Chapter 2

Light

2.1 I - Light as electro-magnetic waves

2.1.1 Options for optics

The study of light and its interaction with matter is referred to as *optics*. It can be carried through at three levels of detail:

1. geometrical optics,
2. physical optics, or
3. quantum-mechanical optics.

Geometrical optics yields the crudest description. It is useful for predicting spatial paths of light that interacts with bodies much larger than its wavelength. More technically, it applies when no diffraction effects have to be reckoned with. A deeper understanding and better prediction of those interactions (e.g. reflection and transmission strengths) follow from physical optics, which considers light as electromagnetic waves. The modern view of quantum electrodynamics, on the other hand, is needed to explain aspects of emission and absorption, where the quanta by which optical, electromagnetic energy is created or destroyed are crucial. These massless quanta or elementary particles are referred to as *photons*.

2.1.2 Electro-magnetic waves

Since computer vision mostly deals with the analysis of reflected and transmitted light at macroscopic scales, but also captures the actual light intensities associated with these effects, physical optics is the choice that best matches its requirements. From the physical optics point of view, light is radiation composed of visible electromagnetic waves. It originates from accelerating electrons. This motion starts a series of interactions between an electric and magnetic field, following Maxwell's equations (see section 2.1.4 for a mathematical formulation). A moving electron constitutes an electric current. A current induces

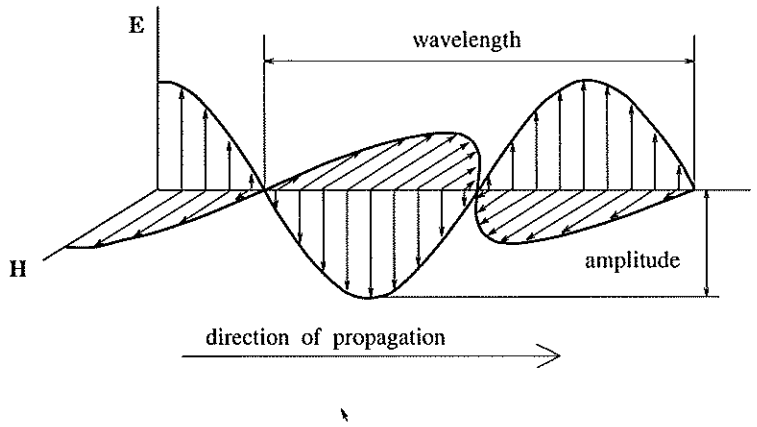


Figure 2.1: *Snapshot of propagating electromagnetic wave. The electric (* $\mathbf{E}$*) and magnetic (* $\mathbf{H}$*) fields are perpendicular. The direction of propagation is orthogonal to both.*

a magnetic field. Since the electron is accelerating, the current changes over time, and hence the magnetic field changes as well. A changing magnetic field generates a changing electric field, that in turn creates a changing magnetic field, etc. In the sequel, the vectors of the electric and magnetic fields will be denoted by $\mathbf{E}$ and $\mathbf{H}$, respectively, and similarly, $\|\mathbf{E}\|$ and $\|\mathbf{H}\|$ will denote their respective magnitudes.

Not just any electric and magnetic fields are compatible, however. For instance, in homogeneous non-conducting media, sinusoidal electric and magnetic fields in perpendicular planes could regenerate eachother, as shown in fig. 2.1. Light coming from a real light source can be considered to be a superposition (i.e. sum) of such *harmonic waves*. Further justification for focusing on this particular type of wave is given in section 2.1.6, where a number of the following of its properties are also proved. The fields vary sinusoidally both spatially and temporally and they are in phase, as is also shown in the figure. In fact, such combined pattern of changes corresponds to sinusoidal field waves that propagate. They do so in a direction that is perpendicular to both the electric and magnetic field vectors, i.e. the waves are *transverse*. At every moment in time both fields will be supposed to be (spatially) constant in planes perpendicular to the direction of propagation, i.e. the waves are *plane* waves. These properties follow from Maxwell's equations.

And there is more. It was one of Maxwell's achievements to show that the mechanism of reciprocal regeneration of the fields also implied that the propagation speed must have a specific value that he recognized as the speed of light. This staggering speed – 299792 km/s (in vacuum) – will be denoted by c_0 [1]. The subscript 0 will be used to indicate values and parameters for vacuum. This velocity had been estimated before Maxwell's time, and the equality of the wave propagation speed and this velocity made him conclude that light was just a type of electromagnetic waves. The above considerations also apply when the light wave is not propagating through vacuum, but through some other translucent medium. The velocity of propagation, however, will be lower (cf. section 2.2.1).

[1] The almost standardized used of c derives from *celer*, Latin for fast.

Looking at fig. 2.1, the following list of parameters is required to fully describe an harmonic wave:

1) wavelength
2) direction
3) amplitude $\|\mathbf{E}\|_{max}$
4) phase
5) direction of polarisation

The *wavelength* is the length of one period of the electric and magnetic fields, as indicated in fig. 2.1. The *direction* of propagation of the wave is perpendicular to both the electric and magnetic field vectors **E** and **H**. The *amplitude* $\|\mathbf{E}\|_{max}$ indicates the maximum of the magnitude $\|\mathbf{E}\|$ attained by the electric field vector **E** (see fig. 2.1). The *phase* determines the position of the maxima of the fields at a certain moment in time. A wave having all parameters equal except for the phase, will at each moment be an exact copy of the original wave, but is shifted by an amount given by the phase difference along the direction of propagation. Waves with the same direction and wavelength and with a fixed phase difference are called *coherent*. In general, the orientation of **E** will change rather randomly about the direction of the wave (and so will the orientation of **H** since these vectors and the direction vector remain perpendicular at all times). If the orientation of **E** remains fixed, however, the light is said to be *planarly* or *linearly polarized* and the *plane of polarization* corresponds to the plane defined by **E** and the direction of propagation (in older books **H** rather than **E** is used to define this plane).

It might seem that the amplitude $\|\mathbf{H}\|_{max}$ of the magnetic field is missing from the list. However, the strengths of the electric and magnetic fields are connected by a simple relation, that depends on the medium:

$$\frac{\|\mathbf{E}\|_{max}}{\|\mathbf{H}\|_{max}} = \frac{\|\mathbf{E}\|}{\|\mathbf{H}\|} = \sqrt{\frac{\mu}{\epsilon}}$$

where μ is the (magnetic) permeability and ϵ the (electric) permittivity of the medium through which the wave propagates. Similarly, the frequency of the wave follows from the wavelength ($c = \lambda\nu$, with c the velocity of light, λ the wavelength, and ν the frequency).

Of the parameters in the list, the first three have a direct bearing on the way they are perceived by the human visual system. The direction determines the viewpoints from where the wave can be observed, the wavelength is related to the observed colour (cf. section 2.1.3), and the amplitude is linked to the intensity (energy) of the wave. Regarding the latter, it can be shown that the flux of electromagnetic power is given by the *Poynting vector* **S**, defined as

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \ .$$

In view of the transversal nature of the wave, this vector points in the direction of propagation. Its magnitude $\|\mathbf{S}\|$ specifies the power per unit area perpendicular to the direction of propagation and is given by

$$\|\mathbf{S}\| = \|\mathbf{E}\| \ \|\mathbf{H}\| = \sqrt{\frac{\epsilon}{\mu}} \ \|\mathbf{E}\|^2$$

or taking the time average $\langle\rangle$ for the sinusoidal variation

$$\langle \|\mathbf{S}\| \rangle = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}\|\mathbf{E}\|^2_{max} \ . \tag{2.1}$$

Phase and direction of polarization have no direct impact on human vision. Yet, many insects, fish, and amphibia are sensitive to polarization as part of their orientation system. Phase and polarization are also relevant for computer vision (chapter 4).

There are two basic types of radiation-related units of measurement:

measurement type		domain of application
radiometric	$\leftrightarrow$	physics
photometric	$\leftrightarrow$	psychophysics

Radiometry deals with the measurement of observer-independent quantities associated with radiant energy, whereas *photometry* involves the measurement of *visually evaluated* radiant energy. Radiometric units deal with any type of electromagnetic radiation, whether or not it is in the visual spectrum. Photometry, however, is restricted to the visible spectrum. Photometric quantities are much more difficult to handle, since they are intended to capture the subjective impression that light patterns create with human observers. Due to differences between observers, but also differences for a single observer (e.g. with age), the definition of photometric units causes severe problems. The problem was circumvented by considering the spectral response of an "average" young human observer. These difficulties are also reflected by the abundance of photometric units that are in parallel use. To make things worse, the use of the terms "radiometry" and "photometry" themselves is a bit fuzzy in the literature.

2.1.3 The wavelength of electromagnetic waves

A separate discussion of the wavelength parameter is in place here, since it is central to what optics and vision are about. Electromagnetic waves come in an extremely broad range of wavelengths and no device is able to capture them all. An overview is given in fig. 2.2. Visible electromagnetic waves have wavelengths in the range $\lambda = 380 - 760\ nm$ [2]. In this *visible range*, different wavelengths are perceived as different colours. Table 2.1 gives a (rather sloppy) relationship between wavelength and observed colour. It is important to emphasize that conversely, it is incorrect to conclude light is of a certain wavelength just by its colour. Since colour is such a tricky aspect of the perception of light, we will devote

[2] It was Isaac Newton (1642-1727) who discovered the multrispectral composition of light. Isaac Newton became professor at Trinity College of Cambridge University in 1668, short after the college could reopen (it had been closed for two years due to a pestilence epidemic). He owed this exceptional position to his teacher, the mathematician Isaac Barrow (1630-1677), who recognized Newton's talents and therefore offered him the teaching of his own course. During the years of the plague Newton already started working on optics at his mother's house, but it was only in 1704 that he published his colour theory in *Opticks*.

