is proportional to the frequency and the electron acceleration, which is proportional to ω^2. The *intensity* of the scattered wave, which is proportional to the square of the electric field, is thus proportional to ω^4. Being that the frequency of the blue light is about 1.7 times the frequency of the red, the intensity of the scattered blue light is $1.7^4 \simeq 10$ times larger than that of the red one. This result is known as the Rayleigh blue sky law, from John Strutt Lord Rayleigh (UK, 1842–1919).

At heights lower than those just considered (of about 100 km), the density of the atmosphere becomes greater and greater, and the molecules oscillate with increasing degrees of coherence with one another. At sea level, the average distance between atmospheric molecules is on the order of nanometers, much smaller than the wavelength. However, the densities of gas with volumes on the order of a wavelength are not constant in time, but statistically fluctuate about their mean

value. Consequently, the distribution of these elements can be considered to be a set of independent centers and the above arguments apply once more, with density fluctuation in the place of single molecules. Other important scattering centers are the microscopic particles and aerosols that are always present in the atmosphere.

When we observe the atmosphere at sunset in the direction of the sun, the colors we see are reds. This is because we are now receiving the fraction of sunlight that has not been scattered by the atmosphere. We can observe that, on a clear day, the sunset is not nearly as colorful. In order for it to be very colorful, small amounts of smoke, dust or water particles need to be present in the atmosphere. These particulates diffuse the short wavelength radiation with much higher efficiency than they do for the longer ones and with the same amount as the thickness of the atmosphere crossed by the light. As the sun approaches the horizon, namely as the light we receive has crossed atmospheres of increasing thicknesses, the color becomes yellow, then orange and finally red.

These phenomena can be simply and spectacularly replicated by using sulfur particles in a water suspension. Figure 5.30 schematically shows the apparatus. The almost point-like source on the left should have a spectrum as similar as possible to the solar spectrum. A white LED with a power of several watts is suitable. The source is in the focus of a converging lens (L in the figure) that produces a parallel light beam crossing a volume of a few liters of distilled water in a glass container. The water should contain as few particulates as possible. The beam is received on a screen S after having crossed the water.

The light beam in the water, assuming it to be perfectly clean and containing no dust, would not be visible from the side. In practice, the presence of some powder is difficult to avoid, and this makes the beam visible, through scattering. Its color is white, because the dust particles are quite large.

We now melt into the water container about 5 g/l of $Na_2S_2O_3$. We obtain the formation of the sodium crystallites, pouring about 2 ml of concentrated sulfuric acid into the water diluted with 100 ml of distilled water per 10 l of water in the container and stirring it well.

A couple of minutes after having done that, we shall see the beam gradually becoming blue. This is because of the growth of small sulfuric scattering centers. After several minutes more, the entire container is now blue. Light has been scattered several times before exiting the container (multiple scattering) by the many centers that are now present.

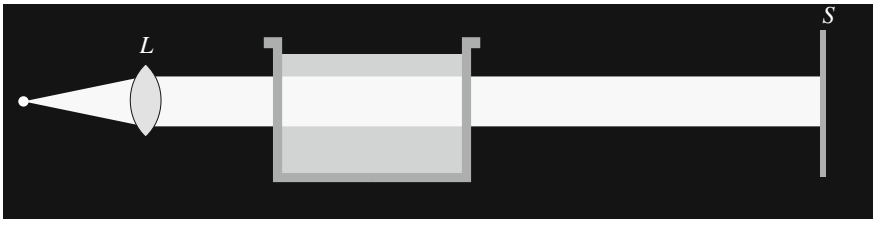


Fig. 5.30 Demonstration apparatus for scattering by randomly distributed centers

In this demonstration, the luminous circle on the screen S simulates the image of the sun through the atmosphere. When the beam becomes blue, the "sun" appears to be yellow, gradually becoming orange, and finally red, as the intensity of the scattered light increases.

5.9 Diffraction by Periodically Distributed Centers

We shall now discuss the Fraunhofer diffraction by a system of periodically distributed centers, in particular, parallel slits. The device is called a diffraction grating, more specifically, a Fraunhofer grating, when the width of the slits is much smaller than their distance. Diffraction gratings are used in optics to split a white light beam into almost monochromatic components traveling in different directions. In general, they are optically flat, mirrored glass sheets, having ridges rather than dark lines on them. Both the total number and the number per unit length of grooves are important, as we shall see. Good gratings have on the order of several hundreds up to one thousand grooves per millimeter, namely one every few micrometer.

Historically, the first diffraction gratings with 660 grooves per millimeter were produced by Henry A. Rowland (USA, 1848–1901) in 1881, using the ruling machine he had invented for that very purpose. The engine employed one main screw to shift the diamond tip used to etch the grating by a very small distance between each line. It is essential that the distance between the lines be kept constant. Rowland succeeded in this, thanks to an ingenious mechanical design, a piece of mechanical art that we shall not describe here. We shall, however, quote his dictum: "No mechanism operates perfectly-its design must make up for imperfections". Several non-mechanical techniques for fabricating diffraction gratings are presently available. We shall give some information on the holographic gratings in Chap. 8.

We shall now find the diffraction pattern of a diffraction grating under Fraunhofer conditions for light, with arguments that are more generally valid, namely for any periodic distribution of identical diffraction centers and for electromagnetic waves of any wavelength (for example, an array of parallel antennas for microwaves).

Consider a plane grating consisting of N straight lines of thickness D equispaced at a distance a and a plane monochromatic wave normally incident on the grating. We study the diffraction pattern under Fraunhofer conditions, namely at a large distance beyond the grating or in the focal plane of a converging lens, as shown in Fig. 5.31. We want to find the light intensity as a function of the angle θ relative to the incoming direction (which is normal to the groove plane).

Each slit can be considered, according to the Huygens-Fresnel principle, as being composed of secondary sources emitting hemispherical wavelets in the semi-space beyond the grating. Being that the secondary sources are on a front wave of the incident wave, they emit in phase with one another. We proceed in a manner similar to the previous section by first adding up the contributions of the

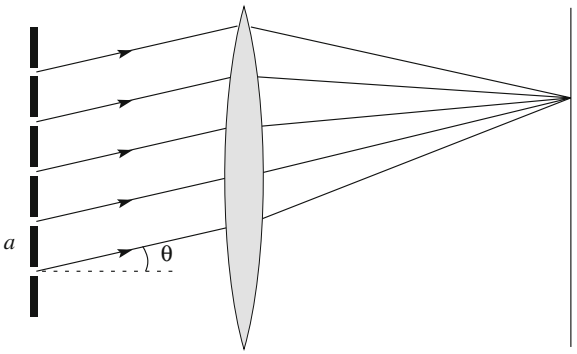


Fig. 5.31 Geometry of the Fraunhofer diffraction of a parallel lines grating

elements of a single slit and then adding up those of all the slits. The diffracted field of the single diffraction center, which is a slit, is given by Eq. (5.15), which we rewrite here for convenience:

$$E_f(\theta) = aD\frac{\sin\,\Phi}{\Phi}\mathrm{e}^{i\omega t}, \tag{5.26}$$

of which we shall take, as usual, the real part.

In adding up the contributions of the different slits, we must consider that the lengths of their path are different. The path difference between the waves from two contiguous slits is clearly $a\,\sin\,\theta$. Consequently, their phase difference when they meet is

$$\phi = ka\sin\theta = \frac{2\pi}{\lambda}a\sin\theta. \tag{5.27}$$

The resulting field is then

$$\begin{aligned} E(\theta) &= E_f(\theta)\mathrm{e}^{i\phi} + E_f(\theta)\mathrm{e}^{i2\phi} + \cdots + E_f(\theta)\mathrm{e}^{i(N-1)\phi} \\ &= E_f(\theta)\left[\mathrm{e}^{i\phi} + \mathrm{e}^{i2\phi} + \cdots + \mathrm{e}^{i(N-1)\phi}\right]. \end{aligned}$$

The result is a product of two factors, as in the previous section, but now the various terms in the latter are not random, but rather linked to one another by a definite phase relation. The sum on the right-hand side is of the type $S = 1 + \alpha + \alpha^2 + \cdots + \alpha^{N-1}$, with $\alpha = \mathrm{e}^{i\phi}$. Namely S being the sum of the first N terms of the geometric series. If the reader does not remember its value, he/she can easily verify that the sum satisfies the relation $\alpha S - S = \alpha^N - 1$. Hence, we have

$$S = \frac{\alpha^N - 1}{\alpha - 1} = \frac{\mathrm{e}^{iN\phi} - 1}{\mathrm{e}^{i\phi} - 1} = \frac{\mathrm{e}^{iN\phi/2}\left(\mathrm{e}^{iN\phi/2} - \mathrm{e}^{-iN\phi/2}\right)}{\mathrm{e}^{i\phi/2}\left(\mathrm{e}^{i\phi/2} - \mathrm{e}^{-i\phi/2}\right)} = \mathrm{e}^{i(N-1)\phi/2}\frac{\sin(N\phi/2)}{\sin(\phi/2)}.$$

Finally, the (complex) expression of the resulting field is

$$E(\theta) = AD\frac{\sin\,\Phi}{\Phi}\frac{\sin(N\phi/2)}{\sin(\phi/2)}e^{i[\omega t+(N-1)\phi/2]}. \tag{5.28}$$

To find the intensity, we must now, as usual, take the square of the real part of this expression and average it over a period. In taking the real part, the exponential factor gives us cos ($\omega t + \beta$), in which β is a constant that we might easily express, but is irrelevant because, in any case, $<\cos^2(\omega t + \beta)> = 1/2$. Calling $I_f(\theta)$ the intensity of a single slit, we conclude that the grating diffraction pattern is

$$I(\theta) = I_f(\theta)\frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)} = I_f(\theta)\frac{sin^2\left(\frac{Nka\sin\theta}{2}\right)}{sin^2\left(\frac{ka\sin\theta}{2}\right)}. \tag{5.29}$$

We now discuss this expression. Let us first notice that the function $I(\theta)$ is the product of two factors. The first factor, $I_f(\theta)$, depends on the shape of the slit, but not on its position. This is the form factor. The second factor depends on the number of slits and on how they are arranged. This is the structure factor.

Let us first consider the case in which the width of the slits is small compared to their distance, namely $D \ll a$. Under these conditions, the form factor $I_f(\theta)$ depends only weakly on θ (the central diffraction maximum is wide) and we can consider it to be constant relative to the structure factor.

Let us consider, under this condition, the simplest case, namely having two slits, $N = 2$. Equation (5.29) gives us

$$I(\theta) = I_f(\theta)\frac{\sin^2\phi}{\sin^2(\phi/2)} = 4I_f(\theta)\cos^2(\phi/2) = I_f(\theta)\cos^2\left(\frac{ka\,\sin\,\theta}{2}\right).$$

We have retrieved the diffraction pattern of the two-slit experiment by Young, Eq. (5.4), with $\phi_2 - \phi_1 = 0$.

Let us now study, in the general case of N slits, the behavior of the structure factor $\frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)}$. First, we see that the interference is completely constructive, and the intensity is a maximum, when the fields resulting from all the slits are in phase with one another, namely when the phases of two contiguous elements differ by an integer multiple of 2π, i.e., for $\phi = 2\pi n$, with n being an integer. All the terms in the sum S thus have the same argument. In terms of the angle θ, the condition is

$$a\,\sin\,\theta_{\max} = n\lambda. \tag{5.30}$$

This means that the condition for a maximum is having a path length difference equal to the integer multiples of the wavelength. All the contributions being in phase, the intensity in the maxima is

$$I(\theta_{\max}) = N^2 I_f(\theta_{\max}). \tag{5.31}$$

We see that the maximum intensity is proportional to the *square* of the number of diffracting centers. Consequently, it is very large. We recall that, for randomly distributed centers, the maximum intensity is proportional to the number of centers. Indeed, they are mutually incoherent in the latter case, coherent in the former.

These maxima are called the *principal maxima* (we shall immediately see why) and n the order of the maximum.

Note that the total number of maxima is finite. The order is limited by the fact that sin θ in Eq. (5.30) obviously cannot be larger than one, and we have

$$n \leq a/\lambda. \tag{5.32}$$

Notice that if, in particular, the distance between the lines a is smaller than the wavelength, only the zero order maximum exists. Light is emitted in a narrow lobe in the forward direction. At the limit at which the distribution of the centers becomes continuous, the transmitted wave has the same direction as the incident one.

Let us now analyze the behavior of the diffraction pattern near the principal maximum of the generic order n, for which $\phi = \phi_{\max} = 2\pi n$. Let us represent each term of the sum S as a rotating vector, as shown in Fig. 5.32 (using only eight vectors, for clarity; in practice, there would be much more). The diffracted intensity is proportional to the square of the resultant vector. In this representation, ϕ is just the angle between two contiguous vectors.

At the maximum, all the vectors rotate one on top of the other, being that their phase difference is an integer multiple of 2π, as in Fig. 5.32a. As we move away from the maximum, and as ϕ becomes different from 0 or from a multiple of 2π, the set of vectors open up like a fan (Fig. 5.32b). The resultant becomes smaller and smaller and vanishes when the fan extends over the entire round angle, namely when the difference between ϕ and $\phi_{\max}$, say $\Delta\phi = \phi - \phi_{\max}$ is $\Delta\phi = 2\pi/N$ (see Fig. 5.32c). Let $\theta_{\min}$ be the corresponding angle and $\theta_{\max}$ the angle corresponding to the principal maximum considered. Then, we have

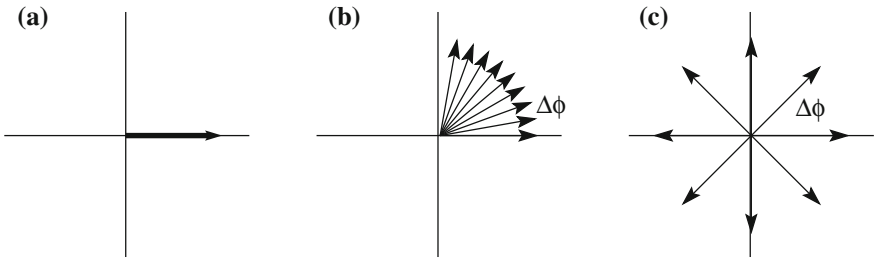


Fig. 5.32 Rotating vector representation of the terms of S: **a** in a principal maximum, **b** between the maximum and the nearest minimum, **c** in the nearest minimum

$$\Delta\phi = \frac{2\pi}{\lambda} a \sin\theta_{\min} - \frac{2\pi}{\lambda} a \sin\theta_{\max} = \frac{2\pi}{N}.$$

If we call $\Delta\theta$ the angular distance between the maximum and minimum, namely $\Delta\theta = \theta_{\min} - \theta_{\max}$, we have

$$\sin(\theta_{\max} + \Delta\theta) - \sin(\theta_{\max}) = \frac{\lambda}{Na}.$$

We can interpret $\Delta\theta$ as being the half-width of the principal maximum. We see that the maxima are all the narrower the greater the number N of slits. If the considered angles are small, as is often but not always the case, the above expression simplifies as

$$\Delta\theta \cong \frac{\lambda}{Na}. \tag{5.33}$$

The angular half-width is inversely proportional to the total number of lines under these conditions. Indeed, with increasing N, the height of the maximum increases as N^2, and its width decreases as $1/N$. The area of the maximum, which is proportional to the diffracted energy, is proportional to the number of slits, as expected.

Consider now two consecutive principal maxima, located, say, at $\theta_{\max,1}$ and $\theta_{\max,2}$. The angular separation between them is $\sin\theta_{\max 2} - \sin\theta_{\max 1} \cong \lambda/a$, which for small angles becomes

$$\theta_{\max 2} - \theta_{\max 1} \cong \lambda/a. \tag{5.34}$$

Hence, for small angles, the half-width of the principal maxima is equal to their distance $\Delta\theta$ divided by the number of slits.

Let us now continue increasing θ beyond the minimum. The resultant of the rotating vectors increases from zero, reaches a first secondary maximum, decreases to zero, increases to a smaller second secondary maximum and so on. In any case, the intensities in the secondary maxima are much smaller than in the principal ones, because the largest fraction of the rotating vectors is spread over the round angle, canceling one another out, and only a small fraction of them can add up in a non-zero resultant. In conclusion, we can state that the vast majority of the diffracted light is in the principal maxima.

The principal use of diffraction grating in optics is analyzing non-monochromatic light. Equation (5.30) tells us that, for every order, the angular position of the principal maximum depends on the wavelength. Consequently, the grating separates a non-monochromatic incident light in a set of diffracted waves at different angles. The separation in the sines of the angles of two given wavelengths is proportional to the order n. Hence, no separation exists for the 0th order. In white light, the central maximum is white, while in the higher orders, colors are separated

in a spectrum, with the red at larger angles than the blue (i.e., in an opposite sense as that for the dispersion phenomena).

An important characteristic of a grating is its *resolving power*, which is a measure of the capability of the grating to separate two wavelengths close to one another. Consider a light containing two components of wavelengths of a small difference, say λ and $\lambda + \delta\lambda$, incident on the grating. For simplicity, consider the light diffracted at a small angle (which is a good approximation for $\theta < 20°$). Let $\delta\theta$ be the angular separation of the maxima relative to the two wavelengths at the nth order. We find $\delta\theta$ by differentiating the maximum condition$\theta \approx \sin\theta = n\lambda/a$, obtaining $\delta\theta \cong (n.a)\delta\lambda$. Taking into account that each maximum has a non-zero half-width $\Delta\theta$ given by Eq. (5.33), we can state that the two components are distinguishable, namely are separated, if $\delta\theta \geq \Delta\theta$, namely if $\frac{n}{a}\delta\lambda \geq \frac{\lambda}{Na}$.

This condition is called the Rayleigh criterion. The resolving power of the grating is defined as the reciprocal of the minimum resolvable specific wavelength difference, namely as

$$\lambda/\delta\lambda = nN. \tag{5.35}$$

We see that the resolving power increases with the total number of lines and with the order of the considered spectrum.

For example, the resolving power of a grating with 500 lines per mm (namely $a = 2$ μm) and a total of $N = 50{,}000$ lines is equal to 100,000 at the second order. This means that it can separate two wavelengths differing by one part in one hundred thousand.

Notice that the number of lines per unit length cannot be too large, that is, their distance a cannot be too small if we want maxima of an order higher than zero. Indeed, we can write Eq. (5.32) as $a \geq \lambda n$. Hence, if we want, for example, to have the second order maxima ($n = 2$) and we take in the round figure $\lambda = 0.5$ μm, we have $a \geq 1$ μm. In conclusion, gratings with more than 1000 lines per mm are not useful for studying visible light.

Up to this point, we have studied the positions and the widths of the principal maxima. We shall now consider their heights. These are determined by the form factor $I_f(\theta)$ in Eq. (5.29), evaluated at the angle of the considered maximum. We recall that $I_f(\theta)$ depends on the width of the single slits.

Figure 5.33 shows the Fraunhofer diffraction pattern of a grating for which, as an example, the distance between lines is four times their width ($a = 4D$). The dotted curve is $I_f(\theta)$. The intensities of the maxima are modulated by the diffraction pattern of the slit. In particular, the 4th order principal maxima (on the two sides) are absent, because their positions coincide with those of the first minima of the slit diffraction pattern.

Figure 5.34 shows two photos taken of a single slit of $D = 54$ μm and two equal parallel slits of the same width illuminated with a laser of $\lambda = 532$ nm at a distance of $a = 108$ μm. One sees how the height of the maxima in the second photo follows the behavior of the single slit diffraction pattern.

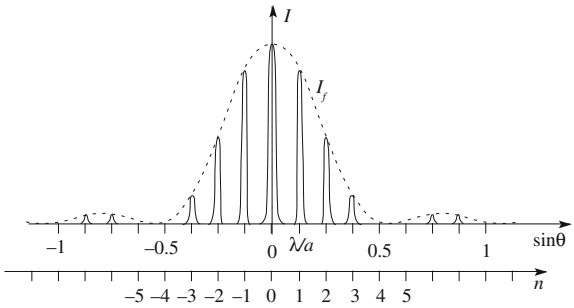


Fig. 5.33 Fraunhofer diffraction pattern of a grating with $a = 4D$

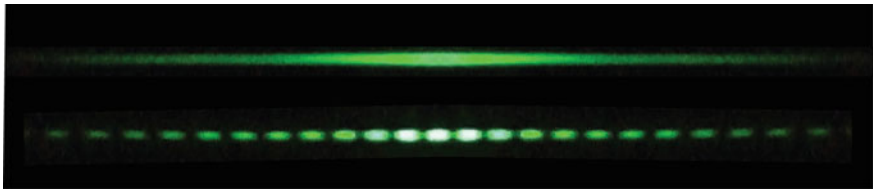


Fig. 5.34 Fraunhofer diffraction patterns of a single slit of width $D = 54$ μm and of two slits of the same width separated by $a = 108$ μm. Light source is a laser at $\lambda = 532$ nm. Photo by A. Bettini

In summary, the period of the grating determines the positions of the maxima, the total number of lines determines the width of the maxima and their height, and the shape of the slit (or, more generally, of the diffraction center) determines the relative values of the heights of the maxima.

The considerations we have just developed are useful when we want to study a periodic structure. Suppose we have a periodic structure of diffraction centers for which we do not know the period and the shape. This might be, for example, a crystal of a mineral. We can shine a monochromatic light of known wavelength on the structure. Measuring the angular positions of the diffraction maxima, we can extract the period of the crystal, while measuring their relative height, we gain information on the shape of the centers.

To be precise, the crystal lattices are periodic structures in three dimensions. Namely, the positions of the diffraction centers depend on three coordinates, rather than one, as with a Fraunhofer grating. The diffraction centers are atoms that may all be of the same type or of a few different types alternating one with the other. The situation is more complex, but still basically similar to that which we have studied. One understands how diffraction experiments using electromagnetic waves with wavelengths on the order of the lattice spacing, namely X-rays, are powerful tools for determining both the shape of the ladder and that of the atoms, or molecules, which act as diffraction centers.

5.10 Diffraction as Spatial Fourier Transform

In this section, we shall take back the Fraunhofer diffraction patterns of the single slit and the multiple slit grating. We shall see how they are just the squares of the Fourier transforms of the transparencies, as a function of the coordinates, and the corresponding diffracting systems (slit and grating, respectively). This conclusion has a general character and will lead us to consider diffraction phenomena from a different point of view, which we shall develop further in Chap. 8. We recall that, in Sect. 2.7, we studied the spatial Fourier transform considering two examples that will be useful now.

Let us start from the slit of infinite extension and width D illuminated by a plane monochromatic wave at normal incidence. The plane in which the slit is open, which we shall call the diaphragm, absorbs the incident wave outside the slit. We take the orthogonal coordinates with the origin in the center of the slit, the y-axis in the direction of the slit, the x-axis perpendicular to it in the plane of the slit and z in the direction of the incident wave. The incident wave has the same phase, which we call α_i, and the same amplitude, which we call A_i, at all the points of the $x\ y$ plane. Let A_t be the amplitude of the wave immediately after the diaphragm. Clearly, A_t is equal to A_i for $-D/2 \leq x \leq D/2$ and zero outside this interval. We can say that the effect of the screen with the diaphragm is that of multiplying the incident amplitude by an *amplitude transmission coefficient*, which we define as the following function of x:

$$T(x) = 1 \text{ for } -D/2 \leq x \leq D/2; \quad T(x) = 0 \text{ for } |x| > D/2. \tag{5.36}$$

We now generalize the concept, calling any surface having different levels of transparency depending on the point a *diaphragm*. For simplicity, consider, for the moment, a transparency function of x alone. This is, for example, the case of a diffraction grating in which the transparency periodically varies between 1 (inside the slits) and 0 (between the slits), as in Fig. 5.35a. We can also think of cases in which $T(x)$ varies through any value between 0 and 1, namely with different levels

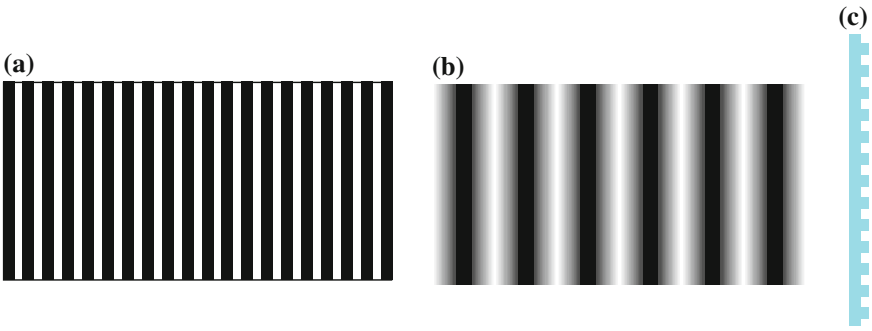


Fig. 5.35 Examples of **a** and **b** amplitude diaphragm (*front view*), **c** phase diaphragm (*side view*)

of grey rather than only black and white (as in Fig. 5.35b), and of cases in which the function is not periodic. More generally, the function T depends on both coordinates. Such is the case with, for example, black and white photos on a slide. We call a diaphragm having the effect of multiplying the amplitude of the incoming wave by a factor $T(x, y)$, which is called the amplitude transparency of the diaphragm, an *amplitude diaphragm*. Note that the thickness of the sheet supporting the amplitude diaphragm is ideally zero so as not to introduce any phase difference between the incoming and outgoing waves.

Consider, as an opposite case, a sheet of refractive index n, which is perfectly transparent and consequently does not change the amplitude of the incident wave. However, its thickness varies from point to point, as shown, for example, in Fig. 5.35c. What is the effect of this diaphragm on our normally incident plane monochromatic wave? The amplitude of the incident wave is

$$E_i(x, y, z, t) = A_i \mathrm{e}^{i(\omega t - kz + \alpha_i)} = A_i \mathrm{e}^{i\alpha_i} \mathrm{e}^{i(\omega t - kz)},$$

where, on the right-hand side, $A_i \mathrm{e}^{i\alpha_i}$ is the complex amplitude. Consider the field in the points x, y of a plane immediately beyond the plate. Its real amplitude is unaltered, being A_i, but its phase is different from the phase at the entrance, because the diaphragm has introduced a phase delay, which is a function of its thickness and hence of the position. We call the phase delay between exit and entrance $\Delta\phi(x, y)$. The transmitted wave is then

$$E_t(x, y, z, t) = A_i \mathrm{e}^{i\alpha_i} \mathrm{e}^{i\Delta\phi(x,y)} \mathrm{e}^{i(\omega t - kz)}.$$

We see that the effect of the diaphragm is that of multiplying the incident complex amplitude by the factor $T(x, y) = e^{i\Delta\phi(x,y)}$. We call this type of diaphragm a *phase diaphragm*. We shall discuss an example of this in Sect. 7.12. The function T is also called the amplitude transparency in this case.

Clearly, the most general diaphragm changes both the amplitude and the phase of the incident wave. We thus speak of *amplitude and phase diaphragm*s. Its amplitude transparency has both an amplitude different from one and a phase different from zero, both being functions of the coordinates.

We shall now discuss two simple examples of amplitude diaphragms having an amplitude transparency function of one coordinate only (x). Consider a normally incoming monochromatic plane wave of amplitude A_i and let $A_t(x)$ be the transmitted amplitude. The amplitude transmission coefficient is then

$$T(x) = A_t(x)/A_i(x). \tag{5.37}$$

Note that the transparency we see looking, for example, through a slide is determined by the ratio between transmitted and incident *intensities* rather than the amplitudes. It is proportional to the absolute square of $T(x)$.

Let us now go back to the slit that we consider to be an amplitude diagram of amplitude transparency given by Eq. (5.36). As we saw in Sect. 5.6, the diffracted amplitude in the direction θ, under Fraunhofer conditions, is proportional to the integral

$$\int_{-D/2}^{+D/2} e^{-ikx\sin\theta} dx,$$

where k is the wave number. Taking into account Eq. (5.36), we can write this integral as

$$\int_{-\infty}^{+\infty} T(x) e^{-ikx\sin\theta} dx.$$

We now note that this expression is proportional to the diffracted field for every diaphragm if we insert its amplitude transmission coefficient $T(x)$. The expression is general.

We now make the important observation that the x component of the wave vector $\mathbf{k}$ of the diffracted wave in the direction θ is $k_x = k\sin\theta$. This expression appears in the integral we have written, and we can write that the diffracted amplitude in the direction θ is proportional to

$$G(k_x) = \int_{-\infty}^{+\infty} T(x) e^{-ik_x x} dx. \tag{5.38}$$

We immediately recognize this expression to be (apart from the irrelevant 2π factor) the space Fourier transform of the amplitude transmission coefficient $T(x)$. In conclusion, the field diffracted in the direction θ by a diaphragm of amplitude transparency $T(x)$ is proportional to the space Fourier transform evaluated at the value of the component $k_x = k\sin\theta$ of the wave vector in that direction. As in the temporal Fourier transform, the variable conjugated to time is the angular frequency, while in the spatial Fourier transform, the variable conjugate to x is the x component of the wave vector.

Coming back to the slit of width D, the integral in Eq. (5.38) has already been calculated in Sect. 2.7. The result is given by Eq. (2.77), which, apart from the 2π factor, is

$$G(k_x) = D\frac{\sin(Dk_x/2)}{Dk_x/2}. \tag{5.39}$$

We immediately recognize the amplitude of the diffracted field in Eq. (5.16). A comparison of Fig. 5.24a with Fig. 2.23 shows that the curve in the latter is equal to the square of the curve in the former.

The physical meaning of the result is as follows. The incident field is a plane wave propagating in the z direction, which is the direction of its wave vector. The x component of the incident vector has a well-defined value, namely $k_x = 0$. The diffracted field is not a plane wave, and it does not have a single propagation direction; rather, it propagates with different amplitudes in different directions. We can think of the diffracted field as being a linear superposition of an infinite number of plane waves (the Fourier integral), each proceeding in its own direction, or, put another way, one for each value of k_x, with amplitudes specified by Eq. (5.39).

A difference with the temporal case is that, in the case of the spatial Fourier transform, both positive and negative values of the Fourier variable (k_x in this case) have a direct physical meaning. They correspond to waves moving to the right and to the left, respectively (thinking of the slit as being vertical).

The new point of view we have gained allows us to understand two important aspects of diffraction phenomena.

We recall the bandwidth theorem established in Chap. 2 for a function of time. The theorem states that the bandwidth of the Fourier transform is inversely proportional to the duration in time. The bandwidth is the range of frequencies that significantly contributes to the spectrum of the function. In the spatial case, consider again the case of the slit. The narrower the slit, namely the narrower the limitation it induces on the intercepted wavefront along the x coordinate, $\Delta x = D$, the wider the interval of values of k_x that are present in the diffracted wave. If we call this interval Δk_x, the relation given by Eq. (2.78) is

$$\Delta x \cdot \Delta k_x = 2\pi. \tag{5.40}$$

The result is completely general. Whatever way the front of a wave is limited in its transversal extension, the propagation direction, or, in other words, the corresponding wave vector component, it is no longer completely defined, to an extent that it is greater the narrower the limitation of the front. In quantum mechanics, the propagation of a particle is described by an associated wave. The wave vector is proportional to the linear momentum of the particle. The expression that exactly corresponds to Eq. (5.40) is the position-momentum uncertainty principle.

A second aspect that we can now easily understand is the following. In Sect. 2.7, we saw that the Fourier transforms of the amplitude transmission coefficient of a diaphragm with opaque and transparent regions and a second one that is transparent where the first is opaque and opaque where the first is transparent have equal amplitudes. For example, the diffraction pattern of a slit and an absorbing strip (Fig. 2.22) with the same width (Fig. 2.24) are equal. This statement is general and is called *Babinet's principle* after Jaques Babinet (France, 1794–1872).

The absolute square of the spatial Fourier transform of a diaphragm can be seen, in a literal sense, through the following procedure. As we know, the diffraction pattern under Fraunhofer conditions of a diaphragm can be seen on a screen in the

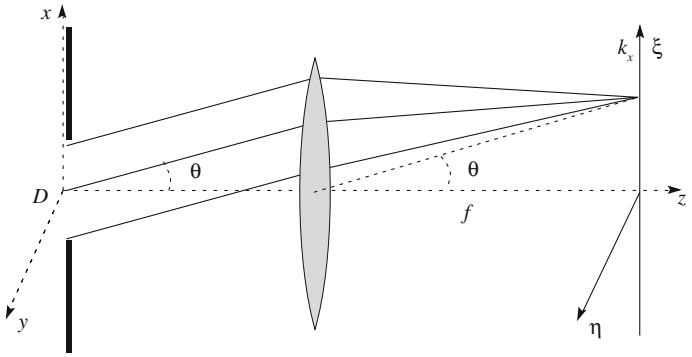


Fig. 5.36 Geometry for observing the spatial Fourier transform of a diaphragm

focal plane of a converging lens located beyond the diaphragm. Figure 5.36 shows the arrangement with a slit as a diaphragm, as an example. Let ξ and η be two orthogonal coordinates on the screen parallel to x and y, respectively, and with origin on the optical axis.

As known from elementary optics (and as we shall see in Chap. 7), parallel rays incident on a lens at the angle θ are focalized by the lens at a point of the focal plane at a distance from the axis proportional to tgθ. Under small angle conditions, we can assume that distance to be proportional to sin θ and then to ξ. Under these conditions, ξ is proportional to $k_x = k\sin\,\theta$.

We can conclude that when we have a monochromatic plane wave (which is spatially and temporally coherent) normally incident on an amplitude diaphragm of amplitude transmission coefficient $T(x, y)$ and we deploy a convergent lens beyond the diaphragm, the amplitude of the field on the focal plane is the spatial Fourier transform of $T(x, y)$. A consequence is that the components of low spatial frequencies are found near the optical axis, while those of higher frequencies are farther from it. If we put a photographic film in the focal plane, the grey level distribution resulting from its exposure is proportional to the light intensity, namely to the absolute square of the Fourier transform. Clearly, the conclusion is valid in general, for any $T(x, y)$, not being limited to the example of the slit.

The following examples may be useful for grasping the physical meaning of a spatial Fourier transform. Consider, as a diaphragm, a slide of a shot you have taken. Remember that $T(x, y)$ is the ratio between the amplitude of the transmitted and incident waves. If the slide has high contrast images, like, for example, the edges of an illuminated wall near black shadows or quite small objects, then $T(x, y)$ varies quite rapidly, at least at some parts of the slide, and its Fourier spectrum contains important components at high spatial frequencies. Contrastingly, if the images are smoother, for example, in a photo of a meadow, the spectrum of $T(x, y)$ is limited to a lower frequency bandwidth.

In Sect. 2.6, we also studied the Fourier transform of an infinite succession of square pulses in time. Clearly, we can think of the analogous situation in space. We have an infinite succession of slits, which is an idealized diffraction grating of infinite extension. The system being periodic, its spectrum is discrete. Let us compare Fig. 2.16, thinking of its analogous situation in space, with Fig. 5.33. We understand that the different components of the spectrum are just the principal diffraction maxima. They appear to be infinitely narrow, because they have been produced by an unrealistic grating having infinite lines.

Let us finally consider the diffraction pattern of a grating with an amplitude transmission coefficient, varying like a sine between 0 (fully transparent) and 1 (fully opaque), as in Fig. 5.35b, namely with

$$T(x) = \frac{1}{2}(1 + \cos kx). \tag{5.41}$$

The Fourier transform of $T(x)$ contains only a constant and the fundamental. Consequently, the principal maxima in the diffraction pattern are those of the orders 0, +1 and −1 only. We shall come back to this issue in Chap. 8, where we shall see, in particular, how such a grating can be fabricated.

Summary

In this chapter, we studied three types of phenomena: interference, diffraction and scattering. They are consequences of the wave nature of light and, ultimately, are substantially different aspects of the same phenomenon. The most important concepts we have learned are the following:

1. Two (or more) monochromatic light sources having a fixed phase relationship with one another produce interference. The total intensity is not the sum of the two intensities, but an interference term that can be positive or negative must be added. In practice, to have interference, one must start from a single source, divide the produced beam in two, have both beams going through different paths, and finally recombine the two of them.
2. The light intensity resulting from two independent sources is the sum of the two intensities taken separately from one another.
3. Spatial and temporal coherence are necessary conditions for observing interference. The degree of spatial coherence between two points in a certain instant is the degree of correlation of the field at the two points in that instant. The degree of temporal coherence between two instants at a certain point is the degree of correlation of the field in the two instants at that point. The bandwidth theorem states that the degree of temporal coherence is inversely proportional to the bandwidth.
4. Under particular conditions, interference is also observed when the coherence conditions are not fulfilled. In these cases, the fringes are localized.

5. Diffraction happens every time the front of an advancing wave is restricted by an obstacle. Diffraction is when light does not propagate in a straight line for reflection or refraction. Particularly relevant cases are diffraction by a slit and by a circular aperture.
6. The pattern due to diffraction by a number of equal centers is the product of a form factor (that is, the diffraction pattern of the single center) and a structure function (that depends on the arrangement of the centers).
7. When the centers are casually distributed, the diffraction pattern is equal to the form factor of the center multiplied by their number.
8. Diffraction gratings are important examples of centers periodically distributed.
9. The important properties of a Fraunhofer grating are:
    a. The largest fraction of the diffracted light is in the principal maxima.
    b. The angles of the principal maxima are increasing functions of the wavelength.
    c. The intensities of the principal maxima are proportional to the square of the total number of lines.
    d. The intensities of the principal maxima are proportional to the intensity of the diffraction pattern of the single slit of the grating at its position.
    e. The width of the maxima is inversely proportional to the total number of lines and to the order.
    f. The spectral resolving power is equal to the number of lines times the order of the considered maximum.
10. The diffraction pattern under Fraunhofer conditions is apportioned to the square of the Fourier transform of the amplitude transparency of the diaphragm.

Problems

5.1 Consider an ultrasonic monochromatic plane wave incident on a screen with two narrow widths separated by 100 mm. The angular distance between interference maxima is 9°. What is the wavelength?

5.2 We conduct the Young two-slit experiment with dichromatic light containing the two wavelengths $\lambda_1 = 450$ nm and $\lambda_2 = 550$ nm. We observe the fringes in the focal plane of a lens. What is the ratio of the distances from the axis of the first order fringes for the two wavelengths? And for the fourth order fringes? And for the zero order?

5.3 Can two harmonic oscillations of different frequency be coherent?

5.4 We want to perform the Young experiment with two pinholes separated by 2 mm and a source having 5 mm diameter emitting monochromatic light of 0.6 μm wavelength at 3 m from the pinholes. Will the experiment work? What is the minimum distance we should locate the primary source, maintaining the other conditions?

5.5 Consider a light monochromatic plane wave incident on the Young slits at the θ with the normal. Show that the smallest value of θ, for which, in the forward direction, there is an interference minimum, is $\lambda/2d$.

5.6 A parallel light beam of $\lambda = 650$ nm is incident on a slit of 2 μm width. Find the angles at which the intensity minima are observed.

5.7 With mercury light, which has the wavelength $\lambda = 546.1$ nm, we observe, with a Fraunhofer grating, the principal maximum of the first order, at 19°. How many lines per millimeter does the grating have?

5.8 What is the maximum diffraction order one can obtain with light of wavelength $\lambda = 0.5$ μm and gratings of 0.1 mm, 0.01 mm e 0.001 mm in turn?

