Chapter 7. Electromagnetic Wave Propagation

This (long!) chapter focuses on the most important effect that follows from the time-dependent Maxwell equations, namely the electromagnetic waves, at this stage avoiding a discussion of their origin, i.e. radiation. I start from the simplest, plane waves in a uniform and isotropic media. The next step is a discussion non-uniform systems, in particular those with sharp boundaries between different materials, which bring in such new effects as wave reflection and refraction. Then I will proceed to the structure of electromagnetic waves propagating along various long, cylindrical structures, called transmission lines - such as coaxial cables, waveguides, and optical fibers. In the end of the chapter, electromagnetic oscillations in final-length fragments of such lines, serving as resonators, are also discussed.

7.1. Plane waves

Let us start from considering a spatial region that does not contain field sources ($\rho = 0$, $\mathbf{j} = 0$), and is filled with a linear, uniform, isotropic medium, which obeys Eqs. (3.38) and (5.110):

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

Moreover, let us assume for a minute that these material equations hold for all frequencies of interest. As was already shown in Sec. 6.7, in this case the Lorenz gauge condition (6.108) allows the Maxwell equations to be recast into wave equations (6.110) for the vector and scalar potentials. However, for most our purposes it is more convenient to use directly the homogeneous Maxwell equations (6.94) for the electric and magnetic fields - which are independent of the gauge choice. After the elementary elimination of **D** and **B** using Eq. (1), [1] these equations take a simple, 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, taking the curl ($\nabla\times$) of each of Eqs. (2a), and using the vector algebra identity (5.31), whose first term, for both **E** and **H**, vanishes due to Eqs. (2b), we get similar wave equations for the electric and magnetic fields:

Electro-magnetic wave equations

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

where parameter v is defined by relation

EM wave velocity

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

with $v^2 = 1/\varepsilon_0\mu_0 \equiv c^2$ in free space.

[1] Though **B** rather then **H** is the actual (microscopically-averaged) magnetic field, it is mathematically more convenient (just as in Sec. 6.2) to use the latter vector in our current discussion, because at sharp media boundaries, **H** obeys the boundary condition (5.118) similar to that for **E** – see Eq. (3.47).

Two vector equations (3) are of course six similar equations for three Cartesian components of two vectors **E** and **H**. Each of these equations allows, in particular, the following solution,

$$f = f(z - vt), \tag{7.5}$$

Plane wave

where z is the Cartesian coordinate along a certain (arbitrary) direction **n**. This solution describes a specific type of a *wave*, i.e. a certain field pattern moving, without deformation, along axis z, with velocity v. According to Eq. (5), each variable f has the same value in each plane perpendicular to the direction **n** of wave propagation, hence the 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 solution (5) into Eqs. (2a), we get

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

Relation between the fields

where constant Z is defined as

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

Wave impedance

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

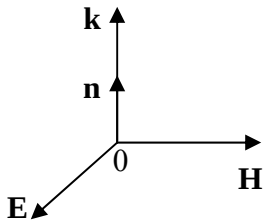


Fig. 7.1. Field vectors in a plane electromagnetic wave propagating along direction **n**.

Second, the field magnitudes are related by constant Z, called the *wave impedance* of the medium. Very soon we will see that the wave 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).[2] In particular, in free space,

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

Wave impedance of free space

Now plugging Eq. (6) into Eqs. (6.104b) and (6.105), we get:

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

Wave's energy per unit volume

[2] In Gaussian units, E and H have the same dimensionality (in particular, in a free-space wave, $E = H$), making the (very useful) notion of the wave impedance less manifestly exposed - and in some textbooks not mentioned at all.

Wave's power per unit area

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

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

$$\mathbf{S} = \mathbf{n}uv. \tag{7.9c}$$

In the view of the Poynting vector paradox discussed in Sec. 6.7 (see Fig. 6.10), one may wonder whether this expression may be interpreted as the actual density of power flow. In contrast to the static situation shown in Fig. 6.7, that limits the electric and magnetic fields to a vicinity of their sources, waves may travel far from them. As a result, they can form *wave packets* of finite length in free space – see Fig. 2.

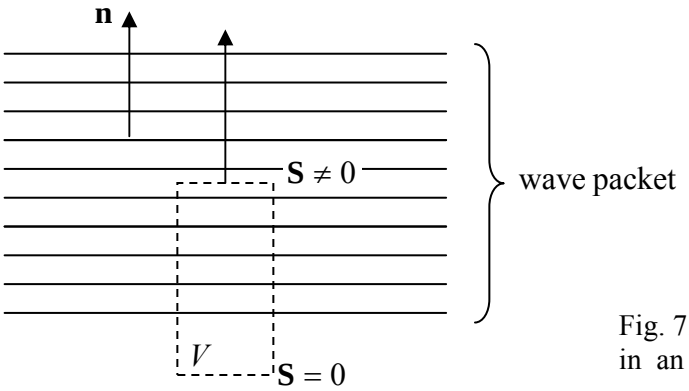


Fig. 7.2. Interpreting the Poynting vector in an electromagnetic wave.

Let us apply the Poynting theorem (6.103) to the cylinder shown by 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. (6.103), the rate of change of the full energy $\mathscr{E}$ inside the volume is $d\mathscr{E}/dt = -SA$ (where A is the lid area), so that 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 Eq. (9) 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 wave impedances of virtually all wave transmission lines, so we may use is and Eq. (9) to get some sense of how 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 (the power density 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 an 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),

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

gives $B = \mu_0 H = E/c \sim 7\times10^{-8}$T, i.e. a magnetic field thousand times less than the Earth field, and about 8 orders of magnitude lower than the field of a typical permanent magnet. A possible interpretation of this huge difference is that the scale of magnetic fields $B \sim E/c$ in the waves is "normal" for

electromagnetism, while that of permanent magnet fields is abnormally high, because they are due to the ferromagnetic alignment of electron spins, essentially quantum objects – see the discussion in Sec. 5.5.

As soon as ε and μ are simple constants, wave speed v is also constant, and Eq. (5) is valid for an arbitrary function f - defined by either initial or boundary conditions. In plain English, a medium with frequency-independent ε and μ supports propagation of plane waves with an arbitrary waveform without either decay (*attenuation*) or 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 see that all our prior results remain valid even in that general case, provided that we limit them to single-frequency (i.e. sinusoidal, or *monochromatic*) waves. Such waves may be most conveniently presented as[3]

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

Mono-chromatic wave

where f_ω is the *complex amplitude* of the wave, and k is its *wave number* (the magnitude of *wave vector* $\mathbf{k} \equiv \mathbf{n}k$), sometimes also 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 T:

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

Spatial and temporal frequencies

Requiring Eq. (11) to be a particular form of Eq. (5), i.e. the argument $(kz - \omega t) \equiv k[z - (\omega/k)t]$ to be proportional to $(z - vt)$, so that $\omega/k = v$, we see that the wave number should equal

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

Dispersion relation

showing that in this "dispersion-free" case the *dispersion relation* $\omega(k)$ is linear.

Now note that Eq. (6) does not claim mean vectors **E** and **H** retain their direction in space. (The simple case when they do is called the *linear polarization* of the wave.) Indeed, nothing in the Maxwell equations prevents, for example, joint rotation of this pair of vectors around the fixed vector **n**, while still keeping all these three vectors perpendicular to each other at all times. An arbitrary rotation law, or even an arbitrary constant frequency of such rotation, however, would violate the single-frequency (monochromatic) character of the elementary sinusoidal wave (11). In order to understand what is the most general type of polarization the wave may have without violating that condition, let us present two Cartesian components of one of these vectors (say, **E**) along any two fixed axes x and y, perpendicular to each other and axis z (i.e. vector **n)**, in the same form as used in Eq. (11):

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

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

[3] Due to the linearity of Eqs (2), operator Re in Eq. (11) may be ignored until the end of almost any calculation. Because of that, the exponential presentation of monochromatic variables is more convenient than manipulation with sine and cosine functions. (See also CM Sec. 4.1.)

In the simplest case when the arguments of the 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:

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

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

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

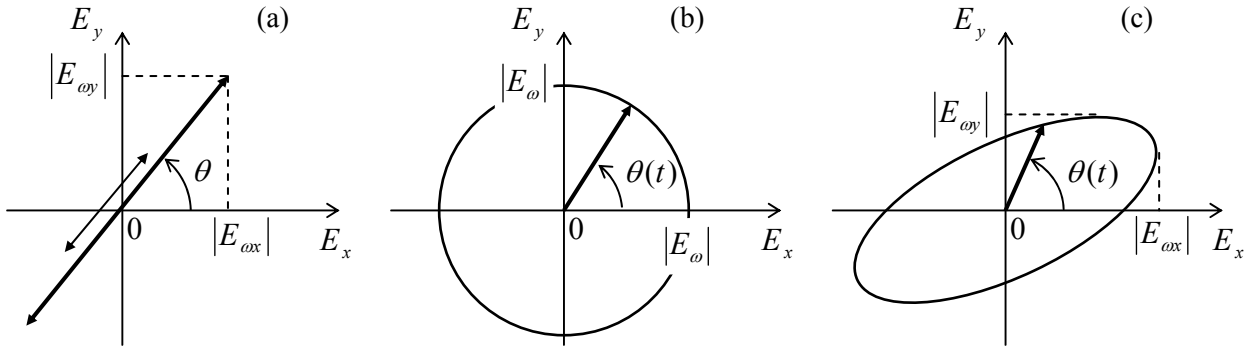


Fig. 7.3. Time evolution of the electric field vector in (a) linearly-polarized, (b) circularly-polarized, and (c) elliptically-polarized waves.

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 $-\pi/2$:

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

In this case

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

This means that on the [x, y] plane, the end of vector **E** moves, with 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*.[4] These particular solutions of the Maxwell equations are very convenient for quantum electrodynamics, because single electromagnetic field quanta with a

[4] In the convention that dominates research and engineering literature (but unfortunately is not universal), the wave is called *right-polarized* (RP) if it is described by the lower sign in Eqs. (18)-(20), and *left-polarized* (LP) in the opposite case. Another popular term for these cases is the "waves of negative / positive *helicity*".

certain (positive or negative) spin direction may be considered as elementary excitations of the corresponding circularly-polarized wave. (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 circularly-polarized waves with opposite helicities, just as Eqs. (14) present 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 electric field vector end moves along an ellipse on the [x, y] plane (Fig. 3c), such wave is called *elliptically polarized*. The eccentricity and orientation of the ellipse are completely described by one complex number, the ratio $E_{\omega x}/E_{\omega y}$, i.e. two real numbers: $|E_{\omega x}/E_{\omega y}|$ and $\varphi = \arg(E_{\omega x}/E_{\omega y})$.

The same information may be expressed via four so-called *Stokes parameters* s_0, s_1, s_2, s_3, which are popular in optics because they may be used for the description of not only completely coherent waves that are discussed here, but also of party coherent or even fully incoherent waves - including the *natural light* emitted by thermal sources like our Sun. In contrast to the notion of coherent waves whose complex amplitudes are considered deterministic numbers, the instant amplitudes of incoherent waves should be treated as stochastic variables.[5]

7.2. Attenuation and dispersion

Now let me show that *any* linear, isotropic medium may be characterized, by complex, frequency-dependent electric permittivity $\varepsilon(\omega)$ and magnetic permeability $\mu(\omega)$. Indeed, starting from electric effects, the electric polarization of realistic elementary dipoles of the medium cannot follow the applied electric field instantly, if the field frequency ω is comparable with those of the internal processes - say, transitions between atomic energy levels. Let us consider the most general law of time evolution of polarization $P(t)$ for the case of arbitrary applied electric field $E(t)$,[6] but for a sufficiently dilute medium, so that the local electric field $\mathbf{E}_{ef}$ (3.63), acting on each elementary dipole, is essentially the microscopically-averaged field $\mathbf{E}$.[7] Then, due to the linear superposition principle, $P(t)$ should be a linear sum (integral) of the values of $E(t')$ at all previous moments of time, $t' < t$, weighed by some function of t and t':

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

Temporal Green's function

The condition $t' < t$, which is implied by this relation, expresses a key principle of physics, the *causal relation* between a cause (in our case, the electric field applied to each dipole) and its effect (the

[5] For further reading about the Stokes parameters, as well as about many optics topics I will not have time to cover (especially the geometrical optics and the diffraction-imposed limits on optical imaging resolution), I can recommend the classical text by M. Born *et al*., *Principles of Optics*, 7th ed., Cambridge U. Press, 1999.

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

[7] Note that this condition (which excludes, in particular, the molecular-field effects discussed in Sec. 3.5) is not mentioned in most E&M textbooks. If the molecular fields are important, Eq. (21) and its corollaries are only valid for the relation between **P** and the effective local electric field $\mathbf{E}_{ef}$.

polarization it creates). Function $G(t, t')$ is called the *temporal Green's function* for the electric polarization.[8] In order to understand its physical sense, let us consider the case when the applied field $E(t)$ is a very short pulse at $t = t_0$, that may be approximated with the Dirac's delta-function:

$$E(t) = \delta(t - t_0) . \tag{7.22}$$

Then Eq. (21) yields just $P(t) = G(t, t_0)$, showing that the Green's function is just the polarization at moment t, created by a unit δ-functional pulse of the applied field at moment t' (Fig. 4). Thus, the temporal G is the exact time analog of the spatial Green's functions $G(\mathbf{r}, \mathbf{r}')$ we have already studied in the electrostatics – see Sec. 2.7.

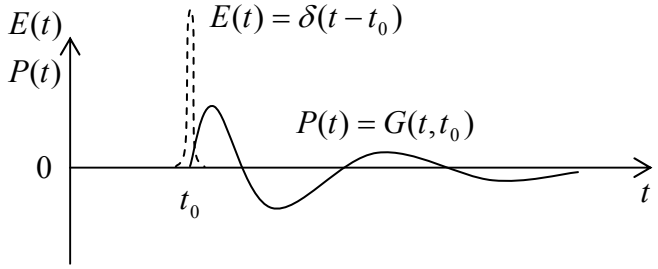


Fig. 7.4. Temporal Green's function for electric polarization (schematically).

What are the general properties of the temporal Green's function? First, the function is evidently real, since the dipole moment $\mathbf{p}$ and hence polarization $\mathbf{P} = n\mathbf{p}$ are real by the definition – see Eq. (3.6). Next, 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 function G oscillates approaching zero) depends on the medium. Finally, if parameters of the medium do not change in time, the polarization response to an electric field pulse should depend not on its absolute timing, but only on the time difference $\theta \equiv t - t'$ between the pulse and observation instants:

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

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

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

The expression in square brackets is of course nothing more that the complex amplitude P_ω of the polarization. This means that though even if the static relation (3.35) $P = \chi_e \varepsilon_0 E$ is invalid for an arbitrary time-dependent process, we may still keep its Fourier analog,

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

[8] A discussion of the temporal Green's functions in application to classical oscillations may be also found in CM Sec. 4.1.

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. (3.38):

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

Complex electric permittivity

Then, according to Eq. (3.36), the permittivity is related to the temporal Green's function by the usual Fourier transform:

$$\varepsilon(\omega) \equiv \varepsilon_0 + \frac{P_\omega}{E_\omega} = \varepsilon_0 + \int_0^\infty G(\theta) e^{i\omega\theta} d\theta . \tag{7.27}$$

It is evident from this expression that $\varepsilon(\omega)$ may be complex,

$$\varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega), \quad \varepsilon'(\omega) = \varepsilon_0 + \int_0^\infty G(\theta) \cos\omega\theta \, d\theta, \quad \varepsilon''(\omega) = \int_0^\infty G(\theta) \sin\omega\theta \, d\theta , \tag{7.28}$$

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.

Absolutely similar arguments show that the linear magnetic properties may be characterized with complex, frequency-dependent permeability $\mu(\omega)$. Now rewriting Eqs. (1) for the complex amplitudes of the fields at a particular frequency, we may 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

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

so that the wave impedance and wave number may be both complex functions of frequency.

This fact has important consequences for the electromagnetic wave propagation. First, plugging the presentation 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.[9] Second, if the waveform is not sinusoidal (and hence should be presented 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.[10]

[9] 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 also due to effects different from absorption.

[10] The reader is probably familiar with the most noticeable effect of the dispersion, namely the difference between that *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 $d^2\omega / d^2 k'$, leads to the deformation (gradual broadening) of the envelope itself. Following tradition,

Let us consider a simple but very important *Lorentz oscillator model* of a dispersive medium.[11] In dilute atomic or molecular systems (including gases), electrons respond to the external electric field especially strongly when frequency ω is close to certain eigenfrequencies ω_j corresponding to the spectrum of quantum transitions of a single atom/molecule. An approximate, phenomenological description of this behavior may be obtained from a classical model of several externally-driven harmonic oscillators with finite damping. For an oscillator, driven by electric field's force $F(t) = qE(t)$, we can write the 2nd Newton law as

Lorentz oscillator model

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

where ω_0 is the own frequency of the oscillator, and δ its damping coefficient. For a sinusoidal field, $E(t) = \mathrm{Re}\,[E_\omega \exp\{-i\omega t\}]$, we can look for a particular, *forced-oscillation* solution in a similar form $x(t) = \mathrm{Re}\,[x_\omega \exp\{-i\omega t\}]$.[12] Plugging this solution into Eq. (30), we can readily find the complex amplitude of these oscillations:

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

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 (27) we get

$$\varepsilon(\omega) = \varepsilon_0 + \frac{nq^2}{m}\frac{1}{(\omega_0^2 - \omega^2) - 2i\omega\delta}\,. \tag{7.32}$$

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

$\varepsilon(\omega)$ in Lorentz oscillator model

$$\varepsilon(\omega) = \varepsilon_0 + n\frac{q^2}{m}\sum_j \frac{f_j}{(\omega_j^2 - \omega^2) - 2i\omega\delta_j}\,, \tag{7.33}$$

where $f_j \equiv n_j/n$ is the fraction of oscillators with eigenfrequency ω_j, so that 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 effect of oscillator resonances 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 frequencies ω_j, where derivative $d\varepsilon'(\omega)/d\omega$ is negative.[13] Thus, for a system of for weakly-damped oscillators, Eq. (33) may be approximated, at most frequencies, as a sum of odd singularities ("poles"):

---

these effects are discussed in more detail in the quantum-mechanics part of my lecture notes (QM Sec. 2.1), because they are the crucial factor of Schrödinger's wave mechanics. (See also CM Sec. 5.3.)