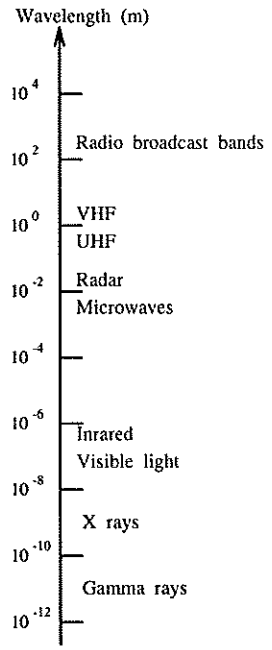


Figure 2.2: *The visible wavelengths only constitute a minute portion of the electromagnetic spectrum. These wavelengths correspond to distinct colours as indicated in table 2.1.*

wavelength (in nm)	colour
380 - 450	violet
450 - 490	blue
490 - 560	green
560 - 590	yellow
590 - 630	orange
630 - 760	red

Table 2.1: *Wavelengths and the corresponding observed colours.*

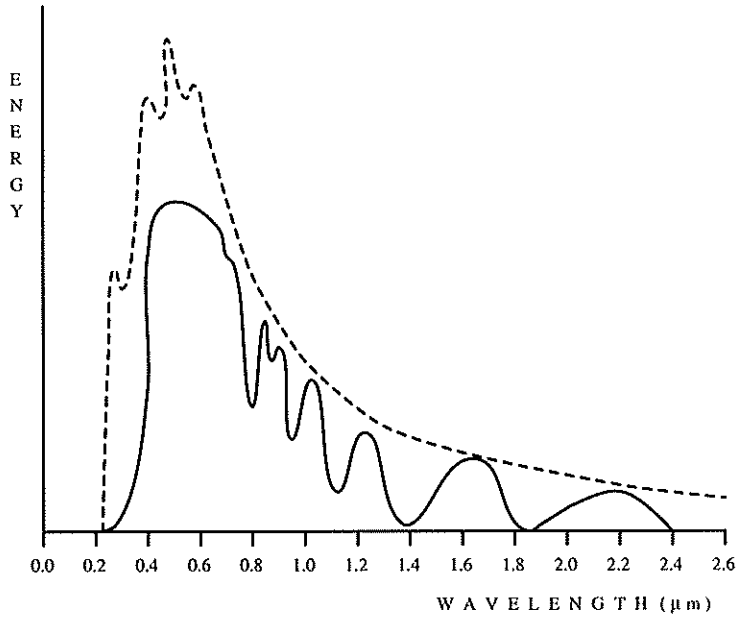


Figure 2.3: *Qualitative sketch of the relative energy for different wavelengths in sun light as observed below the earth's atmosphere (solid line) and above (dashed line).*

separate sections to this phenomenon (chapter 3 for the perception of colour and chapter ?? for the recording and displaying of colours with man-made equipment).

It is not by accident that biological vision systems sensitive to this particular frequency range have developed during evolution. The solar radiation spectrum has a peak around $500nm$. Due to scattering and absorption (see section 2.2.1) in the earth's atmosphere, the solar spectrum is considerably altered but the peak remains centered around that wavelength (green). As figure 2.3 shows, most of the sun's radiation that pierces through the earth atmosphere is within the visible range. The longer infrared wavelengths pass to some extent, but are less appropriate due to problems with thermal noise reducing the signal to noise ratio for such wavelengths. The shorter ultraviolet waves are blocked by the ozone layer (and hopefully for long, if we are not be turned into crisps).

The visible range and the neighbouring ultraviolet and infrared ranges are the ones used most in image processing applications. Together, they are referred to as the *optical range*, because they are handled with the same kind of instruments (lenses, gratings, etc.). The ultraviolet range is further subdivided in UV-A (320-400 nm), UV-B (290-320 nm), and UV-C (200-290 nm). Ultraviolet photons are particularly energetic. UV-C is filtered out completely by the time sun light reaches the earth surface, but UV-B is still present to a limited extent and causes burning of the skin. UV-A is less harmful but causes wrinkles. Also the infrared range is subdivided: near-infrared (0.7-1.3 μm), mid-infrared (1.3-3 μm), and thermal-infrared (beyond 3 μm).

It is important to note that cameras and image acquisition devices will in general not have the same spectral sensitivity as the human eye. As a consequence, wavelengths outside the visible spectrum might be included in the recorded data without this being an explicit

desire by the designer of the vision system. So will the relative sensitivity to different wavelengths not always be in accordance with human vision. Anyway, it is the visible range rather than any other part of the spectrum that is important to computer vision just the same. On the other hand, there are applications where the invisible bands are considered very important or even more important than the visible part, as e.g. the infrared in certain applications of remote sensing or X-rays in the processing of radiographies.

2.1.4 * Maxwell's equations

Maxwell [3] predicted the existence of electromagnetic waves in 1862. This conclusion he derived from a set of equations that he compiled. These equations not only summarized experimental knowledge at his time, but also contained some refinements that he felt necessary on theoretical grounds. Basically, the equations describe the interplay between electric and magnetic fields.

Two isolated charges exert forces onto eachother. The classical framework assumes that every charge generates an electric field, in which other charges will "feel" themselves immersed. The electric field at a certain position is denoted by the *electric field* vector $\mathbf{E}$. If there is a charge q there, it experiences a force $q\mathbf{E}$. If the charge were moving, it might experience another force that also depends on its speed $\mathbf{v}$. This is "explained" by the introduction of another field, the *magnetic induction* $\mathbf{B}$. The force exerted by this field on the charge is $q\mathbf{v} \times \mathbf{B}$. The concept of electromagnetic waves has everything to do with how these fields interact.

The Maxwell equations will not be discussed in their most general form here. It will be assumed that the medium in which the electric and magnetic induction fields exist is isotropic and homogeneous, and that these fields are exclusively due to the light wave, not to free charges or currents already present.
Faraday's induction law states that

$$\oint_C \mathbf{E} \cdot d\mathbf{C} = -\iint \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \ ,$$

i.e. the net change in magnetic induction $\mathbf{B}$ through the surface S with normal vectors $\mathbf{S}$ generates an electric field $\mathbf{E}$ that satisfies the above constrained when integrated along the surface's closed boundary curve C, with tangent vectors $\mathbf{C}$. In the absence of free charges (as assumed here), the electric field lines close onto themselves. The same applies to the magnetic induction. This is expressed by *Gauss' electric law*

$$\oiint_S \epsilon \mathbf{E} \cdot d\mathbf{S} = 0 \ ,$$

[3] James Clerk Maxwell (1832-1879). At Maxwell's time the scientific community already suspected light to be some kind of wave. Attempts to link electricity and light had already been made by Gauss and Riemann, and Faraday had noticed a connection between light and magnetism and had suggested an electromagnetic origin of light. The story goes that on the evening when Maxwell pinpointed the connection between electromagnetism and light, he had a date with his later wife. When she started to talk about the beauty of the stars, good old Maxwell proudly remarked that she was talking to the only person who understood what starlight really was.

and *Gauss's magnetic law*

$$\oiint_S \mu \mathbf{B} \cdot d\mathbf{S} = 0 \ ,$$

with S this time a closed surface, ϵ the electric permittivity, and μ the magnetic permeability of the medium. Finally, *Ampère's law* as modified by Maxwell states that

$$\oint \mathbf{B} \cdot d\mathbf{C} = \mu \iint_S \left(\sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t} \right) \cdot d\mathbf{S} \ ,$$

with σ the electric conductivity. This law implies that not only currents generate magnetic induction fields, but also changes in the electric field.

The magnetic induction $\mathbf{B}$ is often replaced by the *magnetic field* $\mathbf{H} = (1/\mu)\mathbf{B}$. $\mathbf{H}$ leads to expressions that are more symmetric in their electric and magnetic components.

Finally, a pointwise reformulation of the above, (non-general) Maxwell equations is obtained by using Gauss's divergence theorem

$$\oiint_S \mathbf{F} \cdot d\mathbf{S} = \iiint_V (\nabla \cdot \mathbf{F}) \, dV$$

and Stokes' theorem

$$\oint_C \mathbf{F} \cdot d\mathbf{C} = \iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S}$$

for a field of vectors $\mathbf{F}$, with $\nabla \cdot$ the divergence and $\nabla \times$ the curl. Readers not acquainted with these differential operators find their definitions in section 2.1.5. A differential formulation of Maxwell's equations is then found as

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \quad \text{Faraday's law} \tag{2.2}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} \quad \text{Ampère's law, as modified by Maxwell} \tag{2.3}$$

$$\nabla \cdot (\epsilon \mathbf{E}) = 0 \quad \text{Gauss's law for the electric field} \tag{2.4}$$

$$\nabla \cdot (\mu \mathbf{H}) = 0 \quad \text{Gauss's law for the magnetic field} \tag{2.5}$$

Note that the coupling between the electric and magnetic field vectors arises via the first two of the Maxwell equations. In the static case they are completely independent for dielectrics (i.e. non-conducting media for which $\sigma \simeq 0$). But once they change over time, they have to do so in a strongly coupled fashion. Also note that in isotropic and homogeneous media ϵ and μ are constant scalars, and therefore

$$\nabla \cdot \mathbf{E} = 0 \ , \tag{2.6}$$

$$\nabla \cdot \mathbf{H} = 0 \ . \tag{2.7}$$

2.1.5 * Interludium on differential operators

In the previous section we have introduced the triple $\nabla = (\frac{\partial}{\partial X}\frac{\partial}{\partial Y}\frac{\partial}{\partial Z})^T$ of differential operators. Considering this triple as a vector, its inner and vectorial product with other vectors were introduced and called the divergence and curl, resp. The notations were chosen accordingly. As will be shown in section 2.1.6 the analogy between the differential operators and vector calculations allows elegant and compact derivations.

As an example for readers not acquainted with the divergence and curl operators, we let them work on the electric field $\mathbf{E} = (E_X, E_Y, E_Z)^T$ where E_X, E_Y, and E_Z are the components along the spatial coordinate axes X, Y, and Z, resp. Taking $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ as the unit vectors along these axes, one defines

$$\begin{aligned}
\nabla\cdot\mathbf{E} &= \frac{\partial E_X}{\partial X} + \frac{\partial E_Y}{\partial Y} + \frac{\partial E_Z}{\partial Z} \ , \\
\nabla\times\mathbf{E} &= \left(\frac{\partial E_Z}{\partial Y} - \frac{\partial E_Y}{\partial Z}\right)\mathbf{i} + \left(\frac{\partial E_X}{\partial Z} - \frac{\partial E_Z}{\partial X}\right)\mathbf{j} + \left(\frac{\partial E_Y}{\partial X} - \frac{\partial E_X}{\partial Y}\right)\mathbf{k} \ .
\end{aligned}$$

There is another operator, the "Laplacian" ∇^2, that is of interest. This is defined as

$$\begin{aligned}
\nabla^2\mathbf{E} &= \mathbf{i}\nabla^2 E_X + \mathbf{j}\nabla^2 E_Y + \mathbf{k}\nabla^2 E_Z \\
&= \mathbf{i}\left(\frac{\partial^2 E_X}{\partial X^2} + \frac{\partial^2 E_X}{\partial Y^2} + \frac{\partial^2 E_X}{\partial Z^2}\right) + \\
&\quad \mathbf{j}\left(\frac{\partial^2 E_Y}{\partial X^2} + \frac{\partial^2 E_Y}{\partial Y^2} + \frac{\partial^2 E_Y}{\partial Z^2}\right) + \\
&\quad \mathbf{k}\left(\frac{\partial^2 E_Z}{\partial X^2} + \frac{\partial^2 E_Z}{\partial Y^2} + \frac{\partial^2 E_Z}{\partial Z^2}\right)
\end{aligned}$$

Note that this prescription for the Laplacian entails a definition of an operator ∇^2 operating on a vector and an operator denoted by the same symbol but operating on scalar functions. A final, but well-know operator is the *gradient*, which operates on a scalar function, e.g.

$$\nabla E_X = \nabla E_X = \mathbf{i}\frac{E_X}{\partial X} + \mathbf{j}\frac{E_X}{\partial Y} + \mathbf{k}\frac{E_X}{\partial Z}$$

It is also useful to remind the reader of a number of propositions governing the interactions of these operators and that will be applied in the remainder of the course:

1. The divergence of the curl of any vector function is zero everywhere.
2. The curl of the gradient of any scalar function is zero everywhere.
3. The curl of the curl of a vector function $\mathbf{E}$ is
$$\nabla\times(\nabla\times\mathbf{E}) = \nabla(\nabla\cdot\mathbf{E}) - \nabla^2\mathbf{E} \ .$$
4. The divergence of the vector product of two vector functions $\mathbf{E}$ and $\mathbf{H}$ is
$$\nabla\cdot(\mathbf{E}\times\mathbf{H}) = \mathbf{H}\cdot(\nabla\times\mathbf{E}) - \mathbf{E}\cdot(\nabla\times\mathbf{H}) \ .$$

2.1.6 * Derivation of the properties of light

Section 2.1.2 succinctly described the conclusions about light that are implicitly described by the Maxwell equations in the case of homogeneous and isotropic media. These will be made explicit in this section.