5.9 A grating of $N = 500$ lines is illuminated by a sodium lamp. The sodium light contains a doublet (two nearby components) of $\lambda_1 = 589.0$ nm and $\lambda_2 = 589.6$ nm. Are the two lines resolved at the first order? And at the second order?

5.10 We perform the Young two-slit experiment with light of a bandwidth between 450 and 650 nm. How many fringes can we observe?

5.11 State on which of the following characteristics the resolving power of a Fraunhofer grating depends: the light wavelength, the period of the grating, the number of slits, the order of the considered maximum.

5.12 What is the maximum order at which we can observe the sodium yellow light ($\lambda = 589$ nm) with a 2 μm period grating?

5.13 A vertical film of soapy water forming a sledge is illuminated by normally incident white light. We observe the localized fringes through a red glass (that transmits wavelengths around 630 nm), measure their distance and find 3 mm. What shall the period be if we observe the fringes through a blue ($\lambda = 430$ nm) glass?

5.14 What is the minimum number of lines for a grating to be able to resolve in the first order the doublet in the potassium spectrum at $\lambda_1 = 404.4$ nm and $\lambda_2 = 404.7$ nm?

5.15 A slit limits a plane monochromatic wave of wavelength $\lambda = 0.6$ μm in the x direction. In the diffracted wave, the values of k_x range from -10^6 to $+10^6$ m^{-1}. What is the width of the slit? How much is the angular width of the diffraction maximum?

Chapter 6
Polarization

Abstract Polarization phenomena are characteristic of transverse waves, namely when the wave function is a vector normal to the propagation direction. Our focus will be on light waves, but the concepts discussed will have a general character. We shall define the different polarization states of light and establish the relationship between them. Light from thermal sources is not polarized. We then study the phenomena of dichroism, scattering, reflection and birefringence, the structure of the light wave in an anisotropic medium and, finally, optical activity.

To characterize several types of wave, a scalar function is not sufficient, but a vector function is required. In other words, in addition to the amplitude, the wave function also has a direction. So, while a sound wave is characterized only by its amplitude, in the case of a vibrating string, we must also specify the direction of the vibration. The string of a guitar, for example, can vibrate perpendicularly or parallel to the soundboard. The vibrations in any other direction can be expressed as a superposition of those two.

Electric and magnetic fields in a vacuum and in the isotropic media are normal to the propagation direction of the progressive electromagnetic waves. As we know, we can only consider one of these fields, the other being known when the first is known. We shall consider, as usual, the electric field. The polarization phenomena depend on the direction of the electric field. We shall study some of them in this chapter, considering, in particular, light waves. However, the concepts will have a general character.

Note that the word "polarization" has several meanings in physics. Its meaning here is different from the polarization of a dielectric medium.

In Sect. 6.1, we shall define the different polarization states of light and establish the relationships between them. In Sect. 6.2, we shall see that light from thermal sources is not polarized, namely it is not in any of the above-defined polarization states. In Sects. 6.3, 6.5, 6.6 and 6.7, we study four phenomena leading to polarized light, namely dichroism, scattering, reflection and birefringence. In Sect. 6.4, we study the basic instrument capable of determining the polarization state, namely the polarization analyzer, and in Sect. 6.8, the instruments capable of altering the phase

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_6

difference between components of polarized light. In Sect. 6.7, the study of the structure of a plane monochromatic wave in a uniaxial anisotropic medium will teach us that, in such a medium phase and group velocities differ not only in their absolute values but in their directions as well. Finally, in Sect. 6.9, we shall study optical activity, a phenomenon characteristic of the circularly birefringent media, in which the phase velocities of circularly polarized light in clockwise and counter-clockwise directions are different.

6.1 Polarization States of Light

Any transverse wave, like the elastic waves on a string and electromagnetic waves, both progressive and stationary, may be polarized or not, and, in the former case, can be in different states of polarization. Polarization is connected to relations between different directions of the wave function. Consequently, we do not speak of polarization for longitudinal waves, like sound, in which there is only one direction of the wave function. We shall limit our discussion mainly to light and to plane progressive monochromatic waves. The greater part of our conclusions, however, has a more general validity.

Let us consider a progressive plane monochromatic electromagnetic wave propagating in the positive z direction, and let us choose the x and y axes perpendicular to one another and to z. Let ω be the angular frequency and k the wave number of the wave. We shall now study phenomena connected to the direction of the electric field vector. As we learned, the direction of the vector $\mathbf{E}$ in a vacuum and in an isotropic medium is, at every point and in every instant, perpendicular to the propagation direction z, namely in the $x\,y$ plane. In general, this direction varies along the propagation direction and with time. If we take a picture, so to speak, of the electric field at a certain instant, it will show its tip describing a curve within the space, which is, in general, not in a single plane. Similarly, if we look at the field at a certain z, we will see its direction varying with time.

We say that a wave is *linearly polarized* or that it is in a state of *linear polarization* if the direction of $\mathbf{E}$ is the same in every instant and at every point. Clearly, any direction in the xy plane is a possible direction of linear polarization for an electromagnetic wave (in general, for any transverse wave). We define as *base polarization states* the two states of linear polarization in the directions of the x and y axes, namely

$$E_x(z,t) = E_{01}\cos(\omega t - kz + \phi_1), \quad E_y(z,t) = 0 \tag{6.1}$$

and

$$E_x(z,t) = 0, \quad E_y(z,t) = E_{02}\cos(\omega t - kz + \phi_2). \tag{6.2}$$

We see that two constants are present in each of these expressions, namely the oscillation amplitude, which we called E_{01} and E_{02}, and the initial phases in the

origin (namely for $t = 0$ and $z = 0$), which we called ϕ_1 and ϕ_2. As we shall immediately see, every sum of the states in Eqs. (6.1) and (6.2) gives a linear polarized wave (of the considered angular frequency and wave number), and, reciprocally, any linear polarization state can be expressed as a sum of the states in Eqs. (6.1) and (6.2) by suitably choosing the four constants. For these reasons, they are called base states. Note that the choice of the directions x and y is clearly arbitrary, as long as they are normal to the propagation direction and to one another.

We immediately notice that, in practice, there are three independent constants rather than four. Indeed, we can arbitrarily fix one of the phases, because this amounts to a change in the $t = 0$ instant. As usual, what matters is the phase difference, not the absolute values of the phases. We now exploit this arbitrariness so as to put one phase to zero, and write the base states as

$$E_x(z,t) = E_{01}\cos(\omega t - kz), \quad E_y(z,t) = 0 \tag{6.3}$$

and

$$E_x(z,t) = 0, \quad E_y(z,t) = E_{02}\cos(\omega t - kz + \phi). \tag{6.4}$$

We shall now study the polarization states obtained with different combinations of the base states. We start by adding two base states with the same phase, which we can take to be zero, and arbitrary amplitudes. We obtain

$$E_x(z,t) = E_{01}\cos(\omega t - kz), \quad E_y(z,t) = E_{02}\cos(\omega t - kz) \tag{6.5}$$

We see that both components of the field vanish in the same instants, which means that we are dealing with a linear polarization state. The polarization direction is determined by the ratio of the amplitudes on the two axes. More precisely, E_{02}/E_{01} is the tangent of the angle of the polarization direction with the x-axis. Figure 6.1 shows three examples, with the black line representing the oscillation of the field.

In Fig. 6.1a, we have $E_x(z,t) = 0$ and $E_y(z,t) = E_0\cos(\omega t - kz)$, with an arbitrary E_0. In Fig. 6.1b, we have $E_x(z,t) = E_0\cos(\omega t - kz)$ and $E_y(z,t) = (E_0/2)\cos(\omega t - kz)$, and in Fig. 6.1c, $E_x(z,t) = E_0\cos(\omega t - kz)$ and $E_y(z,t) = 0$.

We also obtain a linear polarization state if the phase difference between the base states is π. In this case too, the two components vanish at the same instant. Indeed,

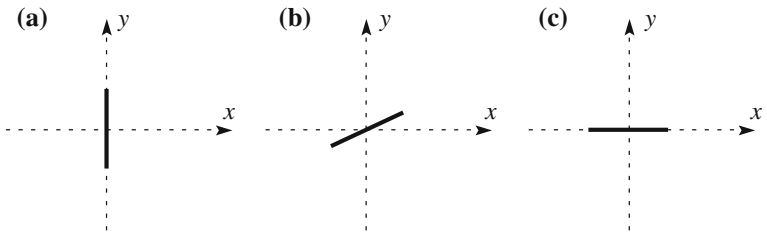


Fig. 6.1 Three examples of linear polarization. **a** $E_{01} = 0$, **b** $E_{02} = E_{01}/2$, **c** $E_{02} = 0$

changing the phase by π is equivalent to changing the sign of the amplitude. When we talk of equal phases, we will always mean equal modulo π.

Let us now consider a second example, namely combining two base states with the same amplitude and a phase difference of $\pm\pi/2$. We have

$$E_x(z,t) = E_0 \cos(\omega t - kz), \quad E_y(z,t) = E_0 \cos(\omega t - kz \pm \pi/2). \tag{6.6}$$

To be concrete, let us consider the case with the plus sign, which we can write as

$$E_x(z,t) = E_0 \cos(\omega t - kz), \quad E_y(z,t) = -E_0 \sin(\omega t - kz). \tag{6.7}$$

Let us look at the wave in the direction opposite to its propagation direction, namely with the wave coming toward us, and to the trajectory, so to speak, of the tip of the electric field projected onto the xy plane. This is given by Eq. (6.7) with $z = 0$. Clearly, the curve is a circle made in a clockwise direction. Similarly, the case of $-\pi/2$ gives us a circle in a counter-clockwise direction. These two states are called *circular polarization states*. We define a state of *right* circular polarization as being one in which we look at the source straight in the direction of the incoming wave and see the electric vector rotating in a counter-clockwise direction, with *left* circular polarization being the same, but clockwise. The reason for the adjectives is that the definition can also be given by stating that, for right polarization, were you to point the thumb of your right hand in the direction of propagation of the wave, the electric vector would be rotating in the direction of your fingers, and similarly for the left polarization. These two states are shown in Fig. 6.2. Note that this convention is the most adopted one in physics, but unfortunately, it is not universal, so that you can find the opposite one in a number of books.

In order to better understand how two linearly polarized waves combine with one another to produce a circularly polarized wave, let us consider now the following example. Figure 6.3a shows a snapshot of a monochromatic wave linearly polarized in a vertical plane. You can think of the mechanical oscillation of a string, or of the electric field of an electromagnetic wave. Figure 6.3b shows a monochromatic wave of the same wavelength and amplitude polarized in the horizontal plane. The latter precedes the former by a quarter of a wavelength. This means that, in a position in which the former has a maximum, the latter is zero.

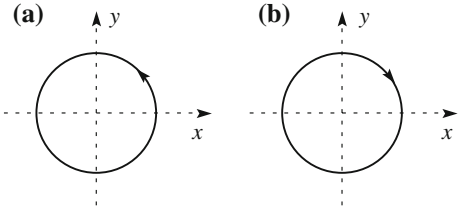


Fig. 6.2 Figures show the trajectories of the tip of the electric field with the propagation direction normal to the figure towards the reader. **a** *Right* circular polarization, **b** *Left* circular polarization

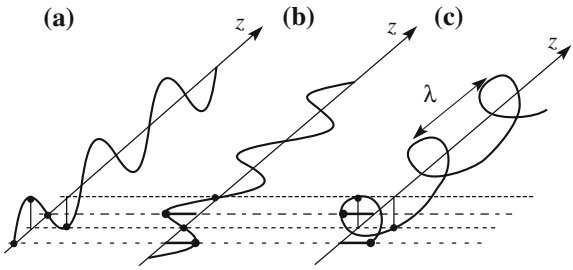


Fig. 6.3 Snapshot of an elastic progressive wave on a string; **a** vertical linear polarization, **b** horizontal linear polarization advanced by a quarter wavelength, **c** combination of (**a**) and (**b**)

In the presence of both waves, being that the elastic medium and the electromagnetic field are both linear, the resulting motion of the string, or the resulting electric field, is the sum of those of the two waves considered separately. Figure 6.3c shows the resulting total displacement, or electric field. The magnitude is constant, while the direction varies linearly with the position. The string has the shape of a helix wrapped at a fixed distance from the axis. The pitch of the helix is the wavelength λ. If we now think of freezing the helix at a certain instant, we obtain a screw. If we want to have the screw advance, we must turn it counter-clockwise, as seen from the source, namely in the direction opposite to that of a normal screw. This is the state we defined as being *left*. If the wave in (b) was delayed one quarter wavelength rather than advanced, we would have obtained a helix would like a normal screw. The polarization would have been *right.*

Let us now move to the immediately more complicated case, namely the sum of two base states with a phase difference of $\pm\pi/2$ and different amplitudes, namely

$$E_x(z,t) = E_{01}\cos(\omega t - kz),\ E_y(z,t) = E_{02}\cos(\omega t - kz \pm \pi/2). \tag{6.8}$$

It easy to see that the trajectory of the tip of the electric field is now an ellipse, having x and y as axes and semi axes E_{01} and E_{02}. The rotation sense is like in the cases of circular polarization. We say that the polarization is elliptical (right or left). Figure 6.4 shows an example of right elliptical polarization.

As a matter of fact, elliptical polarization is the most general state of polarization (linear and circular polarizations being special cases of elliptical polarization), when the axes do not necessarily coincide with the x and y reference axes. We immediately understand this to be true if we think of that fact that the most general motion of a pendulum is about an ellipse and is the combination of two harmonic motions with the same frequency on the two coordinate axes.

Let us prove this, namely that the most general combination of the base states that

$$E_x(z,t) = E_{01}\cos(\omega t - kz),\ E_y(z,t) = E_{02}\cos(\omega t - kz + \phi) \tag{6.9}$$

with arbitrary E_{01}, E_{02} and ϕ being an elliptical polarization. The easiest way is to prove that the resulting "trajectory" of the tip of the electric field is a conic curve. Indeed, we know that the field magnitude is limited (does not go to infinite) and the

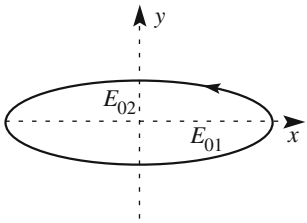


Fig. 6.4 *Right* elliptical polarization. Propagation direction is toward the reader

only limited conic is the ellipse. Now, it is clear that E_x and E_y can be expressed as linear combinations $\cos(\omega t - kz)$ and $\sin(\omega t - kz)$ (the first one is already such). These expressions, which are easy to find but unnecessary for our purposes, are a system of two linear equations that, once solved, gives $\cos(\omega t - kz)$ and $\sin(\omega t - kz)$ as *linear* combinations of E_x and E_y. If we now impose the condition $\cos^2(\omega t - kz) + \sin^2(\omega t - kz) = 1$, we obtain an expression of the type $A + BE_x^2 + CE_y^2 + DE_xE_y = 0$, where A, B, C and D are constants that we do not need to express, because this expression is, in any case, the equation of a conic.

Figure 6.5 shows several examples of elliptic polarization, all with amplitude in the y direction twice as large as that in the x direction and different values of the phase difference.

We have seen that all the polarization states can be expressed as linear combinations, or, we can also say, superpositions, of two base states, which are the linear polarization states along two axes normal to one another. However, the most general polarization state can also be expressed as a superposition of other pairs of base states. The most important are the circular polarization states. We shall not demonstrate this statement, but we shall only show that a linear polarized wave can be expressed as a superposition of two circular polarized waves, one right and one left, having equal amplitudes.

Let us consider a right circularly polarized monochromatic wave moving in the positive z direction. Let the components of its electric field at the point $z = 0$ be given by

$$E_x^R = E_0 \cos \omega t, \quad E_y^R = E_0 \sin \omega t.$$

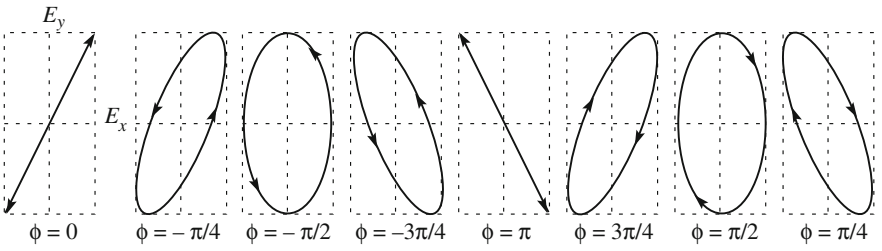


Fig. 6.5 Examples of elliptic polarization with $E_{02} = 2E_{01}$ and phase differences as in the *inserts*

The generic left circularly polarized wave of the same amplitude (and same frequency) is then

$E_x^L = E_0 \cos(\omega t + \alpha), \quad E_y^L = -E_0 \sin(\omega t + \alpha),$

where we took into account a possible initial phase difference α. The resulting field is

$$E_x = E_x^L + E_x^R = E_0[\cos(\omega t) + \cos(\omega t + \alpha)] = 2E_0 \cos(\omega t + \alpha/2)\cos(-\alpha/2)$$
$$E_y = E_y^L + E_y^R = E_0[\sin(\omega t) - \sin(\omega t + \alpha)] = 2E_0 \cos(\omega t + \alpha/2)\sin(-\alpha/2)$$

We see that the ratio $E_y(t)/E_x(t)$ is independent of time, namely that it is

$$E_y(t)/E_x(t) = -\tan \alpha/2. \tag{6.10}$$

We conclude that the resulting field is a linearly polarized oscillation at the angle $-\alpha/2$ with the x-axis.

6.2 Unpolarized Light

We have seen that the most general state of polarization of a wave is the elliptical polarization. However, natural light, such as the light we receive from the sun and the stars, is not usually polarized. Indeed, in the previous section, we considered a monochromatic wave, which has, in particular, an initial phase that is defined once and forever. Contrastingly, as we already discussed, the natural light emitted by *thermal sources* is a mixture of an enormous number of elementary waves, each due to the de-excitation of an atom or a molecule, each having a very short duration, on the order of nanoseconds or less. In other words, natural light waves have a finite coherence time Δt on the order of the duration of the elementary wavelets. As a matter of fact, the light emitted by each atom is in a certain state of polarization. As long as it lasts, there is a fixed phase difference between the two components of the field on the base states. But this is not the case with the total light wave, which is the sum of an enormous number of small waves, chaotically polarized independently of one another, each lasting about Δt. In practice, our instruments take an average on durations much longer than Δt. Under these conditions, we speak of *unpolarized light*. We understand that an ideal perfectly monochromatic wave ($\Delta t = \infty$) is always polarized.

The direction of the field of a linearly polarized light is constant in time. It is not constant for an elliptically polarized light, but it varies in a regular and predictable way. Contrastingly, the direction of the electric field of an unpolarized light varies chaotically in all the directions normal to the propagation direction. All these directions have the same probability. Intermediate cases are the states of *partial polarization*. In this case, there is a direction, modulo π, in which the probability of

finding the electric field is a maximum, while it is a minimum at 90° to it and intermediate in between We shall see in Sect. 6.8 how to distinguish the various cases operationally.

6.3 Dichroism

Dichroism is the property of certain materials, known as dichroic, to present different absorption coefficients to light polarized in the two base polarization states. We can distinguish *linear dichroism* when the absorption coefficients are different for light linearly polarized with directions at 90° to one another, and *circular dichroism* when the absorption coefficients are different between right and left circular polarized light. Here, we discuss linear dichroism, and will discuss circular dichroism in Sect. 6.9.

Some linear dichroic materials are suitable for preparing polarizing sheets of large dimensions (up to several square meters) and consequently are often used in practice. The most important example is the *polaroid*, which is now a common name, owing to it being a trademark of the Polaroid Corporation, produced in the form of a synthetic plastic film. This is one of the outstanding inventions of Edwin H. Land (USA, 1909–1991), who developed it over a period of a few years starting in 1929. In its original form, a polaroid contains submicroscopic crystals of herapathite, which is a salt of iodine and quinine (iodoquinine sulfate, to be exact). The crystals are shaped like needles, about one micrometer long and a dozen of picometers in diameter. The fact that the diameter is much smaller than the light wavelength is very important for minimizing the loss of light through scattering.

The crystals are embedded in a transparent nitrocellulose bath and aligned through stretching during the manufacturing of the film. Basically, one starts by preparing a colloidal dispersion of submicroscopic needles of herapathite in the form of a very viscous mass, which is then extruded through a long and narrow slit. A preferred direction now exists on the film, the crystallites having oriented themselves preferentially in the extrusion direction.

A second type of polaroid invented in 1938 by E. Land is a film of polymer molecules of polyvinyl alcohol (PVA) impregnated with iodine. Once more, the polymer chains are aligned during manufacturing through stretching.

The iodine atoms, present in both cases, have the essential role of delivering electrons that are free to move along the herapatite needles or the polyvinyl chains, making them similar to conducting wires.

We now choose a reference system on the film with the x-axis in the preferred direction of the "wires" and the y-axis normal to them. Consider a plane light wave normally incident on the film. The direction of its electric field $\mathbf{E}$ is, in any case, on the xy plane. The effects of its components E_x and E_y are different. The E_x component acts in the direction in which electrons are free to move over distances much longer than the wavelength. These electrons are accelerated by E_x absorbing energy from the wave. Colliding with other particles or impurities, the electrons re-emit

light in all directions, a very small fraction of which is forward. Consequently, the absorption coefficient is large. Contrastingly, E_y is practically unable to accelerate electrons, being incapable of moving in directions perpendicular to the crystals, or to the polymers, because their diameters are much smaller than the wavelength. Consequently, the absorption coefficient is small.

An ideal linear polarizer transmits without any absorption, namely unaltered, a normally incident plane wave linearly polarized in a certain direction, called the polarization axis of the polarizer, and completely absorbs a wave linearly polarized perpendicularly to the polarization axis. Notice that, with the term polarization *axis*, we mean a direction on the film, rather than a particular line. Namely, in an ideal polarizer, the ratio between the former and the latter absorption coefficient is zero. While such a perfect polarizer cannot be made in practice, the ratio of the absorbing coefficients of a good polarizer can be as small as 1/100 or even less.

In general, the ratio between the absorption coefficients depends on the wavelength. In particular, it grows sharply when the wavelength decreases to the order of magnitude of the diameter of the "wires".

A linear polarizer based on the same concepts can be easily prepared for electromagnetic waves of much longer wavelengths, for radio waves, for example. In this case, we can use an array of multiple straight conducting wires (real macroscopic wires now) stretched parallel to one another, having diameters and distances much smaller than the wavelength of the wave to be polarized.

As a matter of fact, Heinrich Hertz in, in his study of the electromagnetic waves that he had recently discovered, was the first, in 1888, to use such arrays to polarize and analyze the polarization state of the waves. He discovered that an array of copper wires of 1 mm diameter, spaced 3 cm apart, transmitted the incident radio waves of a 66 cm wavelength if the electric field of the radiation was perpendicular to the wires, but did not transmit them if the field was parallel to the wires.

6.4 Analyzers

We shall now start discussing the problem of analyzing the polarization state of a given light wave. Here, we consider linear polarization and non-polarization states, namely the linear polarization analyzers. We shall consider the other polarization states later in the chapter, after having discussed the relevant physical phenomena.

As a matter of fact, any (linear) polarizer is also a (linear) analyzer. Indeed, if we have a linearly polarized light wave normally incident, for example, on a polaroid with the electric field of amplitude E_0 at the angle θ with the axis of the polaroid, only the component $E_0 \cos \theta$ is transmitted. In the arrangement shown in Fig. 6.6, the first polaroid linearly polarizes the unpolarized incident light, while the second polaroid, whose axis is at the generic angle θ to the first one, analyzes the polarization state. If we vary the angle θ between the axes, the intensity of the light transmitted by the second polarizer varies as

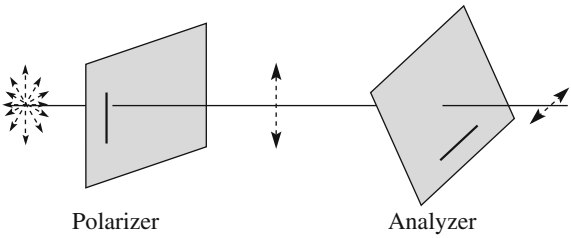


Fig. 6.6 Linear polarizer and analyzer set-up

$$I(\theta) = I_0 \cos^2 \theta, \tag{6.11}$$

where I_0 is the incident intensity (proportional to E_0^2). This is known as Malus' law, after Étienne-Louis Malus (France, 1775–1812), one of the scientists who made the greatest number of contributions to the study of light polarization phenomena.

In a typical experiment, such as the one we have just described, we use two instruments; the first one to prepare the polarization state (the polarizer), the second to analyze it (the analyzer).

To analyze the polarization state of a light wave, we can use a polaroid as an analyzer. We place it normally to the beam and rotate its axis, measuring the transmitted intensity as a function of the angle of the axis with a fixed direction. If the light is linearly polarized, we observe the transmitted intensity going through a maximum, a minimum, another maximum and another minimum, each separated by 90°. The intensity in the minima is compatible with zero. If the light is partially polarized, the intensity in the minima is not zero and that in the maxima is smaller than in the previous case. The *degree of polarization* is defined as the difference between maximum and minimum intensities, divided by their sum. If the light is unpolarized, the transmitted light intensity is the same at every angle. Notice, however, that we would observe the same behavior in the last two cases if the light were elliptically or circularly polarized, respectively. We shall see in Sect. 6.8 how these cases are distinguished.

Polaroid polarizers/analyzers are easily available in the form of polaroid sunglasses. They are produced because, as we shall discuss in the following sections, light reflected from surfaces, such as smooth water, but also a flat road, are generally horizontally polarized. This may create an annoying glare, which is eliminated by the polaroid glasses, whose "lenses" are linear polarizers with vertical axes.

6.5 Polarization by Scattering

In Sect. 5.9, we discussed the scattering of sunlight by molecules in the atmosphere or, at lower altitudes, by density fluctuations, and found the explanation given by Lord Rayleigh for the blue color of the sky. We come back to this process now,

because it also has the effect of polarizing light. It is an example of a phenomenon called *polarization by scattering*.

We briefly repeat here the arguments in Sect. 5.9. Consider an approximately monochromatic light wave incident on an atom. As we have already done several times, we think of the atom as a spherical cloud of negative charge with a massive point at the center of equal and opposite charge. Under the action of the harmonically oscillating electric field, the negative charges are displaced relative to the positive one, in the direction opposite to that of the electric field. This is opposed by the restoring forces internal to the atom. Under these conditions, the atom behaves as a forced oscillator in its stationary motion. Hence, the charge of the atom oscillates at the frequency of the incoming wave.

Each accelerating charge of the atomic oscillators produces an electric field, which, under the conditions we are discussing (here and in the rest of the chapter), is given by Eq. (3.42). We rewrite it here for convenience:

$$\mathbf{E}_{\text{rad}} = -\frac{q}{4\pi\varepsilon_0}\frac{\mathbf{a}_n(t - r/c)}{rc^2}. \tag{6.12}$$

We stress, in particular, that the *direction* of the electric field of the wave emitted by the oscillating charge is equal to that of the projection of its acceleration on the plane *normal to the line of sight*, which is the direction joining the atom with the observation point. Clearly, the acceleration itself has the direction of the electric field of the wave incident on the atom.

Consider a group of atoms at the origin of the axes and let us choose the z-axis in the direction, and sense, of the incoming wave. The direction of its electric field belongs, in any case, to the xy plane, and so, consequently, does the acceleration. Consider an observer looking at the charge from a direction perpendicular to that of the incident wave (in the z direction), say, for example, from the x direction. The electric field of the wave coming from the charge is then normal to the x-axis. On the other hand, the field is in the xy plane, and consequently, it must be in the y direction. Under these conditions, the light wave is linearly polarized. This process is called polarization by scattering.

An everyday example is the light coming from the sky, which is the sunlight scattered by the air molecules, or density fluctuations. If we look in a direction in which light comes to us after having being scattered by 90°, we "see" it completely linearly polarized normally to the scattering plane. Figure 6.7 shows two of these directions, along with a direction of non-polarization and one of partial polarization. In addition, the figure shows the direction of the primary rainbow. In this case, the scattering is by water droplets, but the arguments we have developed still hold. One sees that the angle is such that the polarization is almost complete.

The phenomenon cannot be observed with the naked eye, because our eyes are not very sensitive to polarization. We must look through a polarization analyzer, using, for example, a pair of polaroid sunglasses. Remember that their transmission axis is vertical. If we take the sunglasses in our hand and look through them towards the sky in one of the directions at 90° with the sun, as shown in Fig. 6.7,

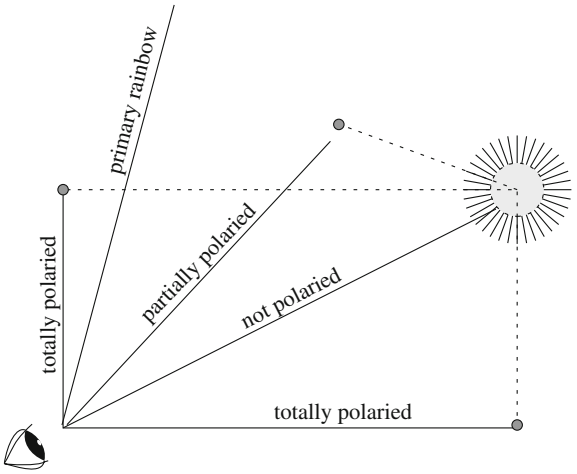


Fig. 6.7 Sunlight scattered by the atmosphere through different angles, and corresponding polarization. The direction of the primary rainbow (Sect. 6.4) is also shown. Linear polarization is normal to the page

and then turn the sunglasses, we observe, in a complete turn, two intensity maxima and two minima separated by 90°, according to Malus' law. If we look in directions corresponding to scattering angles other than 90°, the light is partially polarized. At a zero scattering angle (taking care not to look directly into the sun), the light is not polarized. Similar observations can be done with the light of a rainbow.

Several animals, including bees, ants and a number of crabs, have eyes sensitive to polarization and exploit the polarization of the sky for purposes of for orientation. At the end of the 1940s, the Austrian zoologist Karl von Frisch (Austria, 1866–1982) observed that bees were able to orient themselves even when the sun was not visible. A small clear sky area was enough for the insects. As proof, von Frisch inserted a polarizing filter between the eyes of the insects and the skylight. He observed that when he turned the polarizer, the insects' sense of direction was systematically altered. In this way, he was able to demonstrate that it was just the light polarization being exploited by the bees. Other scientists observed similar behavior in other animal species after him.

6.6 Polarization by Reflection

A third phenomenon capable of polarizing light is the *reflection*. Consider, for example, two media, air and glass or air and water, separated by a flat surface, as shown in Fig. 6.8. We can consider an unpolarized ray incident on the interface as a superposition of two components, each linearly polarized in a direction at 90° to one another, one with the electric field normal and one parallel to the incidence plane (Fig. 6.8a and b, respectively).

In the situation shown in Fig. 6.8a, the field of the incident wave oscillates normally to the incidence plane, which is the plane of the figure. The sources of the

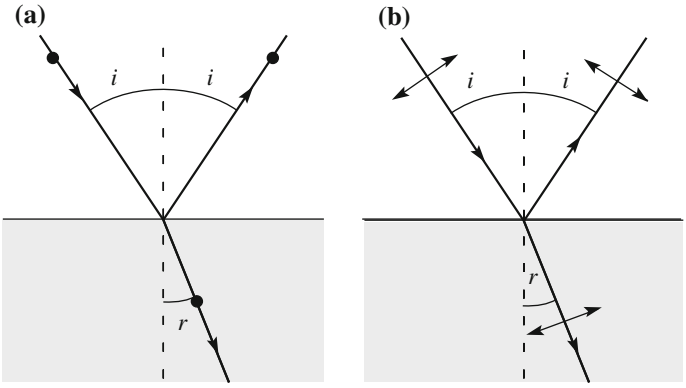


Fig. 6.8 Incident, refracted and reflected rays linearly polarized **a** normally to the incidence plane, **b** parallel to the incidence plane

refracted and reflected waves are the *oscillating charges in the second medium*, namely the glass (or the water). Being that they oscillate as forced to by the incident wave, the direction of oscillation is normal to the plane of the figure as well. This is then also the direction of the electric field of the refracted and reflected waves. The conclusion is independent of the incident angle and, consequently, we do not expect anything peculiar to happen when it varies.

Consider now the component of the incidence wave with the electric field oscillating in the incidence plane, namely the plane of the figure, as in Fig. 6.8b. Clearly, the oscillation direction of the molecules in the glass is in the same plane, but in which direction in this plane? To answer the question, we must fix our attention on the refracted wave. Its electric field is not only in the plane of the figure, but also normal to the propagation direction of the wave. But this must also be the direction of the oscillations of the atomic charges, being that it is parallel to the field in the glass (Fig. 6.8b). Considering now the field of the reflected wave, and remembering Eq. (6.12), we see that it is smaller the smaller the component of the molecule acceleration normal to its propagation direction.

Consider, in particular, the incident direction for which the angle between refracted and reflected rays is exactly 90°, as shown in Fig. 6.9. This is called the *Brewster angle*, after David Brewster (Scotland, 1781–1868), who discovered the effect in the first years of the XIX century. Clearly, the intensity of the reflected ray is zero under these conditions, because the projection of the accelerations of the charges originating the reflected wave on the normal to that direction is null. Under these conditions, the reflected light, from unpolarized incident light, is completely polarized in the direction perpendicular to the incidence plane. For incident angles different from the Brewster angle, the reflected wave polarization is partial.

The Brewster angle, θ_B, is easily calculated starting from the condition $i + r = 90°$. If we call n_1 and n_2 the refractive indices in the first and second medium, respectively, the Snell law gives us $n_1 \sin i = n_2 \sin r$. For $i = \theta_B$, we have

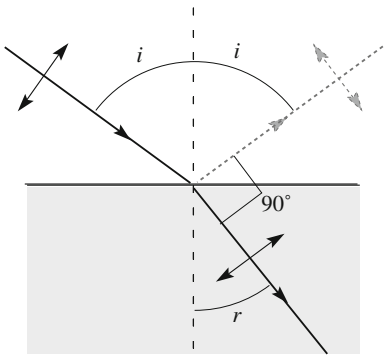


Fig. 6.9 Brewster angle conditions; polarization in the incidence plane

$$n_1 \sin\theta_B = n_2 \sin(90^\circ - \theta_B) = n_2 \cos(\theta_B),$$

which gives us

$$\theta_B = \arctan(n_2/n_1). \tag{6.13}$$

The Brewster angle for the air-glass interface is 56°. The phenomenon can be observed by looking at images reflected on a glass surface, like that of a window or of a showcase, through a pair of polaroid sunglasses and rotating the glasses. A more quantitative check can be done by measuring the intensity of the reflected ray as a function of the incident angle of a beam of light linearly polarized in the incidence plane. One notices a clear minimum at the Brewster angle. Repeating the measurement with an incident beam polarized normally to the incidence plane, no minimum is found.

QUESTION Q 6.1. Calculate the Brewster angle at air-water and water-glass interfaces. □

6.7 Birefringence

Up to now, we have implicitly considered the concepts of the polarization of a wave and the polarization of the corresponding light ray as being equivalent. However, the two concepts are different in principle and coincide only in isotropic media, not in anisotropic media, as we shall now discuss. We start by recalling that, in Chap. 4, we defined and discussed, for a monochromatic wave, the concepts of phase velocity and group velocity. We saw them to be different from one another in a dispersive medium. We were then considering isotropic media, in which we found that the two velocities generally differ in magnitude. In an anisotropic medium, the two velocities generally differ in *direction* as well.

Let us fix our attention on these directions, considering a progressive plane monochromatic wave. The direction of the phase velocity at every point is normal

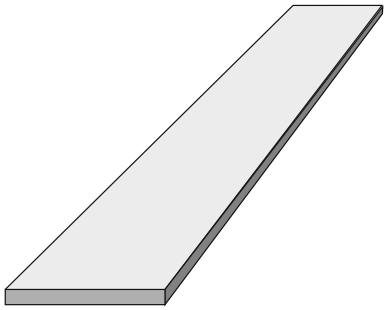


Fig. 6.10 An elastic slab, a birefringent medium for elastic waves

to the wave surface through that point. Indeed, moving along the phase surface, the phase does not vary by definition. The direction of the group velocity is the direction of propagation of the energy transported by the wave. In an isotropic medium, energy propagates normally to the wavefronts. Indeed, this is what intuition suggests, but it is valid only in an isotropic medium.

In this section and the next, we shall analyze the behavior of plane monochromatic and linearly polarized waves progressing in an anisotropic medium. We consider, in particular, the *birefringent* (or *birefractive*) media, in which the phase velocity (or, equivalently, the refractive index) depends on the propagation direction and the state of linear polarization. These media have two refractive indices, one for each properly chosen linear polarization base state.

We start with a mechanical example of birefringent medium for elastic waves. Figure 6.10 shows a long elastic slab, several centimeters wide and a few millimeters thick. Clearly, the restoring forces for a given displacement are much larger in the horizontal than in the vertical direction. Consequently, the phase velocity for vertically polarized waves is much smaller than that for horizontally polarized ones of the same frequency. The situation is similar for electromagnetic waves in birefringent media.

The analysis in Sect. 4.7 showed us that when a plane light wave travels in a transparent dielectric medium, its electric field induces an oscillating dipole moment **p** in the atoms of the medium. The oscillating dipoles emit electromagnetic waves. The sum of all these waves and of the incident one is still a plane wave, as is the incident wave alone, but its phase velocity is not c but rather $\upsilon_p = c/n$, where n is the refractive index. In Sect. 4.7, we were considering an isotropic molecular charge distribution. As a consequence, the displacement of the charges that we called **x** was parallel to the direction of the field of the incident wave. The same was true for the dipole moment $\mathbf{p} = q\mathbf{x}$ induced in every molecule and for the induced dipole moment per unit volume, namely the polarization,[1] $\mathbf{P} = n_p\mathbf{p}$ dove n_p is the number of molecules per unit volume.

[1]Unfortunately, the world "polarization", used in physics, has several different meanings. In this chapter, we are dealing with the "polarization" of light, which we have defined. In this sentence and in the following discussion, "polarization" is also the electric dipole density in the dielectric.

Consider now a non-isotropic molecule. Think, for example, of a diatomic one, made of a negative charge distribution around two massive point-like positive charges at a certain distance from one another. We can imagine that the deformation of the molecule due to a certain value of an applied electric field should be different if the field is in the direction of the line of the two nuclei (the bound) or normal to that. And indeed, it is so. Under these two conditions, when the applied field is parallel or normal to the bond, the displacement is in the direction of the field. However, this is only the case for these directions. If the field is neither normal nor parallel to the bond, the displacement has a direction different from that of the field.

This fact can be easily understood considering the mechanical model shown in Fig. 6.11a. A bead is connected to a rigid standing frame by two pairs of springs. The elastic constants of the springs of each pair are equal to one another, but those of the two pairs are different, say, k_1 in the x-direction and k_2 in the y-direction. Consider, for example, that $k_2 = 2k_1$. If we now apply the electric field force at, say, an angle of 45° with the axes, the displacement component in the x-direction will be twice as large as that in the y-direction, as shown in Fig. 6.11b. The resultant displacement direction is different from that of the electric field. We see that only in the direction of one of the spring pairs do the field and displacement have the same direction. In general, the displacement and induced dipole moment **p** have a direction different from **E**.

