4.19 *Blackbody radiation in a small cavity.* Consider thermal radiation in equilibrium inside such a cubic cavity. Compute the radiation energy density in a cubic cavity of length $L = 1$ μm at $T = 400$ K and compare it with the Planck distribution obtained by assuming that the cavity is very large compared to the wavelength.

4.20 *Entropy of one phonon state.* From eqs. (4.14) and (4.40), show that the entropy, s, of one phonon state having a frequency ω obeys the following relationship:

$$\frac{\hbar\omega}{T} f_0(1 + f_0) = -\frac{\kappa_B T}{\hbar} \frac{\partial s}{\partial \omega}$$

Where f_0 is the Bose–Einstein distribution.

5

Energy Transfer by Waves

The wave–particle duality of matter from quantum mechanics implies that energy carriers have both wave and particle characteristics. One way to think about this duality is that material waves are granular rather than continuous. For example, a phonon wave at frequency ν contains a discrete number of identical waves, each having an energy $h\nu$. A fundamental property of waves is their phase information. A coherent wave has a fixed relationship between two points in space or at two different times. Due to the fixed phase relationship, the superposition of waves from the same source creates interference and diffraction phenomena that are familiar in optics.

The wave characteristics of matter (electrons, phonons, and photons) are important for transport processes at interfaces and in nanostructures. We have seen in previous chapters that the size effects on energy quantization can be considered as a result of the formation of standing waves. In this chapter, we will discuss the reflection of waves at a single interface, and interference and tunneling phenomena in thin films and multilayers. We will make parallel presentations for three major energy carriers: electrons, photons, and phonons. We have discussed rather extensively in chapters 2 and 3 the electron waves based on the Schrödinger equation. Optical wave effects are readily observable and can be understood from classical electrodynamics based on the Maxwell equations, which will be reviewed in this chapter. For phonons, we will adapt a continuum approach based on the acoustic waves, rather than on the discrete lattice dynamics method we used in chapter 3. The acoustic-wave-based approach allows us to treat phonons in parallel with electrons and photons. We will see that wave reflection, interference, and tunneling phenomena can occur for all three types of carriers and the descriptions of these phenomena are also similar (table 1.4), despite the differences in their statistical behavior, dispersion, and origin (table 1.3), as we discussed in previous

chapters. Readers familiar with any of these waves can use the analogy for understanding the other waves.

For macroscale transport processes, however, we seldom consider the phase of material waves. Rather, we treat the entities as particles. Why and when can we do so? Section 5.6 answers these questions and briefly discusses transport in the partially coherent regime.

5.1 Plane Waves

When throwing a stone into water, one can observe a concentric wave propagating outward. Television antennas emit electromagnetic waves that are approximately spherical. Rather than considering these nonplanar waves, we will carry out most of our discussion in this chapter on the basis of plane waves, although the phenomena to be discussed also exist for other forms of waves such as the cylindrical or spherical waves. A *plane wave* is one that has a constant amplitude at any plane perpendicular to the direction of propagation at any fixed time. These waves must satisfy the equation governing their motion. Later, we will discuss these governing equations, such as the Maxwell equations for electromagnetic waves. Before getting into these details, let's first examine some common forms of plane waves. For example, in chapter 2, we showed that the wavefunction of a free electron is [eq. (2.34)]

$$\Psi(x,t) = A_1 e^{-i(\omega t - kx)} + A_2 e^{-i(\omega t + kx)} \tag{5.1}$$

where the first term represents a plane wave traveling in the positive x-direction and the second term in the negative x-direction. These are scalar plane waves because the wavefunction is a scalar. Other waves, such as the electromagnetic field, are vector waves because the electric/magnetic fields have directions. We can express a harmonic, vector plane wave propagating in three-dimensional space as

$$\mathbf{F}(t,\mathbf{r}) = \mathbf{A}\sin(\omega t - \mathbf{k}\bullet\mathbf{r}) \tag{5.2}$$

where $\mathbf{A}$ represents the amplitude and direction of the field (e.g., electric, magnetic, or atomic displacement), ω is the angular frequency, $\mathbf{k}$ [with components (k_x, k_y, k_z)] is the wavevector representing the direction of wave propagation and its spatial periodicity ($|\mathbf{k}| = 2\pi/\lambda$), and $\mathbf{r}$ is the spatial coordinate. Equation (5.2) is a plane wave because all the points $\mathbf{r}$ satisfying $\mathbf{k}\bullet\mathbf{r} = constant$ form a plane perpendicular to the wavevector $\mathbf{k}$, and the field $\mathbf{F}$ is a constant on this plane at any given time.

Very often, it is much more convenient to use the complex representation rather than the sine and cosine representation for the waves. For example, instead of eq. (5.2), we can write $\mathbf{F}$ as

$$\mathbf{F}(t,\mathbf{r}) = \mathbf{A}\exp[-i(\omega t - \mathbf{k}\bullet\mathbf{r})] \tag{5.3}$$

When using such a complex representation, we resort to either the real part or the imaginary part of the final solution as the solution of the problem, in accordance with whether the initial or boundary conditions are represented in terms of cosine (real part) or sine (imaginary part) functions. For example, the imaginary part of $\mathbf{F}$ in eq. (5.3) is

identical to eq. (5.2) and we thus expect that the solution to a physical problem will be the imaginary part of the complex variables used in solving the governing equations.

In the following sections, we will examine three types of waves: the electron wave as a material waves, the electromagnetic wave governing the radiation transfer, and the acoustic wave representing lattice vibration.

5.1.1 Plane Electron Waves

In chapter 2, we dealt extensively with electron waves in planar geometries, such as free electrons and electrons in a potential well. The wavefunction of a plane electron wave propagating along the positive x-direction is

$$\Psi(x,t) = A\exp[-i(\omega t - kx)] \tag{5.4}$$

From the Schrödinger equation, we obtained in chapter 2 the following dispersion relation between the electron energy E and wavevector k

$$k = \sqrt{\frac{2m(E-U)}{\hbar^2}} \tag{5.5}$$

where U is the electrostatic potential. The particle current (or flux) can be calculated from [eq. (2.31)]:

$$\mathbf{J} = \frac{i\hbar}{2m}(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi) = \mathrm{Re}\left[\frac{i\hbar}{m}\Psi\nabla\Psi^*\right] \tag{5.6}$$

As we will see later, this flux expression is similar to the Poynting vector that represents the energy flux of electromagnetic and acoustic waves.

5.1.2 Plane Electromagnetic Waves

In this section, we will introduce the Maxwell equations that govern the propagation of electromagnetic waves. We will show that a plane wave of the form of eq. (5.3) satisfies the Maxwell equations and discuss how to calculate the energy flux of the electromagnetic waves.

An electromagnetic wave in vacuum is characterized by an *electric field vector* $\mathbf{E}$ [$\mathrm{N\,C^{-1}} = \mathrm{V\,m^{-1}}$], and a *magnetic field vector* $\mathbf{H}$ [$\mathrm{C\,m^{-1}\,s^{-1}} = \mathrm{A\,m^{-1}}$]. When the electromagnetic field interacts with a medium, under the force of the electric and magnetic fields the electrons and ions of the atoms in the medium are set into motion. These electrons and ions generate their own electric and magnetic fields that influence each other and are superimposed onto the external fields. For example, the positive ions and negative electrons of an atom under an external field will be deformed from the original equilibrium condition, forming an electrical dipole.* A measure of the capability of the material to respond to the incoming electric field is the electric polarization per unit

*A dipole is a pair of positive charge Q and negative charge $-Q$, separated by a small distance a. The dipole moment of the pair of charges equals $p = Qa$.

volume, or the *dipole moment* per unit volume $\mathbf{P}$ [$\mathrm{C\,m^{-2}}$], which is related to the electric field through the electric susceptibility χ,

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E} \tag{5.7}$$

where ε_0 is the vacuum permittivity, $\varepsilon_0 = 8.85 \times 10^{-12}$ [$\mathrm{C^2\,N^{-1}\,m^{-2}} = \mathrm{F\,m^{-1}}$], and the electric susceptibility is nondimensional. The total field inside the medium is measured by the *electric displacement* $\mathbf{D}$ [$\mathrm{C\,m^{-2}}$], which is a superposition of the contributions from the external electric field and the electric polarization,

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 (1 + \chi)\mathbf{E} = \varepsilon \mathbf{E} \tag{5.8}$$

where ε is called the electrical *permittivity* of the medium.

The electron and ion motion in a medium also induces a magnetic field, which is superimposed onto the external magnetic field. A measure of the total magnetic field inside the medium is called the *magnetic induction* $\mathbf{B}$ [$\mathrm{N\,s\,m^{-1}\,C^{-1}} = \mathrm{N\,A^{-1}\,m^{-1}}$],

$$\mathbf{B} = \mu \mathbf{H} \tag{5.9}$$

where μ is the magnetic *permeability*. In vacuum and in most materials, $\mu = \mu_0 = 4\pi \times 10^{-7}\ \mathrm{N\,s^2\,C^{-2}} = \mathrm{H\,m^{-1}}$. If a material has μ larger than μ_0, it is called paramagnetic, and if $\mu < \mu_0$ it is diamagnetic. Many materials are nonmagnetic, with $\mu = \mu_0$.

The electric dipoles and electric polarization are most appropriate when describing the distortions of electrons bound to ions. The free electrons that are not bound to any atoms will also be set into motion by the external electric field. The motion of free electrons forms a current, which is related to the electric field through Ohm's law,

$$\mathbf{J}_e = \sigma_e \mathbf{E} \tag{5.10}$$

where σ_e [$\mathrm{C^2\,N^{-1}\,m^{-1}\,s^{-1}} = \Omega^{-1}\,\mathrm{m^{-1}}$] is the electrical *conductivity*.

