

Chapter 3

Constitutive Relations

Maxwell's equations define the fields that are generated by currents and charges. However, they do not describe how these currents and charges are generated. Thus, to find a self-consistent solution for the electromagnetic field, Maxwell's equations must be supplemented by relations that describe the behavior of matter under the influence of fields. These material equations are known as constitutive relations.

The constitutive relations express the secondary sources \mathbf{P} and \mathbf{M} in terms of the fields \mathbf{E} and \mathbf{H} , that is $\mathbf{P} = f[\mathbf{E}]$ and $\mathbf{M} = f[\mathbf{H}]$.¹ According to Eq. (1.19) this is equivalent to $\mathbf{D} = f[\mathbf{E}]$ and $\mathbf{B} = f[\mathbf{H}]$. If we expand these relations into power series

$$D = D_0 + \left. \frac{\partial D}{\partial E} \right|_{E=0} E + \frac{1}{2} \left. \frac{\partial^2 D}{\partial E^2} \right|_{E=0} E^2 + \dots \quad (3.1)$$

we find that the lowest-order term depends *linearly* on \mathbf{E} . In most practical situations the interaction of radiation with matter is weak and it suffices to truncate the power series after the linear term. The nonlinear terms come into play when the fields acting on matter become comparable to the atomic Coulomb potential. This is the territory of strong field physics. Here we will entirely focus on the linear properties of matter.

¹In some exotic cases we can have $\mathbf{P} = f[\mathbf{E}, \mathbf{H}]$ and $\mathbf{M} = f[\mathbf{H}, \mathbf{E}]$, which are so-called bi-isotropic or bi-anisotropic materials. These are special cases and won't be discussed here.

3.1 Linear Materials

The most general linear relationship between \mathbf{D} and \mathbf{E} can be written as²

$$\mathbf{D}(\mathbf{r}, t) = \varepsilon_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\varepsilon}(\mathbf{r}-\mathbf{r}', t-t') \mathbf{E}(\mathbf{r}', t') d^3\mathbf{r}' dt', \quad (3.2)$$

which states that the response \mathbf{D} at the location \mathbf{r} and at time t not only depends on the excitation \mathbf{E} at \mathbf{r} and t , but also on the excitation \mathbf{E} in all other locations \mathbf{r}' and all other times t' . The integrals represent summations over all space and over all times. The response function $\tilde{\varepsilon}$ is a tensor of rank two. It maps a vector \mathbf{E} onto a vector \mathbf{D} according to $D_i = \sum_j \tilde{\varepsilon}_{ij} E_j$, where $\{i, j\} \in \{x, y, z\}$. A material is called *temporally dispersive* if its response function at time t depends on previous times. Similarly, a material is called *spatially dispersive* if its response at \mathbf{r} depends also on other locations. A spatially dispersive medium is also designated as a *nonlocal medium*.

Note that Eq. (3.2) is a convolution in space and time. Using the Fourier transform with respect to both time and space, that is,

$$\hat{\mathbf{D}}(k_x, k_y, k_z, \omega) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{D}(x, y, z, t) e^{ik_x x} e^{ik_y y} e^{ik_z z} e^{i\omega t} dx dy dz dt, \quad (3.3)$$

allows us to rewrite Eq. (3.2) as

$$\hat{\mathbf{D}}(\mathbf{k}, \omega) = \varepsilon_0 \varepsilon(\mathbf{k}, \omega) \hat{\mathbf{E}}(\mathbf{k}, \omega), \quad (3.4)$$

where ε is the Fourier transform of $\tilde{\varepsilon}$. Note that the response at (\mathbf{k}, ω) now only depends on the excitation at (\mathbf{k}, ω) and not on neighboring (\mathbf{k}', ω') . Thus, a nonlocal relationship in space and time becomes a *local relationship* in Fourier space! This is the reason why life often is simpler in Fourier space.

The principle of causality states that there can be no response before the excitation took place. Consequently, $\tilde{\varepsilon}(\mathbf{r}-\mathbf{r}', t-t') = 0$ for all $t' > t$. In Fourier space, this causality requirement leads to a relation between $\text{Re}\{\varepsilon(\mathbf{k}, \omega)\}$ and $\text{Im}\{\varepsilon(\mathbf{k}, \omega)\}$, known as the Kramers-Kronig Relations.

²We assume that the response is translation invariant in both time and space.