Even if the majority of the molecules are not isotropic, the anisotropy usually does not appear at the macroscopic level. This is because in gases, liquids and amorphous materials, the directions of the molecules are chaotically distributed. Consequently, even if the induced dipole moment **p** in each of them is not parallel to **E**, the dipole moment per unit volume **P** is parallel to **E**. Indeed, the components normal to **E** add up to zero. As a matter of fact, this happens in crystals too, when the molecules are symmetrically arranged, as in cubic crystals.

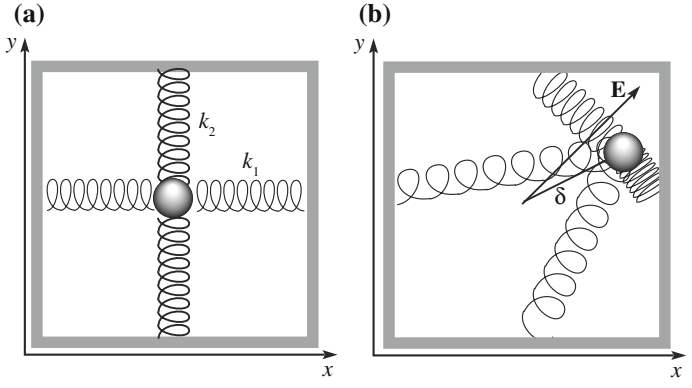


Fig. 6.11 **a** Mechanical model: a bead and two pairs of springs, with elastic constants $k_2 = 2k_1$. **b** displacement for an electric field at 45°

The media we shall deal with can be considered, in any case, to be linear dielectrics, in which P is proportional to E. We now consider macroscopic uniaxial crystals, namely crystals having a single symmetry axis. In this case as well, an "axis" means a direction. A typical case is calcite, which is a calcium carbonate ($CaCO_3$) whose crystal cell is a hexagonal right prism. The symmetry axis of the prism is called the optical axis. Under these conditions, the macroscopic polarizability is different along the optical axis on one side and in any perpendicular direction on the other. If the field is parallel to the axis, the polarization **P** is parallel to **E** with the proportionality constant, say $\varepsilon_0\chi_p$. Namely, we have

$$\mathbf{P} = \varepsilon_0\chi_p\mathbf{E}. \tag{6.14}$$

If **P** is normal to the axis, it is still parallel to **E**, but with a different proportionality constant, say, $\varepsilon_0\chi_n$, namely we have

$$\mathbf{P} = \varepsilon_0\chi_n\mathbf{E}. \tag{6.15}$$

Hence, the medium has two electric susceptibilities, χ_p and χ_n. For reasons of symmetry, it is clear that χ_n is independent of the direction in a plane normal to the axis. As in the mechanical example we gave, if **E** is neither parallel nor perpendicular to the axis, the polarization **P** is not in the direction of **E**.

An interesting example is the Iceland spar, which is a transparent calcite crystal. Its electric susceptibilities for light are $\chi_p = 1.21$ and $\chi_n = 1.74$. Consider, for example, the wave electric field to be directed **E** at 60° to the axis. The components of the polarization vector are $P_p = \varepsilon_0\chi_p E\cos 60° = 0.61\varepsilon_0 E$ and $P_n = \varepsilon_0\chi_n E\sin 60° = 1.51\varepsilon_0 E$. Hence, the angle of **P** to the axis is arctan(1.51/0.61) = 68°30′. The angle between **P** and **E** is 68°30′–60° = 8°30′, as shown in Fig. 6.12.

We now come back to the general discussion and recall that the refractive index, hence the phase velocity, is directly connected to the polarizability. The consequence is that the refractive index in a uniaxial medium depends on the linear polarization state. We shall now determine the structure of a progressive plane monochromatic light wave in a uniaxial medium. The arguments will be similar to those we followed in Sect. 3.6 to study the structure of such a wave in a vacuum. We refer the reader to Sect. 10.7 of the third volume of the course for the Maxwell equations in a linear dielectric medium. We shall, however, revisit the relevant expressions here.

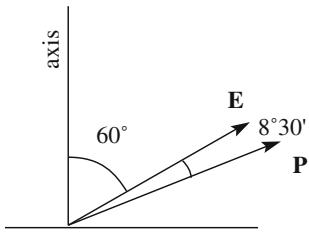


Fig. 6.12 Direction of **P** for **E** at 60° with the axis in an Iceland spar crystal

The relevant fields of an electromagnetic wave in a dielectric are the electric field **E**, the magnetic field **B**, as in a vacuum, and, in addition, the auxiliary field **D**. The latter is called the electric displacement and is given by

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}. \tag{6.16}$$

Under the conditions we are discussing, there are no free charges, but there are polarization charges. We recall that the latter are not sources for **D**, but, as all the electric charges are, are sources for **E**. Consequently, we have

$$\nabla \cdot \mathbf{D} = 0 \tag{6.17}$$

and

$$\nabla \cdot \mathbf{E} \neq 0. \tag{6.18}$$

As for the magnetic field, its divergence is zero, as always, namely

$$\nabla \cdot \mathbf{B} = 0. \tag{6.19}$$

In addition, under all practical circumstances, **B** is the same in a vacuum and in a dielectric, being that the magnetic susceptibility is very close to one. In the absence of conduction currents, as is the case here, the curl of the magnetic field is given by

$$\nabla \times \mathbf{B} = \mu_0 \frac{\partial \mathbf{D}}{\partial t}. \tag{6.20}$$

There are four more vectors to consider. Two of them, the wave vector **k** and the phase velocity $\mathbf{v}_p$, are perpendicular to the wave surface in the direction of the propagation. The group velocity $\mathbf{v}_g$ has the direction of the energy propagation, which, for an electromagnetic field, is the Poynting vector

$$\mathbf{S} = \varepsilon_0 c^2 \mathbf{E} \times \mathbf{B}. \tag{6.21}$$

Let us consider, as we did in Sect. 3.6, a plane monochromatic wave. Its field can be written as

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})}, \quad \mathbf{D} = \mathbf{D}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \alpha)}, \quad \mathbf{B} = \mathbf{B}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \beta)}, \tag{6.22}$$

where we have allowed for the presence of initial phase differences between the fields (even if it is irrelevant for our arguments).

Equation (6.18) gives us

$$\mathbf{k} \cdot \mathbf{E} \neq 0. \tag{6.23}$$

As opposed to in a vacuum, the electric field is *not* perpendicular to the wave vector, or to the phase velocity.

On the other hand, Eqs. (6.19) and (6.20) give us

$$\mathbf{k} \cdot \mathbf{B} = 0, \quad \mathbf{k} \times \mathbf{B} = -\mu_0 \omega \mathbf{D}. \tag{6.24}$$

The magnetic field is, as in a vacuum, perpendicular to the wave vector, but as opposed to in a vacuum, is perpendicular to the electric displacement **D** rather than to **E**. Equation (6.24) also tells us that **D** is perpendicular to **k**, hence its direction belongs to the wave surface.

In conclusion, all four vectors **D**, **E**, **k** and **S** are perpendicular to **B**, and consequently belong to the same plane. It is thus convenient to look at them in this plane, which is the plane of Fig. 6.13. In the figure, *AA* is the wave surface, on which the vector **D** lays, as we have said. **E** and **D** have generally different directions. Let ϕ be the angle between them. The wave vector **k** is perpendicular to **D** and **S** is perpendicular to **E**. Consequently, the angle between **k** and **S**, which is the angle between the wave and group velocities, is ϕ as well.

Namely, the angle between the wave and group velocities is

$$\phi = \arccos \frac{\mathbf{E} \cdot \mathbf{D}}{|\mathbf{E}||\mathbf{D}|}.$$

After the above discussion, it is clear that we must distinguish the concepts of the linear polarization of a wave and the linear polarization of a ray, in which a ray is the trajectory of energy. When we talk of a monochromatic plane wave, we think of the propagation of a wave surface, like *AA* in Fig. 6.13. The direction of its linear polarization is the direction of **D** that belongs to that plane. If the wave is linearly polarized, the direction of **D** does not vary. The concept of polarization of the ray is different, because we are considering the energy propagation. The linear polarization of the ray is the direction of **E**, which is normal to that propagation direction.

As we have already noticed, particular cases exist in which the deformation of the molecules, and consequently the polarization **P**, is parallel to **E**. In these cases,

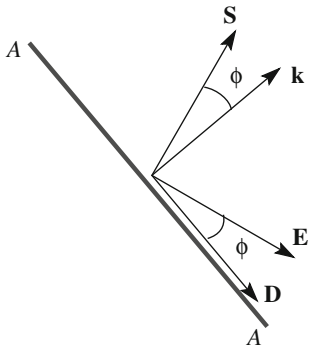


Fig. 6.13 The four vectors **D**, **E**, **k** and **S** in a plane perpendicular to **B**. *AA* is a wave surface

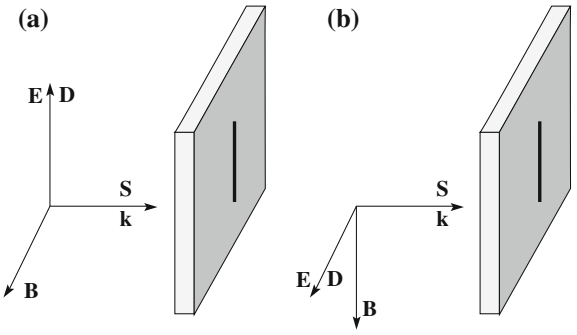


Fig. 6.14 The fields of a plane monochromatic wave with electric field **a** parallel to the axis, **b** normal to the axis

D is parallel to **E** as well, for Eq. (6.16). This is the case when the electric field has the direction of the optical axis (see Fig. 6.14a) or any direction perpendicular to the axis (see Fig. 6.14b). In all other cases, the directions of **E** and **D** are different.

Consider a uniaxial crystal cut like a plate, with faces parallel to the optical axis and a monochromatic plane light wave normally incident on its face, as in Fig. 6.15. There are two cases that are the most relevant: (a) the polarization direction is parallel to the optical axis, (b) the polarization direction is normal to the optical axis. In the former case, **E** and, consequently, **D** are parallel to the axis, while in the latter case, they are normal to the axis. In these two particular cases, wave polarization and ray polarization coincide. This is not true for any other polarization direction.

In the two cases, the polarizability being different, the refractive indices (and the phase velocities) are also different. Let υ_p and $n_p = c/\upsilon_p$ be the phase velocity and index in case (a) and υ_n and $n_n = c/\upsilon_n$ the phase velocity and index in case (b). Note that the footers n and p mark the orientation (normal or parallel) of the electric field relative to the axis.

We shall come back to the crystal faces parallel to the axis in the next section. Here, we consider a crystal whose faces have been cut neither parallel nor perpendicular to the optical axis. The plane of Fig. 6.15 is the plane made by the optical axis and the incidence direction. The latter is perpendicular to the face of the crystal. The direction of the optical axis is the oblique segment in the figure.

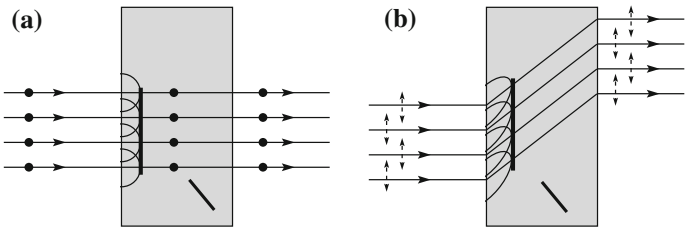


Fig. 6.15 Plane monochromatic wave incident on a uniaxial plate. The optical axis is in the plane of the figure and is shown as an oblique segment. Electric field **a** normal, **b** parallel to the plane of the optical axis

In Fig. 6.15a, the polarization of the incident wave is perpendicular to the plane of the figure. Applying the Huygens-Fresnel principle, we can consider every point of a wave surface in the crystal as a source of secondary wavelets. Each segment of a wavelet propagates in a different direction, but, in any case, normally to the axis. Consequently, all of them have phase velocity v_n. The trace of the wavelet in the plane of the figure is thus a semicircle. The envelope of the wavelets is, consequently, parallel to the wave surface (the surface of equal phase) and the ray has the same direction as the incident ray. Under these conditions, the refracted ray behaves "normally", namely following Snell's law, even when the incidence angle is different from zero, as can be shown. For this reason, it is called the *ordinary ray*.

Consider now the situation in Fig. 6.15b, in which the incident wave is linearly polarized in a direction belonging to the plane defined by the optical axis and the incidence direction. We again consider the points of the wavefront in the crystal as sources of secondary wavelets. However, the different segments of a wavelet now propagate at different angles with the optical axis and, consequently, with different phase velocities. The phase velocity is v_p for the segments having the field **D** (and **E**) parallel to the axis, which, in an argument that requires we pay close attention, propagate perpendicularly to the axis. It is v_n for the segments having the field **D** (and **E**) normal to the axis, which propagate in a direction parallel to the axis. The phase velocity has intermediate values for the other segments. A more in-depth analysis shows that the secondary wavelets are rotation ellipsoids. Figure 6.15b is drawn in the case for $v_p > v_n$. We obtain the new wave surface by taking the envelope of the wavelets, which is in the plane of **D** and **B**. We see that it is parallel to the wave plane of the incident wave. The figure also shows the rays, namely the energy trajectories. We see that the rays are not perpendicular to the wave surfaces. Indeed, the direction of the ray is the direction of $\mathbf{E} \times \mathbf{B}$, and **E** is not parallel to **D**. Under these conditions, the refracted ray does not follow Snell's law and is called the *extraordinary ray*. This phenomenon is called *anomalous refraction*. Under these conditions, the directions of the phase and group velocities are different.

Figure 6.15 shows how the waves and the corresponding rays exist from the plate after having crossed its thickness. If the incident ray is not polarized, it splits into an ordinary and an extraordinary ray. The two rays, polarized at 90° with one another, exit separately from one another, if the thickness of the plate is large enough. We can easily verify their polarizations with an analyzer.

Question Q 6.2. Draft the corresponding diagram in Fig. 6.15 b) for $n_p > n_n$. □

An important example of a uniaxial crystal, which we have already mentioned, is the Iceland spar, a transparent form of calcite. The refractive indices of the ordinary and extraordinary rays are $n_o = 1.658$ and $n_e = 1.486$, respectively (for yellow light, to be precise). Such stones can be found in Iceland, as the name suggests. According to a possibly true but not historically proven tale, Vikings used Iceland spar crystals for navigational purposes. When the sun was invisible, being under the horizon or covered by clouds, but some open spot of sky was available, they measured the polarization of the light from that region of the sky and inferred the position of the sun, much like the bees do.

A curious effect of birefringence has been observed for centuries and used for entertaining in living rooms. Crystals of Iceland spar placed over an image would show it as being doubled (you can find nice images of that on the web).

Polarizers can be built by blocking one of the two rays. An historically important example is the Nicol prism invented by William Nicol (Scotland, 1779–1851) in 1828. It is made of two prisms of calcite, smartly arranged in a way so that the extraordinary ray is transmitted while the ordinary ray is deflected at an angle via total internal reflection.

6.8 Phase Shifters

Let us go back and consider a plate of a birefringent medium whose faces contain the direction of the optical axis. Consider a linearly polarized plane monochromatic wave normally incident on the plate. Let z be this direction. We know that the phase velocity in the crystal is v_p if the electric field **E** of the wave is parallel to the optical axis, v_n if **E** is normal to the axis. This happens in any case in which the polarizability of the medium has a preferential direction, not only in the uniaxial crystals. For example, the plastic films containing long molecular chains, which are arranged with a certain degree of orientation as a result of the film having been produced by extrusion or by passing it between rollers, are birefringent. A preferential direction can also be easily induced in a plastic film simply by pulling it in that direction.

What happens if the electric field of the polarized wave is at a generic angle θ with the optical axis? To answer the question, we must consider the components of **E** parallel and normal to the axis separately, namely $E_p = E \cos\theta$ and $E_n = E \sin\theta$. Initially, when they enter the medium, the phases of the two components are equal, but they become different, with a difference increasing with the path in the medium, because the two phase velocities are different. After having traveled a path Δz, the phase difference is

$$\Delta\phi = \Delta z \cdot k \cdot \left(n_p - n_n\right) = \Delta z \cdot \frac{2\pi}{\lambda} \cdot \left(n_p - n_n\right), \tag{6.25}$$

where k is the wave number and λ is the wave length in a vacuum. As a superposition of two linearly polarized waves at 90° to one another with different phases, the resultant wave is elliptically polarized, as we learned in Sect. 6.1.

Clearly, if Δz is the thickness of the plate, $\Delta\phi$ is the phase difference between the two linearly polarized components at the point of exit from the plate, so that we can obtain any polarization state starting from linear polarization, by suitably choosing both Δz and θ.

For example, if we want a circular polarization we chose $\theta = \pm 45°$ (the sign determines the handiness of the polarization), in order to have the two components of the field on the axes be equal to one another, and Δz such that $\Delta\phi$ is equal to $\pi/2$. Such a plate gives a relative shift of the phases of the two components of a quarter

of a period and is called a *quarter-wave plate*. We understand that if we want circularly polarized light starting from unpolarized light, we must use, in sequence, a linear polarizer (for example, a polaroid) and a quarter-wave plate with its axis at 45° to the axis of the polaroid. Depending on whether this angle is +45° or –45°, we obtain the two circular polarization states. Circular polarizers, namely sandwiches of the two elements glued together, are commercially available. Notice that they work properly at a given wavelength, as one immediately understands by looking at Eq. (6.25).

It is useful to look at the orders of magnitude. We have seen that the indices of the Iceland spar are $n_o = 1.658$ and $n_e = 1.486$, for yellow light. The difference, which is called *birefringence*, is $\Delta n = n_o - n_e = 0.17$ and is one of the largest amongst crystals. The mica, which is a sheet polysilicate, has a birefringence $\Delta n = 3.3 \times 10^{-3}$. Polymethil metracrylates (PMMA) (of which Perspex, Lucite, Plexiglas, etc., are trade names) are transparent plastic plates. The plates have a small birefringence with their axis normal to the face, as a result of the production process. This is because the plastic mixture, still in a viscous state, is poured into basins having the shape of the plate being produced. In the subsequent hardening process, the material contracts, mainly in the vertical direction normal to its free surface, becoming optically anisotropic. The difference between indices is very small, on the order of $\Delta n = 2 \times 10^{-5}$. Hence, a quarter-wave plate for, say, $\lambda = 0.5$ μm, if made of calcite, should have a thickness $\Delta z = 0.74$ μm, which is too small to be practical. Using mica instead, we need the thickness $\Delta z = 38$ μm, which is easily feasible. For this reason, mica was widely used for the scope until the appearance of plastic materials, which are much cheaper. In the somewhat extreme case of the PMMA, the thickness of a quarter-wave plate is as large as $\Delta z = 6.25$ mm.

Figure 6.16 schematically shows the set-up of the crossed polarizers, which is quite useful for studying the birefringence of transparent samples. On an optical bench, one arranges two linear polarizers, two polaroids, for example, with their axes at 90° to one another. The light source may be a common one that produces unpolarized light, but it should be (approximately) monochromatic. The first polarizer produces light linearly polarized at 90° with the axis of the second, which does not let any light through. The two devices may be equal, but their function is different, the first acting as a polarizer, the second as an analyzer. If we now insert a birefringent material, for example, a cellophane or PMMA film, as in Fig. 6.16, we see some light coming through the analyzer. If we rotate the plastic film around the axis of the system, we see two directions at 90° to one another at which no light is transmitted. This happens when the optical axis of the sample is parallel to the axis of the polarizer or to that of the analyzer.

In order to learn more from this type of experiment, we observe that the changes induced on the linearly polarized light by the sample depend on two factors: the angle θ between the electric field $\mathbf{E}$ of the wave and the optical axis of the sample and the phase shift $\Delta\phi$ given by Eq. (6.25). The polarization state of the transmitted light after the sample can be one of the following.

(a) Light is still linearly polarized but possibly in another direction. We can determine if this is the case by turning the analyzer. We should find two positions of maximum intensity 180° from one another, separated (at 90° from each of them) by minima of *zero* intensity.
(b) Light is elliptically polarized. Acting as in the previous case, we should find a qualitatively similar result, but now the intensity in the minima is not zero. The intensity of transmitted light never vanishes.
(c) Light is circularly polarized. Turning the analyzer, the transmitted intensity is constant.

In each case, we can, in an equivalent manner, turn the sample instead of the analyzer.

Since the phase shift $\Delta\phi$ depends on the wavelength, when we operate with a white source and look at the sample through the analyzer, we see it colored. If its thickness is uniform, within a fraction of a wavelength, its color should be uniform as well. There are, however, almost always small thickness variations from point to point, and we see bands of different colors that are the loci of equal thickness. Similar images are seen when the stresses, which are often present as well, have caused small variations in the refractive indices across the sample. This phenomenon is called *stress induced birefringence*. As a matter of fact, one can study the stresses in parts of buildings and in engines making plastic models of these parts and observing them under a load between crossed polaroids. In this way, one learns, even if in a qualitative way, where the stresses are most important.

A spectacular show can be achieved by preparing the environment necessary to grow a crystal and then observing it grow between crossed poloaroids. Shadows and lights of different brilliant colors evolve before your eyes as the crystal takes form.

The device we have discussed essentially consists of an element that prepares the incident beam in a state of known polarization, a birefringent sample to be studied, which is crossed by the beam, and an analyzer to determine how the sample has changed the state of polarization.

A similar set-up, with a birefringent sample of known characteristics and the light incident on it of unknown polarization, allows us to determine the latter. Consider, in particular, the birefringent sample being a quarter-wave plate, with an axis of known direction. By turning the plate, we can distinguish whether the light is elliptically polarized (circularly, in particular) or partially polarized (or not polarized at all).

Another interesting phenomenon is the birefringence induced by an electric field, which is called the Kerr effect, after John Kerr (Scotland, 1824–1907), who discovered it in 1875. Consider a transparent liquid whose molecules have a sizeable permanent electric dipole. When we apply an electric field to the liquid, the dipoles tend to orient in the direction of the field, and the medium becomes optically anisotropic. The effect can be observed by placing the liquid, once again, between crossed polarizers, as in Fig. 6.17.

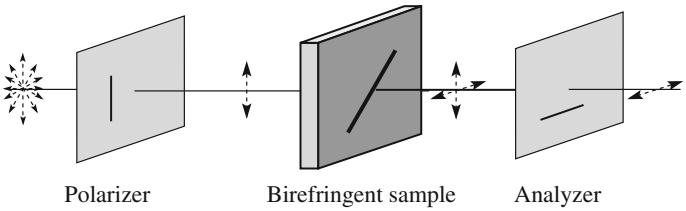


Fig. 6.16 Analyzing the birefringence with crossed polarizers

With this arrangement, we can study how the index difference $\Delta n = n_o - n_e$ varies as a function of the applied electric field intensity E. As a matter fact, the direct observable is the induced phase difference shift $\Delta\phi$, which is found to be proportional to the applied field intensity squared, namely as

$$\Delta\phi = \Delta z 2\pi\beta E^2, \tag{6.26}$$

where Δz is the length of the liquid in the field and β is a constant characteristic of the liquid (and of its thermodynamic state) called the Kerr constant (the factor 2π is there for historic reasons). A liquid with a very high Kerr constant is nitrobenzene, for which $\beta = 2.4 \times 10^{-12}$ mV^{-2}.

Consider, for example, a capacitor $\Delta z = 10$ cm long, having nitrobenzene between the plates. What is the field needed for the system to act like a quarter-wave plate? This is $E = (4\beta\Delta z)^{-1/2} = 1$ MV/m, which is not a difficult value to achieve. For example, if we build the capacitor with a 1 mm gap between the plates, we should apply a potential difference of 1000 V between them.

An important application of the Kerr effect is in transforming oscillations of an electric field into oscillations of light intensity. Indeed, the effect is quite fast, because the typical time taken by the molecules to orient in the varying field is on the order of the nanosecond.

6.9 Optical Activity

Optical activity is the property of certain substances to rotate the polarization direction of a linearly polarized light wave that crosses them. This phenomenon was discovered in 1811 in quartz by François Arago.

Consider again the set-up of crossed polarizers and a sample of the optically active substance under study between them. Under these conditions, we observe light going through the analyzer (which, we must remember, has its axis at 90° with the polarizer). If we now turn the axis of the analyzer, we reach an angle, say, α, at which the light extinguishes. We also find that the α is proportional to the thickness of the medium Δz as $\alpha = \rho\Delta z$, where the proportionality constant ρ is a characteristic of the substance and its thermodynamic state, called the *specific rotation*

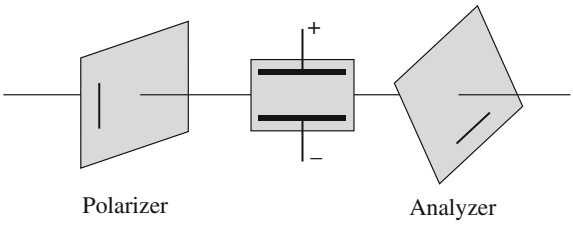


Fig. 6.17 Arrangement of a crossed polarizer for studying the Kerr effect

constant. "Specific" because it is per unit length, "rotation" referring to the plane of the linear polarization. Its usual units are the degree per millimeter.

The phenomenon shows dispersion, namely that ρ is a function (generally decreasing) of the wavelength. For several substances, the phenomenon is conspicuous, with values of ρ on the order of a dozen degrees per millimeter. Particularly active are the solutions of several organic molecules in water, notably the sugars. For the solutions, the rotation of the polarization direction is proportional to the concentration c of the solute, namely $\alpha = \gamma c \Delta z$, where γ is characteristic of the solute, depending on its thermodynamic state. The optical activity of the sugars is employed to control their production, while they are still in the solution, with the optical saccharometers. These instruments determine the sugar concentration in the solution, for example, the glucose in the molasses obtained from sugar canes or beets, from the angle of rotation of the polarization direction of a light beam.

The common property of the optically-active substances is the presence of molecules, or elementary cells if they are crystals, that are not mirror symmetric. This means that the shape of the molecule, or cell, and its mirror image are different, as with, for example, a hand or a screw. For this reason, they are called *chiral molecules*, from the Greek "chir" for hand. Chiral molecules come in two different types (mirror images of one another) called *optical isomers* and denoted with the symbols D (for dextro = right) and L (for laevo = left). A substance with right-handed molecules, namely D, rotates the polarization direction to the right, while a substance with left-handed molecules rotates it to the left. By "toward right", we mean that an observer looking at the light coming toward him/her must turn the analyzer to the right, meaning clockwise, to extinguish the light.

Optical activity exists both for liquids and solid state substances. There is, for example, the levo-quartz and the dextro-quartz, the levotartaric acid and the dextrotartaric acid, etc. The chemical and physical properties of the pair at the macroscopic level are equal, except for the fact of rotating light polarization in opposite directions.

The inorganic optical isomers (like, for example, the dextro- and levo-quartz) are found in nature with the same abundance. This is a consequence of the fact that the forces in action when the crystal (or the molecule) is formed, which are electromagnetic, are invariant under the inversion of the reference axes. Contrastingly, the optically active organic molecules synthetized by the living organisms are found as only one of the isomers, while an artificial synthesis process produces both isomers

in equal amounts. Indeed, in 1849, Louis Pasteur (France, 1822–1895) discovered that the living organisms produce tartaric acid only in the *D* form, while in his laboratory, the chemical synthesis process resulted in a 50/50 mixture of the two isomers. He also discovered the existence of microorganisms capable of feeding and metabolizing only one of the two optical isomers. More examples are the above-mentioned biological glucose that is in the *D* form (called dextrose) and the biological proteins of all living organisms that are made of amminoacids, which are always left isomers. All the DNA molecules are right helixes. How this happened in the evolutionary process of life is clearly a fundamental question, to which, however, we do not yet have an answer.

Let us now give an explanation for the correlation between molecular handiness and optical activity. We start by recalling that a linearly polarized state is the superposition of two circularly polarized states, one left and one right. Consider now a substance that is circularly birefringent, namely that has two different refractive indices (or phase velocities), one for left circularly polarized light, one for right circularly polarized light. Consider a plane monochromatic linearly polarized light wave entering into such a medium. We can think of it as a superposition of two circularly polarized waves with a certain phase difference. While advancing, the phases of the two components proceed at different speeds, and consequently, a phase difference $\Delta\phi$ develops between them, proportionally increasing with the path in the medium. However, as we learned in Sect. 6.1, the superposition of two opposite circular polarization states having a phase difference $\Delta\phi$ is a plane wave with the electric field direction forming the angle $\Delta\phi/2$ with the x-axis. We can conclude that the specific rotation, i.e., the rotation per unit path, is proportional to the difference between the refractive indices for right and left circular polarizations.

Let us now try to understand the molecular origin of the effect, considering, for example, a molecule having the shape of a helix, as is the case with several organic molecules, such as that in Fig. 6.18. Let us consider a circularly polarized wave advancing in the direction of the molecular axis, which we choose as the z-axis. Clearly, in one of the two circular polarization states, the electric field direction changes along the molecule following the direction of the molecular "wire", and consequently acts efficiently on the electrons and in regard to stretching the structure. For this state, the electric polarizability is large. The same does not happen for the opposite circular polarization state, and consequently, the polarizability is smaller.

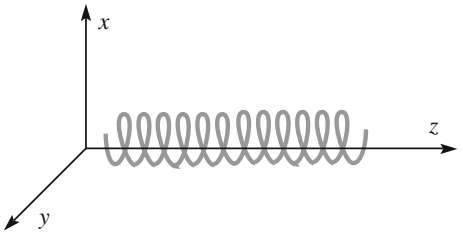


Fig. 6.18 A helix-shaped molecule of an optically active substance

A deeper analysis shows that the same effect is present for any orientation of the helix-shaped molecule and for every shape with specular asymmetry.

Summary

In this chapter, we studied the polarization phenomena of electromagnetic waves, in particular, of light. We learned the following principal concepts.

1. The polarization state of a light wave can be expressed as a superposition of two independent base states. We can take as base states two states of linear polarization at 90° to one another or two circularly polarized states in opposite directions. The most general polarization is elliptical.
2. Natural light from thermal sources (the sun and the stars, for example) is unpolarized. These sources are made of a huge number of molecules that emit light chaotically independently of one another. Consequently, the coherence time is on the order of a nanosecond or less, namely much smaller than the integration time of our sensors.
3. Unpolarized light can be polarized by one of the following processes:
   (a) Dichroism (the property of materials with two absorption coefficients)
   (b) Scattering (light from the sky or from a rainbow)
   (c) Reflection (images on windows and water surfaces)
   (d) Birefringence (the property of materials having two refractive indices)
4. The structure of a progressive plane monochromatic wave in a birefringent medium. The phase and group velocities have different magnitudes and different directions, the **D** field being normal to the former, while the **E** field is normal to the latter.
5. How to build quarter-wave plates and, more generally, introducing a phase shift between polarization states.
6. Optical activity, a property of the circular birefringent materials.

Problems

6.1. Consider the two polarization states $E_x = E_0 \cos(\omega t - kz + \phi_1)$ and $E_x = 2E_0 \cos(\omega t - kz + \phi_2)$. Draw the trajectory of the tip of the electric field for the following different values of the phase difference: ϕ_2–ϕ_1 = 0°, 45°, 90°, 135°, 180°.

6.2. Find the Brewster angle for light reflection off a glass of index n = 1.57.

6.3. A beam of natural light goes through two polaroids, one next to the other. The intensity after the second is ¼ of the incident of the first. What is the angle between the polaroid axes?

6.4. How can you obtain circular polarized light when starting with unpolarized light?

6.5. You measure the polarization of the sun rays reflected off the surface of a lake and find it to be complete. What is the angle of the sun on the horizon?

6.6. You insert, between two crossed polaroids, a third one with its axis at the angle θ with the first. What is the fraction of the light intensity transmitted by the system under these conditions as a function of θ?

6.7. A plate of birefringent material has been cut with faces parallel to the optical axis. Let x be the direction of the axis and y be normal to it, both in the plane of the plate. A plane monochromatic wave is normally incident (z direction). Its field on the first face is $\mathbf{E} = \mathbf{E}_0 \cos(\omega t - kz)$, with the origin of z at the entrance to the first polarizer. $\mathbf{E}$ is directed at the angle α with the x axis. Let δ be the phase shift between the two linear components introduced by the plate. What is the polarization state after the plate in the following cases: (a) $\alpha = 0°$ and $\delta = 180°$, (b) $\alpha = 90°$ and $\delta = 180°$, (c) $\alpha = 30°$ and $\delta = 180°$, (d) $\alpha = 30°$ and $\delta = 90°$, (e) $\alpha = 45°$ and $\delta = 90°$, (f) $\alpha = 045°$ and $\delta = 180°$.

6.8. The refractive indices of quartz are $n_o = 1.544$ and $n_e = 1.533$ at 0.6 μm. Calculate the thickness of the quarter-wave plate.

6.9. The intensity of a circularly polarized light incident on a circular polarizer is I. What is the intensity at the exit?

Chapter 7
Optical Images

Abstract In this chapter, we study the properties of optical images, constantly taking into account the wave nature of light. After having defined the concept of the image, we first discuss the plane mirror and the prism, and then curved mirrors and thin lenses, finding their basic equations and properties. We then deal with aberrations, depth of field, the resolving power and the action of a lens on the phase of the incident wave. We then discuss basic optical instruments, namely the magnifying glass, the telescope and the microscope. Finally, we give a few basic elements of photometric concepts.

In this and the subsequent chapter, we study the properties of optical images. Light consists of electromagnetic waves having wavelengths on the order of tenths of micrometers, namely much smaller than those of everyday objects. Consider a small segment of a wave surface propagating in a homogeneous and isotropic medium. Its trajectory is rectilinear. At the interfaces between different media, it reflects and refringes according to the Snell law, as we saw in Chap. 4. The luminous rays are the trajectories of the energy, which coincide with the propagation trajectories of the elements of the wave surface, because we are considering an isotropic medium. The geometrical optics, or ray optics, is a good approximation every time the dimensions of the surfaces met by light are much larger than its wavelength.

However, even if extremely useful, ray optics is often insufficient for understanding the physical processes at the base of image formation. Indeed, the structure of images is always determined by diffraction. The very concept of a ray is a mathematical, rather than physical, concept. Indeed, thinking of a light ray, one can imagine a narrow beam entering into a dark room through a small hole in a window. The beam is a straight line of a certain diameter. However, if we attempt to further define it to be a better approximation of a ray by making the pinhole narrower and narrower, when its diameter is on the order of the wavelength, the beam spreads out into a wide solid angle, as a result of diffraction. Trying to make the ray better defined, we have destroyed it. Similarly, if we have a light beam incident on a square mirror, we see the reflected beam. But if the size of the mirror is, say,

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_7

1 μm × 1 μm, the reflected ray no longer exists. The reflection law of geometrical optics does not hold under these conditions (again, due to diffraction).

In this chapter, we shall discuss the main aspects of image formation processes, constantly taking into account the wave nature of light. However, we shall employ geometrical optics when it is allowed and useful for simplifying the arguments. The focus on the physical aspects will lead us, on one side, to overlook aspects, so to speak, of the engineering of optics, and on the other, not to consider the psychological aspects of the image perception process. The latter should, however, be taken into account under certain circumstances, because psychological processes may cause us to locate images where they are, in fact, not present, and, contrastingly, prevent us from perceiving images in some cases. We shall warn the reader when we encounter these issues.

After having defined the concept of the image in Sect. 7.1, we shall discuss, in Sect. 7.2, the image formation process in the simplest cases of the plane mirror and the prism (including dispersion). In Sects. 7.3 and 7.4, we shall study the parabolic and the spherical mirrors, and how they form the image of a point source on the axis. In the remaining part of the chapter, we shall deal with thin lenses. We shall study the image formation of point sources from Sects. 7.5–7.7 and subsequently of extended objects under incoherent illumination conditions.

Lenses are subject to imperfections, called aberrations and irregularities, of which we shall give the basic elements in Sects. 7.8 and 7.9. We shall also see that optically "perfect" optical systems do exist.

Optical instruments produce two-dimensional images (on the retina or the image sensor of a camera) of objects in the world that are three-dimensional. In Sect. 7.10, we shall determine what the limits are for such images to be sharp.

Another relevant property of optical systems is their ability to give distinguishable images of two nearby points or, as we say, to resolve them. We shall study resolving power in Sect. 7.11.

In Sect. 7.12, we shall analyze the action of the lens, considered to be a phase diaphragm, on the phase of an incident plane wave. This action is more general, not only of the lens, but also of other image-forming instruments.

In the final sections, we shall present elements of the simplest optical instruments, namely the magnifying glass, the telescope and the microscope, and, finally, a few basic elements of photometric concepts.

7.1 Preliminaries

We begin by noting that a precise definition of the concept of an *optical image* is far from being trivial. Let us start from the observation that reflecting plane surfaces and transparent prisms both deviate the path of an incident light wave, without changing their curvature. Consider, for example, a point source S in the neighborhood of a plane mirror. As Fig. 7.1 shows, the light waves emitted by the source are deviated, reflected, in this case, by the mirror. When the reflected waves hit the

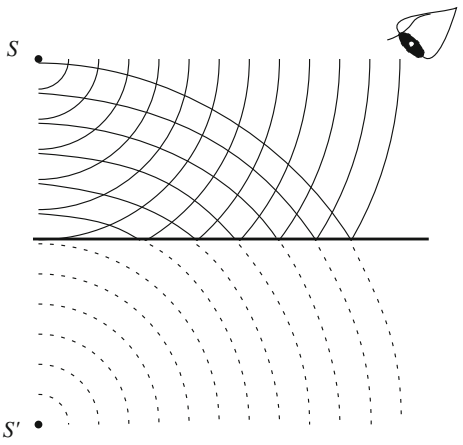


Fig. 7.1 Reflection of light waves from a point source S and its image S'

eye of the observer, he/she, automatically thinking that the light has always traveled in a straight line, sees a luminous point source S' on the other side of the mirror. There is no real source there, however, and we speak of an *image* of the source.

Hence, in this case, the image is the center of the spherical waves reaching the eye of the observer after having been diverted along their path. If the waves were not diverted, we would have spoken of an object, rather than of an image. Clearly, very similar arguments hold for the action of a prism.

Curved, spherical and parabolic mirrors change the curvature of the incident waves. Similarly, if we interpose a lens between a point source and the eye of the observer, the waves reaching his/her eye do not have their curvature center at the source, but at a different point, which is an image of the source. We conclude with the definition of the optical image given by Vasco Ronchi (Italy, 1897–1988), which is as follows: *the image of a point source is the center of the light waves that reach the eye of the observer after having been diverted or deformed along their path*. In this definition, we include plane waves, considering them to be spherical with a center (the source or the image) at an infinite distance.

If instead of a point source, we have an extended light source (like the sun or a lamp) or an illuminated object, we can think of it as being a set of point sources of different intensities and positions. The image is the set of the images of all the point-like elements of the source or of the illuminated body.

Note that the process of image formation in our eyes is not only a physical process but also involves a number of psychological components. Due to these, the observer does not always perceive an image properly. We shall give some examples later on.

