Chapter 7. Electromagnetic Wave Propagation

This (rather extensive) chapter focuses on the most important effect that follows from the time-dependent Maxwell equations, namely the electromagnetic waves, at this stage avoiding the issue of their origin, i.e. of the wave radiation process – which will be the subject of Chapters 8 and 10. We will start from the simplest, plane waves in uniform and isotropic media, and then proceed to a discussion of nonuniform systems, bringing up such effects as reflection and refraction. Then we will discuss the so-called guided waves, propagating along various transmission lines – such as cables, waveguides, and optical fibers. Finally, the end of the chapter is devoted to final-length fragments of such lines, serving as resonant cavities, and to the effects of energy dissipation in transmission lines and cavities.

7.1. Plane waves

Let us start by considering a spatial region that does not contain field sources ($\rho = 0$, $\mathbf{j} = 0$), and is filled with a uniform, isotropic, linear medium, which therefore obeys Eqs. (3.46) and (5.110):

$$\mathbf{D} = \varepsilon\mathbf{E}, \qquad \mathbf{B} = \mu\mathbf{H}. \tag{7.1}$$

Moreover, let us assume for a while that these constitutive equations hold for all frequencies of interest. (Of course, these relations are *exactly* valid for the very important particular case of free space, where we may formally use the macroscopic Maxwell equations (6.100), but with $\varepsilon = \varepsilon_0$ and $\mu = \mu_0$.) As was already shown in Sec. 6.8, in this case, the Lorenz gauge condition (6.117) allows the Maxwell equations to be recast into the wave equations (6.118) for the scalar and vector potentials. However, for most purposes, it is more convenient to use the homogeneous Maxwell equations (6.100) for the electric and magnetic fields – which are independent of the gauge choice. After an elementary elimination of **D** and **B** using Eqs. (1),[1] these equations take a simple, very symmetric form:

Maxwell equations for uniform linear media

$$\nabla\times\mathbf{E} + \mu\frac{\partial\mathbf{H}}{\partial t} = 0, \qquad \nabla\times\mathbf{H} - \varepsilon\frac{\partial\mathbf{E}}{\partial t} = 0, \tag{7.2a}$$

$$\nabla\cdot\mathbf{E} = 0, \qquad \nabla\cdot\mathbf{H} = 0. \tag{7.2b}$$

Now, let us act by the operator $\nabla\times$ on each of Eqs. (2a), i.e. take their curl, and then use the vector algebra identity (5.31). The appearing terms $\nabla\cdot\mathbf{E}$ and $\nabla\cdot\mathbf{H}$ vanish due to Eqs. (2b), so the first terms of Eqs. (2a) turn into the Laplace operators of these vectors (with the minus sign). Now swapping, in the second terms, the operators $\partial/\partial t$ and $\nabla\times$, and using Eqs. (2a) again, we get fully similar wave equations for the electric and magnetic fields:[2]

EM wave equations

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{E} = 0, \qquad \left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{H} = 0, \tag{7.3}$$

[1] Though in a medium, **B** rather than **H** is the actual macroscopic magnetic field, mathematically it is a bit more convenient (just as it was in Sec. 6.2) to use the vector pair {**E**, **H**} in the following discussion, because at sharp media boundaries, it is **H** that obeys the boundary condition (5.117) similar to that for **E** – cf. Eq. (3.37).

[2] The two vector equations (3) are of course just a shorthand for six similar equations for three Cartesian components of **E** and **H**.

where the parameter v is defined as

$$v^2 \equiv \frac{1}{\varepsilon\mu}. \tag{7.4}$$

Wave velocity

with $v^2 = 1/\varepsilon_0\mu_0 \equiv c^2$ in free space – see Eq. (6.120) again.

These equations allow, in particular, solutions of the following type;

$$E \propto H \propto f(z - vt), \tag{7.5}$$

Plane wave

where z is the Cartesian coordinate along a certain (*arbitrary*) direction **n**, and f is an *arbitrary* function of one argument. Note that this solution, first of all, describes a *traveling wave* – meaning a certain field pattern moving, without deformation, along the z-axis, with the constant velocity v. Second, according to Eq. (5), both **E** and **H** have the same values at all points of each plane perpendicular to the direction **n** ≡ $\mathbf{n}_z$ of the wave propagation; hence the second name – *plane wave*.

According to Eqs. (2), the independence of the wave *equations* (3) for vectors **E** and **H** does not mean that their plane-wave *solutions* are independent. Indeed, plugging any solution of the type (5) into Eqs. (2a), we get

$$\mathbf{H} = \frac{\mathbf{n}\times\mathbf{E}}{Z}, \qquad \text{i.e. } \mathbf{E} = Z\,\mathbf{H}\times\mathbf{n}, \tag{7.6}$$

Field vector relation

where

$$Z \equiv \frac{E}{H} = \left(\frac{\mu}{\varepsilon}\right)^{1/2}. \tag{7.7}$$

Wave impedance

The vector relationship (6) means, first of all, that at any point of space and at any time instant, the vectors **E** and **H** are perpendicular not only to the propagation vector **n** (such waves are called *transverse*) but also to each other — see Fig. 1.

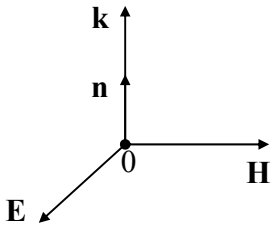


Fig. 7.1. Field vectors in a plane electromagnetic wave propagating along direction **n**.

Second, this equality does not depend on the function f, meaning that the electric and magnetic fields increase and decrease *simultaneously*. Finally, the field magnitudes are related by the constant Z called the *wave impedance* of the medium. Very soon we will see that this impedance plays a pivotal role in many problems, in particular at the wave reflection from the interface between two media. Since the dimensionality of E, in SI units, is V/m, and that of H is A/m, Eq. (7) shows that Z has the dimensionality of V/A, i.e. ohms (Ω).[3] In particular, in free space,

[3] In the Gaussian units, E and H have a similar dimensionality (in particular, in a free-space wave, $E = H$), making the (very useful) notion of the wave impedance less manifestly exposed – so in some older physics textbooks it is not mentioned at all!

Wave impedance of free space

$$Z = Z_0 \equiv \left(\frac{\mu_0}{\varepsilon_0}\right)^{1/2} = 4\pi \times 10^{-7} c \approx 377\ \Omega\,. \tag{7.8}$$

Next, plugging Eq. (6) into Eqs. (6.113) and (6.114), we get:

Wave's energy

$$u = \varepsilon E^2 = \mu H^2\,, \tag{7.9a}$$

Wave's power

$$\mathbf{S} \equiv \mathbf{E} \times \mathbf{H} = \mathbf{n}\frac{E^2}{Z} = \mathbf{n} Z H^2\,, \tag{7.9b}$$

so, according to Eqs. (4) and (7), the wave's energy and power densities are universally related as

$$\mathbf{S} = \mathbf{n} u v\,. \tag{7.9c}$$

In view of the Poynting vector paradox discussed in Sec. 6.8 (see Fig. 6.11), one may wonder whether the last equality may be interpreted as the actual density of power flow. In contrast to the static situation shown in Fig. 6.11, which limits the electric and magnetic fields to the vicinity of their sources, waves may travel far from them. As a result, they can form *wave packets* of a finite length in free space – see Fig. 2.

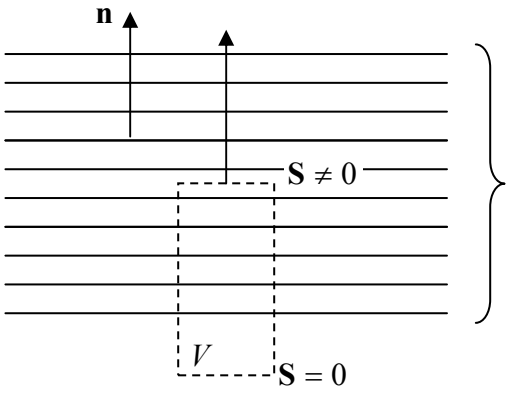


Fig. 7.2. Interpreting the Poynting vector in a plane electromagnetic wave. (Horizontal lines show equal-field planes.)

Let us apply the Poynting theorem (6.111) to the cylinder shown with dashed lines in Fig. 2, with one lid inside the wave packet, and another lid in the region already passed by the wave. Then, according to Eq. (6.111), the rate of change of the full field energy $\mathscr{E}$ inside the volume is $d\mathscr{E}/dt = -SA$ (where A is the lid area), so S may be indeed interpreted as the power flow (per unit area) from the volume. Making a reasonable assumption that the finite length of a sufficiently long wave packet does not affect the physics inside it, we may indeed interpret the **S** given by Eqs. (9b-c) as the power flow density inside a plane electromagnetic wave.

As we will see later in this chapter, the free-space value Z_0 of the wave impedance, given by Eq. (8), establishes the scale of Z of virtually all wave transmission lines, so we may use it, together with Eq. (9), to get a better feeling of how much different are the electric and magnetic field amplitudes in the waves – on the scale of typical electrostatics and magnetostatics experiments. For example, according to Eqs. (9), a wave of a modest intensity $S = 1$ W/m^2 (this is what we get from a usual electric bulb a few meters away from it) has $E \sim (SZ_0)^{1/2} \sim 20$ V/m, quite comparable with the dc field created by a standard AA battery right outside it. On the other hand, the wave's magnetic field $H = (S/Z_0)^{1/2} \approx 0.05$ A/m. For this particular case, the relation following from Eqs. (1), (4), and (7),

$$B = \mu H = \mu \frac{E}{Z} = \mu \frac{E}{(\mu/\varepsilon)^{1/2}} \equiv (\varepsilon\mu)^{1/2} E = \frac{E}{v}, \tag{7.10}$$

gives $B = \mu_0 H = E/c \sim 7\times10^{-8}$T, i.e. a magnetic field a thousand times lower than the Earth's field, and about 7 orders of magnitude lower than the field of a typical permanent magnet. This huge difference may be interpreted as follows: the scale $B \sim E/c$ of magnetic fields in the waves is "normal" for electromagnetism, while the permanent magnet fields are abnormally high because they are due to the ferromagnetic alignment of electron spins, essentially relativistic objects – see the discussion in Sec. 5.5.

The fact that Eq. (5) is valid for an arbitrary function f means, in the standard terminology, that a medium with frequency-independent ε and μ supports the propagation of plane waves without either decay (*attenuation*) or waveform deformation (*dispersion*). However, for any real medium but pure vacuum, this approximation is valid only within limited frequency intervals. We will discuss the effects of attenuation and dispersion in the next section and will see that all our prior formulas remain valid even for an arbitrary linear media, provided that we limit them to single-frequency (i.e. sinusoidal, frequently called *monochromatic*) waves. Such waves may be most conveniently represented as[4]

$$f = \mathrm{Re}\left[f_\omega e^{i(kz - \omega t)} \right], \tag{7.11}$$

Mono-chromatic wave

where f_ω is the *complex amplitude* of the wave, and k is its *wave number* (the magnitude of the *wave vector* $\mathbf{k} \equiv \mathbf{n}k$), sometimes called the *spatial frequency*. The last term is justified by the fact, evident from Eq. (11), that k is related to the wavelength λ exactly as the usual ("temporal") frequency ω is related to the time period $\mathcal{T}$:

$$k = \frac{2\pi}{\lambda}, \qquad \omega = \frac{2\pi}{\mathcal{T}}. \tag{7.12}$$

Spatial and temporal frequencies

In the dispersion-free case (5), the compatibility of that relation with Eq. (11) requires the argument $(kz - \omega t) \equiv k\,[z - (\omega/k)t]$ to be proportional to $(z - vt)$, so $\omega/k = v$, i.e.

$$k = \frac{\omega}{v} \equiv (\varepsilon\mu)^{1/2}\omega, \tag{7.13}$$

Dispersion relation

so in that particular case, the *dispersion relation* $\omega(k)$ is linear.

Now note that Eq. (6) does not mean that the vectors **E** and **H** retain their direction in space. (The wave in which they do, is called *linearly polarized*.[5]) Indeed, nothing in the Maxwell equations prevents, for example, a joint rotation of this vector pair around the fixed vector **n**, while still keeping all these three vectors perpendicular to each other at any instant – see Fig. 1. However, an arbitrary rotation law or even an arbitrary constant frequency of such rotation would violate the single-frequency (monochromatic) character of the elementary sinusoidal wave (11). To understand what is the most general type of polarization the wave may have without violating that condition, let us represent two

[4] As we have already seen in the previous chapter (see also CM Sec. 5.1), such complex-exponential representation of sinusoidally changing variables is more convenient for mathematical manipulation than by using sine and cosine functions, especially because in all linear relations, the operator Re may be omitted (implied) until the very end of the calculation. Note, however, that this is *not* valid for the quadratic forms such as Eqs. (9).

[5] The possibility of different polarizations of electromagnetic waves was discovered (for light) in 1699 by Rasmus Bartholin, a.k.a. Erasmus Bartholinus.

Cartesian components of one of these vectors (say, **E**) along any two fixed axes *x* and *y*, perpendicular to each other and the *z*-axis (i.e. to the vector **n)**, in the same form as used in Eq. (11):

$$E_x = \mathrm{Re}\left[E_{\omega x} e^{i(kz - \omega t)} \right], \qquad E_y = \mathrm{Re}\left[E_{\omega y} e^{i(kz - \omega t)} \right]. \tag{7.14}$$

To keep the wave monochromatic, the complex amplitudes $E_{\omega x}$ and $E_{\omega y}$ have to be constant in time; however, they may have different magnitudes and an arbitrary phase shift between them.

In the simplest case when the arguments of these complex amplitudes are equal,

$$E_{\omega x,y} = \left| E_{\omega x,y} \right| e^{i\varphi}. \tag{7.15}$$

the real field components have the same phase:

$$E_x = \left| E_{\omega x} \right| \cos(kz - \omega t + \varphi), \qquad E_y = \left| E_{\omega y} \right| \cos(kz - \omega t + \varphi), \tag{7.16}$$

so their ratio is constant in time – see Fig. 3a. This means that the wave is linearly polarized, with the polarization plane defined by the relation

$$\tan\theta = \left| E_{\omega y} \right| / \left| E_{\omega x} \right|. \tag{7.17}$$

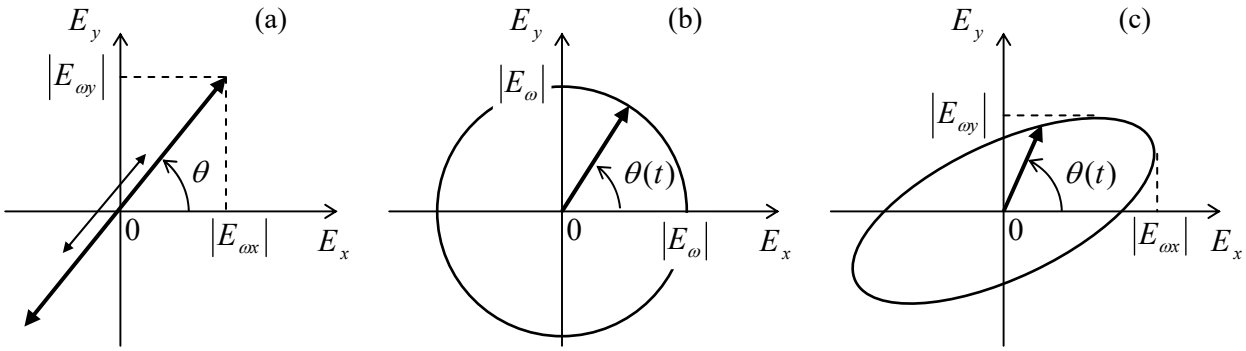


Fig. 7.3. Time evolution of the instantaneous electric field vector in monochromatic waves with: (a) a linear polarization, (b) a circular polarization, and (c) an elliptical polarization.

Another simple case is when the moduli of the complex amplitudes $E_{\omega x}$ and $E_{\omega y}$ are equal, but their phases are shifted by $+\pi/2$ or $-\pi/2$:

$$E_{\omega x} = \left| E_\omega \right| e^{i\varphi}, \quad E_{\omega y} = \left| E_\omega \right| e^{i(\varphi \pm \pi/2)}. \tag{7.18}$$

In this case

$$E_x = \left| E_\omega \right| \cos(kz - \omega t + \varphi), \quad E_y = \left| E_\omega \right| \cos\left(kz - \omega t + \varphi \pm \frac{\pi}{2} \right) \equiv \mp \left| E_\omega \right| \sin(kz - \omega t + \varphi). \tag{7.19}$$

This means that on the wave's plane (normal to **n**), the end of the vector **E** moves, with the wave's frequency ω, either clockwise or counterclockwise around a circle – see Fig. 3b:

$$\theta(t) = \mp(\omega t - \varphi). \tag{7.20}$$

Such waves are called *circularly polarized*. In the dominant convention, the wave is called *right-polarized* (RP) if it is described by the lower sign in Eqs. (18)-(20), i.e. if the vector **ω** of the angular frequency of the field vector's rotation coincides with the wave propagation's direction **n**, and *left-polarized* (LP) in the opposite case. These particular solutions of the Maxwell equations are very convenient for quantum electrodynamics, because single electromagnetic field quanta with a certain (positive or negative) spin direction may be considered elementary excitations of the corresponding circularly polarized wave.[6] (This fact does not exclude, from the quantization scheme, waves of other polarizations, because any monochromatic wave may be presented as a linear combination of two opposite circularly polarized waves – just as Eqs. (14) represent it as a linear combination of two linearly polarized waves.)

Finally, in the general case of arbitrary complex amplitudes $E_{\omega x}$ and $E_{\omega y}$, the field vector's end moves along an ellipse (Fig. 3c); such a wave is called *elliptically polarized*. The elongation ("eccentricity") and orientation of the ellipse are completely described by one complex number, the ratio $E_{\omega x}/E_{\omega y}$, i.e. by two real numbers – for example, $|E_{\omega x}/E_{\omega y}|$ and $\varphi = \arg(E_{\omega x}/E_{\omega y})$.[7]

7.2. Attenuation and dispersion

Let me start the discussion of the dispersion and attenuation effects by considering a particular case of the time evolution of the electric polarization $\mathbf{P}(t)$ of a dilute, non-polar medium, with negligible interaction between its elementary dipoles $\mathbf{p}(t)$. As was discussed in Sec. 3.3, in this case, the local electric field acting on each elementary dipole, may be taken equal to the macroscopic field $\mathbf{E}(t)$. Then, the dipole moment $\mathbf{p}(t)$ may be caused not only by the values of the field $\mathbf{E}$ at the same moment of time (t), but also by those at the earlier moments $t' < t$. Due to the linear superposition principle, the macroscopic polarization $\mathbf{P}(t) = n\mathbf{p}(t)$ should be a sum (practically, an integral) of the values of $\mathbf{E}(t')$ at all moments $t' \le t$, weighed by some function of t and t':[8]

$$P(t) = \int_{-\infty}^{t} E(t')G(t,t')dt'. \qquad (7.21)$$

Temporal Green's function

[6] This issue is closely related to that of the wave's angular momentum; it will be more convenient for me to discuss it later in this chapter (in Sec. 7).

[7] Note that the same information may be expressed via four so-called *Stokes parameters* s_0, s_1, s_2, and s_3, which are popular in practical optics, because they may be used for the description of not only completely coherent waves that were discussed here but also of party coherent or even fully incoherent waves – including the *natural light* emitted by thermal sources such as our Sun. (In contrast to the coherent waves (14), whose complex amplitudes are deterministic numbers, the amplitudes of incoherent waves should be treated as random variables.) For more on the Stokes parameters, as well as many other optics topics I will not have time to cover, I can recommend the classical text by M. Born *et al.*, *Principles of Optics*, 7th ed., Cambridge U. Press, 1999.

[8] In an isotropic media, the vectors **E**, **P**, and hence $\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}$, are all parallel, and for notation simplicity, I will drop the vector sign in the following formulas of this section. I am also assuming that **P** at any point **r** is only dependent on the electric field at the same point, and hence drop the factor $\exp\{ikz\}$, the same for all variables. This last assumption is valid if the wavelength λ is much larger than the elementary dipole's size a. In most systems of interest, the scale of a is atomic ($\sim 10^{-10}$m), so this approximation is valid up to extremely high frequencies, $\omega \sim c/a \sim 10^{18}$ s^{-1}, corresponding to hard X-rays.

The condition $t' \le t$, which is implied by this relation, expresses a keystone principle of all science, the *causal relation* between a cause (in our case, the electric field $E(t')$ applied to each dipole) and its effect (the polarization $P(t)$ it creates). The function $G(t, t')$ is called the *temporal Green's function* for the electric polarization.[9] To reveal its physical sense, let us consider the case when the applied field $E(t)$ is a very short pulse at the moment $t_0 < t$, which may be well approximated with Dirac's delta function:

$$E(t) = \delta(t - t'') . \tag{7.22}$$

Then Eq. (21) yields just $P(t) = G(t, t'')$, so the Green's function $G(t, t')$ is just the polarization at moment t, created by a unit δ-functional pulse of the applied field at moment t' (Fig. 4).

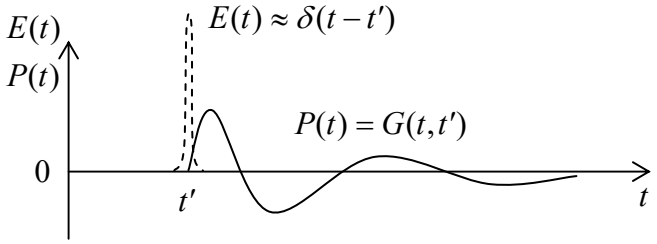


Fig. 7.4. An example of the temporal Green's function for the electric polarization (schematically).

What are the general properties of the temporal Green's function? First, for systems without infinite internal "memory", G should tend to zero at $t - t' \to \infty$, although the type of this approach (e.g., whether the function G oscillates approaching zero, as in Fig. 4, or not) depends on the medium's properties. Second, if the parameters of the medium do not change in time, the polarization response to an electric field pulse should be dependent not on its absolute timing, but only on the time difference $\theta \equiv t - t'$ between the pulse and observation instants, when Eq. (21) is reduced to

$$P(t) = \int_{-\infty}^{t} E(t')G(t - t')dt' \equiv \int_{0}^{\infty} E(t - \theta)G(\theta)d\theta . \tag{7.23}$$

For a sinusoidal waveform, $E(t) = \mathrm{Re}\,[E_\omega e^{-i\omega t}]$, this equation yields

$$P(t) = \mathrm{Re}\int_{0}^{\infty} E_\omega e^{-i\omega(t-\theta)}G(\theta)d\theta \equiv \mathrm{Re}\left[\left(E_\omega \int_{0}^{\infty} G(\theta)e^{i\omega\theta}d\theta\right)e^{-i\omega t}\right]. \tag{7.24}$$

The expression in the last parentheses is of course nothing else than the complex amplitude P_ω of the polarization. This means that though even if the static linear relation (3.43), $P = \chi_e\varepsilon_0 E$, is invalid for an arbitrary time-dependent process, we may still keep its Fourier analog,

$$P_\omega = \chi_e(\omega)\varepsilon_0 E_\omega, \qquad \text{with } \chi_e(\omega) \equiv \frac{1}{\varepsilon_0}\int_{0}^{\infty} G(\theta)e^{i\omega\theta}d\theta , \tag{7.25}$$

for each sinusoidal component of the process, using it as the definition of the frequency-dependent electric susceptibility $\chi_e(\omega)$. Similarly, the frequency-dependent electric permittivity may be defined using the Fourier analog of Eq. (3.46):

[9] The idea of these functions is very similar to that of the spatial Green's functions (see Sec. 2.10), but with a new twist, due to the causality principle. A discussion of the temporal Green's functions in application to classical mechanics (which to some extent overlaps with our current discussion) may be found in CM Sec. 5.1.

$$D_\omega \equiv \varepsilon(\omega) E_\omega . \tag{7.26a}$$

Complex electric permittivity

Then, according to Eq. (3.47), the complex permittivity is related to the temporal Green's function by the usual Fourier transform:

$$\varepsilon(\omega) \equiv \varepsilon_0 + \frac{P_\omega}{E_\omega} = \varepsilon_0 + \int_0^\infty G(\theta) e^{i\omega\theta} d\theta . \tag{7.26b}$$

This relation shows that the function $\varepsilon(\omega)$ may be complex,

$$\varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega), \quad \text{with } \varepsilon'(\omega) = \varepsilon_0 + \int_0^\infty G(\theta)\cos\omega\theta\, d\theta, \quad \varepsilon''(\omega) = \int_0^\infty G(\theta)\sin\omega\theta\, d\theta , \tag{7.27}$$

and that its real part $\varepsilon'(\omega)$ is always an even function of frequency, while the imaginary part $\varepsilon''(\omega)$ is an odd function of ω. Note that though the particular causal relationship (21) between $P(t)$ and $E(t)$ is conditioned by the elementary dipole independence, the frequency-dependent complex electric permittivity $\varepsilon(\omega)$ may be introduced, in a similar way, if *any* two linear combinations of these variables are related by a similar formula.

Absolutely similar arguments show that magnetic properties of a linear, isotropic medium may be characterized by a frequency-dependent, complex permeability $\mu(\omega)$. Now rewriting Eqs. (1) for the complex amplitudes of the fields at a particular frequency, we may readily repeat all calculations of Sec. 1, and verify that all its results are valid for monochromatic waves even for a dispersive (but necessarily linear!) medium. In particular, Eqs. (7) and (13) now become

$$Z(\omega) = \left(\frac{\mu(\omega)}{\varepsilon(\omega)}\right)^{1/2}, \qquad k(\omega) = \omega\left[\varepsilon(\omega)\mu(\omega)\right]^{1/2}, \tag{7.28}$$

Complex Z and k

so the wave impedance and the wave number may be both complex functions of frequency.[10]

This fact has important consequences for electromagnetic wave propagation. First, plugging the representation of the complex wave number as the sum of its real and imaginary parts, $k(\omega) \equiv k'(\omega) + ik''(\omega)$, into Eq. (11):

$$f = \mathrm{Re}\left\{f_\omega e^{i[k(\omega)z - \omega t]}\right\} = e^{-k''(\omega)z}\,\mathrm{Re}\left\{f_\omega e^{i[k'(\omega)z - \omega t]}\right\}, \tag{7.29}$$

we see that $k''(\omega)$ describes the rate of wave *attenuation* in the medium at frequency ω.[11] Second, if the waveform is not sinusoidal (and hence should be represented as a sum of several/many sinusoidal components), the frequency dependence of $k'(\omega)$ provides for wave *dispersion*, i.e. the waveform deformation at the propagation, because the propagation velocity (4) of component waves is now different.[12]

---

[10] The first unambiguous observations of dispersion (for the case of light refraction) were described by Sir Isaac Newton in his *Optics* (1704) – even though this genius has never recognized the wave nature of light!

[11] It may be tempting to attribute this effect to wave *absorption*, i.e. the dissipation of the wave's energy, but we will see very soon that wave attenuation may be due to different effects as well.

[12] The reader is probably familiar with the most noticeable effect of the dispersion: the difference between the *group velocity* $v_{gr} \equiv d\omega/dk'$ giving the speed of the envelope of a wave packet with a narrow frequency spectrum, and the *phase velocity* $v_{ph} \equiv \omega/k'$ of the component waves. The second-order dispersion effect, proportional to $d^2\omega/d^2k'$, leads to the deformation (gradual broadening) of the envelope itself. Following tradition, these effects

As an example of such a dispersive medium, let us consider a simple but very representative *Lorentz oscillator model*.[13] In dilute atomic or molecular systems (e.g., gases), electrons respond to the external electric field especially strongly when its frequency ω is close to certain frequencies ω_j corresponding to the spectrum of quantum interstate transitions of a single atom/molecule. A phenomenological description of this behavior may be obtained from a classical model of several externally driven harmonic oscillators, generally with non-zero damping. For a single oscillator, driven by the electric field's force $F(t) = qE(t)$, we can write the 2nd Newton law as

$$m\left(\ddot{x} + 2\delta_0\dot{x} + \omega_0^2 x\right) = qE(t), \tag{7.30}$$

where ω_0 is the own frequency of the oscillator, and δ_0 is its damping coefficient. For the electric field of a monochromatic wave, $E(t) = \mathrm{Re}\,[E_\omega \exp\{-i\omega t\}]$, we may look for a particular, *forced-oscillation* solution of this equation in a similar form $x(t) = \mathrm{Re}\,[x_\omega \exp\{-i\omega t\}]$.[14] Plugging this solution into Eq. (30), we readily find the complex amplitude of these oscillations:

$$x_\omega = \frac{q}{m}\frac{E_\omega}{\left(\omega_0^2 - \omega^2\right) - 2i\omega\delta_0}. \tag{7.31}$$

Using this result to calculate the complex amplitude of the dipole moment as $p_\omega = qx_\omega$, and then the electric polarization $P_\omega = np_\omega$ of a dilute medium with n independent oscillators for unit volume, for its frequency-dependent permittivity (26) we get

Lorentz oscillator model

$$\varepsilon(\omega) = \varepsilon_0 + n\frac{q^2}{m}\frac{1}{\left(\omega_j^2 - \omega^2\right) - 2i\omega\delta_0}. \tag{7.32}$$

This result may be readily generalized to the case when the system has several types of oscillators with different masses and frequencies:

$$\varepsilon(\omega) = \varepsilon_0 + nq^2\sum_j \frac{f_j}{m_j\left[\left(\omega_j^2 - \omega^2\right) - 2i\omega\delta_j\right]}, \tag{7.33}$$

where $f_j \equiv n_j/n$ is the fraction of oscillators with frequency ω_j, so the sum of all f_j equals 1. Figure 5 shows a typical behavior of the real and imaginary parts of the complex dielectric constant, described by Eq. (33), as functions of frequency. The oscillator resonances' effect is clearly visible, and dominates the media response at $\omega \approx \omega_j$, especially in the case of low damping, $\delta_j << \omega_j$. Note that in the low-damping limit, the imaginary part of the dielectric constant ε'', and hence the wave attenuation k'', are negligibly small at all frequencies besides small vicinities of ω_j, where the derivative $d\varepsilon'(\omega)/d\omega$ is

---

are discussed in more detail in the quantum-mechanics part of this series (QM Sec. 2.2), because they are a crucial factor of Schrödinger's wave mechanics. (See also a brief discussion in CM Sec. 6.3.)