Now, applying the curl operator to eq. 2.2, we find

$$\nabla(\nabla\cdot\mathbf{E}) = \nabla^2\mathbf{E} - \mu\nabla \times \frac{\partial \mathbf{H}}{\partial t} \ .$$

Then, using eqs. 2.3 and 2.6 we obtain

$$\nabla^2\mathbf{E} = \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\sigma\frac{\partial \mathbf{E}}{\partial t} \ . \tag{2.8}$$

Similarly, taking the curl of eq. 2.3 and substituting for $\nabla \times \mathbf{E}$ from eq. 2.2, it follows that

$$\nabla^2\mathbf{H} = \mu\epsilon\frac{\partial^2\mathbf{H}}{\partial t^2} + \mu\sigma\frac{\partial \mathbf{H}}{\partial t} \ . \tag{2.9}$$

Some readers will recognize these equations as the equations of telegraphy. This result makes clear the behavioral *distinction between conductors and dielectrics.* Indeed, for dielectrics $\sigma = 0$ and this equation reduces to the well-known wave equation.

First, consider the case of dielectrics ($\sigma = 0$). In that case, the equations (2.8) and (2.9) simplify to

$$\begin{aligned} \nabla^2\mathbf{E} &= \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} \\ \nabla^2\mathbf{H} &= \mu\epsilon\frac{\partial^2\mathbf{H}}{\partial t^2} \end{aligned}$$

It comes out that a wave like

$$\mathbf{E} = \mathbf{E}_0\cos(\mathbf{k}\cdot\mathbf{X} - \omega t)$$

is a solution of the first equation if

$$\|\mathbf{k}\|^2 = \mu\epsilon\omega^2 \ .$$

Considering the equation

$$\mathbf{k}\cdot\mathbf{X} - \omega t = constant \ ,$$

positions with equal phase at a fixed time t are seen to lie in planes perpendicular to $\mathbf{k}$. Hence, these are plane waves. Planes of a fixed phase are seen to move along the direction $\mathbf{k}$ with time, at a velocity c such that

$$c = \frac{\omega}{\|\mathbf{k}\|} = \frac{1}{\sqrt{\mu\epsilon}} \ .$$

Also note that this particular solution, with $\mathbf{E}_0$ a fixed vector, represents a (linearly) polarized wave. In fact, any form of light can be represented by two orthogonal linearly

polarized waves, so that this isn't a real restriction. A similar type of solution is found for $\mathbf{H}$. Given the linearity of the above equations any linear combinations of such solutions are solutions as well. Hence, we can bring to bear the whole of Fourier analysis on the study of light by decomposing light waves into their harmonic components. This explains the attention paid to plane, harmonic light waves in section 2.1.2. For ease of computation, one usually prefers working with the complex expression

$$\mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{X}-\omega t)} \ , \tag{2.10}$$

where it should be understood that only the real part is endowed with a physical meaning. Note that the above expression satisfies the wave equation under the same conditions that applied to its real part.

Having found this basic type of solution for dielectrics, it is shown to have important implications. Applying the operator $\frac{\partial}{\partial t}$ to a solution of the form (2.10) has the effect of multiplication with $-i\omega$. Similarly, application of the partial derivatives composing ∇ corresponds to multiplication with the corresponding components of $i\mathbf{k}$. Hence, for complex harmonic waves the Maxwell equations can be rewritten

$$\begin{aligned} \mathbf{k}\times\mathbf{E} &= \mu\omega\mathbf{H} \\ \mathbf{k}\times\mathbf{H} &= -\epsilon\omega\mathbf{E} \\ \mathbf{k}\cdot\mathbf{E} &= 0 \\ \mathbf{k}\cdot\mathbf{H} &= 0 \end{aligned}$$

The above equations compactly represent the fact that $\mathbf{k}$, $\mathbf{E}$, and $\mathbf{H}$ must be mutually perpendicular. It follows that light consists of transverse waves. From the constraint that $\|\mathbf{k}\|^2 = \mu\epsilon\omega^2$ either of the first two equations implies that the electric and magnetic field amplitudes are coupled as

$$\frac{\|\mathbf{E}\|}{\|\mathbf{H}\|} = \sqrt{\frac{\mu}{\epsilon}} \ ,$$

as already mentioned in section 2.1.2.

Now, consider the case of conducting materials. The additional term $\mu\sigma\mathbf{E}$ in eq. 2.8 complicates matters. However, starting again from a temptative solution (2.10) but accepting $\mathbf{k}$ to be complex, the constraint that should be satisfied is found to be

$$\mathbf{k}^2 = \mathbf{k}\cdot\mathbf{k} = \mu\epsilon\omega^2 + i\mu\sigma\omega \ .$$

If $\mathbf{k} = \mathbf{k}_R - i\mathbf{k}_I$, then $\mathbf{k}_R$ still represents the direction of propagation, whereas $\mathbf{k}_I$ specifies the gradient of amplitude (direction of fastest attenuation) and results in the attenuation factor $e^{-\mathbf{k}_I\cdot\mathbf{X}}$. Although for this type of solutions the Maxwell equations can be rewritten in a form very similar to the one obtained for dielectrics, the same, simple interpretation no longer applies. $\mathbf{k}$ being complex precludes direct interpretations in terms of orthogonality conditions. In fact, a detailed analysis would show that the waves are no longer transverse and that the fields have non-zero components along the direction of propagation. As shown in section 2.2.2, conductors strongly absorb light (i.e. show a strong attenuation) and light

will hardly enter these materials. Conductors will reflect most of the light that falls on them anyway.

In the previous discussion, it was also silently assumed that the parameters μ, ϵ, and σ are real. In fact, as discussed in sections 2.2.1 and 2.2.2, if these equations are to remain valid for waves with increasingly short wavelengths – and as a matter of fact into the optical spectrum band – then ϵ and σ would have to be complex. Since having both refinements together leads to an overly complicated analysis, either one or the other will be considered. Hence, most of the remaining discussion on optical properties of matter is divided into the case of perfect dielectrics ($\sigma = 0$) and strong conductors (metals). Materials like semi-conductors would require both and are not considered. Note that a complex ϵ introduces attenuation also in the case of dielectrics. This is clarified in section 2.2.2.

2.1.7 Radiometric measurements of light

An electromagnetic wave carries energy, several quantitative aspects of which might be of relevance.

The total *radiant energy* coming from a source and falling on a certain surface, expressed in the usual MKS unit of joules, is significant when e.g. an accumulator type of detector such as a silver halide crystal in photographic film or a CCD element is considered. The energy per unit of time is called radiant power or, more often, *radiant flux*, and is expressed in watts. We are used to selecting light bulbs with sufficient wattage to satisfactorily light a room or workspace.

Often we are interested in energy or power per unit area of the source or the radiated surface. The radiant flux emitted per unit area of a source is the *exitance*, in the past also referred to as *radiant emittance*. The radiant flux incident per unit area of the radiated surface is the *irradiance* [4]. Boths are given in watts/m^2. Radiation from the sun has 1350 watts/m^2 of irradiance in the absence of atmospheric absorption if measured in a plane perpendicular to the rays. From the point of view of the radiated surface, the *radiant intensity* or the radiant flux emitted by the source per unit solid angle might be of more direct use than the exitance. Indeed, a source might radiate different amounts of energy in different directions. The solid angle of a cone of viewing directions with apex at the viewpoint is the area cut by the cone on the sphere with center at the viewpoint and unity radius and is expressed in *steradians*. The solid angle of a hemisphere therefore is $2\pi sr$. The surface patch in fig. 2.4, with area A and at distance R from the viewpoint has

$$\Omega = \frac{A \cos\theta}{R^2},$$

as its solid angle. In this expression θ is the angle between the local surface normal and the line through the patch center and the viewpoint.

If the source is large compared to its distance from the receiver, then the *radiance* or radiant flux emitted per unit solid angle and per unit *foreshortened* source area should be

[4] We prefer this terminology over *radiant illuminance*, which is highly confusing since luminance and illuminance will be seen to be photometric quantitities

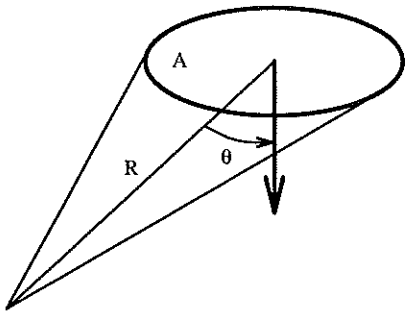


Figure 2.4: *Small surface patch viewed from a distance R under inclination θ.*

used. Foreshortened area is the area as seen from the radiated surface. As an example, the moon is seen as a circular disc, but the actual area on the moon that contributes to an area in the visual field differs strongly from the center towards the rim of the circle. Foreshortened area is obtained by multiplying the source area by the cosine of the angle between a perpendicular to the source surface and the specified direction. In fact, it is the same correction for oblique viewing as already discussed with the calculation of solid angles of small surface patches. Radiance is expressed in $\mathrm{W.m^{-2}.sr^{-1}}$ (watts per square meter per steradian). It is a most useful concept, since it captures the essential information one needs for calculating how much radiation from the source will reach a patch of the radiated surface, given the position and geometry of the source surface with respect to the radiated patch, i.e. for calculating irradiance at that patch. More on the relationship between image irradiance and source (object) radiance is to follow in chapter 4.