Spatial dispersion, i.e. a nonlocal response, is encountered near material surfaces or in objects whose size is comparable with the mean-free path of electrons. In general, nonlocal effects are very difficult to account for. In most cases of interest the effect is very weak and we can safely ignore it. Temporal dispersion, on the other hand, is a widely encountered phenomenon and it is important to take it accurately into account. Thus, we will be mostly concerned with relationships of the sort

$$\hat{\mathbf{D}}(\mathbf{r}, \omega) = \varepsilon_0 \varepsilon(\mathbf{r}, \omega) \hat{\mathbf{E}}(\mathbf{r}, \omega), \quad (3.5)$$

where $\varepsilon(\mathbf{r}, \omega)$ is called the *dielectric function*, also called the relative electric permittivity. Similarly, for the magnetic field we obtain

$$\hat{\mathbf{B}}(\mathbf{r}, \omega) = \mu_0 \mu(\mathbf{r}, \omega) \hat{\mathbf{H}}(\mathbf{r}, \omega), \quad (3.6)$$

with $\mu(\mathbf{r}, \omega)$ being the relative magnetic permeability. Notice that the spectral representation of Maxwell's equations [(2.25)–(2.28)] is formally identical to the complex notation used for time-harmonic fields [(2.31)–(2.34)] . Therefore, Eqs. (3.5) and (3.6) also hold for the complex amplitudes of time-harmonic fields

$$\underline{\mathbf{D}}(\mathbf{r}) = \varepsilon_0 \varepsilon(\mathbf{r}, \omega) \underline{\mathbf{E}}(\mathbf{r}), \quad \underline{\mathbf{B}}(\mathbf{r}) = \mu_0 \mu(\mathbf{r}, \omega) \underline{\mathbf{H}}(\mathbf{r}) \quad (3.7)$$

However, these equations generally do not hold for time-dependent fields $\mathbf{E}(\mathbf{r}, t)$! One can use (3.7) for time-dependent fields only if dispersion can be ignored, that is $\varepsilon(\omega) = \varepsilon$ and $\mu(\omega) = \mu$. The only medium that is strictly dispersion-free is vacuum.

3.1.1 Electric and Magnetic Susceptibilities

The linear relationships (3.7) are often expressed in terms of the electric and magnetic susceptibilities χ_e and χ_m , respectively. These are defined as

$$\underline{\mathbf{P}}(\mathbf{r}) = \varepsilon_0 \chi_e(\mathbf{r}, \omega) \underline{\mathbf{E}}(\mathbf{r}), \quad \underline{\mathbf{M}}(\mathbf{r}) = \chi_m(\mathbf{r}, \omega) \underline{\mathbf{H}}(\mathbf{r}). \quad (3.8)$$

Using the relations (1.19) we find that $\varepsilon = (1 + \chi_e)$ and $\mu = (1 + \chi_m)$.

3.1.2 Conductivity

The conductivity σ relates an induced conduction current $\underline{\mathbf{j}}_{\text{cond}}$ in a linear fashion to an exciting field $\underline{\mathbf{E}}$. Similar to Eq. (3.7), this relationship can be represented as

$$\underline{\mathbf{j}}_{\text{cond}}(\mathbf{r}) = \sigma(\mathbf{r}, \omega) \underline{\mathbf{E}}(\mathbf{r}) . \quad (3.9)$$

It turns out that the conduction current is accounted for by the imaginary part of $\varepsilon(\omega)$ as we shall show in the following.

Maxwell's curl equation for the magnetic field (2.33) reads as

$$\nabla \times \underline{\mathbf{H}}(\mathbf{r}) = -i\omega \underline{\mathbf{D}}(\mathbf{r}) + \underline{\mathbf{j}}_0(\mathbf{r}) . \quad (3.10)$$

We now introduce the linear relationship (3.7) and obtain

$$\begin{aligned} \nabla \times \underline{\mathbf{H}}(\mathbf{r}) &= -i\omega \varepsilon_0 \varepsilon(\mathbf{r}, \omega) \underline{\mathbf{E}}(\mathbf{r}) + \underline{\mathbf{j}}_0(\mathbf{r}) \\ &= -i\omega \varepsilon_0 [\varepsilon'(\mathbf{r}, \omega) + i\varepsilon''(\mathbf{r}, \omega)] \underline{\mathbf{E}}(\mathbf{r}) + \underline{\mathbf{j}}_0(\mathbf{r}) , \end{aligned} \quad (3.11)$$

where we split ε into a real and imaginary part as $\varepsilon = \varepsilon' + i\varepsilon''$. The current associated with ε'' is equal to $\omega \varepsilon_0 \varepsilon''(\mathbf{r}, \omega) \underline{\mathbf{E}}(\mathbf{r})$. Comparison with (3.9) yields

$$\sigma(\mathbf{r}, \omega) = \omega \varepsilon_0 \varepsilon''(\mathbf{r}, \omega) \quad (3.12)$$

Thus, we see that the conductivity acts like the imaginary part of the electric permeability and that we can simply accommodate σ in ε by using a complex dielectric function. In the complex notation one does not distinguish between conduction currents and polarization currents. Energy dissipation is associated with the imaginary part of the dielectric function (ε'') whereas energy storage is associated with its real part (ε').