Lenses and mirrors have necessarily a limited size. Consequently, they always intercept and transmit a segment of the incident wave, and the diverted or deformed waves are not spherical or plane surfaces, but circular *segments* of a plane or sphere. This fact implies that, in any case, diffraction phenomena are present, which are all

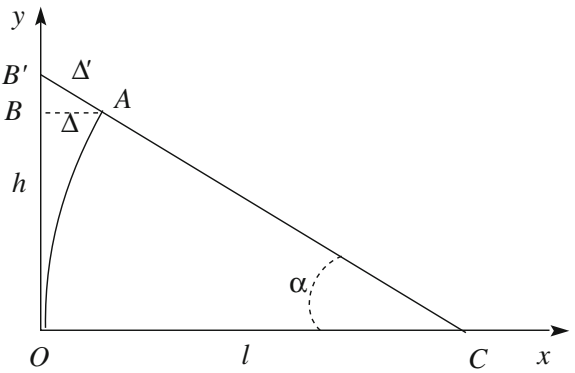


Fig. 7.2 The geometry of paraxial rays

the more important the smaller the dimensions of the wavefronts are relative to the wavelength. Hence, as we shall see, the image of a point is never rigorously a point, but rather a fundamental diffraction pattern.

In this chapter, we consider the processes of image formation under the usual conditions of having thermal light sources, in which the molecules emit their light independently of one another. Images produced with coherent light, like holograms, will be discussed in the next chapter.

The process of image formation by lenses or mirrors almost always only involves light rays forming small angles with the direction of normal incidence on the optical elements. In addition, the rays never stray very far from the axis of the system. Under these conditions, we talk of paraxial rays. We shall now find a simple geometrical formula, which is valid to a good approximation under paraxial conditions. As a matter of fact, this is the only geometrical formula that we shall need.

Consider an arc OA of a circle of center in C and radius l subtending the *small* angle α, as in Fig. 7.2. We want an approximate expression of the distance Δ of the extreme A of the arc from the line tangent to the arc in its other extreme O.

We choose the reference axes as shown in Fig. 7.1. The coordinates of A and C are (Δ, h) and $(l, 0)$, respectively. Imposing their distance as being equal to l, we have

$$AC^2 = (\Delta - l)^2 + h^2 = l^2,$$

which we rewrite, dividing by h^2, as

$$\left(\frac{\Delta}{h}\right)^2 - 2\frac{l}{h}\frac{\Delta}{h} + 1 = 0.$$

Now, $\Delta/h = \tan(\alpha/2)$ is small in our hypothesis of α being small and we can neglect $(\Delta/h)^2$. At the first order in α, we then have

$$\Delta = \frac{h^2}{2l},$$

which is the formula we were looking for. In addition, we shall also need an approximate expression of the distance Δ' (see Fig. 7.1), which is the difference between the length of the hypotenuse and the longer leg of the rectangular triangle OCB'. Indeed, at the first order in α, Δ' is equal to Δ. We can expand this as

$$\Delta = \Delta' \cos\alpha = \Delta'\left(1 - \alpha^2/2 + \cdots\right) \cong \Delta'$$

In conclusion, we have

$$\Delta = \Delta' = \frac{h^2}{2l}. \tag{7.1}$$

Also notice that, always at the first order in α, both distances OB and OB' are equal to h.

To get an idea of the orders of magnitude, let us consider, for example, $\alpha = 10° = 0.17$ rad. In Eq. (7.1), we have neglected terms on the order of α^2, that is, of $(0.17)^2 = 0.03$ relative to the unit. Accuracy within a few percentage points will be enough for our subsequent discussion, but the exact expressions should be used when precision is requested.

7.2 Plane Mirrors and Prisms

There is not too much to say about plane mirrors. We simply recall here that they change the propagation direction of light waves, leaving their curvature unaltered. The image of an extended object consequently appears to be of the same shape and size as the object, the well-known left/right inversion of the vertical mirror apart. Note that the diverted waves are centered, but do not come from the images. Under these conditions, we talk of a *virtual image*. Contrastingly, an image is *real* if the waves come from its points.

We finally note that the position of the image is independent of the wavelength, because the reflection law (Sect. 4.4) does not depend on wavelength.

Prisms, like mirrors, change the propagation direction of the light waves without changing their curvature. Unlike mirrors, however, the deviation they induce does depend on wavelength, as we saw in Sect. 4.4. Consequently, when the incident wave is not monochromatic, after the prism, we have many (infinite) approximately plane waves, each of a wavelength, propagating in different directions. This is the phenomenon of *dispersion* by the prism that we have already encountered and that we will discuss now quantitatively.

Consider a plane wave incident on a prism. The angle between the propagation directions after and before the prism, which we shall call δ, is known as a *deviation*.

The deviation δ depends on the incident angle, and it can be shown (exactly as we did for a water drop in Sect. 4.5) that it initially decreases with increasing incidence angle, reaches a minimum δ_m, and then increases. The minimum deviation conditions happen, as shown in Fig. 7.2, when the incident and transmitted waves lay symmetrically on the two sides of the prism (that does not mean being parallel to the base).

For the sake of simplicity, we shall only consider the minimum deviation condition.

The action of the prism is known once its dihedral angle α and refractive index n are known, under the hypothesis, which we have adopted, that the index of the media on the two sides of the prism is 1. As a matter of fact, we can measure the refractive index of a material by measuring the minimum deviation of a prism of that material.

Let us find the relation between n and δ_m. Consider two points like A and B on the two faces of the prism at the same distance from the vertex V, which we call l. The segment AV of the prism face intercepts a portion of the incident wave, of which AA' in Fig. 7.2 represents a wave surface. On the other side, BB' is a wave surface as well. Consequently, the times taken by the phase to travel the paths $A'VB'$ and AB should be equal. Now, the former is completely in air (index equal to 1), and the latter is completely in the medium of index n, giving us $(A'V + VB')/c = AB/(c/n)$. We can now write

$$AV' + VB' = 2A'V = 2l\cos AVA' = 2l\cos\frac{\pi - (\alpha + \delta_m)}{2} = 2l\sin\frac{\alpha + \delta_m}{2}$$

and

$$AB = 2l\sin\frac{\alpha}{2}.$$

In conclusion, we have

$$n = \frac{\sin\frac{\alpha+\delta_m}{2}}{\sin\frac{\alpha}{2}}, \tag{7.2}$$

which is the equation we were looking for. We shall now study the dependence of the deviation by a prism on the wavelength. For the sake of simplicity, we shall consider the case of the *thin* prisms, namely those having $\alpha \ll 1$. In this case, the minimum deflection condition corresponds to an almost normal incidence on the prism. Under these conditions, δ_m is also small and we can approximate the sine with the angle and write Eq. (7.2) as

$$\delta_m = (n-1)\alpha. \tag{7.3}$$

When the wavelength varies from λ to $\lambda + d\lambda$, the index varies, say, from n to $n + dn$. We obtain the corresponding variation of the minimum deviation by differentiating Eq. (7.3), obtaining

$$d\delta_m = \alpha dn. \tag{7.4}$$

We call the change of deviation per unit wavelength the *angular dispersion* of the prism, namely

$$\frac{d\delta_m}{d\lambda} = \alpha \frac{dn}{d\lambda}. \tag{7.5}$$

If the incident wave is not monochromatic, the components of each wavelength are diverted to a different angle. The dispersion is greater the greater the index dependence on λ.

Consider, as an example, a glass prism with $n = 1.5$ and $\alpha = 20°$ ($\approx$0.35 rad). The minimum deviation is $\delta_m = 10°$ ($\approx$0.17 rad). Indeed, this is an average value. In a typical glass, the difference between the index of the blue ($\lambda = 0.45$ μm) and that of the red ($\lambda = 0.65$ μm) is $dn = 0.01$ (see Sect. 4.4). The corresponding difference in deviation is, for Eq. (7.4), $d\delta = \alpha dn = 3$ mrad. This means, to fix the ideas, that placing a screen at 1 m distance from the prism causes the spectrum to be 3 mm wide from red to blue.

Notice also that, under normal dispersion conditions, namely $dn/d\lambda < 0$, the deviation increases with decreasing wavelength.

The simplified explanation we have just given is sufficient for understanding the importance of the prisms for the analysis of the spectra of the radiation coming from a given source. Such a study gives important pieces of information on the physics of the source. Suppose, for example, we want to study the vibration frequencies of a molecule. We can have a gas composed of these molecules in a transparent container and excite the molecules, for example, with an electric discharge or by heating the gas on a flame. We then limit the light emitted by the gas using a narrow slit parallel to the dihedral angle of a prism located after the slit. In this way, we have a narrow laminar beam incident on the prism. The components of different wavelengths present in the beam are deflected by the prism in different directions, which we can measure by looking at the prism through a telescope rotating on a goniometer. In the case of gases, the observed spectrum consists of a number of "lines" at wavelengths corresponding to the characteristic oscillation frequencies of the molecules.

Coming back to the concept of image, let us consider observing a point source S through the prism. If the source is monochromatic, the waves arrive at our eyes deviated by a certain angle δ, and we see an image of the source displaced on a plane normal to the dihedral angle by a distance proportional to δ (and to the distance of the source). Even now, the image is virtual, and even now, the image of an extended source is equal to the source, with the same geometrical dimensions. However, unlike the plane mirror, if the source is not monochromatic, its image

appears dispersed in monochromatic images, with different colors in different positions.

7.3 Parabolic Mirror

In this section and in the subsequent one, we shall discuss the simplest properties of two types of concave mirror, namely the parabolic and the spherical ones. They have the property to change the curvature of the incident waves or, from the geometric optics point of view, to focus the incident rays. We shall not address the magnification, the aberrations or the resolving power of mirrors, which are similar to those of the lenses that we will discuss in subsequent sections.

A *parabolic mirror* is a paraboloid of revolution.

The first important property of the parabolic mirror is the ability to transform a plane wave incident along the axis in a spherical wave with the center in the *focus* of the paraboloid. We choose the origin of the reference frame in the vertex O of the paraboloid and the x-axis on the geometrical axis. Figure 7.3 represents a section of the paraboloid in a plane containing the x-axis. Let F, having coordinates $(f, 0)$, be the focus and HH the trace of the plane normal to the axis at the distance f from O on the other side with respect to the focus. This is the *directrix* of the parabola.

Consider a plane wave incident on the mirror from the negative x direction. We can think that the different points of the mirror start emitting the reflected wave in the instant in which they are touched by the incident wave. The greater the distance from the axis, the sooner this takes place. For example, the path PA is shorter than the path $P'A'$. It follows that the reflected wave is no longer plane, but is concave, as is $OD'D$ in the figure. We will now show that the reflected wave is spherical (Fig. 7.4).

Consider the points D, D', etc., of the reelected wave. They are such that

$$PA + AD = P'A' + A'D'. \tag{7.6}$$

Now, the points of the paraboloid are by definition equidistant from the directrix HH and from the focus F, namely it is $AB = AF$, $A'B' = A'F$, etc. On the other hand, obviously, we have $PA + AB = P'A' + A'B'$, etc., and consequently $PA + AF = P'A' + A'F$, etc. Finally, taking Eq. (7.6) into account, we conclude that $DF = D'F$ etc., namely D, D' etc., lay on a sphere with center F.

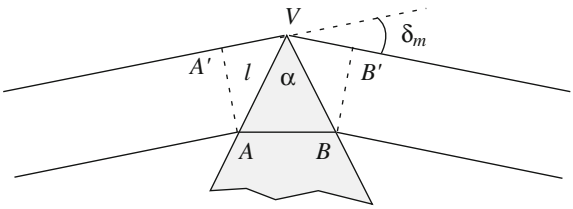


Fig. 7.3 Minimum deviation by a prism

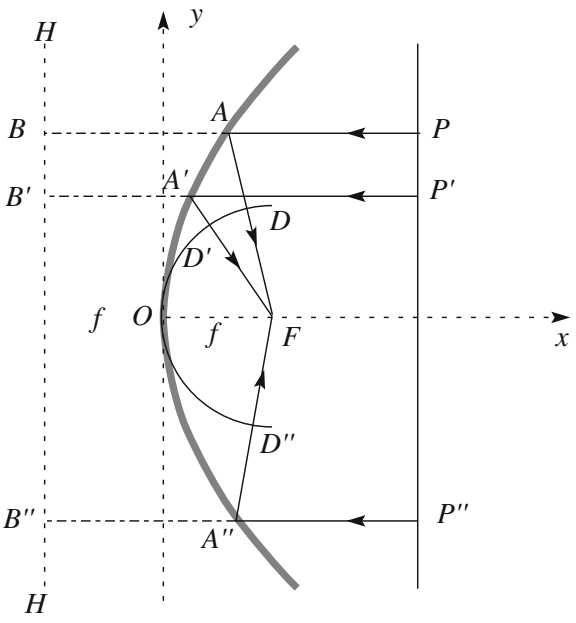


Fig. 7.4 A parabolic mirror, an incident plane wave and the reflected spherical wave at the moment it leaves the mirror

In conclusion, a plane wave incident on a parabolic mirror in the direction of the axis is reflected in a spherical wave with the center in the focus of the paraboloid. The energy transported by the wave concentrates around the focus. Pointing the axis towards the sun, the energy concentration in the focus can be very large, burning and destroying what is there. This is at the origin of the word 'focus.' The geometry of the conic sections and the properties of the parabolic mirrors were well known to the ancient Greeks. Archimedes from Syracuse (Sicily, 287–212 BC) is said to have used parabolic mirrors to burn the Roman ships that were approaching his city.

As we have seen, the reflected wave is a segment of spherical surface with the center in F. The wave segment becomes smaller and smaller as it approaches the center. After having passed the center, the wave becomes a divergent spherical segment. Figure 7.5a shows the converging and diverging waves. Figure 7.5b shows the same process from the point of view of the geometrical optics. The light rays converge in the focus and subsequently diverge from it. The two points of view are equivalent as long as the diffraction phenomena can be neglected.

According to the above given definition, in the focus F, we have an image of a point source on the axis at infinite distance. The waves reaching the observer do come from their center, and consequently, the image is real. As we have just discussed, the energy density in the image is very large. As a matter of fact, it would be infinite if the image were a geometrical point. This is clearly physically

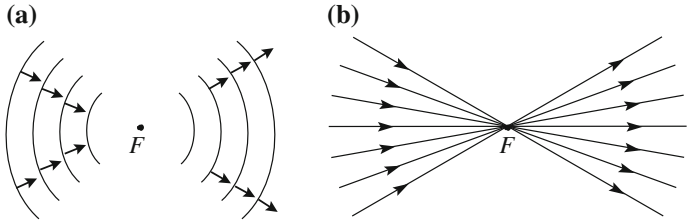


Fig. 7.5 Light converging in the focus and subsequently diverging from it. **a** Wave view, **b** ray view

impossible. Indeed, perfectly point-like images do not exist, because the mirror (and any optical device) has a finite diameter. Consequently, it does not reflect the entire incident wavefront, but only a segment of limited diameter. And we know that diffraction always exists when a wave front is limited. The consequence is that the size of the image can never be null. We shall deal with this issue in Sect. 7.11 in the similar case of the lens.

We note here that an observer located beyond the image does not really see an image, namely a bright spot, in *F*. His/her mind refuses to see something where there is no physical body, but just a vacuum. However, if one puts a frosted glass in *F*, or even a transparent one if it is not too clean, the luminous spot immediately appears. The brain now knows that there is something physical there. This is, in general, the case with real images.

The light coming from a star, and generally from astronomical bodies, is a plane wave. We can thus build a parabolic mirror to concentrate all the light collected by its surface in the focus and place the eyepiece there. The larger the diameter of the mirror, hence, the more luminous the flux intercepted, the further away the stars that one can detect will be. In practice, producing parabolic surfaces of optical quality is extremely difficult, and consequently, telescopes have spherical mirrors, which we discuss in the next section.

Another use of the parabolic mirror is to have it work in the opposite way. If we position a light source in the focus, producing a spherical wave divergent toward the mirror, the reflected wave is a plane wave leaving the mirror along the axis. For this reason, parabolic mirrors are used in the headlights of our cars.

7.4 Spherical Mirror

As we already mentioned, in practical terms, the production of an optical quality parabolic surface is very difficult and expensive, especially if the requested dimensions are large. Spherical mirrors, which can be produced more cheaply and more reliably, are used in their place. Indeed, as we shall now see, a spherical mirror behaves similarly to a parabolic one for paraxial rays. The main difference is

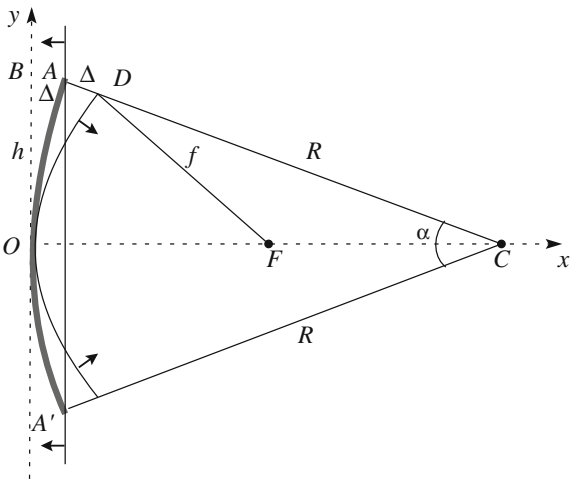


Fig. 7.6 A spherical mirror, the incident wave at the moment it first touches the mirror and the reflected wave at the moment it leaves the mirror

that spherical mirrors suffer from spherical aberration, as we shall see below, while parabolic mirrors do not. As opposed to lenses, mirrors do not suffer from chromatic aberration, because the relation between the reflected and incident ray does not depend on wavelength. The reason for the approximately equal behavior of a spherical and a parabolic mirror is that a spherical cap does not differ much from a parabolic cap if the subtended solid angle is small enough.

In this case as well, the mirror is concave, and consequently, a plane wave incident along the axis touches the rim first. These points are the first to emit the reflected wave, which is consequently concave.

Figure 7.6 shows the intersection of the spherical cap of the mirror with the coordinate plane x, y, where we took the x-axis on the geometrical axis, similarly to Fig. 7.5. A and A' represent the traces of the rim of the mirror, which we assume to be circular, C is the center of the sphere, R is its radius and O is the vertex of the mirror.

Point A of coordinates (Δ, h) is reached by the incident wave first and first starts emitting the reflected wave. When the wave reaches the vertex O, the wave reflected in A has already traveled the path AD, the length of which we call Δ. This is equal to AB as well. The reflected wave is not really spherical, but for small values of the opening angle α, it is approximately so. Let f be the radius of this sphere and F its center (which is obviously on the axis). Now, the distance of point D of the sphere from the y-axis (see Fig. 7.6) is approximately 2Δ. We now apply Eq. (7.1) to the arc OD, obtaining $2\Delta = h^2/(2f)$, and to the arc OA, obtaining $\Delta = h^2/(2R)$. Putting these two results together, we finally get

$$f = \frac{R}{2}. \tag{7.7}$$

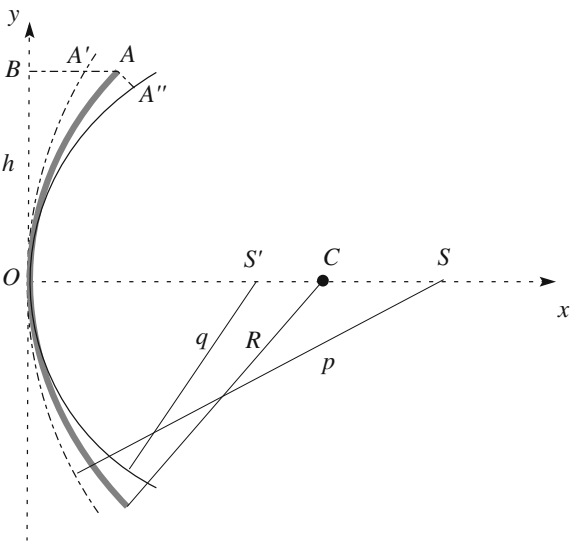


Fig. 7.7 A spherical mirror and the reflected wave at the moment it leaves the mirror

The distance f of the focus from the vertex is the main property of the mirror and is called the *focal length*. Equation (7.7) tells us that the focal length is equal to one half of the curvature radius.

Indeed, the reflected wave is not exactly spherical, its difference from a sphere being on the order of α^2. As a consequence, the portions of the wave at different distances from the axis converge at slightly different positions. The consequence is a distortion of the image, called a spherical aberration, as we already mentioned. We shall discuss this in the case of the lens in Sect. 7.8, which is completely analogous. As with other aberrations, the spherical one can be corrected.

Consider now a point source S that is still on the axis of the mirror, but at a finite rather than infinite distance. Let p be this distance from the vertex O and, to fix the ideas, let it be larger than R. Figure 7.7 shows the reflected wave at the moment its last point leaves the mirror and, along the dotted line, the incident wave as it would have been in the absence of the mirror. Even now, when the incident wave touches at O, the wave reflected at point A on the rim has already made the path AA'' of length equal to AA'.

In this case too, the reflected wave is not spherical, although it is almost so for small values of α. Considering it to be a sphere, let q be its radius and S' its center. We now apply Eq. (7.1) to three circle arcs; first to OA', obtaining

$$A'B = \frac{h^2}{2p}, \tag{7.8}$$

then to OA, obtaining

$$AB = \frac{h^2}{2R}, \tag{7.9}$$

and finally to OA'', considering AA'' parallel to BA, as is approximately true for small values of α, obtaining

$$BA + AA'' = \frac{h^2}{2q}. \tag{7.10}$$

On the other hand, it is $AA' = AA'' = BA - A'B$; hence, we have $BA + AA'' = 2BA - A'B = h^2/(2q)$. Substituting this expression in Eqs. (7.8) and (7.9), we obtain

$$\frac{h^2}{2q} = \frac{2h^2}{2R} - \frac{h^2}{2p},$$

or

$$\frac{1}{p} + \frac{1}{q} = \frac{2}{R}.$$

But the right hand side is just the reciprocal of the focal length for Eq. (7.7), and we write

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f}. \tag{7.11}$$

This important relation gives the relation between the position of the image and the position of the source. It states that the sum of the curvatures at the mirror of the incident and of the reflected waves is a constant, independent of the position of the source, and equal to twice the mirror's curvature.

$1/f$ is called the *optical power*, or simply *power*, of the mirror, while the curvatures at the mirror of the waves are called their *vergences*. Equation (7.11) is called the mirror equation for vergences (we shall see that the same equation holds for thin lenses) and is read by saying that *the sum of the vergences of the incident and reflected waves is constant and equal to the power of the mirror*. Concave mirrors are *con*vergent, meaning that the vergence of the reflected wave is larger than that of the incident wave. We shall not discuss convex mirrors, but simply note that they are *di*vergent. By convention, the focal length, and consequently the power, is positive for converging, negative for diverging mirrors. Vergence and power have the dimension of an inverse length; their unit is called a diopter, which is equal to 1 m^{-1}. We shall come back to that when discussing lenses.

We shall not discuss here the action of the mirror on point sources that are not on the axis, but only mention that Eq. (7.11) still holds. We also mention that the images of extended sources are geometrically similar to their source, but of different size (magnification of the mirror). These properties of mirrors are very similar to those of the lens that we shall discuss in the next sections.

Nature has exploited mirrors in the eyes of certain animals. An example is the Gigantocypris, a crustacean living at depths of 1000 m and more in the oceans, where there is almost no light. This animal is one of the largest of its family, measuring about 25 mm. Its head measures one half of its body, and hosts two large eyes. To increase light collection, the eyes do not use lenses to focus the images on the retina, but rather parabolic mirrors. Mirror eyes have a resolving power smaller than lenses of the same aperture, but are able to produce images of higher luminance (see Sect. 7.16). Other species of mollusks and crustaceans at the same depth have mirror eyes as well.

7.5 Thin Lenses

A lens, in its simplest form, is made of a piece of glass, plastic or another transparent medium, limited by two *spherical coaxial segments* with radiuses R_1 and R_2. Each of the two surfaces can be concave, convex or plane. A lens is said to be thin if its maximum width d is small compared to the other geometrical dimensions, namely the diameter D and its curvature radiuses. In this book, we shall deal only with thin lenses.

A few historical hints will be useful. It has been known since ancient times that, when viewed through almost spherical transparent gems or spherical glass containers full of water, objects appeared much larger and closer, but deformed. While the geometry of mirrors had been fully developed by the Greeks, the first geometrical theory of lenses is credited to Ibn Sahal (Bagdad 940–1000), who flourished during the Arabic enlightening in the Abbasid court. As we already mentioned, he discovered the law of refraction. In 984, he published the book *On the burning instruments*, in which he developed the theory of lenses bound by parabolic and hyperbolic surfaces. As the title suggests, Ibn Sahl's aim was to focus light on a point for burning purposes, as Archimedes had done with mirrors. As a consequence, his theory dealt with the images of point sources on the axis. All these important developments did not lead to any practical exploitation. In the following centuries complete spheres were considered to try to produce images, with poor results affected by severe aberrations. The situation changed with the discovery that transparent bodies limited by spherical *caps* subtending small solid angles worked much better than complete spheres. They were called *lenses* for their resemblance with lentils.

The inventor of the lens is unknown, but manufacture and use of lenses to magnify images, in the form of the magnifying glass, and to correct vision defects, in the form of spectacles, became common in Italy in the second half of the XIII

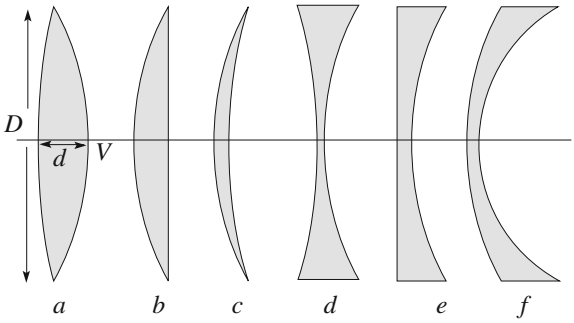


Fig. 7.8 Common types of thin lens. **a** Biconvex, **b** planoconvex, **c** positive meniscus, **d** biconcave, **e** planconcave, **f** negative meniscus

century. Two frescoes by Tommaso da Modena of 1352 portray two prelates of the previous century, both reading, one using a magnifying glass and the other a pair of pince-nez spectacles. The latter (Hugues de Saint Cher) is represented as a cardinal, who is known to have served in this role in Italy from 1244 to 1263. We can conclude that spectacles began to be used to correct sight by no later than the latter date.

These lenses were produced by glass artisans and masters without any knowledge of the underlying theory. For the theory, we must wait until 1611, when Johannes Kepler (Germany, 1571–1630) published his treatise, the *Diotrice*.

Figure 7.8 shows the most common geometrical types of lens. For each surface the point on the axis is called its vertex (like V in Fig. 7.8b). To characterize the action of the lens, it is further necessary to specify the refractive index n of its material and of the two media before and after it. We shall limit our discussion to lenses in air, for which we take the index to be unitary. If, as is usually the case, $n > 1$, the lenses in Fig. 7.8a–c are convergent, while (d), (e) and (f) are divergent. Convergent and divergent lenses are also said to be positive and negative, respectively, from the sign of their focal length, as we shall see.

Let us start by discussing the case of (*a*), in which both surfaces are convex. Let S be a point-like source on the axis at a distance p from the lens (more precisely, from the vertex V of its first surface). Let R_1 and R_2 be the radiuses of the first and second surfaces, respectively.

With reference to Fig. 7.9, Eq. (7.1) gives us

$$d_1 = \frac{(D/2)^2}{2R_1}, \quad d_2 = \frac{(D/2)^2}{2R_2},$$

and, being that the width of the lens is $d = d_1 + d_2$, we have

$$d = \left(\frac{D}{2}\right)^2 \left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{7.12}$$

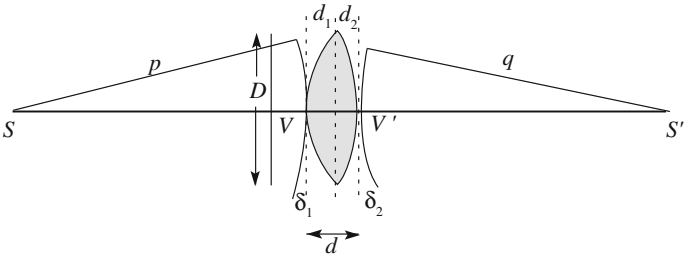


Fig. 7.9 Source S and image S' for a converging lens

The point source S emits a divergent spherical wave, whose radius increases during propagation. When it touches the lens in V, its radius is p and its periphery has a distance from the plane normal to the axis through V of

$$\delta_1 = \frac{(D/2)^2}{2p}.$$

After that, the central part of the wavefront enters the lens, while the periphery is still moving in air. The phase velocity of the central part is now smaller than at the periphery. The wave surface changes shape. Its curvature diminishes, while the wave proceeds. At the exit from the lens, the wave curvature will consequently be smaller than at the entrance. It might go down to zero and even change signs. In the latter case, the center of curvature will move to the right of the lens, as in the example shown in Fig. 7.9.

An analog situation can take place for macroscopic waves, such as those on the surface of water. Think, for example, of a linear wave (analogous in two dimensions to the plane wave in three dimensions) travelling along the surface of the water in a pool. Suppose the depth of the water h to be on the same order as the wavelength. Under these conditions, the phase velocity, according to Eq. (4.12), is proportional to $\sqrt{h}$. We reduce the thickness of the water in a region of the pool, positioning, on its bottom, an obstacle shaped like a small flat hill, all immersed in the water. When the wave front reaches the "hill," its motion is slowed down, the more so the higher the hill. Beyond the obstacle, the wavefront is no longer a straight line, but rather is concave, or circular if the shape of the obstacle is properly designed.

Coming back to light, we notice that the wave surface after the lens is not exactly spherical. However, it is so approximately for small values of D. Considering it to be spherical, let q be its radius and S' its center, which is on the axis. Figure 7.9 shows that, when the wave completely exits from the lens (namely touches it at V'), the distance of its periphery from the plane perpendicular to the axis in V' is

$$\delta_2 = \frac{(D/2)^2}{2q}.$$

We now impose the condition that the two spherical surfaces of centers in S and S' should be wave surfaces. This means that the time taken by the phase to cross the path $\delta_1 + d + \delta_2$ at the speed c must be equal to that which it takes to cross the path d at speed c/n. We then write

$$\frac{\delta_1 + d + \delta_2}{c} = \frac{d\,n}{c}.$$

Simplifying, we get

$$\delta_1 + \delta_2 = (n-1)d. \tag{7.13}$$

By substituting in this equation the expressions of δ_1, δ_2 and d that we found, we have

$$\frac{1}{p} + \frac{1}{q} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{7.14}$$

As we see, the right-hand side depends only on the characteristics of the lens and not on the position of the source. We can then state that the sum of the *curvatures*, namely the *vergences* of the incident wave $1/p$ and of the outgoing one $1/q$, is constant for a given lens, namely independent of the position of S. In particular, the outgoing wave may have zero curvature, namely be plane. We have then $q = \infty$. We indicate with $1/f$ the curvature of the incident wave under this condition. From Eq. (7.14), it is

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{7.15}$$

The length f is the *focal length* of the lens and its reciprocal $1/f$ is its *dioptric power*, or simply its *power*. As we already mentioned, the unit for vergence and dioptric power in the SI is the *diopter* (which is then 1 m^{-1}).

We see immediately that if S is at infinite distance ($p = \infty$), namely if the incident wave is plane, the center of the outgoing wave is at the distance a $q = f$ from V' on the other side of the lens.

This first focus of the lens, F is defined as the point of the axis, which is such that a spherical wave of center in F incident on the lens produces an outgoing plane wave. The second focus F' is the point of the axis that is the center of the spherical outgoing wave when the incident on the axis wave is plane.

In the case we have considered, the two focuses are on the opposite sides of the lens at the same distance from it. This is always the case when the mediums on the two sides of the lens are the same. It can be easily seen, with arguments very similar

to those we have just made, that if the mediums on the two sides are different (as, for example, with the cornea of the eye or swimming under water), the two focal lengths are different. We shall limit the discussion to the case in which the two mediums are the same.

The focal length (or the power) is the fundamental property of the lens. Equation (7.15) gives its expression as a function of the characteristics of the lens, and is called the *lens maker's formula*. Once the lens is made, its focal length is defined, and this is what is needed to link the positions of the source and the image. In fact, we can rewrite Eq. (7.15) as

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f}, \tag{7.16}$$

which is called the *thin lenses equation* or equation for vergences. We already found this for the spherical mirror.

All the results we have just reached considering waves can also be easily reached with geometric optics, in which one considers rays in place of waves, as we shall see in Sec. 7.7. In an isotropic medium, such as those we are considering, the rays are always normal to the wave surfaces. From the point of view of geometric (ray) optics, the situation in Fig. 7.9 corresponds to that in Fig. 7.10.

The rays that leave the source S enter the lens at different points, but all of them are deflected at S'. This can be shown using the Snell law at both surfaces, but we shall not do that.

In conclusion, in the case we have considered, S' is a real image of the source S. As a matter of fact, the image is real, as in Figs. 7.9 and 7.10, only if the radius p of the incident wave (namely the distance of the source S from the lens) is larger than the focal length f. If we now move S closer and closer to the lens, namely we increase the vergence $1/p$ of the incident wave, the vergence $1/q$ of the outgoing wave decreases. The distance q of the image S' increases. Then S is in the first focus, namely when $1/p = 1/f$ the vergence of the outgoing wave is zero. Under these conditions, the image S' is at infinite distance. In the language of geometrical optics, the outgoing rays are parallel to the axis.

If we take the source S still closer to the lens, namely we increase the vergence of the incident wave even more, the lens, so to speak, is no longer able to invert the wave curvature. The center of the outgoing wave is now on the left of the lens, located farther from the lens than S. The outgoing wave is divergent.

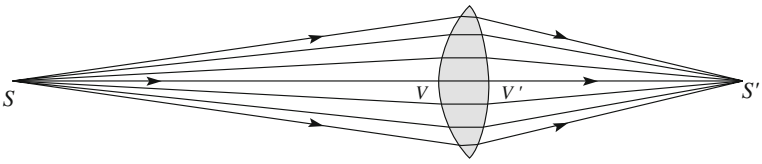


Fig. 7.10 Source and image for a converging lens

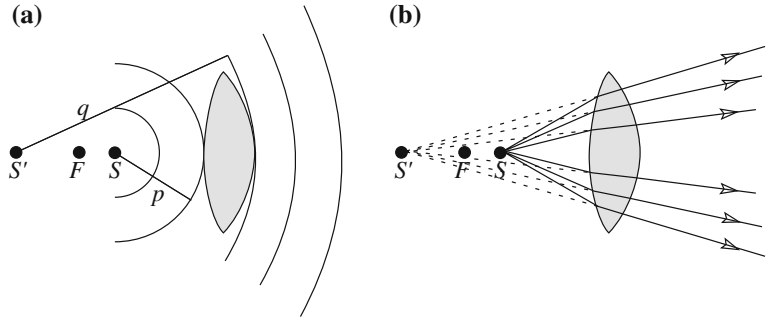


Fig. 7.11 Virtual image of a source closer to the lens than the focus

Equation (7.16) is still valid. In the just described situation in which $p < f$, Eq. (7.16) tells us that $q < 0$. This situation is shown in Fig. 7.11a and, in the language of geometric optics, in Fig. 7.11b. As we see, in this case, S' is a virtual image of the source.

In this section, we have implicitly adopted a few conventions on the signs of the curvatures of the surfaces we have encountered. We shall make them explicit now. First of all, we have considered the incident wave as coming from the left. In addition, the expressions we found can be used, in general, if we take:

(1) the vergence $1/p$ of the incident wave to be positive if its center of curvature is on the left of the lens, negative if it is on the right;
(2) the vergence $1/q$ of the incident wave to be positive if its center of curvature is on the right of the lens, negative if it is on the left;
(3) the focal length f to be positive if the lens is convergent, negative if divergent;
(4) the curvature radius of a surface of the lens to be positive if the surface is convex, negative if concave.

With these conventions, it is easy to extend the argument we made considering a biconvex lens to all the other lens geometries shown in Fig. 7.8. In doing so, one sees that the lenses of Fig. 7.8b, c are *convergent*, meaning that their action is to decrease the vergence of the waves, or, in the language of geometrical optics, to bend the luminous rays *toward the axis*. This happens in any case in which the lens is thicker at its center than at its borders. Contrastingly, the lenses thinner at their center are *divergent*. They increase the vergence of the light waves, or, in other words, bend the luminous rays away from the axis. This is a consequence of the fact that, in these cases, the periphery of the wave slows down more than its center. We leave as an exercise to show that Eqs. (7.14), (7.15) and (7.16) are also valid for divergent lenses (with the sign conventions we adopted).

It is also easy to show the path reversibility property of the lens. This means that, if we put the point light source at S' where we had the image, and consequently, if the wave incident on the lens is equal to the wave that was outgoing, moving in the opposite direction, then the outgoing wave is now equal to the one that was

incoming, moving in the opposite direction of that one. The image is formed at the point at which we had S. Indeed, in our arguments, we have used distances and times taken to cross them, which are independent of the direction of the crossing.

We shall not discuss thick lenses and the optical systems composed of several lenses in this book. Full treatises have been dedicated to the subject. We simply observe that these systems are treated starting with the construction of the image of the first surface. This image is then taken as the source for the next surface and the further corresponding image is found, and so on. In the next section, we shall consider a simple, but useful, example of that.

7.6 Thin Lenses in Contact

Here, we consider the simplest optical system consisting of more than one lens. We have a number of thin lenses, one next to the other in contact with one another, centered on the same axis and all with the same diameter D. Figure 7.12 shows two of them, but our arguments will be valid for any number. Let $f_1, f_2, \ldots$ be their focal distances, n_1, n_2,… their refractive indices, and let the index of the medium in which the system is immersed be equal to 1. Let d_1, d_2,… be their thicknesses. Equation (7.1), with (7.15), gives us

$$d_1 = \frac{D^2}{8f_1(n_1 - 1)}, \quad d_2 = \frac{D^2}{8f_2(n_2 - 1)}, \ldots$$

We now impose the condition of having an outgoing spherical wave when the incident wave is spherical, similar to what we did with Eq. (7.13) for a single lens. The condition is now

$$\delta_1 + \delta_2 + \ldots = (n_1 - 1)d_1 + (n_2 - 1)d_2 + \ldots,$$

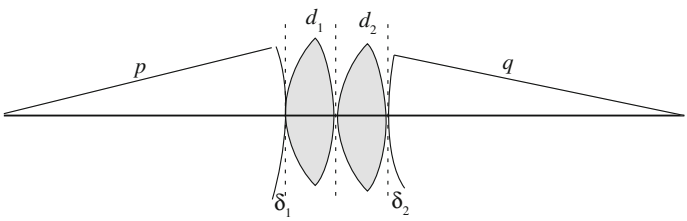


Fig. 7.12 Two thin coaxial lenses of equal diameters in contact

which gives us

$$\frac{D^2}{8p} + \frac{D^2}{8q} = \frac{D^2}{8f_1} + \frac{D^2}{8f_2} + \ldots,$$

that is,

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f_1} + \frac{1}{f_2} + \ldots.$$

We have thus found that the sum of the vergences of the incoming and outgoing waves is independent of the position of the source, namely that it is a characteristic of the system. Hence, the system of centered thin lenses in contact is equivalent to a single lens with a dioptric power equal to the sum of the powers of its component lenses, namely

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} + \ldots, \tag{7.17}$$

7.7 Images of Extended Objects

Up to now, we have considered the image formation of a point-like source located on the axis of the optical system. We shall now study how our eyes and optical instruments produce images of extended objects. In this section, we consider two-dimensional light sources, or illuminated objects, lying in a plane perpendicular to the optical axis of the lens. In Sect. 7.10, we shall discuss three-dimensional objects.