Usually the energy of the radiation will be spread over one or several intervals of the electromagnetic spectrum. The *spectral composition* specifies the energy per wavelength. Putting the integral of the composition curve to 1, the *spectral distribution* is obtained, and only relative data are available in that case. For all the quantities defined in this section, say Q, spectral versions, say Q_λ, can be directly derived as

$$Q_\lambda = \frac{dQ}{d\lambda}.$$

As an example, the spectral exitance is the exitance per wavelength and is expressed in $\mathrm{watts/m^3}$.

2.2 I - Interaction of light and matter

2.2.1 Interaction with matter – a phenomenological account

Through its interaction with matter, light might alter its direction, intensity, state of polarization, and sometimes even its wavelength. Four basic types of interaction will be discussed: *absorption, scattering, refraction,* and *reflection.* In fact, scattering could be regarded the most essential effect that also underlies the others. Yet, in this section,

the discussion will proceed from a phenomenological point of view and these four types correspond to clearly distinct phenomena at a macroscopic scale. Other effects such as *diffraction* and *fluorescence* are not considered. Diffraction is the deviation from straight propagation in the presence of obstacles and fluorescence corresponds to re-emitted radiation after an atom excited by incoming photons returns to its ground state. Both would require an account based on quantum-mechanical optics. Other rather atypical optical characteristics will also not be considered. Some special materials will e.g. reflect different wavelengths very differently. Such behaviour is increasingly used to prevent document falsification and might therefore be relevant in the context of document image processing.

Absorption

As light passes through a medium, even a transparent one such as air or water, it undergoes absorption, as photons collide with particles of matter, give up their energy and disappear. For instance, water absorbs more strongly the longer wavelength, hence its blue colour. The ratio of absorbed power over incoming power is the *absorptance* of the medium. The ratio of power that leaves the (translucent) medium over the original incoming power is referred to as the *transmittance.*

As explained further, absorption normally occurs in specific spectral bands. Where these absorption bands lie depends on the type of material considered. When the distinction is made for individual wavelengths, one speaks of *spectral absorptance* and *spectral transmittance.* In fact, the very existence of most life forms is due to the spectral absorptance characteristics of the atmosphere. Detrimental waves with wavelengths shorter than about 300 nm are absorbed by the ozone layer in the upper strata.

As mentioned in section 2.1.2 and shown in section 2.1.6, the speed of light is determined by the electrical permittivity ϵ and the magnetic permeability μ of the medium through which a light wave propagates and is given by

$$c = \frac{1}{\sqrt{\mu\epsilon}} \ .$$

The ratio of the speed of light in vacuum c_0 to this speed c is called the *index of refraction* $n = c_0/c$. The reason for this name will become clear when refraction is discussed. Contrary to Maxwell's initial thinking about the matter, the speed in a medium that is not a vacuum will depend on the wavelength, since the permittivity is wavelength dependent. Thus, also the index of refraction, which is an important parameter in the interaction between light and matter, depends on wavelength. This section explains this dependence for the case of dielectrics.

A simple electrostatic determination of ϵ as $K\epsilon_0$ with K the medium's dielectric constant K (capacitance of the medium over capacitance of vacuum) will only be correct for low frequencies if the medium contains dipole moments (charge $\times$ distance between the poles). If an electric field $\mathbf{E}$ is applied, these dipoles will orient themselves parallel to the field lines. Some of these dipoles are permanent (e.g. water molecules), but other are induced by the field $\mathbf{E}$ itself on the level of the electrons or ions. Now, suppose the electric

field E starts changing at an ever higher frequency. Initially, the dipoles will manage to realign themselves when the field changes sign. Each type of dipole has its own *natural* or *resonant frequency* below which the dipoles will manage to follow quite well, close to which they start to resonate, and beyond which they will have the opposite phase. Especially dipoles that find their origins in electron distributions have their resonant frequencies in the visible range. Vibrating atoms will yield dipole resonant frequencies that are found rather in the infrared region of the spectrum. Around the resonant frequencies there is an absorption band for which the medium is fairly opaque. There the dipole oscillations get large and dissipation due to work by the field on the charges increases. From a corpuscular point of view, near the resonant frequencies atoms can get excited and in a dense medium they will most probably return to their ground state after dissipating the excess energy thermally. Colourless, transparent media have their resonant frequencies outside the visible region.

No effort will be made here to derive these effects mathematically. The interested reader is referred to section 2.2.2. The following *dispersion equation* describes a simplified model:

$$\frac{\epsilon}{\epsilon_0} = 1 + \sum_j \frac{A_j}{\omega_{0j}^2 - \omega^2} \ .$$

Each dipole type j contributes a term in the sum, where A_j is a constant that doesn't have to concern us any further here, and ω_{0j} is the resonant frequency. Note that

$$n = \frac{c_0}{c} = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \simeq \sqrt{\frac{\epsilon}{\epsilon_0}} \ ,$$

since except for ferromagnetic materials $\mu \simeq \mu_0$. This model therefore predicts the frequency dependence of the index of refraction to a good degree of precision. Fig. 2.5 shows a typical *dispersion curve* $n(\omega)$, as specified by the corresponding dispersion equation. In between the resonant frequencies ω_{0j} the index of refraction increases. This phenomenon is called *normal dispersion.* When crossing such a frequency, the index will decrease, however. The absorption regions correspond to *anomalous dispersion.*

The differences in propagation speed between monochromatic waves of different frequencies has important consequences. If a light wave is composed of several frequencies – any real light wave with its finite extent has to be – then the waveform will definitely change over time because of these differences. The same dependence also forces one to decompose a wave into its sinusoidal components before calculations can be made.

Scattering

Light is scattered as it passes through a transparent or translucent medium. Rays are scattered on striking small particles of matter. Scattering of sunlight by the atmosphere is the reason why the sky is blue (the so-called Tyndall effect) since air primarily scatters the shorter wavelengths. For the same reason, having to pass thicker atmosphere layers at dusk and dawn, sun radiation obtains its crimson hue. As in the case of absorption, the

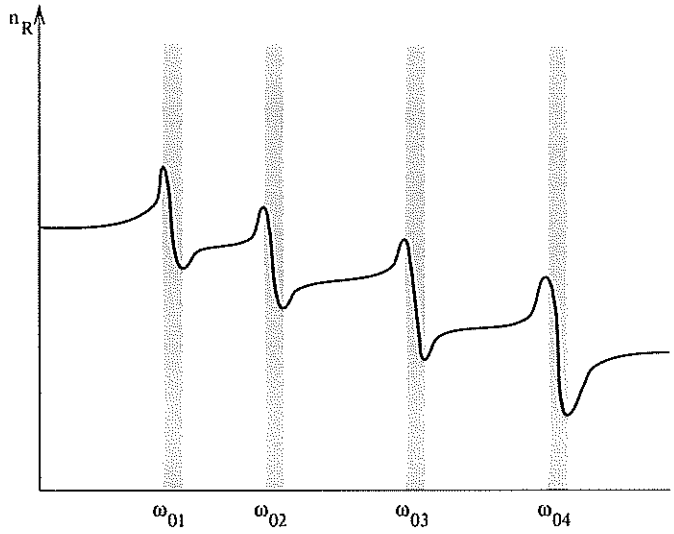


Figure 2.5: *Dependence of the index of refraction n on the angular frequency ω. The shadowed regions are absorption zones around the resonant frequencies.*

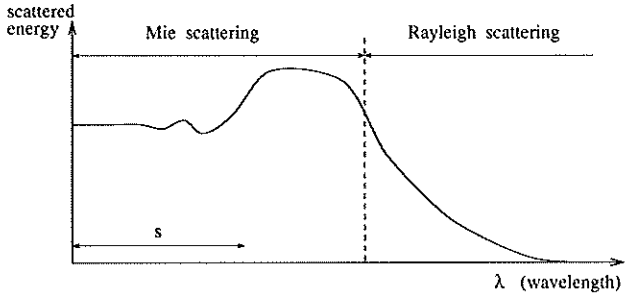


Figure 2.6: *Qualitative sketch of scattered energy as a function of wavelength. The size of the particles causing the scattering is s.*

underlying processes have to do with the interaction between the electric field and bound electrons. The induced electric dipoles oscillate at the frequency of the driving electric field. They radiate electromagnetic waves of that same frequency, i.e. they scatter the light. In dense media these secondary radiations are many and their interference will be such to obtain net propagation along the direction of the incoming wave only. In sparse media such as gases, however, the reradiation (scattering) causes the redirection of light in many directions. Returning to the example of the sky, this is the reason why it isn't simply black when we are not looking in the direction of the sun.

Scattering is often wavelength dependent. The explanation of scattering behaviour requires knowledge of the size s of the particles that cause the scattering. A typical relation between scattered energy and wavelength is given in fig. 2.6. The relative size of the particles s and the wavelength λ of the light is seen to be the determining factor, rather than λ *per se*. When the dimensions of the particles are but 1/10 of the wavelength or

less, then so-called *Rayleigh* [5] *scattering* occurs. The intensity of scattering is proportional to λ^{-4}. As a result mainly the shorter wavelengths are scattered. An example of the effect of such scattering is the blue colour of the sky. Atoms and molecules (with the exception of certain macromolecules) have sizes that give rise to Rayleigh scattering of light. If the particles are larger than 1/10 of the wavelength, one observes *Mie scattering*, of which Rayleigh scattering can be considered a limit case. The intensity of Mie scattering is only weakly dependent on wavelength. As an example, smoke and dust turn the skies over cities and industrial areas sometimes slightly reddish. The particles in a slight atmospheric haze are of the same size as the wavelengths of light. As a consequence, atmospheres with a limited visibility of a few miles say are penetrated much better by infrared light, yielding higher quality remote sensing data at these wavelengths. If the particle size is much larger than the wavelengths, then Mie scattering is completely *non-selective*, i.e. wavelength independent. An example of non-selective scattering is given by clouds, which look white, just as the sunlight they scatter. Due to these larger particles, fog is only slightly better penetrated by infrared than by visible light and virtually no improvement can be obtained. On the other hand, radar signals (SAR: Synthetic Aperture Radar) with wavelengths of several centimeters allow to "look through" clouds and fog. The use of radar is quickly gaining importance in remote sensing, since several areas are almost permanently covered by clouds (e.g. the Amazones). Although such radar signals are composed of wave with wavelengths between 1 mm to 1 m, they are normally referred to as "microwaves".