[13] This example is focused on the frequency dependence of ε rather than μ, because electromagnetic waves interact with "usual" media via their electric field much more than via the magnetic field. Indeed, according to Eq. (7), the magnetic field of the wave is of the order of E/c, so the magnetic component of the Lorentz force (5.10), acting on a non-relativistic particle, $F_m \sim quB \sim (u/c)qE$, is much smaller than that of its electric component, $F_e = qE$, and may be neglected. However, as will be discussed in Sec. 6, forgetting about the possible dispersion of $\mu(\omega)$ may result in missing some remarkable opportunities for manipulating the waves.

[14] If this point and Eq. (30) are not absolutely clear, please see CM Sec. 5.1 for a more detailed discussion.

negative.[15] Thus, for a system of weakly-damped oscillators, Eq. (33) may be well approximated by a sum of singularities ("poles"):

$$\varepsilon(\omega) \approx \varepsilon_0 + n\frac{q^2}{2}\sum_j \frac{f_j}{m_j\omega_j(\omega_j - \omega)}, \qquad \text{for } \delta_j << \left|\omega - \omega_j\right| << \left|\omega_j - \omega_{j'}\right|. \tag{7.34}$$

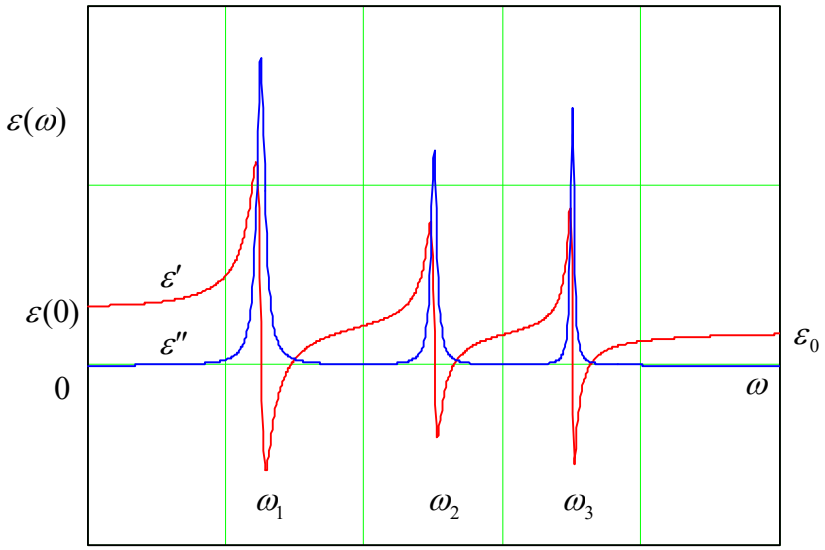


Fig. 7.5. Typical frequency dependence of the real and imaginary parts of the complex electric permittivity, according to the generalized Lorentz oscillator model.

This result is especially important because according to quantum mechanics,[16] Eq. (34) (with all m_j equal) is also valid for a set of non-interacting, similar quantum systems (whose dynamics may be completely different from that of a harmonic oscillator!), provided that ω_j are replaced with frequencies of possible quantum interstate transitions, and coefficients f_j are replaced with the so-called *oscillator strengths* of the transitions – which obey the same *sum rule*, $\Sigma_j f_j = 1$.

At $\omega \to 0$, the imaginary part of the complex permittivity (33) also vanishes (for any δ_j), while its real part approaches its electrostatic ("dc") value

$$\varepsilon(0) = \varepsilon_0 + q^2\sum_j \frac{n_j}{m_j\omega_j^2}. \tag{7.35}$$

Note that according to Eq. (30), the denominator of the fraction in Eq. (35) is just the effective spring constant $\kappa_j = m_j\omega_j^2$ of the j^{th} oscillator, so the oscillator masses m_j as such are actually (and quite naturally) not involved in the static dielectric response.

In the opposite limit of very high frequencies, $\omega >> \omega_j, \delta_j$, the permittivity also becomes real and may be represented as

$$\varepsilon(\omega) = \varepsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right), \qquad \text{where } \omega_p^2 \equiv \frac{q^2}{\varepsilon_0}\sum_j \frac{n_j}{m_j}. \tag{7.36}$$

$\varepsilon(\omega)$ in plasma

[15] In optics, such behavior is called *anomalous dispersion*.

[16] See, e.g., QM Chapters 5-6.

This result is very important because it is also valid at *all* frequencies if all ω_j and δ_j vanish, for example for gases of free charged particles, in particular for *plasmas* – ionized atomic gases, provided that the ion collision effects are negligible. (This is why the parameter ω_p defined by Eq. (36) is called the *plasma frequency*.) Typically, the plasma as a whole is neutral, i.e. the density n of positive atomic ions is equal to that of the free electrons. Since the ratio n_j/m_j for electrons is much higher than that for ions, the general formula (36) for the plasma frequency is usually well approximated by the following simple expression:

$$\omega_p^2 = \frac{ne^2}{\varepsilon_0 m_e}. \tag{7.37}$$

This expression has a simple physical sense: the effective spring constant $\kappa_{ef} \equiv m_e\omega_p^2 = ne^2/\varepsilon_0$ describes the Coulomb force that appears when the electron subsystem of the plasma is shifted, as a whole, from its positive-ion subsystem, thus violating the electroneutrality.[17] Hence, there is no surprise that the function $\varepsilon(\omega)$ given by Eq. (36) vanishes at $\omega = \omega_p$: at this resonance frequency, the polarization electric field **E** may oscillate, i.e. have a non-zero amplitude $E_\omega = D_\omega/\varepsilon(\omega)$, even in the absence of external forces induced by external (stand-alone) charges, i.e. in the absence of the field **D** these charges induce – see Eq. (3.32).

The behavior of electromagnetic waves in a medium that obeys Eq. (36), is very remarkable. If the wave frequency ω is above ω_p, the dielectric constant $\varepsilon(\omega)$ and hence the wave number (28) are positive and real, and waves propagate without attenuation, following the dispersion relation,

Plasma dispersion relation

$$k(\omega) = \omega\left[\varepsilon(\omega)\mu_0\right]^{1/2} = \frac{1}{c}\left(\omega^2 - \omega_p^2\right)^{1/2}, \tag{7.38}$$

whose plot is shown in Fig. 6.

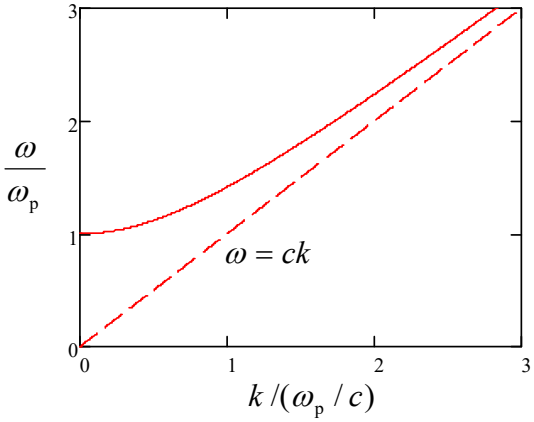


Fig. 7.6. The plasma dispersion law (solid line) in comparison with the linear dispersion in the free space (dashed line).

At $\omega \to \omega_p$ the wave number k tends to zero. Beyond that point (i.e. at $\omega < \omega_p$), we still can use Eq. (38), but it is instrumental to rewrite it in the mathematically equivalent form

[17] Indeed, let us consider such a small shift Δx, perpendicular to the plane surface of a broad, plane slab filled with plasma. The uncompensated ion charges, with equal and opposite surface densities $\sigma = \pm en\Delta x$, that appear at the slab surfaces, create inside it, according to Eq. (2.3), a uniform electric field with $E_x = en\Delta x/\varepsilon_0$. This field exerts the force $-eE_x = -(ne^2/\varepsilon_0)\Delta x \equiv -\kappa_{ef}\Delta x$ on each electron, pulling it back to its equilibrium position.

$$k(\omega) = \frac{i}{c}\left(\omega_{\rm p}^2 - \omega^2\right)^{1/2} = \frac{i}{\delta}, \qquad \text{where } \delta \equiv \frac{c}{\left(\omega_{\rm p}^2 - \omega^2\right)^{1/2}}. \tag{7.39}$$

At $\omega < \omega_{\rm p}$, the so-defined parameter δ is real, and Eq. (29) shows that the electromagnetic field exponentially decreases with distance:

$$f = \mathrm{Re}\, f_\omega e^{i(kz-\omega t)} \equiv \exp\left\{-\frac{z}{\delta}\right\} \mathrm{Re}\, f_\omega e^{-i\omega t}. \tag{7.40}$$

Does this mean that the wave is being absorbed in the plasma? Answering this question is a good pretext to calculate the time average of the Poynting vector $\mathbf{S} = \mathbf{E}\times\mathbf{H}$ of a monochromatic electromagnetic wave in an *arbitrary* dispersive (but still linear and isotropic) medium. First, let us spell out the real fields' time dependences:

$$E(t) = \mathrm{Re}\left[E_\omega e^{-i\omega t}\right] \equiv \frac{1}{2}\left[E_\omega e^{-i\omega t} + \text{c.c.}\right], \qquad H(t) = \mathrm{Re}\left[H_\omega e^{-i\omega t}\right] \equiv \frac{1}{2}\left[\frac{E_\omega}{Z(\omega)} e^{-i\omega t} + \text{c.c.}\right]. \tag{7.41}$$

Now, a straightforward calculation yields[18]

$$\overline{S} = \overline{E(t)H(t)} = \frac{E_\omega E_\omega^*}{4}\left[\frac{1}{Z(\omega)} + \frac{1}{Z^*(\omega)}\right] \equiv \frac{E_\omega E_\omega^*}{2}\mathrm{Re}\frac{1}{Z(\omega)} \equiv \frac{\left|E_\omega\right|^2}{2}\mathrm{Re}\left[\frac{\varepsilon(\omega)}{\mu(\omega)}\right]^{1/2}. \tag{7.42}$$

Let us apply this important general formula to our simple model of plasma at $\omega < \omega_{\rm p}$. In this case, the magnetic permeability equals μ_0, i.e. $\mu(\omega) = \mu_0$ is positive and real, while $\varepsilon(\omega)$ is real and negative, so $1/Z(\omega) = [\varepsilon(\omega)/\mu(\omega)]^{1/2}$ is purely imaginary, and the average Poynting vector (42) vanishes. This means that the energy, on average, does not flow along the z-axis. So, the waves with $\omega < \omega_{\rm p}$ are not absorbed in plasma. (Indeed, the Lorentz model with $\delta_j = 0$ does not describe any energy dissipation mechanism.) Instead, as we will see in the next section, the waves are rather *reflected* from the plasma's boundary, more exactly from its surface layer of a thickness $\sim\delta$.

Note also that in the limit $\omega << \omega_{\rm p}$, Eq. (39) yields

$$\delta \to \frac{c}{\omega_{\rm p}} = \left(\frac{c^2\varepsilon_0 m_{\rm e}}{ne^2}\right)^{1/2} = \left(\frac{m_{\rm e}}{\mu_0 n e^2}\right)^{1/2}. \tag{7.43}$$

But this is just a particular case (for $q = e$, $m = m_{\rm e}$, and $\mu = \mu_0$) of Eq. (6.44) that was derived in Sec. 6.4 for the depth of the magnetic field's penetration into a lossless (collision-free) conductor in the quasistatic approximation. This fact shows again that, as was already discussed in Sec. 6.7, this

[18] For an arbitrary plane wave, the total average power flow may be calculated as an integral of Eq. (42) over all frequencies. By the way, combining this integral and the Poynting theorem (6.111), is it straightforward to prove the following interesting expression for the average electromagnetic energy density of a narrow ($\Delta\omega << \omega$) wave packet propagating in an arbitrary dispersive (but linear and isotropic) medium:

$$\overline{u} = \frac{1}{2}\int_{\text{packet}} \left\{\frac{d\left[\omega\varepsilon'(\omega)\right]}{d\omega} E_\omega E_\omega^* + \frac{d\left[\omega\mu'(\omega)\right]}{d\omega} H_\omega H_\omega^*\right\} d\omega.$$

approximation (in which the displacement currents are neglected) gives an adequate description of the time-dependent phenomena at $\omega << \omega_p$, i.e. at $\delta << c/\omega = 1/k = \lambda/2\pi$.[19]

There are two most important examples of natural plasmas. For the Earth's ionosphere, i.e. the upper part of its atmosphere, which is almost completely ionized by the ultra-violet and X-ray components of the Sun's radiation, the maximum value of n, reached about 300 km over the Earth's surface, is between 10^{10} and 10^{12} m^{-3} (depending on the time of the day and the Sun's activity phase), so the maximum plasma frequency (37) is between 1 and 10 MHz. This is much higher than the particles' typical reciprocal collision time τ^{-1}, so Eq. (38) gives a good description of wave dispersion in this plasma. The effect of reflection of electromagnetic waves with $\omega < \omega_p$ from the ionosphere enables long-range (over-the-globe) radio communications and broadcasting at the so-called *short waves*, with cyclic frequencies of the order of 10 MHz:[20] they may propagate in the flat channel formed by the Earth's surface and the ionosphere, being reflected repeatedly by these parallel "walls". Unfortunately, due to the random variations of the Sun's activity, and hence of ω_p, this natural radio communication channel is not too reliable, and in our age of transworld optical-fiber cables (see Sec. 7 below), its practical importance has diminished.

Another important example of plasmas is free electrons in metals and other conductors. For a typical metal, n is of the order of 10^{23} cm^{-3} $\equiv 10^{29}$ m^{-3}, so Eq. (37) yields $\omega_p \sim 10^{16}$ s^{-1}. This value of ω_p is somewhat higher than the mid-optical frequencies ($\omega \sim 3\times10^{15}$ s^{-1}), explaining why planar, clean metallic surfaces, such as the aluminum and silver films used in mirrors, are so shiny: at these frequencies, their complex permittivity $\varepsilon(\omega)$ is almost exactly real and negative, leading to light reflection, with very little absorption.

The simple model (36), which neglects electron scattering, becomes inadequate at lower frequencies, $\omega\tau \sim 1$. A good phenomenological way of extending the model to the account of scattering is to take, in Eq. (33), the lowest frequency ω_j equal zero (to describe the free electrons), while keeping the damping coefficient δ_0 of this mode larger than zero, to account for the energy dissipation due to their scattering. Then Eq. (33) is reduced to

$$\varepsilon_{ef}(\omega) = \varepsilon_{opt}(\omega) + \frac{n_0 q^2}{m}\frac{1}{-\omega^2 - 2i\omega\delta_0} \equiv \varepsilon_{opt}(\omega) + i\frac{n_0 q^2}{2\delta_0 m\omega}\frac{1}{1 - i\omega/2\delta_0}, \tag{7.44}$$

where the response $\varepsilon_{opt}(\omega)$ at high (in practice, optical) frequencies is still given by Eq. (33), but now with $j > 0$. The result (44) allows for a simple interpretation. To show that, let us incorporate into our calculations the Ohmic conduction of the medium, generalizing Eq. (4.7) as $\mathbf{j}_\omega = \sigma(\omega)\mathbf{E}_\omega$ to account for the possible frequency dependence of the Ohmic conductivity. Plugging this relation into the Fourier image of the relevant macroscopic Maxwell equation, $\nabla\times\mathbf{H}_\omega = \mathbf{j}_\omega - i\omega\mathbf{D}_\omega \equiv \mathbf{j}_\omega - i\omega\varepsilon(\omega)\mathbf{E}_\omega$, we get

$$\nabla\times\mathbf{H}_\omega = \left[\sigma(\omega) - i\omega\varepsilon(\omega)\right]\mathbf{E}_\omega. \tag{7.45}$$

[19] One more convenience of the simple model of a collision-free plasma, which has led us to Eq. (36), is that it may be readily generalized to the case of an additional strong dc magnetic field $\mathbf{B}_0$ (much higher than that of the wave) applied in the direction $\mathbf{n}$ of wave propagation. It is straightforward (and hence left for the reader) to show that such plasma exhibits the *Faraday effect* of the polarization plane's rotation, and hence gives an example of an anisotropic media that violates the Lorentz reciprocity relation (6.121).

[20] These frequencies are an order of magnitude lower than those used for TV and FM-radio broadcasting.

This relation shows that for a monochromatic wave, the addition of the Ohmic current density $\mathbf{j}_\omega$ to the displacement current density is equivalent to the addition of $\sigma(\omega)$ to $-i\omega\varepsilon(\omega)$, i.e. to the following change of the ac electric permittivity:[21]

$$\varepsilon(\omega) \to \varepsilon_{\text{ef}}(\omega) \equiv \varepsilon_{\text{opt}}(\omega) + i\frac{\sigma(\omega)}{\omega}. \tag{7.46}$$

Now the comparison of Eqs. (44) and (46) shows that they coincide if we take

$$\sigma(\omega) = \frac{n_0 q^2 \tau}{m_0}\frac{1}{1-i\omega\tau} \equiv \sigma(0)\frac{1}{1-i\omega\tau}, \tag{7.47}$$

Generalized Drude formula

where the dc conductivity $\sigma(0)$ is described by the Drude formula (4.13), and the phenomenologically introduced coefficient δ_0 is associated with $1/2\tau$. Eq. (47), which is frequently called the *generalized* (or "ac", or "rf") *Drude formula*,[22] gives a very reasonable (semi-quantitative) description of the ac conductivity of many metals almost up to optical frequencies.

Now returning to our discussion of the generalized Lorentz model (33), we see that the frequency dependences of the real (ε') and imaginary (ε'') parts of the complex permittivity it yields are not quite independent. For example, let us have one more look at the resonance peaks in Fig. 5. Each time the real part drops with frequency, $d\varepsilon'/d\omega < 0$, its imaginary part ε'' has a positive peak. Ralph Kronig (in 1926) and Hendrik ("Hans") Kramers (in 1927) independently showed that this is not an occasional coincidence pertinent only to this particular model. Moreover, the full knowledge of the function $\varepsilon'(\omega)$ enables the *calculation* of the function $\varepsilon''(\omega)$, and vice versa. The mathematical reason for this fact is that both these functions are always related to a single real function $G(\theta)$ – see Eqs. (27).

To derive these relations, let us consider Eq. (26b) on the complex frequency plane, $\omega \to \boldsymbol{\omega} \equiv \omega' + i\omega''$:

$$f(\boldsymbol{\omega}) \equiv \varepsilon(\boldsymbol{\omega}) - \varepsilon_0 = \int_0^\infty G(\theta) e^{i\boldsymbol{\omega}\theta} d\theta \equiv \int_0^\infty G(\theta) e^{i\omega'\theta} e^{-\omega''\theta} d\theta. \tag{7.48}$$

For all stable physical systems, $G(\theta)$ has to be finite for all important values of the real integration variable ($\theta > 0$), and tend to zero at $\theta \to 0$ and $\theta \to \infty$. (Indeed, according to Eq. (23), a non-zero $G(0)$ would mean an instantaneous response of the medium to the external force, while $G(\infty) \neq 0$ would mean that it has an infinitely long memory.) Because of that, and thanks to the factor $e^{-\omega''\theta}$, the expression under the integral in Eq. (48) tends to zero at $|\boldsymbol{\omega}| \to \infty$ in all upper half-plane ($\omega'' \geq 0$). As a result, we may claim that the complex function $f(\boldsymbol{\omega})$ given by this relation, is analytical in that half-plane. This fact allows us to apply to it the general *Cauchy integral* formula[23]

$$f(\boldsymbol{\omega}) = \frac{1}{2\pi i}\oint_C f(\boldsymbol{\Omega})\frac{d\boldsymbol{\Omega}}{\boldsymbol{\Omega} - \boldsymbol{\omega}}, \tag{7.49}$$

[21] Alternatively, according to Eq. (45), it is possible (and in the field of infrared spectroscopy, conventional) to attribute the ac response of a medium, at *all* frequencies, to its effective complex conductivity: $\sigma_{\text{ef}}(\omega) \equiv \sigma(\omega) - i\omega\varepsilon(\omega) \equiv -i\omega\varepsilon_{\text{ef}}(\omega)$.

[22] It may be also derived from the Boltzmann kinetic equation in the so-called relaxation-time approximation (RTA) – see, e.g., SM Sec. 6.2.

[23] See, e.g., MA Eq. (15.2).

where $\mathbf{\Omega} \equiv \Omega' + i\Omega''$ is also a complex variable. Let us take the integration contour C of the form shown in Fig. 7, with the radius R of the larger semicircle tending to infinity, and the radius r of the smaller semicircle (around the singular point $\mathbf{\Omega} = \omega$) tending to zero. Due to the exponential decay of $|f(\mathbf{\Omega})|$ at $|\mathbf{\Omega}| \to \infty$, the contribution to the right-hand side of Eq. (49) from the larger semicircle vanishes,[24] while the contribution from the small semicircle, where $\mathbf{\Omega} = \omega + r\exp\{i\varphi\}$, with $-\pi \le \varphi \le 0$, is

$$\lim_{r\to 0} \frac{1}{2\pi i} \int_{\mathbf{\Omega}=\omega+r\exp\{i\varphi\}} f(\mathbf{\Omega}) \frac{d\mathbf{\Omega}}{\mathbf{\Omega}-\omega} = \frac{f(\omega)}{2\pi i} \int_{-\pi}^{0} \frac{ir\exp\{i\varphi\}d\varphi}{r\exp\{i\varphi\}} \equiv \frac{f(\omega)}{2\pi} \int_{-\pi}^{0} d\varphi = \frac{1}{2} f(\omega). \quad (7.50)$$

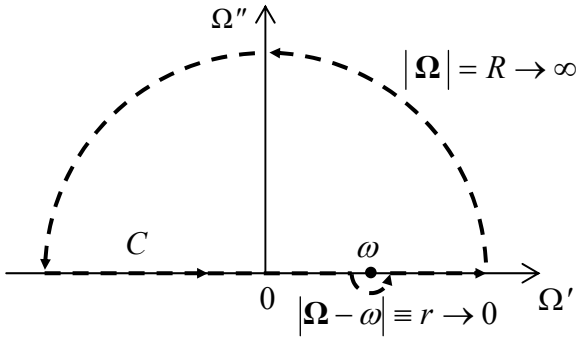


Fig. 7.7. Deriving the Kramers-Kronig dispersion relations.

As a result, for our contour C, Eq. (49) yields

$$f(\omega) = \lim_{r\to 0} \frac{1}{2\pi i} \left(\int_{-\infty}^{\omega-r} + \int_{\omega+r}^{+\infty} \right) f(\Omega) \frac{d\Omega}{\Omega-\omega} + \frac{1}{2} f(\omega)., \quad (7.51)$$

where $\Omega \equiv \Omega'$ on the real axis (where $\Omega'' = 0$). Such an integral, excluding a symmetric infinitesimal vicinity of a pole singularity, is called the *principal value* of the (formally, diverging) integral from $-\infty$ to $+\infty$, and is denoted by the letter P before it.[25] Using this notation, subtracting $f(\omega)/2$ from both parts of Eq. (51), and multiplying them by 2, we get

$$f(\omega) = \frac{1}{\pi i} \mathrm{P} \int_{-\infty}^{+\infty} f(\Omega) \frac{d\Omega}{\Omega-\omega}. \quad (7.52)$$

Now plugging into this complex equality the polarization-related difference $f(\omega) \equiv \varepsilon(\omega) - \varepsilon_0$ in the form $[\varepsilon'(\omega) - \varepsilon_0] + i[\varepsilon''(\omega)]$, and requiring both real and imaginary components of the two sides of Eq. (52) to be equal separately, we get the famous *Kramers-Kronig dispersion relations*

Kramers-Kronig dispersion relations

$$\varepsilon'(\omega) = \varepsilon_0 + \frac{1}{\pi} \mathrm{P} \int_{-\infty}^{+\infty} \varepsilon''(\Omega) \frac{d\Omega}{\Omega-\omega}, \qquad \varepsilon''(\omega) = -\frac{1}{\pi} \mathrm{P} \int_{-\infty}^{+\infty} [\varepsilon'(\Omega) - \varepsilon_0] \frac{d\Omega}{\Omega-\omega}. \quad (7.53)$$

We may use the already mentioned fact that $\varepsilon'(\omega)$ is always an even function, while $\varepsilon''(\omega)$ an odd function of frequency, to rewrite these relations in the following equivalent form,

[24] Strictly speaking, this also requires $|f(\Omega)|$ to decrease faster than Ω^{-1} at the real axis (at $\Omega'' = 0$), but due to the inertia of charged particles, this requirement is fulfilled for all realistic models of dispersion – see, e.g., Eq. (36).

[25] I am typesetting this symbol in a Roman (upright) font, to avoid any possibility of confusion with the medium's polarization.

$$\varepsilon'(\omega) = \varepsilon_0 + \frac{2}{\pi} \mathrm{P} \int_0^{+\infty} \varepsilon''(\Omega) \frac{\Omega d\Omega}{\Omega^2 - \omega^2}, \qquad \varepsilon''(\omega) = -\frac{2\omega}{\pi} \mathrm{P} \int_0^{+\infty} [\varepsilon'(\Omega) - \varepsilon_0] \frac{d\Omega}{\Omega^2 - \omega^2}, \tag{7.54}$$

which is more convenient for most applications, because it involves only physical (positive) frequencies.

Though the Kramers-Kronig relations are "global" in frequency, in certain cases they allow an approximate calculation of dispersion from experimental data for absorption, collected even within a limited ("local") frequency range. Most importantly, if a medium has a sharp absorption peak at some frequency ω_j, we may describe it as

$$\varepsilon''(\omega) \approx c\delta(\omega - \omega_j) + \text{a more smooth function of } \omega, \tag{7.55}$$

and the first of Eqs. (54) immediately gives

$$\varepsilon'(\omega) \approx \varepsilon_0 + \frac{2c}{\pi} \frac{\omega_j}{\omega_j^2 - \omega^2} + \text{another smooth function of } \omega, \tag{7.56}$$

Dispersion near an absorption line

thus predicting the anomalous dispersion near such a point. This calculation shows that such behavior observed in the Lorentz oscillator model (see Fig. 5) is by no means occasional or model-specific.

Let me emphasize again that the Kramers-Kronig relations (53)-(54) are much more general than the Lorentz model (33), and require only a causal linear relation (21) between the polarization $P(t)$ with the electric field $E(t')$.[26] Hence, these relations are also valid for the complex functions relating Fourier images of any cause/effect-related pair of variables. In particular, at a measurement of *any* linear response $r(t)$ of *any* experimental sample to *any* external field $f(t')$, whatever the nature of this response and physics behind it, we may be confident that there is a causal relationship between the variables r and f, so the corresponding complex function $\chi(\omega) \equiv r_\omega/f_\omega$ does obey the Kramers-Kronig relations. However, it is still important to remember that a linear relationship between the Fourier amplitudes of two variables does *not* necessarily imply a causal relationship between them.[27]

7.3. Reflection

The most important new effect arising in nonuniform media is wave *reflection*. Let us start its discussion from the simplest case of a plane electromagnetic wave that is normally incident on a sharp interface between two uniform, linear, isotropic media.

Moreover, let us first consider an even simpler sub-case when one of the two media (say, that located at $z > 0$, see Fig. 8) cannot sustain any electric field at all – as implied, in particular, by the macroscopic model of a perfect conductor – see Eq. (2.1):

[26] Actually, in mathematics, the relations even somewhat more general than Eqs. (53) and valid for an arbitrary analytic function of complex argument (the *Sokhotski-Plemelj theorem*) have been known at least from 1868.

[27] For example, the function $\varphi(\omega) \equiv E_\omega / P_\omega$ in the Lorentz oscillator model, does not obey the Kramers-Kronig relations. This is evident not only physically, from the fact that $E(t)$ is *not* a causal function of $P(t)$, but even mathematically. Indeed, Green's function describing a causal relationship has to tend to zero at small time delays $\theta \equiv t - t'$, so its Fourier image has to tend to zero at $\omega \to \pm\infty$. This is certainly true for the function $f(\omega)$ given by Eq. (32), but not for the reciprocal function $\varphi(\omega) \equiv 1/f(\omega) \propto (\omega^2 - \omega_0^2) - 2i\delta\omega$, which diverges at large frequencies.

$$E\big|_{z\geq 0} = 0. \tag{7.57}$$

This condition is evidently incompatible with the single traveling wave (5). However, this solution may be readily generalized using the fact that the dispersion-free 1D wave equation,

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)E = 0, \tag{7.58}$$

supports waves propagating, with the same speed, in any of two opposite directions. As a result, the following linear superposition of two such waves,

$$E\big|_{z\leq 0} = f(z - vt) - f(-z - vt), \tag{7.59}$$

satisfies both the equation and the boundary condition (57), for an arbitrary function f. The second term on the right-hand side of Eq. (59) may be interpreted as a result of *total reflection* of the incident wave (described by its first term) – in this particular case, with the change of the electric field's sign. This means, in particular, that within the macroscopic model, a conductor acts as a perfect mirror. By the way, since the vector **n** of the reflected wave is opposite to that of the incident one (see the arrows in Fig. 8), Eq. (6) shows that the magnetic field of the wave does *not* change its sign at the reflection:

$$H\big|_{z\leq 0} = \frac{1}{Z}\left[f(z - vt) + f(-z - vt)\right]. \tag{7.60}$$

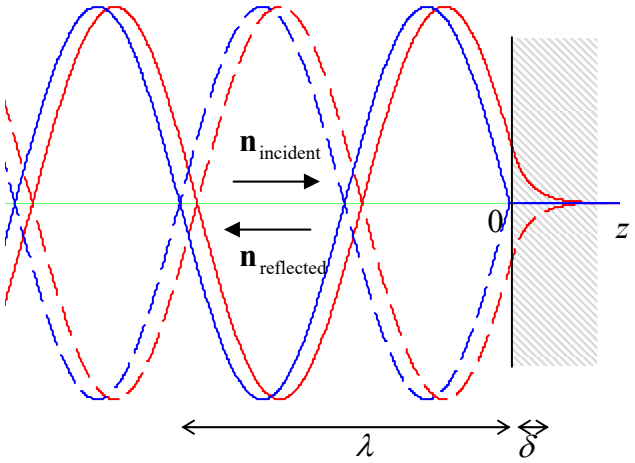


Fig. 7.8. A snapshot of the electric field at the reflection of a sinusoidal wave from a perfect conductor: a realistic pattern (red lines) and its macroscopic, ideal-mirror approximation (blue lines). Dashed lines show the snapshots after a half-period time delay ($\omega\Delta t = \pi$).