The propagation of an electromagnetic wave is governed by the following *Maxwell equations*:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{5.11}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}_e \tag{5.12}$$

$$\nabla \bullet \mathbf{D} = \rho_e \tag{5.13}$$

$$\nabla \bullet \mathbf{B} = 0 \tag{5.14}$$

where ρ_e is the net charge density [$\mathrm{C\,m^{-3}}$]. Equation (5.11) is the Faraday law, which states that a changing magnetic field induces an electric field. Without the first term on the right-hand side, eq. (5.12) is the Ampère law, which says an electric current induces a magnetic field. Maxwell's (1831–1879) ingenuity lies in the first term of the right-hand side of eq. (5.12), which represents the current due to electron oscillation around the ion even though the electrons are not free to move. This additional term, however, places the electric and the magnetic fields at similar positions with respect to time and space, and endows the electromagnetic field with a wave type of behavior, as we will see later. The Maxwell equations govern the propagation of all electromagnetic waves, including, for example, the signals of cellular phones, radios and televisions, lasers, light, thermal radiation, and X-rays, despite the fact that these waves are generated by different sources. The differences between these waves, in terms of propagation, are mainly the wavelength.

We will next demonstrate how to derive a wave type of equation for the electric field. By taking the curl of eq. (5.11), we get

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) \tag{5.15}$$

The left-hand side can be manipulated using the vector identity

$$\nabla \times \nabla \times \mathbf{E} \equiv \nabla(\nabla \bullet \mathbf{E}) - \nabla^2 \mathbf{E} \tag{5.16}$$

For a region free of electrical charge, eq. (5.13) leads to

$$\nabla \bullet \mathbf{D} = 0 \text{ and thus } \nabla \bullet \mathbf{E} = 0 \tag{5.17}$$

We should mention that the second of the above equations is based on the assumption that the dielectric constant ε is independent of space, which is not the case for the photonic crystal that we discussed in chapter 3. Substituting eqs. (5.16) and (5.17) into eq. (5.15) and using $\mathbf{B} = \mu \mathbf{H}$ [eq. (5.9)] yields

$$-\nabla^2 \mathbf{E} = -\mu \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) \tag{5.18}$$

To eliminate $\mathbf{H}$, we substitute eq. (5.12) into the above equation and utilize eqs. (5.8) and (5.10),

$$\nabla^2 \mathbf{E} = \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu\sigma_e \frac{\partial \mathbf{E}}{\partial t} \tag{5.19}$$

If $\sigma_e = 0$, eq. (5.19) is clearly a wave type of equation. The first-order derivative in eq. (5.19) introduces damping to the wave, because of the absorption by free electrons. A similar equation for the magnetic field can be derived. We seek a solution to eq. (5.19) in the form of a plane wave, eq. (5.3),

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \exp[-i(\omega t - \mathbf{k} \bullet \mathbf{r})] \tag{5.20}$$

where $\mathbf{E}_0$ represents both the amplitude and direction of the electric field. Substituting eq. (5.20) into eq. (5.19), we obtain

$$\mathbf{k} \bullet \mathbf{k} = \mu\varepsilon\omega^2 + i\mu\sigma_e\omega = \mu\omega^2[\varepsilon_0(1 + \chi) + i\sigma_e/\omega] = \mu\hat{\varepsilon}\omega^2 \tag{5.21}$$

or

$$|\mathbf{k}| = \frac{N\omega}{c_0} \tag{5.22}$$

with

$$c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}},\ \hat{\varepsilon} = \varepsilon_0(1 + \chi) + i\sigma_e/\omega,\ N = \sqrt{\frac{\mu\hat{\varepsilon}}{\mu_0\varepsilon_0}} = n + i\kappa \tag{5.23}$$

where c_0 is the speed of light in vacuum, $\hat{\varepsilon}$ is called the *complex permittivity*, and N is called the *complex refractive index* or *complex optical constant*.* The real part of N, n, is the usual refractive index of materials. The imaginary part of N, κ, is called the *extinction coefficient* and measures the damping of the electromagnetic field, which arises not only from the absorption of free electrons [the conductivity part in eq. (5.21)], but also from the dipole oscillation of bound electrons and other mechanisms [the susceptibility part of eq. (5.21)]. For nonmagnetic materials ($\mu = \mu_0$),

$$N = \sqrt{\frac{\hat{\varepsilon}}{\varepsilon_0}} = \sqrt{\varepsilon_r} = n + i\kappa \tag{5.24}$$

and $\varepsilon_r = \hat{\varepsilon}/\varepsilon_0$ is called the *dielectric constant* or dielectric function. Neither N nor ε_r is really a constant, as their names suggest, because they are dependent on wavelength. Studies on the wavelength dependence of the dielectric function can lead to insights into the material constituents and energy states. For example, some electron and phonon states can be identified from measuring the dielectric function or complex refractive indices. There exists a large library of complex refractive indices of materials (Palik, 1985).

Substituting eq. (5.22) into eq. (5.20), we see that the electric field of a harmonic electromagnetic wave can be expressed as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \exp\left[-i\omega\left(t - \frac{N}{c_0}\hat{\mathbf{k}} \bullet \mathbf{r}\right)\right] \tag{5.25}$$

where $\hat{\mathbf{k}}$ is the unit vector along the wavevector direction. With the electric field determined, the magnetic field can be computed according to eq. (5.11). One can further prove that the electromagnetic wave is a transverse wave, and that the electric and magnetic fields are perpendicular to each other:

$$\mathbf{E} \perp \mathbf{H} \perp \mathbf{k} \tag{5.26}$$

In the special case that a plane wave is traveling along the x-direction with the electric and magnetic fields pointing in the y- and z-directions, respectively, the electric and magnetic fields can be expressed as

$$\mathbf{E}_y = E_{ya} \exp\left[-i\omega\left(t \mp \frac{Nx}{c_0}\right)\right]\hat{\mathbf{y}} \tag{5.27}$$

$$\mathbf{H}_z = H_{za} \exp\left[-i\omega\left(t \mp \frac{Nx}{c_0}\right)\right]\hat{\mathbf{z}} \tag{5.28}$$

where the minus and plus signs represent waves propagating in the positive and negative x-direction, respectively. Substituting eqs. (5.27) and (5.28) back into the Maxwell equation (5.11), we can see that the electric and the magnetic fields are related:

$$H_{za} = \pm\frac{N}{\mu c_0} E_{ya} \tag{5.29}$$

where the plus and minus signs correspond to those in the exponential factor.

*N is sometimes expressed as $N = n - i\kappa$, coupled with a plane wave of the form $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \exp[i(\omega t - \mathbf{k} \bullet \mathbf{r})]$ rather than eq. (5.20). As long as they are consistent, the end results will be the same.

Now let us see how we can calculate the energy flow associated with electromagnetic fields. We start by manipulating the Maxwell equations. Taking the dot product of $\mathbf{H}$ with eq. (5.11) and the dot product of $\mathbf{E}$ with eq. (5.12), then, subtracting the resulting two equations, we get

$$-\nabla \bullet (\mathbf{E} \times \mathbf{H}) = \frac{\partial}{\partial t}\left(\frac{1}{2}\mu\mathbf{H} \bullet \mathbf{H} + \frac{1}{2}\varepsilon\mathbf{E} \bullet \mathbf{E}\right) + \mathbf{E} \bullet \mathbf{J}_e \tag{5.30}$$

where we have used the vector identity, $\nabla \bullet (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \bullet (\nabla \times \mathbf{E}) - \mathbf{E} \bullet (\nabla \times \mathbf{H})$ We identify the meaning of each term on the right-hand side of eq. (5.30) as

Magnetic field energy density [J m^{-3}]: $\frac{1}{2}\mu\mathbf{H} \bullet \mathbf{H}$ (5.31)

Electric field energy density [J m^{-3}]: $\frac{1}{2}\varepsilon\mathbf{E} \bullet \mathbf{E}$ (5.32)

Joule heating [W m^{-3}]: $\mathbf{E} \bullet \mathbf{J}_e$ (5.33)

To see what $\mathbf{E} \times \mathbf{H}$ means in eq. (5.30), we integrate the equation over a volume,

$$-\iiint \nabla \bullet (\mathbf{E} \times \mathbf{H}) dV = \iiint \left[\frac{\partial}{\partial t}\left(\frac{1}{2}\mu\mathbf{H} \bullet \mathbf{H} + \frac{1}{2}\varepsilon\mathbf{E} \bullet \mathbf{E}\right) + \mathbf{E} \bullet \mathbf{J}_e\right] dV \tag{5.34}$$

and rewrite the left-hand side into a surface integral, using Gauss's theorem,

$$-\oiint (\mathbf{E} \times \mathbf{H}) \bullet \hat{\mathbf{n}} dA = \iiint \left[\frac{\partial}{\partial t}\left(\frac{1}{2}\mu\mathbf{H} \bullet \mathbf{H} + \frac{1}{2}\varepsilon\mathbf{E} \bullet \mathbf{E}\right) + \mathbf{E} \bullet \mathbf{J}_e\right] dV \tag{5.35}$$

where the surface integration is carried out over the surface enclosing the volume and $\hat{\mathbf{n}}$ is the local normal of the surface, pointing outward. The right-hand side of eq. (5.35) represents the rate of change of the stored electromagnetic energy inside the volume plus the heat generated. The first law of thermodynamics requires that this must be supplied by the energy flow into the volume across the boundaries, hence the left-hand side must be this energy flow. Because of the negative sign and the fact that $\hat{\mathbf{n}}$ points outward, the cross product $\mathbf{E} \times \mathbf{H}$ must mean the rate of energy flow into the volume. We call this product the *Poynting vector* $\mathbf{S}$ [W m^{-2}]

$$\boxed{\mathbf{S} = \mathbf{E} \times \mathbf{H}} \tag{5.36}$$

The Poynting vector represents the instantaneous energy flux. It oscillates at twice the frequency of the electromagnetic field. Currently, no electronic devices can measure such a fast signal. What can be measured is the time-averaged Poynting vector

$$\langle \mathbf{S} \rangle = \frac{1}{T}\int_t^{t+T} \mathbf{S} dt' \tag{5.37}$$