[11] This example is focused on the frequency dependence of ε, because electromagnetic waves interact with "usual" media via their electric field much more than via the magnetic field. However, as will be discussed in Sec. 7, forgetting about the possible dispersion of $\mu(\omega)$ might result in missing some remarkable opportunities for manipulating the waves.

[12] If this point is not absolutely clear, please see CM Sec. 3.1.

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

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

This result is especially important because, according to quantum mechanics,[14] Eq. (34) 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*, $\Sigma_j f_j = 1$.

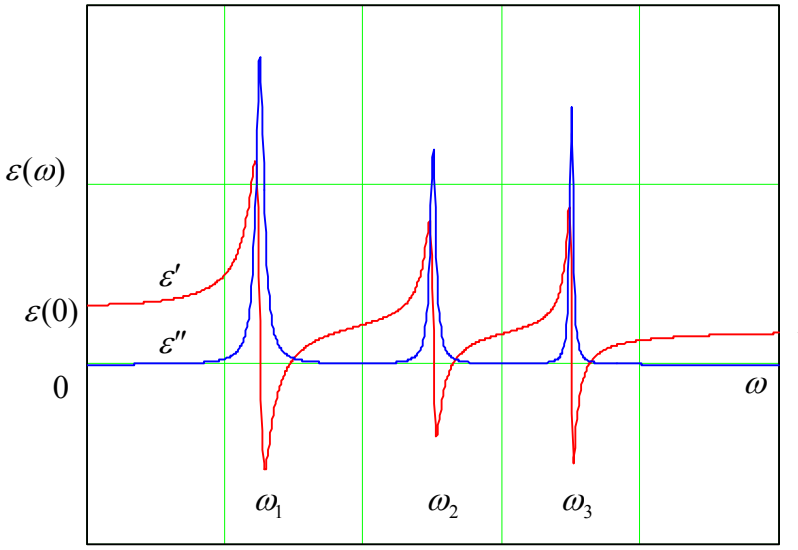


Fig. 7.5. Typical frequency dependence of the real and imaginary parts of the electric permittivity of a media in the Lorentz oscillator model.

At $\omega \to 0$, the imaginary part of the permittivity 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 in Eq. (35) is just the effective spring constant $\kappa_j = m_j\omega_j^2$ of the j^{th} oscillator, so that the oscillator masses m_j as such are actually (and quite naturally) not involved in the static dielectric response.

In the opposite limit $\omega >> \omega_j, \delta_j$, permittivity (33) also becomes real, and may be presented as

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

$\varepsilon(\omega)$ in plasma

The last result is very important, because it is also valid at *all* frequencies if all ω_j and δ_j vanish, i.e. for a gas of free charged particles, in particular for *plasmas* – ionized atomic gases, with negligible collision effects. (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

[14] See, e.g., QM Chapters 5 and 9.

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:

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

This expression has a simple physical sense: the effective spring constant $\kappa_{ef} = m_e\omega_p^2 = ne^2/\varepsilon_0$ describes the Coulomb force that appears when the electron subsystem of a plasma is shifted, as a whole, from its positive-ion subsystem, thus violating the electroneutrality. Indeed, consider such a small shift, Δx, perpendicular to the plane surface of a broad, plane slab filled with plasma. The uncompensated charges, with equal and opposite surface densities $\sigma = \mp en\Delta x$, that appear at the slab surfaces, create inside the it, according to Eq. (2.3), a uniform electric field $E_x = en\Delta x/\varepsilon_0$. This field exerts force $eE = (ne^2/\varepsilon_0)\,\Delta x$ on each positively charged ion. According to the 3rd Newton law, the ions pull each electron back to its equilibrium position with the equal and opposite force $F = -eE = -(ne^2/\varepsilon_0)\,\Delta x$, justifying the above expression for κ_{ef}. Hence it is not surprising that $\varepsilon(\omega)$ described by the first of Eqs. (36) turns into zero at $\omega = \omega_p$: at this resonance frequency, finite free oscillations of charge (and hence of $D = \varepsilon E$) do not require a finite force (and hence E).

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 and hence the wave number (28) are positive and real, and waves propagate without attenuation, following the dispersion relation,

Plasma dispersion relation

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

which is shown in Fig. 6. (As we will see later in this chapter, many wave transmission systems obey such dispersion law as well.)

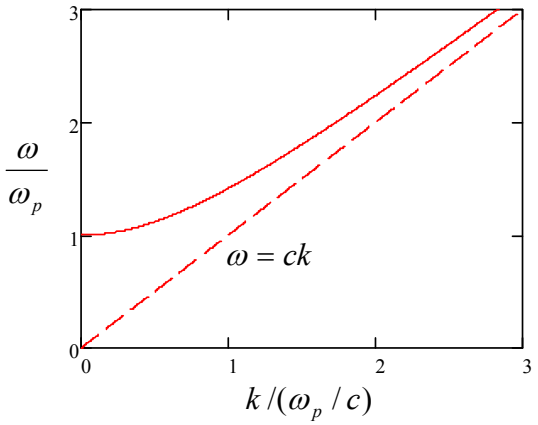


Fig. 7.6. Plasma dispersion law (solid line) in comparison with the linear dispersion of the free space (dashed line).

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

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

According to Eq. (29), this means that the electromagnetic field exponentially decreases with distance:

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

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 **S** = **E**×**H** of a monochromatic electromagnetic wave in an *arbitrary* dispersive (but still linear!) medium. First, let us spell out fields' time dependence:

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

Now, a straightforward calculation yields[15]

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

Let us apply this important general formula to our simple model of plasma at $\omega < \omega_p$. In this case $\mu(\omega) = \mu_0$, i.e. is positive and real, while $\varepsilon(\omega)$ is real and negative, so that $1/Z(\omega) = [\varepsilon(\omega)/\mu(\omega)]^{1/2}$ is purely imaginary, and the average Poynting vector (42) vanishes. This means that energy, on the average, does not flow along axis z – as it would if it was absorbed in plasma. As we will see in the next section, waves with $\omega < \omega_p$ are rather *reflected* from plasma's boundary, without energy loss. Note that in the limit $\omega << \omega_p$, Eq. (39) yields

$$\delta \to \frac{c}{\omega_p} = \left(\frac{c^2\varepsilon_0 m_e}{ne^2}\right)^{1/2} = \left(\frac{m_e}{\mu_0 ne^2}\right)^{1/2}. \tag{7.43}$$

But this is just a particular case (for $q = e$ and $\mu = \mu_0$) of the expression (6.38) that we have derived for the depth of magnetic field penetration into a lossless (collision-free) conductor in the quasistatic approximation. We see again that, as was already discussed in Sec. 6.7, that approximation (in which we neglect the displacement currents) gives an adequate description of the time-dependent phenomena at $\omega << \omega_p$, i.e. at $\delta << c/\omega = 1/k = \lambda/2\pi$.

There are two most important examples of plasmas. For the Earth's ionosphere, i.e. the upper part of the atmosphere that is almost completely ionized by the UV and X-ray components of Sun's radiation, the maximum value of n, reached at about 300 km over the Earth surface, is between 10^{10} and 10^{12} m^{-3} (depending on the time of the day and Sun's activity), so that that the maximum plasma frequency (37) is between 1 and 10 MHz. This is much higher than the particle's reciprocal collision time τ^{-1}, so that Eq. (36) gives a very good description of plasma's electric polarization. The effect of reflection of 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 frequencies of the order of 10 MHz.

[15] 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 (6.103), one can also prove the following interesting expression for the average electromagnetic energy density in an arbitrary dispersive (but linear and isotropic) medium:

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

Such waves may propagate in the flat channel formed by the Earth surface and the ionosphere, reflected repeatedly by these "walls". Unfortunately, due to the random variations of Sun's activity, and hence ω_p, such natural communication channel is not too reliable, and in our age of fiber optics cables its practical importance is diminishing.

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} = 10^{29} m^{-3}, so that Eq. (37) yields $\omega_p \sim 10^{16}$ s^{-1}. Note that this value of ω_p is somewhat higher than mid-optical frequencies ($\omega \sim 3\times10^{15}$ s^{-1}). This explains why planar, even, clean metallic surfaces, such as aluminum and silver films used in mirrors, are so shiny: at these frequencies the permittivity is almost exactly real and negative, leading to light reflection, with very little absorption. However, the considered model, which neglects electron scattering, becomes inadequate at lower frequencies, $\omega\tau \sim 1$.

A phenomenological way of extending the model by account of scattering is to take, in Eq. (33), the lowest eigenfrequency ω_j to be equal zero (to describe free electrons), while keeping the damping coefficient δ_0 of this mode finite, to account for their energy loss due to scattering. Then Eq. (33) is reduced to

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

where response $\varepsilon_{\text{opt}}(\omega)$ at high (in practice, optical) frequencies is still given by Eq. (33), but now with $j \neq 0$.

Result (44) allows for a simple interpretation. To show that, let us incorporate into our calculations the Ohmic conduction, 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 Maxwell equation, $\nabla\times\mathbf{H}_\omega = \mathbf{j}_\omega - i\omega\mathbf{D}_\omega = \mathbf{j}_\omega - i\omega\varepsilon(\omega)\mathbf{E}_\omega$, we get

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

This relation shows that for a sinusoidal process, the addition of the Ohmic current density $\mathbf{j}_\omega$ to the displacement current density is equivalent to addition of $\sigma(\omega)$ to $-i\omega\varepsilon(\omega)$, i.e. to the following change of the ac electric permittivity: [16]

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

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

Generalized Drude formula

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

where the dc conductivity $\sigma(0)$ is described by the Drude formula (4.13), and the phenomenologically introduced coefficient δ_0 is associated with $1/2\tau$. Relation (47), which is frequently called the

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

generalized (or "ac", or "rf") *Drude formula*,[17] gives a very reasonable (semi-quantitative) description of the ac conductivity of many metals almost all the way up to optical frequencies.

7.3. Kramers-Kronig relations

The results for the simple model of dispersion, discussed in the last section, imply that the frequency dependences of the real (ε') and imaginary (ε'') parts of the permittivity 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. R. de L. Kronig in 1926 and H. A. Kramers in 1927 independently showed that this is not an occasional coincidence pertinent only to the Lorentz oscillator model. Moreover, the full knowledge of function $\varepsilon'(\omega)$ allows one to *calculate* function $\varepsilon''(\omega)$, and vice versa. The reason is that both these functions are always related to a single real function $G(\theta)$ by Eqs. (28).

To derive the Kramers-Kronig relations, let us consider Eq. (27) on the complex frequency plane, $\omega \to \omega' + i\omega''$:

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

For all stable physical systems, $G(\theta)$ has to be finite for all important values of the integration variable ($\theta > 0$), and tend to zero at $\theta \to 0$ and $\theta \to \infty$. Because of that, and thanks to factor $e^{-\omega''\theta}$, the expression under the integral tends to zero at $|\omega| \to \infty$ in all upper half-plane ($\omega'' \geq 0$). As a result, we may claim that the complex-variable function $f(\omega)$ is analytical in that half-plane, and allows us to apply to it the *Cauchy integral* formula[18]

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

with the integration contour of the form shown in Fig. 7, with radius R of the larger semicircle tending to infinity, and radius r that of the smaller semicircle (about the singular point $\Omega = \omega$) tending to zero.

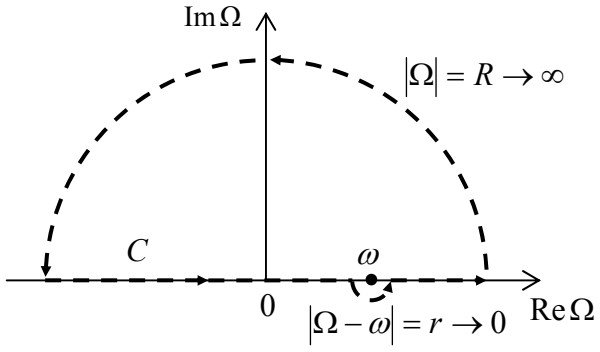


Fig. 7.7. Integration path C used in the Cauchy integral formula to derive the Kramers-Kronig dispersion relations.

[17] 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.

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

Due to the exponential decay of $|f(\Omega)|$ at $|\Omega| \to \infty$, the contribution to the integral from the larger semicircle vanishes,[19] while the contribution from the small semicircle, where $\Omega = \omega + r\exp\{i\varphi\}$, with $-\pi \le \varphi \le 0$, is

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

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

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

Such an integral, excluding a symmetric infinitesimal vicinity of the pole singularity, is called the *principal value* of the (formally, diverging) integral from -∞ to +∞, and is denoted by letter P before it.[20] Using this notation, subtracting $f(\omega)/2$ from both parts of Eq. (48), 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 both parts of Eq. (52) to be equal separately, we get the famous *Kramers-Kronig dispersion relations*

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

Kramers-Kronig dispersion relations

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

$$\varepsilon'(\omega) = \varepsilon_0 + \frac{2}{\pi}\mathrm{P}\int_{0}^{+\infty}\varepsilon''(\Omega)\frac{\Omega d\Omega}{\Omega^2-\omega^2}, \quad \varepsilon''(\omega) = -\frac{2\omega}{\pi}\mathrm{P}\int_{0}^{+\infty}[\varepsilon'(\Omega)-\varepsilon_0]\frac{d\Omega}{\Omega^2-\omega^2}, \quad (7.54)$$

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 in a limited frequency range. For example, if a medium has a sharp absorption peak at some frequency ω_j, we may approximate it as

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

and the first of Eqs. (54) immediately gives

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

[20] I am typesetting this symbol in a Roman font, to exclude any possibility of its confusion with media's polarization.

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

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 (Fig. 5) is by no means occasional or model-specific.

Let me emphasize again that the general, and hence very powerful Kramers-Kronig relations hinge on the causal, linear relation (21) between polarization $P(t)$ with the electric field $E(t')$. 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 applied field $f(t)$, whatever the nature of this response and physics behind it, we may be confident that there is a causal relation between the variables r and f, so that the 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 relation between the Fourier amplitudes of two variables does *not* necessarily imply the causal relationship between them.[21]

7.4. 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 an interface between two uniform, linear, isotropic media.

If the interface is an ideal mirror, the description of reflection is very simple. Indeed, let us assume that one of the two media (say, located at $z > 0$, see Fig. 8) cannot sustain any electric field at all:

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

This condition is evidently incompatible with the single traveling wave (5). However, this solution may be readily corrected 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 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 in Eq. (59) may be interpreted as the *total reflection* of the *incident wave* described by its first term, in this case with the change of electric field's sign. By the way, since vector **n** of the reflected wave is

[21] 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 from the fact that $E(t)$ is *not* a causal function of $P(t)$, but even mathematically. Indeed, the Green's function describing a causal relationship has to tend to zero at small time delays $\theta \equiv t - t'$, so that 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.

opposite to that incident one (see arrows in Fig. 1), Eq. (6) shows that the magnetic field of the wave does not change its sign at the reflection:

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

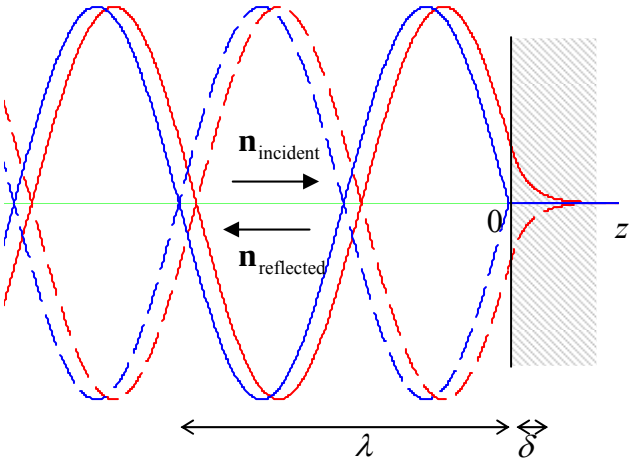


Fig. 7.8. Spatial dependence of electric field at the reflection of a sinusoidal wave from a perfect conductor: the real pattern (red lines) and the crude, ideal-mirror approximation (blue lines). Dashed lines show the patterns after a half-period time delay ($\omega\Delta t = \pi$).

Blue lines in Fig. 8 show the resulting pattern (59) for the simplest, sinusoidal waveform

Wave's total reflection

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

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

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

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

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

As the 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 provides local oscillations of energy, transferring it periodically between the concentrations of the electric and magnetic fields, separated by distance $\Delta z = \pi/2k = \lambda/4$.

For the case of the sinusoidal waves, the reflection effects may be readily explored even for the more general case of dispersive and/or lossy media in which $\varepsilon(\omega)$ and $\mu(\omega)$, and hence the wave vector $k(\omega)$ and wave impedance $Z(\omega)$, defined by Eqs. (28), are certain complex functions of frequency. The "only" new factors we have to account for is that in this case the reflection may not be full, and that

inside the second media we have to use the traveling-wave solution as well. Both these factors may be taken care of by looking for the solution of our boundary problem in the form

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

Wave's partial reflection

and hence, according to Eq. (6),

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

(Indices + and – correspond to, respectively, the media at $z > 0$ and $z < 0$.) Please note the following important features of these relations:

(i) Due to the problem linearity, we could (and did :-) take the complex amplitudes of the reflected and transmitted wave proportional to that (E_ω) of the incident wave, describing them by the dimensionless coefficients R and T. The total reflection from an ideal mirror, that was discussed above, corresponds to the particular case $R = -1$ and $T = 0$.

(ii) Since the incident wave, that we are considering, 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) - in our current problem. However, we would need such a term if the medium at $z > 0$ was non-uniform (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.

(iii) Solution (64)-(65) is sufficient even for the description of the cases when waves cannot propagate at $z \ge 0$, 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 wave number k_+ - see Eq. (29).