The blue lines in Fig. 8 show the resulting pattern (59) for the simplest, monochromatic wave:

Wave's total reflection

$$E\big|_{z\leq 0} = \mathrm{Re}\left[E_\omega e^{i(kz-\omega t)} - E_\omega e^{i(-kz-\omega t)}\right]. \tag{7.61a}$$

Depending on convenience in a particular context, this pattern may be legitimately represented and interpreted either as the linear superposition (61a) of two *traveling* waves or as a single *standing wave*:

$$E\big|_{z\leq 0} = -2\,\mathrm{Im}\left(E_\omega e^{-i\omega t}\right)\sin kz \equiv 2\,\mathrm{Re}\left(iE_\omega e^{-i\omega t}\right)\sin kz \equiv 2\,\mathrm{Re}\left[E_\omega e^{-i\left(\omega t - \pi/2\right)}\right]\sin kz, \tag{7.61b}$$

in which the electric and magnetic field oscillate with the phase shifts by $\pi/2$ both in time and space:

$$H\big|_{z\le 0} = \mathrm{Re}\left[\frac{E_\omega}{Z}e^{i(kz-\omega t)} + \frac{E_\omega}{Z}e^{i(-kz-\omega t)}\right] \equiv 2\,\mathrm{Re}\left(\frac{E_\omega}{Z}e^{-i\omega t}\right)\cos kz\,. \tag{7.62}$$

As a result of this shift, the time average of the Poynting vector's magnitude,

$$S(z,t) = EH = \frac{1}{Z}\mathrm{Re}\left[E_\omega^2 e^{-2i\omega t}\right]\sin 2kz\,, \tag{7.63}$$

equals zero, showing that at the total reflection, there is no *average* power flow. (This is natural because the perfect mirror can neither transmit the wave nor absorb it.) However, Eq. (63) shows that the standing wave features local oscillations of energy, transferring it periodically between the concentrations of the electric and magnetic fields, separated by the distance $\Delta z = \pi/2k = \lambda/4$.

In the case of the sinusoidal waves, the reflection effects may be readily explored even for the more general case of dispersive and/or lossy (but still linear) media in which $\varepsilon(\omega)$ and $\mu(\omega)$, and hence the wave vector $k(\omega)$ and the wave impedance $Z(\omega)$, defined by Eqs. (28), are certain complex functions of frequency. The "only" new factor we have to account for is that in this case, the reflection may not be total, so inside the second medium we have to use the traveling-wave solution as well. This factor may be taken care of by looking for the solution to our boundary problem in the form

$$E\big|_{z\le 0} = \mathrm{Re}\left[E_\omega\left(e^{ik_- z} + R\,e^{-ik_- z}\right)e^{-i\omega t}\right], \qquad E\big|_{z\ge 0} = \mathrm{Re}\left[E_\omega T e^{ik_+ z}e^{-i\omega t}\right], \tag{7.64}$$

Wave's partial reflection

and hence, according to Eq. (6),

$$H\big|_{z\le 0} = \mathrm{Re}\left[\frac{E_\omega}{Z_-(\omega)}\left(e^{ik_- z} - R\,e^{-ik_- z}\right)e^{-i\omega t}\right], \qquad H\big|_{z\ge 0} = \mathrm{Re}\left[\frac{E_\omega}{Z_+(\omega)}T e^{ik_+ z}e^{-i\omega t}\right]. \tag{7.65}$$

(The indices + and – correspond to the media located at $z > 0$ and $z < 0$, respectively.) Please note the following important features of Eqs. (64)-(65):

(i) They satisfy the Maxwell equations in both media. (Historically, the fact that at $z > 0$, these solutions do not include any components proportional to $\exp\{ik_- z\}$, looked surprising and was called the *wave extinction paradox*.)

(ii) Due to the problem's linearity, we could (and did :-) take the complex amplitudes of the reflected and transmitted wave proportional to that (E_ω) of the incident wave, while scaling them with dimensionless, generally complex coefficients R and T. As a comparison of Eqs. (64)-(65) with Eqs. (61)-(62) shows, the total reflection from an ideal mirror corresponds to $R = -1$ and $T = 0$.

(iii) Since in our current problem, the incident wave arrives from one side only (from $z = -\infty$), there is no need to include a term proportional to $\exp\{-ik_+ z\}$ into Eqs. (64)-(65) – even though this term is also a legitimate solution of our wave equation. However, we would need to add such a term if the medium at $z > 0$ had been nonuniform (e.g., had at least one more interface or any other inhomogeneity), because the wave reflected from that additional inhomogeneity would be incident on our interface (located at $z = 0$) from the right.

(iv) Eqs. (64)-(65) may be used even for the description of the cases when waves cannot propagate to $z \ge 0$, for example, a conductor or a plasma with $\omega_p > \omega$. Indeed, the exponential drop of the field amplitude at $z > 0$ in such cases is automatically described by the imaginary part of the wave number k_+ – see Eq. (29).

In order to calculate the coefficients R and T, we need to use boundary conditions at $z = 0$. Since in our current case of the normal incidence, the reflection does not change the transverse character of the partial waves, both vectors **E** and **H** remain tangential to the interface plane (in our notation, $z = 0$). Reviewing the arguments that have led us, in statics, to the boundary conditions (3.37) and (5.117) for these components, we see that they remain valid for the time-dependent situation as well,[28] so for our current case of normal incidence, we may write:

$$E\big|_{z=-0} = E\big|_{z=+0}, \qquad H\big|_{z=-0} = H\big|_{z=+0}. \tag{7.66}$$

Plugging Eqs. (64)-(65) into these conditions, we readily get two equations for the coefficients R and T:

$$1 + R = T, \qquad \frac{1}{Z_-}(1 - R) = \frac{1}{Z_+}T. \tag{7.67}$$

Solving this simple system of linear equations, we get[29]

Reflection and transmission: sharp interface

$$R = \frac{Z_+ - Z_-}{Z_+ + Z_-}, \qquad T = \frac{2Z_+}{Z_+ + Z_-}. \tag{7.68}$$

These formulas are very important, and much more general than one might think because they are applicable for virtually any 1D waves – electromagnetic or not, provided that the impedance Z is defined properly.[30] Since in the general case the wave impedances $Z_\pm$ defined by Eq. (28) with the corresponding indices, are complex functions of frequency, Eqs. (68) show that R and T may have imaginary parts as well. This fact has important consequences at $z < 0$, where the reflected wave, proportional to R, combines ("interferes") with the incident wave. Indeed, with $R = |R|e^{i\varphi}$ (where $\varphi \equiv \arg R$ is a real phase shift), the expression in the parentheses in the first of Eqs. (64) becomes

$$\begin{aligned} e^{ik_- z} + Re^{-ik_- z} &= \left(1 - |R| + |R|\right)e^{ik_- z} + |R|e^{i\varphi}e^{-ik_- z} \\ &\equiv \left(1 - |R|\right)e^{ik_- z} + 2|R|e^{i\varphi/2}\sin\left[k_-(z - \delta_-)\right], \qquad \text{where } \delta_- \equiv \frac{\varphi - \pi}{2k_-}. \end{aligned} \tag{7.69}$$

This means that the field may be represented as a sum of a traveling wave and a standing wave, with an amplitude proportional to $|R|$, shifted by the distance δ_- toward the interface, relative to the ideal-mirror pattern (61b) – see Fig. 8. This effect is frequently used for the experimental measurements of an unknown impedance Z_+ of some medium, provided than Z_- is known – most often, it is the free space, where $Z_- = Z_0$. For that, a small antenna (the *probe*), not disturbing the fields' distribution too much, is placed into the wave field, and the amplitude of the ac voltage induced in it by the wave is measured with a detector (e.g., a semiconductor diode with a nearly-quadratic *I-V* curve), as a function of z (Fig.

[28] For example, the first of Eqs. (66) may be obtained by integrating the full (time-dependent) Maxwell equation $\nabla\times\mathbf{E} + \partial\mathbf{B}/\partial t = 0$ over a narrow and long rectangular contour with dimensions l and d ($d << l$) stretched along the interface. At the application of the Stokes theorem to this integral, the first term gives $\Delta E_\tau l$, while the contribution of the second term is proportional to the product ld, so its contribution at $d/l \to 0$ is negligible. The proof of the second boundary condition is similar – as was already discussed in Sec. 6.2.

[29] Please note that only the media impedances (rather than their wave velocities) are important for reflection in this case! Unfortunately, this fact is not clearly emphasized in some textbooks that discuss only the case $\mu_\pm = \mu_0$, when $Z = (\mu_0/\varepsilon)^{1/2}$ and $v = 1/(\mu_0\varepsilon)^{1/2}$ are proportional to each other.

[30] See, e.g., the discussion of elastic waves of mechanical deformations in CM Secs. 6.3, 6.4, 7.7, and 7.8.

9). From the results of such a measurement, it is straightforward to find both $|R|$ and δ_-, and hence restore the complex R, and then use Eq. (68) to calculate both the modulus and the argument of Z_+. (Before computers became ubiquitous, a specially lined paper called the *Smith chart*, had been frequently used for performing this recalculation graphically; even nowadays, it is still used for result presentation.)

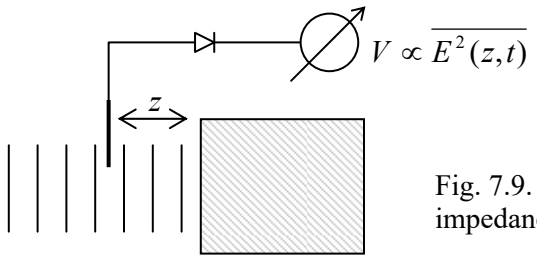


Fig. 7.9. Measurement of the complex impedance of a medium (schematically).

Now let us discuss what these results give for waves incident from the free space ($Z_-(\omega) = Z_0 =$ const, $k_- = k_0 = \omega/c$) onto the surfaces of two particular important media.

(i) For a collision-free plasma (with negligible magnetization) we may use Eq. (36) with $\mu(\omega) = \mu_0$, to represent the impedance (28) in either of two equivalent forms:

$$Z_+ = Z_0 \frac{\omega}{\left(\omega^2 - \omega_p^2\right)^{1/2}} \equiv -iZ_0 \frac{\omega}{\left(\omega_p^2 - \omega^2\right)^{1/2}}. \tag{7.70}$$

The first of these forms is more convenient in the case $\omega > \omega_p$, when the wave vector k_+ and the wave impedance Z_+ of the plasma are real, so a part of the incident wave does propagate into it. Plugging this expression into the latter of Eqs. (68), we see that T is real as well:

$$T = \frac{2\omega}{\omega + \left(\omega^2 - \omega_p^2\right)^{1/2}}. \tag{7.71}$$

Note that according to this formula, and somewhat counter-intuitively, $T > 1$ for any frequency (above ω_p), inviting the question: how can the transmitted wave be more intensive than the incident one that has induced it? To answer this question, we need to compare the powers (rather than the electric field amplitudes) of these two waves, i.e. their average Poynting vectors (42):

$$\overline{S_{\text{incident}}} = \frac{|E_\omega|^2}{2Z_0}, \qquad \overline{S_+} = \frac{|TE_\omega|^2}{2Z_+} = \frac{|E_\omega|^2}{2Z_0} \frac{4\omega\left(\omega^2 - \omega_p^2\right)^{1/2}}{\left[\omega + \left(\omega^2 - \omega_p^2\right)^{1/2}\right]^2}. \tag{7.72}$$

The ratio of these two values[31] is always below 1 (and tends to zero at $\omega \to \omega_p$), so only a fraction of the incident wave power may be transmitted. Hence the result $T > 1$ may be interpreted as follows: an interface between two media may be an *impedance transformer*: it can never transmit more *power* than the incident wave provides, i.e. can only decrease the product $S = EH$, but since the ratio $Z = E/H$ changes at the interface, the amplitude of *one of the fields* may increase at the transmission.

Now let us proceed to case $\omega < \omega_p$ when the waves cannot propagate in the plasma. In this case, the second of the expressions (70) is more convenient, because it immediately shows that Z_+ is purely

[31] This ratio is sometimes also called the "wave transmission coefficient", but to avoid its confusion with the T defined by Eq. (64), it is better to call it the *power transmission coefficient*.

imaginary, while $Z_- = Z_0$ is purely real. This means that $(Z_+ - Z_-) = (Z_+ + Z_-)^*$, i.e. according to the first of Eqs. (68), $|R| = 1$, so the reflection is total, i.e. no incident power (on average) is transferred into the plasma – as was already discussed in Sec. 2. However, the complex R has a finite argument,

$$\varphi \equiv \arg R = 2\arg(Z_+ - Z_0) = -2\tan^{-1}\frac{\omega}{\left(\omega_p^2 - \omega^2\right)^{1/2}}, \tag{7.73}$$

and hence provides a finite spatial shift (69) of the standing wave toward the plasma surface:

$$\delta_- = \frac{\varphi - \pi}{2k_0} = \frac{c}{\omega}\tan^{-1}\frac{\omega}{\left(\omega_p^2 - \omega^2\right)^{1/2}}. \tag{7.74}$$

On the other hand, we already know from Eq. (40) that the solution at $z > 0$ is exponential, with the decay length δ described by Eq. (39). Calculating, from the coefficient T, the exact coefficient before this exponent, it is straightforward to verify that the electric and magnetic fields are indeed continuous at the interface, completing the pattern shown with red lines in Fig. 8. This wave penetration into a fully reflecting material may be experimentally observed, for example, by thinning its sample. Even without solving this problem exactly, it is evident that if the sample's thickness d becomes comparable to δ, a part of the exponential "tail" of the field reaches its second interface, and induces a propagating wave. This is a classical-electromagnetic analog of the quantum-mechanical tunneling through a potential barrier.[32]

Note that at low frequencies, both δ_- and δ tend to the same frequency-independent value,

$$\delta, \delta_- \to \frac{c}{\omega_p} = \left(\frac{c^2\varepsilon_0 m_e}{ne^2}\right)^{1/2} = \left(\frac{m_e}{\mu_0 ne^2}\right)^{1/2}, \qquad \text{at } \frac{\omega}{\omega_p} \to 0, \tag{7.75}$$

which is just the field penetration depth (6.44) calculated for a perfect conductor model (assuming $m = m_e$ and $\mu = \mu_0$) in the quasistatic limit. This is natural, because the condition $\omega << \omega_p$ may be recast as $\lambda_0 \equiv 2\pi c/\omega >> 2\pi c/\omega_p \equiv 2\pi\delta$, i.e. as the quasistatic approximation's validity condition.

(ii) Now let us consider electromagnetic wave's reflection from an Ohmic, non-magnetic conductor. In the simplest low-frequency limit, when $\omega\tau$ is much less than 1, the conductor may be described by a frequency-independent conductivity σ.[33] According to Eq. (46), in this case we can take

$$Z_+ = \left(\frac{\mu_0}{\varepsilon_{opt}(\omega) + i\sigma/\omega}\right)^{1/2}. \tag{7.76}$$

With this substitution, Eqs. (68) immediately give us all the results of interest. In particular, in the most important quasistatic limit when $\delta_s \equiv (2/\mu_0\sigma\omega)^{1/2} << \lambda_0 \equiv 2\pi c/\omega$, i.e. $\sigma/\omega >> \varepsilon_0 \sim \varepsilon_{opt}$, the conductor' impedance is low:

$$Z_+ \approx \left(\frac{\mu_0\omega}{i\sigma}\right)^{1/2} \equiv \pi\left(\frac{2}{i}\right)^{1/2}\frac{\delta_s}{\lambda_0}Z_0, \qquad \text{i.e. } \left|\frac{Z_+}{Z_0}\right| << 1. \tag{7.77}$$

[32] See, e.g., QM Sec. 2.3.

[33] In a typical metal, $\tau \sim 10^{-13}$s, so this approximation works well up to $\omega \sim 10^{13}$ s^{-1}, i.e. up to the far-infrared frequencies.

This impedance is complex, and hence some fraction $\mathscr{f}$ of the incident wave is absorbed by the conductor. The fraction may be found as the ratio of the dissipated power (either calculated, as was done above, from Eqs. (68), or just taken from Eq. (6.36), with the magnetic field amplitude $|H_\omega| = 2|E_\omega|/Z_0$ – see Eq. (62)) to the incident wave's power given by the first of Eqs. (72). The result,

$$\mathscr{f} = \frac{2\omega\delta_s}{c} \equiv 4\pi\frac{\delta_s}{\lambda_0} << 1. \tag{7.78}$$

is used for crude estimates of the energy dissipation in metallic-wall waveguides and resonators. It shows that to keep the energy losses low, the characteristic size of such systems (which gives a scale of the free-space wavelengths λ_0 at which they are used) should be much larger than δ_s. A more detailed theory of these structures, and of the energy loss in them, will be discussed later in this chapter.

7.4. Refraction

Now let us consider the effects arising at a plane interface between two uniform media when the wave's incidence angle θ (Fig. 10) is arbitrary rather than equal to zero as in our previous analysis, for the simplest case of fully transparent media, with real $\varepsilon_\pm(\omega)$ and $\mu_\pm(\omega)$. (For the sake of notation simplicity, in most formulas below, the argument of these functions will be dropped, i.e. just implied.)

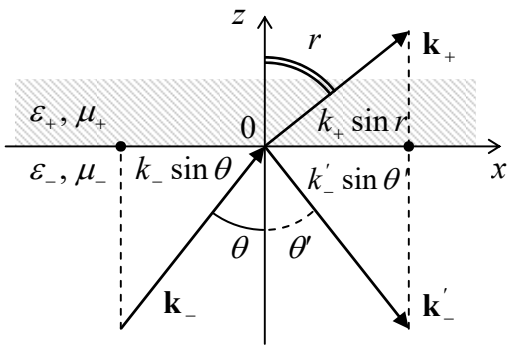


Fig. 7.10. Plane wave's reflection, transmission, and refraction at a plane interface. The plane of the drawing is selected to contain all three wave vectors: $\mathbf{k}_+$, $\mathbf{k}_-$, and $\mathbf{k}'_-$.

In contrast with the case of normal incidence, here the wave vectors $\mathbf{k}_-$, $\mathbf{k}'_-$, and $\mathbf{k}_+$ of the three components (incident, reflected, and transmitted) waves may have different directions. (Such change of the transmitted wave's direction is called *refraction*.) Hence let us start our analysis by writing a general expression for a single plane, monochromatic wave for the case when its wave vector $\mathbf{k}$ has all three Cartesian components, rather than one. An evident generalization of Eq. (11) for this case is

$$f(\mathbf{r},t) = \mathrm{Re}\left[f_\omega e^{i\left(k_x x + k_y y + k_z z\right) - \omega t)}\right] \equiv \mathrm{Re}\left[f_\omega e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}\right]. \tag{7.79}$$

This expression enables a ready analysis of "kinematic" relations, which are independent of the media impedances. Indeed, it is sufficient to notice that to satisfy *any* linear, homogeneous boundary conditions at the interface ($z = 0$), all partial plane waves must have the same temporal and spatial dependence on this plane. Hence if we select the x-z plane so that the vector $\mathbf{k}_-$ lies in it, then $(k_-)_y = 0$, and $\mathbf{k}_+$ and $\mathbf{k}'_-$ cannot have any y-component either, i.e. all three wave vectors lie in the same plane –

that is selected as the plane of the drawing in Fig. 10. Moreover, because of the same reason, their x-components should be equal:

$$k_- \sin\theta = k_-' \sin\theta' = k_+ \sin r\,. \tag{7.80}$$

From here we immediately get two well-known laws: of reflection

Reflection angle

$$\theta' = \theta\,, \tag{7.81}$$

and of refraction:[34]

Snell law

$$\frac{\sin r}{\sin\theta} = \frac{k_-}{k_+}\,. \tag{7.82}$$

In this form, the laws are valid for plane waves of any nature. In optics, the Snell law (82) is frequently represented in the form

$$\frac{\sin r}{\sin\theta} = \frac{n_-}{n_+}\,, \tag{7.83}$$

where $n_\pm$ is the *index of refraction* (also called the "refractive index") of the corresponding medium, defined as its wave number normalized to that of the free space (at the particular wave's frequency):

Index of refraction

$$n_\pm \equiv \frac{k_\pm}{k_0} \equiv \left(\frac{\varepsilon_\pm \mu_\pm}{\varepsilon_0 \mu_0}\right)^{1/2}\,. \tag{7.84}$$

Perhaps the most famous corollary of the Snell law is that if a wave propagates from a medium with a higher index of refraction to that with a lower one (i.e. if $n_- > n_+$ in Fig. 10), for example from water to air, there is always a certain *critical* value θ_c of the incidence angle,

Critical angle

$$\theta_c = \sin^{-1}\frac{n_+}{n_-} \equiv \sin^{-1}\left(\frac{\varepsilon_+ \mu_+}{\varepsilon_- \mu_-}\right)^{1/2}\,, \tag{7.85}$$

at which the refraction angle r (see Fig. 10 again) reaches $\pi/2$. At a larger θ, i.e. within the range $\theta_c < \theta < \pi/2$, the boundary conditions (80) cannot be satisfied by a refracted wave with a real wave vector, so the wave experiences the so-called *total internal reflection*. This effect is very important for practice because it means that dielectric surfaces may be used as optical mirrors, in particular in optical fibers – to be discussed in more detail in Sec. 7 below. This is very fortunate for telecommunication technology because light's reflection from metals is rather imperfect. Indeed, according to Eq. (78), in the optical range ($\lambda_0 \sim 0.5$ μm, i.e. $\omega \sim 10^{15}$ s^{-1}), even the best conductors (with $\sigma \sim 6\times10^8$ S/m and hence the normal skin depth $\delta_s \sim 1.5$ nm) provide power loss of at least a few percent at each reflection.

Note, however, that even within the range $\theta_c < \theta < \pi/2$, the field at $z > 0$ is not identically equal to zero: it penetrates into the lower-n media by a distance of the order of λ_0, exponentially decaying inside it, just as it does at the normal incidence – see Fig. 8. However, at $\theta \neq 0$ the penetrating field still propagates, with the wave number (80), along the interface. Such a field, exponentially dropping in one direction but still propagating as a wave in another direction, is commonly called the *evanescent wave*.

[34] The latter relation is traditionally called the *Snell law*, after a 17th-century astronomer Willebrord Snellius, but it has been traced all the way back to a circa 984 work by Abu Saad al-Ala ibn Sahl. (Claudius Ptolemy who performed pioneering experiments on light refraction in the 2nd century AD, was just one step from this result.)

One more remark: just as at the normal incidence, the field's penetration into another medium causes a phase shift of the reflected wave – see, e.g., Eq. (69) and its discussion. A new feature of this phase shift, arising at $\theta \neq 0$, is that it also has a component parallel to the interface – the so-called *Goos-Hänchen effect*. In geometric optics, this effect leads to an image shift (relative to its position in a perfect mirror) with components both normal and parallel to the interface.

Now let us carry out an analysis of "dynamic" relations that determine amplitudes of the refracted and reflected waves. For this, we need to write explicitly the boundary conditions at the interface (i.e. the plane $z = 0$). Since now the electric and/or magnetic fields may have components normal to the plane, in addition to the continuity of their tangential components, which were repeatedly discussed above,

$$E_{x,y}\big|_{z=-0} = E_{x,y}\big|_{z=+0}, \qquad H_{x,y}\big|_{z=-0} = H_{x,y}\big|_{z=+0}, \tag{7.86}$$

we also need relations for the normal components. As it follows from the homogeneous macroscopic Maxwell equations (6.99b), they are also the same as in statics, i.e. $D_n = \text{const}$, and $B_n = \text{const}$, for our coordinate choice (Fig. 10) giving

$$\varepsilon_- E_z\big|_{z=-0} = \varepsilon_+ E_z\big|_{z=+0}, \qquad \mu_- H_z\big|_{z=-0} = \mu_+ H_z\big|_{z=+0}. \tag{7.87}$$

The expressions of these components via the amplitudes E_ω, RE_ω, and TE_ω of the incident, reflected, and transmitted waves depend on the incident wave's polarization. For example, for a linearly-polarized wave with the electric field vector *normal* to the plane of incidence, i.e. *parallel* to the interface plane, the reflected and refracted waves are similarly polarized – see Fig. 11a.

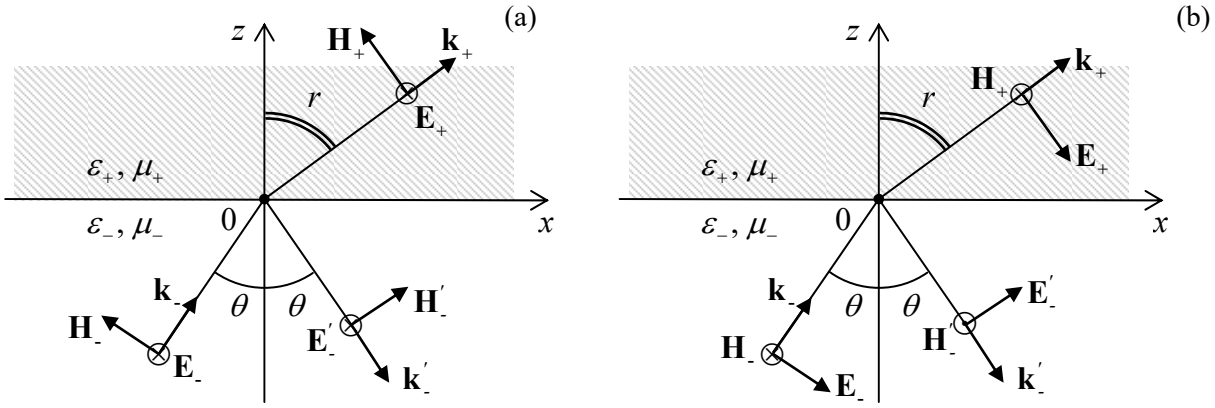


Fig. 7.11. Reflection and refraction at two different linear polarizations of the incident wave.

As a result, all E_z are equal to zero (so the first of Eqs. (87) is inconsequential), while the tangential components of the electric field are equal to their full amplitudes, just as at the normal incidence, so we still can use Eqs. (64) expressing these components via the coefficients R and T. However, at $\theta \neq 0$ the magnetic fields have not only tangential components

$$H_x\big|_{z=-0} = \mathrm{Re}\left[\frac{E_\omega}{Z_-}(1-R)\cos\theta\, e^{-i\omega t}\right], \qquad H_x\big|_{z=+0} = \mathrm{Re}\left[\frac{E_\omega}{Z_+}T\cos r\; e^{-i\omega t}\right], \tag{7.88}$$

but also normal components (see Fig. 11a):

$$H_z\big|_{z=-0} = \mathrm{Re}\left[\frac{E_\omega}{Z_-}(1+R)\sin\theta\ e^{-i\omega t}\right], \qquad H_z\big|_{z=+0} = \mathrm{Re}\left[\frac{E_\omega}{Z_+}T\sin r\ e^{-i\omega t}\right]. \tag{7.89}$$

Plugging these expressions into the boundary conditions expressed by Eqs. (86) (in this case, for the *y*-components only) and the second of Eqs. (87), we get *three* equations for *two* unknown coefficients *R* and *T*. However, two of these equations duplicate each other because of the Snell law, and we get just two independent equations,

$$1+R=T, \qquad \frac{1}{Z_-}(1-R)\cos\theta = \frac{1}{Z_+}T\cos r, \tag{7.90}$$

which are a very natural generalization of Eqs. (67), with the replacements $Z_- \to Z_-\cos r$, $Z_+ \to Z_+\cos\theta$. As a result, we can immediately use Eq. (68) to write the solution of the system (90):[35]

$$R = \frac{Z_+\cos\theta - Z_-\cos r}{Z_+\cos\theta + Z_-\cos r}, \qquad T = \frac{2Z_+\cos\theta}{Z_+\cos\theta + Z_-\cos r}. \tag{7.91a}$$

If we want to express these coefficients via the angle of incidence alone, we should use the Snell law (82) to eliminate the angle *r*, getting frequently quoted bulkier expressions:

$$R = \frac{Z_+\cos\theta - Z_-\left[1-(k_-/k_+)^2\sin^2\theta\right]^{1/2}}{Z_+\cos\theta + Z_-\left[1-(k_-/k_+)^2\sin^2\theta\right]^{1/2}}, \qquad T = \frac{2Z_+\cos\theta}{Z_+\cos\theta + Z_-\left[1-(k_-/k_+)^2\sin^2\theta\right]^{1/2}}. \tag{7.91b}$$

However, conceptually it is preferable to use the kinematic relation (82) and the dynamic relations (91a) separately, because Eq. (91b) obscures the very important physical fact that the ratio of $k_\pm$, i.e. of the wave velocities of the two media, is only involved in the Snell law, while Eqs. (91b) explicitly include only the wave impedances – just as in the case of normal incidence.