An extended object may be a primary light source, such as the sun or a street lamp, or may be illuminated by a light source and scatter its light, like the moon, the sky or common objects. In both cases, we can think of the object as being made of many, as a matter of fact, infinite, point-like sources, each in a certain position, having a certain intensity and a certain frequency spectrum (a certain color, we can say). We then find the position of the image of each of these point sources and its intensity. Finally, we build the extended object image as the set of these images. We can perform this mental decomposition and recombination process because, under the usual conditions of dealing with thermal sources, the point sources are incoherent with one another.

Contrastingly, coherence conditions exist when objects are illuminated by laser light, but also, in some instances, in the microscope. The objects at which one looks through a microscope are strongly illuminated from below by a device, called a light condenser. The condenser consists of an intense light source, usually of small dimensions, a diaphragm used to adjust the light intensity, and a lens that concentrates the light on the object. When the diaphragm is very narrow, the

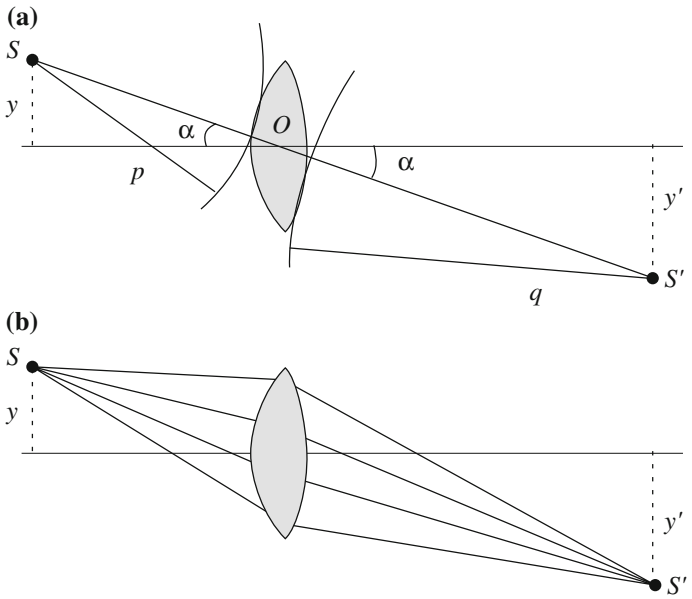


Fig. 7.13 Image formation of an off-axis point source. **a** Wave view, **b** ray view

transmitted light is spatially coherent and the different details of the object scatter light with fixed phase relations with one another. In this chapter, we shall discuss image formation under the usual conditions of incoherent illumination. Coherent illumination will be studied in the next chapter.

Obviously, if the object is extended, not all its points are on the axis. We start by considering a point source S that is not on the axis and determine the location of its image S'. This is shown in Fig. 7.13a from the wave point of view, and in Fig. 7.13b, from the geometric optics view.

To fix the ideas, let us think of our source as having the shape of a luminous wire normal to the axis. The point source S in Fig. 7.13 is a point of the wire, and S' is its image. To see the image, let us place a white screen on the other side of the lens normal to the axis. We choose the distance considering that one of the points of the object is on the axis and we know where its image falls. We place the screen in that position. If we look at the screen, we see an image geometrically similar to that of the wire, namely with the same shape, but possibly different dimensions. This means two things: first that the light emitted by each point of the wire is focalized by the lens at an image *point*, second that the geometrical relations between the image points are geometrically similar to those of the corresponding ones in the object.

We express this property by saying that the lens provides undistorted images of the objects. This property is true under the hypothesis of dealing with *paraxial rays*. We recall that this amounts to the following two assumptions; (a) the angle under

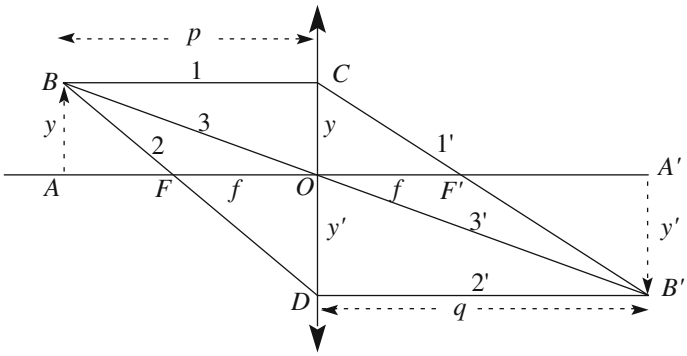


Fig. 7.14 Ray optic geometry for the image formation by a lens

which the object is seen by the lens (α in Fig. 7.13) is small, (b) the distance of the rays from the axis is always small. We shall always work under these conditions and continue to assume, in addition, as we did in the previous sections, that the beam opening angles, namely the angles under which both the source and the image see the lens, are small as well.

Let us now compare Fig. 7.13 with Fig. 7.9. The situations are similar, with the line SOS' being in the place of the axis in the former. Indeed, we obtain the situation we are now considering by taking that in Fig. 7.9 and simply rotating the lens by the angle $-\alpha$. By intuition, we understand that nothing changes appreciably if α is small. Clearly, we should continue to have an image in S' when we rotate the lens a bit.

Let us now see how to construct the image of the wire. In Fig. 7.14, the wire is represented by a dotted arrow of length y. The simplest argument uses geometrical optics. In Fig. 7.14, F and F' are the focuses, and we have adopted the usual convention to represent the (thin) lens as a double-arrowed segment normal to the axis. Under this convention, the two tips are directed outside if the lens is convergent, inside if it is divergent.

Let us look for the image of an off-axis point of our object, for example, its extreme B. One ray that is easy to track is the one running parallel to the axis, namely ray 1 in the figure. Indeed, we know that, beyond the lens, it goes through the second focus F'. Well, the image must lie on the ray we have called $1'$ in the figure. To find it, we need a second ray. As a matter of fact, there are two other rays that are easily drawn. We shall consider both of them for redundancy.

Ray 2 is the ray through the first focus F, which, we know, travels parallel to the axis beyond the lens. This is $2'$ in the figure. The image B' of B must belong both to $1'$ and $2'$ and, consequently, is their intersection. An alternative way to reach the same conclusion is to consider the ray through the center of the lens, which is 3 in the figure. The lens does not deflect this ray, because its two surfaces are parallel planes at the point where the ray crosses. Consequently, the emerging ray has the same direction as the incident one, being only laterally displaced. But the lateral

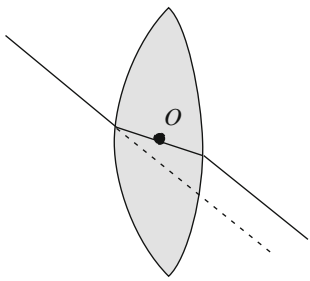


Fig. 7.15 A ray through the center of the lens

displacement is also negligible if the thickness of the lens is small, as it is, being that the lens is thin. This is shown in Fig. 7.15.

Looking at the geometry in Fig. 7.14, we can find the thin lenses equation in Eq. (7.16) with arguments from geometrical optics. The triangles CDB' and COF' being similar, we have

$$\frac{y+y'}{q}=\frac{y}{f} \tag{7.18}$$

Triangles CDB and DOF being similar as well, we also write

$$\frac{y+y'}{p}=\frac{y'}{f}. \tag{7.19}$$

Merging the two equations, we obtain

$$(y+y')\left(\frac{1}{p}+\frac{1}{q}\right)=\frac{y+y'}{f}, \tag{7.20}$$

namely

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f},$$

which is the thin lenses equation. But we can get more from the geometry. As we see in Fig. 7.14, the length of the image, namely y', is different from the length of the object y. The ratio of the two, $m = y'/y$, is independent of the position of the object and is called the *transversal magnification* of the lens. Indeed, taking the ratio of Eqs. (7.18) and (7.19), we have

$$m=\frac{y'}{y}=\frac{q}{p}. \tag{7.21}$$

We also notice that, under the discussed conditions, the image is inverted (upside down) compared to the object. In other cases, which we leave to the reader as an

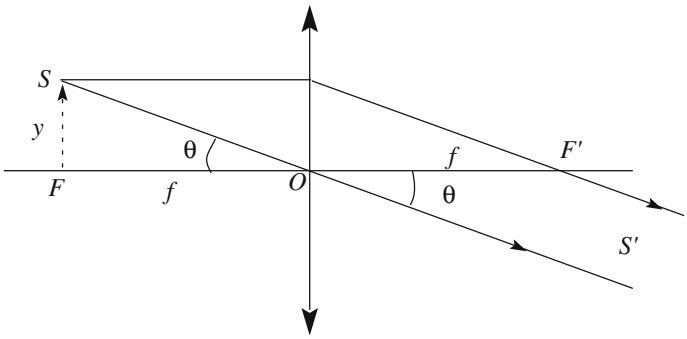


Fig. 7.16 Source in the first focal plane at the distance y from the axis

exercise, the image is erect. Equation (7.21) is valid in any case, provided we add the following sign convention to the list from the preceding section: (5) the transverse dimensions y of the objects should be taken as positive upwards and negative downwards, those of the images y' should be taken as positive downwards and negative upwards.

The planes perpendicular to the axis through the focuses are called the *focal planes*. The focal planes have two important properties that we now find using a geometrical construction similar to that of Fig. 7.14.

Consider a point source S on the first focal plane at the distance y from the axis, as in Fig. 7.16. The figure clearly shows that the light from S is transformed by the lens in a parallel beam forming, with the axis, the angle θ given by

$$\tan\theta = \frac{y}{f}. \tag{7.22}$$

In the paraxial rays approximation, the angles are small, and we can write approximately

$$\theta \cong \frac{y}{f}. \tag{7.23}$$

The second property of the focal planes is relative to the opposite path direction. Figure 7.17 shows a parallel beam incident at the angle θ with the axis, namely from a point source at infinite distance in that direction. The beam is focalized by the lens at the point of the back focal plane at the distance y' from the axis given by

$$y' = f\tan\theta, \tag{7.24}$$

which becomes, in the paraxial rays approximation,

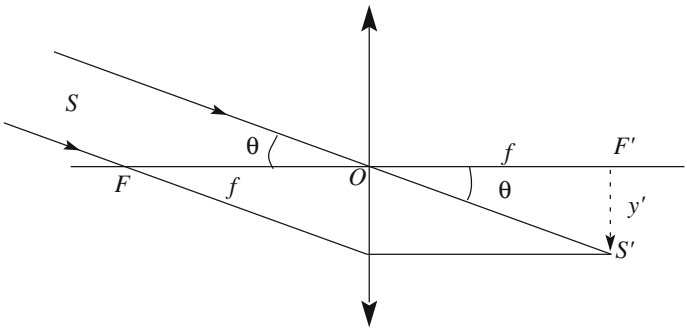


Fig. 7.17 Image of a parallel beam incident at an angle

$$y' \cong f\theta. \tag{7.25}$$

This is easily understood looking at Fig. 7.17 and is an immediate consequence of the path reversibility.

7.8 Aberrations

The light waves we have considered up until now have been monochromatic and perfectly spherical or plane. It will not be surprising to discover that these ideal conditions are seldom met in practice. Indeed, any difference relative to the ideal case may be relevant if it is on the order of the wavelength, which, as we know, is about one half of a micrometer. We can distinguish two types of deviations from perfection, called *aberrations* and *irregularities*. We shall discuss the former in this section, the latter in the next one.

The study of aberrations in optical instruments is indeed one of the most important chapters in optics. Every aberration can be controlled and corrected up to the requested level by using appropriate combinations of optical surfaces and media. Indeed, the objective lenses of a good microscope, telescope and photo-camera are composed of several lenses with surfaces of different curvatures and glasses of different refractive indices. Here, it will be enough to briefly discuss the most important aberrations. In addition, we shall find a criterion for establishing when the remaining differences from the perfect shape are so small as to be irrelevant.

Let us start by considering the consequences of the non-monochromaticity of light. As a matter of fact, under the majority of circumstances, light is white. In Sect. 7.5, we found the image of a point source considering the phase velocity of the wave in the lens. Equation (7.15) shows us that the power of the lens is different for different wavelengths. Namely, upon exiting the lens, the waves corresponding to different wavelengths have different curvatures. Their centers fall in different

positions. As a result, a non-monochromatic point source has several (infinite) images, each of a different color in a slightly different position. This unwanted effect is known as *chromatic aberration*. Due to chromatic aberration, the image of a white point appears as a luminous spot, white in the center, colored on the borders. The simplest way to correct a chromatic aberration is using two lenses, one positive and one negative, cemented together one next to the other, made of different glasses. The two glasses (crown and flint, to be precise) have different dispersion properties and are designed so that the chromatic aberration of one cancels out that of the other. This can be done exactly for two wavelengths, which are chosen at the extremes of the spectrum, namely in the blue and the red. This system is called an achromatic doublet. Its invention should be credited to John Dolland (UK, 1706–1761), who patented it in 1758. This was almost one and a half century after Galileo Galilei had given birth to astronomical observations with the telescope and 82 years after the Cassini-Rømer astronomical determination of the speed of light. The idea can be brought forward by building a triplet, with which the chromatic aberration can be corrected exactly at three wavelengths.

In Sect. 7.5, we took the approximation to consider the wave outgoing from the lens as spherical, but it is not exactly so. In fact, the outgoing wave does not really have a single center. The lines normal to its surface, namely the rays in an isotropic medium, do not converge at a single point, but envelope a surface called a *caustic*. A section of the surface, which is clearly symmetric about the axis, is shown in Fig. 7.18. Namely, the rays closer in angle to the axis (paraxial rays) meet at points of the axis farther from the lens than those farther from the axis (marginal rays), as shown in the figure. This is the *spherical aberration*. The aberration is all the more important the larger the opening angle of the incident wave.

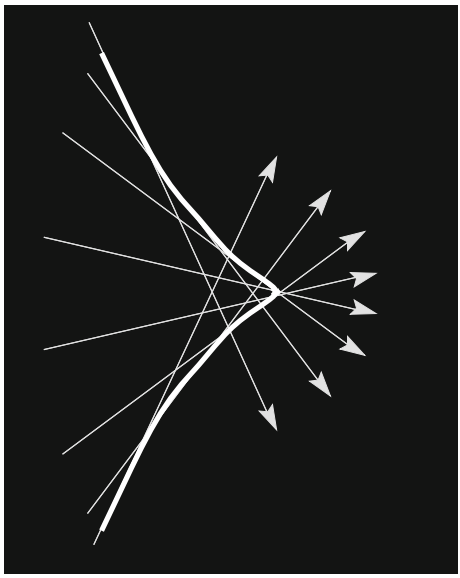


Fig. 7.18 The caustic

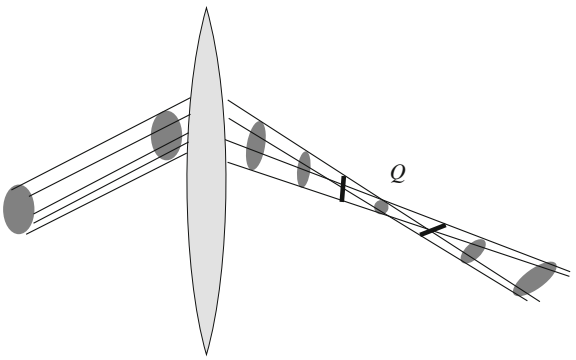


Fig. 7.19 Astigmatism

We can easily observe the caustic on a sheet of paper near to the image point having a beam of light of large angular aperture incident on a lens.

The spherical aberration can be corrected using systems of lenses one next to the other, properly designed using the aberration theory, and computer codes based on it.

Astigmatism is an aberration occurring, even with beams of narrow angular aperture, when the incidence angle is large. To observe the aberration, we should prepare a narrow parallel light beam and have it incident on the lens at a large angle with the axis, as shown in Fig. 7.19. If the lens were perfect, all the rays would converge at a single point in the back focal plane, which is Q in the figure. Indeed, this does not happen. The figure shows a number of sections along the beam after the lens. Initially, the section is an ellipse that, moving forward, gradually becomes flatter up to the point of being a line element. Further forward along the beam, the section is again an ellipse, whose minor axis gradually grows to become a small circle. This is called the circle of least confusion. Beyond the circle, the section is again an ellipse, now with the direction of the major axis, being that the one ahead of the circle was the direction of the minor axis, and further down, a line segment normal to the first one.

The *coma* aberration is a combination of astigmatism and spherical aberration, which becomes important when the incident beam has both a large angular aperture cone and an average direction at a large angle with the axis. The name comes from the shape of the image of a point-source that appears as a blur, resembling a comet.

We shall not discuss further aberrations here, but shall pose the following problem. The consequence of any aberration is a wave surface that is not perfectly spherical, but rather differs from that to some extent. Clearly, the smaller this difference, the smaller the aberration. Well, we shall now see that a value of this difference exists below which the non-spherical wave cannot be distinguished from the spherical one. Below this value, the aberration is not observable and, as a consequence, is irrelevant.

Let us fix our attention on the physical meaning of a wave surface. We have defined it as the locus of the points at which the phase has the same value. However, this is a geometrical surface. To give it a physical meaning, we should

define the set of operations to be performed to localize it. The most accurate method for doing that is to employ an interference phenomenon. We produce a "reference wave," of which we know the phase to be a function of position and time, and make it interfere with the wave surface under examination. Looking at the resulting interference pattern, we can determine the phase difference between the waves, and hence the unknown phase, from our knowledge of the reference wave. In the maxima of the interference fringes, the two waves have the same phase (modulo 2π), and in the minima, opposite phases. Suppose now that we change the phase of the wave under analysis at all the space points by the same quantity $\Delta\phi$, without changing the reference phases. We will see the entire system of fringes rigidly shifting perpendicularly to their direction. If, for example, $\Delta\phi = \pi$, the dark fringes will take the place of the luminous ones and vice versa. This is clearly observable. However, if $\Delta\phi$ is small enough, the lateral shift of the fringes is so small as to be unobservable. The limit depends somewhat on the sensitivity of our instrument, but in practice, a shift of the fringes of one quarter of the distance between two dark fringes is at the visibility limit. We shall consequently define the minimum observable phase difference as $\Delta\phi = 2\pi/4 = \pi/2$.

In conclusion, we are able to determine the phase at each point of a wave within a precision of a $\pi/2$, but not smaller. Symmetrically, if we want to localize a point at which the phase has a certain value, we can do so only with an uncertainty $\lambda/4$ in the direction normal to the wave surface. This means that, physically, the wave surface is not a geometrical surface, but is defined within a thickness equal to $\lambda/4$. Hence, any surface differing from a spherical surface by elevations or depressions of height smaller than $\lambda/4$, as shown in Fig. 7.20, cannot be distinguished from the perfect spherical surface. This statement is called the *quarter-wave criterion* and was established in 1893 by John William Strutt, Baron of Rayleigh (UK, 1842–1919).

In conclusion, if the aberrations of an optical system have been corrected to the point that the resulting wave surfaces differ from the ideal ones by less than $\lambda/4$, the difference is not measurable. We say that the system is *optically perfect*. As a matter of fact, "perfection" rarely exists in physics, but it does exist in optics, being a consequence of the non-zero wavelength.

The quarter-wave criterion has a general validity and we shall use it in different instances in the subsequent sections.

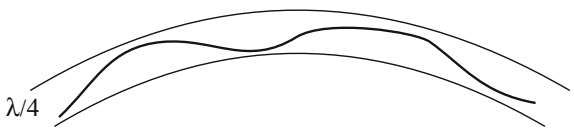


Fig. 7.20 The irregular shape of a wave surface cannot be distinguished from the perfect one if the differences are smaller than $\lambda/4$

7.9 Irregularities

The surfaces of lenses and mirrors must be worked with the precision needed to avoid deforming the transmitted, or reflected, waves. In these cases, we speak of irregularities instead of aberrations. It is a question of semantics, the substance of the problem being the same.

We shall now employ the quarter-wave Rayleigh criterion to evaluate the maximum value of the irregularities on the surface of a lens for optical perfection.

Consider, for the sake of simplicity, a plane wave normally incident on a plate with parallel faces of refractive index n_1. Let $n_2 < n_1$ be the index of the medium after the lens (if it is air, then $n_2 \simeq 1$). Our arguments can be immediately extended to spherical waves and spherical boundary surfaces. Suppose that the second surface of the plate has an irregularity, say, a bump of height δ (Fig. 7.21). The segment of the wave beyond the plate corresponding to the top of the bump has traveled a path length longer through its medium by δ relative to the segments outside the bump. On the length δ, the velocity was c/n_1, while the other parts of the wave were traveling at the speed c/n_2. It follows that the wave surface at the top of the bump lags behind the rest by a length

$$\left(\frac{\delta}{c/n_1} - \frac{\delta}{c/n_2}\right) c = \delta(n_2 - n_1).$$

We can then state, for the quarter-wave criterion, that the deformation is not observable if $\delta(n_2 - n_1) \leq \lambda/4$, namely if the irregularity is

$$\delta \leq \frac{\lambda}{4(n_1 - n_2)}. \tag{7.26}$$

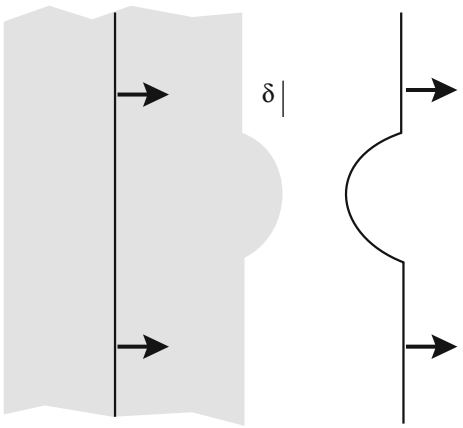


Fig. 7.21 A surface irregularity at the interface between media and the consequent distorted outgoing wave surface

For example, in the case of glass with the typical value $n_1 = 1.5$ immersed in air ($n_2 = 1$), the condition is

$$\delta \leq \lambda/2. \tag{7.27}$$

Namely, the irregularity should be corrected at better than about 0.2 μm. Interferometric methods give us control if this is the case. Notice that the maximum allowed value for δ is that much smaller the larger the index difference $(n_2 - n_1)$ between the lens and the medium in front of it. Surfaces satisfying Eq. (7.27) are called optical surfaces or optically perfect surfaces.

Here, we propose an interesting example. Nature has been unable to work out the external surface of the cornea, which is the first lens of our eyes, as optically perfect. Well, she then performed a trick; the first surface facing the air is not the cornea, but a tear film, whose surface is made perfect by surface tension (and is periodically renewed by blinking). In this way, the index difference between tear film ($n_2 = 1.36$) and cornea ($n_1 = 1.38$) is as small as $n_2 - n_1 = 0.02$, and the maximum allowed irregularity of the cornea surface is about 5 μm.

We will just mention here that modern interferometric measurements of the objective lenses used by G. Galilei in the telescope he developed in 1609 and used for his astronomical discoveries have shown them to be optically perfect at a single wavelength (see Sect. 7.14).

With similar arguments, we can establish the tolerable limits for having an optically perfect mirror surface. Let us consider a plane wave normally incident on a plane mirror having a small bump of height δ, as shown in Fig. 7.22.

The deformation of the reflected wave is clearly 2δ. Hence, the surface is optically flat if $2\delta \leq \lambda/4$, namely if

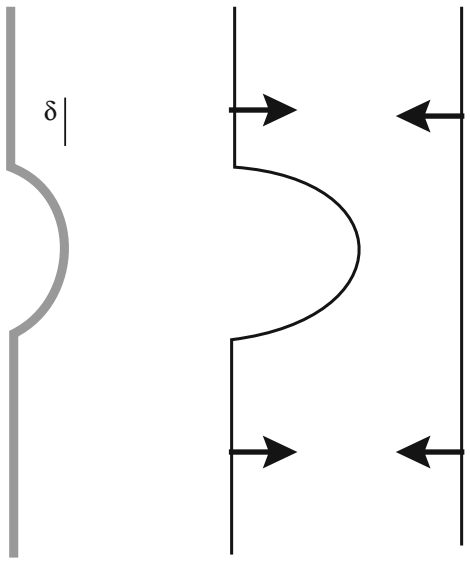


Fig. 7.22 A mirror surface irregularity and the consequent distorted reflected wave surface

$$\delta \leq \lambda/8. \tag{7.28}$$

We see that the limit is substantially more demanding than it would be for a lens surface.

7.10 Depth of Field and Depth of Focus

Optical instruments, like lenses and mirrors, and the eyes of living beings are used to produce images of three-dimensional objects on a two-dimensional sensitive surface. Let us consider, to be specific, a thin lens. We know that if f is the focal length of the lens and p is the distance between the lens and the sensitive surface, only the points in the object space at the distance q, given by the thin lenses equation in Eq. (7.16), should form a neat image, or, as we say, be in focus. The images of the points of the objects farther or closer than p will fall in front or behind the sensitive surface. As a consequence, the images of these points will be small disks, rather than points, of larger diameters for larger differences from p of the object point. However, in practice, in every image forming system, there is always a larger or smaller interval of distances about the nominal value p within which the images will appear of a good quality. As a matter of fact, both possible residual aberrations and diffraction, which cannot be eliminated, result in the image of a point never being a point, but rather a small disk.

We call the above-mentioned range of distances in front of the lens around the nominal p containing the objects that produce an acceptable image on a sensitive surface at the distance q after the lens the *depth of field*. Reciprocally, we call the interval of distances behind the lens around the nominal q wherein the sensitive surface can be placed to produce an acceptable image the *depth of focus*. The two concepts are different, but closely correlated to one another. So much so that the term 'depth of focus' is often used with both meanings.

Clearly, the definitions just given have a certain degree of arbitrariness. However, the very fact that light is a wave puts a non-zero lower limit to the depth of field (and of focus). Indeed, for a given optical instrument and a given wavelength, it is impossible *in principle* to establish whether the distances of two point sources are different or not, if this difference is smaller than a certain value, which we shall now find.

Let us consider a lens of diameter D and two monochromatic point sources S_1 and S_2 on the axis at distances p_1 and p_2 from the lens, respectively. Let λ be their wavelength. The waves emitted by the two sources are spheres of radiuses p_1 and p_2 when they touch the lens on the axis, as shown in Fig. 7.23. The quarter-wave criterion tells us that the two waves are indistinguishable if the maximum distance between them is less than $\lambda/4$. Under these conditions, the images of S_1 and S_2 are not resolved. To find these conditions, we use, once more, Eq. (7.1) with reference

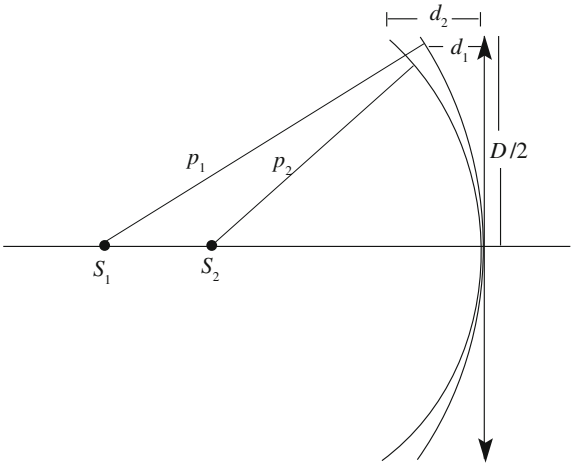


Fig. 7.23 The waves emitted by two point sources on the axis when they touch the lens

to the two arcs in Fig. 7.23. Namely, we write $d_1 = D^2/(8p_1)$ and $d_2 = D^2/(8p_2)$. The quarter-wave criterion gives us the condition

$$d_2 - d_1 = \frac{D^2}{8}\left(\frac{1}{p_2} - \frac{1}{p_1}\right) \leq \frac{\lambda}{4}. \tag{7.29}$$

A useful approximate expression can be written when p_1 and p_2 are not very different from one another. In this case, let p be their mean value and Δp the absolute value of their difference. We can approximate $1/p_2 - 1/p_1 \approx \Delta p/p^2$ and Eq. (7.29) becomes

$$\Delta p \leq \frac{p^2}{D^2} 2\lambda = \frac{2\lambda}{\alpha^2}$$

where $\alpha \simeq D/p$ is the angle under which the lens is seen from the sources, which we have assumed to be small. In this approximation, the depth of field is

$$\text{DOF} = \frac{p^2}{D^2} 2\lambda = \frac{2\lambda}{\alpha^2}. \tag{7.30}$$

Let us discuss the result. The depth of field depends on two elements. It is proportional to the wavelength of the light, because the uncertainty intrinsic to the wave phenomena is larger for larger wavelengths. It is inversely proportional to the square of the lens diameter, or, in other words, to the solid angle under which the lens is seen by the object. In general, optical systems include an adjustable circular diaphragm. Clearly, if we diminish the diameter of the diaphragm, we gain in depth of field (losing luminosity).

Suppose now that the source S_1 is at infinite distance, namely let us take $p_1 \to \infty$. The wave incident from S_1 is plane. Hence, the image of S_1 falls on the back focal plane of the lens. We now ask how close we can have another source, say, S_2, to have its image still be on the focal plane. This means that the incident spherical wave of radius p_2 must be indistinguishable from a plane wave. The answer is given by Eq. (7.29) for $p_1 \to \infty$. We have

$$p_2 \geq \frac{D^2}{2\lambda}. \tag{7.31}$$

Any object at a distance above this limit is, for the lens, at infinity. This distance is called the *hyperfocal distance*. Once again, the hyperfocal distance is smaller (closer objects are still in focus when the lens is focused at infinity) when its diameter is smaller.

Let us consider, as an example, the hyperfocal length of the human eye. One of the optical elements in the eye, the lens, has a focal length that can be adjusted (within certain limits). We do that unconsciously, tightening the muscle that changes its curvature. When the muscle is relaxed, the focal length is at its largest and the images of far off objects fall on the retina. The question is, until when does the muscle remain relaxed when the object gets closer and closer? (Indeed, at shorter distances, the image remains in focus because we unconsciously change the power o the lens). Note that the human eye has a diaphragm, namely the pupil, whose opening varies depending on the illumination. Taking a typical $D = 2$ mm and, in a round number, $\lambda = 0.5$ μm, the hyperfocal length is 4 m.

Question Q 7.1. What is the ratio of the hyperfocal distances of two lenses of diameters 1 mm and 1 cm?□

In concluding this section, we call the attention of the reader to the fact that the definitions we gave of depth of field and hyperfocal distance have been based on the limits imposed by the wave nature of light. In the presence of aberrations, the depth of field is larger and the hyperfocal length is shorter.

7.11 Resolving Power

Consider a plane monochromatic wave incident on a lens in the direction of the axis. The lens transmits only a part of the wavefront, namely a spherical cup having the diameter D of the lens (or of the diaphragm, if smaller). As we know, diffraction always exists when a waveform is limited. As a consequence, even in the absence of aberrations, the wave surface after the lens is not exactly spherical with its center in the focus. Indeed, in the focal plane, we have the Fraunhofer diffraction pattern of the circular aperture that is the opening of the lens. If we put a white screen on the focal plane, we do not see a point, but rather the fundamental diffraction pattern, namely the bright Airy disk surrounded by the very weak rings of the secondary maxima. In practice, the rings can be neglected under the usual conditions. We

recall that the radius of the Airy disk, namely the radius of the first dark ring, is seen from the lens under the angle

$$\Gamma = 1.22\frac{\lambda}{D}. \tag{7.32}$$

Being that $\lambda/D \ll 1$, the radius of the disk is

$$\Sigma = 1.22f\frac{\lambda}{D}. \tag{7.33}$$

This radius is usually very small as well, and the Airy disk often appears almost as a point. This is the case, for example, with the pictures we shoot with our cameras. In principle, however, the wave nature of light implies that the images of the objects are not made of points, but of Airy disks. The optical images have an intrinsic granularity determined by diffraction. This implies that it is not possible to resolve details smaller than a "grain" diameter.

Let us look more closely at this important issue. Consider two monochromatic point sources at great distances, sending two plane waves to the lens, one in the direction of the axis, one at an angle θ with it, as shown in Fig. 7.24. The distance σ between the two images, which are on the focal plane, is $\sigma = \theta f$. Now, if σ is small enough, the two Airy disks of the two images may overlap so much as to be indistinguishable from a single disk. We say that the images are not resolved. The criterion for the limit resolution was given, again, by Lord Rayleigh, and states that "the images of two point sources *of equal intensity* are just resolvable when the center of the diffraction pattern of one falls over the first minimum of the diffraction pattern of the other." We have emphasized "of equal intensity," because when the intensities are very different, more distance is needed to be able to resolve the images.

Everything that we have said about plane waves is easily extended to spherical waves. The only difference is that the image plane is not the focal plane, but a plane at the distance $q \neq f$ beyond the lens. The half aperture angle Γ given by Eq. (7.32) does not depend on q and, as a consequence, the radius of the diffraction disk varies, being equal to $1.22q\lambda/D$.

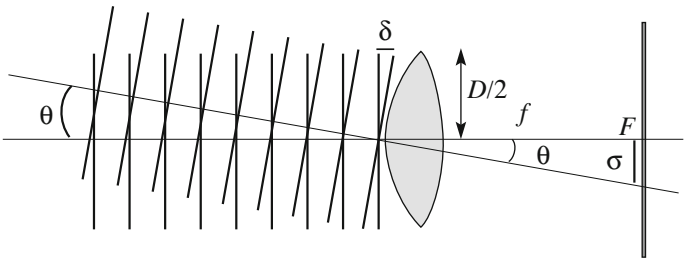


Fig. 7.24 Two incident plane waves, one along the axis, one at an angle θ

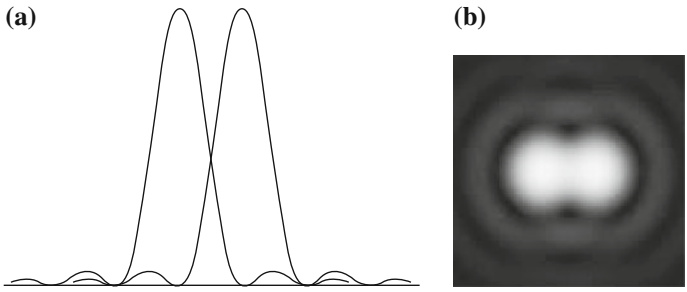


Fig. 7.25 **a** Intensity profiles of the images of two point sources at the resolution limit; **b** corresponding images

Figure 7.25a shows the diffraction intensity profiles for two equal point sources at the resolution limit. Figure 7.25b is a photograph of the corresponding images.

In conclusion, the images of two point sources of the same intensity are resolved if they are seen by the lens under an angle larger than Γ given by Eq. (7.32). We call the reciprocal of this angle the *resolving power* of the lens, namely

$$\frac{1}{\Gamma} = \frac{D}{1.22\lambda}. \tag{7.34}$$

Hence, the resolving power is higher if the diameter of the lens is larger and if the wavelength is smaller. Even now, one factor depends on the instrument, one on the light.

It is convenient to appreciate the orders of magnitude. Consider, for example, the human eye. Let D be the diameter of the pupil of which we take the typical value to be $D \approx 2$ mm. With $\lambda \approx 0.5$ μm, the resolving power is $\Gamma = 1.22\lambda/D = 0.3$ mrad, which is about $1'$. It is then said that the human eye resolves at 1 min. This means that two points seen under an angle smaller than a minute are not resolved. The images are formed on the retina, where the sensitive elements are cells of about 2 μm diameter. Indeed, the sensitive elements should be small in order to detect the details, but it would be of no use to have them smaller than the Airy disk of a point source. The dimensions of the sensitive elements of our eyes are just these.

Question Q 7.2. What is the resolving power of a telescope having one meter diameter objective?□

From the discussion in which we have just engaged, we can answer another question, namely what do we mean when we talk of a point source? Well, the answer immediately follows from the above discussion: a point source is a source seen under an angle smaller than $1.22\lambda/D$. Stars, for example, are very large objects, but they are point-like to our telescopes, because they are very far away and the angle under which we see their diameter is smaller than that limit.

We conclude the section with a further analysis of what we have established. Let us go back to the situation shown in Fig. 7.24, in which two plane waves are

incident on the lens and an angle θ with one another. If $\theta < 1.22\lambda/D$, the two sources are not resolved, namely the two wavefronts are not distinguishable.

We have considered a similar situation in Sect. 7.8, in which we exploited interference to distinguish two wavefronts and established them to be distinguishable if the maximum distance between them is larger than $\lambda/4$ (quarter-wave criterion). Here, we found a different criterion based on diffraction. Let us compare the two criteria.

Figure 7.24 shows that the maximum distance between the two wavefronts is $\delta = D\theta/2$. We stated that they are distinguishable if $\theta > 1.22\lambda/D$, namely if

$$\delta \geq 1.22\frac{\lambda}{2} = 0.61\lambda. \tag{7.35}$$

We conclude that the diffraction-based criterion is less sensitive than the interference-based criterion by a factor of 2.44. This is a consequence of having now considered only the central maximum of the diffraction pattern, neglecting the rings. This, however, allows us to use white light, while the interference-based methods require monochromatic light.

7.12 Nature of the Lens Action

In Sect. 7.5, we saw that the action of a lens on a light wave is to change the form of the wave surface. This is the result of the phase velocity in the lens being different from that in the surrounding medium and of the path length inside the lens being a correct function of the distance from the axis. An in-depth analysis will now allow us to understand the fundamental aspects of the processes leading to the image of a point-like object (and of any object, by generalization). We shall then be able to see how the process can be implemented with means different from those of a lens.

Consider a plane monochromatic wave normally incident on a thin lens, which we take, to be specific, as plano-convex. Let R be the radius of its second surface and d its thickness (in the center). We take the origin of the reference axes in the center of the lens, the z-axis on the optical axis and the x and y axes on the lens plane, as in Fig. 7.26. Let n be the refractive index of the lens and let the lens be immersed in air (refractive index equal to 1). The electric field of the incident wave can be written as

$$E = E_0 e^{i(\omega t - kz + \alpha)} = E_0 e^{i\alpha} e^{i(\omega t - kz)}, \tag{7.36}$$

where α is the initial phase and the quantity $A_i = E_0 e^{i\alpha}$ is the complex amplitude of the wave. Here, we consider the lens to be a *phase diaphragm*. Indeed, the lens is transparent and does not change the absolute value of the amplitude; it has a radially varying thickness, and consequently changes the phase as a function of the distance from the axis ρ. The amplitude transmission coefficient is a function of ρ that we

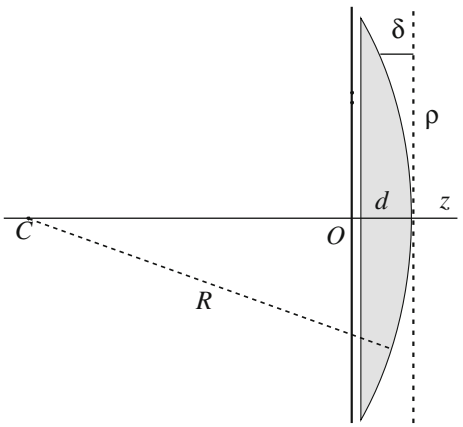


Fig. 7.26 A lens like a phase diaphragm

call $T(\rho)$. Let us find it. The path of the wave runs partially through air and partially through the glass. The latter, δ in the figure, is a function of ρ. As usual, the angles are small, and we can use Eq. (7.1), giving us $\delta = \rho^2/(2R)$. The phase delay $\Delta\phi$ introduced by the lens at ρ is then

$$\Delta\phi(\rho) = kn(d-\delta) + k\delta = k(n-1)\frac{\rho^2}{2R} + knd. \tag{7.37}$$

Calling $\Delta\phi_0 = knd$ the phase delay in the center, the complex amplitude at the exit is

$$A_o(\rho) = A_i e^{-i\left(\Delta\phi_0 + \frac{k\rho^2 n-1}{2\ R}\right)}. \tag{7.38}$$

Clearly, the effect of $\Delta\phi_0$ is to introduce a phase delay independent of ρ that is just like changing the initial phase everywhere, and is consequently immaterial in the image formation process. The amplitude transmission coefficient is the ratio between the amplitudes after and before the diaphragm. Hence, it is

$$T(\rho) = e^{-i\left(\Delta\phi_0 + \frac{k\rho^2 n-1}{2\ R}\right)}.$$

Here, we recognize the expression of the focal length in Eq. (7.15), which, in the present case, is $f = R/(n-1)$, and write the transmission coefficient as

$$T(\rho) = e^{-i\Delta\phi_0} e^{-i\frac{k\rho^2}{2f}}. \tag{7.39}$$

We have conducted our reasoning using a particular thin lens, but the conclusion in Eq. (7.39) is valid for any of the shapes of Fig. 7.8, as is easily established.