If the complex representation of **E** and **H** is used, as is the case most of the time, it can be shown that the time-averaged Poynting vector can be calculated from

$$\langle \mathbf{S} \rangle = \frac{1}{2}\mathrm{Re}(\mathbf{E}_c \times \mathbf{H}_c^*) \tag{5.38}$$

where the subscript c is used, only in this equation, to emphasize the complex representation of **E** and **H**, and the superscript * means the complex conjugate. Because we use the complex representation most of the time, we will drop the subscript c whenever it is clear, as we have been doing so far. The time-averaged Poynting vector expression, eq. (5.38), is similar to the particle flux expression, eq. (5.6), for quantum mechanical waves. As we move on, we will see more similarities between these waves.

As an example, we consider a plane wave propagating in the positive x-direction as given by eqs. (5.27) and (5.28). The corresponding Poynting vector is

$$\begin{aligned}
\langle \mathbf{S} \rangle &= \frac{1}{2}\mathrm{Re}(\mathbf{E}_c \times \mathbf{H}_c^*) \\
&= \frac{1}{2}\mathrm{Re}\left\{E_{ya} H_{za}^* \exp\left[-i\omega\left(t - \frac{Nx}{c_0}\right)\right] \exp\left[i\omega\left(t - \frac{N^* x}{c_0}\right)\right]\right\} (\hat{\mathbf{y}} \times \hat{\mathbf{z}}) \\
&= \frac{1}{2}\mathrm{Re}\left\{E_{ya} \frac{N^*}{\mu c_0} E_{ya}^* \exp\left[\frac{i\omega(N - N^*)_x}{c_0}\right]\right\} \hat{\mathbf{x}} \\
&= \frac{1}{2}\mathrm{Re}\left\{\frac{n - i\kappa}{\mu c_0} \exp\left[-\frac{2\omega\kappa x}{c_0}\right]\right\} E_{ya}^2 \hat{\mathbf{x}} \\
&= \frac{1}{2}\frac{n}{\mu c_0} \exp\left[-\frac{4\pi\kappa x}{\lambda_0}\right] E_{ya}^2 \hat{\mathbf{x}} \\
&= \frac{1}{2}\frac{n}{\mu c_0} e^{-\alpha x} E_{ya}^2 \hat{\mathbf{x}}
\end{aligned} \tag{5.39}$$

where λ_0 is the wavelength in vacuum and we have used eq. (5.29) to replace H_{za} by E_{ya}, and

$$\alpha = \frac{4\pi\kappa}{\lambda_0} \; [m^{-1}] \tag{5.40}$$

is called the *absorption coefficient*. Its inverse, $\delta = 1/\alpha$, is called the *skin depth*. Equation (5.39) shows that as the electromagnetic field propagates, the energy decays exponentially. The skin depth is where the energy flux has dropped by e^{-1}. With a $\kappa \sim 0.1$, the skin depth is of the order of one wavelength λ_0. For optical fibers, a low absorption coefficient is essential. For a $\delta \sim 1000$ m and $\lambda_0 = 1.55$ μm, which is the wavelength used in long-distance optical communication, κ, must be less than 10^{-10}. For metals, κ is usually large in the range from visible to far infrared, and thus electromagnetic fields usually do not penetrate far into metal.

In closing this section, we comment further on the relationship between **E**, **H**, and **k**. Equation (5.26) says that they form an orthogonal set. Usually, it is further understood that this set follows the so-called right-hand rule: with the right hand fully extended and the thumb perpendicular to the four other fingers, close the hand by turning the four fingers, which originally pointed in the **E**-field direction, toward the **H**-field direction; the thumb then points in the **k**-direction. In artificial materials with both negative ε and negative μ, however, the Maxwell equations actually require that **E**, **H**, and **k** follow the left-hand rule (Veselago, 1968). Such left-handed materials may have interesting properties such as negative refractive index arising from taking the negative root of eq. (5.21), and could focus light to a spot much smaller than the wavelength (Pendry, 2000; Shelby et al., 2001).

5.1.3 Plane Acoustic Waves

Having explored quantum mechanical material waves and electromagnetic waves in the form of plane waves, we examine in this section the appropriate form of plane lattice waves. We discussed in chapter 3 lattice waves based on a simple one-dimensional lattice chain model, treating atoms as discrete points. In the long wavelength range, we can neglect the atomic structure and treat the medium as a continuum. The lattice wave in this range can be described by acoustic wave equations. Our discussion of lattice waves in this section will be based on acoustic wave equations, which resemble strongly the electromagnetic waves dealt with in the previous section.

In the continuum representation, the acoustic wave propagation can be described in terms of the local medium *displacement*, **u**, from its equilibrium position, or, more often, the velocity **v** of this displacement,

$$\mathbf{v} = \frac{d\mathbf{u}}{dt} \tag{5.41}$$

The displacement can be related to the *strain tensor* $\overline{\overline{S}}$, where the "=" above the symbol means that it is a second-rank tensor, which can be represented by a (3×3) matrix with 9 components, S_{ij} $(i, j = 1, 2, 3)$, calculated from

$$S_{ij}(\mathbf{r}, t) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{5.42}$$

The strain can be further related to the stress tensor $\overline{\overline{\sigma}}$, which is again a second-rank tensor, through the generalized Hooke law. The force acting on any surface with a norm $\hat{\mathbf{n}}$ is $\mathbf{F} = \overline{\overline{\sigma}} \bullet \hat{\mathbf{n}}$, where the product of a tensor with a vector is carried out using the following matrix product rule:

$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} \tag{5.43}$$

Using eqs. (5.41)–(5.43) and Newton's second law of motion, a relationship between the stress tensor and displacement can be obtained and finally expressed in terms of the displacement velocity. The acoustic wave equations thus obtained in their general form are very complicated in an anisotropic medium with damping (Auld, 1990), because the constitutive relationship between the stress and the strain involves the elastic stiffness tensor and the viscosity tensor (representing damping), both are fourth-rank tensors with

81 components, although symmetry requirements render many of the matrix components to be zero. We will not list the acoustic wave equations but focus only on the special case when the medium is isotropic with no damping, for which only two constants are needed for the stiffness tensor. These two constants, denoted here as λ_L and μ_L, are called the Lame constants. In this case, the acoustic wave equations are significantly simplified. In particular, if we assume a plane wave of the form $\mathbf{v}\exp[-i(\omega t - k\hat{\mathbf{k}}\bullet\mathbf{r})]$, where k is the magnitude of the wavevector and $\hat{\mathbf{k}}$ is the unit wavevector, the acoustic wave equations lead to the following eigenvalue equation (Auld, 1990)

$$k^2\begin{bmatrix} c_{11}\hat{k}_x^2+\mu_L(1-\hat{k}_x^2) & (\lambda_L+\mu_L)\hat{k}_x\hat{k}_y & (\lambda_L+\mu_L)\hat{k}_x\hat{k}_z \\ (\lambda_L+\mu_L)\hat{k}_y\hat{k}_x & c_{11}k_y^2+\mu_L(1-\hat{k}_y^2) & (\lambda_L+\mu_L)\hat{k}_y\hat{k}_z \\ (\lambda_L+\mu_L)\hat{k}_z\hat{k}_x & (\lambda_L+\mu_L)\hat{k}_z\hat{k}_y & c_{11}\hat{k}_z^2+\mu_L(1-k_z^2)\end{bmatrix}\begin{bmatrix}v_x\\v_y\\v_z\end{bmatrix}=\rho\omega^2\begin{bmatrix}v_x\\v_y\\v_z\end{bmatrix} \tag{5.44}$$

where $c_{11}=\lambda_L+2\mu_L$. The above eigenvalue equation determines the dispersion relation between ω and $\mathbf{k}$, and is called the *Christoffel* equation.