In order to find coefficients R and T, we need to use boundary conditions at $z = 0$. Since the reflection does not change the transverse character of the partial waves, at the normal incidence both vectors **E** and **H** remain tangential to the interface plane (in our notation, $z = 0$). Reviewing the arguments that has led us, in statics, to boundary conditions (3.47) and (5.118) for these components, we see that they remain valid for the time-dependent situation as well,[22] so that for our current case of purely transverse waves we can write:

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

Plugging Eqs. (64)-(65) into these conditions, we get

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

[22] For example, the first of conditions (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. In the Stokes theorem, the first term gives $\Delta E_\parallel l$, which the contribution of the second term is proportional to product dl and vanishes as $d/l \to 0$. The proof of the second boundary condition is similar – as was already discussed in Sec. 6.2.

Solving this simple system of equations, we get[23]

Reflection and transmission at a 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 may think, because they are applicable for virtually any 1D waves - electromagnetic or not, if only the impedance Z is defined in a proper way.[24] 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 coefficients R and T may have imaginary parts as well. This fact has most important consequences at $z < 0$ where the reflected wave, proportional to R, interferes with the incident wave. Indeed, plugging $R = |R|e^{i\varphi}$ (where $\varphi \equiv \arg R$ is a real phase shift) into the expression in parentheses in the first of Eqs. (64), we may rewrite it as

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

This means that the field may be presented as a sum of a traveling wave and a standing wave, with amplitude proportional to $|R|$, shifted by distance δ_- toward the interface, relatively to the ideal-mirror pattern (61b). This effect is frequently used for the experimental measurements of an unknown impedance Z_+ of some medium, provided than Z_- is known (e.g., for the free space, $Z_- = Z_0$). For that, a small antenna (the *probe*), not disturbing the field distribution too much, is placed into the wave field, and the amplitude of the ac voltage induced in it by the wave in the probe is measured by some detector (e.g., a semiconductor diode with a quadratic I-V curve), as a function of z (Fig. 9). From this measurement, it is straightforward to find both $|R|$ and δ_-, and hence restore complex R, and then use Eq. (68) to calculate both modulus and argument of Z_+.[25]

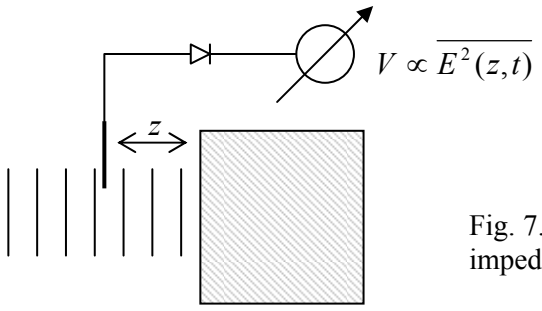


Fig. 7.9. Measurement of the complex impedance of a medium (schematically).

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

[23] Please note that only the media impedances (rather than wave velocities) are important for the 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.

[24] See, e.g., the discussion of elastic waves of mechanical deformations in CM Secs. 5.3, 5.4, 7.7, and 7.8.

[25] Before the advent of computers, specially lined paper (called the *Smith chart*) was commercially available for performing this recalculation graphically; it is occasionally used even nowadays for result presentation.

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

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

The former expression is more convenient in the case $\omega > \omega_p$, when the wave vector k_+ and the wave impedance Z_+ of plasma are real, so that a part of the incident wave propagates into the plasma. Plugging this expression into the latter of Eqs. (68), we see that the transmission coefficient is real:

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

Note that according to this formula, somewhat counter-intuitively, $T > 1$ for any frequency (above ω_p). How can the transmitted wave be more intensive than the incident one that has induced it? For a better understanding of this result, let us compare the powers (rather than amplitudes) of these two waves, i.e. their average Poynting vectors (42):

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

It is easy to see that the ratio of these two values[26] is always below 1 (and tends to zero at $\omega \to \omega_p$), so that only a fraction of the incident wave power may be transferred. Hence the result $T > 1$ may be interpreted as follows: the interface between two media also works as an *impedance transformer*: though it can never transfer 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 transfer.

Now let us proceed to case $\omega < \omega_p$, when the waves cannot propagate in the plasma. In this case, the latter of expressions (70) is more convenient, because it immediately shows that Z_+ is purely 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 that the reflection is total, i.e. no incident power (on the average) is transferred into the plasma – as was already discussed in Sec. 2. However, the complex R has a finite argument,

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

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 that is described by Eq. (39). Calculating, from coefficient T, the exact coefficient before this exponent, it is straightforward to verify that the electric and magnetic fields are indeed

[26] This ratio is sometimes also called the transmission coefficient, but in order to avoid its confusion with T, it is better to call it the *power transmission coefficient*.

continuous at the interface, forming the pattern shown by red lines in Fig. 8. This penetration may be experimentally observed, for example, by bringing close to the interface the surface of another material transparent as frequency ω. Even without solving this problem exactly, it is evident that if the distance between these two interfaces becomes comparable to δ, a part of the exponential "tail" of the field is picked up by the second material, and induces a propagating wave. This is an electromagnetic analog of the quantum-mechanical tunneling through a potential barrier.[27]

Note that at $\omega << \omega_p$, both δ_- and δ are reduced to the same frequency-independent value,

$$\delta, \delta_- \to \frac{c}{\omega_p} = \left(\frac{c^2 \varepsilon_0 m_e}{n e^2} \right)^{1/2} = \left(\frac{m_e}{\mu_0 n e^2} \right)^{1/2}, \tag{7.75}$$

which is just the field penetration depth δ (6.38) 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 $\lambda_0 = 2\pi c/\omega >> 2\pi c/\omega_p = 2\pi\delta$.

(ii) Now let us consider electromagnetic wave reflection from a nonmagnetic 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.[28] According to Eq. (46), in this case we can take

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

With this substitution, Eqs. (68) immediately give us all the results of interest. In particular, they show that now R is complex, and hence some fraction F of the incident wave is absorbed by the conductor. Using Eq. (42), we may calculate the fraction to be

$$F \equiv \frac{\overline{S_+\big|_{z=+0}}}{\overline{S_{\text{incident}}}} = |T|^2 \, \text{Re} \frac{Z_0}{Z_+}. \tag{7.77}$$

(Since power flow S_+ into the conductor depends on z, tending to zero at distances $z \sim \delta$, it is important to calculate it directly at the interface to account for the absorption in the whole volume of the conductor.) Restricting ourselves, for the sake of simplicity, to the most important quasistatic limit, i.e. to $Z_+ = (\mu_0\omega/i\sigma)^{1/2}$, and using Eq. (6.27) to express the impedance via the skin depth, $Z_+ = \pi(2/i)^{1/2}(\delta_s/\lambda_0)Z_0$, we see that $|Z_+| << Z_0$, so that, according to Eq. (68), $T \approx 2Z_+/Z_0$ and

Wave's absorption in conductor's surface

$$F \approx \frac{4|Z_+|^2}{Z_0^2} \text{Re} \frac{Z_0}{Z_+} = 2\frac{\delta_s}{\lambda_0} << 1. \tag{7.78}$$

Thus the absorbed power scales as the ratio of the skin depth to the free-space wavelength. This important result is widely used for the semi-qualitative evaluation of power losses in metallic waveguides and resonators, and immediately shows that in order to keep the losses low, the characteristic size of such systems (that gives a scale of the free-space wavelengths λ_0, at which they are

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

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

used) should be much larger than δ_s. A more detailed theory of these structures will be discussed later in this chapter.

7.5. Refraction

Now let us consider the effects arising at the plane interface if the wave 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 $\varepsilon_\pm$ and $\mu_\pm$.

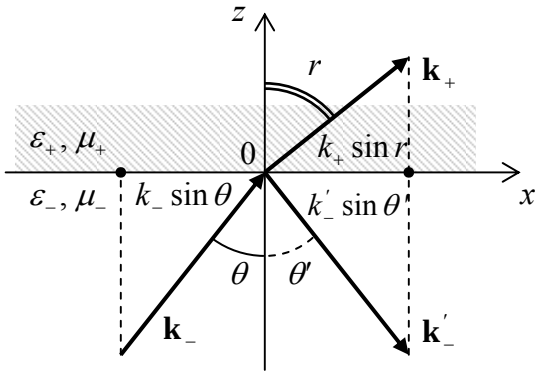


Fig. 7.10. Plane wave reflection, transmission, and refraction at a plane interface. The plane of drawing is selected to contain all three wave vectors $\mathbf{k}_+$, $\mathbf{k}_-$, and $\mathbf{k}_-'$.

In contrast with the case of normal incidence, here the wave vectors $\mathbf{k}_-$, $\mathbf{k}_-'$, and $\mathbf{k}_+$ of the three component (incident, reflected, and transmitted) waves may have different directions. Hence now we have to start our analysis with writing a general expression for a single plane, monochromatic wave for the case when its wave vector $\mathbf{k}$ has all 3 Cartesian components, rather than one. An evident generalization of Eq. (11) to this case is

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

This relation enables a ready analysis of "kinematic" relations that are independent of the media impedances. Indeed, it is sufficient to notice that in order to satisfy *any* linear, homogeneous boundary conditions at the interface ($z = 0$), all waves have the same temporal and spatial dependence on this plane. Hence if we select plane xz so that 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 vectors lie in the same plane - that is selected as the plane of drawing of Fig. 10. Moreover, due to 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 have the well-known laws of reflection

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

Reflection angle

and refraction:[29]

[29] This relation is traditionally called the *Snell law*, after a 17th century's author W. Snellius, though it has been traced back to a circa 984 manuscript by Abu Saad al-Ala ibn Sahl.

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 presented 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 so that of the free space (at wave's frequency):

Index of refraction

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

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 into air, there is always a certain *critical* value θ_c of the incidence angle,

Critical angle

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

at which angle r reaches $\pi/2$. At a larger θ, i.e. within the range $\theta_c < \theta < \pi/2$, the boundary conditions cannot be satisfied with a refracted wave with a real wave vector, so that the wave experiences the so-called *total internal reflection*. This effect is very important for practice, because it shows that dielectric surfaces may be used as mirrors, in particular in optical fibers - to be discussed in more detail in Sec. 8 below. This is very fortunate for all the telecommunication technology, because the light 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 relatively high losses $F \sim 1\%$ 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: just as it does at the normal incidence ($\theta = 0$), it penetrates into the less dense media by a distance of the order of λ_0, exponentially decaying inside it. At $\theta \neq 0$ the penetrating field still changes sinusoidally, with wave number (80), along the interface. Such a field, exponentially dropping in one direction but still propagating as a wave in another direction, is frequently called the *evanescent wave*.

One more remark: just as at the normal incidence, the field 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 called the *Goos-Hänchen effect*. In geometric optics, this effect leads to an image shift (relative to that its position in a perfect mirror) with components both normal and parallel to the interface.

Now let us carry out an analysis of the "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. 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 we have repeatedly discussed,

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

we also need relations for the normal components. As it follows from the homogeneous macroscopic Maxwell equations (6.94b), they are also the same as in statics (D_n = const, B_n = const), for our reference frame choice (Fig. 10) giving

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

The expressions of these components via 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 *perpendicular* to the plane of incidence (Fig. 11a), i.e. *parallel* to the interface plane, the reflected and refracted waves are similarly polarized.

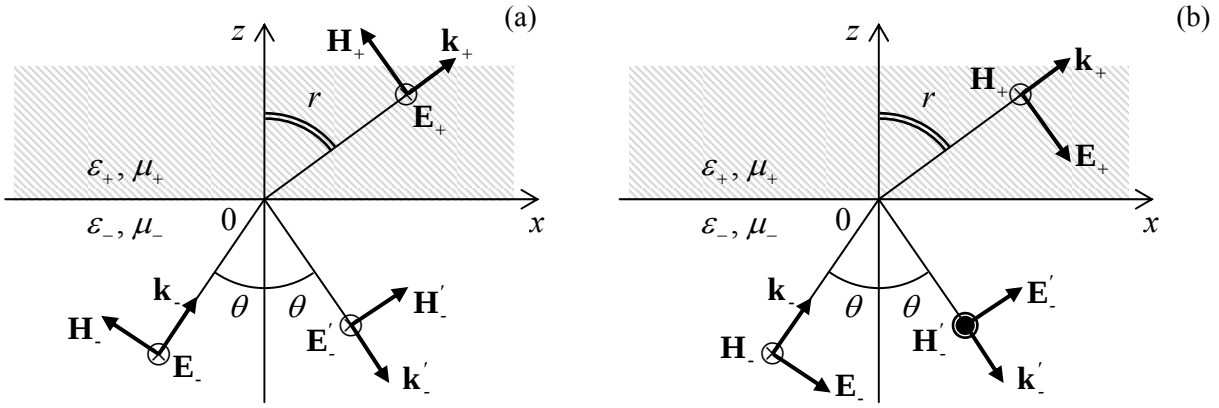


Fig. 7.11. Reflection and refraction at two different linear polarizations of the incident wave.

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

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

but also normal components (Fig. 11a):

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

Plugging these expressions into the boundary conditions expressed by Eqs. (86) (in this case, for 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 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 system (90):[30]

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

If we want to express the coefficients via the angle of incidence alone, we should use the Snell law (82) to eliminate angle r, getting

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

However, my strong preference is to use the kinematic relation (82) and dynamic relations (91a) separately, because Eq. (91b) obscures the very important physical fact that and the ratio of $k_\pm$, i.e. of the wave velocities of the two media, is only involved in the Snell law (79), while the dynamic relations essentially include only the ratio of 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 two media have not only tangential components,

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

but also normal components (Fig. 11b):

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

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):

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

or, alternatively, using the Snell law:

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

[30] Note that we may calculate the reflection and transmission coefficients R' and T' for the wave traveling in the opposite direction just by making 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 $Z_\pm$.

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*.[31] Most textbooks are quick to point out that there is a major difference between these cases: while for the electric field polarization within the plane of incidence (Fig. 11b), the reflected wave amplitude (proportional to coefficient R) turns to zero at a special value of θ (the so-called *Brewster angle*):[32]

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

while there is no such angle in the opposite case (Fig. 11a).[33] However, 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). They 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. $\varepsilon_+ = \varepsilon_-$), 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 nonvanishing amplitude for any θ.

Such account of both media parameters on an equal footing is especially necessary to describe the so-called *negative refraction* effects.[34] As was shown in Sec. 2, in a medium with electric-field-driven resonances, function $\varepsilon(\omega)$ may be almost real and negative, at least within limited frequency intervals – see, in particular, Eq. (34) and Fig. 5. As have already been discussed, if, at these frequencies, function $\mu(\omega)$ is real and positive, then $k^2(\omega) = \omega^2\varepsilon(\omega)\mu(\omega) < 0$, and k may be presented as i/δ with real δ, meaning the exponential field decay into the medium. However, let consider the case when both $\varepsilon(\omega) < 0$ *and* $\mu(\omega) < 0$ at a certain frequency. (This is evidently possible in a medium with both **E**-driven and **H**-driven resonances, at proper relations between their eigenfrequencies.) Since in this case $k^2(\omega) = \omega^2\varepsilon(\omega)\mu(\omega) > 0$, the wave vector is real, so that Eq. (79) describes a traveling wave, and one could think that there is nothing new in this case. Not quite so!

---

[31] After A.-J. Fresnel (1788-1827), one of the pioneers of the wave optics, who is credited, among many other contributions (see in particular Ch. 8), for the concept of light as a purely transverse wave.

[32] A very simple interpretation of Eq. (93) is based on the fact that, together with the Snell law (82), it gives $r + \theta = \pi/2$. As a result, vector $\mathbf{E}_+$ is parallel to vector $\mathbf{k}_-'$, and hence oscillating dipoles of medium at $z > 0$ do not have the component which could induce the transverse electric field $\mathbf{E}_-'$ of the reflected wave.

[33] 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 practical application of the effect is a partial reduction of undesirable glare from wet surfaces (for the water/air interface, $n_+/n_- \approx 1.33$, giving $\theta_B \approx 50°$) by making car light covers and sunglasses of vertically-polarizing materials.

[34] 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 have only recently become a subject of intensive scientific research and engineering development.

First of all, for a sinusoidal, plane wave (79), 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 vector **H**. Namely, if both ε and μ are positive, these equations are satisfied with mutually orthogonal vectors **E**, **H**, and **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 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 Eq. (6.97), that does not involve media parameters, this means that for a plane wave in a left-hand material, the Poynting vector **S** = **E**×**H**, i.e. of the energy flow, is directed *opposite* to the wave vector **k**.

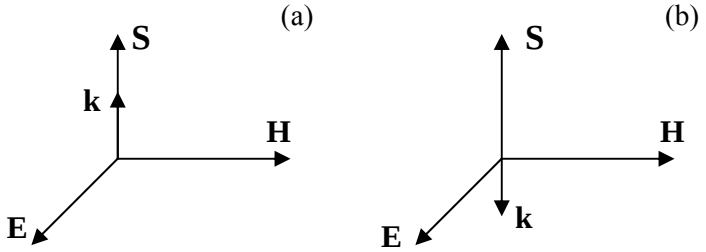


Fig. 7.12. Directions of main vectors of a plane wave inside a medium with (a) positive and (b) negative ε and μ.

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

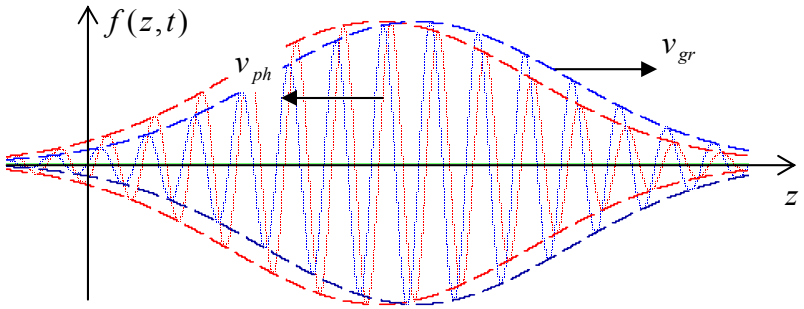


Fig. 7.13. Example of a wave packet moving along axis z with a negative phase velocity, but positive group velocity. Blue lines show a packet snapshot a short time interval after the first snapshot (red lines).