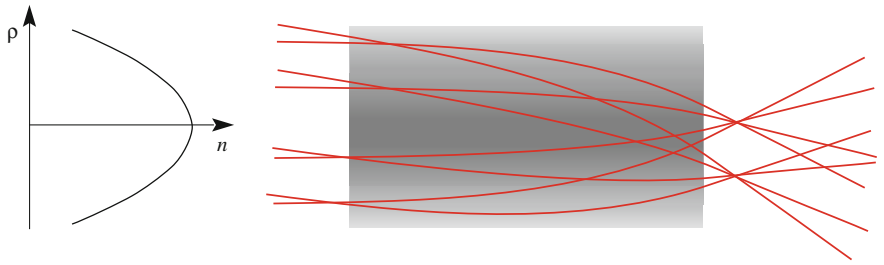


Fig. 7.27 A graded index lens and its effect on light beams incident in different directions. On the left panel, the refractive index as a function of the distance from the axis

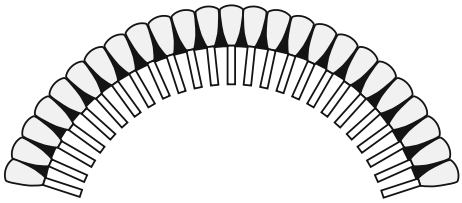


Fig. 7.28 Scheme of the composite eye of the horse shoe crab, showing the GRIN elements and the sensors

In conclusion, to obtain the focalizing effect, and the image formation as a consequence, an instrument must produce a phase delay *proportional to the square distance from the axis.*

A lens produces a phase delay by having a thickness function of the distance from the axis. The same effect can be reached with a transparent plate or cylinder of constant thickness, but with a refractive index varying proportionally to ρ^2. Figure 7.27 shows such a cylinder and two parallel beams incident at different angles. Such an element is called a *graded index lens*, or GRIN.

Nature employs this principle for the eyes of some aquatic insects and crustaceans. In water, in fact, the difference between the refractive indices outside and inside the eye may be too small to produce appreciable deviations of the rays by the curvature of the eye surfaces. By 1891, Sigmund Exner (Austria, 1846–1926) had already described the lateral eyes of the horseshoe crab *Lumulus polyphemus*. The eye is composite, being an array of elementary imaging elements, called ommatidia, as shown in Fig. 7.28. Each ommatidium is a graded index lens with a focal plane at its end where the light sensor is located.

GRINs have several applications in modern technology as well. They are particularly useful where many very small lenses need to be mounted together on a plane or other surface. This is the case, for example, with photocopiers and scanners.

7.13 Magnifying Glass

Optical instruments are used to observe images as enlarged or with better resolution than could be obtained with the naked eye. In this section and in the subsequent two, we shall discuss the principles of the three basic instruments: the magnifying glass, the dioptric telescope and the microscope. Indeed, magnification and resolution are the two basic properties of any optical instrument. Both must be properly considered. For example, it useless to magnify a photograph at increasing levels, in hopes of seeing the leaves on very far trees. Even when the granularity of the film can be ignored, above a certain magnification, the images of the details that we magnify become blurred, because we are already observing the Airy disks, which we keep magnifying as well.

The light waves emerging from an optical instrument enter into the eye of the observer to produce the final image on his/her retina. Consequently, in any case, the observer's eye is an integral part of the optical system. Here, we can think of the eye as being a converging system of lenses, one next to the other, and a sensitive film, namely the retina. The system is equipped with a pupil, having a diameter varying, depending on the illumination conditions, between about 2 mm and 1 cm. The focal length of the eye can be adjusted within certain limits. Some quantitative data, with reference to a "normal" eye, will be useful. When the system is relaxed, the retina is on the focal plane, and we see sharp images of objects at distances equal to or larger than the hyperfocal distance, which we saw to be about 4 m. The eye can be adjusted to focus on retina objects as close as 150 mm, but below 250 mm, sight becomes difficult. To see the details of a small object, the best distance at which to look at it is about 250 mm. Consequently, the *least distance of distinct view* (LDDV) or, alternatively, the *reference seeing distance* (RSD), is defined as $L_d = 250$ mm.

Let us consider the magnifying glass, which is a simple convergent lens, of focal length f, placed before the eye that we use to look at small objects. Let y be the size of the object. If we put it in the focal plane of the lens, the waves emerging from the lens are plane and the eye can look in its relaxed state without effort. As shown by Fig. 7.29, under these conditions, we see the object under an angle θ', which, considering it to be small, is given by $\theta' \cong y/f$. We now compare this angle with the angle under which we would see the object if it were at the shortest distance for a distinct view, which is, obviously, $\theta \cong y/L_d$.

Clearly, $\theta' > \theta$ if the focal length of the lens is smaller than L_d or, in other words, if its power is larger than about four diopters. We define the *angular magnification* of the magnifying glass as the quantity

$$J = \frac{\theta'}{\theta} = \frac{L_d}{f}. \tag{7.40}$$

In practice, considering that the diameter of a magnifying lens is typically several centimeters, it is difficult to produce lenses with very short focal lengths,

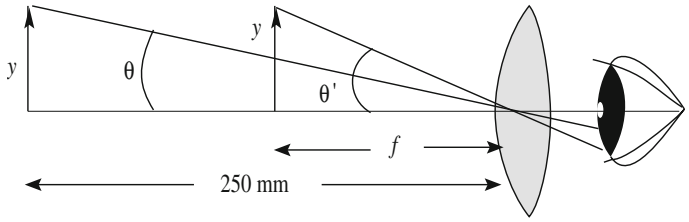


Fig. 7.29 An object view directly at the LDDV and through a magnifying glass

say, smaller than one centimeter. The most common magnifying glasses have dioptric power between 10 and 40 diopters, and consequently magnify between 2.5 and 10 times.

7.14 Telescope

The telescope, as the name suggests, is an instrument intended for the observation of distant objects. There are basically two types of telescope: the catoptric telescope, consisting of two lenses separated by a tube, the objective and the eyepiece, and the reflecting telescope, having a converging mirror in place of the objective. Here, we shall only discuss the former.

From the historical point of view, the origins of the telescope are unknown, but they are certainly to be found outside the academic world. Very likely, some spectacle maker, while testing the quality of his lenses, looking through two of them one next to the other in his hands, noticed that objects far away appeared magnified. Leonardo da Vinci (Italy, 1452–1519), in the *Atlantic code*, already mentions "*spectacles to make the moon larger*," and in 1538, Girolamo Fracastoro (Italy, 1478–1553) writes: "*If somebody looks through two spectacle lenses, one above the other, he will see everything much larger and closer*". The first written record of a magnification system of two lenses, one positive and one negative, is found in the 1589 edition of the treatise "*Magia naturalis*" by Giovanni Battista Della Porta (Italy, 1535–1615). In 1634, I. Beeckman (the rector of the Latin School of Dordrecht, Holland) wrote that a spectacle-maker, from which he was receiving lectures in optical technology, had reported to him of a telescope brought to the Netherlands from Italy that bore the date of 1590. Subsequently, telescopes were being built in the Netherlands by no later than 1604. These telescopes were basically toys, sold at fairs for amusement.[1]

Galileo Galilei (Italy, 1564–1642) was not the inventor of the telescope, but he is the originator of the telescope as a *scientific instrument*. As soon as news reached

[1]For historical details and documents, see, e.g., A. Van Helden, *The invention of the telescope*, Trans. Am. Phil. Soc., New Ser. Vol. 67, n. 4, pp. 1–67 (1977).

him in 1609 of the Dutch gadgets, he understood he might be able to transform the idea into a powerful scientific instrument. It was hard work, done without theoretical support in an experimental trial and error process, guided by logic and profiting of the glass technology that was particularly advanced in Venice. He finally succeeded in building and testing telescopes of magnification of 20 and even 30 and with diffraction limited lenses. The latter specification is of fundamental importance for guaranteeing the resolving power corresponding to the diameter of the lens, but required an ad hoc development of sophisticated technologies and testing procedures.

The theory of the telescope is credited to Johannes Kepler (Germany, 1571–1630). In 1610, having received one of the telescopes made by Galilei and used it for astronomical observation, Kepler thought *he had to* explain how it worked. Within a few months, Kepler published a treatise in 1611, *Dioptrice*, in which he not only mathematically explained the operation of the Galilei telescope, but, more generally, developed what we now know as geometrical optics. Among other things, he invented a telescope consisting of a positive objective and a positive eyepiece (the Keplerian telescope, also called an astronomical telescope), while the Galilean configuration, as those of all its predecessors, was made of a positive and a negative lens. We shall now describe the principles of the Keplerian telescope.

Like the magnifying glass, the telescope increases, by much larger factors, the *angle* under which the objects are seen. Let us start from the magnifying glass, whose magnification, as we saw, is limited by the maximum dioptric power feasible in practice. Note that the image seen by the eye through the glass is virtual. In order to obtain higher magnification, even with somewhat greater focal lengths, we might think of looking instead at a real image. Figure 7.30 shows how this idea might be implemented. The object is the segment S_1S_2 at the distance p from the lens. Its real image $S'_1S'_2$ is at the distance q beyond the lens. If we observe this image from, say, a distance r, we see it under the angle $\theta' \approx y'/r$.

Let us compare this situation to that which would occur without the lens. The eye would see the segment S_1S_2 under the angle $\theta \approx y/(p + q+r)$. Considering that we are now looking at far away objects and, hence, it is $p \gg q + r$, and we can write in a good approximation $\theta \approx y/p = y'/q$. In conclusion, the angular magnification under these conditions is

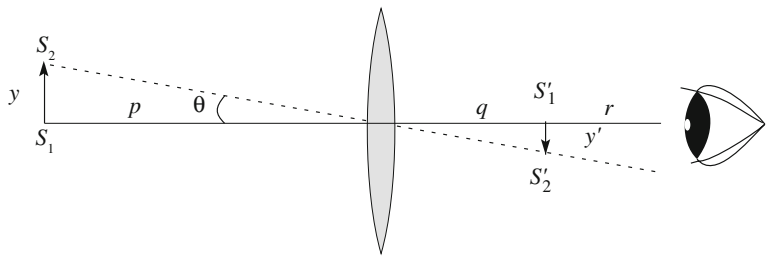


Fig. 7.30 Geometry for looking at the real image of a lens

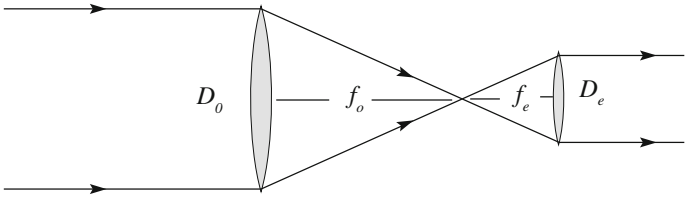


Fig. 7.31 The geometry of the Keplerian telescope

$$J = \frac{\theta'}{\theta} = \frac{q}{r}. \tag{7.41}$$

We see that the smaller the distance r from which we observe S'_1 and S'_2, the larger their angular separation. However, in practice, we cannot bring our eye too close, because it will not be able to adjust itself and the image will be blurry. We can solve this problem by employing, beyond the converging lens considered so far, which is called the objective, a second lens called the eyepiece, which we position between the real image of the objective and the eye. The distance between the lenses, namely the length of the instrument, depends, in general, on the distance of the object to be observed (and on the sight of the observer). The two lenses are soldered at the extremes of two tubes. One of the tubes enters into the other at their open ends. We then adjust the length, sliding one tube inside the other. We shall limit the discussion to the case of objects at infinity, namely the case of astronomical observations, in which, as we shall now see, the distance between eyepiece and objective is fixed. Figure 7.31 shows this situation.

Under these conditions, the objective produces the image on its back focal plane, namely at the distance f_o equal to its focal length. We position the eyepiece, of which we call f_e the focal length, such that its first focus coincides with the second focus of the objective. In this way, the eye is reached by a plane wave, adjusted to infinity and relaxed. Note that the diameter of the eyepiece is designed to intercept the complete beam transmitted by the objective, and no more. Indeed, making it larger would be useless, and making it smaller would reduce the resolving power, as we shall now see.

We are now in a situation similar to that of Fig. 7.30, with the eyepiece in the place of the eye. The angular magnification is then given by Eq. (7.41) posing $q = f_o$ and $r = f_e$, namely it is $J = f_o/f_e$. For the similarity of the two triangles in Fig. 7.31, we can also write the magnification in terms of the diameters D_o and D_e as

$$J = \frac{f_o}{f_e} = \frac{D_o}{D_e}. \tag{7.42}$$

In conclusion, the telescope allows us to observe two point sources under an angle larger than that with which they would be seen with the naked eye by the ratio of the diameters of the objective and the eyepiece. This is not sufficient, however, to claim that the two points are perceived as distinguished, or, as we say, resolved. As we saw in Sect. 7.11, to be resolved, the two points must be seen under an angle larger than

$$\Gamma = 1.22 \frac{\lambda}{D_o}. \tag{7.43}$$

The *resolving power* of the telescope is larger the smaller this angle is, and consequently is defined as equal to its reciprocal

$$\frac{1}{\Gamma} = \frac{D_o}{1.22\lambda}. \tag{7.44}$$

We see that the resolving power is proportional to the diameter of the objective. The reason for this is that the objective transmits only a section of the incident wave, which is a circle of diameter D_o, producing diffraction. Clearly, in order not to degrade the resolving power, the entire wavefront transmitted by the objective must reach the retina of the eye (or the sensitive surface) without further limitations. As we already noticed, and as seen in Fig. 7.31, the eyepiece is designed to transmit the entire front. It is also evident that it is useless making the diameter of the eyepiece larger than the pupil of the eye. These considerations fix, for a given diameter of the objective, the minimum magnification of a telescope.

To fix the orders of magnitude, consider that the diameter of the pupil with good illumination during the day is 2–3 mm, and in the dark at night, 6–8 mm. Consider, for example, a telescope for astronomical nocturnal observations with an objective diameter D_o = 100 mm. Its minimum angular magnification is about 15.

In comparison with the naked eye, the resolving power of a telescope is larger in the ratio of the diameters of its objective and the pupil of the eye. As objectives with diameters on the order of meters can be manufactured, it is clear that resolving powers hundreds of times larger than that of the eye can be reached. Consider, for example, a quite small objective of D_o = 100 mm. Its resolving power is $1.22\lambda/D_o$ that, with λ = 0.5 μm, is equal to 6 μrad, which is about one arcsec. The large astronomical telescopes have a mirror in place of the objective, for which the same arguments hold. Mirrors can be produced, in practice, with much larger diameters, up to many meters, than the lenses.

Clearly, the entire above discussion holds under the hypothesis that objective and eyepiece are diffraction-limited, namely optically perfect. All the aberrations should be corrected at this level. This is achieved by designing systems of lenses.

7.15 Microscope

Like the telescope, the compound microscope is composed of an objective and an eyepiece, but unlike the telescope, is designed to magnify very near, small objects.

The far objects, like stars, are seen through the telescope at definite angles. Their sizes and their distances are not directly measurable. As a matter of fact, astronomers always measure the angles between the different heavenly bodies. Contrastingly, in the case of the microscope, the sizes of the objects are to be defined precisely, while the angles under which they are seen may vary substantially when the objects are approached or moved away even slightly. Note, in addition, that in the microscope, the angles between the rays may be large. Consequently, we cannot use the small angle approximation making the angles equal to their sines or tangents, as we often have done.

We take here the approximation of considering both the objective and the eyepiece as being thin lenses. They are not, but, as in the case of the telescope, this simplifying hypothesis is enough for us to appreciate the operational principles. Figure 7.32 represents the system from the geometric optics point of view.

The object of transverse dimensions y is positioned a little beyond the focal plane of the objective, the focal length of which we indicate with f_o, and we can write $p \approx f_o$. The objective forms the image at the distance q. Consequently, its transverse dimensions are $y' = yq/p = yq/f_o$.

The objectives have small diameters, on the order of a millimeter, and short focal distances, producing a sizeable magnification. In this case too, the real image formed by the objective is seen through the eyepiece, which is located beyond the image at a distance equal to its focal length f_e. Under these conditions, it produces a plane wave that can be seen without effort by a relaxed eye. The eyepiece produces a further magnification. Indeed, the image is seen under the angle θ' given by

$$\tan\theta' = \frac{y'}{f_e} = \frac{y}{f_o}\frac{q}{f_e}$$

We now compare this, similarly, but not exactly, to what we did for the magnifying glass, this tangent with the tangent of the angle under which we would have seen the object without the microscope at the shortest distance of distinct view $L_d = 0.25$ m, namely $\tan\theta = y/L_d$. The magnification of the microscope is then

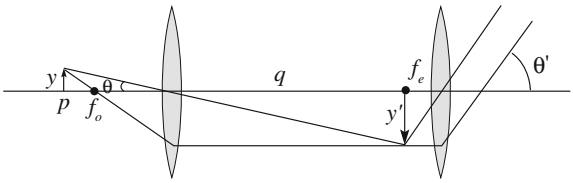


Fig. 7.32 The geometry of the compound microscope

$$J = \frac{\tan\theta'}{\tan\theta} = \frac{L_d}{f_o}\frac{q}{f_e}. \tag{7.45}$$

The magnification is the product of two factors. The first one, L_d/f_o, is due to the objective, the second one, q/f_e, to the eyepiece.

To realize the orders of magnitude, let us consider, for example, an objective of focal length $f_o = 2$ mm and an eyepiece with $f_e = 15$ mm. If we take a typical value $q = 150$ mm, the magnification of the objective is 125 and that of the eyepiece is 10, for a total of 1250.

As should be obvious by now, in order to see very small details, we need to have an adequate resolving power, beyond magnification. Resolving power is, for the microscope, the smallest *distance* Σ between two points at which they are seen as distinct. As we already stated, the angles under which the objects are seen by the microscope are not small in general, and we cannot express Σ as the product of the angle Γ given by Eq. (7.32) times the distance p at which the object is located. It is, however, useful to start from such a small angle approximation, as well as seeing the difference with the correct result. In this rough approximation, the minimum distance between two distinguishable point objects at the distance p is $\Sigma \approx 1.22\lambda p/D$, where D is the diameter of the objective. It is usual to express Σ as a function of the angle α under which the object sees the radius $D/2$ of the objective. This angle, which we are considering to be small, is approximately $\alpha \approx D/(2p)$, and we have

$$\Sigma \approx \frac{1.22\lambda}{2\alpha} = \frac{0.61\lambda}{\alpha}.$$

Let us now find the correct expression, also valid when a is not small. Let us consider two point sources such as A and B shown in Fig. 7.33 and let ask ourselves what is the minimum distance Σ between them so as to have their images resolved. The figure shows the circular waves produced by A and B at the moment they touch the lens. As we discussed at the end of Sect. 7.11, the optical system can distinguish the two waves, provided their maximum distance is $\delta > 0.61\lambda$, as in Eq. (7.35).

We see in Fig. 7.31 that, considering that Σ is small anyhow, the distances from A to O and B to O are practically equal, and consequently, the two waves reach O contemporarily. In that instant, their separation is a maximum near the edge of the lens, namely it is $AC–BC$. This distance is equal to $\Sigma \sin\alpha$. We then have

$$\Sigma = \frac{0.61\lambda}{\sin\alpha},$$

which is now exact, reducing to the previously found expression for small angles, namely for $\sin\alpha \approx \alpha$.

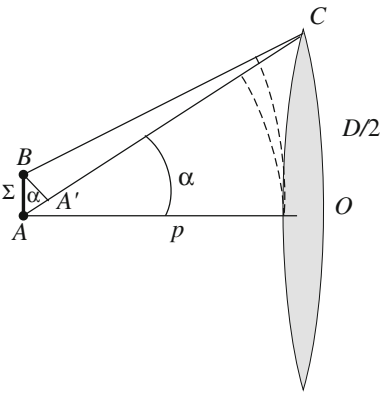


Fig. 7.33 Geometry for two point sources near one another and the objective lens

A "trick" to increase the resolving power of the microscope is to interpose a drop of oil between the objective and the object, with refractive index of, say, n. Under these circumstances, we must substitute the wavelength in the oil in the formula we just found, namely λ/n, in place of λ, which would be in a vacuum. Finally, the resolving power of the microscope is

$$\Sigma = \frac{0.61\lambda}{n \sin \alpha}. \tag{7.46}$$

The quantity

$$NA = n \sin \alpha. \tag{7.47}$$

is called the *numerical aperture* of the objective.

Σ is the minimum distance between two object points to be resolvable. Its reciprocal, namely Σ^{-1}, is the *resolving power of the microscope*. Let us look at the orders of magnitude. Since n is never larger than about 1.5 and $\sin\alpha$ is smaller than one, with $\lambda = 0.55$ μm, we can reach, under the best conditions, a resolving power of $\Sigma^{-1} = (1/0.2)\mu\text{m}^{-1}$. We see that we can resolve at better than half a wavelength.

The resolving power can be increased by decreasing the wavelength of the employed radiation. This explains the use of electronic microscopes, which employ electron beams with wavelengths thousands of times smaller than light.

What we have stated is valid only if the aberrations of the objective and the eyepiece are corrected at the diffraction limit. This cannot be done with thin lenses. Both objectives and eyepieces are complex optical systems containing many optical surfaces (up to a dozen for a good objective).

7.16 Photometric Quantities

An important chapter of optics deals with quantities that are related to the characteristics of human vision, called *photometric quantities*. An "average" human eye can sense electromagnetic radiation in a "window of visibility" extending in wavelength from about 380 to 780 nm. Within this interval, the visual perception for a given energy flux entering into the eye depends on the wavelength. We shall define a *sensitivity function* that is substantially the ratio between the visual sensation and the energy collected by the retina. The eye sensitivity function is not an objective physical quantity, but contains physiological elements as well. First of all, it is very different under daylight conditions and at low light levels, such as those during the night. This is because different types of sensor cells are responsible for the two types of vision, called cones and rods, respectively. Under low levels of luminance, vision is provided by the rods, which are more sensitive but colorblind. We do not see colors at night. This is called *scotoptic vision*. Under daylight conditions, *photoptic vision*, as it is called, is provided by the cones (a more precise definition will be provided later on). We have three types of cones, with peak sensitivity in three different regions of the spectrum, providing us with the perception of colors. The maximum sensitivity is at 437 nm for the "blue" cones, at 533 nm for the "green" cones, and at 564 nm for the "red" cones (as they are so called, even if peak sensitivity lies more in the yellow than in the red). We shall only deal with photoptic vision in the following.

In addition, the sensitivity function varies from individual to individual, and also, for any given individual, as a function of fatigue, health status, age, etc. However, a standard mean sensitivity function of the wavelength, $V(\lambda)$, has been determined to be an average based on several measurements with light of different wavelengths on "normal" individuals at rest and with eyes adapted at the daylight level. This function is shown in Fig. 7.34. Its maximum is in the green at $\lambda = 555$ nm corresponding to 540 THz frequency. Notice, for example, that eye sensitivity is two orders of magnitude smaller both in the red and the violet than in the green, meaning that, in red and blue, we need a hundred times or so more incoming energy than in green for the same level of perception. We shall come to the right scale of the diagram soon.

In the following, we shall consider two sets of quantities, which are in one to one correspondence with one another. The corresponding units are related by the photoptic eye sensitivity function. The *physical* properties of light, and, more generally, of electromagnetic radiation, are measured in *radiometric units*, such as, for example, the energy flux (or radiant power) Φ. The corresponding *photometric* quantities, such as, for example, the luminous flux Φ_L, are measured in *photometric units* that take into account the sensitivity function.

Consider a point light source in a transparent, homogeneous, isotropic and indefinitely extended medium. Under these conditions, the wave surfaces are spheres with centers in the source. The energy transported by the waves crosses surfaces whose area grows as the square of the distance from the source r. Consequently,

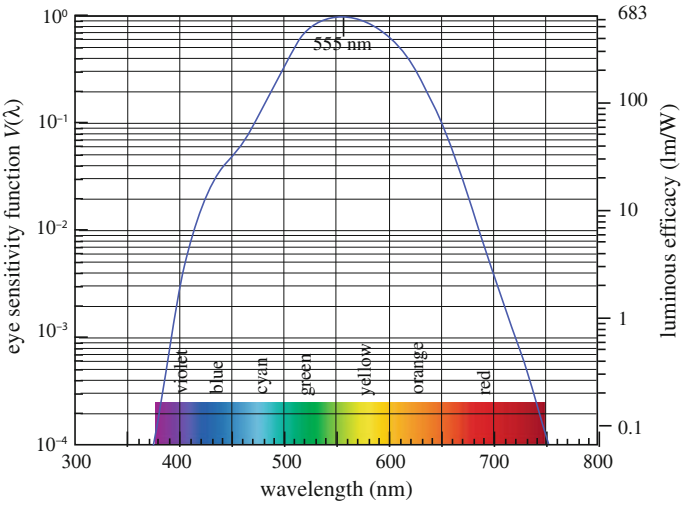


Fig. 7.34 Human eye standard photoptic sensitivity function (left-hand ordinate) and luminous efficacy (right-hand scale) versus wavelength. Note the logarithmic vertical scale

the wave *radiant intensity*, which is the energy flux through a surface divided by its area, decreases as $1/r^2$. Since the energy propagation directions are radial, the energy flux is the same through all the sections of a given solid angle with a vertex in the source. We can then talk of the flux $d\Phi$ sent by the source in the solid angle $d\Omega$. If $d\sigma$ is the normal section of $d\Omega$ at the distance r, the energy intensity of the wave at the distance r is, by definition,

$$I = \frac{d\Phi}{d\sigma} = \frac{d\Phi}{r^2 d\Omega} = \frac{J}{r^2},$$

where we have introduced the proportionality constant

$$J = \frac{d\Phi}{d\Omega}, \tag{7.48}$$

which depends on the source alone and is greater the greater the power emitted by the source. J is called the *radiant emission intensity* of the source (not to be confused with the wave intensity). Its physical dimensions are a power per unit solid angle and its measurements units are W/sr.

The corresponding photometric unit is the *luminous intensity* of the source, namely

$$J_L = \frac{d\Phi_L}{d\Omega}. \tag{7.49}$$

In the general case of non point-like sources, the emission is not generally isotropic, and J and J_L are functions of the emission direction.

The luminous intensity is one of the seven base quantities of the SI. Its unit is the *candela*, with the symbol cd. The other photometric units are derived from the candela. The definition is:

The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540 THz and has a radiant intensity in that direction of 1/683 W per steradian.

Note that 540 THz corresponds to the maximum of $V(\lambda)$, namely to $\lambda = 555$ nm. As a historical curiosity, the "candela" comes from the original, and now obsolete, definition. One candela was defined as the luminous intensity emitted by a plumber's candle of specified dimensions and chemical composition. This candle was similar to those used by plumbers in the XIX century to melt the lead solder when joining water pipes. The value of 1/683 W sr^{-1} in the present definition was chosen to have the new candela be as close as possible to the old one.

A connected photometric quantity is the *luminous flux*. The unitary luminous flux is the flux sent in one steradian solid angle by a source having the light intensity of one candela. The unit is the *lumen*, with the symbol lm. Its definition then follows from the definition of the candela. It can be explicitly written as: *a monochromatic light source emitting a radiant power of (1/683) watt at 555 nm has a luminous flux of one lumen (lm).*

Note that a source of luminous intensity of 1 cd isotropically emitting in the entire solid angle gives a luminous flux of 4π lm = 12.57 lm.

Spectral luminous efficacy is the ratio between any photometric quantity and the corresponding radiometric quantity as a function of wavelength. It is usually defined as the ratio between the photometric flux and radiant power in the same solid angle, namely as

$$\eta = \frac{\Phi_L}{\Phi}. \tag{7.50}$$

Its measurement unit is the lumen per watt (lm/W).

It follows from the definition of the candela that the spectral luminous efficacy for monochromatic radiation of wavelength 555 nm (or frequency of 540 THz) is exactly, namely by definition, 683 lm/W. The right vertical scale of Fig. 7.34 allows one to read the curve of luminous efficacy as a function of the wavelength.

Consider, for example, an incandescent light bulb, consisting of a glass bulb containing, in a vacuum, a wire filament that is brought to high temperature by the passing of an electric current through it. The bulb was invented by Thomas Alva Edison (USA, 1847–1931) in 1879. These bulbs, after becoming commercial at the beginning of the XIX century, have been the dominant electric light for more than a century. The efficacy of an incandescent lamp is typically 16 lm/W, meaning that a radiative flux of 1 W emitted in a given solid angle corresponds to a luminous flux of 16 lm. Let us consider a lamp of luminous intensity of 40 cd, and let us assume, in a rough approximation, that it emits isotropically. The luminous flux in the entire

solid angle is then $4\pi \times 40 \approx 500$ lm, while the absorbed electric power is 500/16 = 31 W. Incandescent bulbs are presently being replaced by light sources of much higher efficacy, like fluorescence bulbs and LEDs, which have typical efficacy levels of 50–100 lm/W.

When we consider an image on a surface like the retina or the film in a photo camera, the page of a book we are reading or the computer screen, an important quantity is the *illuminance* of the surface, which is the luminous flux incident on the unit area of that surface. Let dS be a surface element, not necessarily normal to the light propagation direction, and $d\Phi_L$ be the luminous flux through that element. Illuminance is defined as

$$E_L = \frac{d\Phi_L}{dS}. \tag{7.51}$$

The measurement unit of illuminance is the *lux*, which is one lumen per square meter, lm m^{-2}. Namely, one lux is the illuminance of a surface of one square meter receiving a luminous flux of one lumen. For example, the illuminance under a full moon is about 1 lx, under home lighting, between 30 and 300 lx.

The radiometric quantity corresponding to the illuminance is the *irradiance*, which is the radiant power (or energy flux) per unit area of the surface, namely

$$E = \frac{d\Phi}{dS}. \tag{7.52}$$

The irradiance is measured in watts per square meter (Wm^{-2}).

Consider now a luminous source of non-zero surface. Let dS be an element of its surface. Let $d\Phi_L$ be the luminous flux emitted by dS in the solid angle $d\Omega$ about a direction forming the angle θ with the normal $\mathbf{n}$ to dS, as shown in Fig. 7.35.

It is experimentally found that, for many sources, $d\Phi_L$ is approximately proportional to $d\Omega$ and to the projection of dS on the normal to the considered emission direction, namely to $d\sigma = dS\cos\theta$. We then have

$$d\Phi_L = BdS\cos\theta d\Omega. \tag{7.53}$$

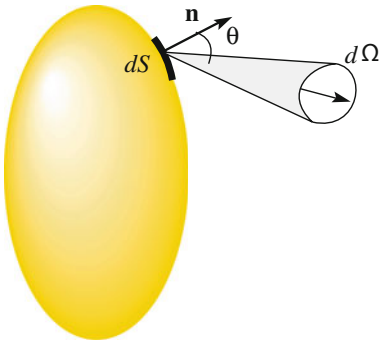


Fig. 7.35 Geometry of light emission by an element of the source surface

If the mentioned relation was of strict proportionality, the quantity we have called B would be constant for the varying emission direction θ. Such a source is said to be a *lambertian source*. In practice, for the majority of sources, B is not strictly constant, but varies slowly with θ. In any case, this quantity, namely

$$B = \frac{d\Phi_L}{d\sigma d\Omega} = \frac{d\Phi_L}{dS \cos\theta d\Omega}. \tag{7.54}$$

is called the *luminance* in the considered direction of the source surface element. The complete definition is as follows.

The luminance is the luminous flux emitted (directly or scattered) in a certain direction per unit solid angle by the surface having unitary area projection on the normal to the considered direction.

The measurement unit of the luminance is the candela per square meter, namely cd m^{-2}, which is called the *nit* (nt), from the Latin *niteo*, meaning 'shining.' For example, the luminance of a computer display typically ranges between 100 and 500 cd m^{-2}. As another example, the surfaces of LEDs that are on the order of a square millimeter have luminance values ranging from the thousands to several millions cd m^{-2}.

We can now say that a photoptic vision regime is defined as the presence of luminance levels larger than 3 cd m^{-2}.

The radiant quantity corresponding to the luminance is the *radiance*.

Question Q 7.3. Consider a 100 W incandescent light bulb giving a luminous flux of 1600 lm. Assume the light to be emitted isotropically. (a) What is the luminous efficiency of the bulb? (b) What is the illuminance on a desk located 2.0 m below the bulb? (c) Is the vision from the desk photoptic? (d) What is the luminous intensity of the bulb?□

An example of a lambertian source is the moon. Indeed, the disc of the moon appears uniformly luminous. Let us consider two surface elements of the same apparent areas, such as those shown in Fig. 7.36. We draw one of the elements near the center of the disk. Its area is equal to the projected area. The element sends the light received from the sun to our eyes and scattered in the direction normal to the moon's surface. The other element in the figure is closer to the border. The area of the emitting surface is larger than the apparent area by the factor l/cos θ. As we see it with the same brilliance, we must conclude that the light emitted (scattered, in this case) at the angle θ is proportional to cos θ.

Several of the most common furnishing lamps are lambertian as well. To achieve that, architects enclose the light bulb in uniformly diffusing glass surfaces.

As we said, the measurement unit for luminance is the nit. To get an idea of its value, consider that the luminance of the moon's surface is 2.5 knit, while that of the clear sky during the day is about 10 knit and that of the sun's surface 1.5 Gnit (do not look at it).

Table 7.1 summarizes the photometric and radiometric quantities and their units.

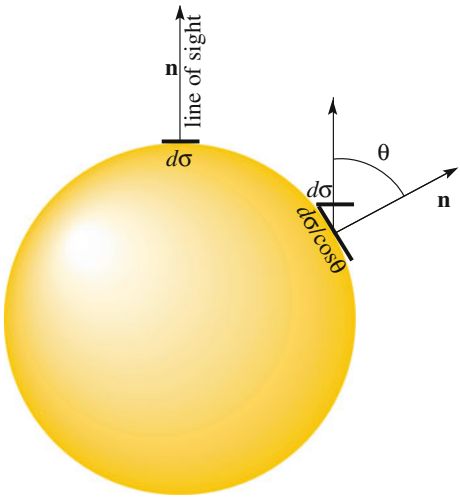


Fig. 7.36 Surface elements of a Lambertian source

Table 7.1 Photometric and radiometric quantities and units

Photometric unit	Dimensions	Radiometric unit	Dimensions
Luminous flux	lm	Radiant flux	W
Luminous intensity	lm sr^{-1} = cd	Radiant intensity	W sr^{-1}
Illuminance	lm m^{-2} = lux	Irradiance	W m^{-2}
Luminance	cd m^{-2} = nit	Radiance	W sr^{-1} m^{-2}

7.17 Properties of Images

The purpose of the eyes and of optical instruments is the formation of an image. The final image falls on a sensitive surface, such as the retina, in the case of the eye, and photographic film or an electronic image sensor, in the respective cases of older and more modern models of photo camera.

We can collectively call all of them *image sensors*. Their purpose is to detect and convey (to our brain or to the electronic memory of the camera) the elements that compose the image. Every sensor must collect a minimum amount of energy to detect an image element. As we already mentioned several times, this implies that the incident energy flux must be integrated over a non-zero area over a non-zero time interval.

As a consequence of the former issue, the image, as recorded by any sensor, is made of discrete *picture elements*, or *pixels*, for short. The consequence is a limited resolving power for the image sensor, namely an upper limit on the detectable spatial frequency. Consider, for example, taking a picture of a grating of alternated black and white strips. Clearly, in the recorded image, the strips will not be resolved if their spatial period is smaller than twice the pixel size.

Examples of elementary sensors are as follows. The sensitive elements of our retina, as already mentioned, are the cones and the rods, of about 2 μm diameter. The image sensors of digital cameras are arrays of solid-state electronic elements (of the CCD and CMOS types) that transform the energy deposited by the incoming light into an electric charge pulse, which is subsequently amplified and processed by microelectronic circuits. The typical size of a pixel is 2–10 μm, corresponding to a resolving power of 500–100 mm^{-1}. The photographic emulsion consists of silver halide crystals dispersed in gelatin, coated onto a substrate of glass or film. The light-sensitive elements are the halide grains. After exposure, the emulsion is treated in a series of chemical processes that transform the grains that have received sufficient luminous energy into metallic silver. Consequently, the grain size determines both the sensitivity and the resolution, one inversely correlated to the other. Emulsions for normal photographic purposes produce pixel sizes ranging from 0.5 to 2 μm, corresponding to resolutions from 2000 to 500 lines per millimeter. As we shall see in the next chapter, this is not sufficient for the recording of holograms. For this purpose, *holographic emulsions* with grain sizes between 10 and 100 nm (spatial frequency up to 100,000–10,000 mm^{-1}) are produced.

Once the area of the pixel is defined, the sensor element still needs to integrate the incoming irradiance (or illuminance, in the case of the eye) over a certain time interval Δt, called the *exposure time*. The radiometric quantity is the *exposure*, which is

$$e = \int_{\Delta t} E(t)dt. \tag{7.55}$$

Being that it is an energy, it is measured in joules (J).

The photometric quantity, which is relevant for the human eye, is the *photometric exposure*

$$e_L = \int_{\Delta t} E_L(t)dt, \tag{7.56}$$

which is measured in lux times second (lx s).

The sensitivity of a sensor is the minimum energy it needs to produce a signal. This is called its *energy threshold*. For example, the energy threshold of the rods of the human eye is, in order of magnitude, 10^{-17} J, corresponding to some 30 photons being absorbed. The thresholds of the electronic sensors (both CCD and CMOS) range around similar values. The threshold of the photographic emulsion depends strongly on the grain size, typically being between two and three orders of magnitude higher.

With a complex instrument, beyond the final image, which we have discussed, other intermediate images, real or virtual, exist inside the instrument itself. Think, for example, of the telescope or the compound microscope. These intermediate images produced by the first optical elements (the objective in the above examples)

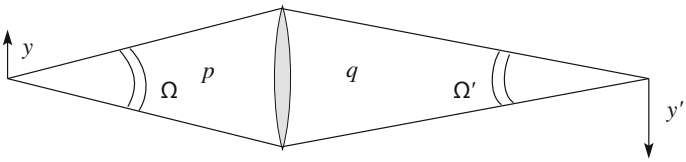


Fig. 7.37 The source, the lens and the image

are also the sources of light waves for the subsequent elements of the instrument (the eyepiece in the examples). For these intermediate sources, the most important quantity is the *luminance* for the eye and the *radiance* for the instruments.