In the opposite case of the linear polarization of the electric field within the plane of incidence (Fig. 11b), it is the magnetic field that does not have a normal component, so it is now the second of Eqs. (87) that does not participate in the solution. However, now the electric fields in the two media have not only tangential components,

$$E_x\big|_{z=-0} = \mathrm{Re}\left[E_\omega(1+R)\cos\theta\ e^{-i\omega t}\right], \qquad E_x\big|_{z=+0} = \mathrm{Re}\left[E_\omega T\cos r\ e^{-i\omega t}\right], \tag{7.92}$$

but also normal components (Fig. 11b):

$$E_z\big|_{z=-0} = E_\omega(-1+R)\sin\theta, \qquad E_z\big|_{z=+0} = -E_\omega T\sin r. \tag{7.93}$$

As a result, instead of Eqs. (90), the reflection and transmission coefficients are related as

$$(1+R)\cos\theta = T\cos r, \qquad \frac{1}{Z_-}(1-R) = \frac{1}{Z_+}T. \tag{7.94}$$

Again, the solution of this system may be immediately written using the analogy with Eq. (67):

[35] Note that we may calculate the reflection and transmission coefficients *R'* and *T'* for the wave traveling in the opposite direction just by making the following parameter swaps: $Z_+ \leftrightarrow Z_-$ and $\theta \leftrightarrow r$, and that the resulting coefficients satisfy the following *Stokes relations*: $R' = -R$, and $R^2 + TT' = 1$, for any $Z_\pm$.

$$R = \frac{Z_+ \cos r - Z_- \cos\theta}{Z_+ \cos r + Z_- \cos\theta}, \qquad T = \frac{2Z_+ \cos\theta}{Z_+ \cos r + Z_- \cos\theta}, \tag{7.95a}$$

or, alternatively, using the Snell law, in a more bulky form:

$$R = \frac{Z_+\left[1-(k_-/k_+)^2\sin^2\theta\right]^{1/2} - Z_-\cos\theta}{Z_+\left[1-(k_-/k_+)^2\sin^2\theta\right]^{1/2} + Z_-\cos\theta}, \qquad T = \frac{2Z_+\cos\theta}{Z_+\left[1-(k_-/k_+)^2\sin^2\theta\right]^{1/2} + Z_-\cos\theta}. \tag{7.95b}$$

For the particular case $\mu_+ = \mu_- = \mu_0$, when $Z_+/Z_- = (\varepsilon_-/\varepsilon_+)^{1/2} = k_-/k_+ = n_-/n_+$ (which is approximately correct for traditional optical media), Eqs. (91b) and (95b) are called the *Fresnel formulas*.[36] Most textbooks are quick to point out that there is a major difference between them: while for the electric field polarization within the plane of incidence (Fig. 11b), the reflected wave's amplitude (proportional to the coefficient R) turns to zero[37] at a special value of θ (called the *Brewster angle*):[38]

$$\theta_B = \tan^{-1}\frac{n_+}{n_-}, \tag{7.96}$$

while there is no such angle in the opposite case (shown in Fig. 11a). However, note that this statement, as well as Eq. (96), is true only for the case $\mu_+ = \mu_-$. In the general case of different ε and μ, Eqs. (91) and (95) show that the reflected wave vanishes at $\theta = \theta_B$ with

$$\tan^2\theta_B = \frac{\varepsilon_-\mu_+ - \varepsilon_+\mu_-}{\varepsilon_+\mu_+ - \varepsilon_-\mu_-} \times \begin{cases} (\mu_+/\mu_-), & \text{for } \mathbf{E} \perp \mathbf{n}_z \text{ (Fig.11a)}, \\ (-\varepsilon_+/\varepsilon_-), & \text{for } \mathbf{H} \perp \mathbf{n}_z \text{ (Fig.11b)}. \end{cases} \tag{7.97}$$

Brewster angle

Note the natural $\varepsilon \leftrightarrow \mu$ symmetry of these relations, resulting from the $\mathbf{E} \leftrightarrow \mathbf{H}$ symmetry for these two polarization cases (Fig. 11). These formulas also show that for any set of parameters of the two media (with $\varepsilon_\pm$, $\mu_\pm > 0$), $\tan^2\theta_B$ is positive (and hence a real Brewster angle θ_B exists) only for one of these two polarizations. In particular, if the interface is due to the change of μ alone (i.e. if $\varepsilon_+ = \varepsilon_-$), the first of Eqs. (97) is reduced to the simple form (96) again, while for the polarization shown in Fig. 11b, there is no Brewster angle, i.e. the reflected wave has a non-zero amplitude for any θ.

Such an account of both media parameters, ε and μ, on an equal footing is necessary to describe several interesting effects. The first of them is the so-called *negative refraction*.[39] As was shown in Sec.

[36] Named after Augustin-Jean Fresnel (1788-1827), one of the wave optics pioneers, who is credited, among many other contributions (see, in particular, discussions in Ch. 8), for the concept of light as a purely transverse wave.

[37] This effect is used in practice to obtain linearly polarized light, with the electric field vector perpendicular to the plane of incidence, from the natural light with its random polarization. An even more widespread application of the effect is a partial reduction of undesirable glare from wet pavement (for the water/air interface, $n_+/n_- \approx 1.33$, giving $\theta_B \approx 50°$) by covering glasses and car headlights with thin vertically-polarizing layers.

[38] A very simple interpretation of Eq. (96) is based on the fact that, together with the Snell law (82), it gives $r + \theta = \pi/2$. As a result, the vector $\mathbf{E}_+$ is parallel to the vector $\mathbf{k}'_-$, and hence the oscillating electric dipoles of the medium do not have the Cartesian component that could induce the transverse electric field $\mathbf{E}'_-$ of the potential reflected wave.

[39] Despite some important background theoretical work by A. Schuster (1904), L. Mandelstam (1945), D. Sivikhin (1957), and especially V. Veselago (1966-67), the negative refractivity effects became a subject of intensive scientific research and engineering development only in the 2000s.

2, in a medium with electric-field-driven resonances, the function $\varepsilon(\omega)$ may be almost real and negative, at least within limited frequency intervals – see, in particular, Eq. (34) and Fig. 5. As has already been discussed, if, at these frequencies, the function $\mu(\omega)$ is real and positive, then $k^2(\omega) = \omega^2\varepsilon(\omega)\mu(\omega) < 0$, and k may be represented as i/δ with a real δ, meaning the exponential field decay into the medium. However, let us consider the case when both $\varepsilon(\omega) < 0$ *and* $\mu(\omega) < 0$ at a certain frequency. (This is possible in a medium with both **E**-driven and **H**-driven resonances, at a proper choice of their resonance frequencies.) Since in this case $k^2(\omega) = \omega^2\varepsilon(\omega)\mu(\omega) > 0$, the wave vector is real, Eq. (79) describes a traveling wave, and one could think that there is nothing new in this case. Not so!

First of all, for a sinusoidal plane wave (79), the operator ∇ is equivalent to the multiplication by $i\mathbf{k}$. As the Maxwell equations (2a) show, this means that at a fixed direction of vectors **E** and **k**, the simultaneous reversal of signs of ε and μ means the reversal of the direction of the vector **H**. Namely, if both ε and μ are positive, these equations are satisfied with mutually orthogonal vectors {**E**, **H**, **k**} forming the usual, *right-hand* system (see Fig. 1 and Fig. 12a), the name stemming from the popular "right-hand rule" used to determine the vector product's direction. However, if both ε and μ are negative, the vectors form a *left-hand* system – see Fig. 12b. (Due to this fact, the media with $\varepsilon < 0$ and $\mu < 0$ are frequently called the *left-handed materials*, LHM for short.) According to the basic relation (6.114), which does not involve media parameters, this means that for a plane wave in a left-hand material, the Poynting vector $\mathbf{S} = \mathbf{E}\times\mathbf{H}$, i.e. the energy flow, is directed *opposite* to the wave vector **k**.

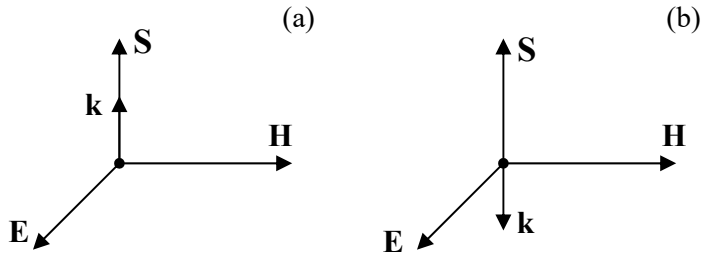


Fig. 7.12. Directions of the main vectors of a plane wave inside a medium with (a) positive and (b) negative values of ε and μ.

This fact may look strange but is in no contradiction with any fundamental principle. Let me remind you that, according to the definition of the vector **k**, its direction shows the direction of the *phase* velocity $v_{ph} = \omega/k$ of a sinusoidal (and hence infinitely long) wave, which cannot be used, for example, for signaling. Such signaling (by sending wave packets – see Fig. 13) is possible only with the *group* velocity $v_{gr} = d\omega/dk$. This velocity in left-hand materials is always directed (as in the right-hand materials) along the vector **S**, i.e. along the wave's energy flow.

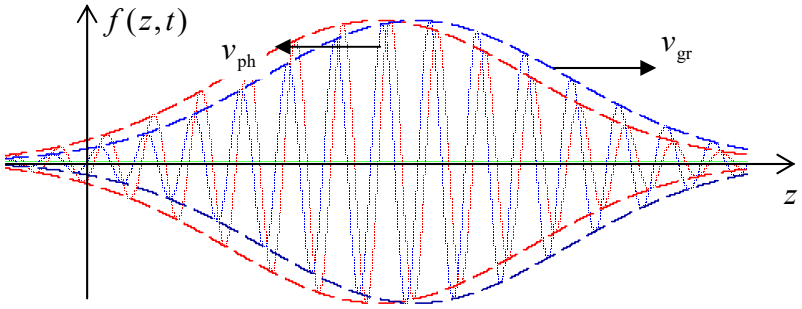


Fig. 7.13. An example of a wave packet moving along axis z with a negative phase velocity, but positive group velocity. Blue lines show a packet's snapshot a short time interval after the first snapshot (red lines).

Perhaps the most fascinating effect possible with left-hand materials is the wave refraction at their interfaces with the usual, right-handed materials – first predicted by V. Veselago in 1960. Consider the example shown in Fig. 14a. In the incident wave, arriving from a usual material, the directions of the vectors $\mathbf{k}_-$ and $\mathbf{S}_-$ coincide, and so they are in the reflected wave with vectors $\mathbf{k}'_-$ and $\mathbf{S}'_-$. This means that the electric and magnetic fields in the interface plane ($z = 0$) are, at our choice of the coordinate axes, proportional to $\exp\{ik_x x\}$, with a positive component $k_x = k_- \cos\theta$. To satisfy any linear boundary conditions, the refracted wave, propagating into the left-handed material, has to match that dependence, i.e. have a positive x-component of its wave vector $\mathbf{k}_+$. But in this medium, this vector has to be antiparallel to the vector $\mathbf{S}$, which in turn should be directed out of the interface, because it represents the power flow from the interface into the material's bulk. These conditions cannot be reconciled by the refracted wave propagating along the usual Snell-law direction (shown with the dashed line in Fig. 13a), but are all satisfied at refraction in the direction given by Snell's angle with the opposite sign. (Hence the term "negative refraction").[40]

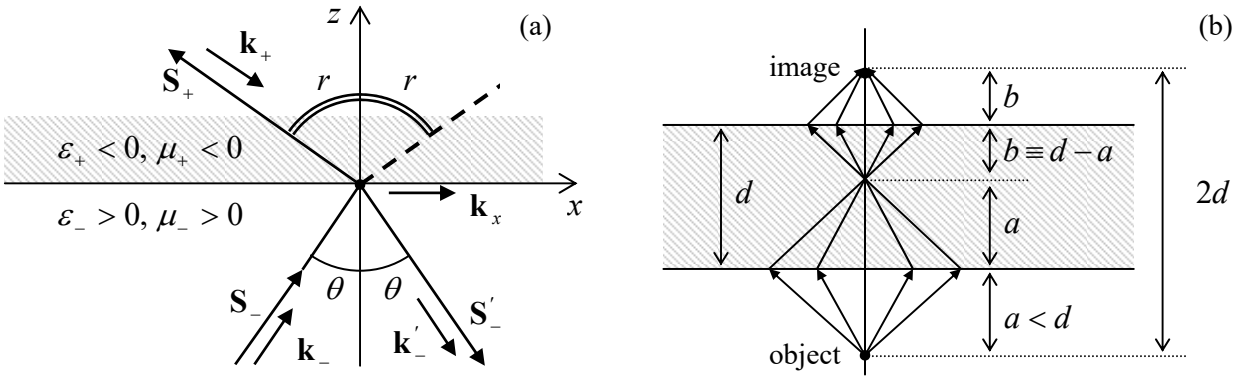


Fig. 7.14. Negative refraction: (a) waves at the interface between media with positive and negative values of $\varepsilon\mu$, and (b) the hypothetical *perfect lens*: a parallel plate made of a material with $\varepsilon = -\varepsilon_0$ and $\mu = -\mu_0$.

In order to understand how unusual the results of the negative refraction may be, let us consider a parallel slab of thickness d, made of a hypothetical left-handed material with exactly selected values $\varepsilon = -\varepsilon_0$, and $\mu = -\mu_0$ (see Fig. 14b). For such a material, placed in free space, the refraction angle $r = -\theta$, so the rays from a point source, located in free space at a distance $a < d$ from the slab's surface, propagate as shown on that panel, i.e. all meet again at the distance a beyond the surface, and then continue to propagate to the second surface of the slab. Repeating our discussion for this surface, we see that a point's image is also formed beyond the slab, at distance $2a + 2b = 2a + 2(d - a) = 2d$ from the object.

Superficially, this system looks like a usual lens, but the well-known lens formula, which relates a and b with the focal length f, is *not* satisfied. (In particular, a parallel beam is *not* focused into a point at any finite distance.) As an additional difference from the usual lens, the system shown in Fig. 14b *does not reflect* any part of the incident light. Indeed, it is straightforward to check that for all the above formulas for R and T to be valid, the sign of the wave impedance Z in left-handed materials has to be kept positive. Thus, for our particular choice of parameters ($\varepsilon = -\varepsilon_0$, $\mu = -\mu_0$), Eqs. (91a) and (95a) are

[40] In some publications inspired by this fact, the left-hand materials are prescribed a negative index of refraction n. However, this prescription should be treated with care. For example, it complies with the first form of Eq. (84), but not its second form, and the sign of n, in contrast to that of the wave vector $\mathbf{k}$, is a matter of convention.

valid with $Z_+ = Z_- = Z_0$ and $\cos r = \cos \theta = 1$, giving $R = 0$ for any linear polarization, and hence for any other wave polarization – circular, elliptic, natural, etc.

The perfect lens suggestion has triggered a wave of efforts to implement left-hand materials experimentally. (Attempts to find such materials in nature have failed so far.) Most progress in this direction has been achieved using the so-called *metamaterials*, which are essentially quasi-periodic arrays of specially designed electromagnetic resonators, ideally with high density $n >> \lambda^{-3}$. For example, Fig. 15 shows the metamaterial that was used for the first demonstration of negative refractivity in the microwave region – for ~10-GHz waves.[41] It combines straight strips of a metallic film, working as lumped resonators with a large electric dipole moment (and hence strongly coupled to the wave's electric field **E**), and several almost-closed film loops (so-called *split rings*), working as lumped resonators with large magnetic dipole moments, strongly coupled to the field **H**. The negative refractivity is achieved by designing the resonance frequencies close to each other. More recently, metamaterials with negative refractivity were demonstrated in the optical range as well,[42] although to the best of my knowledge, their relatively large absorption still prevents practical applications.

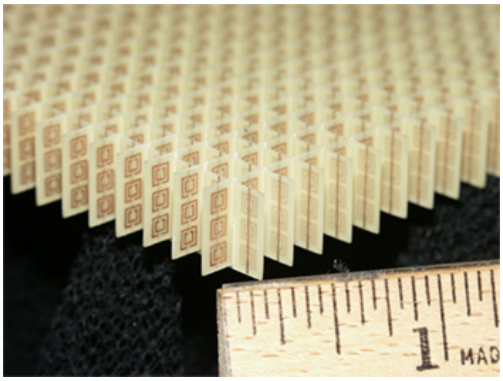


Fig. 7.15. An artificial left-hand material providing negative refraction at microwave frequencies ~10 GHz. The original by Jeffrey D. Wilson (in the public domain) is available at https://en.wikipedia.org/wiki/Metamaterial.

This progress has stimulated the development of other potential uses of metamaterials (not necessarily the left-handed ones), in particular, designs of nonuniform systems with handcrafted distributions $\varepsilon(\mathbf{r}, \omega)$ and $\mu(\mathbf{r}, \omega)$ that may provide electromagnetic wave propagation along the desired paths, e.g., around a certain region of space, making it virtually invisible for an external observer – so far, within very limited frequency ranges.[43]

As was mentioned in Sec. 5.5, another way to reach negative values of $\mu(\omega)$ is to place a ferromagnetic material into such an external dc magnetic field that the frequency ω_r of the ferromagnetic resonance is somewhat lower than ω. If thin layers of such a material (e.g., nickel) are interleaved with layers of a non-magnetic good conductor (such as copper), the average value of $\mu(\omega)$ of the resulting metamaterial may be positive but substantially below μ_0. According to Eq. (6.33), the skin-depth δ_s of such a material may be larger than that of the good conductor alone, enforcing a more uniform distribution of the ac current flowing along the layers, and hence making the energy losses lower than in the good conductor alone. This effect may be useful, in particular, for electronic circuit interconnects.[44]

[41] R. Shelby *et al.*, Science **292**, 77 (2001); J. Wilson and Z. Schwartz, *Appl. Phys. Lett.* **86**, 021113 (2005).
[42] See, e.g., J. Valentine *et al.*, *Nature* **455**, 376 (2008).
[43] For a review of such "invisibility cloaks", see, e.g., B. Wood, *Comptes Rendus Physique* **10**, 379 (2009).
[44] See, for example, N. Sato *et al.*, *J. Appl. Phys.* **111**, 07A501 (2012), and references therein.

7.5. Transmission lines: TEM waves

So far, we have analyzed plane electromagnetic waves, implying that their cross-section is infinite – evidently, an unrealistic assumption. The cross-section may be limited, still sustaining wave propagation, using *wave transmission lines*:[45] long, uniform structures made of either good conductors or dielectrics. Let us first discuss the first option, using the following simplifying assumptions:

(i) the structure is a cylinder (not necessarily with a round cross-section, see Fig. 16) filled with a usual (right-handed), uniform dielectric material with negligible energy losses ($\varepsilon'' = \mu'' = 0$), and

(ii) the wave attenuation due to the skin effect is also negligibly low. (As Eq. (78) indicates, for that the characteristic size a of the line's cross-section has to be much larger than the skin-depth δ_s of its wall material. The energy dissipation effects will be analyzed in Sec. 9 below.)

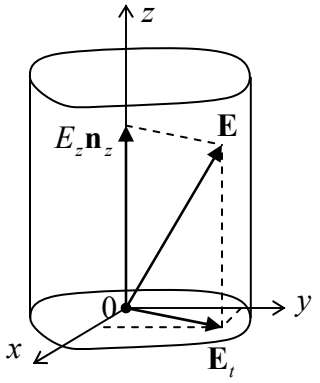


Fig. 7.16. Electric field's decomposition in a transmission line (in particular, a waveguide).

With such exclusion of energy losses, we may look for a particular solution of the macroscopic Maxwell equations in the form of a monochromatic wave traveling along the line:

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\left[\mathbf{E}_\omega(x,y)e^{i(k_z z-\omega t)}\right], \qquad \mathbf{H}(\mathbf{r},t) = \mathrm{Re}\left[\mathbf{H}_\omega(x,y)e^{i(k_z z-\omega t)}\right], \tag{7.98}$$

with real k_z, where the z-axis is directed along the transmission line – see Fig. 16. Note that this form allows a substantial coordinate dependence of the electric and magnetic field within the plane [x, y] of the transmission line's cross-section, as well as nonvanishing longitudinal components E_z and/or H_z of the fields, so the solution (98) is substantially more general than the plane waves discussed above. We will see in a minute that as a result, the parameter k_z may be very much different from its plane-wave value (13), $k \equiv \omega(\varepsilon\mu)^{1/2}$, in the same material, at the same frequency.

In order to describe these effects quantitatively, let us decompose the complex amplitudes of the wave's fields into their longitudinal and transverse components (Fig. 16):[46]

$$\mathbf{E}_\omega = E_z\mathbf{n}_z + \mathbf{E}_t, \quad \mathbf{H}_\omega = H_z\mathbf{n}_z + \mathbf{H}_t. \tag{7.99}$$

[45] Another popular term is the *waveguide*, but it is typically reserved for the transmission lines with singly-connected cross-sections, to be analyzed in the next section. The first structure for guiding waves was proposed by J. J. Thomson in 1893, and experimentally tested by O. Lodge in 1894.

[46] For the notation simplicity, I am dropping index ω in the complex amplitudes of the field components, and also have dropped the argument ω in k_z and Z, even though these parameters may depend on the wave's frequency rather substantially – see below.

Plugging Eqs. (98)-(99) into the source-free Maxwell equations (2), and requiring the longitudinal and transverse components to be balanced separately, we get

$$\begin{aligned} &ik_z \mathbf{n}_z \times \mathbf{E}_t - i\omega\mu \mathbf{H}_t = -\nabla_t \times (E_z \mathbf{n}_z), && ik_z \mathbf{n}_z \times \mathbf{H}_t + i\omega\varepsilon \mathbf{E}_t = -\nabla_t \times (H_z \mathbf{n}_z), \\ &\nabla_t \times \mathbf{E}_t = i\omega\mu H_z \mathbf{n}_z, && \nabla_t \times \mathbf{H}_t = -i\varepsilon\omega E_z \mathbf{n}_z, \\ &\nabla_t \cdot \mathbf{E}_t = -ik_z E_z, && \nabla_t \cdot \mathbf{H}_t = -ik_z H_z. \end{aligned} \tag{7.100}$$

where ∇_t is the 2D del operator acting in the transverse plane $[x, y]$ only, i.e. the usual ∇, but with $\partial/\partial z = 0$. The system (100) looks even bulkier than the original equations (2), but it is much simpler for analysis. Indeed, by eliminating the transverse components from these equations (or, even simpler, just by plugging Eq. (99) into Eqs. (3) and keeping only their z-components), we get a pair of self-consistent equations for the longitudinal components of the fields,[47]

2D Helmholtz equations for E_z and H_z

$$\left(\nabla_t^2 + k_t^2\right) E_z = 0, \qquad \left(\nabla_t^2 + k_t^2\right) H_z = 0, \tag{7.101}$$

where k is still defined by Eq. (13), $k \equiv (\varepsilon\mu)^{1/2}\omega$, while

Wave vector component balance

$$k_t^2 \equiv k^2 - k_z^2 = \omega^2 \varepsilon\mu - k_z^2. \tag{7.102}$$

After the distributions $E_z(x,y)$ and $H_z(x,y)$ have been found from these equations, they provide right-hand sides for the rather simple, closed system of equations (100) for the transverse components of field vectors. Moreover, as we will see below, each of the following three types of solutions:

(i) with $E_z = 0$ and $H_z = 0$ (called the *transverse electromagnetic*, or *TEM waves*),

(ii) with $E_z = 0$, but $H_z \neq 0$ (called either the *TE waves* or, more frequently, *H-modes*), and

(iii) with $E_z \neq 0$, but $H_z = 0$ (the so-called *TM waves* or *E-modes*),

has its own dispersion law and hence its own wave propagation velocity; as a result, these *modes* (i.e. the field distribution patterns) may be considered separately.

In the balance of this section, we will focus on the simplest, TEM waves (i), with *no longitudinal components* of either field. For them, the top two equations of the system (100) immediately give Eqs. (6) and (13), and $k_z = k$. In plain English, this means that $\mathbf{E} = \mathbf{E}_t$ and $\mathbf{H} = \mathbf{H}_t$ are proportional to each other and are mutually perpendicular (just as in the plane wave) at each point of the cross-section and that the TEM wave's impedance $Z \equiv E/H$ and dispersion law $\omega(k)$, and hence the propagation speed, are the same as in a plane wave in the same material. In particular, if ε and μ are frequency-independent within a certain frequency range, the dispersion law within this range is linear, $\omega = k/(\varepsilon\mu)^{1/2}$, and the wave's speed does not depend on its frequency. For practical applications to telecommunications, this is a very important advantage of the TEM waves over their TM and TE counterparts – to be discussed in the next sections.

Unfortunately for practice, such waves cannot propagate in every transmission line. To show this, let us have a look at the two last lines of Eqs. (100). For the TEM waves ($E_z = 0$, $H_z = 0$, $k_z = k$), they are reduced to merely

[47] The wave equation represented in the form (101), even with the 3D Laplace operator, is called the *Helmholtz equation*, named after Hermann von Helmholtz (1821-1894) – the mentor of H. Hertz and M. Planck, among many others.

$$\nabla_t \times \mathbf{E}_t = 0, \qquad \nabla_t \times \mathbf{H}_t = 0,$$
$$\nabla_t \cdot \mathbf{E}_t = 0, \qquad \nabla_t \cdot \mathbf{H}_t = 0. \tag{7.103}$$

Within the coarse-grain description of the conducting walls of the line (i. e., neglecting not only the screening depth but also the skin depth in comparison with the cross-section dimensions), we have to require that inside them, $\mathbf{E} = \mathbf{H} = 0$. Close to a wall but outside it, the normal component E_n of the electric field may be different from zero, because surface charges may sustain its jump – see Sec. 2.1, in particular Eq. (2.3). Similarly, the tangential component H_τ of the magnetic field may have a finite jump at the surface due to skin currents – see Sec. 6.3, in particular Eq. (6.38). However, the tangential component of the electric field and the normal component of the magnetic field cannot experience such jumps, and to have them equal to zero inside the walls they have to equal zero just outside the walls as well:

$$\mathbf{E}_\tau = 0, \qquad H_n = 0\,. \tag{7.104}$$

But the left columns of Eqs. (103)-(104) coincide with the formulation of the 2D boundary problem of electrostatics for the electric field induced by electric charges of the conducting walls, with the only difference that in our current case, the value of ε actually means $\varepsilon(\omega)$. Similarly, the right columns of those relations coincide with the formulation of the 2D boundary problem of magnetostatics for the magnetic field induced by currents in the walls, with $\mu \to \mu(\omega)$, with the only difference is that in our current coarse-grain approximation, the magnetic fields cannot penetrate into the conductors.

Now we immediately see that in waveguides with a singly-connected wall, for example, a hollow conducting tube (see, e.g., Fig. 16), the TEM waves are impossible, because there is no way to create a non-zero electrostatic field inside a conductor with such cross-section. However, such fields (and hence the TEM waves) are possible in structures with cross-sections consisting of two or more disconnected (galvanically-insulated) parts – see, e.g., Fig. 17.

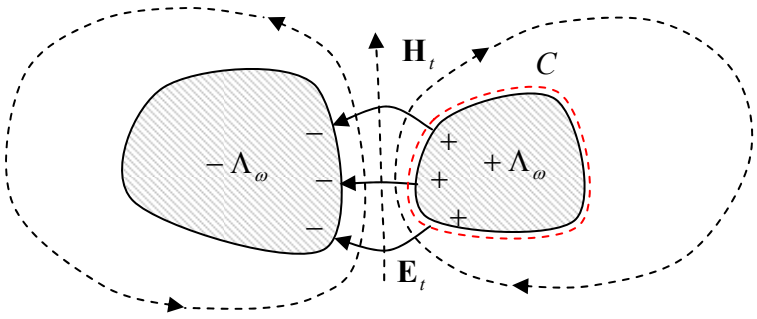


Fig. 7.17. An example of the cross-section of a transmission line that may support the TEM wave propagation.

In order to derive "global" relations for such a transmission line, let us consider the contour C drawn very close to the surface of one of its conductors – see, e.g., the red dashed line in Fig. 17. We can consider it, on one hand, as the cross-section of a cylindrically-shaped Gaussian volume of a certain elementary length $dz << \lambda \equiv 2\pi/k$. Using the generalized Gauss law (3.34), we get

$$\oint_C (\mathbf{E}_t)_n\, dr = \frac{\Lambda_\omega}{\varepsilon}, \tag{7.105}$$

where Λ_ω (not to be confused with the wavelength λ!) is the complex amplitude of the linear density of the electric charge of the conductor. On the other hand, the same contour C may be used in the generalized Ampère law (5.116) to write

$$\oint_C (\mathbf{H}_t)_\tau \, dr = I_\omega , \tag{7.106}$$

where I_ω is the total current flowing along the conductor (or rather its complex amplitude). But, as was mentioned above, in the TEM wave the ratio E_t/H_t of the field components participating in these two integrals is constant and equal to $Z = (\mu/\varepsilon)^{1/2}$, so Eqs. (105)-(106) give the following simple relation between the "global" variables of the conductor:

$$I_\omega = \frac{\Lambda_\omega / \varepsilon}{Z} \equiv \frac{\Lambda_\omega}{(\varepsilon\mu)^{1/2}} \equiv \frac{\omega}{k}\Lambda_\omega . \tag{7.107}$$

This important relation may be also obtained in a different way; let me describe it as well, because (as we will see below) it has an independent heuristic value. Let us consider a small segment $dz \ll \lambda = 2\pi/k$ of the line's conductor, and apply the electric charge conservation law (4.1) to the *instant* values of the linear charge density and current. The cancellation of dz in both parts yields

$$\frac{\partial \Lambda(z,t)}{\partial t} = -\frac{\partial I(z,t)}{\partial z} . \tag{7.108}$$

If we accept the sinusoidal waveform, $\exp\{i(kz - \omega t)\}$, for both these variables, we immediately recover Eq. (107) for their complex amplitudes, showing that this relation expresses just the charge continuity law.

The global equation (108) may be made more specific in the case when the frequency dependence of ε and μ is negligible, and the transmission line consists of just two isolated conductors – see, e.g., Fig. 17. In this case, to have the wave localized in the space near the two conductors, we need a sufficiently fast decrease of its electric field at large distances. For that, their linear charge densities for each value of z should be equal and opposite, and we can simply relate them to the potential difference V between the conductors:

$$\frac{\Lambda(z,t)}{V(z,t)} = C_0 , \tag{7.109}$$

where C_0 is the mutual capacitance of the conductors (per unit length) – which was repeatedly discussed in Chapter 2. Then Eq. (108) takes the following form:

$$C_0 \frac{\partial V(z,t)}{\partial t} = -\frac{\partial I(z,t)}{\partial z} . \tag{7.110}$$

Next, let us consider the contour shown with the red dashed line in Fig. 18 (which shows a different cross-section of the transmission line – by a plane containing the wave propagation axis z), and apply to it the Faraday induction law (6.3).

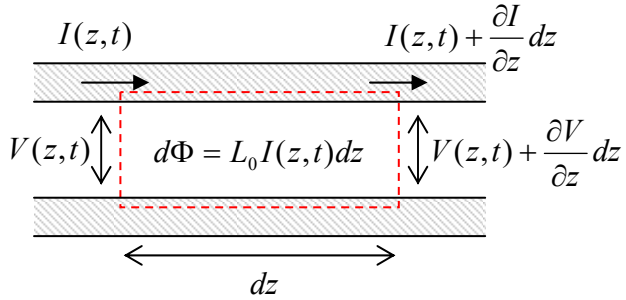


Fig. 7.18. Electric current, magnetic flux, and voltage in a two-conductor transmission line.

Since, in the coarse-grain approximation, the electric field inside the conductors (in Fig. 18, on the horizontal segments of the contour) vanishes, the total e.m.f. equals the difference of the voltages V at the ends of the segment dz, while the only sources of the magnetic flux through the area limited by the contour are the (equal and opposite) currents $\pm I$ in the conductors, we can use Eq. (5.70) to express the flux. As a result, by canceling dz in both parts of the equation, we get

$$L_0 \frac{\partial I(z,t)}{\partial t} = -\frac{\partial V(z,t)}{\partial z}, \tag{7.111}$$

where L_0 is the mutual inductance of the conductors per unit length. The only difference between this L_0 and the dc mutual inductances discussed in Chapter 5 is that at the high frequencies we are analyzing now, L_0 should be calculated neglecting the magnetic field penetration into the conductors. (In the dc case, we had the same situation for superconductor electrodes within their coarse-grain, ideal-diamagnet description.)