Maybe 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. Consider the example shown in Fig. 14a. In the incident wave, coming from the usual material, the directions of vectors **k** and **S** coincide, and so they are in the reflected wave characterized by vectors **k**' and S'. This means that the electric and magnetic fields in the interface plane (z = 0) are, at our choice of coordinates, proportional to $\exp\{ik_x x\}$, with positive component $k_x = k\cos\theta$. In order to satisfy any linear boundary conditions, the refracted wave, going into the left-handed material, should 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 vector $\mathbf{S}$ that, in turn, should be directed out of the interface, because it presents the power flow from the interface into the material bulk. These conditions cannot be reconciled by the refracted wave propagating along the usual Snell-law direction (shown by the dashed line in Fig. 13a), but are all satisfied at refraction in the direction given by Snell's angle with negative sign. (Hence the term "negative refraction").[35]

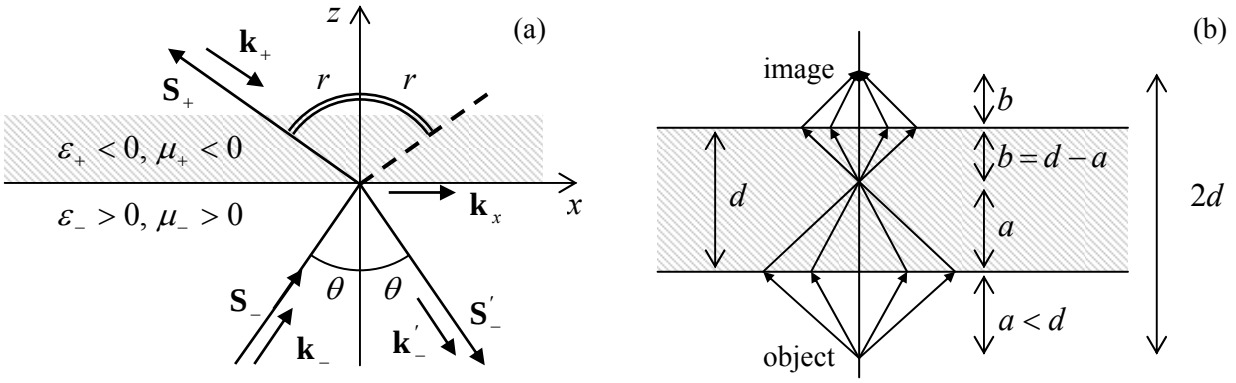


Fig. 7.14. Negative refraction: (a) waves at the interface between media with positive and negative values of $\varepsilon\mu$, and (b) the hypothetical *perfect lens*: a parallel plate made of a material with $\varepsilon = -\varepsilon_0$ and $\mu = -\mu_0$.

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 $\varepsilon = -\varepsilon_0$, $\mu = -\mu_0$ (Fig. 14b), placed in free space. For such a material, the refraction angle $r = -\theta$, so that the rays from a point source, located at a distance $a < d$ from the slab, propagate as shown in that figure, i.e. all meet again at distance a inside the plate, 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 plate at distance $2a + 2b = 2a + 2(d - a) = 2d$ from the object. Superficially, this looks like the 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 in order for all 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 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 found 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,

[35] Inspired by this fact, in some publications 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 wave vector $\mathbf{k}$, is the matter of convention.

Fig. 15a shows the metamaterial that was used for the first demonstration of negative refractivity in the microwave region, i.e. a few-GHz frequencies – see Fig. 15b. It combines straight strips of a metallic film, working as lumped resonators with a large electric dipole moment (hence strongly coupled to wave's electric field **E**), and several almost-closed film loops (so-called *split rings*), working as lumped resonators with large magnetic dipole moments, coupled to field **H**. By designing the resonance frequencies close to each other, the negative refractivity may be achieved – see the black line in Fig. 15b, which shows experimental data. Recently, the negative refractivity was demonstrated in the optical range, albeit at relatively large absorption that spoils all potentially useful features of the left-handed materials.

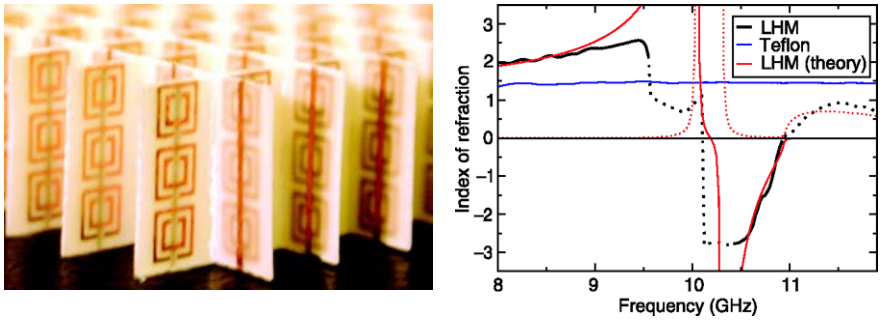


Fig. 7.15. The first artificial left-hand material with experimentally demonstrated negative refraction in a microwave region. Adapted from R. Shelby *et al.*, *Science* **292**, 77 (2001). © AAAS.

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 engineered distributions $\varepsilon(\mathbf{r}, \omega)$ and $\mu(\mathbf{r}, \omega)$, which may provide electromagnetic wave propagation along the desired paths, e.g. around a certain region of space (Fig. 16), making it virtually invisible for an external observer - so far, within a limited frequency range, and a certain wave polarization only. Due to these restrictions, the practical value of this work on such *invisibility cloaks* in not yet clear (at least to this author); but so much attention is focused on this issue[36] that the situation should become much more clear in just a few years.

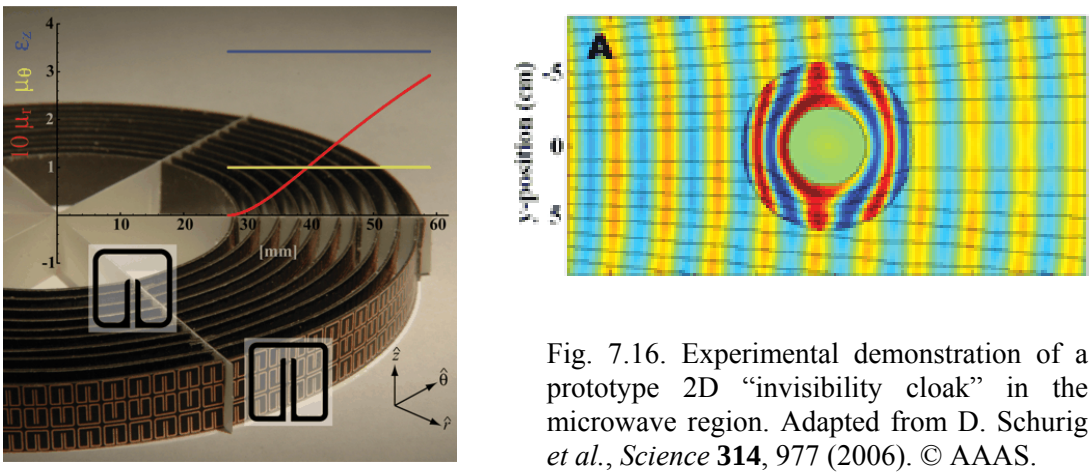


Fig. 7.16. Experimental demonstration of a prototype 2D "invisibility cloak" in the microwave region. Adapted from D. Schurig *et al.*, *Science* **314**, 977 (2006). © AAAS.

[36] For a recent review, see, e.g., B. Wood, *Comptes Rendus Physique* **10**, 379 (2009).

7.6. Transmission lines: TEM waves

So far, we have analyzed plane the electromagnetic waves with infinite cross-section. The cross-section may be limited, still sustaining wave propagation, using w*ave transmission lines* (also called *waveguides*): cylindrically-shaped structures made of either good conductors or dielectrics. Let us first discuss the first option. In order to keep our analysis (relatively :-) simple, let us assume that:

(i) the structure is a cylinder (not necessarily with a round cross-section, see Fig. 17) filled with a usual (right-handed), uniform dielectric material with negligible losses: $\varepsilon = \varepsilon' > 0$, $\mu = \mu' > 0$, 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 waveguide's cross-section has to be much larger than the skin-depth δ_s of its wall material. The effect of skin-effect losses will be analyzed in Sec. 10 below.)

After such exclusion of attenuation, we may look for a particular solution of the Maxwell equations in the form of a monochromatic wave traveling along the waveguide:

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

with real k_z. Note that this form allows an account for a substantial coordinate dependence of the electric and magnetic field in the plane $\{x,y\}$ of the waveguide's cross-section, as well as for longitudinal components of the fields, so that solution (98) is substantially more complex than the plane waves we have discussed above. We will see in a minute that as a result of this dependence, constant k_z may be very much different from the plane-wave value (13), $k \equiv \omega(\varepsilon\mu)^{1/2}$, in the same material.

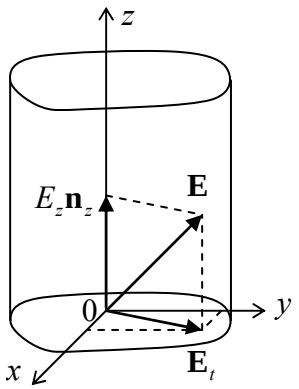


Fig. 7.17. Decomposition of the electric field in a waveguide.

In order to describe these effects explicitly, let us decompose the complex amplitudes of the fields into the longitudinal and transverse components (Fig. 17)[37]

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

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

[37] Note that for the notation simplicity, I am dropping index ω in the complex amplitudes of the field components, and later will drop argument ω in k_z and Z, though they may depend on the wave frequency rather substantially – see below.

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

where ∇_t is the 2D del operator acting in the transverse plane $[x, y]$ only. These equations may look even more bulky than the original Maxwell equations, but actually are much simpler for analysis. Indeed, eliminating the transverse components from these equations (or, even simpler, just plugging Eq. (99) into Eqs. (3) and keeping just their z-components), we may get a pair of self-consistent equations for the longitudinal components of the fields, [38]

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), $k = (\varepsilon\mu)^{1/2}\omega$, and

Wave vector component balance

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

After distributions $E_z(x,y)$ and $H_z(x,y)$ have been found from these equations, they provide right-hand parts for 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*, or *TEM waves*),

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

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

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

Let us start with the simplest, TEM waves with no longitudinal components of either field. For them, the top two equations of 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 mutually perpendicular (just as in the plane wave) at each point of the cross-section, and that the TEM wave impedance $Z \equiv E/\mathrm{H}$ and dispersion law $\omega(k)$, and hence the propagation speed, are the same as in a plane wave in the material filling the waveguide. In particular, if ε and μ are frequency-independent within a certain frequency range, the dispersion law is linear, $\omega = k/(\varepsilon\mu)^{1/2}$, and wave's speed does not depend on its frequency. For practical applications to telecommunications, this is a very important advantage of TEM waves over their TM and TE counterparts – to be discussed below.

Unfortunately, such waves cannot propagate in every waveguide. In order 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 yield

$$\begin{aligned} &\nabla_t\times\mathbf{E}_t = 0, && \nabla_t\times\mathbf{H}_t = 0, \\ &\nabla_t\cdot\mathbf{E}_t = 0, && \nabla_t\cdot\mathbf{H}_t = 0. \end{aligned} \tag{7.103}$$

In the macroscopic approximation of the boundary conditions (i. e., neglecting the screening and skin depths), we have to require that the wave does not penetrate the walls, so that inside them, $\mathbf{E} = \mathbf{H} = 0$. Close to the wall but inside the waveguide, the normal component E_n of the electric field may be

[38] The wave equation presented in the form (101) is called the (in our particular case, 2D) *Helmholtz equation*, after H. von Helmholtz (1821-1894) - the mentor of H. Hertz and M. Planck, among many others.

different from zero, because surface charges may sustain its jump (see Sec. 2.1). Similarly, the tangential component H_τ of the magnetic field may have a finite jump at the surface due to skin currents. However, the tangential component of the electric field and the normal component of magnetic field cannot experience such jump, and in order to have them vanishing inside the walls they have to equal zero near the walls inside the waveguide as well:

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

But the left columns of Eqs. (103) and (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 should be replaced with $\varepsilon(\omega)$. 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 $\mu = \mu(\omega)$. The only difference is that in our current case the magnetic fields should not penetrate inside the conductors.

Now we immediately see that in waveguides with a singly-connected wall topology (see, e.g., the particular example shown in Fig. 17), TEM waves are impossible, because there is no way to create a finite electrostatic field inside a conductor with such cross-section. Fortunately, such fields (and hence TEM waves) are possible in structures with cross-sections consisting of two or more disconnected (dc-insulated) parts – see, e.g., Fig. 18. (Such structures are more frequently called the transmission lines rather than waveguides, the last term being mostly reserved for the lines with singly-connected cross-sections of the walls.)

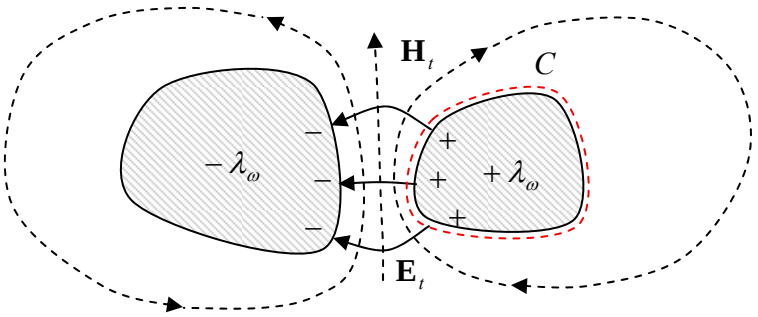


Fig. 7.18. Example of the cross-section of a transmission line that may support the TEM wave propagation.

Now we can readily derive some "global" relations for each conductor, independent on the exact shape of its cross-section. Indeed, consider contour C drawn very close to the conductor's surface (see, e.g., the red dashed line in Fig. 18). First, we can consider it as a cross-section of a cylindrical Gaussian volume of certain length $dz << \lambda \equiv 2\pi/k$. Using the generalized Gauss law (3.29), get

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

where λ_ω (not to be confused with wavelength λ!) is the linear density of electric charge of the conductor. Second, the same contour C may be used in the generalized Ampère law (5.131) 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 that Eqs. (105)-(106) give the following simple relation between the "global" characteristics of the conductor:

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

This relation may be also obtained by a different means; let me describe it, because it has an independent value. Let us consider a small segment $dz << \lambda = 2\pi/k$ of the conductor (limited by the red dashed line in Fig. 18) 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, so that the result just expresses the charge continuity law. However, Eq. (108) is valid for any waveform.)

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. 18). In this case, in order to have the wave well localized in the space near the two conductors, we need a sufficiently fast convergence of its electric field at large distances.[39] 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 – that was repeatedly discussed in Chapter 2. Then Eq. (108) takes the 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. 19 (which shows a 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). Since the electric field is zero inside the conductors (in Fig. 19, on the horizontal parts of the contour), the total e.m.f. equals the difference of voltages V at the end of the segment dz, while the only source 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 it. As a result, canceling dz in both parts of the equation, we get

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

[39] The alternative is to have a virtually plane wave, which propagates along the transmission line conductors, and whose fields are just slightly deformed in their vicinity. Such a wave cannot be "guided" by the conductors, and hardly deserves the name of a "wave *in* the waveguide".

where L_0 is the mutual inductance of the conductors per unit length. The only difference between 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 its penetration into the conductors. (In the dc case, we had the same situation for superconductor electrodes, within their crude, ideal-diamagnetic description.)

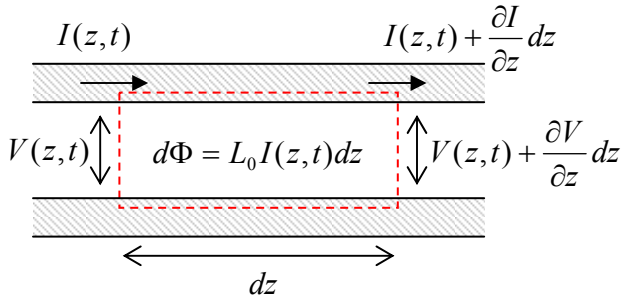


Fig. 7.19. Electric current, magnetic flux, and voltage in a two-conductor transmission line.

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) a 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 the dispersion-free TEM wave propagation. Again, this equation is only valid within the frequency range where the frequency dependence of both ε and μ is negligible. If it 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 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.

Two linear equations (113) are consistent only if

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

L_0C_0 product invariance

Besides the fact we have already known (that the TEM wave speed is the same as that of the plane wave), Eq. (114) gives us a result that I confess I have 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 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 relation stems from the fact that both the electric and magnetic fields may be expressed via the solution of a 2D Laplace equation for system's cross-section.

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

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

Transmission line's TEM Impedance

that is called the *transmission line's impedance*. This parameter has the same dimensionality (in SI units, ohms) 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 cross-section's geometry, while Z does not. In particular, Z_W is the only important parameter of a transmission line for matching with a lumped load circuit (Fig. 20) in the important case when both the cable cross-section's size and the load's linear dimensions are much smaller than the wavelength. (The ability of TEM lines to have such a small cross-section is their another important advantage.) 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 two Kirchhoff' s laws for 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), load's impedance $Z_L(\omega)$ should be real and equal to Z_W - but not necessarily to Z.

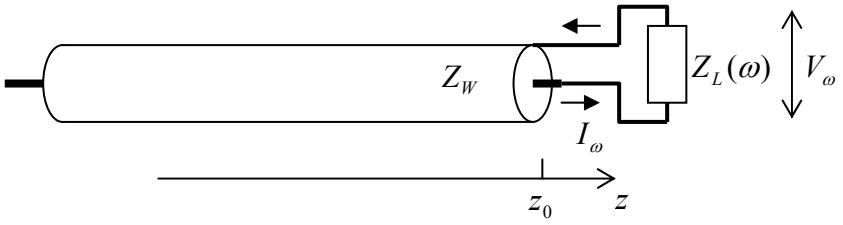


Fig. 7.20. Transmission line impedance matching.