Consider a plane transverse wave propagating along the z-direction and vibrating in the x-direction,

$$\mathbf{v}_T = A_T e^{-i(\omega t-k_T z)}\hat{\mathbf{x}} \tag{5.45}$$

Substituting eq. (5.45) into (5.44) leads to the following solution

$$\mu_L k_T^2=\rho\omega^2 \text{ or } \omega = v_T k_T \tag{5.46}$$

where $v_T=(\mu_L/\rho)^{1/2}$ is the velocity of the wave. There is an identical transverse wave vibrating in the y-direction since the medium is isotropic, as is the case when we consider phonons in three-dimensional crystals on the basis of lattice-dynamics calculations. A longitudinal wave that vibrates along the z-direction,

$$\mathbf{v}_L = A_L e^{-i(\omega t-k_L z)}\hat{\mathbf{z}} \tag{5.47}$$

also satisfies eq. (5.44). Substituting eq. (5.47) into (5.44) yields the following dispersion relation

$$k_L^2 c_{11}=\rho\omega^2 \text{ or } \omega = v_L k_L \tag{5.48}$$

where the longitudinal wave velocity $v_L=(c_{11}/\rho)^{1/2}=[(\lambda_L+2\mu_L)/\rho]^{1/2}$. Thus, unlike electromagnetic waves, which are transverse and always have the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ perpendicular to the wave propagation direction $\hat{\mathbf{k}}$, acoustic waves can exist in both transverse and longitudinal forms. Sound propagation in gases is a longitudinal wave. For a plane wave propagating along an arbitrary direction $\hat{\mathbf{k}}$, we can express the transverse and longitudinal plane waves in the general form

$$\mathbf{v}_{T1}=\hat{\mathbf{a}}A_{T1}e^{-i(\omega t-k_T\hat{\mathbf{k}}\bullet\mathbf{r})},\ (\hat{\mathbf{a}}\bullet\hat{\mathbf{k}}=0) \tag{5.49}$$

$$\mathbf{v}_{T2}=\hat{\mathbf{a}}\times\hat{\mathbf{k}}A_{T2}e^{-i(\omega t-k_T\hat{\mathbf{k}}\bullet\mathbf{r})} \tag{5.50}$$

$$\mathbf{v}_L=\hat{\mathbf{k}}A_L e^{-i(\omega t-k_L\hat{\mathbf{k}}\bullet\mathbf{r})} \tag{5.51}$$

where eq. (5.49) defines a transverse acoustic wave with displacement direction $\hat{\mathbf{a}}$ perpendicular to $\hat{\mathbf{k}}$, and eq. (5.50) denotes another transverse acoustic wave with displacement normal to both $\hat{\mathbf{a}}$ and $\hat{\mathbf{k}}$. Based on the displacement velocity, components of the stress tensor can be calculated from

$$\frac{\partial}{\partial t}\begin{bmatrix}\sigma_{xx}\\\sigma_{yy}\\\sigma_{zz}\\\sigma_{yz}\\\sigma_{xz}\\\sigma_{xy}\end{bmatrix}=\begin{bmatrix}c_{11}&c_{12}&c_{12}&0&0&0\\c_{12}&c_{11}&c_{12}&0&0&0\\c_{12}&c_{12}&c_{11}&0&0&0\\0&0&0&c_{44}&0&0\\0&0&0&0&c_{44}&0\\0&0&0&0&0&c_{44}\end{bmatrix}\begin{bmatrix}\partial v_x/\partial x\\\partial v_y/\partial y\\\partial v_z/\partial z\\\partial v_y/\partial z+\partial v_z/\partial y\\\partial v_x/\partial z+\partial v_z/\partial x\\\partial v_x/\partial y+\partial v_y/\partial x\end{bmatrix} \tag{5.52}$$

where $c_{12}=\lambda_L$ and $c_{44}=\mu_L$. Equation (5.52) is a form of the Hooke law for an isotropic medium without damping. The symmetry relations $\sigma_{xy}=\sigma_{yx}$, $\sigma_{yz}=\sigma_{zy}$, and $\sigma_{xz}=\sigma_{zx}$ can be used to obtain all nine components of the second-order stress tensor.

The time-averaged power carried by the acoustic wave can be calculated from the *acoustic Poynting vector* [W m^{-2}]

$$\mathbf{J}_a=-\frac{1}{2}\mathrm{Re}(\mathbf{v}^*\bullet\bar{\bar{\sigma}}) \tag{5.53}$$

which is again similar to that of an electromagnetic wave [eq. (5.38)] and of quantum mechanical material waves [eq. (5.6)].

The above discussion on acoustic waves and electromagnetic waves is clearly very sketchy and also mathematically involved. The main purpose is to get the reader familiar with the plane waves propagating along an arbitrary wavevector $\mathbf{k}$ direction, as represented by eq. (5.3) and the flux carried by the plane waves. In the next section, we will examine how these plane waves behave at an interface.

5.2 Interface Reflection and Refraction of a Plane Wave

When a wave, be it an electron, photon, or phonon wave, meets a boundary, it will be reflected and refracted. In this section, we will determine how much of the incoming wave is reflected and how much is refracted, by imposing boundary conditions for these waves. The reader will find that although the electron, photon, and phonon waves are quite different in nature, the expressions for their respective interface reflectivity and transmissivity, which we will define soon, are similar.

5.2.1 Electron Waves

At an interface between two materials, an electrical potential generally exists as shown in figure 3.28. We examine the transport of a plane electron wave across an interface represented by a step potential as shown in figure 5.1. Although the more general case of a wave with any arbitrary angle of incidence can be treated, we focus only on the normal incidence case, that is, when the wave is traveling along the z-direction.

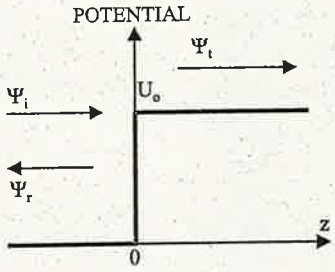


Figure 5.1 Reflection and transmission of an electron wave at an interface, caused by the potential barrier at the interface.

The wavefunctions of the incident, Ψ_i, the reflected, Ψ_r, and the transmitted waves Ψ_t, can be expressed as

$$\Psi_i = A_i e^{-i(\omega t - k_1 z)},\ \Psi_r = A_r e^{-i(\omega t + k_1 z)},\ \Psi_t = A_t e^{-i(\omega t - k_2 z)} \tag{5.54}$$

where k_1 and k_2 are the electron wavevectors in the two media, respectively, and

$$\omega = \frac{E}{\hbar},\ k_1 = \sqrt{\frac{2mE}{\hbar^2}},\ k_2 = \sqrt{\frac{2m\,(E - U_0)}{\hbar^2}} \tag{5.55}$$

These wavefunctions have been obtained in sections 2.3.1 and 3.2.1. Using the boundary conditions on the continuity of the wavefunction and its first-order derivative, similar to eqs. (3.25) and (3.26), we obtain the *reflection* and *transmission coefficients*

$$r = \frac{A_r}{A_i} = \frac{k_1 - k_2}{k_1 + k_2} \text{ and } t = \frac{A_t}{A_i} = \frac{2k_1}{k_1 + k_2} \tag{5.56}$$

The corresponding *reflectivity* and *transmissivity* of the particle flux are

$$\left.\begin{aligned} R &= \frac{J_r}{J_i} = \left|\frac{k_1 - k_2}{k_1 + k_2}\right|^2 \\ \tau &= \frac{J_t}{J_i} = \frac{\mathrm{Re}\left[k_2^* A_t A_t^*\right]}{\mathrm{Re}\left[k_1^* A_i A_i^*\right]} = \frac{4\,\mathrm{Re}(k_1 k_2^*)}{|k_1 + k_2|^2} = 1 - R \end{aligned}\right\} \tag{5.57}$$

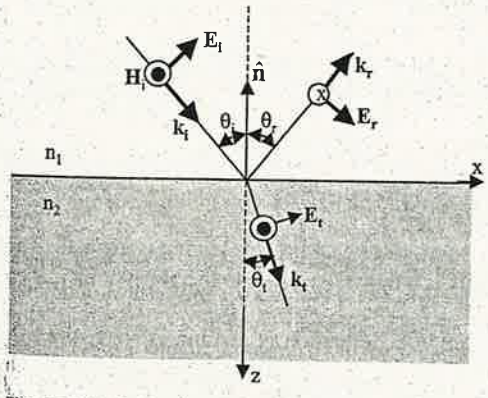
where the incident, reflected, and transmitted particle fluxes (J_i, J_r, and J_t) are calculated from eq. (5.6).

When $E < U_0$, it can be shown that $R = 1$ and $\tau = 0$, which means that all electrons are reflected (total reflection). In this case, however, the wavefunction Ψ_t is not zero, as one can easily show by substituting k_2 from eq. (5.55), which gives an imaginary k_2 when $E < U_0$, into eq. (5.54). The wavefunction Ψ_t thus obtained decays exponentially from the interface but is nonetheless not zero. This transmitted wave, however, does not carry a net particle flux across the boundary ($J_t = 0$), and it is called an *evanescent wave*, which will be discussed in section 5.4.

When $E > U_0$, electrons will be partially reflected and partially transmitted, whereas classical mechanics would lead to 100% transmission without reflection. Making a mundane daily analogy, this means that if one throws a stone at a wall, there is some probability that the stone will be reflected back even if it is thrown higher than the wall. Such a scenario is clearly contrary to our common experience but is possible in the quantum world for very small particles such as electrons. In fact, the reflection phenomenon occurs for all waves, as we will see later. Thus, wave mechanics differs significantly from particle mechanics at a single interface.

5.2.2 Electromagnetic Waves

The reflection and refraction of light at an interface is a more familiar process for many readers, and shares many similarities to the behavior of electron waves at an interface. Although we considered only the case of normal incidence for electron waves, we will treat here the more general case of oblique incidence of an electromagnetic wave onto an interface. As shown in figure 5.2, a plane electromagnetic wave propagates along direction $\mathbf{k}_i$ (wavevector direction) and meets an interface with norm $\hat{\mathbf{n}}$. The reflected wave and refracted wave propagate along the $\mathbf{k}_r$ and $\mathbf{k}_t$ directions, respectively. We call the plane formed by $\mathbf{k}_i$ and $\hat{\mathbf{n}}$ the plane of incidence, and the angle formed between $\hat{\mathbf{n}}$ and $\mathbf{k}_i$ the angle of incidence. The electromagnetic wave is a transverse wave, so the **E**-field and the **H**-field can have any orientation in the plane normal to **k**. We can decompose the electric field into two components, one parallel to the plane of incidence and the other perpendicular to the interface. When an electric field is parallel to the plane of incidence, as is the case of figure 5.2, it cannot be parallel to the interface unless the angle of incidence is zero. Its conjugate magnetic field component, in this case pointing out of the paper, is perpendicular to the plane of incidence and is thus always parallel to the interface. This wave is called a transverse magnetic wave, or TM-polarized wave. Sometimes the notations p (parallel polarized) and $/\!/$ (relative to E) are also used. If the electrical field component is perpendicular to the plane of incidence, the wave is called a transverse electric wave or TE-polarized. Notations TE, s (perpendicularly polarized), and $\perp$ are often used interchangeably. We will limit our discussion to positive media, that is, those with refractive indices of the form of eq. (5.24) for which **E**, **H**, and **k** obey the right-hand rule.