The system of Eqs. (110) and (111) is frequently called the *telegrapher's equations*. Combined, they give for any "global" variable f (either V, or I, or Λ) the usual 1D wave equation,

$$\frac{\partial^2 f}{\partial z^2} - L_0 C_0 \frac{\partial^2 f}{\partial t^2} = 0, \tag{7.112}$$

which describes dispersion-free TEM wave's propagation. Again, this equation is only valid within the frequency range where the frequency dependence of both ε and μ is negligible. If this is not so, the global approach may still be used for sinusoidal waves $f = \mathrm{Re}[f_\omega \exp\{i(kz - \omega t)\}]$. Repeating the above arguments, instead of Eqs. (110)-(111) we get a more general system of two algebraic equations

$$\omega C_0 V_\omega = k I_\omega, \qquad \omega L_0 I_\omega = k V_\omega, \tag{7.113}$$

in which $L_0 \propto \mu$ and $C_0 \propto \varepsilon$ may now depend on frequency. These equations are consistent only if

$$\boxed{L_0 C_0 = \frac{k^2}{\omega^2} \equiv \frac{1}{v^2} \equiv \varepsilon\mu}. \tag{7.114}$$

L_0C_0 product invariance

Besides the fact we have already known (that the TEM wave's speed is the same as that of the plane wave), Eq. (114) gives us the result that I confess was not emphasized enough in Chapter 5: the product L_0C_0 does not depend on the shape or size of line's cross-section, provided that the magnetic field's penetration into the conductors is negligible). Hence, if we have calculated the mutual capacitance C_0 of a system of two cylindrical conductors, the result immediately gives us their mutual inductance: $L_0 = \varepsilon\mu/C_0$. This relationship stems from the fact that both the electric and magnetic fields may be expressed via the solution of the same 2D Laplace equation for the system's cross-section.

With Eq. (114) satisfied, any of Eqs. (113) gives the same result for the following ratio:

$$\boxed{Z_W \equiv \frac{V_\omega}{I_\omega} = \left(\frac{L_0}{C_0}\right)^{1/2}}, \tag{7.115}$$

Transmission line's TEM Impedance

which is called the *transmission line's impedance*. This parameter has the same dimensionality (in SI units – ohms, denoted Ω) as the wave impedance (7),

$$Z \equiv \frac{E_\omega}{H_\omega} = \left(\frac{\mu}{\varepsilon}\right)^{1/2}, \tag{7.116}$$

but these parameters should not be confused, because Z_W depends on the cross-section's geometry, while Z does not. In particular, Z_W is the only important parameter of a transmission line for its matching with a lumped load circuit (Fig. 19) in the important case when both the cable cross-section's size and the load's linear dimensions are much smaller than the wavelength.[48]

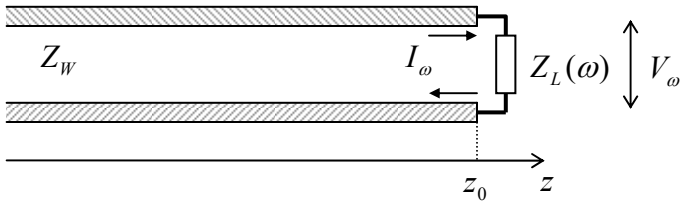


Fig. 7.19. Passive, lumped termination of a TEM transmission line.

Indeed, in this case, we may consider the load in the quasistatic limit and write

$$V_\omega(z_0) = Z_L(\omega) I_\omega(z_0) \ , \tag{7.117}$$

where $Z_L(\omega)$ is the (generally complex) impedance of the load. Taking $V(z,t)$ and $I(z,t)$ in the form similar to Eqs. (61) and (62), and writing the two Kirchhoff's circuit laws for the point $z = z_0$, we get for the reflection coefficient a result similar to Eq. (68):

$$R = \frac{Z_L(\omega) - Z_W}{Z_L(\omega) + Z_W}. \tag{7.118}$$

This formula shows that for the perfect matching (i.e. the total wave absorption in the load), the load's impedance $Z_L(\omega)$ should be real and equal to Z_W – but not necessarily to Z.

As an example, let us consider one of the simplest (and most practically important) transmission lines: the coaxial cable (Fig. 20).[49]

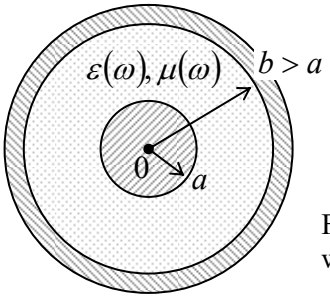


Fig. 7. 20. The cross-section of a coaxial cable with (possibly, dispersive) dielectric filling.

For this geometry, we already know the expressions for both C_0 and L_0,[50] though they have to be modified for the account of arbitrary dielectric and magnetic constants, and the magnetic field's non-penetration into the conductors. As a result of this (elementary) modification, we get the formulas,

[48] The ability of TEM lines to have such a small cross-section is another important practical advantage.
[49] It was invented by the same O. Heaviside in 1880.
[50] See, respectively, Eqs. (2.49) and (5.79).

$$C_0 = \frac{2\pi\varepsilon}{\ln(b/a)}, \qquad L_0 = \frac{\mu}{2\pi}\ln(b/a), \tag{7.119}$$

Coaxial cable's C_0 and L_0

illustrating that the universal relationship (114) is indeed valid. For the cable's impedance (115), Eqs. (119) yield a geometry-dependent value

$$Z_W = \left(\frac{\mu}{\varepsilon}\right)^{1/2}\frac{\ln(b/a)}{2\pi} \equiv Z\frac{\ln(b/a)}{2\pi} \neq Z\,. \tag{7.120}$$

For the standard TV antenna cables (such as RG-6/U, with $b/a \sim 3$, $\varepsilon/\varepsilon_0 \approx 2.2$), $Z_W = 75\ \Omega$, while for most computer component connections, coaxial cables with $Z_W = 50\ \Omega$ (such as RG-58/U) are prescribed by electronic engineering standards. Such cables are broadly used for the transmission of electromagnetic waves with frequencies up to 1 GHz over distances of a few km, and up to ~20 GHz on the tabletop scale (a few meters), limited by wave attenuation – see Sec. 9 below.

Moreover, the following two facts enable a wide application, in electrical engineering and physical experiment, of coaxial-cable-*like* systems. First, as Eq. (5.78) shows, in a cable with $a << b$, most energy of the wave is localized near the internal conductor. Second, the theory to be discussed in the next section shows that excitation of other (H- and E-) waves in the cable is impossible until the wavelength λ becomes smaller than $\sim\pi(a + b)$. As a result, the TEM mode propagation in a cable with $a << b < \lambda/\pi$ is not much affected even if the internal conductor is not straight, but bent – for example, into a helix – see, e.g., Fig. 21.

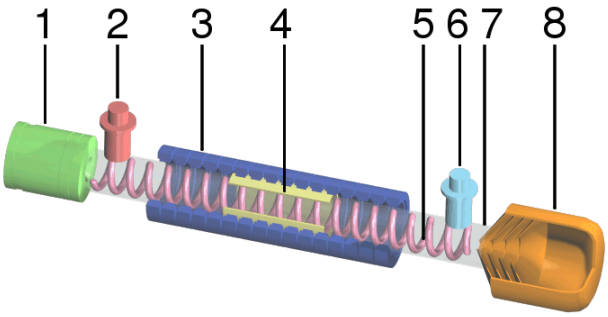


Fig. 7.21. A typical traveling-wave tube: (1) electron gun, (2) ac input, (3) beam-focusing magnets, (4) wave attenuator, (5) helix coil, (6) ac output, (7) vacuum tube, (8) electron collector. Adapted from https://en.wikipedia.org/wiki/Traveling-wave_tube under the Creative Commons BY-SA 3.0 license.

In such a system, called the *traveling-wave tube* (TWT), a quasi-TEM wave propagates with velocity $v \approx c$ along the helix's length, so the velocity's component along the cable's axis may be made close to the velocity $u << c$ of the electron beam moving ballistically along the tube's axis, enabling their effective interaction, and as a result, a length-accumulating amplification of the wave.[51]

Another important example of a TEM transmission line is a set of two parallel wires. In the form of *twisted pairs*,[52] they allow communications, in particular long-range telephone and DSL Internet

[51] Despite the current prevalence of semiconductor devices in electronics, TWTs are still used in satellite TV and radio systems, because they may work at very high microwave power – e.g., up to 200W at 20 GHz and pulsed 50W at 200 GHz. Very unfortunately, in this course, I will not have time/space to discuss even the (rather elegant) basic theory of such devices. The reader interested in this field may be referred, for example, to the detailed monograph by J. Whitaker, *Power Vacuum Tubes Handbook*, 3rd ed., CRC Press, 2017.

[52] Such twisting, around the line's direction axis, reduces the crosstalk between adjacent lines, and the parasitic radiation at their bends.

connections, at frequencies up to a few hundred kHz, as well as relatively short, multi-line Ethernet and TV cables at frequencies up to ~ 1 GHz, limited mostly by the mutual interference ("crosstalk") between the individual lines of the same cable, and the unintentional radiation of the wave into the environment.

7.6. Waveguides: *H* and *E* waves

Let us now return to Eqs. (100) and explore the *H*- and *E*-waves – with, respectively, either H_z or E_z different from zero. At the first sight, they may seem more complex. However, Eqs. (101), which determine the distribution of these longitudinal components over the cross-section, are just the 2D Helmholtz equations for scalar functions. For simple cross-section geometries, they may be readily solved using the methods discussed for the Laplace equation in Chapter 2, in particular the variable separation. After the solution of such an equation has been found, the transverse components of the fields may be calculated by differentiation, using the simple formulas,

$$\mathbf{E}_t = \frac{i}{k_t^2}\left[k_z \nabla_t E_z - kZ(\mathbf{n}_z \times \nabla_t H_z)\right], \qquad \mathbf{H}_t = \frac{i}{k_t^2}\left[k_z \nabla_t H_z + \frac{k}{Z}(\mathbf{n}_z \times \nabla_t E_z)\right], \tag{7.121}$$

which follow from the first line of Eqs. (100).[53]

In comparison with the boundary problems of electro- and magnetostatics, the only conceptually new feature of Eqs. (101) is that they form the so-called *eigenproblems*, with typically many solutions (*eigenfunctions*), each describing a specific wave mode, and corresponding to a specific *eigenvalue* of the parameter k_t. The good news here is that these values of k_t are determined by this 2D boundary problem and hence do not depend on k_z. As a result, the dispersion law $\omega(k_z)$ of any mode, which follows from the last form of Eq. (102),

Universal dispersion relation

$$\omega = \left(\frac{k_z^2 + k_t^2}{\varepsilon\mu}\right)^{1/2} \equiv \left(v^2 k_z^2 + \omega_c^2\right)^{1/2}, \tag{7.122}$$

is functionally similar for all modes. It is also similar to that of plane waves in plasma (see Eq. (38), Fig. 6, and their discussion in Sec. 2), with the only difference that the speed in light *c* is generally replaced with $v = 1/(\varepsilon\mu)^{1/2}$ – the speed of the plane or TEM waves in the medium filling the waveguide, and that ω_p is replaced with the so-called *cutoff frequency*

$$\omega_c \equiv v k_t, \tag{7.123}$$

specific for each mode. (As Eq. (101) implies, and as we will see from several examples below, k_t has the order of $1/a$, where *a* is the characteristic dimension of the waveguide's cross-section, so the critical value of the free-space wavelength $\lambda \equiv 2\pi c/\omega$ is of the order of *a*.) Below the cutoff frequency of each particular mode, this wave cannot propagate in the waveguide.[54] As a result, the modes with the *lowest*

[53] For the derivation of Eqs. (121), one of these two linear equations should be first vector-multiplied by $\mathbf{n}_z$. Note also that this approach could not be used to analyze the TEM waves, because for them $k_t = 0$, $E_z = 0$, $H_z = 0$, and Eqs. (121) yield uncertainty.

[54] An interesting twist in the ideas of electromagnetic metamaterials (mentioned in Sec. 5 above) is the so-called *ε-near-zero* materials, designed to have the effective product $\varepsilon\mu$ much lower than $\varepsilon_0\mu_0$ within certain frequency ranges. Since at these frequencies, the speed *v* (4) becomes much lower than *c*, the cutoff frequency (123)

values of ω_c present special practical interest, because the choice of the signal frequency ω between the two lowest values of the cutoff frequency (123) guarantees that the waves propagate in the form of only one mode, with the lowest k_t. Such a choice enables engineers to simplify the excitation of the desired mode by wave generators and to avoid the unintentional transfer of electromagnetic wave energy to undesirable modes by (virtually unavoidable) small inhomogeneities of the system.

The boundary conditions for the Helmholtz equations (101) depend on the propagating wave type. For the *E*-modes, with $H_z = 0$ but $E_z \neq 0$, the condition $E_\tau = 0$ immediately gives

$$E_z\big|_C = 0, \tag{7.124}$$

where *C* is the inner contour limiting the conducting wall's cross-section. For the *H*-modes, with $E_z = 0$ but $H_z \neq 0$, the boundary condition is slightly less obvious and may be obtained using, for example, the second equation of the system (100), vector-multiplied by $\mathbf{n}_z$. Indeed, for the component normal to the conductor surface, the result of such multiplication is

$$ik_z(\mathbf{H}_t)_n - i\frac{k}{Z}\left(\mathbf{n}_z \times \mathbf{E}_t\right)_n = \frac{\partial H_z}{\partial n}. \tag{7.125}$$

But the first term on the left-hand side of this relation must be zero on the wall surface, because of the second of Eqs. (104), while according to the first of Eqs. (104), the vector $\mathbf{E}_t$ in the second term cannot have a component tangential to the wall. As a result, the vector product in that term cannot have a normal component, so the term should equal zero as well, and Eq. (125) is reduced to

$$\frac{\partial H_z}{\partial n}\Big|_C = 0\,. \tag{7.126}$$

Let us see how all this machinery works for a simple but practically important case of a metallic-wall waveguide with a rectangular cross-section – see Fig. 22

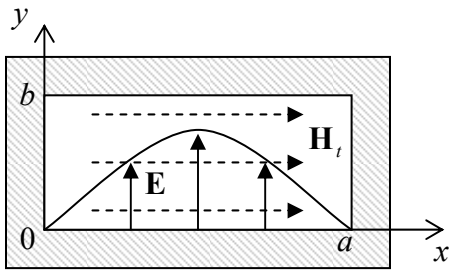


Fig. 7.22. A rectangular waveguide, and the transverse field distribution in its fundamental mode H_{10} (schematically).

In the natural Cartesian coordinates shown in this figure, both Eqs. (101) take the simple form

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_t^2\right) f = 0, \qquad \text{where } f = \begin{cases} E_z, & \text{for } E\text{-modes}, \\ H_z, & \text{for } H\text{-modes}. \end{cases} \tag{7.127}$$

From Chapter 2, we know that the most effective way of solving such equations in a rectangular region is the variable separation, in which the general solution is represented as a sum of partial solutions of the type

---

virtually vanishes. As a result, the waves may "tunnel" through very narrow sections of metallic waveguides filled with such materials – see, e.g., M. Silveirinha and N. Engheta, *Phys. Rev. Lett.* **97**, 157403 (2006).

$$f = X(x)Y(y)\,. \tag{7.128}$$

Plugging this expression into Eq. (127), and dividing each term by XY, we get the equation,

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + k_t^2 = 0\,, \tag{7.129}$$

which should be satisfied for all values of x and y within the waveguide's interior. This is only possible if each term of the sum equals a constant. Taking the X-term and Y-term constants in the form $(-k_x^2)$ and $(-k_y^2)$, respectfully, and solving the corresponding ordinary differential equations,[55] for the eigenfunction (128) we get

$$f = (c_x \cos k_x x + s_x \sin k_x x)(c_y \cos k_y y + s_y \sin k_y y), \qquad \text{with } k_x^2 + k_y^2 = k_t^2\,, \tag{7.130}$$

where the constants c and s should be found from the boundary conditions. Here the difference between the H-modes and E-modes kicks in.

For the H-modes, Eq. (130) is valid for H_z, and we should use the boundary condition (126) on all metallic walls of the waveguide, i.e. at $x = 0$ and a; and $y = 0$ and b – see Fig. 22. As a result, we get very simple expressions for eigenfunctions and eigenvalues:

$$(H_z)_{nm} = H_l \cos\frac{\pi n x}{a}\cos\frac{\pi m y}{b}\,, \tag{7.131}$$

$$k_x = \frac{\pi n}{a}, \quad k_y = \frac{\pi m}{b}, \quad \text{so that } (k_t)_{nm} = \left(k_x^2 + k_y^2\right)^{1/2} = \pi\left[\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2\right]^{1/2}, \tag{7.132}$$

where H_l is the longitudinal field's amplitude, and n and m are two integer numbers – each of them arbitrary besides that they cannot be equal to zero simultaneously.[56] Assuming, just for certainty, that $a \geq b$ (as shown in Fig. 22), we see that the lowest eigenvalue of k_t and hence the lowest cutoff frequency (123) are achieved for the so-called H_{10} *mode* with $n = 1$ and $m = 0$, and hence with

Fundamental mode's cutoff

$$\boxed{(k_t)_{10} = \frac{\pi}{a}}, \tag{7.133}$$

thus confirming our prior estimate of k_t.

Depending on the a/b ratio, the second-lowest k_t (and hence ω_c) belongs to either the H_{11} mode with $n = 1$ and $m = 1$:

$$(k_t)_{11} = \pi\left(\frac{1}{a^2} + \frac{1}{b^2}\right)^{1/2} \equiv \left[1 + \left(\frac{a}{b}\right)^2\right]^{1/2}(k_t)_{10}\,, \tag{7.134}$$

or to the H_{20} mode with $n = 2$ and $m = 0$:

[55] Let me hope that the solution of equations of the type $d^2X/dx^2 + k_x^2X = 0$ does not present any problem for the reader, at least due to their prior experience with problems such as standing waves on a guitar string, wavefunctions in a flat 1D quantum well, or (with the replacement $x \to t$) a classical harmonic oscillator.

[56] Otherwise, the function $H_z(x,y)$ would be constant, so, according to Eq. (121), the transverse components of the electric and magnetic field would equal zero. As a result, as the last two lines of Eqs. (100) show, the whole field would be zero for any $k_z \neq 0$.

$$(k_t)_{20} = \frac{2\pi}{a} \equiv 2(k_t)_{10}. \tag{7.135}$$

These values become equal at $a/b = \sqrt{3} \approx 1.7$; in practical waveguides, the a/b ratio is not too far from this value. For example, in the standard X-band (~10-GHz) waveguide WR90, $a \approx 2.3$ cm ($f_c \equiv \omega_c/2\pi \approx$ 6.5 GHz), and $b \approx 1.0$ cm.

Now let us have a look at the alternative *E*-modes. For them, we still should use the general solution (130) with $f = E_z$, but now with the boundary condition (124). This gives us the eigenfunctions

$$(E_z)_{nm} = E_l \sin\frac{\pi n x}{a} \sin\frac{\pi m y}{b}, \tag{7.136}$$

and the same eigenvalue spectrum (132) as for the *H* modes. However, now neither n nor m can be equal to zero; otherwise, Eq. (136) would give the trivial solution $E_z(x,y) = 0$. Hence the lowest cutoff frequency of TM waves is achieved at the so-called E_{11} *mode* with n =1, m = 1, and with the eigenvalue given by Eq. (134), always higher than $(k_t)_{10}$.

Thus the fundamental H_{10} mode is certainly the most important wave in rectangular waveguides; let us have a better look at this field distribution. Plugging the corresponding solution (131) with n = 1 and m = 0 into the general relation (121), we easily get

$$(H_x)_{10} = -i\frac{k_z a}{\pi} H_l \sin\frac{\pi x}{a}, \qquad (H_y)_{10} = 0, \tag{7.137}$$

$$(E_x)_{10} = 0, \qquad (E_y)_{10} = i\frac{ka}{\pi} Z H_l \sin\frac{\pi x}{a}. \tag{7.138}$$

This field distribution is (schematically) shown in Fig. 22. Neither of the fields depends on the coordinate y – the feature very convenient, in particular, for microwave experiments with small samples. The electric field has only one (in Fig. 22, vertical) component that vanishes at the side walls and reaches its maximum at the waveguide's center; its field lines are straight, starting and ending on wall surface charges (whose distribution propagates along the waveguide together with the wave). In contrast, the magnetic field has two non-zero components (H_x and H_z), and its field lines are shaped as horizontal loops wrapped around the electric field maxima.

An important question is whether the H_{10} wave may be usefully characterized by a unique impedance introduced similarly to Z_W of the TEM modes – see Eq. (115). The answer is *not*, because the main value of Z_W is a convenient description of the impedance matching of a transmission line with a lumped load – see Fig. 19 and Eq. (118). As was discussed above, such a simple description is possible (i.e., does not depend on the exact geometry of the connection) only if both dimensions of the line's cross-section are much less than λ. But for the H_{10} wave (and more generally, any non-TEM mode) this is impossible – see, e.g., Eq. (129): its lowest frequency corresponds to the TEM wavelength $\lambda_{max} = 2\pi/(k_t)_{min} = 2\pi/(k_t)_{10} = 2a$. (The reader is challenged to find a simple interpretation of this equality.)

Now let us consider metallic-wall waveguides with a round cross-section (Fig. 23a). In this single-connected geometry, the TEM waves are impossible again, while for the analysis of *H*-modes and *E*-modes, the polar coordinates $\{\rho, \varphi\}$ are most natural. In these coordinates, the 2D Helmholtz equation (101) takes the following form:

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}+k_t^2\right]f=0, \qquad \text{where } f=\begin{cases}E_z, & \text{for } E\text{ - modes,}\\ H_z, & \text{for } H\text{ - modes.}\end{cases} \tag{7.139}$$

Separating the variables as $f = \mathcal{R}(\rho)\mathcal{F}(\varphi)$, we get

$$\frac{1}{\rho\mathcal{R}}\frac{d}{d\rho}\left(\rho\frac{d\mathcal{R}}{d\rho}\right)+\frac{1}{\rho^2\mathcal{F}}\frac{d^2\mathcal{F}}{d\varphi^2}+k_t^2=0\,. \tag{7.140}$$

But this is exactly the Eq. (2.127) that was studied in Sec. 2.7 in the context of electrostatics, just with a replacement of notation: $\gamma \to k_t$. So we already know that to have 2π-periodic functions $\mathcal{F}(\varphi)$ and finite values $\mathcal{R}(0)$ (which are evidently necessary for our current case – see Fig. 23a), the general solution must have the form given by Eq. (2.136), i.e. the eigenfunctions are expressed via integer-order Bessel functions of the first kind:

$$f_{nm}=J_n(k_{nm}\rho)(c_n\cos n\varphi+s_n\sin n\varphi)\equiv \text{const}\times J_n(k_{nm}\rho)\cos n(\varphi-\varphi_0), \tag{7.141}$$

with the eigenvalues k_{nm} of the transverse wave number k_t to be determined from appropriate boundary conditions, and an arbitrary constant φ_0.

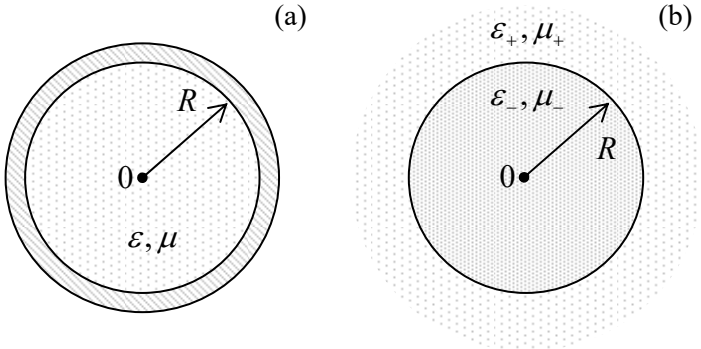


Fig. 7.23. (a) Metallic and (b) dielectric waveguides with circular cross-sections.

As for the rectangular waveguide, let us start with the H-modes ($f = H_z$). Then the boundary condition on the wall surface ($\rho = R$) is given by Eq. (126), which, for the solution (141), takes the form

$$\frac{d}{d\xi}J_n(\xi)=0, \qquad \text{where } \xi\equiv kR\,. \tag{7.142}$$

This means that the eigenvalues of Eq. (139) are

$$k_t=k_{nm}=\frac{\xi'_{nm}}{R}, \tag{7.143}$$

where ξ'_{nm} is the m^{th} zero of the function $dJ_n(\xi)/d\xi$. Approximate values of these zeros for several lowest n and m may be read out from Fig. 2.18; their more accurate values are given in Table 1 below.

Table 7.1. Zeros ξ'_{nm} of the function $dJ_n(\xi)/d\xi$ for a few lowest values of the Bessel function's index n and the root's number m.

	$m = 1$	2	3
$n = 0$	3.83171	7.015587	10.1735
1	1.84118	5.33144	8.53632
2	3.05424	6.70613	9.96947
3	4.20119	8.01524	11.34592

The table shows, in particular, that the lowest of the zeros is $\xi'_{11} \approx 1.84$.[57] Thus, perhaps a bit counter-intuitively, the fundamental mode, providing the lowest cutoff frequency $\omega_c = vk_{nm}$, is H_{11}, corresponding to $n = 1$ rather than $n = 0$:

$$H_z = H_l J_1\left(\xi'_{11}\frac{\rho}{R}\right)\cos(\varphi - \varphi_0). \tag{7.144}$$

It has the transverse wave number is $k_t = k_{11} = \xi'_{11}/R \approx 1.84/R$, and hence the cutoff frequency corresponding to the TEM wavelength $\lambda_{max} = 2\pi/k_{11} \approx 3.41\ R$. Thus the ratio of λ_{max} to the waveguide's diameter $2R$ is about 1.7, i.e. is close to the ratio $\lambda_{max}/a = 2$ for the rectangular waveguide. The origin of this proximity is clear from Fig. 24, which shows the transverse field distribution in the H_{11} mode. (It may be readily calculated from Eqs. (121) with $E_z = 0$ and H_z given by Eq. (144).)

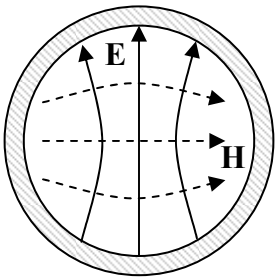


Fig. 7.24. Transverse field components in the fundamental H_{11} mode of a metallic, circular waveguide (schematically).

One can see that the field structure is actually very similar to that of the fundamental mode in the rectangular waveguide, shown in Fig. 22, despite the different nomenclature (which is due to the different coordinate system used for the solution). However, note the arbitrary constant angle φ_0, indicating that in circular waveguides, the transverse field's polarization is arbitrary. For some practical applications, such degeneracy of these "quasi-linearly-polarized" waves creates problems; some of them may be avoided by using waves with circular polarization.

As Table 1 shows, the next lowest H-mode is H_{21}, for which $k_t = k_{21} = \xi'_{21}/R \approx 3.05/R$, almost twice larger than that of the fundamental mode, and only then comes the first mode with no angular dependence of any field, H_{01}, with $k_t = k_{01} = \xi'_{01}/R \approx 3.83/R$,[58] followed by several angle-dependent modes: H_{31}, H_{12}, etc.

For the E modes, we may still use Eq. (141) (with $f = E_z$), but with the boundary condition (124) at $\rho = R$. This gives the following equation for the problem eigenvalues:

$$J_n(k_{nm}R) = 0, \quad \text{i.e. } k_{nm} = \frac{\xi_{nm}}{R}, \tag{7.145}$$

where ξ_{nm} is the m^{th} zero of function $J_n(\xi)$ – see Table 2.1. That table shows that the lowest k_t is equal to $\xi_{01}/R \approx 2.405/R$. Hence the corresponding mode (E_{01}), with no angular dependence of its fields, e.g.

[57] Mathematically, the lowest root of Eq. (142) with $n = 0$ equals 0. However, it would yield $k = 0$ and hence a constant field H_z, which, according to the first of Eqs. (121), would give a vanishing electric field.

[58] The electric field lines in the H_{01} mode (as well as all higher H_{0m} modes) are directed straight from the symmetry axis to the walls, reminding those of the TEM waves in the coaxial cable. Due to this property, these modes provide, at $\omega >> \omega_c$, much lower energy losses (see Sec. 9 below) than the fundamental H_{11} mode, and are sometimes used in practice, despite the inconvenience of working in the multimode frequency range.

$$E_z = E_l J_0\left(\xi_{01}\frac{\rho}{R}\right), \tag{7.146}$$

has the second-lowest cutoff frequency, ~30% higher than that of the fundamental mode H_{11}.

Finally, let us discuss one more topic of general importance – the number N of electromagnetic modes that may propagate in a waveguide within a certain range of relatively large frequencies $\omega >> \omega_c$. It is easy to calculate for a rectangular waveguide, with its simple expressions (132) for the eigenvalues of $\{k_x, k_y\}$. Indeed, these expressions describe a rectangular mesh on the $[k_x, k_y]$ plane, so each point corresponds to the plane area $\Delta A_k = (\pi/a)(\pi/b)$, and the number of modes in a large k-plane area $A_k >> \Delta A_k$ is $N = A_k/\Delta A_k = abA_k/\pi^2 = AA_k/\pi^2$, where A is the waveguide's cross-section area.[59] However, it is frequently more convenient to discuss transverse wave vectors $\mathbf{k}_t$ of arbitrary direction, i.e. with an arbitrary sign of their components k_x and k_y. Taking into account that the opposite values of each component actually give the same wave, the actual number of different modes of each type (E- or H-) is a factor of $2^2 \equiv 4$ lower than was calculated above. This means that the number of modes of *both* types is

$$N = 2\frac{A_k A}{(2\pi)^2}. \tag{7.147}$$

Let me leave it for the reader to find hand-waving (but convincing :-) arguments that this *mode counting rule* is valid for waveguides with cross-sections of any shape, and any boundary conditions on the walls, provided that $N >> 1$.