As an example, let us consider one of the simplest (and the most important) transmission lines: the coaxial cable (Fig. 21).[40]

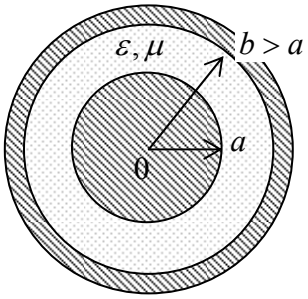


Fig. 7. 21. Cross-section of a coaxial cable with arbitrary (possibly, dispersive) dielectric filling.

For this geometry, we already know expressions for both L_0 and C_0, though they have to be modified for the dielectric constant and the magnetic field non-penetration into the conductors. After that modification,

[40] The coaxial cable was first patented by O. Heaviside in 1880.

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

Coaxial cable's C_0 and L_0

So, the universal relation (114) is indeed valid! For cable's impedance (115), Eqs. (119) yield

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

For standard TV antenna cables (such as RG-6/U, with $b/a \sim 3$, $\varepsilon/\varepsilon_0 \approx 2.2$), $Z_W = 75$ ohms, while for most computer component connections, cables with $Z_W = 50$ ohms (such as RG-58/U) are prescribed by electronic engineering standards. Such cables are broadly used for transfer of electromagnetic waves with frequencies (limited mostly by cable attenuation; see Sec. 10 below) up to 1 GHz over distances of a few km, and up to ~20 GHz on the tabletop scale (a few meters).

Another important example of TEM transmission lines is the set of two parallel wires. In the form of *twisted pairs*,[41] they allow communications, in particular long-range telephone and DSL Internet connections, at frequencies up to a few hundred kHz, as well as relatively short Ethernet and TV cables at frequencies up to ~ 1 GHz, limited mostly by the mutual interference and parasitic radiation effects.

7.7. *H* and *E* waves in metallic waveguides

Let us now return to Eqs. (100) and explore the TE and TM waves - with, respectively, either H_z or E_z different from zero. At the first sight, they may seem more complex. However, equations (101), which determine the distribution of these longitudinal components over the cross-section, are just 2D Helmholtz equations for scalar functions. For simple cross-section geometries may be 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 two equations in the first line of Eqs. (100).[42]

In comparison with the electro- and magnetostatics problems, the only conceptually new feature of Eqs. (101), with appropriate boundary conditions, 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 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 each mode, that follows from the last form of Eq. (102),

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

Universal dispersion relation

[41] The twisting reduces mutual induction ("crosstalk") between the lines, and parasitic radiation at their bends.

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

is functionally the same as that of plane waves in a plasma (see Eq. (38), Fig. 6, and their discussion), with the only differences that c is now replaced with $v = 1/(\varepsilon\mu)^{1/2}$, the speed of plane (or any TEM) waves in the medium filling the waveguide, and ω_p is replaced with the so-called *cutoff frequency*

$$\omega_c \equiv vk_t, \tag{7.123}$$

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 waveguide's cross-section, so that the critical value of the free-space wavelength is of the order of a.) Below the cutoff frequency of each particular mode, it cannot propagate in the waveguide. [43] As a result, modes with the *lowest* values of ω_c present special practical interest, because the choice of the signal frequency ω between two lowest values of cutoff frequency guarantees that the waves propagate in the form of only one mode, with the lowest k_t. Such a choice allows to simplify the excitation of the desired mode by wave generators, and to avoid the parasitic transfer of electromagnetic wave energy to undesirable modes by (unavoidable) small inhomogeneities of the system.

The boundary conditions for the Helmholtz equations (101) depend on the propagating wave type. For TM waves (i.e. E modes, with $H_z = 0$ but $E_z \neq 0$), in the macroscopic approximation the boundary condition $E_\tau = 0$ immediately gives

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

where C is the contour limiting the conducting wall's cross-section. For TE waves (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 system (100), vector-multiplied by $\mathbf{n}_z$. Indeed, for the component perpendicular to the conductor surface the equation gives

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

But the first term in the left-hand part of this equation must be zero on the wall surface, because of the second of Eqs. (103), while according to the first of Eqs. (103), 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 that 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 what does this approach give for a simple but practically important example of a metallic-wall waveguide with a rectangular cross-section. In this case it is natural to use the Cartesian coordinates shown in Fig. 22, so that both Eqs. (101) take the simple form

[43] An interesting recent 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) virtually vanishes. As a result, 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).

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+k_t^2\right)f=0,\quad f=\begin{cases}E_z, & \text{for TM waves,}\\ H_z, & \text{for TE waves.}\end{cases} \tag{7.127}$$

From Chapter 2 we know that the most effective way of solution of 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

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

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

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

that 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,[44] for eigenfunction (128) we get

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

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

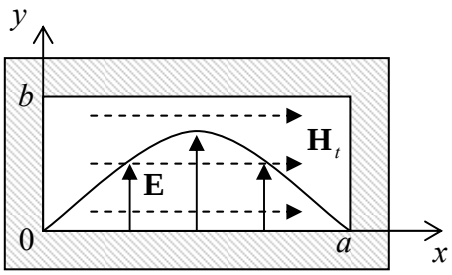


Fig. 7.22. Rectangular waveguide, and the transverse field distribution in the basic mode H_{10} (schematically).

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

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

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

[44] Let me hope that the solution of equations of the type $d^2X/dx^2+k_x^2X=0$ does not present a problem for the reader, due to his or her 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.

where H_l is the longitudinal field amplitude, and n and m are two arbitrary integer numbers, besides that they cannot equal to zero simultaneously. (Otherwise, function $H_z(x,y)$ would be constant, so that, 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 $k_z \neq 0$.) Assuming, 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), is achieved for the so-called H_{10} *mode* with $n = 1$ and $m = 0$, and hence

Basic mode's cutoff

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

(thus confirming our prior estimate of k_t).

Depending on the a/b ratio, the second lowest k_t and cutoff frequency belong to either the H_{11} mode with $n = 1$ and $m = 1$:

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

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

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

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 waveguide WR90 with $a \approx 2.3$ cm ($f_c \equiv \omega_c/2\pi \approx 6.5$ GHz), $b \approx 1.0$ cm.

Now let us have a fast look at alternative TM waves (E modes). For them, we may still should use the general solution (130) with $f = E_z$, but now with boundary condition (124). This gives us 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 provided by the so-called E_{11} *mode* with n =1, $m = 1$, and the eigenvalue is again given by Eq. (134).

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

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

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

This field distribution is (schematically) shown in Fig. 22. Neither of the fields depends on the vertical coordinate – which is very convenient, in particular, for microwave experiments with small samples.

The electric field has only one (vertical) component that vanishes at the side walls and reaches maximum at 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 nonvanishing 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 similar 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 the transmission line with a lumped load – see Fig. 20 and Eq. (118). As was discussed above, such simple description is possible (i.e., does not depend on the exact geometry of the connection) only if both dimensions of 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 $\lambda_{\max} = 2\pi/(k_t)_{\min} = 2\pi/(k_t)_{10} = 2a$.[45]

Now let us consider metallic waveguides with round cross-section (Fig. 23a). In this single-connected geometry, again, the TEM waves are impossible, 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 form

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}+k_t^2\right]f=0, \quad f=\begin{cases}E_z, & \text{for TM waves,}\\ H_z, & \text{for TE waves.}\end{cases} \tag{7.139}$$

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

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

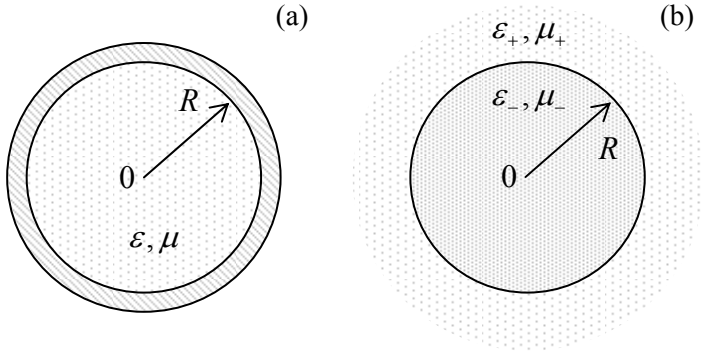


Fig. 7.23. (a) Metallic and (b) dielectric waveguides with circular cross-sections.

But this is exactly the Eq. (2.127) that was studied in the context of electrostatics, just with a replacement of notation: $\gamma \to k_t$. So we already know that in order to have 2π-periodic functions $\mathcal{F}(\varphi)$, and finite values $\mathcal{R}(0)$ (which are necessary for our current case – see Fig. 23a), the general solution is

[45] The reader is encouraged to find a simple interpretation of this equality.

given by Eq. (2.136), i.e. the eigenfunctions may be expressed via integer-order Bessel functions of the first kind:[46]

$$f_{nm} = \text{const} \times J_n(k_{nm}\rho)e^{in\varphi}, \tag{7.141}$$

with eigenvalues k_{nm} of the transverse wave number k_t to be determined from appropriate boundary conditions.

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

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

This means that eigenvalues of Eq. (139) are

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

where ξ'_{nm} is the m^{th} root of function $dJ_n(\xi)/d\xi$. The approximate values of these roots for several lowest n and m may be read out from the plots in Fig. 2.16; their more accurate values are presented in Table 1 below.

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

It shows, in particular, that the lowest of the roots is $\xi'_{11} \approx 1.84$. Thus, a bit counter-intuitively, the basic mode, providing the lowest cutoff frequency $\omega_c = vk_{nm}$, is H_{11} corresponding to $n = 1$ rather than $n = 0$:[47]

$$H_z = H_l J_1\left(\xi'_{11}\frac{\rho}{R}\right)e^{i\varphi}, \tag{7.144}$$

with the transverse wave vector $k_t = k_{11} = \xi'_{11}/R \approx 1.84/R$, and hence the cutoff frequency corresponding to the TEM wavelength $\lambda_{\max} = 2\pi/k_{11} \approx 3.41\,R$. Thus the ratio of $\lambda_{\max}$ to the waveguide diameter $2R$ is

[46] In Chapter 2, it was natural to take the angular dependence in the sin-cos form, which is equivalent to adding a similar term with $n \to -n$ to the right-hand part of Eq. (141). However, since the functions f we are discussing now are already complex, it is easier to do calculations in the exponential form - though it is vital to restore real fields before calculating any of their nonlinear forms, e.g., the wave power.

[47] The lowest root of Eq. (142) with $n = 0$, i.e. ξ'_{00}, equals 0, and would yield $k = 0$ and hence a constant field H_z, which, according to the first of Eqs. (121), would give vanishing electric field.

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 the real part of Eq. (144).)

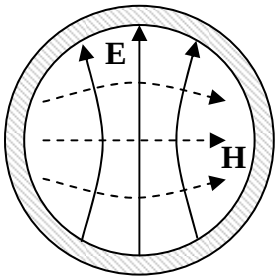


Fig. 7.24. Transverse field components in the basic H_{11} mode of a metallic, circular waveguide (schematically).

One can see that the field structure is actually very similar to that of the basic mode in the rectangular waveguide, shown in Fig. 22, despite the different nomenclature (due to the different type of used coordinates). However, note the arbitrary argument of complex constant H_l in Eq. (144), indicating that in circular waveguides the transverse field polarization is arbitrary. For some practical applications, the degeneracy of these "quasi-linearly-polarized" waves creates problems; they may be avoided by using waves with circular polarization.[48]

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 basic mode, and only then comes the first mode with no angular dependence of the any field, H_{01}, with $k_t = k_{01} = \xi'_{01}/R \approx 3.83/R$.[49]

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

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

where ξ_{nm} is the m-th root of function $J_n(\xi)$ – see Table 2.1. The table shows that the lowest k_t equals to $\xi_{01}/R \approx 2.405/R$. Hence the corresponding mode (E_{01}), with

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

has the second lowest cutoff frequency, approximately 30% higher than that of the basic 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 $\omega >> \omega_c$. This 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

[48] Actually, Eq. (144) does describe a circularly polarized wave, while the real and imaginary parts of this expression describing two mutually perpendicular quasi-linearly-polarized waves.

[49] Electric field lines in the H_{01} mode (as well as all higher H_{0m} modes) are directed straight from the axis to the walls, reminding those of TEM waves in the coaxial cable. Due to this property, these modes provide, at $\omega >> \omega_c$, much lower power losses (see Sec. 10 below) than the fundamental H_{11} mode, and are sometimes used in practice, despite all inconveniences of working in the multimode frequency range.

that 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.[50] However, it is frequently more convenient to discuss transverse wave vectors $\mathbf{k}_t$ of arbitrary direction, i.e. with arbitrary sign 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 4 lower than was calculated above. This means that the number of modes of *both* types is

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

It may be convincingly argued that this *mode counting rule* is valid for waveguides with cross-section of any shape, and any boundary conditions on the walls, provided that $N >> 1$.

7.8. Dielectric waveguides and optical fibers

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

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

(In most cases the difference is achieved due to that in the dielectric constant, $\varepsilon_- < \varepsilon_+$, while magnetically both materials are almost passive: $\mu_- \approx \mu_+ \approx \mu_0$, and I will assume that in my narrative.) The idea of the waveguide operation may be readily understood in the case when wavelength λ is much smaller than the characteristic size R of core's cross-section. If this "geometric optics" limit, at the distances of the order of λ from the core-to-cladding interface, which determines the wave reflection, we can consider the interface as a plane. As we know from Sec. 5, if angle θ of plane wave incidence on such an interface is larger than the critical value θ_c specified by Eq. (82), the wave is totally reflected. As a result, the waves launched into the fiber core at such "grazing" angles, propagate inside the core, repeatedly reflected from the cladding – see Fig. 25.

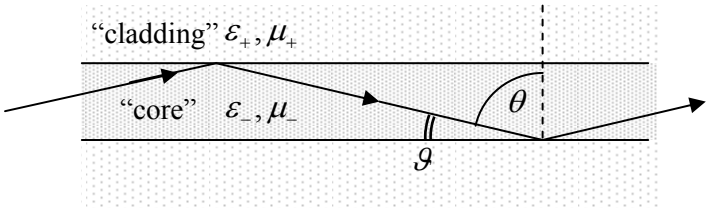


Fig. 7.25. Wave propagation in a thick optical fiber.

The most important type of dielectric waveguides are *optical fibers*.[51] Due to a heroic technological effort, in about three decades starting from the mid-1960s, the attenuation of glass fibers

[50] This formula ignores the fact that, according to our analysis, some modes (with $n = 0$ *and* $m = 0$ for H modes, and $n = 0$ *or* $m = 0$ for E modes, are forbidden. However, for $N >> 1$, the associated corrections of Eq. (91) are negligible.

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

has been decreased from the values of the order of 20 db/km (typical for the window glass) to the fantastically low values about 0.2 db/km (meaning a virtually perfect transparency of 10-km-long fiber segments!) – see Fig. 26a. It is remarkable that this ultralow power loss may be combined with an extremely low frequency dispersion, especially for near-infrared waves (Fig. 26b). In conjunction with the development of inexpensive erbium-based quantum amplifiers, this breakthrough has enabled inter-continental (undersea), broadband[52] optical cables, which are the backbone of all the modern telecommunication infrastructure. The only bad news is that these breakthroughs were achieved for just one kind of materials (silica-based glasses)[53] within a very narrow range of their chemical composition. As a result, the dielectric constants $\varepsilon_{\pm}/\varepsilon_0$ of the cladding and core of practical optical fibers are both close to 2.2 ($n_{\pm} \approx 1.5$) and are very close to each other, so that the relative difference of the refraction indices,

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

is typically below 0.5%, thus limiting the fiber bandwidth – see below.

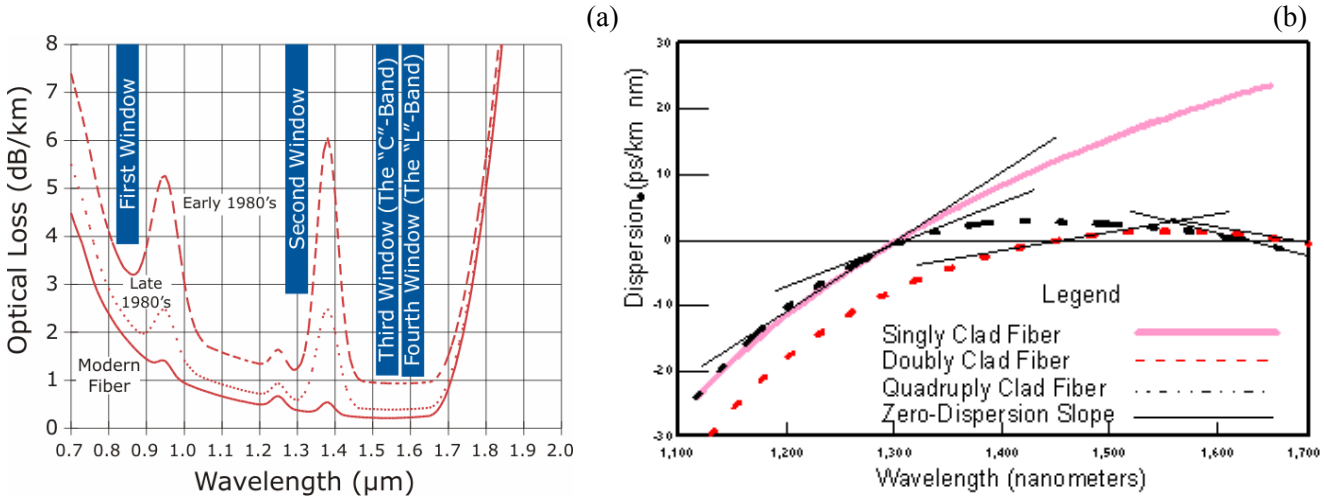


Fig. 7.26. (a) Attenuation and (b) dispersion of representative single-mode optical fibers. (Adapted, respectively, from http://olson-technology.com and http://www.timbercon.com.)