Scattering also depends on the polarization of the incoming light. As mentioned earlier, scattering is the result of reradiation by induced dipoles. All these dipoles have parallel axes and oscillate at the same frequency if they are spatially sufficiently close to experience the same electric field. The dipoles maximally radiate at right angles to their axis, and radiation drops when leaving that plane and vanishes completely in the direction that corresponds to their axis. Since unpolarized light can be modeled as the sum of two incoherent, perpendicularly polarized waves, it suffices to investigate the two polarization states. Consider fig. 2.7. For each of these two direction of polarization, there are directions in which no scattering occurs. Viewing from these directions, the light will be polarized since only one of the polarization states yields scatter. These effects can be easily verified with Polaroid sunglasses. When turning your back towards the sun and looking directly before you at the sky under a viewing angle that is perpendicular to the direction between you and the sun, the light scattered towards you will be horizontally polarized. The sunglasses have a vertical transmission axis (i.e. only vertically polarized light will pass, see chapter 4) and as a result the sky looks brighter when tilting your head.

[5] John William Strutt Rayleigh (1842-1919), not to be confused with his son physicist Robert John Strutt Rayleigh, succeeded Maxwell as Professor of Theoretical Physics at Cambridge University in 1879. One of his achievements was to predict the presence of argon gas in the atmosphere. He won the Nobel price for Physics in 1904.

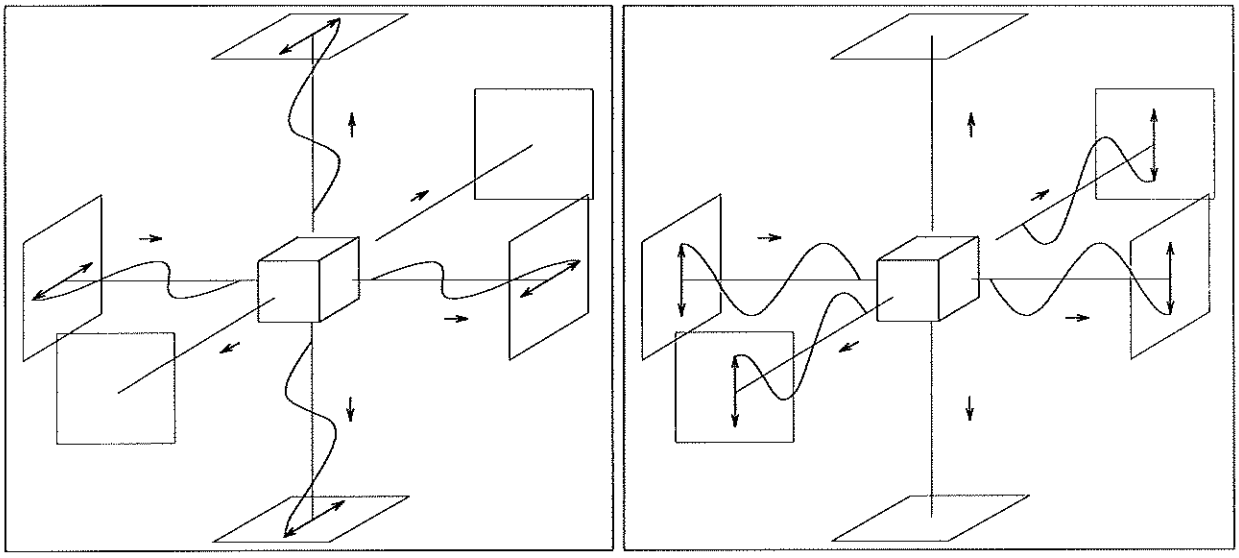


Figure 2.7: *Scattering depends on the direction of polarization (see text).*

Reflection

When light strikes a surface, its energy is not only refracted into the second medium, but some of it is reflected. The ratio of reflected power and incoming power is the *reflectance*. A matte black surface absorbs most of the light falling on it and reflects little, while a shiny surface does the opposite.

As with absorptance and transmittance, it is possible to look at different wavelengths separately. One obtains the *spectral reflectance*. The chlorophyll in a leaf absorbs the red light components more strongly, while the green components are to a higher degree reflected. So the leaf looks green.

Now suppose a plane light wave reaches a planar interface (planar at the scale of the wavelength!) between two media with different indices of refraction. At such an interface, only part of the light will be reflected, except for grazing incidence. The reflectance depends on several parameters. Several of them are illustrated in fig. 2.12. One of them is the angle θ_i that the direction of propagation of the incoming wave makes with the normal to the interface, the so-called *angle of incidence*. The direction of propagation for the reflected wave is simply the mirror direction with respect to the interface normal, i.e. the *angle of reflection* θ_r equals θ_i. Also of importance are the indices of refraction of the two media. The relation between these three parameters is expressed by the *Fresnel coefficients*, derived in section 2.2.2, which specify how much of the incoming light is reflected. Usually, computer vision deals with situation where the incoming wave is propagating through air and impinges on some other material. In that case, almost always will the index of refraction for the second medium be larger, since that of air is only 1.000293. This is assumed to be the case, i.e. this section only discusses so-called *external reflection*. The case of incidence on a material with lower index of refraction, *internal reflection*, is only discussed in section 2.2.2. Finally, the influence of the direction of polarization should be

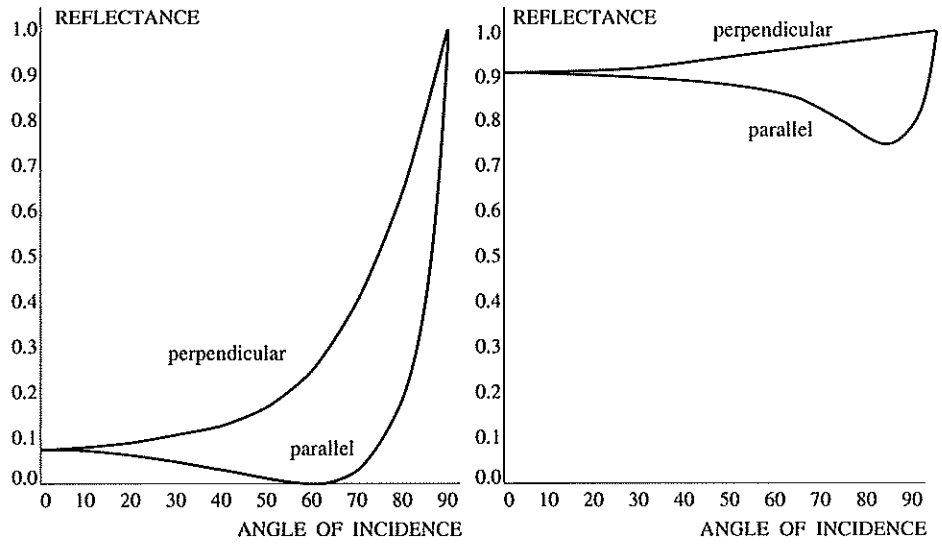
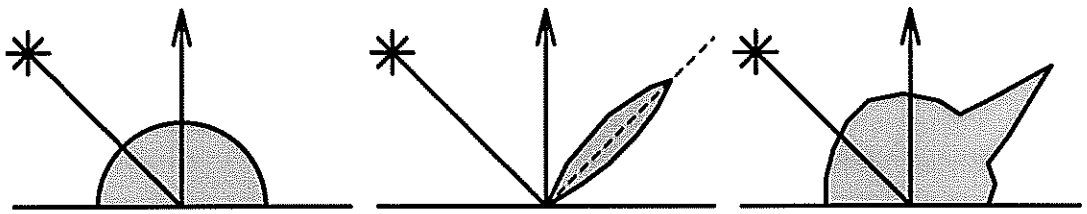


Figure 2.8: *Left: reflectance curves for a dielectric as a function of the angle of incidence for two directions of planar polarization. Right: same for a conductor (aluminium).*

mentioned. The direction of propagation of the incoming wave together with the interface normal define a plane, referred to as the *plane of incidence.* At the interface, the incoming wave can be decomposed into a component with polarization perpendicular to that plane (i.e. $\mathbf{E}_{i\perp}$ in fig. 2.12), and a component parallel to the plane ($\mathbf{E}_{i\parallel}$ in fig. 2.12). The reflectance for these components will differ. Fig. 2.8 illustrates the reflectance as a function of the angle of incidence and the polarization direction for two materials. On the left, the reflectance is shown for glass (a dielectric medium with an index of refraction of about 1.5). On the right, the reflectance is given for aluminium (example of a conductor). As one can see, the distinction between dielectrics and conductors is very evident, as already suggested by section 2.1.6 and corroborated by section 2.2.2. Note that for dielectrics there is an angle of incidence, called the *Brewster angle*, for which the parallel component is not reflected. The Brewster angle is typically close to 60°.

The type of reflection discussed so far, is *specular reflection* and is illustrated in fig. 2.9(b). This occurs when the surface is smooth in optical terms (deviations from planarity have to be smaller than the wavelength). Mirrors are the ultimate example of this, but metallic surfaces will in general also have a strong specular reflection.
Alternatively, a macroscopically planar interface between materials may be rather rough when viewed at the scale of optical wavelengths. The surface can then be modeled as a multitude of small, planar facets oriented differently, each yielding specular reflection according to their own normal orientations. Light will then also be reflected in different directions. The prototypical case of such *diffuse* reflection is the so-called Lambertian surface [6] where light is reflected in all directions with relative strengths given by the cosine

[6] After Johann Heinrich Lambert (1728-1777), German mathematician and physicist of French origin. *Photometria* (1760) was probably his most important contribution to physics. Also noteworthy is the onset



(a) (b) (c)

Figure 2.9: *Polar diagrams of perceived surface brightness for a given angle of light incidence: (a) purely diffuse reflection (Lambertian surface), (b) specular reflection, (c) mixed type.*

of the angle made with the surface normal. A Lambertian surface therefore has orientation independent radiance (the cosines of the reflectance and area foreshortening compensate) and the observed irradiance will be proportional to the cosine of the angle between the direction of illumination and the surface normal. Most real surfaces will show a mixture of these reflection types, as illustrated in fig. 2.9.

Refraction

When rays of light pass from a medium of one index of refraction to a medium of a different index of refraction, they are bent, or refracted, unless they strike the boundary between the two media perpendicularly. Refraction occurs, for example, at the boundary between air and water. If somebody is standing in the water, the part of the legs in the water will seem to be displaced with respect to the part above the water surface.

As in the case of reflection, the amount of light that is refracted into the second medium does not only depend on the index of refraction, but also on the angle of incidence and the polarization. In fact, the proportion of refracted irradiance can be derived easily as one minus the reflectance, which was given in fig. 2.8. Assuming that the interface between two media is optically smooth, the direction of refraction θ_t (see also fig. 2.12) with respect to the interface normal is given by *Snell's law*:

$$n_1 \sin\theta_i = n_2 \sin\theta_t \ ,$$

with n_1 the index of refraction of the medium left by the wave, and n_2 the index of refraction of the medium entered. Note again that usually the former will be air and hence $n_1 \simeq 1$. This law is derived in section 2.2.2. Because the second medium will normally have a larger index of refraction, θ_t will typically be smaller than θ_i. The index of refraction will typically depend on wavelength and therefore also the angles of refraction (and reflection) will be wavelength dependent. This explains the origin of the index's name. Note the important qualitative difference with reflection: there the redirection of light is wavelength independent.

to non-Euclidean geometry in his work. Lagrange and Gauss were inspired by him.