Symbol Convention:

- ◉ Field Going Out of Paper
- ⊗ Field Going Into Paper

E-Field In the Plane of Incidence:

TM Wave = // Wave = p Wave

H-Field In the Plane of Incidence:

TE Wave = ⊥ Wave = s Wave

Figure 5.2 Reflection and refraction of an electromagnetic field at an interface.

We need to establish boundary conditions for the electric and magnetic fields to determine the reflection and transmission at the interface. By applying the Maxwell equations to a very thin control volume surrounding an interface, the following *boundary conditions* can be obtained (Born and Wolf, 1980)

$$\hat{\mathbf{n}} \bullet (\mathbf{D}_2 - \mathbf{D}_1) = \rho_s \tag{5.58}$$

$$\hat{\mathbf{n}} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \tag{5.59}$$

$$\hat{\mathbf{n}} \bullet (\mathbf{B}_2 - \mathbf{B}_1) = 0 \tag{5.60}$$

$$\hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = J_s \tag{5.61}$$

where ρ_s [$\mathrm{C\,m^{-2}}$] and $\mathbf{J}_s$ [$\mathrm{A\,m^{-1}}$] are the net surface charge density and the surface current density, respectively, $\mathbf{E}_1$ and $\mathbf{E}_2$ are the total electric fields on the two sides of the interface, and similarly $\mathbf{H}_1$ and $\mathbf{H}_2$ are the total magnetic fields on the two sides of the interface. To obtain the "total" $\mathbf{E}$ and $\mathbf{H}$ for side 1 of figure 5.2, we need to sum up the incident and the reflected fields. Equation (5.58) means that the difference between the normal components of the electric displacements across the interface must be equal to interface charge density, while eq. (5.59) means that the tangential components of the electric field must be continuous. Equation (5.60) says that the normal components of the magnetic induction must be continuous, while eq. (5.61) means that the difference of the tangential components of the magnetic field across a surface equals the surface current density.

With the above boundary conditions, we can determine the amount of reflection and transmission of an incident electromagnetic wave onto a surface. We consider a plane TM wave incident onto a surface at an incident angle θ_i. The wavevector directions of the incident, reflected, and transmitted waves are $(\sin\theta_i,\ 0,\ \cos\theta_i)$, $(\sin\theta_r,\ 0,\ -\cos\theta_r)$, and $(\sin\theta_t,\ 0,\ \cos\theta_t)$, respectively. Using a plane wave of the form of eq. (5.25), the incident, reflected, and transmitted electric fields can be expressed as

$$\mathbf{E}_{//i} \exp\left[-i\omega\left(t - \frac{n_1 x \sin\theta_i + n_1 z \cos\theta_i}{c_0}\right)\right] \tag{5.62}$$

$$\mathbf{E}_{//r} \exp\left[-i\omega\left(t - \frac{n_1 x \sin\theta_r - n_1 z \cos\theta_r}{c_0}\right)\right] \tag{5.63}$$

$$\mathbf{E}_{//t} \exp\left[-i\omega\left(t - \frac{n_2 x \sin\theta_t - n_2 z \cos\theta_t}{c_0}\right)\right] \tag{5.64}$$

respectively. Here, we temporarily assume that the refractive indices are real. The subscript "//" means that the electric field is polarized parallel to the plane of incidence (TM wave as shown in figure 5.2).

Some readers may ask how to determine the direction of $\mathbf{E}_r$ and $\mathbf{H}_r$ in figure 5.2. The answer is that a correct assumption of the direction is not important as long as both $\mathbf{E}_r$ and $\mathbf{H}_r$ follow the right-hand rule. The signs in the final results will take care of the directions. Notice the sign change in eq. (5.63) before z in the exponent due to the change in the wave propagation direction for the reflected wave. Based on a similar derivation of eq. (5.29), the magnitude of the magnetic field, which is pointing out of the plane of the paper, is related to the electric field by

$$H_y = \frac{n}{\mu c_0} E_{//}(\text{forward}), \quad H_y = -\frac{n}{\mu c_0} E_{//}(\text{backward}) \tag{5.65}$$

where the "forward" denotes waves propagating along the positive z-direction (incident and refracted waves) and "backward" applies to the reflected wave propagating along the negative z-direction, the subscript "y" of H denotes that H points perpendicular to the plane of incidence, in the y-direction, and $E_{//}$ is the magnitude of vector $\mathbf{E}_{//}$. In figure 5.2 we show the reflected magnetic fields pointing into the paper because of the negative sign in eq. (5.65) for the reflected wave. In reality, the actual sign change may be in the reflected electric field rather than the magnetic field. As long as one is consistent with the mathematical operations, the end result will give the correct sign.

We consider a surface free of net charge and current, and take this surface as $z = 0$. To determine the magnitudes of the reflected and refracted fields, we need consider only the continuity of the tangential components. The boundary conditions on the normal components will be automatically satisfied. Applying the continuity of the electric field, and noting that the electric field of a TM wave is not perpendicular to the surface, we use the component along the x-direction,

$$\cos\theta_i E_{//i} \exp\left[i\omega\frac{n_1 x \sin\theta_i}{c_0}\right] + \cos\theta_r E_{//r} \exp\left[i\omega\frac{n_1 x \sin\theta_r}{c_0}\right]$$
$$= \cos\theta_t E_{//t} \exp\left[i\omega\frac{n_2 x \sin\theta_t}{c_0}\right] \tag{5.66}$$

where we have dropped the factor of $e^{-i\omega t}$ since it is contained in all the terms and cancels out. Since x can take any value, the above equation is valid only when the exponents are equal. This gives

$$n_1 \sin\theta_i = n_1 \sin\theta_r = n_2 \sin\theta_t \tag{5.67}$$

which leads to the *Snell law* for reflection and refraction

$$\theta_i = \theta_r \text{ and } n_1 \sin\theta_i = n_2 \sin\theta_t \tag{5.68}$$

Substituting eqs. (5.67) and (5.68) back into eq. (5.66) leads to

$$\cos\theta_i E_{//i} + \cos\theta_i E_{//r} = \cos\theta_t E_{//t} \tag{5.69}$$

which gives one equation relating $E_{//i}$, $E_{//r}$, and $E_{//t}$. Another relation can be obtained on the basis of the continuity of the tangential components of the magnetic field at the interface, since there is no surface current. On the basis of eq. (5.65), we can write the continuity of the tangential component of the magnetic field, eq. (5.61), as

$$n_1 E_{//i} - n_1 E_{//r} = n_2 E_{//t} \tag{5.70}$$

Solving eqs. (5.69) and (5.70), we obtain the reflection coefficient $r_{//}$ and transmission coefficient $t_{//}$ for a TM wave as

$$r_{//} = \frac{E_{//r}}{E_{//i}} = \frac{-n_2 \cos\theta_i + n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} \tag{5.71}$$

$$t_{//} = \frac{E_{//t}}{E_{//i}} = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t} \tag{5.72}$$

Similarly, for a TE wave,

$$r_{\perp} = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} \tag{5.73}$$

$$t_{\perp} = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_t} \tag{5.74}$$

Equations (5.71)–(5.74) are called the *Fresnel coefficients* of reflection and transmission.

The above discussion assumes real refractive indices for both media. If any of the media is absorbing, one can prove that the expressions for the Snell law and the Fresnel coefficients remain the same except that we should replace n by the complex refractive index N. This statement can raise questions. For example, the Snell law becomes $n_1 \sin\theta_i = N_2 \sin\theta_t$. If N_2 is complex, the angle of refraction θ_t is also complex. What does a complex angle mean? To answer this question, one can substitute the complex angle into the transmitted wave expression, eq. (5.64), and see that in this case the constant amplitude surface does not coincide with the constant phase surface. Such a wave with different constant amplitude and phase surfaces is called an inhomogeneous wave. The proof is left as an exercise.