Practical optical fibers come in two flavors: *multi-mode* and *single-mode* ones. Multi-mode fibers, used for transfer of high optical power (up to as much as ~10 watts), have relatively thick cores, with a diameter $2R$ of the order of 50 μm, much larger than $\lambda \sim 1$ μm. In this case, the "geometric-optics" picture of the wave propagation discussed above is quantitatively correct, and we may use it to calculate the number of quasi-plane-wave modes that may propagate in the fiber. Indeed, for the complementary angle (Fig. 25)

[52] Each frequency band shown in Fig. 26a, at a typical signal-to-noise ratio $S/N > 10^5$ (50 db), corresponds to the Shannon bandwidth $\Delta f \log_2(S/N)$ exceeding 10^{14} bits per second, five orders of magnitude (!) higher than that of a modern Ethernet cable. And this is only per one fiber; an optical cable may have hundreds of them.

[53] The silica-based fibers were suggested in 1966 by C. Kao (the 2009 Nobel Prize in physics), but the very idea of using optical fibers for communications may be traced back to at least the 1963 work by J. Nishizawa.

$$\vartheta \equiv \frac{\pi}{2} - \theta , \tag{7.150}$$

Eq. (82) gives the propagation condition

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

For the case $\Delta << 1$, when the incidence angles $\theta > \theta_c$ of all propagating waves are close to $\pi/2$, and hence the complimentary angles are small, we can keep only two first terms in the Taylor expansion of the left-hand part 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. close to 5°. Due to this smallness, we can 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 number N of propagating modes:

Number of 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 $N \approx 150$.

The largest problem with using multi-mode fibers for communications is their high *geometric dispersion*, i.e. the difference of the mode propagation speed, which is usually characterized in terms of the signal delay time difference (traditionally measured in picoseconds per kilometer) between the fastest and the slowest mode. Within the geometric optics approximation, the difference of 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

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

For the example considered above, the TEM wave speed $v = c/n \approx 2\times10^8$ 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.) Such disastrous dispersion should be compared with *chromatic dispersion* that is due to the frequency dependence of $\varepsilon_\pm$, and has the steepness $(dt/d\lambda)/l$ of the order of 10 ps/km·nm (see the solid pink line in Fig. 26b). One can see that through the whole frequency band ($d\lambda \sim 100$ nm) the total chromatic dispersion dt/l is of the order of only 1,000 ps/km.

Due to the large geometric dispersion, the multimode fibers are used for signal transfer over only short distances (~ 100 m), while long-range communications are based on single-mode fibers, with thin cores (typically with diameters $2R \sim 5$ 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 we should get back to the Maxwell equations. In particular, this analysis should take into an explicit

account the evanescent wave propagating in the cladding, because its penetration depth may be comparable with R.[54]

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

$$f_- = f_l J_n(k_t \rho) e^{in\varphi}, \text{ with } k_t^2 = k_-^2 - k_z^2 > 0, \qquad k_-^2 \equiv \omega^2 \varepsilon_- \mu_-, \tag{7.156}$$

where amplitudes f_l (i.e., E_l and H_l) may be complex to account for the possible angular shift between these components. On the other hand, for the evanescent wave in the cladding, we may rewrite Eq. (102) as

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

Figure 27 illustrates the relation between k_t, κ_t, k_z, and $k_\pm$; note that the following sum,

$$k_t^2 + \kappa_t^2 \equiv \omega^2 (\varepsilon_- - \varepsilon_+) \mu_0, \tag{7.158}$$

Universal relation between k_t and κ_t

is fixed (at fixed frequency) and, for typical fibers, very small ($\sim 2\Delta k^2 << k^2$). By the way, Fig. 27 shows that neither of k_t and κ_t can be larger than $\omega[(\varepsilon_- - \varepsilon_+)\mu_0]^{1/2} = k\Delta^{1/2}$. In particular, this means that the depth $\delta = 1/\kappa_t$ of wave penetration into the cladding is at least $1/k\Delta^{1/2} = \lambda/2\pi\Delta^{1/2} >> \lambda/2\pi$. This is why the cladding layers in practical optical fibers are made as thick as ~50 μm, so that only a negligibly small tail of this evanescent wave field reaches their outer surfaces.

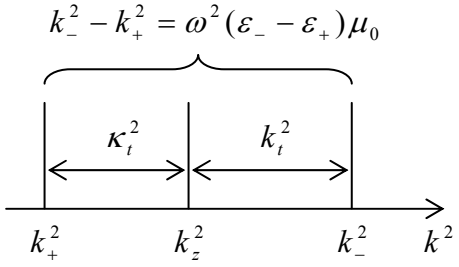


Fig. 7.27. Relation between the transverse exponents k_t and κ_t for waves in optical fibers.

In the polar coordinates, Eq. (157) becomes

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

instead of Eq. (139). From Sec. 2.5 we know that the eigenfunctions of Eq. (159) are the products of the angular factor $\exp\{in\varphi\}$ by a linear combination of the modified Bessel functions I_n and K_n, shown in

[54] I believe that the following calculation is important – both for practice, and as a good example of Maxwell theory application. However, its results will not be used in the following sections/chapters of the course, so that if the reader is not interested in this topic, he or she may safely jump to the beginning Sec. 9.

Fig. 2.20, now of argument $\kappa_t\rho$. In our case, the fields should vanish at $\rho \to \infty$, so that only the latter functions (of the second kind) can participate:

$$f_+ \propto K_n(\kappa_t \rho)e^{in\varphi}. \tag{7.160}$$

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 fields first calculated from using Eqs. (121). Such a conceptually simple, but a bit bulky calculation (which I am leaving for reader's exercise :-), yields a system of two linear, homogeneous equations for complex amplitudes E_l and H_l, that 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 prime means the derivative of each function over its full argument: $k_t\rho$ for J_n, and $\kappa_t\rho$ for K_n.

For any given frequency ω, the system of Eqs. (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 larger ratio 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 fields having just one finite longitudinal component (either E_z or H_z), and the *HE* modes are the usual *E* waves, while the *EH* modes are the *H* waves. For the *H* modes, the characteristic equation is reduced to the requirement that the second parentheses in the left-hand part of Eq. (161) equals to zero. Using the identities $J'_0 = -J_1$ and $K'_0 = -K_1$, this equation may be rewritten as

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

Using the simple relation between k_t and κ_t given by Eq. (158), we may plot both parts of Eq. (162) as a function of the same argument, say, $\xi \equiv k_tR$ – see Fig. 28.

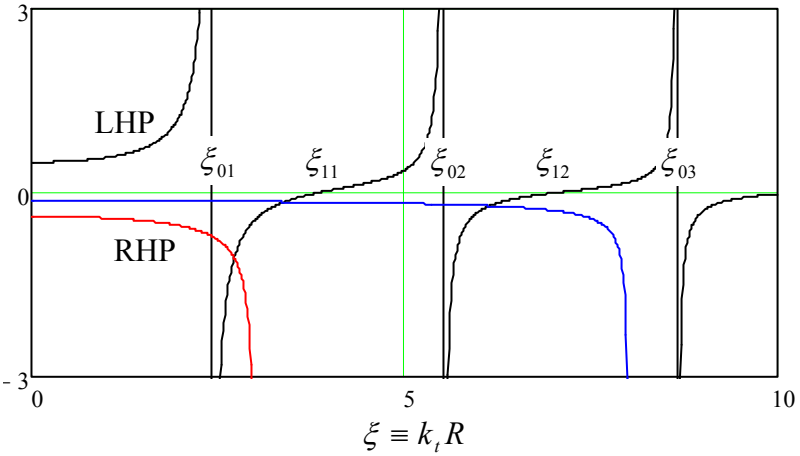


Fig. 7.28. Two sides of the characteristic equation (162), plotted as a function of k_tR, for two values of its dimensionless parameter: $V = 8$ (blue line) and $V = 3$ (red line). Note that according to Eq. (158), the argument of functions K_0 and K_1 is just $\kappa_tR = [V^2 - (k_tR)^2]^{1/2} = (V^2 - \xi^2)^{1/2}$.

The right-hand part of Eq. (162) depends not only on ξ but also on the dimensionless parameter V defined as the normalized right-hand part of Eq. (158):

$$V^2 \equiv \omega^2(\varepsilon_- - \varepsilon_+)\mu_0 R^2 \approx 2\Delta k_\pm^2 R^2 . \quad (7.163)$$

(According to Eq. (155), if $V >> 1$, it gives the doubled number of the fiber modes – the conclusion confirmed by Fig. 28, taking into account that it describes only the H modes.) Since the ratio K_1/K_0 is positive for all values of their argument (see, e.g., the right panel of Fig. 2.20), the right-hand part of Eq. (162) is always negative, so that the equation may have solutions only in the intervals where the ratio J_1/J_0 is negative, i.e. at

$$\xi_{01} < k_t R < \xi_{11}, \quad \xi_{02} < k_t R < \xi_{12}, \ldots , \quad (7.164)$$

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

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

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 in the left-hand part 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 waveguides, with no waves possible at frequencies below ω_c, but this is not so. The basic reason for the difference is that in metallic waveguides, the approach to ω_c results in the divergence of the longitudinal wavelength $\lambda_z \equiv 2\pi/k_z$. On the contrary, in dielectric waveguides this approach leaves λ_z finite ($k_z \to k_+$). Due to this difference, a certain linear superposition of HE and EH modes with $n = 1$ can propagate at frequencies well below the cutoff frequency for $n = 0$, which we have just calculated.[55] 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 *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 axis x parallel to the transverse component of the magnetic field vector, so that $E_x|_{\rho=0} = 0$, 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 z-components of the field, but also their x- and y-components) are proportional to $J_0(k_t\rho)$, with zero coefficients for E_x and H_y:

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

LP_{01} mode's fields distribution

Now we can readily calculate the longitudinal components, using the last two equations of Eqs. (100):

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

where I have used 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

[55] This fact becomes less surprising if we recall that in the circular metallic waveguide, discussed in Sec. 7, the lowest mode (H_{11}, Fig. 23) also corresponded to $n = 1$ rather than $n = 0$.

type (141) with $n = 1$. Note also that if $k_t << k_z$ (this relation is always true if $\Delta << 1$ – see Fig. 27), the longitudinal components of the fields are much smaller than their transverse counterparts, so that 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 in order to ensure the continuity of the fields at the core-to-cladding interface ($\rho = R$), we need to have a similar angular dependence of these components at $\rho \geq R$. The longitudinal components of the fields are tangential to the interface and thus should be continuous. Using the solutions similar to Eq. (160) with $n = 1$, we get

$$E_z = -i\frac{k_t}{k_z}\frac{J_1(k_t R)}{K_1(\kappa_t R)}E_0 K_1(\kappa_t \rho)\sin\varphi, \quad H_z = -i\frac{k_t}{k_z}\frac{J_1(k_t R)}{K_1(\kappa_t R)}H_0 K_1(\kappa_t \rho)\cos\varphi, \quad \text{for } \rho \geq R. \quad (7.168)$$

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 $\varepsilon_+ \approx \varepsilon_-$,[56] we can satisfy these conditions with the following solutions

$$E_x = 0, \quad E_y = \frac{J_0(k_t \rho)}{K_0(\kappa_t \rho)}E_0 K_0(\kappa_t \rho), \quad H_x = \frac{J_0(k_t \rho)}{K_0(k_t \rho)}H_0 K_0(\kappa_t \rho), \quad H_y = 0, \quad \text{for } \rho \geq R. \quad (7.169)$$

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

$$E_z = \frac{1}{-ik_z}\frac{\partial E_y}{\partial y} = -i\frac{\kappa_t}{k_z}\frac{J_0(k_t R)}{K_0(\kappa_t R)}E_0 K_1(\kappa_t \rho)\sin\varphi,$$
$$H_z = \frac{1}{-ik_z}\frac{\partial H_x}{\partial x} = -i\frac{\kappa_t}{k_z}\frac{J_0(k_t R)}{K_0(\kappa_t R)}H_0 K_1(\kappa_t \rho)\cos\varphi, \quad \text{for } \rho \geq R. \quad (7.170)$$

We see that this equation provides the same functional dependence of the fields as Eqs. (166), i.e. the internal and external fields are compatible, but their amplitudes coincide only if

LP_{01} mode's characteristic equation

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

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. 29. Indeed, its right-hand part is always positive, and the left-hand part tends to zero at $k_t R \to 0$. Due to this, Eq. (171) may have a solution for arbitrary small values of parameter V, defined by Eq. (159), i.e. for *arbitrary low frequencies*. This is why this mode is used in practical single-mode fibers: there are no other modes that can propagate at $\omega < \omega_c$, so that the geometric dispersion problem is avoided.

It is easy to use the Bessel function approximations given by the first term of the expansion (2.132) and also Eq. (2.157) to show that in the limit $V \to 0$ (i.e. $V << 1$), $\kappa_t R$ tends to zero much faster

[56] This is the core assumption of this approximate theory which accounts only for the most important effect of the difference of dielectric constants ε_+ and ε_-: the opposite signs of the differences $(k_+^2 - k_z^2) = k_t^2$ and $(k_-^2 - k_z^2) = -\kappa_t^2$. For more discussion of accuracy of this approximation and some exact results, let me refer the interested reader 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, which was cited above.

than $k_t R \approx V$: $\kappa_t R \to 2\exp\{-1/V\} << V$. 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 core. The drawback of this feature is that it requires very thick cladding, in order to avoid energy losses in outer ("buffer" and "jacket") layers that defend the silica components from the elements, but lack their low optical absorption. Due to this reason, the core radius is usually selected so that parameter V is just slightly less than the critical value $V_c = \xi_{01} \approx 2.4$ for higher modes, thus ensuring the single-mode operation and eliminating the geometric dispersion problem.

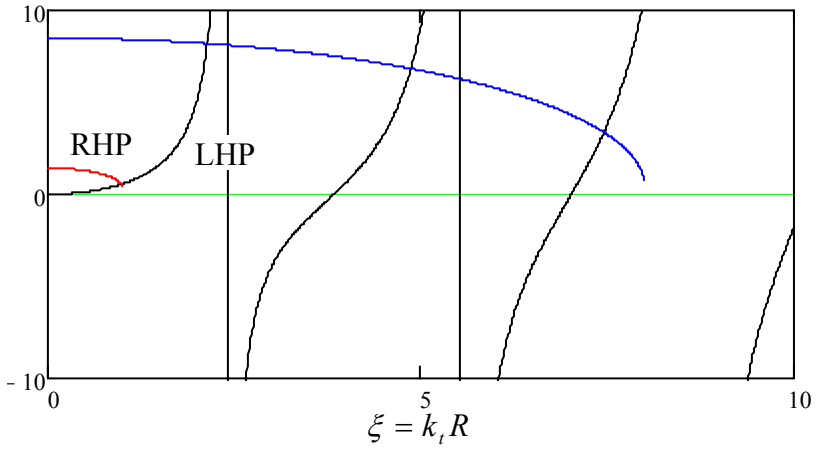


Fig. 7.29. Two sides of the characteristic equation (167) for the LP_{01} mode, plotted as a function of $k_t R$, for two values of the dimensionless parameter: $V = 8$ (blue line) and $V = 1$ (red line).

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

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

where $\zeta \equiv - [(d^2\varepsilon/d\rho^2)/\varepsilon]_{\rho=0}$ is a positive constant characterizing the fiber composition gradient.[57] 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, but with the space-dependent transverse wave vector: [58]

$$\left[\nabla_t^2 + k_t^2(\rho)\right] f = 0, \quad \text{where } k_t^2(\rho) \equiv k^2(\rho) - k_z^2 = \omega^2 \varepsilon(\rho)\mu_0 - k_z^2 = k_t^2(0)\left(1 - \frac{\zeta}{2}\rho^2\right). \tag{7.173}$$

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

[57] For an axially-symmetric fiber with a smooth function $\varepsilon(\rho)$, the *first* derivative $d\varepsilon/d\rho$ should vanish at $\rho = 0$.
[58] Such approach is invalid at arbitrary (large) ζ. Indeed, 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 $\nabla\varepsilon$.

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+k_t^2(0)\left(1-\frac{\varsigma}{2}x^2-\frac{\varsigma}{2}y^2\right)\right]f=0, \tag{7.174}$$

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

$$\frac{d^2X}{dX^2}+\frac{d^2Y''}{dY^2}+k_t^2(0)\left(1-\frac{\varsigma}{2}x^2-\frac{\varsigma}{2}y^2\right)=0, \tag{7.175}$$

so that functions X and Y obey the same similar differential equation,

$$\frac{d^2f}{dx^2}+k_x^2\left(1-\frac{\varsigma}{2}x^2\right)f=0, \quad f=\begin{cases}X,\\ Y.\end{cases} \tag{7.176}$$

with the separation constants satisfying the following relation:

$$k_x^2+k_y^2=k_t^2(0)=\omega^2\varepsilon(0)\mu_0-k_z^2. \tag{7.177}$$

Equation (176) is well known from the elementary quantum mechanics, because the Schrödinger equation for the perhaps most important quantum system, a 1D harmonic oscillator, may be rewritten in this form. Their eigenvalues are described by a simple formula

$$(k_x)_n=\left(\frac{\varsigma}{2}\right)^{1/2}(2n+1), \qquad (k_y)_m=\left(\frac{\varsigma}{2}\right)^{1/2}(2m+1), \qquad n,m=0,1,2,... \tag{7.178}$$

but eigenfunctions $X_n(x)$ and $Y_m(y)$ have to be expressed via not quite elementary functions - the Hermite polynomials.[59] For our 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 cutoff frequency:

$$\omega_c^2\varepsilon(0)\mu_0=(k_x^2)_0+(k_y^2)_0=\varsigma. \tag{7.179}$$

(Note that at $\zeta\to 0$, the cutoff frequency tends to zero, as it should be for a wave in a uniform medium.) The eigenfunctions corresponding to the lowest eigenvalues are simple:

$$f_0(x)=\text{const}\times\exp\left\{-\frac{\varsigma x^2}{4}\right\}, \tag{7.180}$$

so that the field distribution follows the Gaussian ("bell curve") function

$$f_0(\rho)=f_0(0)\exp\left\{-\frac{\varsigma(x^2+y^2)}{4}\right\}=f_0(0)\exp\left\{-\frac{\varsigma\rho^2}{4}\right\}. \tag{7.181}$$

This is the so-called *Gaussian beam*, very convenient for some applications. Still, the graded-index fibers have higher attenuation than their step-index counterparts, and are not used as broadly.