7.7. Dielectric waveguides, optical fibers, and paraxial beams

Now let us discuss electromagnetic wave propagation in *dielectric waveguides*. The simplest, *step-index* waveguide (see Figs. 23b and 25) consists of an inner *core* and an outer shell (in the optical fiber technology lingo, called *cladding*) with a higher wave propagation speed, i.e. a lower index of refraction:

$$v_+ > v_-, \quad \text{i.e. } n_+ < n_-, \quad k_+ < k_-, \quad \varepsilon_+\mu_+ < \varepsilon_-\mu_-. \tag{7.148}$$

at the same frequency. (In most cases the difference is achieved due to that in the electric permittivity, $\varepsilon_+ < \varepsilon_-$, while magnetically both materials are virtually passive: $\mu_- \approx \mu_+ \approx \mu_0$, so their refraction indices $n_\pm$, defined by Eq. (84), are very close to $(\varepsilon_\pm/\varepsilon_0)^{1/2}$; I will limit my discussion to this approximation.)

The basic idea of the waveguide's operation may be readily understood in the limit when the wavelength λ is much smaller than the characteristic size R of the core's cross-section. In this "geometric-optics" limit, at distances of the order of λ from the core-cladding interface, which provides the wave reflection, we can neglect the interface's curvature and approximate its geometry with a plane. As we know from Sec. 4, if the angle θ of the wave's incidence on such a plane interface is larger than the critical value θ_c specified by Eq. (85), the wave is totally reflected. As a result, the waves launched into the fiber core at such "grazing" angles, propagate inside the core, being repeatedly reflected from the cladding – see Fig. 25.

[59] This formula ignores the fact that, according to the above analysis, some modes (with $n = 0$ *and* $m = 0$ for the H modes, and $n = 0$ *or* $m = 0$ for the E modes) are forbidden. However, for $N >> 1$, the associated corrections of Eq. (147) are negligible.

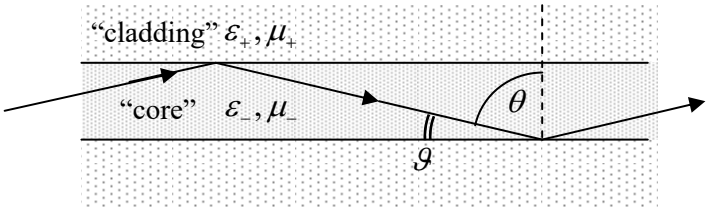


Fig. 7.25. Wave propagation in a thick optical fiber at $\theta > \theta_c$.

The most important type of dielectric waveguides is *optical fibers*.[60] Due to a heroic technological effort over three decades starting from the mid-1960s, the attenuation of such fibers has been decreased from values of the order of 20 db/km (typical for a window glass) to the fantastically low values of ~0.2 db/km (meaning virtually perfect transparency of 10-km-long fiber segments!), combined with the extremely low plane-wave ("chromatic") dispersion below 10 ps/km·nm.[61] In conjunction with the development of inexpensive erbium-based quantum amplifiers, this breakthrough has enabled inter-city and inter-continental (undersea), broadband[62] optical cables, which are the backbone of all modern telecommunication infrastructure.

The only bad news is that these breakthroughs were achieved for just one kind of materials (silica-based glasses)[63] within a very narrow range of their chemical composition. As a result, the dielectric constants $\kappa_\pm \equiv \varepsilon_\pm/\varepsilon_0$ of the cladding and core of practical optical fibers are both close to 2.2 (giving $n_\pm \approx 1.5$) and hence very close to each other, so the relative difference of the refraction indices,

$$\Delta \equiv \frac{n_- - n_+}{n_-} = \frac{\varepsilon_-^{1/2} - \varepsilon_+^{1/2}}{\varepsilon_-^{1/2}} \approx \frac{\varepsilon_- - \varepsilon_+}{2\varepsilon_\pm}, \tag{7.149}$$

is typically below 0.5%. This factor limits the fiber bandwidth. Indeed, let us use the geometric-optics picture to calculate the number of quasi-plane-wave modes that may propagate in the fiber. For the complementary angle (Fig. 25)

$$\vartheta \equiv \frac{\pi}{2} - \theta, \qquad \text{so that } \sin\theta = \cos\vartheta, \tag{7.150}$$

Eq. (85) gives the following propagation condition:

$$\cos\vartheta > \frac{n_+}{n_-} = 1 - \Delta \cdot \tag{7.151}$$

[60] For a comprehensive discussion of this vital technology see, e.g., A. Yariv and P. Yeh, *Photonics*, 6th ed., Oxford U. Press, 2007.

[61] Both these parameters have their best values not in the visible light range (with wavelengths from 380 to 740 nm), but in the near-infrared, with the attenuation lowest between approximately 1,500 and 1,630 nm. As a result, most modern communication systems use two spectral windows – the so-called C-band (1,530-1,565 nm) and L-band (1,570-1,610 nm) within that range.

[62] Each of the spectral bands mentioned above, at a typical signal-to-noise ratio $S/N > 10^5$, corresponds to the Shannon bandwidth $\Delta f \log_2(S/N)$ exceeding 10^{14} bits per second, some five orders of magnitude (!) higher than that of a modern Ethernet cable. The practically usable bandwidth of each fiber is somewhat lower, but a typical optical cable, with many fibers in parallel, has a proportionately higher aggregate bandwidth. A relatively recent (circa 2017) example is the C-band transatlantic (6,600-km-long) cable *Marea*, with eight fiber pairs and an aggregate useable bandwidth of 160 terabits per second.

[63] The silica-based fibers were developed in 1966 by an industrial research group led by Charles Kao (who shared the 2009 Nobel Prize in physics), but the very idea of using optical fibers for long-range communications may be traced back at least to the 1963 work by Jun-ichi Nishizawa – who also invented semiconductor lasers.

In the limit $\Delta << 1$, when the incidence angles $\theta > \theta_c$ of all propagating waves are very close to $\pi/2$, and hence the complementary angles are ϑ small, we may keep only two first terms in the Taylor expansion of the left-hand side of Eq. (151) and get

$$\vartheta_{\max}^2 \approx 2\Delta . \tag{7.152}$$

(Even for the higher-end value $\Delta = 0.005$, this critical angle is only ~0.1 radian, i.e. is close to 5°.) Due to this smallness, we may approximate the maximum transverse component of the wave vector as

$$(k_t)_{\max} = k(\sin\vartheta)_{\max} \approx k\vartheta_{\max} \approx \sqrt{2}k\Delta , \tag{7.153}$$

and use Eq. (147) to calculate the number N of propagating modes:

$$N \approx 2\frac{(\pi R^2)(\pi k^2 \vartheta_{\max}^2)}{(2\pi)^2} = (kR)^2\Delta . \tag{7.154}$$

For typical values $k = 0.73\times10^7$ m^{-1} (corresponding to the free-space wavelength $\lambda_0 = n\lambda = 2\pi n/k \approx 1.3$ μm), $R = 25$ μm, and $\Delta = 0.005$, this formula gives $N \approx 150$.

Now we can calculate the *geometric dispersion* of such a fiber, i.e. the difference in the mode propagation speed, which is commonly characterized in terms of the difference between the wave delay times (traditionally measured in picoseconds per kilometer) of the fastest and slowest modes. Within the geometric optics approximation, the difference in time delays of the fastest mode (with $k_z = k$) and the slowest mode (with $k_z = k\sin\theta_c$) at distance l is

$$\Delta t = \Delta\left(\frac{l}{v_z}\right) = \Delta\left(\frac{k_z l}{\omega}\right) = \frac{l}{\omega}\Delta k_z = \frac{l}{v}(1-\sin\theta_c) = \frac{l}{v}\left(1-\frac{n_+}{n_-}\right) \equiv \frac{l}{v}\Delta . \tag{7.155}$$

For the example considered above, the TEM wave's speed in the glass, $v = c/n \approx 2\times10^8$ m/s, and the geometric dispersion $\Delta t/l$ is close to 25 ps/m, i.e. 25,000 ps/km. (This means, for example, that a 1-ns pulse, being distributed between the modes, would spread to a ~25-ns pulse after passing a just 1-km fiber segment.) This result should be compared with the chromatic dispersion mentioned above, below 10 ps/km·nm, which gives dt/l is of the order of only 1,000 ps/km in the whole communication band $d\lambda$ ~ 100 nm. Due to this high geometric dispersion, such relatively thick ($2R$ ~ 50 nm) *multi-mode fibers* are used for the transfer of signals over only short distances below ~ 100 m. (As compensation, they may carry relatively large power, beyond 10 mW, without being damaged by the field.)

Long-range telecommunications are based on *single-mode fibers*, with thin cores (typically with diameters $2R$ ~ 5 μm, i. e. of the order of $\lambda/\Delta^{1/2}$). For such structures, Eq. (154) yields $N \sim 1$, but in this case, the geometric optics approximation is not quantitatively valid, and for the fiber analysis, we should get back to the Maxwell equations. In particular, this analysis should take into explicit account the evanescent wave in the cladding, because its penetration depth may be comparable with R.[64]

[64] The following quantitative analysis of the single-mode fibers is very valuable – both for practice and as a very good example of Maxwell equations' solution. However, I have to confess that its results will not be used in the following parts of the course. So, if the reader is not interested in this topic, they may safely jump to the text following Eq. (181). (I believe that the discussion of the angular momentum of electromagnetic radiation, starting at that point, is compulsory for every professional physicist.)

Since the cross-section of an optical fiber lacks metallic walls, the Maxwell equations describing them cannot be exactly satisfied with either TEM-wave, or *H*-mode, or *E*-mode solutions. Instead, the fibers can carry the so-called *HE* and *EH* modes, with both vectors **H** and **E** having longitudinal components simultaneously. In such modes, both E_z and H_z inside the core ($\rho \le R$) have a form similar to Eq. (141):

$$f_- = f_l J_n(k_t \rho)\cos n(\varphi - \varphi_0), \qquad \text{where } k_t^2 = k_-^2 - k_z^2 > 0, \text{ and } k_-^2 \equiv \omega^2 \varepsilon_- \mu_-, \tag{7.156}$$

where the constant angles φ_0 may be different for each field. On the other hand, for the evanescent wave in the cladding, we may rewrite Eqs. (101) as

$$\left(\nabla^2 - \kappa_t^2\right) f_+ = 0, \qquad \text{where } \kappa_t^2 \equiv k_z^2 - k_+^2 > 0, \quad \text{and } k_+^2 \equiv \omega^2 \varepsilon_+ \mu_+ . \tag{7.157}$$

Figure 26 illustrates these relations between k_t, κ_t, k_z, and $k_\pm$; note that the following sum,

Universal relation between k_t and κ_t

$$\boxed{k_t^2 + \kappa_t^2 = \omega^2(\varepsilon_- - \varepsilon_+)\mu_0 = 2k^2\Delta,} \tag{7.158}$$

is fixed (at a given frequency) and, for typical fibers, is very small ($<< k^2$). In particular, Fig. 26 shows that neither k_t nor κ_t can be larger than $\omega[(\varepsilon_- - \varepsilon_+)\mu_0]^{1/2} = (2\Delta)^{1/2} k$. This means that the depth $\delta = 1/\kappa_t$ of the wave penetration into the cladding is at least $1/k(2\Delta)^{1/2} = \lambda/2\pi(2\Delta)^{1/2} >> \lambda/2\pi$. This is why the cladding layers in practical optical fibers are made as thick as ~50 μm, so only a negligibly small tail of this evanescent wave field reaches their outer surfaces.

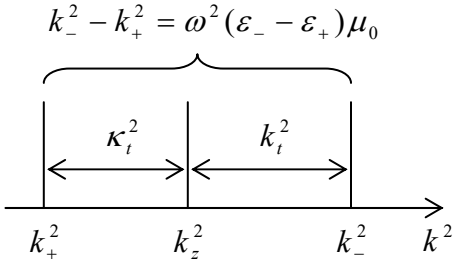


Fig. 7.26. The relation between the transverse exponents k_t and κ_t for waves in optical fibers.

In the polar coordinates, Eq. (157) becomes

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2} - \kappa_t^2\right] f_+ = 0, \tag{7.159}$$

- the equation to be compared with Eq. (139) for the circular metallic-wall waveguide. From Sec. 2.7, we know that the eigenfunctions of Eq. (159) are the products of the sine and cosine functions of $n\varphi$ by a linear combination of the modified Bessel functions I_n and K_n, shown in Fig. 2.22, now of the argument $\kappa_t\rho$. The fields have to vanish at $\rho \to \infty$, so only the latter functions (of the second kind) can participate in the solution:

$$f_+ \propto K_n(\kappa_t \rho)\cos n(\varphi - \varphi_0). \tag{7.160}$$

Now we have to reconcile Eqs. (156) and (160), using the boundary conditions at $\rho = R$ for both longitudinal and transverse components of both fields, with the latter components first calculated using Eqs. (121). Such a conceptually simple, but a bit bulky calculation (which I am leaving for the reader's exercise), yields a system of two homogeneous linear equations for the complex amplitudes E_l and H_l, which are compatible if

$$\left(\frac{k_-^2}{k_t}\frac{J_n'}{J_n}+\frac{k_+^2}{\kappa_t}\frac{K_n'}{K_n}\right)\left(\frac{1}{k_t}\frac{J_n'}{J_n}+\frac{1}{\kappa_t}\frac{K_n'}{K_n}\right)=\frac{n^2}{R^2}\left(\frac{k_-^2}{k_t^2}+\frac{k_+^2}{\kappa_t^2}\right)\left(\frac{1}{k_t^2}+\frac{1}{\kappa_t^2}\right), \tag{7.161}$$

where the prime signs denote the derivatives of each Bessel function over its full argument: $k_t\rho$ for J_n, and $\kappa_t\rho$ for K_n, and the functions and their derivatives are taken at $\rho = R$.

For any given frequency ω, the system of equations (158) and (161) determines the values of k_t and κ_t, and hence k_z. Actually, for any $n > 0$, this system provides two different solutions: one corresponding to the so-called *HE* wave, with a larger ratio of E_z/H_z, and the *EH* wave, with a smaller value of that ratio. For angular-symmetric modes with $n = 0$ (for whom we might naively expect the lowest cutoff frequency), the equations may be satisfied by the fields having just one non-zero longitudinal component (either E_z or H_z), so the *HE* wave are the usual *E*-modes, while the *EH* modes are the *H*-waves. For the *H*-modes, the characteristic equation is reduced to the requirement that the expression in the second parentheses on the left-hand side of Eq. (161) is equal to zero. Using the Bessel function identities $J_0' = -J_1$ and $K_0' = -K_1$, this equation may be rewritten in a simpler form:

$$\frac{1}{k_t}\frac{J_1(k_tR)}{J_0(k_tR)}=-\frac{1}{\kappa_t}\frac{K_1(\kappa_tR)}{K_0(\kappa_tR)}. \tag{7.162}$$

Using the universal relation between k_t and κ_t given by Eq. (158), we may plot both sides of Eq. (162) as functions of the same argument, say, $\xi \equiv k_tR$ – see Fig. 27.

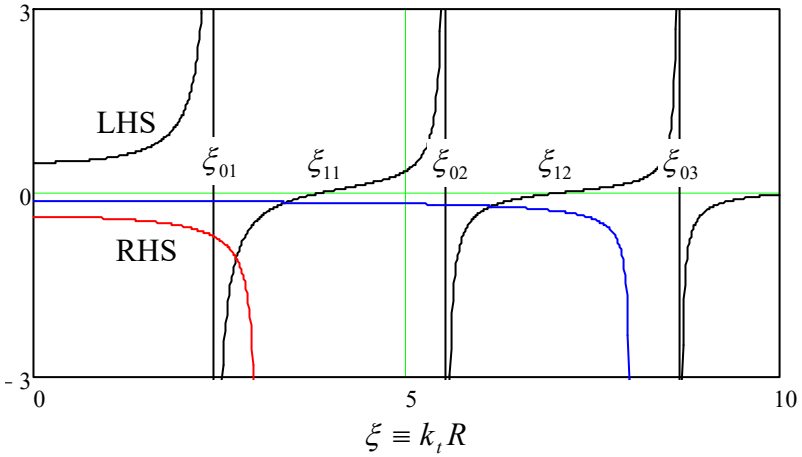


Fig. 7.27. Two sides of the characteristic equation (162), plotted as functions of k_tR, for two values of its dimensionless parameter: $\mathscr{V} = 8$ (blue line) and $\mathscr{V} = 3$ (red line). Note that according to Eq. (158), the argument of the functions K_0 and K_1 is $\kappa_tR = [\mathscr{V}^2 - (k_tR)^2]^{1/2} \equiv (\mathscr{V}^2 - \xi^2)^{1/2}$.

The right-hand side of Eq. (162) depends not only on ξ, but also on the dimensionless parameter $\mathscr{V}$ defined as the normalized right-hand side of Eq. (158):

$$\mathscr{V}^2 \equiv \omega^2(\varepsilon_- - \varepsilon_+)\mu_0R^2 \approx 2\Delta\, k_\pm^2R^2. \tag{7.163}$$

(According to Eq. (154), if $\mathscr{V} >> 1$, it gives twice the number N of the fiber modes – the conclusion confirmed by Fig. 27, taking into account that it describes only the *H*-modes.) Since the ratio K_1/K_0 is positive for all values of the functions' argument (see, e.g., the right panel of Fig. 2.22), the right-hand side of Eq. (162) is always negative, so the equation may have solutions only in the intervals where the ratio J_1/J_0 is negative, i.e. at

$$\xi_{01} < k_tR < \xi_{11}, \quad \xi_{02} < k_tR < \xi_{12}, \ldots, \tag{7.164}$$

where ξ_{nm} is the m-th zero of the function $J_n(\xi)$ – see Table 2.1. The right-hand side of the characteristic equation (162) diverges at $\kappa_t R \to 0$, i.e. at $k_t R \to \mathscr{V}$, so no solutions are possible if $\mathscr{V}$ is below the critical value $\mathscr{V}_c = \xi_{01} \approx 2.405$. At this cutoff point, Eq. (163) yields $k_\pm \approx \xi_{01}/R(2\Delta)^{1/2}$. Hence, the cutoff frequency of the lowest H mode corresponds to the TEM wavelength

$$\lambda_{\max} = \frac{2\pi R}{\xi_{01}}(2\Delta)^{1/2} \approx 3.7 R\Delta^{1/2}. \tag{7.165}$$

For typical parameters $\Delta = 0.005$ and $R = 2.5$ μm, this result yields $\lambda_{\max} \sim 0.65$ μm, corresponding to the free-space wavelength $\lambda_0 \sim 1$ μm. A similar analysis of the first parentheses on the left-hand side of Eq. (161) shows that at $\Delta \to 0$, the cutoff frequency for the E modes is similar.

This situation may look exactly like that in metallic-wall waveguides, with no waves possible at frequencies below ω_c, but this is not so. The basic reason for the difference is that in the metallic waveguides, the approach to ω_c results in the divergence of the longitudinal wavelength $\lambda_z \equiv 2\pi/k_z$. On the other hand, in dielectric waveguides, the approach leaves λ_z finite ($k_z \to k_+$). Due to this difference, a certain linear superposition of HE and EH waves with $n = 1$ can propagate at frequencies well below the cutoff frequency for $n = 0$, which we have just calculated.[65] This mode, in the limit $\varepsilon_+ \approx \varepsilon_-$ (i.e. $\Delta << 1$) allows a very interesting and simple description using the *Cartesian* (rather than polar) components of the fields, but still expressed as functions of the *polar* coordinates ρ and φ. The reason is that this mode is very close to a linearly polarized TEM wave. (Due to this reason, this mode is referred to as LP_{01}.)

Let us select the x-axis parallel to the transverse component of the magnetic field vector at $\rho = 0$, so $E_x|_{\rho=0} = 0$, but $E_y|_{\rho=0} \neq 0$, and $H_x|_{\rho=0} \neq 0$, but $H_y|_{\rho=0} = 0$. The only suitable solutions of the 2D Helmholtz equation (that should be obeyed not only by the z-components of the fields but also their x- and y-components) are proportional to $J_0(k_t\rho)$, with zero coefficients for E_x and H_y:

$$E_x = 0, \quad E_y = E_0 J_0(k_t\rho), \quad H_x = H_0 J_0(k_t\rho), \quad H_y = 0, \quad \text{for } \rho \le R. \tag{7.166}$$

LP_{01} mode

Now we can use the last two equations of Eqs. (100) to calculate the longitudinal components of the fields:

$$E_z = \frac{1}{-ik_z}\frac{\partial E_y}{\partial y} = -i\frac{k_t}{k_z}E_0 J_1(k_t\rho)\sin\varphi, \qquad H_z = \frac{1}{-ik_z}\frac{\partial H_x}{\partial x} = -i\frac{k_t}{k_z}H_0 J_1(k_t\rho)\cos\varphi, \tag{7.167}$$

where I have used the following mathematical identities: $J_0 = -J_1'$, $\partial\rho/\partial x = x/\rho = \cos\varphi$, and $\partial\rho/\partial y = y/\rho = \sin\varphi$. As a sanity check, we see that the longitudinal component or each field is a (legitimate!) eigenfunction of the type (141), with $n = 1$. Note also that if $k_t << k_z$ (this relation is always true if $\Delta << 1$ – see either Eq. (158) or Fig. 26), the longitudinal components of the fields are much smaller than their transverse counterparts, so the wave is indeed very close to the TEM one. Because of that, the ratio of the electric and magnetic field amplitudes is also close to that in the TEM wave: $E_0/H_0 \approx Z_- \approx Z_+$.

Now to satisfy the boundary conditions at the core-to-cladding interface ($\rho = R$), we need to have a similar angular dependence of these components at $\rho \ge R$. The longitudinal components of the fields

[65] This fact becomes less surprising if we recall that in the circular metallic waveguide, discussed in Sec. 6, the fundamental mode (H_{11}, see Fig. 23) also corresponded to $n = 1$ rather than $n = 0$.

are tangential to the interface and thus should be continuous. Using the solutions similar to Eq. (160) with $n = 1$, we get

$$E_z = -i\frac{k_t}{k_z}\frac{J_1(k_tR)}{K_1(\kappa_tR)}E_0K_1(\kappa_t\rho)\sin\varphi, \quad H_z = -i\frac{k_t}{k_z}\frac{J_1(k_tR)}{K_1(\kappa_tR)}H_0K_1(\kappa_t\rho)\cos\varphi, \quad \text{for } \rho \geq R. \quad (7.168)$$

For the transverse components, we should require the continuity of the normal magnetic field μH_n, for our simple field structure equal to just $\mu H_x\cos\varphi$, of the tangential electric field $E_\tau = E_y\sin\varphi$, and of the normal component of $D_n = \varepsilon E_n = \varepsilon E_y\cos\varphi$. Assuming that $\mu_- = \mu_+ = \mu_0$, and $\varepsilon_+ \approx \varepsilon_-$[66] we can satisfy these conditions with the following solutions:

$$E_x = 0, \quad E_y = \frac{J_0(k_tR)}{K_0(\kappa_tR)}E_0K_0(\kappa_t\rho), \quad H_x = \frac{J_0(k_tR)}{K_0(k_tR)}H_0K_0(\kappa_t\rho), \quad H_y = 0, \quad \text{for } \rho \geq R. \quad (7.169)$$

From here, we can calculate components from E_z and H_z, using the same approach as for $\rho \leq R$:

$$E_z = \frac{1}{-ik_z}\frac{\partial E_y}{\partial y} = -i\frac{\kappa_t}{k_z}\frac{J_0(k_tR)}{K_0(\kappa_tR)}E_0K_1(\kappa_t\rho)\sin\varphi,$$
$$H_z = \frac{1}{-ik_z}\frac{\partial H_x}{\partial x} = -i\frac{\kappa_t}{k_z}\frac{J_0(k_tR)}{K_0(\kappa_tR)}H_0K_1(\kappa_t\rho)\cos\varphi, \quad \text{for } \rho \geq R. \quad (7.170)$$

These relations provide the same functional dependence of the fields as Eqs. (167), i.e. the internal and external fields are compatible, but their amplitudes at the interface coincide only if

LP_{01} mode: characteristic equation

$$k_t\frac{J_1(k_tR)}{J_0(k_tR)} = \kappa_t\frac{K_1(\kappa_tR)}{K_0(\kappa_tR)}. \quad (7.171)$$

This characteristic equation (which may be also derived from Eq. (161) with $n = 1$ in the limit $\Delta \to 0$) looks close to Eq. (162), but functionally is much different from it – see Fig. 28. Indeed, its right-hand side is always positive, and the left-hand side tends to zero at $k_tR \to 0$. As a result, Eq. (171) may have a solution for arbitrary small values of the parameter $\mathscr{V}$ defined by Eq. (163), i.e. for *arbitrary low frequencies* (large wavelengths). This is why this mode is used in practical single-mode fibers: there are no other modes with wavelengths larger than the λ_{max} given by Eq. (165), so they cannot be unintentionally excited on small inhomogeneities of the fiber.

It is easy to use the Bessel function approximations by the first terms of the Taylor expansions (2.132) and (2.157) to show that in the limit $\mathscr{V} \to 0$, κ_tR tends to zero much faster than $k_tR \approx \mathscr{V}$: $\kappa_tR \to 2\exp\{-1/\mathscr{V}\} << \mathscr{V}$. This means that the scale $\rho_c \equiv 1/\kappa_t$ of the radial distribution of the LP_{01} wave's fields in the cladding becomes very large. In this limit, this mode may be interpreted as a virtually TEM wave propagating in the cladding, just slightly deformed (and guided) by the fiber's core. The drawback of this feature is that it requires very thick cladding, to avoid energy losses in its outer ("buffer" and "jacket") layers that defend the silica layers from the elements and mechanical damages, but lack their

[66] This is the core assumption of this approximate theory, which accounts only for the most important effect of the small difference of the dielectric constants ε_+ and ε_-: the difference between $(k_+^2 - k_z^2) = k_t^2 > 0$ and $(k_-^2 - k_z^2) = -\kappa_t^2 < 0$. For more discussion of the accuracy of this approximation and some exact results, the interested reader may be referred either to the monograph by A. Snyder and D. Love, *Optical Waveguide Theory*, Chapman and Hill, 1983, or to Chapter 3 and Appendix B in the monograph by Yariv and Yeh, that was cited above.

low optical absorption. Due to this reason, the core radius is usually selected so that the parameter $\mathscr{V}$ is just slightly less than the critical value $\mathscr{V}_c = \xi_{01} \approx 2.4$ for higher modes, thus ensuring the single-mode operation.

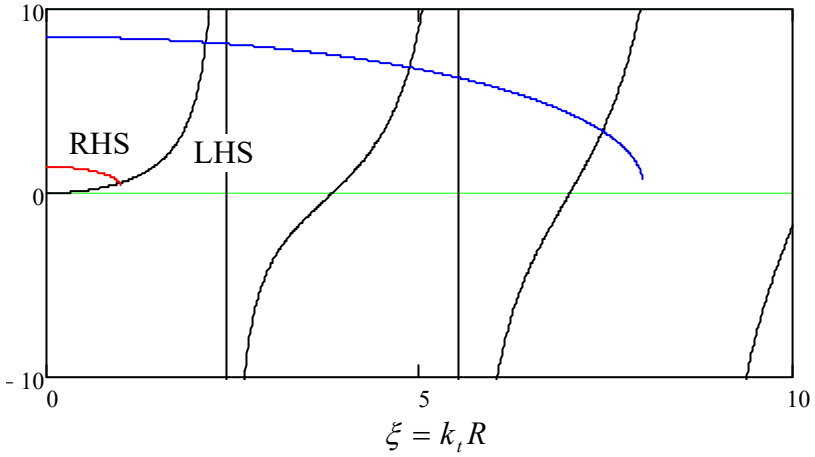


Fig. 7.28. Two sides of the characteristic equation (171) for the LP_{01} mode, plotted as a function of $k_t R$, for two values of the dimensionless parameter: $\mathscr{V} = 8$ (blue line) and $\mathscr{V} = 1$ (red line).

In order to reduce the field spread into the cladding, the step-index fibers discussed above may be replaced with *graded-index* fibers whose dielectric constant ε is gradually and slowly decreased from the center to the periphery.[67] Keeping only the main two terms in the Taylor expansion of the function $\varepsilon(\rho)$ at $\rho = 0$, we may approximate such reduction as[68]

$$\varepsilon(\rho) \approx \varepsilon(0)\left(1 - \zeta\rho^2\right), \tag{7.172}$$

where $\zeta \equiv -[(d^2\varepsilon/d\rho^2)/2\varepsilon]_{\rho=0}$ is a positive constant characterizing the fiber composition gradient. Moreover, if this constant is sufficiently small ($\zeta << k^2$), the field distribution across the fiber's cross-section may be described by the same 2D Helmholtz equation (101), but with a space-dependent transverse wave vector:[69]

$$\left[\nabla_t^2 + k_t^2(\rho)\right] f = 0, \tag{7.173}$$

where

$$k_t^2(\rho) = k^2(\rho) - k_z^2 \equiv k_t^2(0) - k^2(0)\zeta\rho^2, \quad \text{and} \quad k^2(0) \equiv \omega^2\varepsilon(0)\mu_0 .$$

Surprisingly for such an axially-symmetric problem, because of its special dependence on the radius, this equation may be most readily solved in the Cartesian coordinates. Indeed, rewriting it as

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_t^2(0) - k^2(0)\zeta\left(x^2 + y^2\right)\right] f = 0 , \tag{7.174}$$

and separating the variables as $f = X(x)Y(y)$, we get

[67] Due to the difficulty of fabrication of graded-index fibers with wave attenuation below a few dm/km, they are not used as broadly as the step-index ones.

[68] For an axially-symmetric smooth function $\varepsilon(\rho)$, the *first* derivative $d\varepsilon/d\rho$ always vanishes at $\rho = 0$, so Eq. (172) does not have a term linear in ρ.