The Fresnel coefficients give the magnitudes of reflected and transmitted fields. To calculate the energy flux going across the interface, we need to examine the Poynting vector. For a TM wave, we have

$$\begin{aligned}\langle \mathbf{S} \rangle &= \frac{1}{2}\mathrm{Re}(\mathbf{E} \times \mathbf{H}^*) = -\frac{1}{2}\mathrm{Re}(E_z H_y^*)\hat{\mathbf{x}} + \frac{1}{2}\mathrm{Re}(E_x H_y^*)\hat{\mathbf{z}} \\ &= \frac{n}{2\mu c_0} E_{//}^2 \sin\theta\, \hat{\mathbf{x}} + \frac{n}{2\mu c_0} E_{//}^2 \cos\theta\, \hat{\mathbf{z}} \\ &= \langle S_x \rangle \hat{\mathbf{x}} + \langle S_z \rangle \hat{\mathbf{z}}\end{aligned}$$

Since only the component in the z-direction goes across the interface, we define the reflectivity and transmissivity, based on the power arriving per unit area normal to the interface, $S_{i,z}$, as

$$\left.\begin{aligned}\text{Reflectivity:}\quad R_{//} &= \frac{S_{//r,z}}{S_{//i,z}} = \frac{S_{//r}}{S_{//i}} = |r_{//}|^2 \\ R_{\perp} &= |r_{\perp}|^2\end{aligned}\right\} \tag{5.75}$$

$$\left.\begin{aligned}\text{Transmissivity:}\quad \tau_{//} &= \frac{S_{//t,z}}{S_{//i,z}} = \frac{S_{//t}}{S_{//i}} = \frac{\mathrm{Re}(N_2^* \cos\theta_t)}{\mathrm{Re}(N_1^* \cos\theta_i)} |t_{//}|^2 \\ \tau_{\perp} &= \frac{\mathrm{Re}(N_2 \cos\theta_t)}{\mathrm{Re}(N_1 \cos\theta_i)} |t_{\perp}|^2\end{aligned}\right\} \tag{5.76}$$

Equations (5.75) and (5.76) apply to the cases when either or both of the two media are absorbing. When medium 1 is non-absorbing, it can be shown that

$$R + \tau = 1 \tag{5.77}$$

for both TM and TE waves. However, when medium 1 is also absorbing, it can be shown that the above intuitive expression is no longer valid. This is because of the interference of the incident and reflected waves (Knittl, 1976).

At normal incidence, the reflectivity can be simplified to

$$R = R_{//} = R_{\perp} = \left|\frac{k_1 - k_2}{k_1 + k_2}\right|^2 = \left|\frac{N_2 - n_1}{N_2 + n_1}\right|^2 = \frac{(n_2 - n_1)^2 + (\kappa_2 - \kappa_1)^2}{(n_2 + n_1)^2 + (\kappa_2 + \kappa_1)^2} \tag{5.78}$$

which is identical to eq. (5.57) for the reflectivity of an electron wave. For an air/glass interface, where $n = 1$ for air and $n \approx 1.45$ for glass between $\lambda = 0.5$ and $0.6\ \mu\mathrm{m}$, the reflectivity at normal incidence is ~3.4%. For an air/silicon interface (n for Si ≈ 4 between $\lambda = 0.5$ and $0.6\ \mu\mathrm{m}$), the reflectivity is ~ 36% at normal incidence.

In figure 5.3, we show the reflectivity for typical dielectric and metallic materials as a function of the angle of incidence. The figure shows that the reflectivity depends on the polarization of the incident radiation. The reflectivity for a TM wave can be zero for dielectric materials. From eq. (5.71), this happens when the numerator equals zero, that

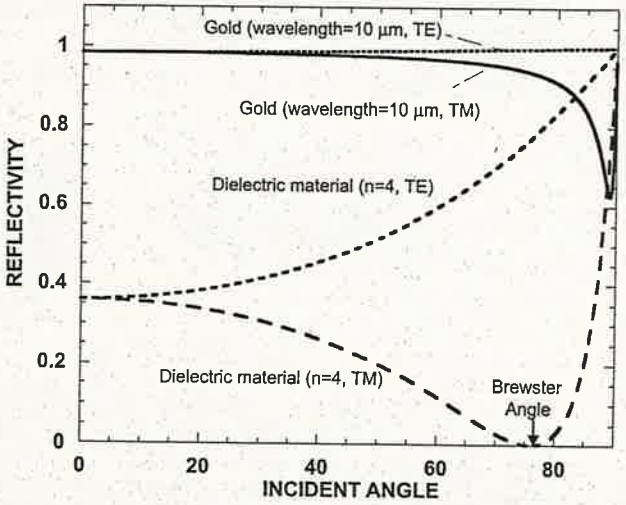


Figure 5.3 Reflectivity as a functionof the angle of incidence for a dielectric material with $n = 4$ and for gold with $N = 10.8 + i51.6$.

is, when $\sin(2\theta_i) = \sin(2\theta_t)$ or $\theta_t = \pi/2 - \theta_i$. Substituting this condition into the Snell law, eq. (5.67), leads to the following incidence angle, θ_B, at which the reflectivity of the TM mode is zero,

$$\tan\theta_B = n_2/n_1 \tag{5.79}$$

This angle is called the *Brewster angle*. At this angle, only TE waves are reflected. This phenomenon can be exploited to control the polarization of white light and is also the cause of shiny dark (seemingly wet) surfaces on the freeway on a sunny day.

Another interesting situation is when the refractive index of medium 1 (incident side) is larger than that of medium 2. Because the maximum angle of the refracted wave is $\theta_t = 90^\circ$, there exists an angle of incidence above which no real solution for θ_t exists. This *critical angle* happens when, according to the Snell law,

$$n_1 \sin\theta_c = n_2 \sin 90^\circ \text{ or } \theta_c = \arcsin(n_2/n_1) \tag{5.80}$$

Above this angle, the reflectivity equals one; that is, all the incident energy is reflected. This is called total internal reflection and is the basis of waveguides that confine the photon waves laterally, as in an optical fiber and a semiconductor laser. An optical fiber has a core region and a cladding layer [figure 5.4(a)]. The refractive index in the core region is higher than in the cladding layer. If light is launched into the fiber at an angle of incidence (relative to the core/cladding interface) larger than the critical angle, the light will be bounced inside the core only without leakage, thus traveling a long distance along the fiber if the absorption coefficient of the core is small. However, if the angle of incidence is smaller than the critical angle, the light can escape the fiber core. In a semiconductor laser, light is emitted through electron–hole recombination inside the active region. The emitted light spreads over the core region and is confined by cladding layers that have a lower refractive index than the core [figure 5.4(b)]. We mention here, however, that even though the reflectivity is one and transmissivity is zero for a wave incident above the critical angle, there is still a nonzero electromagnetic

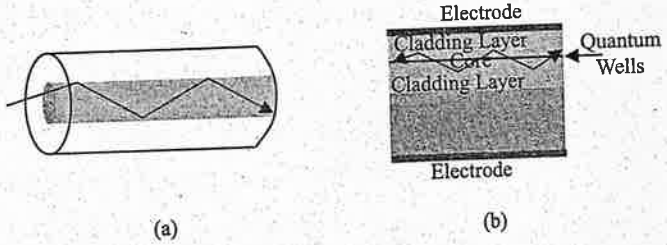


Figure 5.4 Examples of waveguides based on total internal reflection. (a) An optical fiber is made of a core (a few microns wide) and a cladding layer in concentric cylindrical configuration. (b) An edge-emitting semiconductor layer has a planar geometry. Light is emitted inside the quantum wells (a few tens of angstroms) through electron–hole recombination. The emitted light is confined inside the core (~ microns in total thickness) by total internal reflection effects. Some of the emitted light comes out from the edges because the side surfaces (mirrors or facets) has a nonzero transmissivity.

wave in medium 2. This is called the *evanescent wave*, similar to the evanescent electron wave we mentioned in the previous section. We will discuss the evanescent wave in more detail in section 5.4.

Example 5.1

A 0.5 W laser with a beam diameter of 1 mm and wavelength 0.5 μm is directed at normal incidence to a piece of aluminum. The complex refractive index of aluminum at 0.5 μm is $0.769 + i6.08$. Determine the distribution of heat generation inside the aluminum.

Solution: From eq. (5.39), we know that the Poynting vector, thus the intensity of the laser beam, decreases exponentially inside the film. The distribution of the laser intensity can be expressed as

$$I = (1-R)I_i e^{-\alpha x}\ [\mathrm{W\,m^{-2}}] \tag{E5.1.1}$$

where x is the coordinate perpendicular to the surface, R is the reflectivity, and α the absorption coefficient. The decrease in intensity is converted into heat. So, the heat generation distribution is

$$\dot{q} = -\frac{dI}{dx} = (1-R)\alpha I_i e^{-\alpha x}\ [\mathrm{W\,m^{-3}}] \tag{E5.1.2}$$

We can calculate R and α as

$$R = \left|\frac{1-N}{1+N}\right|^2 = \left|\frac{0.231 - 6.08i}{1.769 + 6.08i}\right|^2 = 0.923 \tag{E5.1.3}$$

$$\alpha = \frac{4\pi\kappa}{\lambda} = \frac{4\pi \times 6.08}{0.5 \times 10^{-6}\ \mathrm{m}} = 1.53 \times 10^8\ \mathrm{m^{-1}} \tag{E5.1.4}$$

Substituting these values and $I_i = 0.5\ \mathrm{W}/(\pi \times 0.001^2/4) = 6.34 \times 10^5\ \mathrm{W\,m^{-2}}$ into the heat generation distribution expression leads to the final answer as

$$q = 7.5 \times 10^{12}\, e^{-\alpha x}\ [\mathrm{W\,m^{-3}}] \tag{E5.1.5}$$

Comment. Because α is very large, heat absorption occurs only in the region near the surface.

With the expression for the surface reflectivity and transmissivity, the emissivity of the surface at the same wavelength can be readily calculated. If medium 2 is semi-infinite, the energy transmitted into the medium will eventually be absorbed and thus the absorptivity equals the transmissivity at the interface. From Kirchoff's law, the emissivity equals the absorptivity at the same incident direction and the same wavelength.