We shall now discuss the illuminance and luminance of the images of light sources (which may also be illuminated objects, obviously). The situation is completely different for extended sources endpoint sources.

Let us consider an extended source in front of a lens. In Fig. 7.37, y represents the diameter of the source, which is at the distance p from the lens. Let q be the distance of the image beyond the lens and y' its diameter. As we know, $y'/y = q/p$. Consider now the areas, normal to the axis, of corresponding elements of source and image. Let us call them σ and σ', respectively. Clearly, their ratio is

$$\frac{\sigma'}{\sigma} = \frac{q^2}{p^2}. \tag{7.57}$$

Calling Ω and Ω' the solid angles under which the lens is seen from the source and from the image, respectively, we have that

$$\Omega p^2 = \Omega' q^2 \tag{7.58}$$

and hence, for Eq. (7.57), we have

$$\Omega\sigma = \Omega'\sigma' \tag{7.59}$$

Calling Φ_L the luminous flux sent by the source to the lens, namely in the solid angle Ω, the luminance of the source is

$$B = \frac{\Phi_L}{\Omega\sigma}. \tag{7.60}$$

Similarly, the luminance of the image is

$$B' = \frac{\Phi'_L}{\Omega'\sigma'}. \tag{7.61}$$

where Φ'_L is the luminous flux *transmitted* by the lens. If the lens were to transmit the entire incident flux, Φ'_L would be equal to Φ_L. In general, it is somewhat smaller

due mainly to reflections at the surfaces and possibly absorption. We can say that $\Phi'_L = \tau\Phi_L$, where $\tau \leq 1$ is the lens transmission coefficient. For Eq. (7.58), we have

$$B' = \tau B. \tag{7.62}$$

This important equation states that *the luminance (sometimes called the brightness) of the image can never be larger than that of the source*.

It is therefore pointless to try to produce, starting from a given source, images of larger luminance. Indeed, it is possible to have an image of luminous intensity larger than that of the source, but only at the price of increasing the apparent emitting area or solid angle.

The illuminance of the image can be immediately obtained multiplying the luminance by the solid angle of the incoming luminous flux, namely Ω' in our case. We then have that

$$E'_L = B'\Omega' = \tau B\Omega'. \tag{7.63}$$

We see that the illuminance of the image is that much greater the greater the luminance of the source and the solid angle under which the lens is seen from the image.

Consider now a point source. In the above discussion of extended sources, we have neglected the diffraction effects. Namely, we have assumed a rectilinear propagation of energy. This is permissible for surface elements that are not too small. It is no longer so for point sources, such as, for example, stars.

Figure 7.38 shows a point source and its image beyond the lens. The latter is not a point, but a fundamental diffraction pattern. With a very good approximation, we can consider it to consist of the Airy disk, in which the largest part of the energy is located. For the sake of simplicity, we shall consider the Airy disk as being uniformly illuminated. If q is the distance of the image beyond the lens and D is the lens diameter, the radius of the Airy disk is $\rho = 1.22\lambda q/D$. The area of the disk is then

$$\sigma' = 2.5\pi\rho^2 = 2.5\pi\frac{\lambda^2 q^2}{D^2}.$$

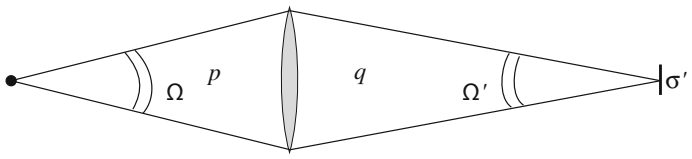


Fig. 7.38 A point source and its image, with the Airy disk

Note that we can speak of the luminance of the image, because it is extended, due to diffraction, but cannot speak of the luminance of the source, being that it is point-like. The luminance of the image is

$$B' = \frac{\Phi'_L}{\Omega' \sigma'}.$$

Now, $\Phi'_L = \tau \Phi_L$ and we can express Φ_L as the illuminance E_{L0} of the lens due to the source times the area S of the lens, namely as $\Phi_L = SE_{L0}$. But it is $S = \Omega' q^2$, and hence, we have

$$B' = \frac{\Phi'_L}{\Omega' \sigma'} = \frac{\tau E_{L0} S}{\Omega' \sigma'} = \frac{\tau E_{L0} q^2}{\sigma'} = \frac{\tau E_{L0} D^2}{2.5 \pi \lambda^2}$$

In conclusion, the luminance of the image is

$$B' = \frac{\tau E_{L0} D^2}{2.5 \pi \lambda^2}. \tag{7.64}$$

We see that the luminance of a point source is proportional to the square of the lens diameter, namely to its area, or also to the amount of light that the lens is able to collect for a given illuminance. This is completely different from the case of the extended source. In addition, we see that the source affects the image through the illuminance E_{L0} it produces on the lens. Finally, in Eq. (7.64), the wavelength is present to remind us that the phenomenon is ruled by diffraction.

Once more, we obtain the illuminance of the image by multiplying its luminance by the solid angle of the incoming flux. We have

$$E'_L = \frac{\tau E_{L0} D^2 \Omega'}{2.5 \pi \lambda^2}. \tag{7.65}$$

We understand how the telescope allows us to see objects too faint to be visible with the naked eye. The illuminance of the retina at the point of the image of a star produced by the optical system consisting of telescope plus eye is that much greater the greater the useful diameter of the system. Take note of the fact that Ω' is practically fixed for the eye. Contrastingly, if we use the telescope to look at terrestrial objects, which are always extended, their images appear enlarged but not brighter (with higher luminance) than the objects themselves.

Summary

In this chapter, we studied the physical processes leading to the formation of images under usual incoherent illumination conditions. The principal concepts we have learned are as follows.

1. The concept of the image of a point source, which is the center of the waves reaching the eye after having been redirected or deformed.

2. The angular dispersion of a prism.
3. The focal length and the power of a lens.
4. The image formation of extended objects. Under the usual conditions, the objects can be thought of as being composed of point sources independent of one another.
5. The quarter wave criterion.
6. The aberrations and irregularities, and how to correct them for the purpose of optical "perfection".
7. The depth of field.
8. The resolving power.
9. The action of a lens on the phase of the light wave and the roots of the image formation process.
10. The simplest optical instruments.
11. The basic photometric quantities and their relations with radiometric quantities.
12. The luminance of the images of the extended endpoint-like sources.

Problems

7.1. A tree has needle-like leaves at a distance of 1 cm from one another. What is the maximum distance at which they are resolved by the naked eye?
7.2. Is the deviation angle of a glass prism larger for blue or red light?
7.3. Two lenses of power of +8 diopters and −4 diopters are in contact, one next to the other. What is the focal length of the system?
7.4. From which of the following quantities does the resolving power of a telescope depend? Light wavelength, distance of the object, power of the objective, diameter of the objective, diameter of the eyepiece.
7.5. The eyepiece of a telescope contributes to the magnification. Can it help in increasing the resolving power? Why?
7.6. Consider a converging thin lens of dioptric power of +4 diopters. Draw the image of an arrow with the tail on the axis before the lens and normal to the axis and in the following positions, $p = 50$ cm, 25 and 10 cm. Repeat the construction with a lens of −4 diopters.
7.7. Does the parabolic mirror suffer from the spherical aberration?
7.8. Find the dioptric powers of the lenses of refractive index $n = 1.5$ with the following surface radiuses. (a) $R_1 = 15$ cm, $R_2 = 25$ cm; (b) $R_1 = 15$ cm, $R_2 = \infty$; (c) $R_1 = 15$ cm, $R_2 = -25$ cm.
7.9. Two biconcave lenses have the same shape, but different refractive indices, equal to 1.5 and 1.7, respectively. What is the ratio of their focal lengths?
7.10. Does the field depth depend each of the following quantities? The wavelength, the diameter of the lens, its focal length.
7.11. With the eye adjusted to infinity, one observes an object through a magnifying glass of focal length $f = 5$ cm located directly after the eye. What is the angular magnification?
7.12. The point source S is located on the axis at the distance $p = 1$ m from the lens. The image S' is at the distance $q = 1$ m from the lens as well, as shown

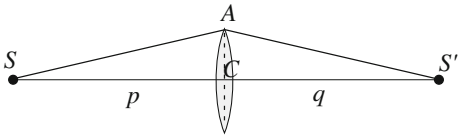


Fig. 7.39 A point on the axis source and its image

in Fig. 7.39. The radius of the lens is $CA = 0.1$ m. The refractive index of the glass is $n = 1.5$. What is the thickness of the lens in C?

7.13. A thin lens forms the images of two point sources, one being red and the other blue. What is, in a round figure, the ratio between the luminance of the images?

7.14. The closest star is 4×10^{16} m away. Suppose we want to build a reflecting telescope capable of seeing whether the star has a planet orbiting at the same distance as the earth from the sun (about 150 Gm). What should the minimum diameter of the mirror be?

7.15. A luminous source has a surface A. A lens collects the emitted luminous flux in a cone of vertex angle of 10° and forms a real image whose area is 1/10 of the source area. What is the opening angle of the light emerging from the image?

7.16. A point source emits electromagnetic radiation of wavelength between 0.8 and 1 μm with radiant intensity of 600 W/sr. What is the luminous intensity?

Chapter 8
Images and Diffraction

Abstract In this chapter, we discuss two subjects in which diffraction is the dominant process in image formation. The first argument, developed in the first two sections, is the Abbe theory of image formation. The theory is valid in general, but is easier to understand under conditions of coherence. The second argument, developed in the subsequent sections, deals with those actual three-dimensional images known as holograms. In a hologram, both the amplitude and the phase of the wave produced by the object are recorded, using coherent illumination.

In this chapter, we discuss two subjects in which diffraction is the dominant process in image formation.

The first argument, developed in the first two sections, is the Abbe theory of image formation. The theory is valid in general, but is easier to understand under conditions of coherence. We shall start from the conclusions we reached in Sect. 5.10, where we showed that the Fourier diffraction pattern of a diaphragm is proportional to the square of the spatial Fourier transform of its amplitude transmission coefficient. We shall show how the image formation process can be thought of as the succession of a Fourier transform, in going from the lens to its focal plane, and of an anti-transform, in going from the focal plane to the image plane. This process has been summarized in the single sentence (quoted by Zernike): "the microscope image is the interference effect of a diffraction phenomenon".

The second argument deals with those actual three-dimensional images known as holograms. A hologram is, in its true substance, a diffraction grating, on which both the amplitude and the phase of the wave originated by the object are encoded. In order to understand the processes of recording the hologram of an object, and subsequently of reconstructing its image, we shall proceed step by step. We shall start by dealing with the straight strips sine transparency grating that diffracts in the first order only. We shall see how such a grating can be recorded in a photographic process by having two monochromatic plane light waves incident at an angle with one another interfere on a high-resolution photographic plate. We shall then see that

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0_8

the grating obtained by developing the plate can be used to reproduce one of the light waves that we used to produce it.

We shall then study a circular grating, namely the Soret zone plate, which, working with diffraction, has characteristics similar to a lens, but with many focal lengths. We shall subsequently discuss the Gabor grating, which is similar, but with sinusoidally varying transparency. It works like a lens, with only one focal length, but one that is both positive and negative. It can be recorded with a photographic process similar to the one previously considered, having now the interference on the photographic plate of a plane wave and a circular one scattered by a point-like object. We shall see that, again, we can reconstruct the image of the point object, illuminating the grating with a plane monochromatic wave.

Finally, we shall describe how to record holograms of three-dimensional objects, encoding both the amplitude and the phase of the object wave, and how to play them back to reconstruct their three-dimensional images.

8.1 Abbe Theory of Image Formation

The purpose of optical instruments is to form optical images of objects. We shall now examine the different stages of the image formation process in greater depth. We shall consider the image formation process by a thin lens. This is the simplest case, but is sufficient to highlight the fundamental aspects we want to discuss. We shall assume here the light incident on the lens to be *coherent*. This is the case if the light beam is produced by a laser and, under particular conditions, in a microscope. As already mentioned, these conditions occur when the diaphragm in the light condenser is narrow and intensely illuminated. We shall now discuss the simplest elements of the Abbe theory of image formation that was developed in 1866 by Ernst Karl Abbe (Germany, 1840–1905), just to improve the understanding and quality of the images in a microscope. The Abbe theory can be generalized to illumination under non-coherent light, but we shall not extend our discussion to that.

Let us consider a two-dimensional object, such as a common photographic slide or a grating or a biological specimen on a microscope slide, and let us illuminate it with a normally incident monochromatic plane wave. The light that passes through the object is focused by the objective of the microscope, which we consider to be a thin lens, forming the image.

In the language of Sect. 5.10, our object is a diaphragm. If it changes by a position-dependent factor the amplitude of the incoming wave is an amplitude diaphragm, if it changes the phase by a position-dependent phase-shift, it is a phase diaphragm, and if it has the two effects simultaneously, it is an amplitude and phase diaphragm.

To be specific, let the object be a grating with alternate absorbing and transparent straight strips, shown in Fig. 8.1. As we already learned in Sect. 7.12, the amplitude distribution on the back focal plane of the lens is proportional to the spatial Fourier transform of the amplitude transmission coefficient of the object. In our case, the

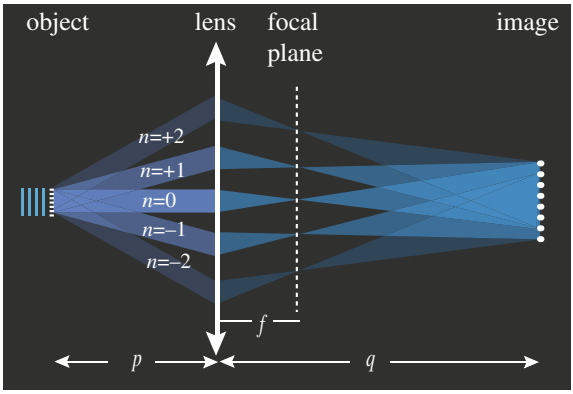


Fig. 8.1 Image formation a parallel lines grating

transmission coefficient is (almost) periodic and its Fourier transform consists of the discrete sequence of the principal diffraction maxima. Their total number is determined by the diameter of the lens. We show five of them in the example in Fig. 8.1. Their spatial frequency is an increasing function of the distance from the axis.

This is the first part of the Abbe theory. The second part tells us how the image is formed starting from the Fourier transform.

Using the Huygens-Fresnel principle, we can think of the luminous wave beyond the focal plane as originating from fictitious sources at the interference maxima. In other words, the image plane receives light from the sources on the luminous regions on the focal plane that are coherent with one another. The result of the interference from these sources is the image of the object exactly at the distance q beyond the lens, which is such that

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f}. \tag{8.1}$$

This can be rigorously demonstrated, but we will be satisfied with the result. We see that the second process can be thought of as a spatial Fourier anti-transform.

We can draw this conclusion, which is valid in general. The image formation process comes about in two stages. In the first stage, the lens produces the spatial Fourier transform of the amplitude transmission coefficient of the object in its back focal plane; in the second stage, the propagation of the wave from the focal to the image plane performs the anti-transform, giving back an image that is similar to the object. Abbe's theory has been summarized in a single sentence (quoted by Zernike): "*microscope image is the interference effect of a diffraction phenomenon*."

Consider now some simple experimental consequences of the Abbe theory. We first observe how the theory explains that a certain degradation of the image relative to the object cannot be avoided. It is indeed clear that, due to the finite diameter of the lenses, not all the Fourier components can get through an optical instrument. The higher frequency components are too far from the optical axis and are lost in

the subsequent image formation. As a consequence, for example, in the image of the grating considered above, the edges between the transparent and opaque strips will not be as sharp as in the object, due to the absence of the higher frequency components. We can check this with the following simple experiment.

Using an optical bench, we put a circular iris diaphragm of variable diameter on the focal plane of the lens and a white screen in the image plane to observe the image. We start with the iris completely open and observe that the images of the lines are quite sharp. When we close the iris gradually, we observe that every time a diffraction order (namely a Fourier component) is excluded, the contrast of the image suddenly decreases. When only the central and the two first order maxima are let through, the image is a sine grating, namely the light illuminance is the square of a sine function plus a constant. If we go further and exclude the first order maxima too, letting only the central one through, the image disappears completely; our screen is evenly illuminated. Indeed, the central maximum does not carry any information at all on the image.

Let us now consider, as an object, a luminous point, for example, a pinhole transmitting the light emanating from a source in front of it, as in a microscope. Its Fourier transform extends up to very high frequencies, up to infinite frequencies in the ideal case of a geometric point source.

We can consider the Fourier transform of the pinhole to be a constant on the focal plane. The finite diameter of the lens allows the passage of spatial frequencies up to a maximum value that we call k_c and block the higher frequencies. The subsequent process leading to the image will be the Fourier antitransform of a function of k that is constant for $k \leq k_c$ and zero for $k > k_c$. Remember now that the antitransform operation is equal to the transform, an irrelevant sign apart here. Clearly then, the image is the diffraction pattern of the lens' aperture. We understand better that every optical instrument has a limited resolving power, as we found in Sect. 7.11.

Note that what we have said so far, and what we shall say in the following, is valid only if the lens is diffraction-limited, not in the presence of larger aberrations.

Having learned where the Fourier transform is physically located, we can think of modifying the image by acting on the transform, by depressing or boosting some of the frequencies. This operation is called *spatial filtering*.

For example, if we want to enhance the contrast of the small details of the object in the image, we can block or attenuate the lowest spatial frequencies by inserting a transparent film into the focal plane with a totally or partially absorbing small disc in the center. If you do that, you will see luminous images of the small objects of enhanced contrast on a dark background. Indeed, you have eliminated not only the lowest frequencies, but also the zero frequency Fourier component, which is responsible for the background illuminance. As a matter of fact, dark field techniques have been developed by microscopists, in particular, for the observation of live and unstained biological specimens.

Conversely, if we want to make a photographic image smoother, namely to reduce the contrast in comparison to the object, we can introduce an iris diaphragm on the focal plane, of the proper diameter.

We can also consider more complex filters. We could, for example, remove unwanted spatial frequencies by placing a transparent slide in the focal plane with an opaque annulus whose inner and outer radii define the frequency band that we want to delete. One can use this method, for example, to eliminate the dithering present in some images, such as those which appear in newspapers.

On the other hand, spatial filtering, which is always unwillingly present to some degree, may alter the final image to such an extent as to bring in details that were not present in the original. In this case, we speak of false details. As an example, let us go back to the object consisting of a Fraunhofer grating. Suppose we block all of the odd diffraction orders (or all the even ones but the zero-order) with suitable obstacles in the focal plane. The resulting image is a grating with twice as many lines as the original! The distance in the focal plane between two successive interference maxima is, in fact, as we know, inversely proportional to the pitch of the grating.

The examples we have given should be sufficient to convince the reader of the many possibilities offered by the spatial filtering technique. In the next section, we shall discuss an important example of application in the filtering of images in a microscope, namely the phase contrast microscope.

8.2 Phase Contrast Microscope

As a relevant example of spatial filtering, we shall now discuss the working principle of the phase contrast microscope, invented by Frits Zernike (Netherlands 1888–1966) in 1934.

The observation of microorganisms and other biological specimens under a microscope almost always presents the problem that the object under observation is transparent, often as much as the water in which it is immersed. Under these conditions, the contrast is not sufficient to see the object. The most commonly used procedure for making it visible is staining, meaning covering it with a dye that is absorbed by its biological tissue. This procedure has several drawbacks, in particular, that it cannot be used to study living microorganisms, because the dye kills them. The phase contrast technique is a brilliant solution to this problem.

First of all, we must operate the microscope under the coherent illumination conditions of the specimen. As we already mentioned, we achieve that with a strong illumination of the iris diaphragm in the condenser, closing it to a small diameter. Under these conditions, the iris is a point source, S, located in the forward focus of the converging lens of the condenser (see Fig. 8.4). As a consequence, the light wave incident on the glass-plate supporting the specimen is a plane wave. The phase of each monochromatic component of the light, which is white, does not depend on its position on the plate.

Zernike started from the consideration that the specimens generally have refractive indices that are different, even if not by much, from the index of the water in which they are immersed. We can thus consider the glass-plate with the specimen

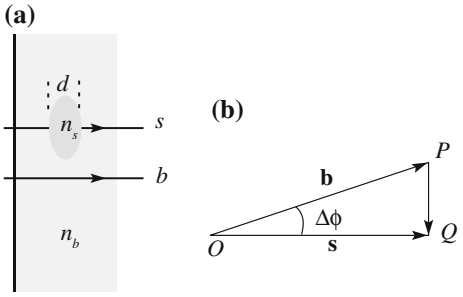


Fig. 8.2 **a** Cross-section of the microscope glass-plate showing a film of water and the specimen; **b** wave field vectors at the exit for the two paths shown in **a**

to be a phase diaphragm that does not act much on the amplitude, but can act substantially on its phase. Indeed, the *phase* of the transmitted wave at the points where it crosses the microorganism, namely our *specimen*, is appreciably different from that where it went through water alone, say from the *background*.

Figure 8.2a shows a cross-section of the glass-plate supporting a microorganism, represented by a small ellipse in a water film. Let n_s and n_b be the refractive indices of the specimen and of the water, respectively, and d the thickness of the specimen. Consider two paths, s crossing the microorganism and b in the water. The phase difference between them at the exit is

$$\Delta\phi = (n_s - n_b)k, \tag{8.2}$$

where k is the light wave number in a vacuum.

Now, the eye, and any other recording instrument, is sensitive to the amplitude, but not to the phase. However, if we were able to transform the phase modification in an amplitude modification, we could make the object visible. Figure 8.2b shows the starting point, namely two rotating vectors representing the wave field after the glass-plate; **b** at the points of water, which we call the *background*, and **s** at the points of the microorganism, which we call the *signal*. The effect of the microorganism is to introduce a phase difference, say $\Delta\phi$, between the phase of the output wave at and outside the microorganism. The amplitude remains substantially unaltered. Then the effect of the microorganism is to add the vector PQ to the background vector **b**, as shown in Fig. 8.2b. Being that $\Delta\phi$ is always small, the angle between PQ and **b** is about 90°. The magnitude of the signal vector **s** is not very different from that of **b**. However, if we advance or delay the phase of the background relative to the signal by 90°, we obtain one of the situations represented in Fig. 8.3. The signal **s** = **b** + PQ now has an amplitude sensibly different from that of the background, being brighter or darker in the two cases.

But how can we change the phase of the background relative to the signal? Here, the Abbe theory helps us, allowing us to work in the spatial frequency domain rather than in the space domain. Indeed, the Fourier transform of the background contains mainly low spatial frequency components. Contrastingly, the Fourier transform of the signal, being that of a small object, extends to much higher frequencies. Consequently, in the back focal plane of the objective, the signal and

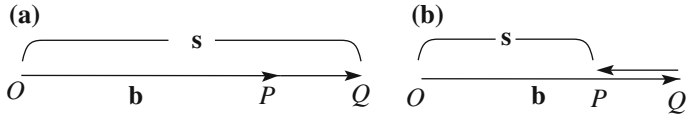


Fig. 8.3 Wave field after having advanced in **a**, retarded in **b** the phase of PQ by 90° relative to that of ***b***

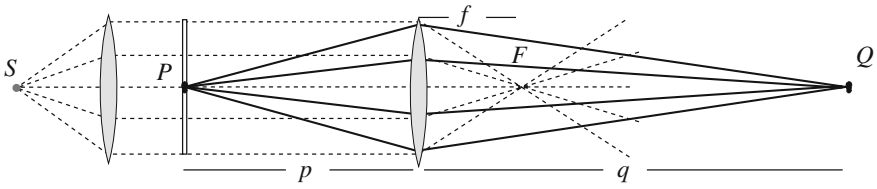


Fig. 8.4 Geometric optics of the background and object rays

the background are separated. The light of the background is near to the axis, while the light of the signal extends farther out.

Figure 8.4 shows the situation from the point of view of geometric optics. On the left, we have the light condenser. The point S is the strongly illuminated iris in the focus of the lens producing the parallel illuminating beam, as we already mentioned. The point P represents the microorganism on the microscope slide. Its distance from the objective (which we approximate with a thin lens) is p. The rays crossing the slide outside P, namely the background, join in the back focus of the objective. The rays crossing P converge after the lens to the image Q at the distance q given by the thin lens equation. As a consequence, they cross the focal plane spread over a wide zone.

We can produce a phase shift of π with a transparent plate, whose thickness d and refractive index n are chosen to produce such a phase delay, namely with $knd = \pi$. We insert this *phase-strip*, as Zernike called it, in the back focus of the objective.

Let I_s and I_b be the signal and background intensities, respectively. We define as the intensity contrast (between signal and background) the quantity

$$\gamma = \frac{|I_s - I_b|}{I_s + I_b}. \tag{8.3}$$

Note the absolute value in the numerator. Indeed, the intensity of the signal may be larger (clear field) or smaller (dark field) than that of the background. The intensities I_s and I_b are proportional to the squares of the magnitudes of the vectors **s** and **b**. Now, as seen in Fig. 8.2, it is $PQ = b\Delta\phi$. Thus, in the case of Fig. 8.3a, for example, we have $\mathbf{s}^2 = |\mathbf{b} + PQ|^2 = b^2(1 + \Delta\phi)^2 \cong b^2(1 + 2\Delta\phi)$, and thus

$$\gamma = 2\Delta\phi, \tag{8.4}$$

and similarly for Fig. 8.3b. Now, the human eye can distinguish an object on a smooth background down to a contrast of 0.02, but for a comfortable observation, one needs about $\gamma = 0.1$. This corresponds to a phase difference of 50 mrad. Hence, taking $\lambda = 0.5$ μm, we have the condition $(n_s - n_b)d > 4$ nm. If the index difference is, for example, $n_s - n_b = 0.1$, which is not an unusual condition, the minimum observable thickness is 40 nm. This is a completely adequate value, taking into account that the minimum observable size, at the limit of the resolving power, is about one third of a wavelength, namely about 200 nm.

8.3 Sine Grating

In this section, we begin the study, which will continue until the end of the chapter, of gratings that, if illuminated with light of appropriate characteristics, produce images. We shall obtain such gratings through a photographic process, starting with the objects of which we shall subsequently produce images. The images produced in this way have encoded their complete information on the object wave, both on its amplitude and on its phase. They are holograms.

Let us start by recalling a few concepts. In Sect. 7.12, we saw that a convergent lens placed beyond a two-dimensional object, which we called a diaphragm, illuminated by a plane normally incident wave, produces the spatial Fourier transform of the amplitude transmission coefficient of that diaphragm in its back focal plane. From this point of view, if, for example, the diaphragm is a Fraunhofer grating, the principal maxima of its diffraction pattern are the discrete components of the spectrum of the transparency of the grating. To be precise, the Fourier spectrum would be discrete, namely the maxima would have zero width, if the grating were infinitely extended. In practice, the maxima have a width that is inversely proportional to the number of lines, namely to the extension of the grating. Pay attention to the fact that there are two maxima for each order, one for positive spatial frequency and one for negative, one on the right of the central maximum and one on the left.

Let us now look for a grating that produces only first order maxima (order +1 and −1). This is the Fourier transform of a grating, whose amplitude transparency varies as a sine function. We already considered such a grating in Sect. 2.7 and already noticed that, in practice, we must add a constant to the transparency function to avoid it having non-physical negative values. The most general expression of the amplitude transparency of such a sine grating, or diaphragm, is the following

$$T(x) = \frac{1}{1+b}(1 - b\cos hx). \tag{8.5}$$

Here, it is $h = 2\pi/a$, where a is the spatial period of the grating in x and, obviously, $1/a$ is the number of periods per unit length x. Hence, h is the spatial

frequency of the grating. The parameter b, which is, in any case, between 0 and 1, represents the *modulation* of the diaphragm. It measures the relative importance of the sinusoidal over the continuum component. Finally, the normalization coefficient $1/(1 + b)$ has been chosen in order to have $0 \leq T(x) \leq 1$ in any case, as it should.

Let us now have a plane monochromatic wave normally incident on the diaphragm positioned in front of a lens, in order to be under the Fraunhofer conditions. Equation (5.38) then tells us that the amplitude of the diffracted wave is proportional to the Fourier transform of $T(x)$, namely to

$$G(k_x) = \frac{1}{1+b} \int_{-\infty}^{+\infty} \mathrm{e}^{-ik_x x}(1 - b\cos hx)dx. \tag{8.6}$$

Without any calculation, we observe that we are considering the transform of a periodic function (if the grating is ideal and infinitely extended). The transform is then discrete. The function is the sum of a constant term, whose transform falls at $k_x = 0$, and a term proportional to $\cos hx$, whose transform has two equal values at $k_x = +h$ and $k_x = -h$. Clearly, the three terms correspond in the diffraction pattern to the maxima of order 0, +1 and −1, respectively. The two first order maxima have the same height. Clearly, the ratio between their height and the height of the 0 order maximum is smaller the smaller the modulation b is.

In practice, the real gratings are not infinitely extended. The Fourier transform is no longer discrete. The maxima are in the same positions as for the infinite grating, but they have a non-zero width, which is larger for smaller grid extensions.

Let us now see how to produce physically a grating with the amplitude transmission coefficient of Eq. (8.5) (beyond what we already found in Sect. 8.1). Pay attention to the fact that Eq. (8.5) is for the *amplitude*. Namely, it is the amplitude, not the intensity, which we want to be a sine function.

We shall manufacture the grating through a photographic method, producing an image on a photographic plate capable of recording it. The plate is shown as a dark vertical rectangle in Fig. 8.5 on the right. To produce the image, we need a coherent monochromatic light wave of sufficiently wide front. As shown in Fig. 8.5, we start from a laser beam, which is monochromatic and coherent, but is not wide enough. Indeed, the typical diameter of a laser beam is on the order of a millimeter. A beam expander is a system of two converging lenses, one next to the other, in the geometry of the Keplerian telescope, namely with the back focus of the first lens

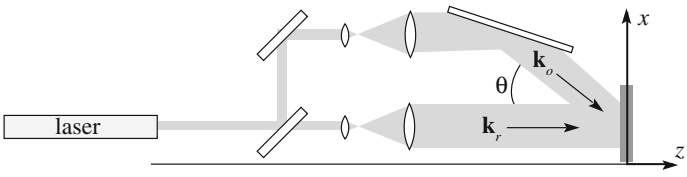


Fig. 8.5 How to produce sine grating

coinciding with the forward focus of the second, as shown in the figure. The beam expander transforms an incident parallel beam in an outgoing parallel beam of diameter larger by a factor equal to the ratio between the focal lengths. We shall often use beam expanders in our subsequent work.

As shown in the figure, we divide the still narrow laser beam into two parts, proceeding in different directions, using a beam splitter, and then expand both of them with two beam expanders. One of the beams, which we shall call the *reference beam*, is incident on the photographic film directly in the normal direction. The second beam, which we call the *object beam*, is deflected by a mirror so as to have it incident on the photographic plate at an angle, say, θ, with the reference beam. The reason for these names will soon be clear. We choose a reference frame with the z-axis in the direction of the reference beam and x and y on the photographic plate.

The field of the reference beam is $E_r e^{i(\omega t - kz)}$ and that of the object wave is $E_0 e^{i(\omega t - kz\cos\theta - kx\sin\theta)}$. The field on the photographic film is their sum at $z = 0$, namely

$$E = E_0 e^{i(\omega t - kx\sin\theta)} + E_r e^{i\omega t}$$

The illuminance is obtained by taking the average over time of the square of the real part. The real part at $z = 0$ squared is

$$\begin{aligned}(\mathrm{Re}\, E)^2 &= [E_0\cos(\omega t - kx\sin\theta) + E_r\cos\omega t]^2 \\ &= E_0^2\cos^2(\omega t - kx\sin\theta) + E_r^2\cos^2(\omega t) \\ &\quad + E_0E_r\cos(2\omega t - kx\sin\theta) + E_0E_r\cos(-kx\sin\theta)\end{aligned}$$

Now, taking the average over a period, we obtain

$$I \propto \frac{E_r^2}{2} + \frac{E_0^2}{2} + E_rE_0\cos(kx\sin\theta). \tag{8.7}$$

In practice, the intensity of the reference wave is always much larger than that of the object wave, namely having $E_r^2 \gg E_o^2$. Under these conditions, Eq. (8.7) becomes approximately

$$I \propto \frac{E_r^2}{2}\left[1 + 2\frac{E_0}{E_r}\cos(kx\sin\theta)\right]. \tag{8.8}$$

We now expose the film to this illuminance for a time interval Δt long enough to have the needed exposure and subsequently develop and fix the film with the required chemical processes. The resulting developed film will be darker at the points that have received more light. Without entering into the details of the developing process, we simply state that it can be controlled to produce an *amplitude* transparency inversely proportional to the exposure, namely to I in Eq. (8.8), in a good approximation. The amplitude transparency is then

$$T \propto \frac{1}{1 + 2\frac{E_0}{E_r}\cos(kx \sin\,\theta)},$$

Finally, as we have $E_0/E_r \ll 1$, we can write approximately

$$T \propto 1 - 2\frac{E_0}{E_r}\cos(kx \sin\,\theta). \tag{8.9}$$

So, we have obtained the sine grating, as we wanted. It has a (small) modulation equal to

$$b = 2\frac{E_0}{E_r} \tag{8.10}$$

and a period $a = 2\pi/(k \sin\,\theta)$, which is more usefully expressed in terms of the wavelength of the light as

$$a = \frac{\lambda}{\sin\,\theta}. \tag{8.11}$$

We see that the period of our sine grating is small, on the order of the wavelength. It is somewhat larger if we work with a small angle between the beams. For example, if $\theta = 15°$, that is, $\sin\,\theta \approx 0.26$, with $\lambda = 0.5$ μm, we have $a = 1.9$ μm or, as we say, 520 periods per mm. Clearly, one must use a very high-resolution holographic emulsion (see Sect. 7.17).

We conclude the section with an important observation, which will be useful in the subsequent sections. Let us illuminate our grating with a plane monochromatic wave equal to the reference wave during the recording process. Well, under these conditions, and according to that which we stated at the beginning of the section, beyond the grating, we have three approximately plane waves, those of the diffraction orders 0, +1 and −1. The zero order wave can be considered as the incident wave of reduced amplitude. The wave of order +1 propagates in the exact same direction θ under which, in the recording process, the object wave was incoming. Namely, we have *reconstructed* the object wave. As matter of fact, our grating is the simplest possible hologram. It is the hologram of a point at infinite distance. In addition, the wave of order −1 propagates in the symmetrical direction $-\theta$.

8.4 Fresnel Zones

In this textbook, we have discussed diffraction phenomena under the Fraunhofer conditions, which we defined in Sect. 5.6, in particular, with Eq. (5.14). When these conditions are not satisfied, namely in near field, we speak of Fresnel conditions.

The description of the Fresnel diffraction phenomena is, in general, mathematically more complex than in the Fraunhofer case. We shall consider here only one effect, namely how images can be formed without using lenses, mirrors or GRINs. To make the discussion simpler, we shall consider a monochromatic light, which we can obtain using a laser. As we shall see, the images obtained in this way contain the complete information both on the amplitude and on the phase of the object wave. Contrastingly, the information in the images obtained with devices operating with non-coherent light is limited to the intensity of the object wave. These are the afore-mentioned *holograms*, which were theoretically developed by Dennis Gabor (Hungary and UK, 1900–1979) between 1948 and 1950. Lasers had not yet been invented at that point, but holograms became an important element of modern optics the moment that laser coherent light became easily available.

A hologram is basically a diffraction grating under Fresnel conditions. We shall start by considering a simpler grating, namely the zone plate, which is based on the theory of diffraction developed by Augustin-Jean Fresnel between 1815 and 1819.

We start here by defining the *Fresnel zones*. Consider a monochromatic plane light wave propagating in the positive direction of the z axis. The xy plane is then on a wavefront. Consider the point C on the z axis of co-ordinates (0, 0, z), as in Fig. 8.6. The light wave in C can be thought of, for the Huygens-Fresnel principle, as being the result of the contributions of infinite secondary elementary sources on the xy plane. All their phases are equal.

Due to the symmetry of the problem, all the sources laying at the same distance ρ from the origin are under the same conditions. Consequently, we can divide the xy plane into circular zones of radii ρ and $\rho + d\rho$ centered on the origin. The amplitude of the secondary wave emitted by each zone shall be proportional to $2\pi\rho\, d\rho$ and shall reach C with the phase $\phi(\rho)$, which is a function of ρ. The phase depends on the length of path PC, which is equal to $\sqrt{z^2 + \rho^2}$. What does matter, as usual, are the phase differences, between, say, the wave corresponding to $\rho = 0$, namely $\phi(0)$, and the one at the generic ρ, namely $\phi(\rho)$.

With reference to Fig. 8.6, we have $\phi(\rho) - \phi(0) = -k\,PQ$ where k is the wave number. In all the interesting cases, ρ is small compared to z. Then, for Eq. (7.1), we can write $PQ = \rho^2/(2z)$, obtaining

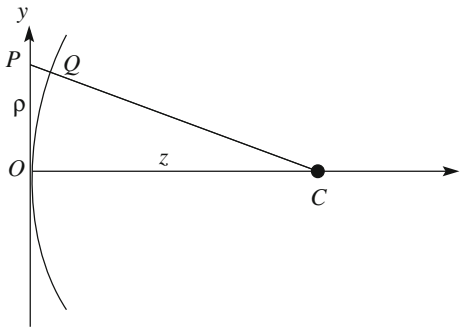


Fig. 8.6 Geometry used to define the Fresnel zones

$$\phi(\rho) - \phi(0) = -k\,\frac{\rho^2}{2z}. \tag{8.12}$$

We define the zones in which ϕ varies by π as *Fresnel zones* relative to the point C. In other words, the radius ρ_m of the mth Fresnel zone is such that $|\phi(\rho_m) - \phi(0)| = m\pi$, or

$$\rho_m^2 = \frac{2\pi mz}{k} = mz\lambda \tag{8.13}$$

and, hence,

$$\rho_m = \sqrt{mz\lambda}. \tag{8.14}$$

We see that the rays of the Fresnel zones increase *as the square roots of the integer numbers* do likewise.

We now consider laying a screen on the xy plane that completely absorbs the incident light. We then open a circular aperture in the screen centered in O, of radius r, which we assume to be capable of varying. We measure the light intensity in C while we gradually increase r.

Starting from $r = 0$, the light intensity in C initially increases, as we are including an increasing number of contributions that are substantially in phase with one another. Indeed, as can be seen in Fig. 8.6, the distance between the sphere of center C and the plane xy is close to zero for small values of ρ.

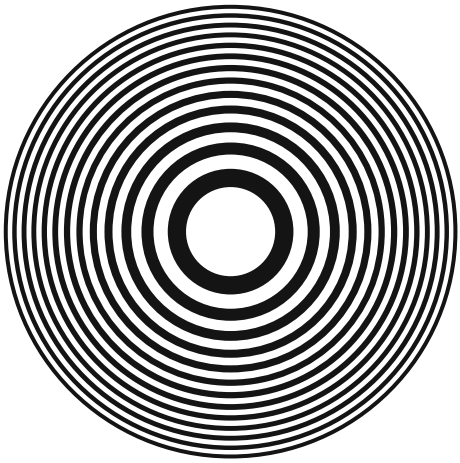
When r reaches the radium ρ_1 of the first Fresnel zone, the phase difference between the contributions of the center and the periphery of our aperture reaches the value of π. Consequently, if we increase r beyond ρ_1, we start adding contributions opposite in phase to those near O, and the intensity in C decreases. The decrease continues up to the point at which r is equal to ρ_2 of the second Fresnel zone. We have a minimum here. Continuing beyond ρ_2, we add contributions of the same sign as those of the first zone. And so it goes on; increasing r, we obtain a sequence of alternate maxima and minima.