[69] This approach is invalid at arbitrary (large) ζ because in the macroscopic Maxwell equations, $\varepsilon(\mathbf{r})$ is under the differentiation sign, and the exact Helmholtz-type equations for fields have additional terms containing $\nabla\varepsilon$.

$$\frac{1}{X}\frac{d^2X}{dx^2}+\frac{1}{Y}\frac{d^2Y}{dy^2}+k_t^2(0)-k^2(0)\zeta\left(x^2+y^2\right)=0, \tag{7.175}$$

so the functions X and Y obey similar differential equations, for example

$$\frac{d^2X}{dx^2}+\left[k_x^2-k^2(0)\zeta x^2\right]X=0, \tag{7.176}$$

with the separation constants satisfying the following condition:

$$k_x^2+k_y^2=k_t^2(0)\equiv k^2(0)-k_z^2. \tag{7.177}$$

The ordinary differential equation (176) is well known in quantum mechanics, because the stationary Schrödinger equation for one of the most important basic quantum systems, a 1D harmonic oscillator, may be rewritten in this form. Its eigenvalues are very simple:

$$(k_x^2)_n=k(0)\zeta^{1/2}(2n+1),\qquad (k_y^2)_m=k(0)\zeta^{1/2}(2m+1),\qquad \text{with } n,m=0,1,2,..., \tag{7.178}$$

but the corresponding eigenfunctions $X_n(x)$ and $Y_m(y)$ are expressed via not quite elementary functions – the Hermite polynomials.[70] For most practical purposes, however, the lowest eigenfunctions $X_0(x)$ and $Y_0(y)$ are sufficient, because they correspond to the lowest $k_{x,y}$, and hence the lowest

$$\left[k_t^2(0)\right]_{\min}=(k_x^2)_0+(k_y^2)_0=2k(0)\zeta^{1/2}, \tag{7.179}$$

and the lowest cutoff frequency. As may be readily verified by the substitution to Eq. (176), the eigenfunctions corresponding to this fundamental mode are also simple:

$$X_0(x)=\text{const}\times\exp\left\{-\frac{k(0)\zeta^{1/2}x^2}{2}\right\}, \tag{7.180}$$

and similarly for $Y_0(y)$, so the field distribution follows the Gaussian function

$$f_0(\rho)=f_0(0)\exp\left\{-\frac{k(0)\zeta^{1/2}\rho^2}{2}\right\}\equiv f_0(0)\exp\left\{-\frac{\rho^2}{2a^2}\right\},\quad \text{with } a\equiv 1/k^{1/2}(0)\zeta^{1/4}, \tag{7.181}$$

where $a >> 1/k(0)$ has the sense of the effective width of the field's extension in the radial direction, normal to the wave's propagation axis z. This is the so-called *Gaussian beam*, very convenient for some applications.

The Gaussian beam (181) is just one example of the so-called *paraxial beams*, which may be represented as a result of modulation of a plane wave with a wave number k, by an axially-symmetric *envelope function* $f(\rho)$, where $\boldsymbol{\rho}\equiv\{x, y\}$, with a relatively large effective radius $a >> 1/k$.[71] Such beams give me a convenient opportunity to deliver on the promise made in Sec. 1: calculate the angular momentum $\mathbf{L}$ of a circularly polarized wave propagating in free space, and prove its fundamental relation to the wave's energy U. Let us start from the calculation of U for a paraxial beam (with an

[70] See, e.g., QM Sec. 2.9.

[71] Note that propagating in a uniform medium, i.e. outside of grade-index fibers or other focusing systems, such beams gradually increase their width a due to diffraction – the effect to be analyzed in the next chapter.

arbitrary, but spatially localized envelope *f*) of a circularly polarized wave, with the transverse electric field components given by Eq. (19):

$$E_x = E_0 f(\rho)\cos\psi, \qquad E_y = \mp E_0 f(\rho)\sin\psi, \tag{7.182a}$$

where E_0 is the real amplitude of the wave's electric field at the propagation axis, $\psi \equiv kz - \omega t + \varphi$ is its total phase, and the two signs correspond to two possible directions of the circular polarization.[72] According to Eq. (6), the corresponding transverse components of the magnetic field are

$$H_x = \pm\frac{E_0}{Z_0} f(\rho)\sin\psi, \qquad H_y = \frac{E_0}{Z_0} f(\rho)\cos\psi\,. \tag{7.182b}$$

These expressions are sufficient to calculate the energy density (6.113) of the wave,[73]

$$u = \frac{\varepsilon_0\left(E_x^2 + E_y^2\right)}{2} + \frac{\mu_0\left(H_x^2 + H_y^2\right)}{2} = \frac{\varepsilon_0 E_0^2 f^2}{2} + \frac{\mu_0 E_0^2 f^2}{2Z_0^2} \equiv \varepsilon_0 E_0^2 f^2, \tag{7.183}$$

and hence the full energy (per unit length in the direction *z* of the wave's propagation) of the beam:

$$U = \int u d^2 r \equiv 2\pi\int_0^\infty u\rho d\rho = 2\pi\varepsilon_0 E_0^2 \int_0^\infty f^2 \rho d\rho\,. \tag{7.184}$$

However, the transverse fields (182) are insufficient to calculate a non-zero average of **L**. Indeed, following the angular moment's definition in mechanics,[74] $\mathbf{L} \equiv \mathbf{r}\times\mathbf{p}$, where **p** is the particle's (linear) momentum, we may use Eq. (6.115) for the electromagnetic field momentum's density **g** in free space, to define the field's angular momentum's density as

$$\boxed{\mathbf{l} \equiv \mathbf{r}\times\mathbf{g} \equiv \frac{1}{c^2}\mathbf{r}\times\mathbf{S} \equiv \frac{1}{c^2}\mathbf{r}\times(\mathbf{E}\times\mathbf{H}).} \tag{7.185}$$

EM field's angular momentum

Let us use the familiar *bac minus cab* rule of the vector algebra[75] to transform this expression to

$$\mathbf{l} = \frac{1}{c^2}\left[\mathbf{E}(\mathbf{r}\cdot\mathbf{H}) - \mathbf{H}(\mathbf{r}\cdot\mathbf{E})\right] \equiv \frac{1}{c^2}\left\{\mathbf{n}_z\left[E_z(\mathbf{r}\cdot\mathbf{H}) - H_z(\mathbf{r}\cdot\mathbf{E})\right] + \left[\mathbf{E}_t(\mathbf{r}\cdot\mathbf{H}) - \mathbf{H}_t(\mathbf{r}\cdot\mathbf{E})\right]\right\}. \tag{7.186}$$

If the field is purely transverse ($E_z = H_z = 0$), as it is in a strictly plane wave, the first square brackets in the last expression vanish, while the second bracket gives an azimuthal component of **l**, which oscillates in time and vanishes at its time averaging. (This is exactly the reason why I have not tried to calculate **L** during our first discussion of the circularly polarized waves in Sec. 1.)

---

[72] For our task of calculating two *quadratic* forms of the fields (**L** and *U*), their real representation (182) is more convenient than the complex-exponent one. However, for *linear* manipulations, the latter representation of the circularly polarized waves, $\mathbf{E}_t = E_0 f(\rho)\mathrm{Re}[(\mathbf{n}_x \pm i\mathbf{n}_y)\exp\{i\psi\}]$, $\mathbf{H}_t = (E_0/Z_0)f(\rho)\mathrm{Re}[(\mp i\mathbf{n}_x + \mathbf{n}_y)\exp\{i\psi\}]$, is usually more convenient, and is broadly used.

[73] Note that, in contrast to a linearly-polarized wave (16), the energy density of a circularly-polarized wave does not depend on the full phase ψ – in particular, on *t* at fixed *z*, or vice versa. This is natural because its field vectors rotate (keeping their magnitude) rather than oscillate – see Fig. 3b.

[74] See, e.g., CM Eq. (1.31).

[75] See, e.g., MA Eq. (7.5).

Fortunately, our discussion of optical fibers, in particular, the derivation of Eqs. (167), (168), and (170) gives us a clear clue on how to resolve this paradox. If the envelope function $f(\rho)$ differs from a constant, the transverse wave components (182) alone do *not* satisfy the Maxwell equations (2b), which necessitates longitudinal components E_z and H_z of the fields, with[76]

$$\frac{\partial E_z}{\partial z} = -\frac{\partial E_x}{\partial x} - \frac{\partial E_y}{\partial y}, \qquad \frac{\partial H_z}{\partial z} = -\frac{\partial H_x}{\partial x} - \frac{\partial H_y}{\partial y}. \tag{7.187}$$

However, as these expressions show, if the envelope function f changes very slowly in the sense $df/d\rho \sim f/a << kf$, the longitudinal components are very small and do not have a back effect on the transverse components. Hence, the above calculation of U is still valid (asymptotically, at $ka \to 0$), and we may still use Eqs. (182) on the right-hand side of Eqs. (187),

$$\frac{\partial E_z}{\partial z} = E_0\left(-\frac{\partial f}{\partial x}\cos\psi \pm \frac{\partial f}{\partial x}\sin\psi\right), \qquad \frac{\partial H_z}{\partial z} = \frac{E_0}{Z_0}\left(\mp\frac{\partial f}{\partial x}\sin\psi - \frac{\partial f}{\partial x}\cos\psi\right), \tag{7.188}$$

and integrate them over z as

$$\begin{aligned} E_z &= E_0\int\left(-\frac{\partial f}{\partial x}\cos\psi \pm \frac{\partial f}{\partial x}\sin\psi\right)dz = \frac{E_0}{k}\left(-\frac{\partial f}{\partial x}\int\cos\psi\, d\psi \pm \frac{\partial f}{\partial x}\int\sin\psi\, d\psi\right) \\ &\equiv \frac{E_0}{k}\left(-\frac{\partial f}{\partial x}\sin\psi \mp \frac{\partial f}{\partial x}\cos\psi\right). \end{aligned} \tag{7.189a}$$

Here the integration constant is taken for zero because no wave field component may have a time-independent part. Integrating, absolutely similarly, the second of Eqs. (188), we get

$$H_z = \frac{E_0}{kZ_0}\left(\pm\frac{\partial f}{\partial x}\cos\psi - \frac{\partial f}{\partial y}\sin\psi\right). \tag{7.189b}$$

With the same approximation, we may calculate the longitudinal (z–) component of $\mathbf{l}$, given by the first term of Eq. (186), keeping only the dominating, transverse fields (182) in the scalar products:

$$l_z = E_z\left(\mathbf{r}\cdot\mathbf{H}_t\right) - H_z\left(\mathbf{r}\cdot\mathbf{E}_t\right) \equiv E_z\left(xH_x + yH_y\right) - H_z\left(xE_x + yE_y\right). \tag{7.190}$$

Plugging in Eqs. (182) and (189), and taking into account that in free space, $k = \omega/c$, and hence $1/Z_0c^2k = \varepsilon_0/\omega$, we get:

$$l_z = \mp\frac{\varepsilon_0E_0^2}{\omega}\left(xf\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}\right) \equiv \mp\frac{\varepsilon_0E_0^2}{2\omega}\left[x\frac{\partial\left(f^2\right)}{\partial x} + y\frac{\partial\left(f^2\right)}{\partial y}\right] \equiv \mp\frac{\varepsilon_0E_0^2}{2\omega}\boldsymbol{\rho}\cdot\boldsymbol{\nabla}\left(f^2\right) \equiv \mp\frac{\varepsilon_0E_0^2}{2\omega}\rho\frac{d\left(f^2\right)}{d\rho}. \tag{7.191}$$

Hence the total angular momentum of the beam (per unit length), is

$$L_z = \int l_z d^2r \equiv 2\pi\int_0^\infty l_z\rho d\rho = \mp\pi\frac{\varepsilon_0E_0^2}{\omega}\int_0^\infty \rho^2\frac{d\left(f^2\right)}{d\rho}d\rho \equiv \mp\pi\frac{\varepsilon_0E_0^2}{\omega}\int_{\rho=0}^{\rho=\infty}\rho^2 d\left(f^2\right). \tag{7.192}$$

Taking this integral by parts, with the assumption that $\rho f \to 0$ at $\rho \to 0$ and $\rho \to \infty$ (as it is true for the Gaussian beam (181) and all realistic paraxial beams), we finally get

[76] The complex-exponential versions of these equalities are given by the bottom line of Eq. (100).

$$L_z = \pm\pi \frac{\varepsilon_0 E_0^2}{\omega} \int_0^\infty f^2 d\left(\rho^2\right) \equiv \pm 2\pi \frac{\varepsilon_0 E_0^2}{\omega} \int_0^\infty f^2 \rho d\rho \,. \tag{7.193}$$

Now comparing this expression with Eq, (184), we see that remarkably, the ratio L_z/U does not depend on the shape and the width of the beam (and of course on the wave's amplitude E_0), so these parameters are very simply and universally related:

$$\boxed{L_z = \pm \frac{U}{\omega}} \,. \tag{7.194}$$

Angular momentum at circular polarization

Since this relation is valid in the plane-wave limit $a \to \infty$, it may be attributed to plane waves as well, with the understanding that in reality, they always have some width ("aperture") restriction.

As the reader certainly knows, in quantum mechanics the energy excitations of any harmonic oscillator of frequency ω are quantized in the units of $\hbar\omega$, while the internal angular momentum of a particle is quantized in the units of $s\hbar$, where s is its spin. In this context, the classical relation (194) is used in quantum electrodynamics as the basis for treating the electromagnetic field excitation quanta (*photons*) as some sort of quantum particles with spin $s = 1$. (Such integer spin also fits the Bose-Einstein statistics of the electromagnetic radiation.)

Unfortunately, I do not have time/space for a further discussion of the (very interesting) physics of paraxial beams but cannot help noticing, at least in passing, the very curious effect of *helical waves* – the beams carrying not only the "spin" momentum (194), but also an additional "orbital" angular momentum. The distribution of their energy in space is not monotonic, as it is in the Gaussian beam (181), but reminds several threads twisted around the propagation axis – hence the term "helical".[77] Mathematically, their field structure is described by the *associate Laguerre polynomials* – the same special functions that are used for the quantum-mechanical description of hydrogen-like atoms.[78] Presently, there are efforts to use such beams for the so-called *orbital angular momentum* (OAM) multiplexing for high-rate information transmission.[79]

7.8. Resonant cavities

Generally, *resonators* are structures that may sustain oscillations (in electrodynamics, of the electromagnetic field) even without an external source, until the oscillation amplitude slowly decreases in time due to unavoidable energy losses. If the resonator quality (described by the so-called *Q-factor*, which will be defined and discussed in the next section) is high, $Q >> 1$, this decay takes many oscillation periods. Alternatively, high-Q resonators may sustain high oscillating fields permanently, if driven by relatively weak incident waves. In contrast to lumped-element resonators, say, the well-known *LC tank circuit*, the subject of this section is *resonant cavities* (or "distributed resonators") limited by either conducting or dielectric walls that contain distributed standing waves inside them.

[77] Noticing such solutions of the Maxwell equations may be traced back to at least a 1943 theoretical work by J. Humblet; however, this issue had not been discussed in literature too much until experiments carried out in 1992 – see, e.g. L. Allen *et al.*, *Optical Angular Momentum;* IOP, 2003.

[78] See, e.g., QM Sec. 3.7.

[79] See, e.g., J. Wang *et al.*, *Nature Photonics* **6**, 488 (2012).

Conceptually the simplest resonant cavity is the *Fabry-Pérot interferometer*[80] that may be obtained by placing two well-conducting planes parallel to each other.[81] Indeed, in Sec. 3 we have seen that if a plane wave is normally incident on such a "perfect mirror", located at $z = 0$, its reflection, at negligible skin depth, results in a standing wave described by Eq. (61b):

$$E(z,t) = \mathrm{Re}\left(2E_\omega e^{-i\omega t + i\pi/2}\right)\sin kz\,. \tag{7.195}$$

This wave would not change if we place the second mirror (isolating the segment of length l from the external wave source) at any position $z = l$ with $\sin kl = 0$, i.e. with

$$kl = p\pi, \quad \text{where } p = 1, 2, ... \tag{7.196}$$

This condition, which determines the spectrum of *own* (or *resonance*, or *eigen-*) *frequencies* of the resonator of fixed length l,

$$\omega_p = vk_p = \frac{\pi v}{a}p, \qquad \text{with } v = \frac{1}{(\varepsilon\mu)^{1/2}}, \tag{7.197}$$

has a simple physical sense: the resonator's length l equals exactly p half-waves of the frequency ω_p. Though this is all very simple, please note a considerable change of philosophy from what we have been doing in the previous sections: the main task of the resonator's analysis is finding its own frequencies ω_p, which are now determined by the system's geometry rather than by an external wave source.

Before we move to cavities of more complex shapes, let us use Eq. (62) to represent the magnetic field in the Fabry-Pérot interferometer:

$$H(z,t) = \mathrm{Re}\left(2\frac{E_\omega}{Z}e^{-i\omega t}\right)\cos kz\,. \tag{7.198}$$

Expressions (195) and (198) show that in contrast to traveling waves, each field of the standing wave changes simultaneously (proportionately) at all points of the Fabry-Pérot resonator, turning to zero everywhere twice a period. At these instants, the energy of the corresponding field vanishes, but the total energy of the two fields stays constant because the counterpart field oscillates with the phase shift $\pi/2$. Such behavior is typical for all electromagnetic resonators.

A more technical remark is that we can readily get the same results (195)-(198) by solving the Maxwell equations from scratch. For example, we already know that in the absence of dispersion, losses, and sources, they are reduced to wave equations (3) for any field components. For the Fabry-Pérot resonator's analysis, we can use the 1D form of these equations, say, for the transverse component of the electric field:

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)E = 0, \tag{7.199}$$

and solve it as a part of an eigenvalue problem with the corresponding boundary conditions. Indeed, by separating time and space variables as $E(z, t) = Z(z)\mathcal{T}(t)$, we obtain

[80] This device, named after its inventors, Charles Fabry and Alfred Pérot; is also called the *Fabry-Pérot etalon* (meaning "gauge"), because of its initial usage for light wavelength measurements.

[81] The resonators formed by well-conducting (usually, metallic) walls are frequently called *resonant cavities*.

$$\frac{1}{Z}\frac{d^2Z}{dz^2} - \frac{1}{v^2}\frac{1}{\mathcal{T}}\frac{d^2\mathcal{T}}{dt^2} = 0. \tag{7.200}$$

Calling the separation constant k^2, we get two similar ordinary differential equations,

$$\frac{d^2Z}{dz^2} + k^2 Z = 0, \qquad \frac{d^2\mathcal{T}}{dt^2} + k^2 v^2 \mathcal{T} = 0, \tag{7.201}$$

both with sinusoidal solutions, so the product $Z(z)\mathcal{T}(t)$ is a standing wave with the wave vector k and frequency $\omega = kv$. (In this form, the equations are valid even in the presence of dispersion, but with a frequency-dependent wave speed: $v^2 = 1/\varepsilon(\omega)\mu(\omega)$.) Now using the boundary conditions $E(0, t) = E(l, t) = 0$,[82] we get the eigenvalue spectrum for k_p and hence for $\omega_p = vk_p$, given by Eqs. (196) and (197).

Lessons from this simple case study may be readily generalized to any cavity formed as a transmission line's section:[83] there are two approaches to finding the resonant frequency spectrum:

(i) We may look at a traveling wave solution and find where reflecting mirrors may be inserted without affecting the wave's structure.

(ii) We may solve the general 3D wave equations,

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) f(\mathbf{r}, t) = 0, \tag{7.202}$$

for field components, as an eigenvalue problem with appropriate boundary conditions. If the system's parameters (and hence the coefficient v) do not change in time, the spatial and temporal variables of Eq. (202) may be *always* separated by taking

$$f(\mathbf{r}, t) = \sum_k \mathcal{f}_k(\mathbf{r})\mathcal{T}_k(t), \tag{7.203}$$

where each function $\mathcal{T}_k(t)$ *always* obeys the same equation as in Eq. (201), having the sinusoidal solution of frequency $\omega_k = vk$. Plugging this solution back into Eqs. (202), for the spatial distribution of the field, we get the *3D Helmholtz equation*,

$$\boxed{\left(\nabla^2 + k^2\right)\mathcal{f}_k(\mathbf{r}) = 0,} \tag{7.204}$$

3D Helmholtz equation

whose eigenfunctions $\mathcal{f}_k(\mathbf{r})$ may be much more involved, especially for non-symmetric geometries.

Let us use these approaches to find the resonant frequency spectrum of a few simple, but practically important cavities. First of all, the first method is completely sufficient for the analysis of any resonator formed as a fragment of a uniform TEM transmission line (e.g., a coaxial cable), confined with two conducting lids normal to the line's direction. Indeed, since in such lines $k_z = k = \omega/v$, and the electric field is perpendicular to the propagation axis, e.g., parallel to the lid surface, the boundary conditions are exactly the same as in the Fabry-Pérot resonator, and we again arrive at the eigenfrequency spectrum (197).

[82] This is of course the expression of the first of the general boundary conditions (104). The second of these conditions (for the magnetic field) is satisfied automatically for the transverse waves we are considering.

[83] The resonators may have different geometries as well, and in many cases, only the second approach may be used.

Now let us analyze a slightly more complex system: a rectangular metallic-wall cavity of volume $a \times b \times l$ – see Fig. 29. To use the first approach outlined above, let us consider the resonator as a finite-length ($\Delta z = l$) section of the rectangular waveguide extended along the z-axis, which was analyzed in detail in Sec. 6. As a reminder, at $a < b$, in the fundamental H_{10} traveling wave mode, both vectors **E** and **H** do not depend on y, with **E** having only a y-component. In contrast, **H** has two components, H_x and H_z, with the phase shift $\pi/2$ between them, and with H_x having the same phase as E_y – see Eqs. (131), (137), and (138). Hence, if a plane perpendicular to the z-axis, is placed so that the electric field vanishes on it, H_x also vanishes, so both boundary conditions (104), pertinent to a perfect metallic wall, are fulfilled simultaneously.

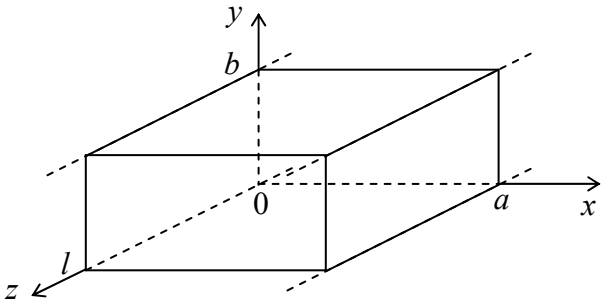


Fig. 7.29. Rectangular metallic-wall resonator as a finite section of a waveguide with the cross-section shown in Fig. 22.

As a result, the H_{10} wave would not be perturbed by two metallic walls separated by an integer number of half-wavelengths $\lambda_z/2$ corresponding to the wave number given by the combination of Eqs. (102) and (133):

$$k_z = \left(k^2 - k_t^2\right)^{1/2} = \left(\frac{\omega^2}{v^2} - \frac{\pi^2}{a^2}\right). \tag{7.205}$$

Using this expression, we see that the smallest of these distances, $l = \lambda_z/2 = \pi/k_z$, gives the resonance frequency[84]

$$\omega_{101} = v\left[\left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{l}\right)^2\right]^{1/2}, \tag{7.206}$$

where the indices of ω show the numbers of half-waves along each dimension of the system, in the order [a, b, l]. This is the lowest ("fundamental") frequency of the resonator (if $b < a, l$).

The field distribution in this mode is close to that in the corresponding waveguide mode H_{10} (Fig. 22), with the important difference that the magnetic and electric fields are now shifted by phase $\pi/2$ both in space and time, just as in the Fabry-Pérot resonator – see Eqs. (195) and (198). Such a time shift allows for a very simple interpretation of the H_{101} mode, which is especially adequate for very flat resonators, with $b << a, l$. At the instant when the electric field reaches its maximum (Fig. 30a), i.e. when the magnetic field vanishes in the whole volume, the surface electric charge of the broadest (in Fig. 30, horizontal) walls of the resonator is largest, being localized mostly near the centers of the walls. At the immediate later times, the walls start to recharge via surface currents, whose density J is largest in the side walls, and reaches its maximal value in a quarter of the oscillation period $\mathcal{T} = 2\pi/\omega_{101}$ – see Fig. 30b. The currents generate the vortex magnetic field, with looped field lines in the plane of the

[84] In most electrical engineering handbooks, the index corresponding to the shortest side of the resonator is listed last, so the fundamental mode is nominated as H_{110} and its eigenfrequency as ω_{110}.

broadest face of the resonator. The surface currents continue to flow in this direction until (in one more quarter period) the broader walls of the resonator are fully recharged in the polarity opposite to that shown in Fig. 30a. After that, the surface currents start to flow in the direction opposite to that shown in Fig. 30b. This process, which repeats again and again, is conceptually similar to the well-known oscillations in a lumped *LC* circuit, with the role of (now, distributed) capacitance played mostly by the broadest walls of the resonator, and that of (now, distributed) inductance, by its narrower walls.

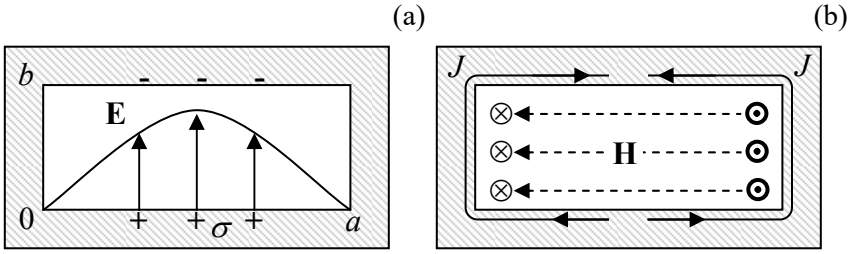


Fig. 7.30. Fields, charges, and currents in the fundamental (H_{101}) mode of a rectangular metallic resonator, at two instants separated by $\Delta t = \pi/2\omega_{101}$ – schematically.

In order to generalize Eq. (206) to higher oscillation modes, the second of the approaches discussed above is more prudent. Separating the variables in the Helmholtz equation (204) as $\mathcal{R}(\mathbf{r}) = X(x)Y(y)Z(z)$, we see that X, Y, and Z have to be either sinusoidal or cosinusoidal functions of their arguments, with the wave vector components satisfying the characteristic equation

$$k_x^2 + k_y^2 + k_z^2 = k^2 \equiv \frac{\omega^2}{v^2}. \tag{7.207}$$

In contrast to the wave propagation problem, now we are dealing with standing waves along all three dimensions, and have to satisfy the macroscopic boundary conditions (104) on all sets of parallel walls. It is straightforward to check that these conditions ($E_\tau = 0$, $H_n = 0$) are fulfilled at the following field component distribution:

$$\begin{aligned} E_x &= E_1 \cos k_x x \sin k_y y \sin k_z z, & H_x &= H_1 \sin k_x x \cos k_y y \cos k_z z, \\ E_y &= E_2 \sin k_x x \cos k_y y \sin k_z z, & H_y &= H_2 \cos k_x x \sin k_y y \cos k_z z, \\ E_z &= E_3 \sin k_x x \sin k_y y \cos k_z z, & H_z &= H_3 \cos k_x x \cos k_y y \sin k_z z, \end{aligned} \tag{7.208}$$

with each of the wave vector components having an equidistant spectrum similar to Eq. (196):

$$k_x = \frac{\pi n}{a}, \quad k_y = \frac{\pi m}{b}, \quad k_z = \frac{\pi p}{l}, \tag{7.209}$$

so the full spectrum of resonance frequencies is given by the following formula:

$$\omega_{nmp} = vk = v\left[\left(\frac{\pi n}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2 + \left(\frac{\pi p}{l}\right)^2\right]^{1/2}, \tag{7.210}$$

which is a natural generalization of Eq. (206). Note, however, that of the three integers m, n, and p, at least two have to be different from zero to keep the fields (208) from vanishing at all points.

We may use Eq. (210), in particular, to evaluate the number of different modes in a relatively small range $d^3k \ll k^3$ of the wave vector space volume that is, on the other hand, much larger than the reciprocal volume, $1/V = 1/abl$, of the cavity. Taking into account that each eigenfrequency (210), with

$nml \neq 0$, corresponds to two field modes with different polarizations,[85] the argumentation absolutely similar to the one used for the 2D case at the end of Sec. 7, yields

Oscillation mode density

$$dN = 2V\frac{d^3k}{(2\pi)^3}. \tag{7.211}$$

This property, valid for resonators of arbitrary shape, is broadly used in classical and quantum statistical physics,[86] in the following form. If some electromagnetic mode functional $f(\mathbf{k})$ is a smooth function of the wave vector $\mathbf{k}$, and the volume V is large enough, then Eq. (211) may be used to approximate the sum of the functional's values over the modes by an integral:

$$\sum_{\mathbf{k}} f(\mathbf{k}) \approx \int_N f(\mathbf{k})dN \equiv \int_{\mathbf{k}} f(\mathbf{k})\frac{dN}{d^3k}d^3k = 2\frac{V}{(2\pi)^3}\int_{\mathbf{k}} f(\mathbf{k})d^3k\,. \tag{7.212}$$

Leaving similar analyses of resonant cavities of some other simple shapes for the reader's exercises, let me finish this section by noting that low-loss resonators may be also formed by finite-length sections of not only metallic-wall waveguides of various cross-sections but also of dielectric waveguides. Moreover, even a simple slab of a dielectric material with a μ/ε ratio substantially different from that of its environment (say, of the free space) may be used as a high-Q Fabry-Pérot interferometer (Fig. 31), due to an effective wave reflection from its surfaces at the normal and especially an inclined incidence – see, respectively, Eqs. (68), and Eqs. (91) and (95).

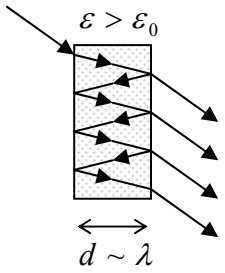


Fig. 7.31. A dielectric Fabry-Pérot interferometer.

Actually, such dielectric Fabry-Pérot interferometers are frequently more convenient for practical purposes than metallic-wall resonators, not only due to possibly lower losses (especially in the optical range) but also due to a natural coupling to the environment, which offers a ready way of wave insertion and extraction – see Fig. 31 again. The backside of the same medal is that this coupling to the environment provides an additional mechanism of power losses, limiting the resonance's quality factor – see the next section.

7.9. Energy loss effects

The inevitable energy losses ("dissipation") in passive media lead, in two basic situations, to two different effects. In a long transmission line fed by a constant wave source, the losses lead to a gradual

[85] This fact becomes evident from plugging Eqs. (208) into the Maxwell equation $\nabla\cdot\mathbf{E} = 0$. The resulting equation, $k_xE_1 + k_yE_2 + k_zE_3 = 0$, with the discrete, equidistant spectrum (209) for each wave vector component, may be always satisfied by two linearly independent sets of the constants $E_{1,2,3}$.