5.2.3 Acoustic Waves

The reflection and refraction of acoustic waves can be treated similarly to the electromagnetic waves. The continuity requirements for acoustic waves are that the displacement velocity and the force at the interface must be continuous:

$$\sum \mathbf{v}_1 = \sum \mathbf{v}_2 \text{ and } \sum \bar{\bar{\sigma}}_1 \bullet \hat{\mathbf{n}} = \sum \bar{\bar{\sigma}}_2 \bullet \hat{\mathbf{n}} \tag{5.81}$$

where $\hat{\mathbf{n}}$ is the norm of the interface and $\bar{\bar{\sigma}} \bullet \hat{\mathbf{n}}$ is the force acting on the interface, as expressed by eq. (5.43), and the summation is over all the fields (for example, the incident and reflected) on each side. While continuity of force is always required, even in a lattice dynamics simulation, continuity of displacement velocity is not always true at the atomic scale. For example, the two atoms at the two sides of an interface can have different displacement velocities. In the long wavelength limit, however, the continuum assumption is reasonable and eq. (5.81) is valid.

Using the boundary conditions and plane acoustic waves of the form of eqs. (5.49)–(5.51), one can again derive expressions for the reflectivity and transmissivity of acoustic waves at an interface as for electromagnetic and electron waves. The derivation, however, is more complex because acoustic waves have three polarizations and one must consider the possibility of coupling among these polarizations—for example, whether a longitudinal wave can excite a transverse component in the reflected and transmitted waves (Auld, 1990). The simplest case is when the medium is isotropic and the incident wave is a transverse wave with displacement polarized in the direction perpendicular to the plane of incidence (called a horizontally polarized shear wave or SH wave). In this case, only one SH reflected wave and one SH transmitted wave are excited. From the boundary conditions and following a similar procedure as for optical waves, one can derive the reflection and transmission coefficients for an SH wave as

$$\left.\begin{aligned} r_s &= \frac{v_r}{v_i} = \frac{Z_1 \cos\theta_i - Z_2 \cos\theta_t}{Z_1 \cos\theta_i + Z_2 \cos\theta_t} \\ t_s &= \frac{v_t}{v_i} = \frac{2Z_1 \cos\theta_i}{Z_1 \cos\theta_i + Z_2 \cos\theta_t} \end{aligned}\right\} \tag{5.82}$$

where $Z = (\rho c_{44})^{1/2} = \rho v_T$ is the *acoustic impedance*, which plays a similar role to the optical refractive index. These equations are identical to the Fresnel coefficients of a TE wave [eqs. (5.73) and (5.74)] with the acoustic impedances replacing the refractive indices. The relation between the incident angle θ_i and the refraction angle θ_t is given by the Snell law, which assumes the following form for an SH incident wave:

$$\frac{\sin\theta_i}{v_{T1}} = \frac{\sin\theta_t}{v_{T2}} \tag{5.83}$$

On the basis of the reflection and tranmission coefficients and the acoustic Poynting vector, eq. (5.53), one can calculate the energy reflectivity and transmissivity for acoustic waves. At normal incidence, the acoustic reflectivity for an SH wave is

$$R_s = \left|\frac{Z_1 - Z_2}{Z_1 + Z_2}\right|^2 \tag{5.84}$$

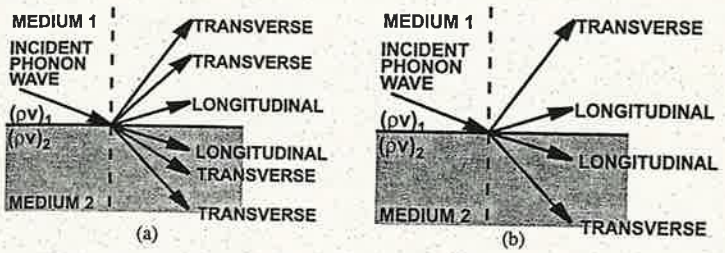


Figure 5.5 Reflection and refraction of acoustic waves in (a) anisotropic and (b) isotropic media, for longitudinally polarized waves (L waves) or transversely polarized waves with the displacement vector in the plane of incidence (vertically polarized shearwave or SV waves).

For a transverse wave polarized in the plane of incidence (vertically polarized shear wave or SV wave) and for a longitudinally polarized incident wave (L wave), coupling of different polarizations can occur. In general, an incident wave can excite three reflected waves and three transmitted waves, as shown in figure 5.5(a). For isotropic media, the two transverse waves are degenerate, so the picture reduces to 5.5(b). Correspondingly, the Snell law looks slightly different,

$$\frac{\sin\theta_i}{v_i} = \frac{\sin\theta_{rL}}{v_{L1}} = \frac{\sin\theta_{rT}}{v_{T1}} = \frac{\sin\theta_{tL}}{v_{L2}} = \frac{\sin\theta_{tT}}{v_{T2}} \tag{5.85}$$

where subscripts r and t represent the reflected and transmitted components, and L and T represent longitudinal (L) and transverse modes (SH and SV).

For isotropic media, the reflection and transmission coefficients can be found by solving a 4×4 matrix equation (Auld, 1990). An example of phonon reflectivity and transmissivity at an interface between two isotropic media with acoustic properties similar to those of Si and Ge is shown in figure 5.6 (Chen, 1999). It shows

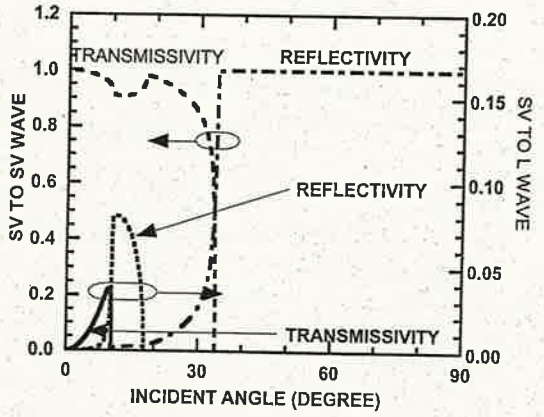


Figure 5.6 Phonon transmissivity and reflectivity at an interface similar to Si and Ge for an acoustic wave (incident from the Ge side) with the displacement vector polarized in the plane of incidence (Chen, 1999), that is, an SV incident wave. Note that L waves are also excited.

that part of the transverse incident wave is converted into a longitudinal wave. This is called mode conversion. Total reflection of acoustic waves can also occur and several critical angles can exist, as one can infer from eq. (5.85), for refracted waves of different polarizations. In figure 5.6, the total reflection of an SV wave from a germanium-like medium into a silicon-like medium occurs at 33°. Above this angle, an evanescent wave exists in the silicon side, similar to the evanescent electron and photon waves.

5.2.4 Thermal Boundary Resistance

The reflection of waves reduces the number of forward-going carriers (electrons, phonons, and photons) and is thus a source of resistance to the carrier flow. For the current flow, electron interface reflection creates an additional electrical resistance. For photons, highly reflective materials are used as radiation shields for thermal insulation, as exemplified in multilayer thermal insulation materials (Tien and Cunnington, 1973). The reflection of acoustic waves results in a resistance to heat flow, called the thermal boundary resistance. This is well known in cryogenics, where thermal boundary resistance was first discovered to exist between liquid helium and solid walls and is called the Kapitza resistance (Kapitza, 1941; Swartz and Pohl, 1989). The same phenomenon also occurs for two solid interfaces and was first treated theoretically by Little (1959).

We take figure 5.7(a) as a model system, where a heat current flows across the interface from medium 1 into medium 2, the media being maintained at two different temperatures. We neglect the temperature drop inside each material due to phonon scattering, but will come back to this issue in chapter 7. We first count the phonon heat flux going from medium 1 into medium 2, $q_{1\to2}$ [W m^{-2}], which can be written as

$$q_{1\to2} = \sum_{p=1}^{3m}\left[\frac{1}{V_1}\sum_{k_x=-k_{\max}}^{k_{\max}}\sum_{k_y=-k_{\max}}^{k_{\max}}\sum_{k_z=0}^{k_{\max}} v_{z1}\hbar\omega\tau_{12}f(\omega,T_{e1})\right] \tag{5.86}$$

where T_{e1} represents the temperature of the phonons coming toward the interface and $f(\omega, T_{e1})$ is the Bose–Einstein distribution for phonons at T_{e1}, and τ_{12} is the phonon transmissivity from medium 1 into medium 2. The summations are over all the wavevectors of k_x, k_y, and positive wavevector of k_z, so that only phonons coming toward the interface are counted. The maximum wavevector is π/a; that is, the boundaries of the first Brillouin zone. Equation (5.86) is a mixture of the phonon concept we learnt in chapter 3 and the acoustic wave propagation discussed in this chapter. The transmissivity can, for example, be calculated on the basis of the discussion in the previous section. The term $\hbar\omega f(\omega, T_{e1})$ inside eq. (5.86) represents the average energy per quantum state and the summation over the wavevectors denotes the allowable quantum states of medium 1 within a volume V_1. The division by the volume is to obtain the energy density per unit volume. These phonons are propagating toward the interface at a speed v_{z1}—the phonon velocity along the z-direction. The summation over p denotes the phonon polarizations, that is, $3m$ if there are m atoms per basis. To carry out the summation over wavevectors

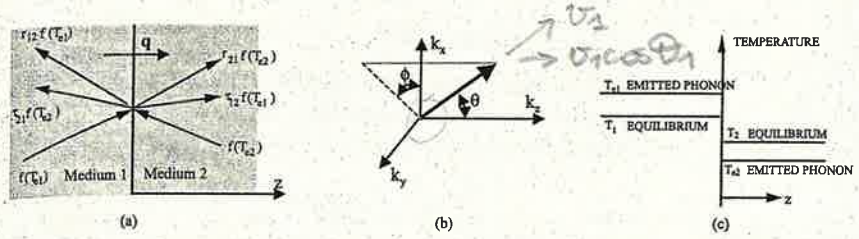


Figure 5.7 (a) Transmission and reflection of phonons at an interface. (b) Coordinate system for thermal boundary resistance evaluation. (c) Difference between the emitted phonon temperature and the local equivalent equilibrium temperature.