2.2.2 * Interaction with matter – a mathematical account

The index of refraction for dielectrics

In eq. 2.3 the influence of **E** on free conduction electrons was included through the conductivity. However, also bound electrons undergo the influence of **E**. As will be explained, both these effects of **E** are responsible for absorption phenomena. **E**'s interaction with bound electrons might lead to the *polarization* of the material (not to be confused with the polarization of light!). Electrons can move away from their equilibrium positions, thereby creating electric dipole moments (charge × distance between the poles), referred to as *induced dipoles.* Sometimes a material also contains permanent dipoles (e.g. water molecules). **E** will orient these dipoles parallel to itself. As a consequence, the effective field **E** will be changed. This difference between the field that would exist in vacuum (with permittivity ϵ_0) and the resulting field, was actually captured by introducing a different permittivity ϵ for the medium. Indeed, for a resulting polarization field **P**, ϵ is such that

$$\epsilon \mathbf{E} = \epsilon_0 \mathbf{E} + \mathbf{P} \ ,$$

and its introduction is equivalent to the addition of the polarization in eq. 2.3, when it is assumed that the polarization field is simply found by scaling the electric field. A simple electrostatic determination of ϵ as $K\epsilon_0$ with K the medium's dielectric constant K will only be correct for low frequencies, as already mentioned in section 2.2.1. A more sophisticated analysis is needed.

Considering the electrons with their charge $-e$ and displacement **r** from equilibrium, the resulting macroscopic polarization **P** with N electrons per unit volume is

$$\mathbf{P} = -Ne\mathbf{r} \ .$$

The bound electrons will approximately behave as damped harmonic oscillators, with

$$m\frac{d^2\mathbf{r}}{dt^2} + m\gamma\frac{d\mathbf{r}}{dt} + L\mathbf{r} = -e\mathbf{E}$$

as the equation of their motion. In this equation, m is the mass of the electron, the second term represents frictional damping, and the third term stands for the elastic return force. Note that the influence of the magnetic field on the moving electron is neglected since it typically is much smaller than $e\mathbf{E}$. Now, if the field **E** varies harmonically with time by a factor $e^{-i\omega t}$, then

$$\left(-m\omega^2 - i\omega m\gamma + L\right)\mathbf{r} = -e\mathbf{E} \ .$$

Hence, introducing the *resonance frequency* $\omega_0 = \sqrt{L/m}$

$$\mathbf{P} = \frac{Ne^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}\mathbf{E} \ . \tag{2.11}$$

Note the phase reversal beyond the resonant frequency. Around the resonant frequency, the dipole oscillations get large and dissipation due to work by the field on the charges

increases. As presently shown, close to this frequency there is an absorption band for which the medium is fairly opaque.

In section 2.1.6 it was shown that, in the absence of conduction electrons, light waves must satisfy a simple wave equation and therefore can be considered superpositions of harmonic waves. Here, a more detailed analysis shows that at least around the resonant frequencies, ϵ has to be considered a complex number with a non-negligible imaginary component, and therefore, as in the case of conductors, an extinction factor appears. Using eq. (2.11) the wave equation can be rewritten

$$\begin{aligned}
\nabla^2\mathbf{E} &= \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} \\
&= \mu\left(\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2}+\frac{\partial^2\mathbf{P}}{\partial t^2}\right) \\
&= \mu\left(\epsilon_0+\frac{Ne^2/m}{\omega_0^2-\omega^2-i\omega\gamma}\right)\frac{\partial^2\mathbf{E}}{\partial t^2} \\
&\simeq \frac{1}{c_0^2}\left(1+\frac{Ne^2/m\epsilon_0}{\omega_0^2-\omega^2-i\omega\gamma}\right)\frac{\partial^2\mathbf{E}}{\partial t^2}\ ,
\end{aligned} \tag{2.12}$$

using the fact that except for ferromagnetic materials $\mu \simeq \mu_0$. In analogy with earlier analysis, a complex wavenumber $\mathbf{k} = \mathbf{k}_R + i\mathbf{k}_I$ is introduced, such that

$$\mathbf{k}\cdot\mathbf{k} = (\omega\mu\epsilon)^2 = \frac{\omega^2}{c_0^2}\left(1+\frac{Ne^2/m\epsilon_0}{\omega_0^2-\omega^2-i\omega\gamma}\right)\ .$$

Note that $\mathbf{k}_R$ and $\mathbf{k}_I$ are not necessarily parallel vectors. A possible solution for eq. 2.12 is then

$$\mathbf{E} = \mathbf{E}_0 e^{i((\mathbf{k}_R+i\mathbf{k}_I)\cdot\mathbf{X}-\omega t)} = \mathbf{E}_0 e^{-\mathbf{k}_I\cdot\mathbf{X}} e^{i(\mathbf{k}_R\cdot\mathbf{X}-\omega t)}\ .$$

Of course, a general light wave would consist of a superposition of such waves. Notice the absorption factor that comes with the imaginary component. Since the energy in the wave at a given point is proportional to $\|\mathbf{E}\|^2$, $2\|\mathbf{k}_I\|$ is the *coefficient of absorption.* The speed of the wave is still determined by $\|\mathbf{k}_R\|$ and ω. As can be seen from eq. (2.12), for frequencies sufficiently far from the resonant frequency ω_0 the real part of ϵ will be dominant, but close to this frequency, the imaginary part and hence also the extinction of the wave are the largest. In analogy with the pure harmonic case, one defines the the complex index of refraction as

$$n = \frac{c_0}{\omega}\|\mathbf{k}_R\| + i\frac{c_0}{\omega}\|\mathbf{k}_I\| = \frac{c_0}{c} + i\frac{c_0}{\omega}\|\mathbf{k}_I\|\ .$$

The real part still represents the speed of the wave in vacuum over the speed in the actual medium.

In reality, different electrons may be bound in different ways. Instead of having a single resonant frequency, each type of bounding will introduce one of its own. This leads to an

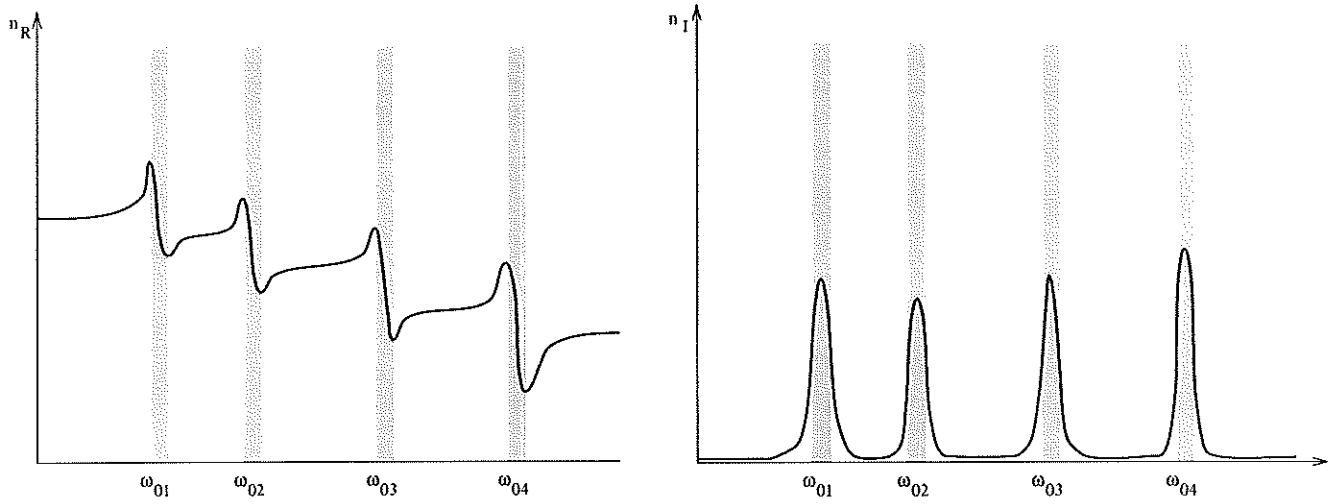


Figure 2.10: *Dependence of the real (left) and imaginary (right) components of the index of refraction n on the angular frequency ω. The shadowed regions are absorption zones around the resonant frequencies.*

expression for the squared index of refraction n^2 of the form

$$n^2 = \frac{\epsilon\mu}{\epsilon_0\mu_0} \simeq \frac{\epsilon}{\epsilon_0} = 1 + \frac{Ne^2}{m\epsilon_0} \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 - i\gamma_j\omega} \ ,$$

with the summation taken over the different bounding types j, with f_j the fraction of electrons corresponding to each type, and again using the fact that except for ferromagnetic materials $\mu \simeq \mu_0$. The real and imaginary components n_R and n_I are shown as a function of frequency in fig. 2.10. The real component was already shown in section 2.2.1 but is shown again for ease of reference. As already noted, $n_R(\omega)$ is the so-called *dispersion curve.* Note how the absorption bands correspond to maxima of n_I.

In summary, the first term on the right-hand side in Ampère's equation (2.3) should be understood to also take the effect of polarization into account. The foregoing analysis has shown that a number of practically important effects find their origin in this phenomenon. Examples are the different speeds for waves of different wavelengths and the existence of absorption bands.

The index of refraction for conductors

In the case of a good conductor, there are conduction (i.e. unbound) electrons. The resulting current density J cannot be simply determined as $\mathbf{J} = \sigma\mathbf{E}$, with σ the static conductivity. Again, as in the case of dielectrics, the time dynamic nature of the electric field has to be taken account of. The following analysis shows that this refinement amounts to making σ complex.

Since the conduction electrons are unbound, the elastic restoring force considered for the bound electrons in the previous case can be removed from the electron's equation of

motion:

$$m\frac{d\mathbf{v}}{dt} + f\mathbf{v} = -e\mathbf{E}$$

where $\mathbf{v}$ stands for the electron's velocity, m its mass, $-e$ its charge, and f a constant that represents frictional dissipation associated with the motion. At a global scale, with N the number of *conduction* electrons per unit volume, the current density $\mathbf{J} = -Ne\mathbf{v}$ satisfies

$$m\frac{d\mathbf{J}}{dt} + f\mathbf{J} = Ne^2\mathbf{E} \ .$$

If a time dependence $e^{-i\omega t}$ is assumed for both $\mathbf{J}$ and $\mathbf{E}$, then

$$(-i\omega m + f)\ \mathbf{J} = Ne^2\mathbf{E}$$

and hence,

$$\mathbf{J} = \frac{Ne^2}{f - i\omega m}\mathbf{E} \ . \tag{2.13}$$

Since for $\omega = 0$ we know that $\mathbf{J} = \sigma_{static}\mathbf{E}$, it follows that $\sigma_{static} = Ne^2/f$. The bottomline is that in eqs. (2.8) and (2.9) σ has to be given the complex value of eq. (2.13).