Speaking of the Gaussian beams (or more generally, any beams with axially-symmetric profile $f_0(\rho)$), I cannot help noticing the very curious option of forming so-called *helical waves* with complex amplitude $f_0(\rho)\exp\{il\varphi\}$, where l is an integer constant, and φ is the azimuthal angle (so that in our notation $x = \rho\cos\varphi$, $y = \rho\sin\varphi$). Let me leave it for reader's exercise to prove that the electromagnetic

[59] See, e.g., QM Sec. 2.6.

field of such a wave has an angular momentum vector $\mathbf{L} = L_z\mathbf{n}_z$, with L_z proportional to l.[60] Quantization of the helical field gives $L_z = l\hbar$ per photon. The case $l = \pm 1$ is possible for infinite-width beams (i.e. plane waves) and means their circular polarization, quantum-mechanically corresponding to spin ±1 of their photons - see the discussion in the end of Sec. 1. In contrast, the implementation of higher values of $|l|$ requires space-limited beams (with $f_0 \to 0$ at $\rho \to \infty$) and may be interpreted as giving the wave an additional "orbital" angular momentum.[61]

7.9. Resonators

Resonators are the distributed oscillators, i.e. structures that may sustain standing waves (in electrodynamics, oscillations of the electric and magnetic field at each point) even without a 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, this decay takes many oscillation periods. Alternatively, high-Q resonators may sustain oscillating fields permanently, if fed with a relatively weak incident wave.

Conceptually the simplest resonator is the *Fabry-Pérot interferometer*[62] that may be obtained by placing two well-conducting planes parallel to each other.[63] Indeed, in Sec. 1 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. (61) – that may be rewritten as

$$E(z,t) = \mathrm{Re}\left(2E_\omega e^{-i\omega t + i\pi/2}\right)\sin kz\,. \tag{7.182}$$

Hence the wave would not change if we had suddenly put 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.

$$kl = p\pi, \quad \text{where } p = 1, 2, ... \tag{7.183}$$

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

$$\omega_p = vk_p = \frac{\pi v}{a}p, \quad v = \frac{1}{(\varepsilon\mu)^{1/2}}, \tag{7.184}$$

[60] This task should be easier after reviewing results of field's momentum analysis in Sec. 9.8, in particular Eqs. (9.235) and (9.237).

[61] Theoretically, the possibility of separating of the angular momentum of an electromagnetic wave to the "spin" and "orbital" parts may be traced back to at least the 1943 work by J. Humblet; however, this issue had not been discussed in literature too much until the spectacular 1992 experiments by L. Allen *et al.* who demonstrated a simple way of generating such helical optical beams. (For reviews of this and later work see, e.g., G. Molina-Terriza *et al*., *Nature Physics* **3**, 305 (2007) and/or L. Marrucchi *et al*., *J. Opt*. **13**, 064001 (2011), and references therein.) Presently there are efforts to use this approach for so-called "orbital angular moment (OAM) multiplexing" of waves for high-rate information transmission – see, e.g., J. Wang *et al*., *Nature Photonics* **6**, 488 (2012).

[62] The device is named after its inventors, M. Fabry and A. Pérot; and is also called the *Fabry-Pérot etalon* (meaning "gauge"), because of its initial usage for the light wavelength measurement.

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

has a simple physical sense: the resonator length l equals exactly p half-waves of 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 in resonator analysis is finding its eigenfrequencies ω_p that are now determined by the system geometry rather than by an external wave source.

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

$$H(z,t) = \mathrm{Re}\left(2\frac{E_\omega}{Z}e^{-i\omega t}\right)\cos kz \quad . \tag{7.185}$$

Expressions (182) and (185) 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 those instants the electric field energy of the resonator vanishes, but the total energy stays constant, because the magnetic field oscillates (also simultaneously at all points) with the phase shift $\pi/2$. Such behavior is typical for all electromagnetic resonators.

Another, more technical remark is that we can readily get the same results (182)-(185) by solving the Maxwell equations from the 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 their 1D form, say, for the transverse component of the electric field:

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)E = 0, \tag{7.186}$$

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

$$\frac{1}{Z}\frac{d^2Z}{dz^2} - \frac{1}{v^2}\frac{1}{\mathcal{T}}\frac{d^2\mathcal{T}}{dt^2} = 0\,. \tag{7.187}$$

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

$$\frac{d^2Z}{dz^2} + k^2 Z = 0, \tag{7.188}$$

$$\frac{d^2\mathcal{T}}{dt^2} + k^2 v^2 \mathcal{T} = 0, \tag{7.189}$$

both with sinusoidal solutions, so that their product is a standing wave with a wave vector k and frequency $\omega = kv$, which may be presented by Eq. (182).[64] Now using the boundary conditions $E(0, t) = E(l, t) = 0$,[65] we get the eigenvalue spectrum for k_p and hence for $\omega_p = vk_p$, given by Eqs. (183) and (184).

[64] In this form, the equations are valid even in the presence of dispersion, but with the frequency-dependent wave speed: $v^2 = 1/\varepsilon(\omega)\mu(\omega)$.

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

Lessons from this simple case study may be readily generalized for an arbitrary resonator: there are (at least :-) two methods of finding the eigenfrequency spectrum:

(i) We may look at a traveling wave solution and find where reflecting mirrors may be inserted without affecting the wave's structure. Unfortunately, this method is limited to simple geometries.

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

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) f(\mathbf{r},t) = 0\,, \tag{7.190}$$

for field components, as an eigenvalue problem with appropriate boundary conditions. If system parameters (and hence coefficient v) do not change in time, the spatial and temporal variables of Eq. (185) may be *always* separated by taking

$$f(\mathbf{r},t) = \mathcal{R}(\mathbf{r})\mathcal{T}(t)\,, \tag{7.191}$$

where function $\mathcal{T}(t)$ *always* obeys the same equation (189), having the sinusoidal solution of frequency $\omega = vk$. Plugging this solution back into Eq. (190), for the spatial distribution of the field we get the *3D Helmholtz equation*,

$$\boxed{\left(\nabla^2 + k^2\right)\mathcal{R}(\mathbf{r}) = 0\,,} \tag{7.192}$$

3D Helmholtz equation

whose solution (for non-symmetric geometries) may be much more complex.

Let us use these methods to find the eigenfrequency spectrum of a few simple, but practically important resonators. 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) between two conducting lids perpendicular to the line 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 (184).

Now let us analyze a slightly more complex system: a rectangular metallic-wall cavity of volume $a\times b\times l$ – see Fig. 30. In order to use the first method, let us consider the resonator as a finite-length ($\Delta z = l$) of the rectangular waveguide stretched along axis z, which was analyzed in detail in Sec. 7. As a reminder, for $a < b$, in the basic H_{10} traveling wave mode, both **E** and **H** do not depend on y, with vector **E** having only y-component. On the contrary, vector **H** has both components H_x and H_z, with the phase shift $\pi/2$ between them, with component H_x having the same phase as E_y – see Eqs. (131), (137), and (138). Hence, if a plane, perpendicular to axis z, is placed so that the electric field vanishes on it, H_x also vanishes, so that all the boundary conditions (104) pertinent to a perfect metallic wall are fulfilled simultaneously.

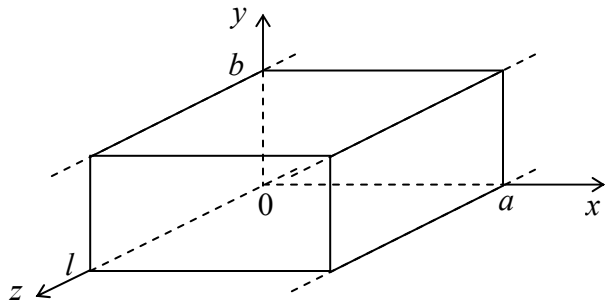


Fig. 7.30. Rectangular metallic resonator as a finite section of a waveguide with the cross-section shown in Fig. 25.

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

$$k_z = \left(k^2 - k_t^2\right)^{1/2} = \left(\frac{\omega^2}{v^2} - \frac{\pi^2}{a^2}\right). \tag{7.193}$$

Using this expression, we see that the smallest of these distances, $l = \lambda_z/2 = \pi/k_z$, gives resonance frequency[66]

Basic mode's frequency

$$\omega_{101} = v\left[\left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{l}\right)^2\right]^{1/2}, \tag{7.194}$$

with the indices showing the number of half-waves along each dimension of the system. This is the lowest (fundamental) eigenfrequency 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 phases of the magnetic and electric fields are shifted by phase $\pi/2$ both in space and time, just as in the Fabry-Pérot resonator – see Eqs. (182) and (185). Such time shift allows for a very simple interpretation of the H_{101} mode that is especially adequate for very flat resonators, with $b << a, l$. At the instant when the electric field reaches maximum (Fig. 31a), i.e. the magnetic field vanishes in the whole volume, the surface electric charge of the walls (with density $\sigma = E_n/\varepsilon$) is largest, being localized mostly in the middle of the broadest (in Fig. 31, horizontal) faces of the resonator. At 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 period of the oscillation period of frequency ω_{101} – see Fig. 31b. The currents generate the vortex magnetic field, with looped field lines in the plane of the broadest face. 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. 31a. After that, the surface currents stat to flow in the direction opposite to that shown in Fig. 31b. This process, that 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 faces of the resonator, and that of distributed inductance, mostly by its narrower walls.

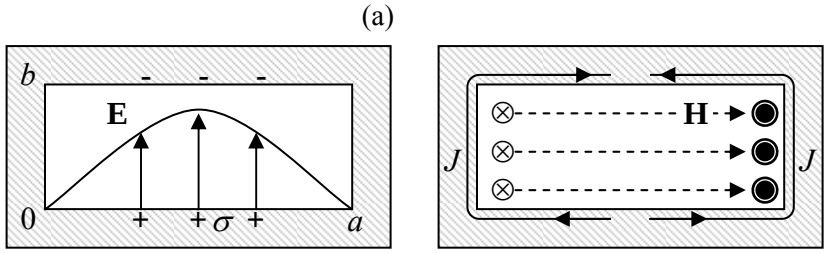


Fig. 7.31. Fields, charges, and currents in the basic H_{101} mode of a rectangular metallic resonator, at two instants separated by $\Delta t = \pi/2\omega_{101}$ - schematically.

In order to generalize result (194) to higher oscillation modes, the second method discussed above is more prudent. Separating variables as $\mathcal{E}(\mathbf{r}) = X(x)Y(y)Z(z)$ in the Helmholtz equation (192), we

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

see that X, Y, and Z have to be sinusoidal functions of their arguments, with wave vector components satisfying the characteristic equation

$$k_x^2 + k_y^2 + k_z^2 = k^2 \equiv \frac{\omega^2}{v^2}. \tag{7.195}$$

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

$$\begin{aligned} E_x &= E_1 \cos k_x x \sin k_y y \sin k_z z, \quad H_x = H_1 \sin k_x x \cos k_y y \cos k_z z, \\ E_y &= E_2 \sin k_x x \cos k_y y \sin k_z z, \quad H_y = H_2 \cos k_x x \sin k_y y \cos k_z z, \\ E_z &= E_3 \sin k_x x \sin k_y y \cos k_z z, \quad H_z = H_3 \cos k_x x \cos k_y y \sin k_z z, \end{aligned} \tag{7.196}$$

with each of the wave vector components having the equidistant spectrum similar to the one given by Eq. (193):

$$k_x = \frac{\pi n}{a}, \quad k_y = \frac{\pi m}{b}, \quad k_z = \frac{\pi p}{l}, \tag{7.197}$$

so that the full spectrum of eigenfrequencies is given by the following formula,

$$\omega_{nmp} = vk = v\left[\left(\frac{\pi n}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2 + \left(\frac{\pi p}{l}\right)^2\right]^{1/2}, \tag{7.198}$$

which is a natural generalization of Eq. (194). Note, however, that of 3 integers m, n, and p at least two have to be different from zero, in order to keep the fields (196) nonvanishing.

Let us use Eq. (199) to evaluate the number of different modes in a relatively small region $d^3k \ll k^3$ (which is still much larger than the reciprocal volume, $1/V = 1/abl$, of the resonator) of the wave vector space. Taking into account that each eigenfrequency (198), with $nml \neq 0$, corresponds to two field modes with different polarizations,[67] the argumentation absolutely similar to the one used in the end of Sec. 7 for the 2D case yields

$$\boxed{dN = 2V\frac{d^3k}{(2\pi)^3}}. \tag{7.199}$$

Oscillation mode density

This property, valid for resonators of arbitrary shape, is broadly used in classical and quantum statistical physics,[68] in the following form. If some electromagnetic mode property, $f(\mathbf{k})$, is a smooth function of the wave vector, and volume V is large enough, then Eq. (199) may be used to approximate the sum over the modes by an integral:

[67] This fact becomes evident from plugging Eq. (196) 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 (197) for each wave vector component, may be satisfied by two linearly independent sets of constants $E_{1,2,3}$.

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

$$\sum_{\mathbf{k}} f(\mathbf{k}) \approx \int_N f(\mathbf{k}) dN = \int_{\mathbf{k}} f(\mathbf{k}) \frac{dN}{d^3k} d^3k = 2\frac{V}{(2\pi)^3} \int_{\mathbf{k}} f(\mathbf{k}) d^3k \,. \tag{7.200}$$

Finally, note that low-loss resonators may be also formed by finite-length sections of not only metallic waveguides with different cross-sections, but also of the dielectric waveguides. Moreover, even the a simple slab of a dielectric material with a μ/ε ratio substantially different from that of its environment (say, the free space) may be used as a high-Q Fabry-Pérot interferometer (Fig. 32), due to an effective wave reflection from its surfaces at normal and especially inclined incidence – see, respectively, Eqs. (68) and Eqs. (91) and (95).

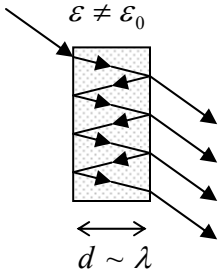


Fig. 7.32. Dielectric Fabry-Pérot interferometer.

Actually, such dielectric Fabry-Pérot interferometer is frequently more convenient for practical purposes than a metallic resonator, due to its natural coupling to environment, that enables a ready way of wave insertion and extraction. The back side of the same medal is that this coupling to environment provides an additional mechanism of power losses, limiting the resonance quality – see the next section.

7.10. Energy loss effects

Inevitable energy losses ("power dissipation") in passive media lead, in two different situations, to two different effects. In a long transmission line fed by a constant wave source at one end, the losses lead to a gradual *attenuation* of the wave, i.e. to the decrease of its amplitude, and hence power $\mathcal{P}$, with the distance z along the line. In linear materials, the losses are proportional to the wave amplitude squared, i.e. to the time- average of the power itself, so that the energy balance on a small segment dz takes the form

$$d\mathcal{P} = -\frac{d\mathcal{P}_{\text{loss}}}{dz} dz = -\alpha \mathcal{P} dz \,. \tag{7.201}$$

Coefficient α, participating in the last form of Eq. (201) and defined by relation

$$\alpha \equiv \frac{d\mathcal{P}_{\text{loss}} / dz}{\mathcal{P}} \,, \tag{7.202}$$

is called the *attenuation constant*.[69] Comparing the evident solution of Eq. (201),

[69] In engineering, attenuation is frequently measured in *decibels per meter* (acronymed as db/m or just dbm):

$$\alpha\Big|_{\text{db/m}} \equiv 10 \log_{10} \frac{\mathcal{P}(z=0)}{\mathcal{P}(z=1\,\text{m})} = 10 \log_{10} e^{\alpha[1/\text{m}]} = \frac{10}{\ln 10} \alpha \left[\text{m}^{-1}\right] \approx 4.34\, \alpha \left[\text{m}^{-1}\right].$$

$$\mathcal{P}(z) = \mathcal{P}(0)e^{-\alpha z}, \tag{7.203}$$

Wave attenuation

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

$$\alpha = 2\,\mathrm{Im}\,k_z, \tag{7.204}$$

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 $\alpha << \mid k_z \mid \approx \mathrm{Re}\,k_z$, its effects on the field distributions 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, energy losses in the filling dielectric, and the skin effect in conducting walls, are independent and additive.

The dielectric losses are especially simple to describe. Indeed, a review of our calculations in Secs. 6-8 shows that all of them remain valid if either $\varepsilon(\omega)$, or $\mu(\omega)$, or both, and hence $k(\omega)$, have small imaginary parts:

$$k'' = \omega\,\mathrm{Im}\left[\varepsilon^{1/2}(\omega)\mu^{1/2}(\omega)\right] << k'. \tag{7.205}$$

In TEM transmission lines, $k = k_z$, and hence Eq. (205) yields

$$\alpha_{\text{dielectric}} = 2k'' = 2\omega\,\mathrm{Im}\left[\varepsilon^{1/2}(\omega)\mu^{1/2}(\omega)\right]. \tag{7.206}$$

Energy loss in filling dielectric

For dielectric waveguides, in particular optical fibers, these losses are the main attenuation mechanism. As we already know from Sec. 8, in practical optical fibers $\kappa_t R >> 1$, i.e. most of the field propagates (as the evanescent wave) in the cladding, and the wave mode is very close to TEM. This is why it is sufficient to use Eq. (206) for the cladding material alone.

In waveguides with non-TEM waves, we can readily use the relations between k_z and k derived above to re-calculate k'' into Im k_z. (Note that as such re-calculation, values of k_t stay real, because they are just the eigenvalues of the Helmholtz equation (101), which does not include k.).