With the definitions of ε and μ , the wave equations for the complex fields $\underline{\mathbf{E}}(\mathbf{r})$ and $\underline{\mathbf{H}}(\mathbf{r})$ in linear media are

$$\nabla \times \mu(\mathbf{r}, \omega)^{-1} \nabla \times \underline{\mathbf{E}}(\mathbf{r}) - k_0^2 \varepsilon(\mathbf{r}, \omega) \underline{\mathbf{E}}(\mathbf{r}) = i\omega \mu_0 \underline{\mathbf{j}}_0(\mathbf{r}) \quad (3.13)$$

$$\nabla \times \varepsilon(\mathbf{r}, \omega)^{-1} \nabla \times \underline{\mathbf{H}}(\mathbf{r}) - k_0^2 \mu(\mathbf{r}, \omega) \underline{\mathbf{H}}(\mathbf{r}) = \nabla \times \varepsilon(\mathbf{r}, \omega)^{-1} \underline{\mathbf{j}}_0(\mathbf{r}) \quad (3.14)$$

where $k_0 = \omega/c$ denotes the vacuum wavenumber. Note that these equations are also valid for anisotropic media, *i.e.* if ε and μ are tensors. We emphasize that Eqs. (3.13) and (3.14) are expressed solely in terms of the current density $\underline{\mathbf{j}}_0$, which accounts for the source current density and any secondary sources not accounted for by ε and μ , respectively.

If μ is isotropic and spatially independent, that is $\mu(\mathbf{r}, \omega) = \mu(\omega)$, then we can multiply Eq. (3.13) on both sides with μ . Furthermore, if there are no sources we can drop \mathbf{j}_0 and obtain the Helmholtz equation (2.12), but with the difference that now $k^2 = k_0^2 \varepsilon \mu$, that is,

$$\nabla^2 \underline{\mathbf{E}}(\mathbf{r}) + k^2 \underline{\mathbf{E}}(\mathbf{r}) = \nabla^2 \underline{\mathbf{E}}(\mathbf{r}) + k_0^2 n^2 \underline{\mathbf{E}}(\mathbf{r}) = 0 \quad (3.15)$$

where $n = \sqrt{\varepsilon \mu}$ is called *index of refraction*.

Conductors

The conductivity σ is a measure for how good a conductor is. For example, quartz has a conductivity of $\sigma_{\text{SiO}_2} = 10^{-16} \text{ A / V m}$, and the conductivity of copper is $\sigma_{\text{Au}} = 10^8 \text{ A / V m}$. These values are different by 24 orders of magnitude! There are hardly any other physical parameters with a comparable dynamic range.

The net charge density ρ inside a conductor is zero, no matter whether it transports a current or not. To see this, we assume that the current is dominated by conduction, that is $\mathbf{j} = \sigma \mathbf{E}$, and insert it into the charge conservation law $\nabla \cdot \mathbf{j} + \partial \rho / \partial t = 0$. Using Gauss' law $\nabla \cdot \mathbf{E} = \rho / (\varepsilon_0 \varepsilon)$ yields

$$\frac{\partial}{\partial t} \rho(t) = -\frac{\sigma}{\varepsilon_0 \varepsilon} \rho(t), \quad (3.16)$$

which has the solution

$$\rho(t) = \rho(t=0) e^{-t \sigma / (\varepsilon_0 \varepsilon)}. \quad (3.17)$$

Thus, any charge inside the conductor dissipates within a time of $T_\rho = \varepsilon_0 \varepsilon / \sigma$. For a perfect conductor, $\sigma \rightarrow \infty$ and hence $\rho(t) = 0$. For realistic conductors with finite σ the characteristic time is $T_\rho \sim 10^{-19} \text{ s}$, which is so short that it can be neglected. Notice that we have assumed that the conductivity σ is dispersion free, that is, σ

has no memory. This is, of course, an approximation.

When a charge moves through a conductor it undergoes collisions with the lattice. After a collision event, the charge is accelerated by the external field until it is slowed down by the next collision. For good conductors (copper), the time between collisions is typically in the order of $\tau \sim 10^{-14}$ s. The sequence of acceleration and deceleration events results in a finite velocity \mathbf{v}_d for the charge, called the *drift velocity*. The current density due to a charge density moving at finite speed is $\mathbf{j} = q \mathbf{v}_d n$, where n is the charge density, *i.e.* the number of charges per unit volume.³ The drift velocity \mathbf{v}_d is proportional to the driving field \mathbf{E} and the proportionality constant is called *mobility* μ .⁴ Thus, $\sigma = nq\mu$.