Taking

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{X}-\omega t)} = \mathbf{E}_0 e^{-\mathbf{k}_I\cdot\mathbf{X}} e^{i(\mathbf{k}_R\cdot\mathbf{X}-\omega t)} \ .$$

with $\mathbf{k}$ complex for a trial solution, it is found that $\|\mathbf{k}\|$ must satisfy

$$\mathbf{k}\cdot\mathbf{k} = \epsilon\mu\omega^2 + \frac{i\omega\mu Ne^2}{f - i\omega m} \ .$$

As in the case of conductors, the result is also a complex index of refraction n

$$n = n_R + in_I = \frac{c_0}{\omega}\left(\|\mathbf{k}_R\| + i\|\mathbf{k}_I\|\right) \ .$$

The derivation of explicit, algebraic expressions for n_R and n_I is rather cumbersome, but fig. 2.11 gives a qualitative sketch of the dependence of n_R and n_I on the frequency of the wave. Near to what is called the *plasma frequency* ω_p both components will be very close in value for typical metals. Below this frequency the behaviour can be described as highly opaque and strongly reflecting. Above this frequency a metal will become increasingly transparent.

It is interesting to note that in the foregoing analysis also the absorption by bound charges can be brought to bear, by again considering ϵ to be complex. This model is the most complete one and is especially relevant for poor conductors and semiconductors. In that case both terms for unbound and bound charges will contribute to the expression for the complex index of refraction with squared value

$$n^2 = 1 - \frac{Ne^2}{m\epsilon_0\omega^2 + i\epsilon_0 f\omega} + \frac{Ne^2}{m\epsilon_0}\sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 - i\gamma_j\omega} \ .$$

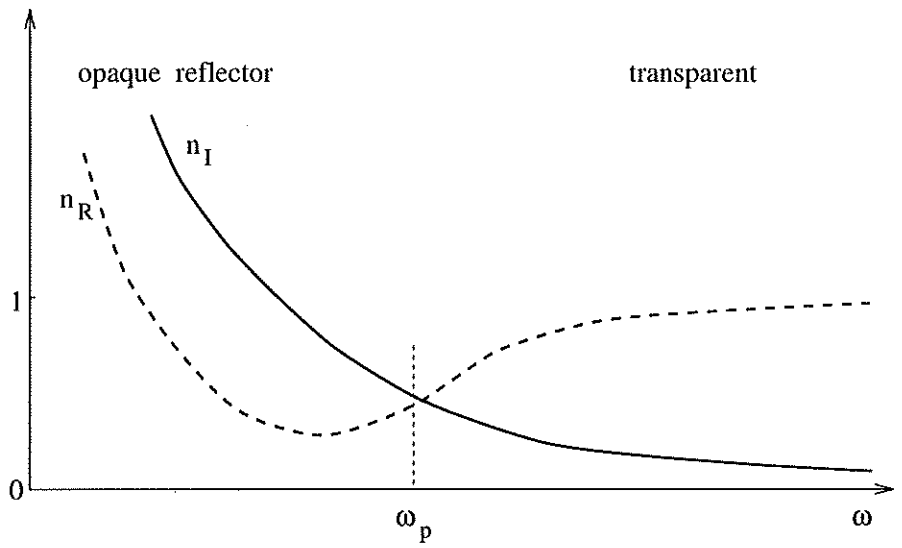


Figure 2.11: *Real (n_R) and imagninary (n_I) components of the index of refraction for a typical metal.*

Reflection and refraction at a boundary between dielectrics

In fig. 2.8 a qualitative discussion was given of how light is reflected at a planar boundary between two dielectrics, depending on its state of polarization and angle of incidence. Here, these results will be given a theoretical underpinning. Consider fig. 2.12. A planar boundary separates two dielectric media with refraction indices n_1 and n_2. An incoming light wave is seen to be incident from the left ($\mathbf{E}_i$) and its propagation vector $\mathbf{k}_i$ makes an angle θ_i with the normal $\mathbf{n}$ of the boundary plane. The vectors $\mathbf{k}_i$ and $\mathbf{n}$ together define a plane that is called the *plane of incidence*. Part of the wave is reflected back into the first medium ($\mathbf{E}_r$) and part is refracted into the second medium ($\mathbf{E}_t$). For the sake of the further analysis, the electric field vectors of the different waves are decomposed into a component within the plane defined by the normal to the boundary and their direction of propagation (subscript $\|$) and perpendicular to that plane (subscript $\perp$). In the sequel, the relation between these vectors at the boundary will be considered.

As usual by now, we restrict attention to the harmonic wave type

$$\mathbf{E}_i = \mathbf{E}_{0i} e^{i(\mathbf{k}_i \cdot \mathbf{X} - \omega_i t)} \ .$$

In fact, there is no advantage in using the complex form for the subsequent discussion, and therefore the real component will be considered only:

$$\mathbf{E}_i = \mathbf{E}_{0i} \cos\left(\mathbf{k}_i \cdot \mathbf{X} - \omega_i t + \epsilon_i\right) \ .$$

Notice that a non-zero phase ϵ_i has been introduced as well. The reason is that the origin won't be chosen arbitrarily this time, but at the point of incidence at the boundary. At the point where the wave hits the surface, a reflected and a transmitted (refracted) wave

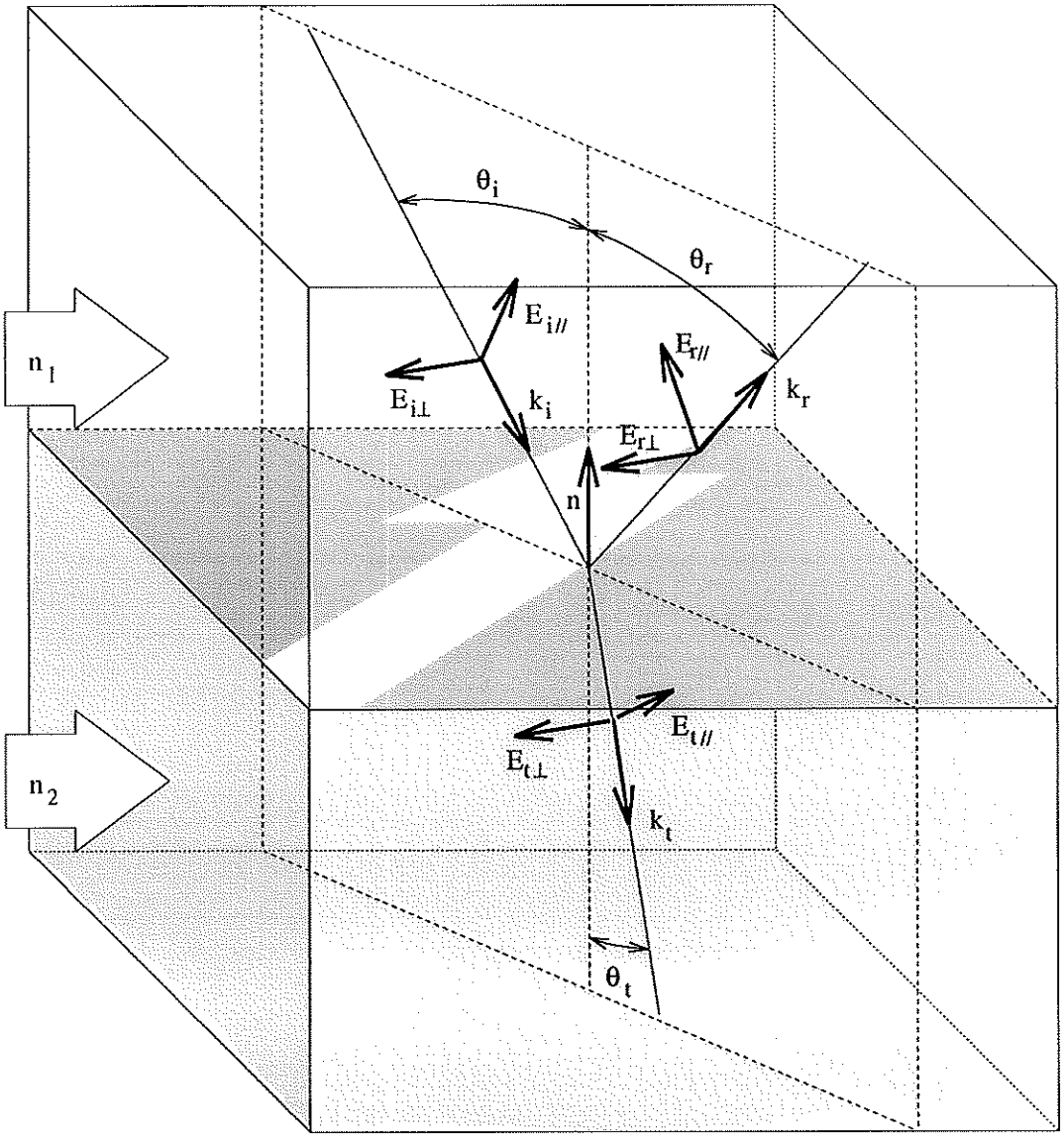


Figure 2.12: *Planar boundary between two media with refraction indices n_1 and n_2. The electric fields $\mathbf{E}_i$ of the incoming light, $\mathbf{E}_r$ of the reflected light, and $\mathbf{E}_t$ of the refracted light are decomposed into two linearly polarized components. One component lies in the plane of incidence, the other is perpendicular to this plane.*

are initiated with equations

$$\begin{aligned} \mathbf{E}_r &= \mathbf{E}_{0r} \cos\left(\mathbf{k}_r \cdot \mathbf{X} - \omega_r t + \epsilon_r\right) \ , \\ \mathbf{E}_t &= \mathbf{E}_{0t} \cos\left(\mathbf{k}_t \cdot \mathbf{X} - \omega_t t + \epsilon_t\right) \ . \end{aligned}$$

These fields are not independent, but should obey a number of *boundary conditions*. First consider a cylinder with its axis through the point of incidence and parallel to $\mathbf{n}$. Imagine that the axis of the cylinder is shortened and that its top and bottom planes approach the boundary from both sides. Then, eq. 2.6 shows that in the limit, the flux of $\epsilon\mathbf{E}$ through top and bottom have to be equal and therefore, the net component of $\epsilon\mathbf{E}$ perpendicular to the boundary has to be the same on both sides. Note that ϵ can not simply be neglected here, since the two media will typically have different values for ϵ. Similarly, one can think of a loop that encloses the point of incidence, runs along a straight path through one medium and closes through the other following a parallel path in the opposite direction. Again bringing the two paths very close to the boundary, and making the links passing through the boundary on eiter side very short, eq. 2.2 implies that the net component of $\mathbf{E}$ along any direction parallel to the boundary has to be the same on either side. Similar conclusions can be drawn for $\mathbf{H}$ from eqs. (2.3) and (2.7). The tangential component of $\mathbf{H}$ is continuous when crossing the boundary, as well as the vertical component of $\mu\mathbf{H}$.