In waveguides and transmission lines with metallic conductors, much higher energy losses may come from the skin effect. Let us calculate them, assuming that we know the field distribution in the wave, in particular, the tangential component H of the magnetic field at conductor surface. Then, if the wavelength λ is much larger than δ_s, as it usually is,[70] we may use the results of the quasistatic approximation derived in Sec. 6.2, in particular Eqs. (6.27)-(6.28) for the relation between the complex amplitudes of the current density in the conductor and the tangential magnetic field

$$j_\omega(x) = k_- H_\omega(x), \quad k_- = -\frac{(1-i)}{\delta_s}, \quad \delta_s^2 = \frac{2}{\mu\omega\sigma}. \tag{7.207}$$

The power loss density (per unit volume) may be now calculated by time averaging of Eq. (4.39):

$$\mathcal{P}_{\text{loss}}(x) = \frac{|j_\omega(x)|^2}{2\sigma} = \frac{|k_-|^2 |H_\omega(x)|^2}{2\sigma} = \frac{|H_\omega(x)|^2}{\delta_s^2\sigma}, \tag{7.208}$$

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

and its integration along the normal to the surface (through all the skin depth), using the exponential law (6.26). This (elementary) integration yields the following power loss per unit area:[71]

Energy loss in metallic walls

$$\frac{d\mathcal{P}_{\rm loss}}{dA} \equiv \int_0^\infty \mathcal{P}_{\rm loss}(x)dx = \left|H_\omega(0)\right|^2 \frac{\mu\omega\delta_s}{4}. \tag{7.209}$$

The total power loss $d\mathcal{P}_{\rm loss}/dz$ per unit length of a waveguide, i.e. the right-hand part of Eq. (201), now may be calculated by the integration of the ratio $\mathcal{P}_{\rm loss}/A$ along the contour(s) limiting the cross-section of all conductors of the line. Since our calculation is only valid for low losses, we may ignore their effect on the field distribution, so that the unperturbed distribution may be used both in Eq. (209), i.e. the nominator of Eq. (202), and also for the calculation of the average propagating power, i.e. the denominator of Eq. (202), 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 TEM transmission lines, the coaxial cable (Fig. 19). As we already know from Sec. 6, in the absence of losses, the distribution of TEM mode fields is the same as in statics, namely:

$$H_z = 0, \quad H_\rho = 0, \quad H_\varphi(\rho) = H_0\frac{a}{\rho}, \tag{7.210}$$

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

$$E_z = 0, \quad E_\rho(\rho) = ZH_\varphi(\rho) = ZH_0\frac{a}{\rho}, \quad E_\varphi = 0, \quad Z \equiv \left(\frac{\mu}{\varepsilon}\right)^{1/2}. \tag{7.211}$$

Now we can, neglecting losses for now, use Eq. (42) to calculate the time-averaged Poynting vector

$$\overline{S} = \frac{Z\left|H_\varphi(\rho)\right|^2}{2} = \frac{Z\left|H_0\right|^2}{2}\left(\frac{a}{\rho}\right)^2, \tag{7.212}$$

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

$$\mathcal{P} = \int_A \overline{S}d^2r = \frac{Z\left|H_0\right|^2 a^2}{2}2\pi\int_a^b \frac{\rho d\rho}{\rho^2} = \pi Z\left|H_0\right|^2 a^2 \ln\frac{b}{a}. \tag{7.213}$$

For the particular case of the coaxial cable (Fig. 19), the contours limiting the wall cross-sections are circles of radii $\rho = a$ (where the surface field amplitude $H_\omega(0)$ equals, in our notation, H_0), and $\rho = b$ (where, according to Eq. (204), the field is a factor of b/a lower). As a result, for the power loss per unit length, Eq. (209) yields

$$\frac{d\mathcal{P}_{\rm loss}}{dz} = \left(2\pi a\left|H_0\right|^2 + 2\pi b\left|H_0\frac{a}{b}\right|^2\right)\frac{\mu_0\omega\delta_s}{4} = \frac{\pi}{2}a\left(1+\frac{a}{b}\right)\mu\omega\delta_s\left|H_0\right|^2. \tag{7.214}$$

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. (213)-(214) into the definition (202) of α, to calculate the part of the attenuation constant associated with the skin effect:

[71] For a normally-incident plane wave, this formula would bring us back to Eq. (78).

$$\alpha_{\rm skin} \equiv \frac{d\mathcal{P}_{\rm loss}/dz}{\mathcal{P}} = \frac{1}{2\ln(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right)\frac{\mu\omega\delta_s}{Z} = \frac{k\delta_s}{2\ln(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right). \tag{7.215}$$

We see that the relative (dimensionless) attenuation, α/k, scales approximately as the ratio $\delta_s/\min[a, b]$. This result is should be compared with Eq. (78) for the normal incidence of plane waves on a conducting surface.

Let us evaluate α for the standard TV cable RG-6/U (with copper conductors of diameters $2a$ = 1 mm, $2b$ = 4.7 mm, and $\varepsilon \approx 2.2\ \varepsilon_0$, $\mu \approx \mu_0$). According to Eq. (6.27a), for f = 100 MHz ($\omega \approx 6.3\times10^8$ 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 that the attenuation length scale $\lambda\!\!\!^{-} \equiv 1/\alpha$ is about 60 m. Hence the attenuation in a cable connecting a roof TV antenna to a TV set in the same house 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, an attempt to use the same cable in the X-band ($f \sim 10$ GHz) is more problematic. Indeed, though the skin depth $\delta_s \propto \omega^{-1/2}$ decreases with frequency, the wave length drops, i.e. k increases, even faster ($k \propto \omega$), so that the attenuation $\alpha_{\rm skin} \propto \omega^{1/2}$ becomes close to 0.16 m, and $\lambda\!\!\!^{-}$ to ~6 m. This is why at such frequencies, it is more customary 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. (209) and the results derived in Sec. 9, for reader's exercise.

The power loss effect on free oscillations *in resonators* is different: there it leads to a gradual decay of oscillation energy $\mathcal{E}$ in time. The useful measure of this decay, called the *Q factor*, may be introduced by writing the temporal analog of Eq. (201):

$$d\mathcal{E} = -\mathcal{P}_{\rm loss}dt = -\frac{\omega}{Q}\mathcal{E}dt\,, \tag{7.216}$$

where ω in the eigenfrequency in the loss-free limit, and the dimensional Q factor is defined by a relation parallel to Eq. (202): [72]

$$\boxed{\frac{\omega}{Q} \equiv \frac{\mathcal{P}_{\rm loss}}{\mathcal{E}}\,.} \tag{7.217}$$

Q-factor's definition

The solution to Eq. (216),

$$\boxed{\mathcal{E}(t) = \mathcal{E}(0)e^{-t/\tau}, \quad \text{with } \tau \equiv \frac{Q}{\omega} = \frac{Q/2\pi}{\omega/2\pi} = \frac{QT}{2\pi},} \tag{7.218}$$

Oscillation energy decay

which is an evident temporal analog of Eq. (203), shows the physical meaning of the Q factor: the characteristic time τ of the oscillation energy decay is $(Q/2\pi)$ times longer than the oscillation period $T = 2\pi/\omega$. (Another interpretation of Q comes from the relation[73]

$$\boxed{Q = \frac{\omega}{\Delta\omega},} \tag{7.219}$$

FWHM bandwidth

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

[73] See, e.g., CM Sec. 4.1.

where $\Delta\omega$ is the so-called *FWHM* [74] 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 forced oscillations induced in the resonator by an input signal is twice lower than its resonant value.)

In the important particular case of resonators formed by insertion of metallic walls into a TEM transmission line of 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 if the line attenuation coefficient α is already known. In fact, as was discussed in Sec. 9 above, the standing waves in such a resonator, of the length given by Eq. (183): $l = p(\lambda/2)$ with $p = 1, 2,\ldots$, may be understood as an overlap of two TEM waves running in opposite directions, or in other words, a traveling wave and its reflection from one of the ends, the whole roundtrip taking time $\Delta t = 2l/v = p\lambda/v = 2\pi p/\omega = pT$. According to Eq. (201), at this distance the wave's power should drop by $\exp\{-2\alpha l\} = \exp\{-p\alpha\lambda\}$. On the other hand, the same decay may be viewed as happening in time, and according to Eq. (216), result in the drop by $\exp\{-\Delta t/\tau\} = \exp\{-(pT)/(Q/\omega)\} = \exp\{-2\pi p/Q\}$. Comparing these two exponents, we get

Q vs. α

$$Q = \frac{2\pi}{\alpha\lambda} = \frac{k}{\alpha}. \tag{7.220}$$

This simple relation neglects the losses at wave reflection from the walls limiting the resonator length. Such approximation is indeed legitimate at $a << \lambda$; 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 smaller than that given by Eq. (220), and needs to be calculated directly. 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 nominator and denominator of the right-hand part of Eq. (217) may be calculated neglecting the effects of the power loss on the field distribution in the resonator. I am leaving such a calculation, for the simplest (rectangular and circular) resonators, for reader's exercise.

To conclude this chapter, the last remark: in some resonators (including certain dielectric resonators and metallic resonators with holes in their walls), additional losses due to wave radiation into the environment are also possible. In some simple cases (say, the Fabry-Pérot interferometer shown in Fig. 32) the calculation of these *radiative losses* is straightforward, but sometimes it requires more elaborated approaches, which will be discussed in the next chapter.

7.11. Exercise problems

7.1.* Find the temporal Green's function of a medium whose complex dielectric constant obeys Eq. (32), using:

(i) the Fourier transform, and
(ii) the direct solution of Eq. (30), which describes the corresponding model of the medium.

Hint: For the Fourier transform, you may like to use the Cauchy integral.[75]

7.2. The electric polarization of a material responds in the following way to an electric field step:[76]

[74] This is the acronym for "Full Width at Half-Maximum".
[75] See, e.g., MA Eq. (15.2).

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

where τ is a positive constant. 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 dielectric constant $\varepsilon(\omega)$ for 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 Lorentz oscillator model of an atom, given by Eq. (30), to calculate the average potential energy of the atom 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 $E_\omega(\mathbf{r})$. Discuss the conditions of validity of your result.

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

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

Discuss the conditions of validity of your result.

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

7.7. As was briefly discussed in Sec. 2,[77] a wave pulse of a finite but relatively large spatial extension $\Delta r >> \lambda \equiv 2\pi/k$ may be represented with a *wave packet* – a sum of sinusoidal waves with wave vectors $\mathbf{k}$ within a relatively narrow interval. Consider an electromagnetic plane wave packet of this type, with the electric field distribution

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

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

---

[76] This function $E(t)$ is of course proportional to the well-known step function θ - see, e.g., MA Eq. (14.3). I am not using this notion just to avoid a possible confusion between two different uses of the Greek letter θ.

[77] And in more detail in CM Sec. 5.3, and especially in QM Sec. 2.1.

7.8.* Analyze the effect of a constant, uniform magnetic field $\mathbf{B}_0$, parallel to the direction **n** of electromagnetic wave propagation, on the wave dispersion in plasma, within the same simple model that was used in the lecture notes for derivation of Eq. (7.38). (Limit your analysis to relatively weak waves, whose magnetic field is negligible in comparison with $\mathbf{B}_0$.)

Hint: You may like to represent the incident wave as a linear superposition of two circularly polarized waves, with the left- and right-hand polarization.

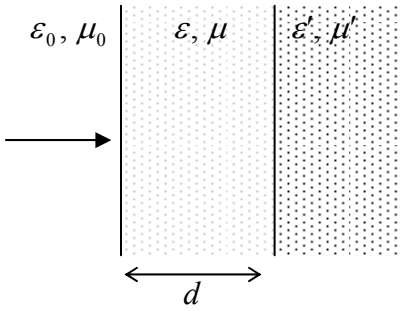
7.9. A monochromatic, plane electromagnetic wave is normally incident from free space on a uniform slab of a material with electric permittivity ε and magnetic permeability μ, with the slab thickness d comparable with the wavelength.

(i) Calculate the power transmission coefficient $\mathcal{T}$, i.e. the fraction of the incident 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 function $\mathcal{T}(\omega)$ depend on the film thickness d and the wave impedance $Z = (\mu/\varepsilon)^{1/2}$ of its material?

7.10. A monochromatic, plane electromagnetic wave with free-space wave number k_0 is normally incident on a plane conducting film of thickness $d \sim \delta_s << 1/k_0$. Calculate the power transmission coefficient of the system, i.e. the fraction of incident wave's power propagating beyond the film. Analyze the result in the limits of small and large ratios d/δ_s.

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

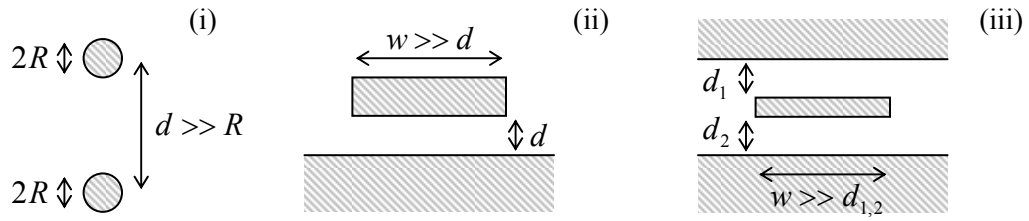


7.12. A monochromatic, plane wave is incident from inside a medium with $\varepsilon\mu > \varepsilon_0\mu_0$ on its plane surface, at the angle of incidence θ larger than the critical angle $\theta_c = \sin^{-1}(\varepsilon_0\mu_0/\varepsilon\mu)^{1/2}$. Calculate the depth δ of the evanescent wave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

7.13. Analyze the possibility of propagation of surface electromagnetic waves along a plane 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 propagation direction. (After solving the problem, justify this mode choice.)

7.14. Calculate the characteristic impedance Z_W of the long, straight TEM transmission lines formed by metallic electrodes with cross-sections shown in Fig. below:



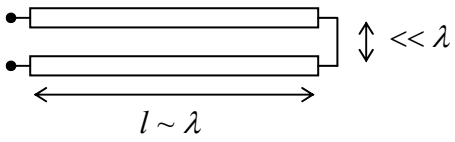
(i) two round, parallel wires, separated by distance $d >> R$,
(ii) *microstrip line* of width $w >> d$,
(iii) *stripline* with $w >> d_1 \sim d_2$,

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

7.15. Modify results of Problem 10(ii) for a superconductor microstrip line, taking into account the magnetic field penetration into both the strip and the ground plane.

7.16.* What lumped ac circuit would be equivalent to the system shown in Fig. 20, with incident wave's power $\mathcal{P}_i$? Assume that the wave reflected from the load circuit does not return to it.

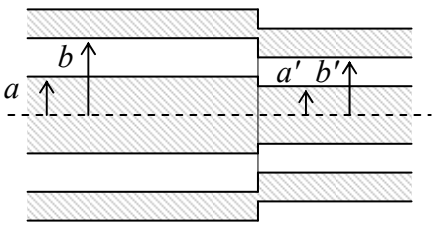
7.17. Find the lumped ac circuit equivalent to a loss-free TEM transmission line of length $l \sim \lambda$, with a small cross-section area $A << \lambda^2$, as "seen" (measured) from one end, if the line's conductors are galvanically connected ("shortened") at the other end – see Fig. on the right.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



7.18. 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 .

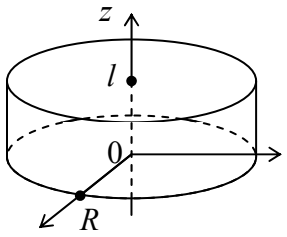
7.19.* For a metallic coaxial cable with the circular cross-section (Fig. 21), find the lowest non-TEM mode and calculate its cutoff frequency.

7.20. Two coaxial cable sections are connected coaxially - see Fig. on the right, which shows 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?



7.21.* Use the recipe outlined in Sec. 8 to prove the characteristic equation (161) for the *HE* and *EH* modes in a round, step-index optical fiber.

7.22. Find the lowest eigenfrequencies, and corresponding oscillation modes, of a round cylindrical resonator (see Fig. on the right) with perfectly conducting walls.



7.23. A plane, monochromatic wave propagates through a medium whose Ohmic conductance σ dominates the power losses, while the electric and magnetic polarization effects are negligible. Calculate the wave attenuation coefficient and relate the result with some calculation carried out in Chapter 6.

7.24. Generalize the telegrapher's equations (110)-(111) by taking into account small energy losses in:

(i) transmission line's conductors, and
(ii) the media separating the conductors,

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

7.25. Calculate the skin-effect contribution to the attenuation coefficient α, defined by Eq. (202), for the basic (H_{10}) mode propagating in a waveguide with the rectangular cross-section – see Fig. 22. Use the results to evaluate α and L for 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 estimate with that, made in Sec. 10, for a standard coaxial cable, for the same frequency.

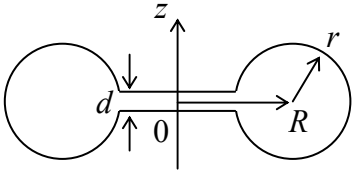
7.26.* Calculate the skin-effect contribution to the attenuation coefficient α of

(i) the basic (H_{11}) mode, and
(ii) the H_{01} mode

in a metallic waveguide with the circular cross-section (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.

7.27. For a rectangular metallic-wall resonator with dimensions $a \times b \times l$ ($b \le a, l$), calculate the Q-factor in the fundamental (lowest) oscillation mode, due to the skin-effect losses in the walls. Evaluate the factor (and the lowest eigenfrequency) for a $23 \times 23 \times 10$ mm^3 resonator with copper walls, at room temperature.

7.28.* Calculate the lowest eigenfrequency and Q factor (due to the skin-effect losses) of the toroidal (axially-symmetric) resonator with metallic walls and interior's cross-section shown in Fig. on the right, within the limit $d << r, R$.



7.29. Express the contribution to the damping coefficient (the reciprocal Q-factor) of a resonator, due to small energy losses in the dielectric that fills it, via dielectric's complex functions $\varepsilon(\omega)$ and $\mu(\omega)$ of the material.

7.30. For the dielectric Fabry-Pérot resonator (Fig. 32) with the normal wave incidence, find the Q-factor due to radiation losses in the limit of strong impedance mismatch ($Z >> Z_0$), using two methods:

(i) from the energy balance, using Eq. (217), and
(ii) from the frequency dependence of the power transmission coefficient, using Eq. (219).

Compare the results.