In a good conductor the polarization current $\partial \mathbf{P} / \partial t$ due to bound charges can be neglected because it is much smaller than the conduction current \mathbf{j} . In terms of a complex dielectric constant $\epsilon = \epsilon' + i\epsilon''$ (c.f. Equation 3.12) this implies that $\epsilon'' \gg |\epsilon'|$, that is, $\sigma \gg |\omega \epsilon_0 \epsilon'|$. Evidently, the higher the frequency ω is, the more

³The charge density n should not to be confused with the index of refraction n .

⁴Not to be confused with the magnetic permeability μ .

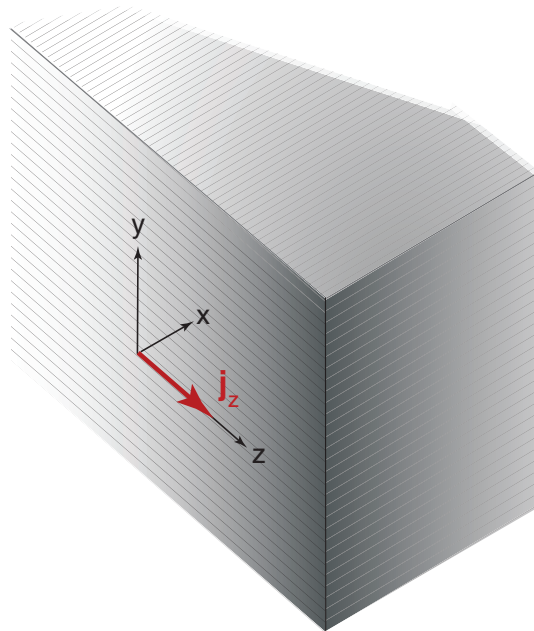


Figure 3.1: Current density $\mathbf{j} = j_z \mathbf{n}_z$ flowing along the surface of a conductor.

challenging it gets to fulfill this condition. In fact, at optical frequencies, metals are no longer good conductors and they are dominated by the polarization current. At lower frequencies, however, it is legitimate to ignore the polarization current when dealing with good conductors. Ignoring $\partial\mathbf{P}/\partial t$ is equivalent to ignoring the real part of the complex dielectric function (3.12), which implies $k^2 = (\omega/c)^2\mu\varepsilon \approx i(\omega/c)^2(\mu\sigma/\omega\varepsilon_0)$. Consequently, the Helmholtz equation (3.15) reads as

$$\nabla^2 \underline{\mathbf{E}}(\mathbf{r}) + i\omega\sigma\mu_0\mu \underline{\mathbf{E}}(\mathbf{r}) = 0, \quad (3.18)$$

and because of $\underline{\mathbf{j}} = \sigma \underline{\mathbf{E}}$ the same equation holds for the current density $\underline{\mathbf{j}}$. Note that we're using complex equation and that $\mathbf{j}(\mathbf{r}, t) = \text{Re}\{\underline{\mathbf{j}}(\mathbf{r}) \exp[-i\omega t]\}$.

Let us now consider a semi-infinite conductor, as illustrated in Fig. 3.1. This situation corresponds to a small section of a wire's surface. The conductor has a surface at $x = 0$ and transports a current $\underline{\mathbf{j}}(\mathbf{r})$ in the z direction. Because of the invariance in y and z we set $\underline{j}_z(x) = A \exp[Bx]$ and insert into Eq. (3.18). Noting that $\sqrt{i} = (1+i)/\sqrt{2}$ we obtain

$$\underline{j}_z(x) = \underline{j}_z(x=0) e^{-(1-i)x/D_s} \quad \text{with} \quad D_s = \sqrt{\frac{2}{\sigma\mu_0\mu\omega}}. \quad (3.19)$$

The length D_s is called the *skin depth*. Since $|\underline{j}_z(x)/\underline{j}_z(0)| = \exp[-x/D_s]$ it describes the penetration of fields and currents into the metal. Evidently, for a perfect conductor ($\sigma \rightarrow \infty$) the skin depth becomes $D_s = 0$, that is, all the current is transported on the surface of the metal. D_s also decreases with increasing frequency ω , but eventually the result (3.19) becomes inaccurate because the polarization current $\partial\mathbf{P}/\partial t$ becomes stronger than the conduction current. In wires of finite diameter the skin depth also depends on the curvature of the wire. At low frequencies, the conductance of a wire scales with the cross-section of the wire, but at high frequencies, the current is confined to the surface of the wire and hence the conductance scales with the circumference.