[86] See, e.g., QM Sec. 1.1 and SM Sec. 2.6.

attenuation of the wave, i.e. to a decrease of its amplitude, and hence its power $\mathscr{P}$, with the growing distance z from the source. In linear materials, the power losses are proportional to the power $\mathscr{P}$ carried by the wave, so the energy balance on a small segment dz takes the form

$$d\mathscr{P}_{\text{loss}} \equiv -d\mathscr{P} \equiv -\frac{d\mathscr{P}}{dz}dz = \alpha\mathscr{P}dz\,. \tag{7.213}$$

The coefficient α participating in the last form of Eq. (213), and hence defined as

$$\alpha \equiv -\frac{d\mathscr{P}/dz}{\mathscr{P}}, \tag{7.214}$$

is called the *attenuation constant*.[87] Comparing the solution of Eq. (213),

$$\boxed{\mathscr{P}(z) = \mathscr{P}(0)e^{-\alpha z}\,,} \tag{7.215}$$

Wave's attenuation

with Eq. (29), where k is replaced with k_z, we see that α may be expressed as

$$\alpha = 2\,\text{Im}\,k_z\,, \tag{7.216}$$

where k_z is the component of the wave vector along the transmission line. In the most important limit when the losses are low in the sense $\alpha << \left|\ k_z\ \right| \approx \text{Re}\ k_z$, its effects on the field distribution along the line's cross-section are negligible, making the calculation of α rather straightforward. In particular, in this limit, the contributions to attenuation from two major sources, the energy losses in the filling dielectric and the skin-effect losses in conducting walls, are independent and additive.

The dielectric losses are especially simple to describe. Indeed, a review of our calculations in Secs. 5-7 shows that all of them remain valid if either $\varepsilon(\omega)$, or $\mu(\omega)$, or both, and hence $k(\omega)$, have small imaginary parts:

$$k'' = \omega\,\text{Im}\left[\varepsilon^{1/2}(\omega)\mu^{1/2}(\omega)\right] << k'\,. \tag{7.217}$$

In TEM transmission lines, $k_z = k$, and hence Eq. (216) yields

$$\boxed{\alpha_{\text{filling}} = 2k'' = 2\omega\,\text{Im}\left[\varepsilon^{1/2}(\omega)\mu^{1/2}(\omega)\right].} \tag{7.218}$$

Attenuation due to filling

For dielectric waveguides, in particular optical fibers, these losses are the main attenuation mechanism. As was discussed in Sec. 7, in practical optical fibers $\kappa_t R >> 1$, i.e. most of the field propagates (as an evanescent wave) in the cladding, with a field distribution very close to the TEM wave. This is why Eq. (218) is approximately valid if it is applied to the cladding material alone. In waveguides with non-TEM waves, we can use the relations between k_z and k, derived in the previous sections, to re-calculate k'' into Im k_z. (Note that at this recalculation, the values of k_t have to be kept real, because they are just the eigenvalues of the Helmholtz equation (101), which does not include the filling media parameters.).

---

[87] In engineering, wave attenuation is most frequently measured in *decibels per meter*, abbreviated as db/m (the term not to be confused with dBm standing for decibel-milliwatt):

$$\alpha_{\text{db/m}} \equiv 10\,\log_{10}\frac{\mathscr{P}(z)}{\mathscr{P}(z+1\,\text{m})} = 10\,\log_{10}e^{\alpha[1/\text{m}]} \equiv \frac{10}{\ln 10}\alpha\left[\text{m}^{-1}\right] \approx 4.343\,\alpha\left[\text{m}^{-1}\right].$$

Alternatively, it is sometimes measured in *nepers per meter* (Np/m) defined as $\alpha_{\text{np/m}} \equiv \alpha/2$, so $\alpha_{\text{db/m}} \approx 8.686\ \alpha_{\text{Np/m}}$.

In transmission lines and waveguides and with metallic walls, higher energy losses may come from the skin effect. If the wavelength λ is much larger than δ_s, as it usually is,[88] the losses may be readily evaluated using Eq. (6.36):

$$\frac{d\mathscr{P}_{\rm loss}}{dA} \equiv -\frac{d\mathscr{P}}{dA} = H_{\rm wall}^2 \frac{\mu\omega\delta_s}{4}, \tag{7.219}$$

where $H_{\rm wall}$ is the real amplitude of the tangential component of the magnetic field at the wall's surface. The total power loss $\mathscr{P}_{\rm loss}/dz$ per unit length of a waveguide, i.e. the right-hand side of Eq. (213), now may be calculated by the integration of this expression along the contour(s) limiting the cross-section of all conducting walls. Since our calculation is only valid for low losses, we may ignore their effect on the field distribution, so the unperturbed distributions may be used both in Eq. (219), i.e. in the numerator of Eq. (214), and also for the calculation of the average propagating power, i.e. the denominator of Eq. (214) – as the integral of the Poynting vector over the cross-section of the waveguide.

Let us see how this approach works for the TEM mode in one of the simplest transmission lines, the coaxial cable (Fig. 20). As we already know from Sec. 5, in the coarse-grain approximation, implying negligible power loss, the TEM mode field distributions between the two conductors are the same as in statics, namely:

$$H_z = 0, \qquad H_\rho = 0, \qquad H_\varphi(\rho) = H_0 \frac{a}{\rho}, \tag{7.220}$$

where H_0 is the field's amplitude on the surface of the inner conductor, and

$$E_z = 0, \qquad E_\rho(\rho) = ZH_\varphi(\rho) = ZH_0 \frac{a}{\rho}, \qquad E_\varphi = 0, \qquad \text{where } Z \equiv \left(\frac{\mu}{\varepsilon}\right)^{1/2}. \tag{7.221}$$

Neglecting the power losses for a minute, we may plug these expressions into Eq. (42) to calculate the time-averaged Poynting vector:

$$\overline{S} = \frac{Z\left|H_\varphi(\rho)\right|^2}{2} = \frac{Z\left|H_0\right|^2}{2}\left(\frac{a}{\rho}\right)^2, \tag{7.222}$$

and from it, the total wave power flow through the cross-section:

$$\mathscr{P} = \int_A \overline{S} d^2 r = \frac{Z\left|H_0\right|^2 a^2}{2} 2\pi \int_a^b \frac{\rho d\rho}{\rho^2} = \pi Z \left|H_0\right|^2 a^2 \ln\frac{b}{a}. \tag{7.223}$$

Next, for this particular system (Fig. 20), the contours limiting the wall cross-section are circles of radii $\rho = a$ (where the surface field amplitude $H_{\rm walls}$ equals, in our notation, H_0), and $\rho = b$ (where, according to Eq. (214), the field is a factor of b/a lower). As a result, for the power loss per unit length, Eq. (219) yields

$$\frac{d\mathscr{P}_{\rm loss}}{dz} = \oint_{C_a+C_b} \frac{d\mathscr{P}_{\rm loss}}{dA} dl = \left(2\pi a\left|H_0\right|^2 + 2\pi b\left|H_0\frac{a}{b}\right|^2\right)\frac{\mu_0\omega\delta_s}{4} = \frac{\pi}{2}a\left(1+\frac{a}{b}\right)\mu\omega\delta_s\left|H_0\right|^2. \tag{7.224}$$

[88] As follows from Eq. (78), which may be used for crude estimates even in cases of arbitrary wave incidence, this condition is necessary for low attenuation: $\alpha << k$ only if $\ell << 1$.

Note that at $a << b$, the losses in the inner conductor dominate, despite its smaller surface, because of the higher surface field.

Now we may plug Eqs. (223) and (224) into the definition (214) of α, to calculate the skin-effect contribution to the attenuation constant:

$$\alpha_{\rm skin} \equiv \frac{d\mathscr{P}_{\rm loss}/dz}{\mathscr{P}} = \frac{1}{2\ln(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right)\frac{\mu\omega\delta_{\rm s}}{Z} \equiv \frac{k\delta_{\rm s}}{2\ln(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right). \tag{7.225}$$

This result shows that the relative (dimensionless) attenuation, α/k, scales approximately as the ratio $\delta_{\rm s}/\min[a, b]$, in a semi-quantitative agreement with the plane-wave result (78).

Let us use this result to evaluate α for the standard TV cable RG-6/U, with copper conductors of diameters $2a = 1$ mm, $2b = 4.7$ mm, and $\varepsilon \approx 2.2\varepsilon_0$ and $\mu \approx \mu_0$. According to Eq. (6.33), for $f = 100$ MHz (i.e. $\omega \approx 6.3\times10^8$ s^{-1}) the skin depth of pure copper at room temperature (with $\sigma \approx 6.0\times10^7$ S/m) is close to 6.5×10^{-6} m, while $k = \omega(\varepsilon\mu)^{1/2} = (\varepsilon/\varepsilon_0)^{1/2}(\omega/c) \approx 3.1$ m^{-1}. As a result, the attenuation is rather low: $\alpha_{\rm skin} \approx 0.016$ m^{-1}, so the attenuation length scale $l_{\rm d} \equiv 1/\alpha$ is about 60 m. Hence the attenuation in a cable connecting a roof TV antenna or a cable distribution box to a TV set is not a big problem, though using a worse conductor, e.g., steel, would make the losses rather noticeable. (Hence the current worldwide shortage of copper.) However, the use of such cable in the X-band ($f \sim 10$ GHz) is more problematic. Indeed, though the skin depth $\delta_{\rm s} \propto \omega^{-1/2}$ decreases with frequency, the wavelength drops, i.e. k increases, even faster ($k \propto \omega$), so the attenuation $\alpha_{\rm skin} \propto \omega^{1/2}$ becomes close to 0.16 m^{-1}, i.e. $l_{\rm d}$ to ~6 m. This is why at such frequencies, it may be necessary to use rectangular waveguides, with their larger internal dimensions a, $b \sim 1/k$, and hence lower attenuation. Let me leave the calculation of this attenuation, using Eq. (219) and the results derived in Sec. 7, for the reader's exercise.

The main effect of dissipation on free oscillations in *resonators* is different: here it leads to a gradual decay of the oscillating fields' energy U in time. A useful dimensionless measure of this decay, called the *Q factor*, is commonly defined by writing the following temporal analog of Eq. (213):[89]

$$-dU \equiv \mathscr{P}_{\rm loss}dt = \frac{\omega}{Q}Udt, \tag{7.226}$$

where ω in the resonance frequency in the loss-free limit, and

$$\boxed{\frac{\omega}{Q} \equiv \frac{\mathscr{P}_{\rm loss}}{U}}. \tag{7.227}$$ Q-factor

The solution of Eq. (226),

$$U(t) = U(0)e^{-t/\tau}, \qquad \text{with } \tau \equiv \frac{Q}{\omega} \equiv \frac{Q/2\pi}{\omega/2\pi} = \frac{Q\mathcal{T}}{2\pi}, \tag{7.228}$$

which is the temporal analog of Eq. (215), shows the physical meaning of the Q-factor: the characteristic time τ of the oscillation energy's decay is $(Q/2\pi)$ times longer than the oscillation period $\mathcal{T} = 2\pi/\omega$. (Another useful interpretation of Q comes from the universal relation[90]

[89] As losses grow, the oscillation waveform deviates from the sinusoidal one, and the very notion of "oscillation frequency" becomes vague. As a result, the parameter Q is well-defined only if it is much higher than 1.

[90] See, e.g., CM Sec. 5.1.

$$Q = \frac{\omega}{\Delta\omega}, \tag{7.229}$$

where $\Delta\omega$ is the so-called *FWHM* [91] *bandwidth* of the resonance, namely the difference between the two values of the external signal frequency, one above and one below ω, at which the energy of the oscillations induced in the resonator by an input signal is twice lower than its resonance value.)

In the important particular case of a resonant cavity formed by the insertion of metallic walls into a TEM transmission line of a small cross-section (with the linear size scale a much less than the wavelength λ), there is no need to calculate the Q-factor directly, provided that the line attenuation coefficient α is already known. In fact, as was discussed in Sec. 8 above, the standing waves in such a resonator, of the length given by Eq. (196): $l = p(\lambda/2)$ with $p = 1, 2,\ldots$, may be understood as an overlap of two TEM waves propagating in opposite directions, or in other words, a traveling wave plus its reflection from one of the ends, the whole roundtrip taking time $\Delta t = 2l/v = p\lambda/v = 2\pi p/\omega = p\mathcal{T}$. According to Eq. (215), at this distance, the wave's power drops by a factor of $\exp\{-2\alpha l\} = \exp\{-p\alpha\lambda\}$. On the other hand, the same decay may be viewed as taking place in time, and according to Eq. (228), results in the drop by a factor of $\exp\{-\Delta t/\tau\} = \exp\{-(p\mathcal{T})/(Q/\omega)\} = \exp\{-2\pi p/Q\}$. Comparing these two exponents, we get

Q vs. α

$$Q = \frac{2\pi}{\alpha\lambda} = \frac{k}{\alpha}. \tag{7.230}$$

This simple relation neglects the losses at the wave reflection from the walls limiting the resonator length. This approximation is indeed legitimate at $l >> \lambda$; if this relation is violated, or if we are dealing with more complex resonator modes (such as those based on the reflection of E or H waves), the Q-factor may be different from that given by Eq. (230), and needs to be calculated directly from Eq. (227). A substantial relief for such a direct calculation is that, just at the calculation of small attenuation in waveguides, in the low-loss limit ($Q >> 1$), both the numerator and denominator of the right-hand side of that formula may be calculated neglecting the effects of the energy loss on the field distribution in the resonator. I am leaving such a calculation, for a few simple resonant cavities, including the rectangular and the circular ones, for the reader's exercise.

To conclude this chapter, the last remark: in some distributed resonators (including certain dielectric resonators[92] and metallic cavities with holes in their walls), additional energy losses due to the wave radiation into the environment are also possible. In some simple cases (say, the Fabry-Pérot interferometer shown in Fig. 31), the calculation of these *radiative losses* is straightforward, but sometimes it requires more elaborated approaches that will be discussed in the next chapter.

7.10. Exercise problems

7.1.* Find the temporal Green's function of a medium whose complex permittivity $\varepsilon(\omega)$ obeys the Lorentz oscillator model given by Eq. (32), by using:

(i) the Fourier transform of the underlying Eq. (7.30), and

[91] FWHM is the acronym for *Full Width at Half-Maximum*.

[92] Due to the absence of metallic walls and the associated skin-effect losses, some microwave dielectric resonators (e.g., those based on pure sapphire crystals cooled to helium temperatures) may feature Q-factors as high as 10^9 – see, e.g., D. Creedon *et al.*, *Appl. Phys.* **98**, 222903 (2011).

(ii) the direct solution of that equation.

Hint: For the Fourier-transform approach, you may like to use the Cauchy integral.[93]

7.2. The electric polarization of some material responds to an electric field step[94] in the following way:

$$P(t) = \varepsilon_1 E_0 \left(1 - e^{-t/\tau}\right), \qquad \text{if } E(t) = E_0 \times \begin{cases} 0, & \text{for } t < 0, \\ 1, & \text{for } 0 < t, \end{cases}$$

where $\tau > 0$ and ε_1 are some constants. Calculate the complex permittivity $\varepsilon(\omega)$ of this material, and discuss a possible simple physical model giving such dielectric response.

7.3. Calculate the complex permittivity $\varepsilon(\omega)$ of a material whose dielectric-response Green's function defined by Eq. (23), is

$$G(\theta) = G_0 \left(1 - e^{-\theta/\tau}\right),$$

with some positive constants G_0 and τ. What is the difference between this dielectric response and the apparently similar one considered in the previous problem?

7.4. Use the oscillator model of an atom, given by Eq. (30), to calculate its average potential energy in a uniform, sinusoidal ac electric field, and use the result to calculate the potential profile created for the atom by a standing electromagnetic wave with the electric field amplitude $E_\omega(\mathbf{r})$.

7.5. The solution of the previous problem shows that a standing electromagnetic wave may exert a time-averaged force on an otherwise free non-relativistic charged particle. Reveal the physics of this force by writing and solving the equations of motion of such a particle in:

(i) a linearly-polarized monochromatic plane traveling wave, and
(ii) a similar but standing wave.

7.6. Use the first of Eqs. (54) to relate the integral $\int_0^\infty \varepsilon''(\Omega)\Omega d\Omega$ to the plasma frequency for the Lorentz oscillator model of a system of non-interacting particles.

7.7. Prove that Eq. (6.42) cannot be correct for all frequencies, and suggest its correction making the result compatible with both the causality principle and the physical model (6.39).

7.8. Calculate, sketch, and discuss the dispersion relation for electromagnetic waves propagating in a medium described by the Lorentz oscillator model (32), for the case of negligible damping.

7.9. As was briefly discussed in Sec. 2,[95] a wave pulse of a finite but relatively large spatial extension $\Delta z >> \lambda \equiv 2\pi/k$ may be formed as a *wave packet* – a sum of sinusoidal waves with wave

[93] See, e.g., MA Eq. (15.2).
[94] This function $E(t)$ is of course proportional to the well-known Heaviside step function $\theta(t)$ – see, e.g., MA Eq. (14.3). I am not using this notion here just to avoid confusion between two different uses of the Greek letter θ.
[95] For even more detail, see CM Sec. 5.3 and especially QM Sec. 2.2.

vectors **k** within a relatively narrow interval. Consider an electromagnetic plane wave packet of this type, with the electric field distribution

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\int_{-\infty}^{+\infty} \mathbf{E}_k e^{i(kz-\omega_k t)}dk, \qquad \text{with } k = \omega_k\left[\varepsilon(\omega_k)\mu(\omega_k)\right]^{1/2},$$

propagating along the z-axis in an isotropic, linear, and dissipation-free (but not necessarily dispersion-free) medium. Express the full energy of the packet (per unit area of the wave's front) via the complex amplitudes $\mathbf{E}_k$, and discuss its dependence on time.

7.10. Prove the Lorentz reciprocity relation (6.121) for a linear isotropic medium.

7.11.* A plane wave of frequency ω is normally incident, from free space, on a plane surface of a collision-free plasma with the electron density growing linearly and slowly with the distance from the surface: $n = \gamma z$ for $z \geq 0$, where $\gamma > 0$ is a small constant. Calculate the functional form of the resulting standing wave's "tail" inside the plasma.

7.12.* Analyze the effect of a time-independent uniform magnetic field $\mathbf{B}_0$, parallel to the direction **n** of an electromagnetic wave propagation, on the wave's dispersion in plasma, within the same simple model that was used in Sec. 2 for the derivation of Eq. (38). (Limit your analysis to relatively weak waves, whose magnetic field is much smaller than $\mathbf{B}_0$.)

Hint: You may like to represent the incident wave as a linear superposition of two circularly polarized waves, with opposite polarization directions.

7.13. A monochromatic plane electromagnetic wave is normally incident, from free space, on a uniform slab with electric permittivity ε and magnetic permeability μ, with the slab's thickness d comparable with the wavelength.

(i) Calculate the power transmission coefficient $\mathcal{T}$, i.e. the fraction of the incident wave's power, that is transmitted through the slab.

(ii) Assuming that ε and μ are frequency-independent and positive, analyze in detail the frequency dependence of $\mathcal{T}$. In particular, how does the function $\mathcal{T}(\omega)$ depend on the slab's thickness d and the wave impedance $Z \equiv (\mu/\varepsilon)^{1/2}$ of its material?

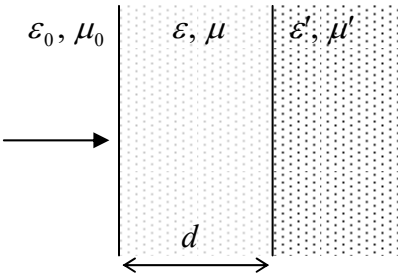
7.14. A plane electromagnetic wave with a free-space wave number k_0 is normally incident on a planar conducting film of thickness $d \sim \delta_s << 1/k_0$. Calculate the power transmission coefficient of the system and analyze the result in the limits of small and large values of the ratio d/δ_s.

7.15. One of the results of the previous problem's solution was the following expression for the coefficient of power transmission of a plane electromagnetic wave through a thin conducting film of thickness $d << \delta_s, \lambda$, at normal incidence:

$$\mathcal{T} = \frac{1}{\left(1 + Z_0/2R_\square\right)^2},$$

where $R_\square \equiv 1/\sigma d$ is the sheet resistance ("resistance per square") of the film. Derive this formula in a simpler way, utilizing the smallness of d from the very beginning. Also, calculate the power reflection coefficient $\mathcal{R}$, compare it with $\mathcal{T}$, and comment.

7.16. A plane wave of frequency ω is normally incident, from free space, on a plane surface of a material with real electric permittivity ε' and magnetic permeability μ'. To minimize the wave's reflection from the surface, it may be covered with a layer, of thickness d, of another transparent material – see the figure on the right. Calculate the optimal values of ε, μ, and d.



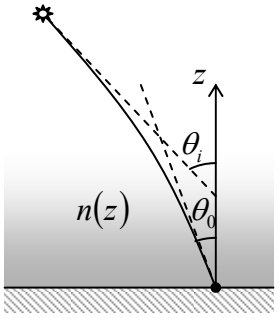
7.17. A monochromatic plane wave is incident from inside a medium with $\varepsilon\mu > \varepsilon_0\mu_0$ on its planar surface, at an incidence angle θ larger than the critical angle $\theta_c = \sin^{-1}(\varepsilon_0\mu_0/\varepsilon\mu)^{1/2}$. Calculate the depth δ of the evanescent wave penetration into the free space, and analyze its dependence on θ. Does the result depend on the wave's polarization?

7.18. Calculate the critical angle θ_c for a wave of frequency ω, incident from free space upon a planar surface of a plasma with electron density n, and discuss the implications of the result for ultraviolet and X-ray optics.

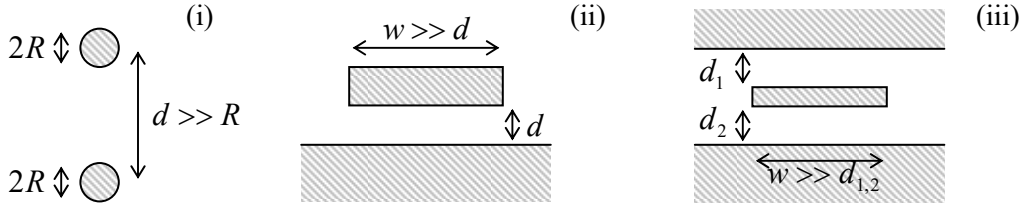
7.19. Analyze the possibility of propagation of surface electromagnetic waves along a planar boundary between plasma and free space. In particular, calculate and analyze the dispersion relation of the waves.

Hint: Assume that the magnetic field of the wave is parallel to the boundary and perpendicular to the wave's propagation direction. (After solving the problem, justify this mode choice.)

7.20. Light from a very distant source arrives to an observer through a plane layer of nonuniform medium with a certain 1D gradient of its refraction index $n(z)$, at angle θ_0 – see the figure on the right. What is the genuine direction θ_i to the source, if $n(z) \to 1$ at $z \to \infty$? (This problem is evidently important for high-precision astronomical measurements at the Earth's surface.)



7.21. Calculate the TEM impedance Z_W of uniform TEM transmission lines with well-conducting electrodes and the cross-sections shown in the figure below:



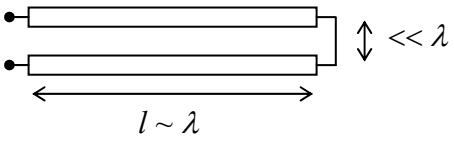
(i) two parallel round wires separated by distance $d >> R$,
(ii) a *microstrip line* of width $w >> d$,
(iii) a *stripline* with $w >> d_1 \sim d_2$,

in all cases using the coarse-grain boundary conditions on conductor surfaces. Assume that the conductors are embedded into a linear dielectric with constant ε and μ.

7.22. Modify the solution of Task (ii) of the previous problem for a superconductor microstrip line, taking into account the magnetic field's penetration into both the strip and the ground plane.

7.23.* What lumped ac circuit would be equivalent to the TEM-line system shown in Fig. 19, with an incident wave's power $\mathscr{P}_i$? Assume that the wave reflected from the lumped load circuit does not return to it.

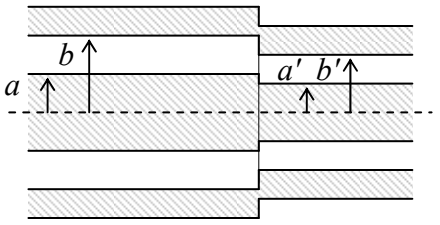
7.24. Find the lumped ac circuit equivalent to a loss-free TEM transmission line of length $l \sim \lambda$, with a small cross-section area $A << \lambda^2$, as "seen" (measured) from one end, if the line's conductors are galvanically connected ("shortened") at the other end – see the figure on the right. Discuss the result's dependence on the signal frequency.



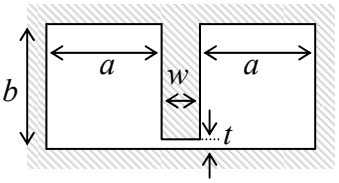
7.25. Represent the fundamental H_{10} wave in a rectangular waveguide (Fig. 22) with a sum of two plane waves, and discuss the physics behind such a representation.

7.26.* For the coaxial cable (see, e.g., Fig. 20), find the lowest non-TEM mode and calculate its cutoff frequency.

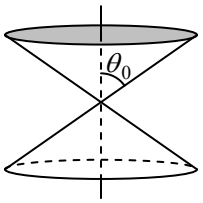
7.27. Two coaxial cable sections are connected coaxially – see the figure on the right, which shows the system's cut along its symmetry axis. Relations (118) and (120) seem to imply that if the ratios b/a of these sections are equal, their impedance matching is perfect, i.e. a TEM wave incident from one side on the connection would pass it without any reflection at all: $R = 0$. Is this statement correct?



7.28. Calculate the cutoff frequencies ω_c of the fundamental mode and the next lowest mode in the so-called *ridge waveguide* with the cross-section shown in the figure on the right, in the limit $t << a, b, w$. Briefly discuss possible advantages and drawbacks of such waveguides for signal transfer and physical experiment.



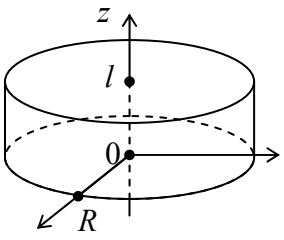
7.29. Prove that TEM-like waves may propagate, in the radial direction, in the free space between two coaxial, round, well-conducting cones – see the figure on the right. Can this system be characterized by a certain transmission line impedance Z_W, as defined by Eq. (115)?



7.30. Use the recipe outlined in Sec. 7 to prove the characteristic equation (161) for the *HE* and *EH* waves in step-index optical fibers with a round cross-section.

7.31. Derive an approximate equation describing spatial variations of the complex amplitude of a general monochromatic paraxial beam propagating in a uniform medium, for the case when these variations are sufficiently slow. Is the Gaussian beam described by Eq. (181) one of the possible solutions of this equation? Give your interpretation of the last result.

7.32. Calculate the lowest resonance frequencies and the corresponding field distributions of standing electromagnetic waves inside a round cylindrical cavity with well-conducting walls (see the figure on the right), neglecting the skin depth δ_s in comparison with l and R.



7.33. Analyze electromagnetic waves that may propagate inside a relatively narrow gap between two well-conducting concentric spherical shells of radii R and $R + d$, in the limit $d << R$.

(i) Within the coarse-grain approximation, derive the 2D equation describing such waves with relatively large wavelengths $\lambda \sim R >> d$.

(ii) Calculate the lowest resonance frequencies of the system.

7.34. A molecule with an electric polarizability α is placed inside an otherwise empty macroscopic cavity with well-conducting walls. Express the resulting shifts of its resonance frequencies via the unperturbed field distribution in the corresponding mode.

7.35. A plane monochromatic wave propagates through a medium with an Ohmic conductivity σ and negligible electric and magnetic polarization effects. Calculate the wave's attenuation and relate the result to a certain calculation carried out in Chapter 6.

7.36. Generalize the telegrapher's equations (110)-(111) by accounting for small energy losses:

(i) in the transmission line's conductors, and

(ii) in the medium separating the conductors,

using their simplest (Ohmic) models. Formulate the conditions of validity of the resulting equations.

7.37. Calculate the skin-effect contribution to the attenuation constant α of a TEM wave in the microstrip line discussed in Problem 21 (ii).

7.38. Calculate the skin-effect contribution to the attenuation coefficient α defined by Eq. (214), for the fundamental (H_{10}) mode propagating in a waveguide with well-conducting walls, of a rectangular cross-section – see Fig.22. Use the results to evaluate the wave decay length $l_d \equiv 1/\alpha$ of a 10 GHz wave in the standard X-band waveguide WR-90 (with copper walls, $a = 23$ mm, $b = 10$ mm, and no dielectric filling), at room temperature. Compare the result with that (obtained in Sec. 9) for the standard TV coaxial cable, at the same frequency.

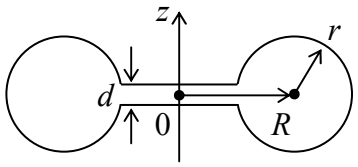
7.39. Calculate the skin-effect contribution to the attenuation coefficient α of

(i) the fundamental (H_{11}) wave, and
(ii) the H_{01} wave,

in a conductor-wall waveguide with the circular cross-section (see Fig. 23a), and analyze the low-frequency ($\omega \to \omega_c$) and high-frequency ($\omega >> \omega_c$) behaviors of α for each of these modes.

7.40. For a rectangular cavity of dimensions $a \times b \times l$, with $b \le a, l$, calculate the Q-factor of the fundamental oscillation mode, due to the skin-effect losses in its conducting walls. Evaluate the factor for a $23 \times 23 \times 10$ mm^3 cavity with copper walls, at room temperature.

7.41.* Calculate the lowest eigenfrequency and the Q-factor (due to the skin-effect losses) of the axially symmetric toroidal cavity with well-conducting walls and the interior's cross-section shown in the figure on the right, for the case $d << r, R$.[96]



7.42. Express the contribution to the damping coefficient (the reciprocal Q-factor) of a resonant cavity, by small energy losses in the dielectric that fills it, via the complex functions $\varepsilon(\omega)$ and $\mu(\omega)$ of the material.

7.43. For the dielectric Fabry-Pérot resonator (Fig. 31) with the normal wave incidence, calculate the Q-factor due to radiation losses, in the limit of a strong impedance mismatch ($Z >> Z_0$), by using two approaches:

(i) from the energy balance, using Eq. (227), and
(ii) from the frequency dependence of the power transmission coefficient, using Eq. (229).

Compare the results.

[96] Such resonators are broadly used in particle accelerators and also in vacuum electron devices for high-power microwave amplification and generation (e.g., the so-called *klystrons*), where the electric field has to be concentrated in the region of charged particle passage – typically, along the symmetry axis (in the figure above, the z-axis), through a pair of small holes in the cavity's walls, which do not affect the field distribution substantially.

This page is intentionally left blank