To know the amplitudes of the contributions of the different zones precisely, we should take into account the obliquity factor introduced, specifically by Fresnel, in the Huygens-Fresnel principle. Without entering into the details, we simply mention that the amplitude in C of the secondary wave emitted between ρ and $\rho + \mathrm{d}\rho$ slowly decreases when ρ increases.

8.5 Zone Plate

In this section, we discuss the zone plate. The plate was described for the first time by Jacques Louis Soret (Switzerland, 1827–1890) in 1875. A zone plate is a grating with perfectly transparent zones alternated with perfectly absorbing ones, having

Fig. 8.7 The zone plate



radii equal to those of the Fresnel zones, namely ρ_1, ρ_2, ρ_3,.... The plate blocks all the even Fresnel zones and lets through the odd ones (or vice versa). In this way, the phases of all the selected contributions interfere constructively in C and the resulting intensity is high at that point. Figure 8.7 shows a zone plate.

As always, it is useful to look at the orders of magnitude. Let, for example, $z = 1$ m be the distance of the reference point C and $\lambda = 0.5$ μm the wavelength of the light, and let us consider the radii of the 50th and 51st zones, which, for Eq. (8.14), are, respectively, $\rho_{50} = 5.00$ mm and $\rho_{51} = 5.05$ mm. Hence, the distance between two consecutive zones at this order is only 50 μm and is clearly smaller and smaller with increasing orders.

Recalling Eq. (8.14) giving the radii of the Fresnel zones, we state that a zone plate is a circular grating with alternatively transparent and absorbing circular zones, periodic in the square radius of the zones ρ^2 with period (in ρ^2) equal to

$$\Pi = 2z\lambda = 2\rho_1^2. \tag{8.15}$$

The factor 2 corresponds to the fact that the period is of *two* Fresnel zones; z is the distance of point C from the plate.

Inverting the argument, suppose now that we have a zone plate with a given period Π. The radii of the transparent and absorbing zones are in the sequence

$$\rho_m = \sqrt{\frac{m\Pi}{2}} = \sqrt{mz\lambda}. \tag{8.16}$$

If we now send a plane monochromatic light wave normally on our plate, the wave will be focused at the point of the axis at the distance z. This distance is implicitly given by Eq. (8.16) as a function of the radii of the zones. We see that the action of the zone plate is similar to that of a converging lens with focal length z. In analogy to the lens, we call this distance f_1 (we shall soon see the reason for the

subscript). The focal length depends on the geometry of the plate and on the wavelength. We can express the focal length in terms of the period or, equivalently, of the first radius. Namely, it is

$$f_1 = \frac{\Pi}{2\lambda} = \frac{\rho_1^2}{\lambda}. \tag{8.17}$$

There are other similarities between the zone plate and the lens, but also differences. The most important difference is that the plate has not one, but rather infinite focal lengths, both positive and negative.

Indeed, going back to the arguments we made to maximize the intensity at a given point C, we see that the effect can also be obtained letting the first three Fresnel zones be open, the next three closed, and so on. Again, all the transmitted contributions interfere constructively, but each of them will be in amplitude of (about) 1/3 as before, because the contribution of the first third will cancel that of the second approximately. The period in ρ^2 of this plate is three times the period of the previous one. Inverting the argument, a zone plate of a given period focuses a plane wave incident on the axis not only at a point of the axis at distance f_1 but also at another one at the distance $f_2 = f_1/3$. The intensity of the corresponding maximum is lower than that in f_1 by a factor of (about) 1/9.

The same argument is obviously valid for any group of odd numbers of Fresnel zones. We can then state that the plate has an infinite sequence of focal lengths. The terms of the sequence are called orders. The focal lengths are given by the expression

$$f_{n+1} = \frac{\Pi}{2(2n+1)\lambda} = \frac{\rho_1^2}{(2n+1)\lambda} \quad n = 0, 1, 2, \ldots \tag{8.18}$$

The intensity of the maxima decreases rather quickly with the order n as $1/n^2$.

However, there is more. Indeed, there are as many focuses at the coordinates $-f_{n+1}$ on the negative z-axis, secularly symmetrical about the grating. In other words, in the diffracted wave, there are both spherical waves converging towards their centers at the distances of Eq. (8.18) with $z > 0$, but also spherical waves with centers in the symmetric points with $z < 0$ that propagate as if they originated at those points. We can state that the zone plate produces a sequence of *real images* (at $z > 0$) and one of *virtual images* (at $z < 0$) of a point at a finite distance on the axis.

To prove what we have just stated, let us consider Fig. 8.8. The zone plate is in the xy plane and a plane monochromatic wave is incident from the left along the z-axis. Consider a spherical surface of radius f_1 with center in C' symmetric, with respect to the plane xy, to the point C, which contains the first order focus. This surface is not a wavefront, because there is no diffracted wave at the left of the grating. However, if we propagate that surface as if it were a wavefront, it will become real in the region $z > 0$. We can then consider our surface as a virtual wave originated in C'. Looking at Fig. 8.8, one easily understands that the phase delay at a point P of the plane $x\,y$ is the same as that which we considered in the previous section with reference to Fig. 8.6. There, the conditions for positive interference

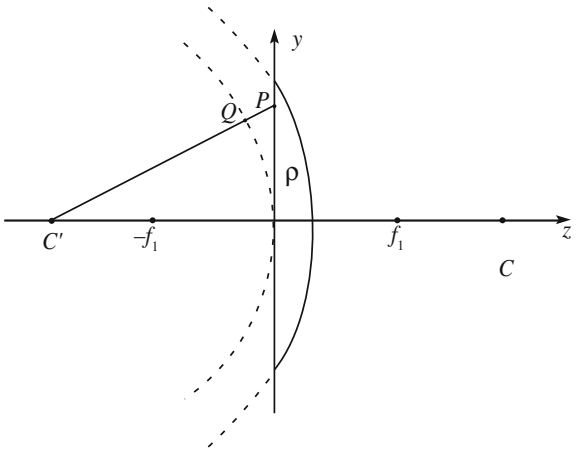


Fig. 8.8 Schematic of the geometry of the zone plate action for the first virtual image

were present in C, and now the same is valid for C'. A similar argument is obviously valid for all the focuses.

There are also analogies between the zone plate and the Fraunhofer grating. To see that, let us start by considering the zone plate divided into many sectors, each under a small angle at the vertex. Each sector contains a series of circular arcs, alternatively transparent and opaque. We approximate them as straight segments. The sector is thus similar to a Fraunhofer grating with a varying period. We intuitively understand that each sector gives a contribution similar to that represented in Fig. 8.9, and that we can obtain the total contribution of the zone plate by rotating the figure around the z-axis.

Note that Eq. (8.18) shows that the focal lengths depend on the wavelength, as is obvious considering that they are the result of an interference effect. Consequently, the image of a white source at infinite distance on the axis is not unique, but

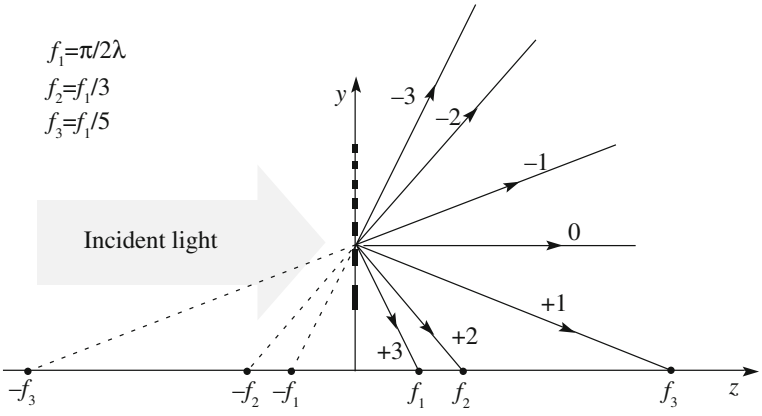


Fig. 8.9 Analogy between the Fraunhofer and zone gratings

dispersed in the different colors. In the language of the lenses, the zone plate suffers from a strong chromatic aberration.

Finally, we notice that a zone plate having opaque and transparent zones in inverted positions with respect to the one we just considered behaves similarly to the latter. In particular, the focal lengths are equal.

The zone plate produces images by exploiting the diffraction phenomena. In the next section, we shall see that it also behaves similarly to a lens for light sources at finite distances, both on the axis and otherwise.

8.6 Action of the Zone Plate on a Spherical Wave

We have seen how a zone plate acting on a plane monochromatic wave incident on its axis behaves like a "multiple" lens (i.e., with several focal lengths). This is the case for axial and paraxial spherical waves as well, as we shall now see. Consider a monochromatic spherical wave with center in A on the axis of the plate, at its left at the distance p. The arguments will be similar to those we developed in Sect. 8.4 for the plane wave. The geometry, analogous to Fig. 8.6, is now shown in Fig. 8.10.

Consider a generic point B on the axis on the right side of the grating, at a distance that we call q. Using the Huygens-Fresnel principle, we can consider the wave in B as being the result of the contributions of secondary sources on the xy plane. We calculate the phase difference in B between the wave generated by a secondary source at the distance ρ from the axis (point P in the figure) and the wave generated by the secondary source in the center O.

The path difference between APB and AOB is the sum of the paths RP and PQ. In the small angle approximation, we can now write

$$\phi(\rho) - \phi(0) = -k\,\frac{\rho^2}{2p} - k\,\frac{\rho^2}{2q} = -k\,\frac{\rho^2}{2}\left(\frac{1}{p} + \frac{1}{q}\right). \tag{8.19}$$

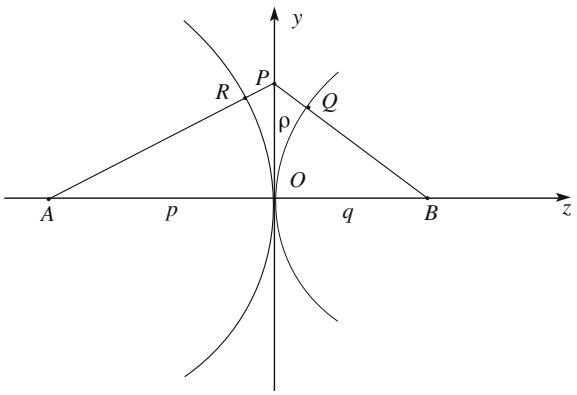


Fig. 8.10 Incident and transmitted spherical waves on the axis of a zone plate

The following argument is identical to the one in Sect. 8.4, with $(1/p + 1/q)$ in place of $1/z$, as one sees comparing Eq. (8.19) with Eq. (8.12). We can thus conclude that a light wave with its source on the axis in A will be focused by a zone plate in a series of points along the z axis with coordinates q_n given by

$$\frac{1}{p} + \frac{1}{q_n} = \frac{1}{f_n}. \tag{8.20}$$

The zone plate behaves like a multiple focal length lens for spherical waves with centers on the axis.

Notice that in Eq. (8.20), the images at $q_n > 0$ are on the opposite side of the zone plate relative to the object, and are real images, while those at $q_n < 0$ are on the same side as the object, and are virtual.

Consider now an off-axis source, like A in Fig. 8.11, at the distance p on the left side of the zone plate. If all the angles of the rays with the axis are small, one can show, in a manner similar to that which we used for the lens, that the interference maximum is at the point B, on the line through A and O at the distance q_1 from O such that

$$\frac{1}{p} + \frac{1}{q_1} = \frac{1}{f_1}.$$

The zone plate forms a real image of A in B. In addition, it can be similarly shown that that there is a succession of real images on the right side of the zone plate and virtual ones on its left side along the line AB at the distances q_n given by Eq. (8.20). In conclusion, a zone plate behaves like a multiple lens with both positive and negative focal lengths. As with a lens, it can produce an image, or, more precisely, many images, of extended objects.

Zone plates are used, in practice, to form images of objects when dealing with radiations for which lenses cannot be fabricated. This is the case with electromagnetic waves of wavelengths much larger than those of visible light, like

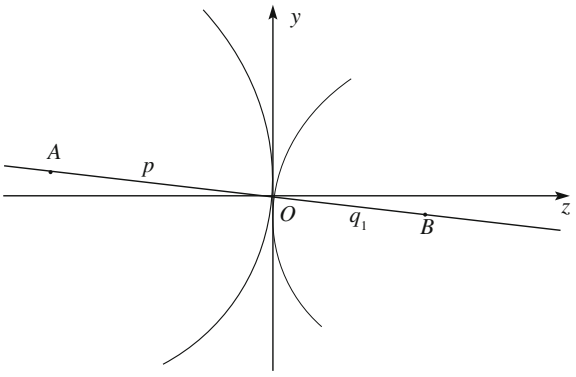


Fig. 8.11 Incident and transmitted spherical waves off axis of a zone plate

microwaves (wavelengths on the order of centimeters), or much shorter ones, like the extreme ultraviolet (100–10 nm) and X-rays (10–0.1 nm). At atomic and subatomic scales, particles are associated with waves. A mono-energetic beam of atoms or molecules is also a monochromatic atomic or molecular wave. Zone plates are used as focusing elements for them. Other examples are those of acoustic waves, for which zone plates also find applications for focusing purposes.

8.7 Camera Obscura

The simplest zone plate one can imagine consists of a single zone. Clearly, the properties of the image-forming properties of the zone plate discussed in the previous sections remain valid, even with reduced luminance, if only one Fresnel zone is open, say, the central one. This is simply a circular hole. Being that the hole is usually small, but not necessarily, as we shall see, one talks of a *pinhole camera*. The *camera obscura* is based on the same principles.

Figure 8.12 shows the basic concepts. Let r be the radius of the circular hole in an opaque screen and let us have a monochromatic plane wave of wavelength λ normally incident on it. Under these conditions, only the first focal length, namely f_1, is relevant, and we call it simply f.

Beyond the screen, we shall have a segment of spherical wave converging to the axis with curvature radius

$$f = \frac{\rho_1^2}{\lambda} = \frac{r^2}{\lambda}. \tag{8.21}$$

The corresponding dioptric power, in terms of the diameter d of the hole, is

$$\frac{1}{f} = 4\frac{\lambda}{d^2}. \tag{8.22}$$

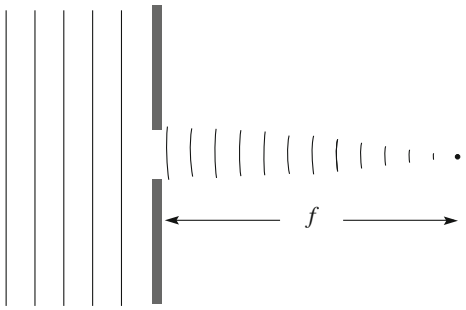


Fig. 8.12 Action of a pinhole on a normally incident monochromatic plane wave

If the incident wave is spherical with its center on the axis at the distance p from the hole, the outgoing wave will be approximately spherical with its center on the axis at the distance q such that

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f}. \tag{8.23}$$

Let us look at the orders of magnitude. Equation (8.22) shows us that, in order to have powers on the order of several diopters, the hole should be rather small. If, for example, $d = 0.5$ mm and $\lambda = 0.5$ μm, the dioptric power is about 8 m^{-1}. Objects at infinite distance (large distances, in practice) give an image at 125 mm beyond the hole. If the diameter is ten times larger, $d = 5$ mm, its power is 0.08 m^{-1} and the images of far objects are 12.5 m beyond the hole.

One difference between the pinhole and the zone plate is that the image produced by the former is the result of diffraction, while that produced by the latter is due not only to diffraction, but also to the interference between the waves coming from the different zones. The interference maxima are narrower for larger numbers of zones, as we have seen. Consequently, the focal lengths are more sharply defined if the zones are great in number. Now, we consider a single zone, and the focal length given by Eq. (8.21) is not sharply defined at all. In other words, the depth of focus is large. As a consequence, even when pinholes are used in white light, as is usually the case, the effect of the wavelength dependence of the focal length (chromatic aberration) shown by Eq. (8.21) is not very relevant.

We also note that the imaging action also exists if the aperture is smaller than the first Fresnel zone. Indeed, in this case too, for example, for an incident plane wave, the Huygens-Fresnel wavelets from all the elements of the surface of the aperture interfere constructively on the focal plane. Consequently, the hole can have any shape, provided it is contained in the central zone.

We can easily observe the imaging capability of a pinhole by drilling a hole in a card with a needle and then observing well-illuminated objects through the pinhole in front of one eye. It will appear to works like a magnifying glass. You can also fold one finger, leaving a small aperture between its joints and look through it.

The camera obscura (‘dark chamber’ in Latin) is an image-producing device known since ancient times, based on the same principle. It is schematically shown in Fig. 8.13.

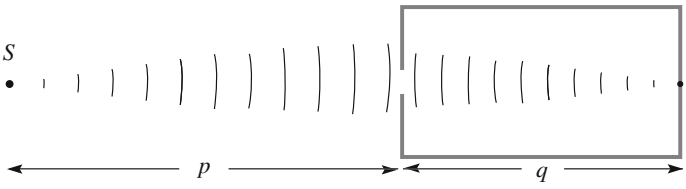


Fig. 8.13 The camera obscura principle

The camera obscura is a light, tight box with a pinhole on one side and a frosted piece of glass on the opposite side, created to enable looking at the images from the outside. Alternatively, one can open a narrow hole in the window shutter of an otherwise dark room and look at the images on the opposite wall.

Notice that the image formation in the camera obscura is a genuine diffraction process, due to the wave nature of light, and not to the rectilinear propagation of rays, as is often claimed. This conclusion can be easily reached with the experiment represented in Fig. 8.13. Assuming the length q of the chamber to be on the order of a few meters, we observe clear images with a hole about 0.5 mm in diameter. If the diameter is 20–50 μm, the images are very imperfect, and there will be no image at all, save for a uniformly illuminated screen, with a 1 μm diameter hole.

Let us now explain a common observation. Looking under a tree on a sunny day, you observe shadows and illuminated regions. There are shadows that geometrically replicate the shapes of the leaves and irregular light patterns projecting the gaps in the foliage. The geometric similitude of the shadow is the case with the leaves that are relatively close, in practice, within a few meters. However, if the tree is tall enough, there are also luminous discs, several centimeters in diameter, as shown in Fig. 8.14a. These are pinhole images of the sun. The "pinholes" are the gaps in the foliage, located far enough apart to be equal or smaller than the central Fresnel zone.

If you do not believe this, we can check it as follows. Wait for a partially clouded sky. The light should be strong enough to produce shadows alternating with bright regions, but the sun should be behind a thin cloud, making its image very fuzzy. If you look behind trees under these conditions, you will see bright regions and shadows, but no circular image.

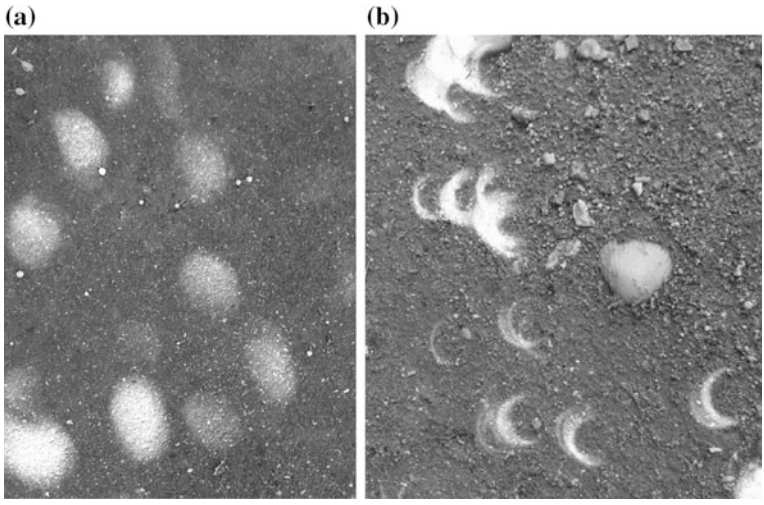


Fig. 8.14 "Pinhole" images of the sun cast on the ground by the gaps between the leaves of a tree, **a** under usual conditions. Picture by the author, **b** during a partial eclipse. Picture by Paul D. Maley

More spectacular is the phenomenon that occurs during a partial eclipse. The images are not round, but rather crescent-shaped, as is the sun. Figure 8.14b shows an example. If you are really, very lucky and observe an annular eclipse, then you will see luminous rings (easily found on the web).

QUESTION Q 8.1. Having observed images of the sun of diameter $D = 5$ cm under a tree, estimate the height of the gaps generating those images [Hint. The sun is seen under an angle of about ½° from earth].

8.8 Gabor Grating

The Gabor grating, which we shall now discuss, is one of the simplest possible holograms, namely the hologram of a luminous point at finite distance. It is useful here to recall three results we reached in previous sections. First, the Fraunhofer diffraction pattern of a grating of parallel strips alternating between completely absorbing and completely transmitting has an infinite discrete series of principal maxima, which correspond to the spatial Fourier components of the amplitude transmission coefficient of the grating. The transmission coefficient is a periodic "square wave" with alternate values of 0 and 1. Second, the diffraction pattern of a similar grating, but with amplitude transmission coefficient varying as a sine, plus a constant, only has the first order maxima, beyond the central one, corresponding to the components of its spatial Fourier transform. Third, the zone plate, which is a grating of circular zones alternating between completely absorbing and completely transmitting produces a series of diffraction maxima of increasing order on its axis. We now observe that the amplitude transmission coefficient of the zone plate, which depends on the distance ρ from the center, is, like the first case above, a "square wave" periodic in ρ^2 with alternate values of 0 and 1 (every half a period), as shown in Fig. 8.15a.

In our discussion of the zone plate in Sect. 8.6, we only considered the phase differences between the Huygens-Fresnel waves emitted by the different zones and found the points on the axis where their interference is constructive. The reason for the presence of the multiple focuses should be traced to the fact that the contributions of the different parts of a single Fresnel zone are not exactly in phase with

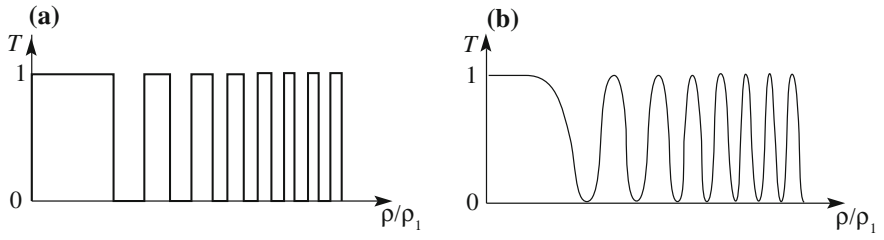


Fig. 8.15 Amplitude transmission of **a** a zone plate, **b** a Gabor grating

one another. As a matter of fact, as one can imagine from the three points recalled above, a zone plate with amplitude transmission coefficient T that is a sine (or cosine) function of ρ^2 produces only two diffracted spherical waves, namely the waves on the order +1 and −1.

As usual, to avoid negative values of the transparency T, we must add a constant term. Then, we have

$$T(\rho) = a\left[1 - b\cos\left(\frac{2\pi}{\Pi}\rho^2\right)\right], \tag{8.24}$$

where Π is the period in ρ^2 and b is the modulation. Figure 8.15b shows the transparency amplitude as a function of the distance from the center ρ relative to the radius of the first zone, namely of ρ/ρ_1. We shall call this zone plate a Gabor grating.

We state, in conclusion, that a plane monochromatic wave of wavelength λ normally incident on a grating whose amplitude transmission coefficient is the same as in Eq. (8.24) gives origin to two spherical waves with centers at two points of the axis, symmetrically located about the grating. Their distance from the grating is

$$f_1 = \frac{\Pi}{2\lambda} = \frac{\rho_1^2}{\lambda}, \tag{8.25}$$

In addition, after the grating, there is also a plane wave that we may consider simply to be a fraction of the incident wave. This wave is the consequence of the constant term (order zero) in Eq. (8.24) and is more intense the smaller the modulation b is. The centers of the two spherical waves are two images, one virtual and one real, of the source producing the incident wave.

How can we produce a Gabor grating? We can do that in much the same manner as we produced the parallel lines sine grating with a photographic process in Sect. 8.3. Now, we must produce interference between a plane wave and a spherical wave. We start from a laser beam and expand it with a beam expander, producing a monochromatic plane wave, which we show in Fig. 8.16. We position the photographic film perpendicularly to the propagation direction of the plane wave. We chose the x and y axes of the reference frame on the film and the z-axis in the direction of the incident wave. We put a small reflecting sphere at a point P on the z-axis at the coordinate $-z_P$ in front of the film. Through diffraction, the small sphere produces a spherical wave, which we shall call the *object wave*, that is coherent with the incident plane wave, which we call the *reference wave*.

Let us calculate the illuminance of the photographic film exposed under these conditions, due to the interference of the two waves. Figure 8.17 shows the relevant geometry. Let Δ_O be the phase difference of the two waves in O, which is a constant, having no consequence in what follows. The phase difference between the two waves at the generic point Q at the distance ρ from the axis is

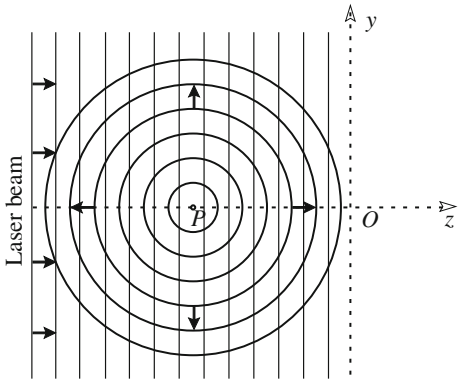


Fig. 8.16 A plane and a spherical wave for recording a Gabor grating

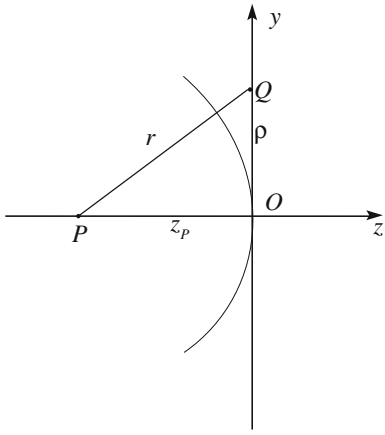


Fig. 8.17 Geometry of the interference of the plane (reference) and spherical (object) waves

$\Delta\phi = \Delta_O + k(r - z_P)$, where r is the distance between P and Q and k is the wave number. For small angles, namely for small values of ρ, we can approximate as usual, writing

$$\Delta\phi = \Delta_O + \frac{k\rho^2}{2z_P}.$$

With a calculation completely similar to that developed in Sect. 8.3, one establishes that the illuminance of the film as a function of ρ is given by

$$I \propto E_r^2 + E_o^2 + 2E_r E_o \cos\left(\frac{k\rho^2}{2z_P} + \Delta_O\right), \tag{8.26}$$

where E_r and E_o are the amplitudes of the reference and object waves, respectively. This expression is analogous to Eq. (8.7) for the straight lines sine grating. Under the usual conditions, the intensity of the object wave is much smaller than that of

the reference beam, namely $E_o/E_r \ll 1$. Neglecting terms in $(E_o/E_r)^2$, we can approximate Eq. (8.26) with

$$I \propto E_r^2 \left[1 + 2\frac{E_o}{E_r}\cos\left(\frac{k\rho^2}{2z_P} + \Delta_O\right)\right].$$

Following the same photographic process we mentioned in Sect. 8.3, we obtain a grating with an amplitude transmission coefficient inversely proportional to the illuminance, namely

$$T \propto 1/I \propto 1 - 2\frac{E_o}{E_r}\cos\left(\frac{k\rho^2}{2z_P} + \Delta_O\right). \tag{8.27}$$

We have now the Gabor sine grating that we wanted. In particular, its period in ρ^2 is

$$\Pi = 2\lambda z_P, \tag{8.28}$$

which depends, as expected, both on the wavelength and on the distance of the object from the grating.

We shall now use the Gabor grating to reconstruct the object wave that was present in the recording process. To do that, we produce an expanded laser beam, exactly equal to the reference wave we used in the recording process, and have it normally incident on our grating, as shown in Fig. 8.18. We shall call it the *reconstruction wave*. According to the conclusions of the previous section, beyond the grating, there are three waves, produced by diffraction. One is a fraction of the incoming wave that is of no interest here, the others are two spherical waves with centers in the two focuses, namely on the two sides of the grating, both at the same distance

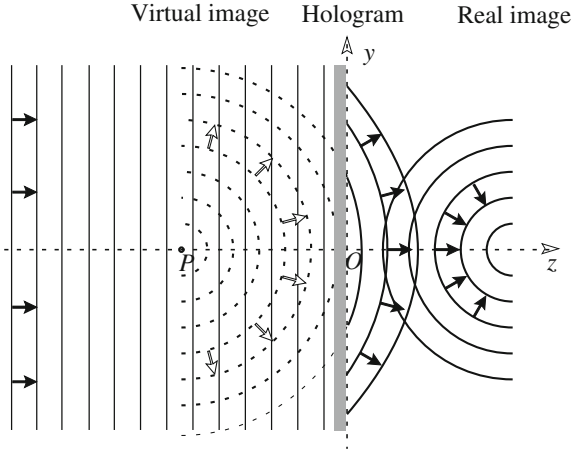


Fig. 8.18 Reconstruction of the wavefront

$$z_i = \frac{\Pi}{2\lambda} = z_P. \tag{8.29}$$

The wave from the forward focus is from a virtual image which is exactly where the object, namely the small sphere, was during the recording process. We have reconstructed the wavefront of the object.

The grating that we have recorded is similar to a photogram and is called a *hologram* of the point object P. If we illuminate it, as in Fig. 8.18, with a wave equal to the reference wave, and we look through it from the positive z side, we see the original object in the position it held when it existed during the recording process.

8.9 Holograms

We are now ready to record a hologram and subsequently to reconstruct the image it encodes. In the previous section, we recorded the hologram of a point-like object. Let us now substitute a three-dimensional object for the point source and illuminate it, as we did previously, with a reference plane monochromatic wave. Each point of the object will now scatter light as the small sphere in the previous section did, producing its Gabor grating on the photographic plate. The hologram thus obtained will hence be a coherent superposition of many Gabor gratings, each with its own center, intensity and phase. The hologram does not resemble the original object in any way, but contains complete information on both the amplitude and the phase of the object wave. The distance from the hologram (namely the z coordinate) of each point of the object is encoded in the period of the corresponding Gabor grating, while the other two coordinates x and y are encoded in the position of the center of the grating and the intensity in the modulation.

Let us now illuminate the hologram with the coherent reconstruction wave, which is a wave equal to the reference wave in the recording process, as we did in Sect. 8.9 for the point source. The wavefront of the light which was emitted by the object during recording will be faithfully reconstructed. Two three-dimensional images of the object, one real and one virtual, are formed at the two sides of the hologram, as shown in Fig. 8.19.

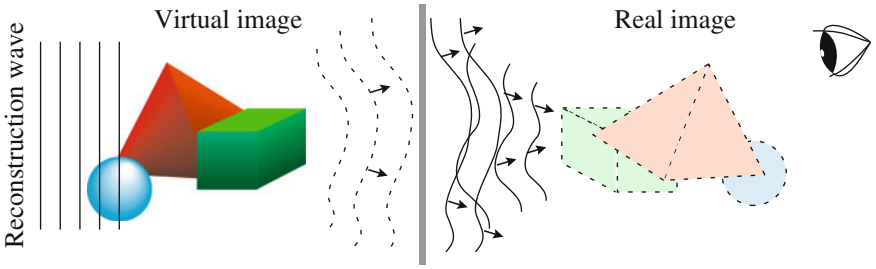


Fig. 8.19 Virtual and real images reconstructed in an online hologram

The virtual image can be observed by looking through the hologram from the side opposite to that of the incoming reconstruction wave, as shown in the figure. The real image cannot be immediately seen, for the already mentioned psychological reasons. To observe it, one has simply to position a white paper sheet in the image and scan through. Notice that, while the virtual image is a faithful reconstruction of the object, the real image is somewhat "inverted", but not like it would be in a mirror. We say that the virtual image is stereoscopic and the real one pseudoscopic.

Note that the virtual images (the interesting ones) that we observe are really three-dimensional. If you look through the hologram to the virtual image and move your head from the right to the left, you will see the image changing exactly as it would in the case of a real object, due to the changing parallax. You will also be able to see details that were not visible from the previous position, because they were behind some other detail. That is, everything will proceed as if you were observing a truly three-dimensional object through a window, which is the hologram.

Note, in addition, that if, for example, you cover a portion of any of a diffraction grating (Fraunhofer, Soret or Gabor) with a sheet of paper, the remaining portion gives rise to the same diffracted waves of the complete pattern, only with a loss in resolution (that is, with greater angular width of diffraction maxima). Hence, if you cover a portion of the hologram with a sheet of paper (or illuminate that portion only, reducing the diameter of the reconstruction beam), you will continue to see the complete three-dimensional image of the object, just through a smaller window. Even if, in particular, you leave free (or illuminate with the reference wave) only a small portion of the hologram, you will still be able to observe the complete image under different perspectives as if through a hole in a window.

The reason for all of this lies in the fact that the information contained in a hologram is, as noted above, much greater than that contained in a common photograph. The photographic plate, in both cases, records the intensity of the incident light wave. Consequently, a common photograph contains information on the amplitude alone. The information on the phase has been lost. In the case of the hologram, it is the existence of the reference wave and the state of its being coherent with the object wave that allows for transforming, in an interference process, the phase variations from point to point into amplitude variations. The consequent illuminance pattern is recorded on the photographic plate. Hence, the hologram records both the phase and the amplitude of the wave produced by the object, that is, it encodes all information on the light wave it emits, unlike an ordinary photograph, which contains only a part of the information. Indeed, the word comes from the Greek *olos* for "all, entire".

The geometrical configuration of the beams we have considered so far in the recording and subsequent observation of a hologram is called an *in line* configuration. It is simple, but has the disadvantage that, in reconstruction, the waves from both the virtual and the real images are present simultaneously in the same region of space. Under these conditions, some degradation of the images cannot be avoided. Consequently, in practice, one uses a somewhat different arrangement, in which

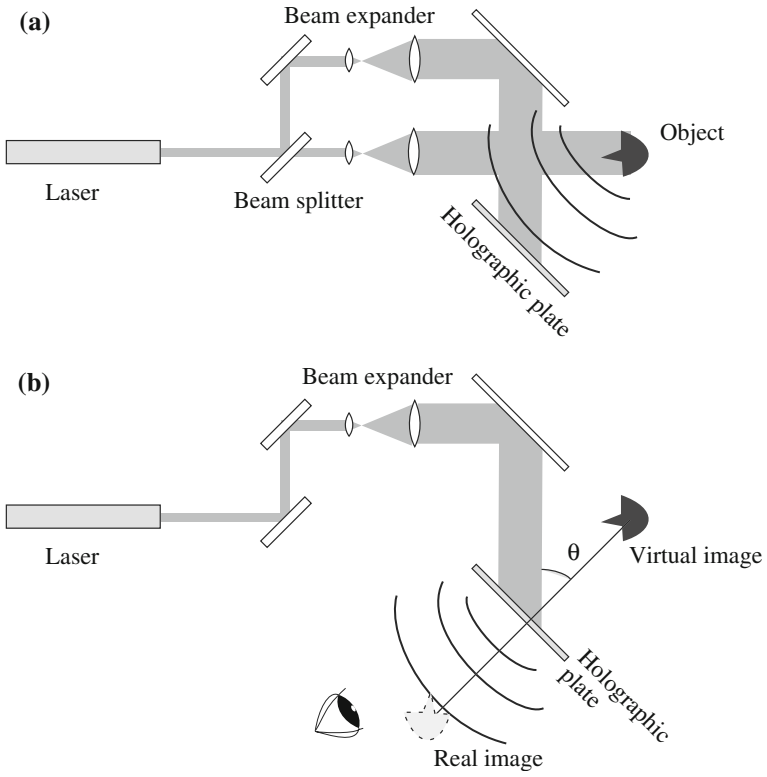


Fig. 8.20 **a** hologram recording process, **b** hologram viewing process

there is a non-zero angle θ between the reference beam direction and the average one of the light from the object. Figure 8.20a shows how this can be done. As you can see in Fig. 8.20b, in the subsequent reconstruction process, the two images, real and virtual, are better separated.

Recording a hologram requires, in practice, some important cautions. In fact, a hologram is the record of an interference pattern. The spacing of the fringes resulting from the interference between object and reference beams ranges between several micrometers and tenths of a micrometer (i.e., spatial frequencies ranging from several hundred to several thousand lines per millimeter).

The recording medium should be able to fully resolve the fringes. There are several recording media available, the most common one being the silver halide photographic emulsion, which is the one with the best resolution (see Sect. 7.17). Holography grade emulsions must be used, having a grain size in the 10–100 nm range. The standard photographic emulsions cannot be used, because their resolution is insufficient (see Sect. 7.17). In addition, the response of the recording medium should be flat over the full range of spatial frequencies to be recorded. Another important parameter is the exposure needed for recording. The exposure

per unit area is as low as 10–20 J m^{-2} for holographic film and much larger, about 1000 J m^{-2}, for photoresists. All the optical components, the photographic plate and the object itself, must be mounted rigidly, and any vibration during exposure must be avoided. Just purposefully think that a $\lambda/2$ shift of the photographic plate replaces the clear fringes with the dark ones. A further caution concerns the cleanliness of all the optical surfaces in the system. Just a speck of dust on the surface of a lens or on a mirror is enough, with the waves they scatter, to corrupt the even shape of the incident wavefront.

Summary

In this chapter, we studied two arguments, the Abbe theory of image formation and how to record and look at holograms, which produce genuine three-dimensional images. We learned the following principal concepts:

1. Image formation by a lens is a two-step process. The first step is the spatial Fourier transform of the object luminance produced by diffraction in the back focal plane, the second step is the anti-transform in the image plane produced by the interference between the light coming from different orders of the diffraction pattern.
2. The spatial filtering in the focal plane, the phase contrast microscope, in particular.
3. We have reached the concept of the hologram through a step-by-step process, considering, one after the other, the diffraction by a grating of parallel fully absorbing and fully transmitting stripes, a grating of parallel stripes with sinusoidal transparency, a Fresnel zone plate with fully absorbing and fully transmitting alternate zones, and a Gabor zone plate with zones of sinusoidal transparency. We have seen how to photographically produce such gratings.
4. How to record and look at the holograms of three-dimensional objects.

Problems

8.1 What are the radii of the first five Fresnel zones and of the zones 100 and 101 for a plane wave of wavelength λ = 500 nm and an observation point at the distance of 0.5 m?

8.2. What is the focal length of a camera obscura having a pinhole with a 0.2 mm diameter for light of 0.4 μm wavelength?

8.3. A Gabor grating focalizes a plane wave of wavelength λ = 670 nm at 10 cm distance. What is the radius of the first dark ring? And that of the thousandth one?

8.4. What is the hyperfocal length of a circular aperture of 2 mm diameter at 0.5 μm wavelength?

Index

A. Bettini, *A Course in Classical Physics 4 - Waves and Light*,
Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-48329-0