Given the forementioned boundary conditions and the considered form of the waves, it is immediately clear that the conditions can only be satisfied at all times if

$$\begin{aligned} \omega_i &= \omega_r = \omega_t \ , \\ \epsilon_i &= \epsilon_r = \epsilon_t \ . \end{aligned}$$

and also at every point of the boundary if

$$\mathbf{k}_i \cdot \mathbf{X} = \mathbf{k}_r \cdot \mathbf{X} = \mathbf{k}_t \cdot \mathbf{X} \tag{2.14}$$

irrespective of the choice of the point $\mathbf{X}$ on the boundary. From the latter equations it immediately follows that for all $\mathbf{X}$ of the boundary

$$(\mathbf{k}_i - \mathbf{k}_r) \cdot \mathbf{X} = (\mathbf{k}_i - \mathbf{k}_t) \cdot \mathbf{X} = 0$$

and therefore $\mathbf{k}_i$, $\mathbf{k}_r$, and $\mathbf{k}_t$ are coplanar and the plane they define is perpendicular to the boundary. These facts are already taken into account in fig. 2.12 as it is. Hence, the planes defined by the boundary normal and each of the directions of propagation are one and the same and all perpendicular components of the electric field vectors are parallel.

With the constraints of the previous paragraph established, the three angles θ_i, θ_r, and θ_t are considered. Using $c = \omega / \|\mathbf{k}\|$, eqs. 2.14 implies that

$$\|\mathbf{k}_i\| \sin\theta_i = \frac{\omega}{c_i} \sin\theta_i = \frac{\omega}{c_r} \sin\theta_r = \|\mathbf{k}_r\| \sin\theta_r \ .$$

Moreover, since the incident and the reflected wave propagate through the same medium, $c_i = c_r$ and therefore

$$\theta_i = \theta_r \ ,$$

which is the familiar law of specular reflection. Similarly,

$$\|\mathbf{k}_i\| \sin\theta_i = \frac{\omega}{c_i} \sin\theta_i = \frac{\omega}{c_t} \sin\theta_t = \|\mathbf{k}_t\| \sin\theta_t$$

implies Snell's law

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_2}{n_1} \ .$$

First, the **E**-components parallel to the plane of incidence are considered. The boundary conditions yield

$$\begin{array}{rcrcl} \|\mathbf{E}_{i\|}\| \cos\theta_i & - & \|\mathbf{E}_{r\|}\| \cos\theta_i & = & \|\mathbf{E}_{t\|}\| \cos\theta_t \\ \|\mathbf{E}_{i\|}\| \sin\theta_i \epsilon_1 & + & \|\mathbf{E}_{r\|}\| \sin\theta_i \epsilon_1 & = & \|\mathbf{E}_{t\|}\| \sin\theta_t \epsilon_2 \end{array}$$

as constraints. Using

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_2}{n_1} \simeq \sqrt{\frac{\epsilon_2}{\epsilon_1}}$$

one finds that

$$\begin{aligned} \frac{\|\mathbf{E}_{r\|}\|}{\|\mathbf{E}_{i\|}\|} &= \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} \ , \\ &= \frac{(n_2/n_1)^2 \cos\theta_i - \sqrt{(n_2/n_1)^2 - \sin^2\theta_i}}{(n_2/n_1)^2 \cos\theta_i + \sqrt{(n_2/n_1)^2 - \sin^2\theta_i}} \ , \\ &= \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \ . \end{aligned}$$

The components of **E** perpendicular to the plane of incidence are now analysed. The boundary constraints on **E** only seem to imply that

$$\|\mathbf{E}_{i\perp}\| + \|\mathbf{E}_{r\perp}\| = \|\mathbf{E}_{t\perp}\| \ .$$

However, since the amplitudes of **E** and **H** are coupled, the boundary conditions on **H** actually restrict the consistency of the **E**s as well. The directions of the **H**s are found by considering that **k**, **E**, and **H** form a right-handed system as shown in fig. 2.1. Expressing the continuity of the tangent component of **H** and reformulating the constraint in terms of the amplitude of **E** one finds

$$\|\mathbf{E}_{i\perp}\| \sqrt{\frac{\epsilon_1}{\mu_1}} \cos\theta_i - \|\mathbf{E}_{r\perp}\| \sqrt{\frac{\epsilon_1}{\mu_1}} \cos\theta_r = \|\mathbf{E}_{t\perp}\| \sqrt{\frac{\epsilon_2}{\mu_2}} \cos\theta_t \ .$$

Again using the approximation $n_2/n_1 \simeq \sqrt{\frac{\epsilon_2}{\epsilon_1}}$ it now follows that

$$\begin{aligned} \frac{\|\mathbf{E}_{r\perp}\|}{\|\mathbf{E}_{i\perp}\|} &= \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} \ , \\ &= \frac{(n_1/n_2) \cos\theta_i - \sqrt{1 - (n_1/n_2)^2 \sin^2\theta_i}}{(n_1/n_2) \cos\theta_i + \sqrt{1 - (n_1/n_2)^2 \sin^2\theta_i}} \ , \\ &= -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \ . \end{aligned}$$

Note that the ratio of reflected over incident power per unit area of the boundary is obtained by squaring the expressions for the amplitude ratios $\|\mathbf{E}_r\|/\|\mathbf{E}_i\|$. This follows from the results for the Poynting vector in section 2.1.2, the equality $\theta_i = \theta_r$ of the angles of incidence and reflection and the fact that the incident and reflected waves propagate through the same material and therefore are determined by the same permittivity and permeability. This ratio of powers has been drawn in the left part of fig. 2.8 for the case where $n_1 < n_2$ (so-called *external reflection*). This figure should be complemented with the case where $n_1 > n_2$ (so-called *internal reflection*), i.e. when the light is incident from a medium with higher optical density. The major reason for not simply discarding this case in this course is its relevance for optical fibers, where the central part consists of material of higher refraction index than the surrounding layers or cladding, and compound lenses. As a consequence of Snell's law $\theta_i > \theta_t$ if $n_i < n_t$. Therefore, θ_t can never reach 90°. If $n_i > n_t$, however, θ_t will reach 90° before θ_i does, at the *critical angle* for which $\sin\theta_i = n_2/n_1$. For glasses with their refraction index around 1.5, this critical angle is about 40°. Beyond this angle

$$\frac{\|\mathbf{E}_{r\perp}\|}{\|\mathbf{E}_{i\perp}\|} = \frac{\cos\theta_i - i\sqrt{\sin^2\theta_i - (n_2/n_1)^2}}{\cos\theta_i + i\sqrt{\sin^2\theta_i - (n_2/n_1)^2}},$$

$$\frac{\|\mathbf{E}_{r\|}\|}{\|\mathbf{E}_{i\|}\|} = \frac{(n_2/n_1)^2\cos\theta_i - i\sqrt{\sin^2\theta_i - (n_2/n_1)^2}}{(n_2/n_1)^2\cos\theta_i + i\sqrt{\sin^2\theta_i - (n_2/n_1)^2}},$$

and muliplication with the complex conjugate values immediately shows that both complex numbers have magnitude 1. Hence, beyond the critical angle, all the light is reflected, as shown in fig. 2.13.

The ratio of refracted over incident power is found as

$$\frac{n_2\cos\theta_t}{n_1\cos\theta_i}\left(\frac{E_t}{E_i}\right)^2$$

where again the assumption $\mu_1 \simeq \mu_2$ was made.

The foregoing discussion is mainly important for its explanation of the qualitative results found in practical experiments. However, these results are theoretical idealizations that might differ substantially from experimental results in quantitative terms. This is due to impurity of the materials, surface finishing operations, etc.

Reflection and refraction at a dielectric/metal boundary

Not available yet.

2.3 * Further reading

A good starting point for the student interested in phenomena of light and optics is the book by Falk *et al.* [2]. This book is narrative and non-mathematical in flavour, but contains

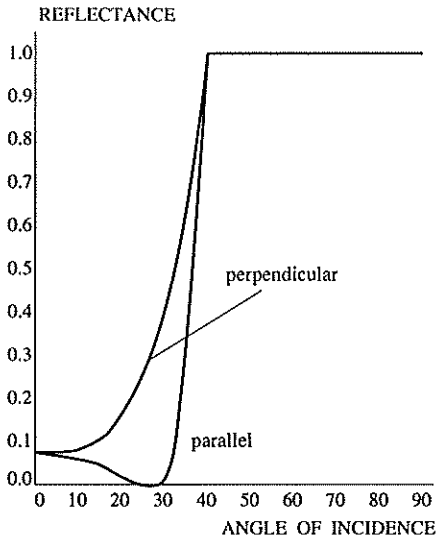


Figure 2.13: *Reflectance curves for a dielectric in the case of internal reflection, as a function of the angle of incidence and for polarization directions parallel and perpendicular to the plane of incidence. This figure complements the left part of fig. 2.8, which shows the results for the more usual case of external reflection.*

a wealth of illustrations and examples from daily experience. The book by Hecht [4] is a good alternative for those who look for a more thorough introduction. It is self-contained and covers a broad gamut of subjects. It has marvelous illustrations and doesn't shy away from the mathematical models needed to explain the different effects.
Books that go somewhat deeper into mathematical issues are e.g. Fowles [3] and Matveev [5]. Especially the former gives concise and elegant derivations.
A detailed discussion of radiometric and photometric units and measurements is provided in [3].
Those with a keen interest in the modeling of diffuse reflection will find a detailed discussion in Beckmann and Spizzochino [1], but could also consult books in the realm of computer graphics, where references to refined mathematical models will be found.

Bibliography

[1] P. Beckmann and A. Spizzochino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Pegamon, NY, 1963

[2] D. Falk, D. Brill, and D. Stork, *Seeing the Light, optics in nature, photography, color, vision, and holography*, John Wiley & Sons, NY, 1986

[3] G. Fowles, *Introduction to modern optics*, Dover, 1989

[4] E. Hecht, *Optics*, Addison-Wesley, 1987

[5] A. Matveev, *Optics*, Mir Publishers, Moscow, 1988

[6] R. McCluney, *Introduction to Radiometry and Photometry*, Artech House, Boston, 1994