in eq. (5.86), we will convert it into an integration using the differential density-of-states concept outlined in section 3.4.4. For an isotropic medium, we have

$$q_{1\to2} = \iint_{\Omega_1>2\pi}\left[\int_0^{\omega_{D1}} \tau_{12}v_{z1}\hbar\omega f(\omega,T_{e1})\frac{D_1(\omega)}{4\pi}d\omega\right]d\Omega_1$$
$$= \frac{1}{4\pi}\int_0^{2\pi}d\phi_1\int_0^{\pi/2}\sin\theta_1 d\theta_1\left[\int_0^{\omega_{D1}} v_1\cos\theta_1\tau_{12}(\omega,\phi_1,\theta_1)\hbar\omega f(\omega,T_{e1})D_1(\omega)d\omega\right] \tag{5.87}$$

where ω_{D1} is the Debye frequency of medium 1 and the angles are denoted in figure 5.7(b). We use the notation $\Omega_1 > 2\pi$ to represent the half-space solid angle toward the interface from medium 1. We should point out, however, that the Debye assumption is not really necessary for the derivation. Similarly, there exists a phonon heat flux from medium 2 to medium 1. The net heat flux from material 1 into material 2 is the difference of the two,

$$q = \iint_{\Omega_1\geq2\pi}\left[\int v_1\cos\theta_1\hbar\omega f(\omega,T_{e1})\tau_{12}(\omega,\phi_1,\theta_1)D_1(\omega)/4\pi d\omega\right]d\Omega_1$$
$$-\iint_{\Omega_2\leq2\pi}\left[\int v_2\cos\theta_2\hbar\omega f(\omega,T_{e2})\tau_{21}(\omega,\phi_2,\theta_2)D_2(\omega)/4\pi d\omega\right]d\Omega_2 \tag{5.88}$$

where we have suppressed the integration limits and used $d\Omega = \sin\theta\, d\theta\, d\phi$ for simplicity in notation; details can be worked out by following eq. (5.87). At thermal equilibrium, that is, when $T_{e1} = T_{e2}$, the net heat flux is zero, so that

$$\int_{\Omega_1\geq2\pi}\left[\int v_1\cos\theta_1\hbar\omega f(\omega,T_{e1})\tau_{12}(\omega,\phi_1,\theta_1)D_1(\omega)d\omega/4\pi\right]d\Omega_1$$
$$= \int_{\Omega_2\leq2\pi}\left[\int v_2\cos\theta_2\hbar\omega f(\omega,T_{e1})\tau_{21}(\omega,\phi_2,\theta_2)D_2(\omega)d\omega/4\pi\right]d\Omega_2 \tag{5.89}$$

Equation (5.89) is an example of the *principle of detailed balance*, which requires that no net flux of any kind (heat, particle, or charge) crosses an interface if the system is at equilibrium. Using eq. (5.89), we can write eq. (5.88) as

$$q = \int_{\Omega_1 \geq 2\pi} \left[\int v_1 \cos\theta_1 \hbar\omega [f(\omega, T_{e1}) - f(\omega, T_{e2})] \tau_{12}(\omega, \phi_1, \theta_1) D_1(\omega)/4\pi d\omega \right] d\Omega_1 \tag{5.90}$$

Thus, with the help of the principle of detailed balance, we can express the flux with properties in one of the two media plus, of course, the transmissivity. This greatly simplifies the calculations. When the difference between T_{e1} and T_{e2} is small, eq. (5.90) can be further written as

$$q = \frac{T_{e1} - T_{e2}}{\Re_e} \tag{5.91}$$

where $\Re_e$ is the *specific thermal boundary resistance* [K m^2 W^{-1}],

$$\frac{1}{\Re_e} = \frac{1}{4\pi} \int_0^{2\pi} d\phi_1 \int_0^{\pi/2} d\theta_1 \int_0^{\omega_{D1}} v_1 \cos\theta_1 \sin\theta_1 \hbar\omega \frac{\partial f(\omega, T)}{\partial T} \tau_{12}(\omega, \phi_1, \theta_1) D_1(\omega) d\omega$$
$$= \frac{1}{2} \int_0^1 \left[\int_0^{\omega_{D1}} v_1 C_1(\omega) \tau_{12}(\omega, \mu_1) d\omega \right] \mu_1 d\mu_1 \tag{5.92}$$

where $\mu = \cos\theta$ is the directional cosine and $C_1(\omega)[= \hbar\omega D(\omega)\partial f/\partial T]$ is the spectral (mode) specific heat. The second equation in (5.92) used the fact that for an isotropic medium τ_{12} is independent of Φ. When the temperature is low and τ is independent of frequency, it can be shown from the above equation that

$$\Re_e \propto T_e^{-3} \tag{5.93}$$

because of the T^3 dependence of the specific heat. The agreement of eq. (5.93) with experimental results depends on how the transmissivity τ is calculated. Modeling results based on acoustic wave relations such as eq. (5.82) generally agree well with experimental results at very low temperatures, as shown in figure 5.8, and such a model is called the acoustic mismatch model (Little, 1959). At higher temperatures, however, the experimental results deviate from the acoustic mismatch model. This is because the thermal boundary resistance, calculated on the basis of the phonon transmissivity discussed in section 5.2.3, requires that phonon scattering at the interface be specular and elastic; that is, the phonon frequency does not change and the Snell law is obeyed. These conditions are satisfied at very low temperatures when the phonon wavelengths are long, but are more difficult to satisfy at higher temperatures when the phonon wavelengths contributing dominantly to heat transfer are comparable or shorter than the surface roughness. At room temperature, the average phonon wavelength is ~ 10–20 Å in most

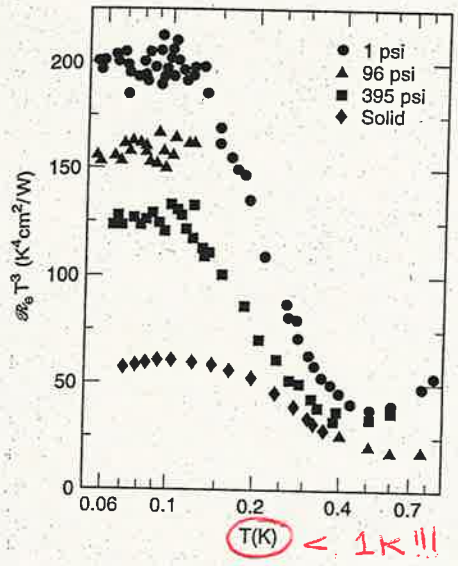


Figure 5.8 Thermal boundary resistance at low temperature (Swartz and Pohl, 1989; courtesy of APS Associate Publisher).

materials. A few atomic layers mixing will create a rough interface that no longer scatters phonons specularly.

Despite intense research, there is no generally accepted way to calculate the thermal boundary resistance at higher temperatures. One rather crude model is called the diffuse mismatch model (Swartz and Pohl, 1989), which assumes that phonons emerging from the interface do not really bear any relationship with their origin; in other words, one cannot tell which side they come from. This assumption implies that

$$R_{12} = \tau_{21} \text{ or } 1 - \tau_{12} = \tau_{21} \tag{5.94}$$

where the second equation comes from the energy conservation identity $R_{12} + \tau_{12} = 1$. We recall again that subscript 12 means from medium 1 into medium 2 and vice versa. By substituting the above relationship into eq. (5.89) and further assuming a linear dispersion for each acoustic wave polarization, Swartz and Pohl (1989) obtained

$$\tau_{d12} = \frac{1/v_2^2}{1/v_1^2 + 1/v_2^2} \tag{5.95}$$

where $1/v^2$ comes from the product of the density of states [eq. (3.55)] and the velocity. This relation is valid at low temperatures. At higher temperatures, a similar treatment leads to (Dames and Chen, 2004),

$$\tau_{d12}(T_e) = \frac{v_2 U_2(T_e)}{v_1 U_1(T_e) + v_2 U_2(T_e)} = \frac{1}{1 + [v_1 U_1 / v_2 U_2]} \tag{5.96}$$

where U is the volumetric internal energy [eq. (4.73)],

$$U = \int_0^{\omega_D} \hbar\omega f(\omega, T_e) D(\omega) d\omega = \int_0^{T_e} C(T) dT \tag{5.97}$$

Chen (1998) assumed that the specific heat is independent of temperature and further expressed eq. (5.96) in terms of the specific heat. Equation (5.96), however, is more accurate and includes all other simplified cases, including eq. (5.95). Using the Debye approximation, the ratio in the denominator of eq. (5.96) can be expressed as

$$\frac{v_1 U_1}{v_2 U_2} = \frac{1/v_1^2 \int_0^{\omega_{D1}} \omega^3 f(\omega, T_e) d\omega}{1/v_1^2 \int_0^{\omega_{D2}} \omega^3 f(\omega, T_e) d\omega} \tag{5.98}$$

We should caution, however, that the diffuse mismatch model discussed above is a very crude approximation and is clearly not valid when the two materials are very similar. In this case, the transmissivity should approach one but eq. (5.96) predicts a transmissivity approaching 0.5.

The case when the two materials are similar, and in the limit of an identical material with an imaginary interface, poses another dilemma for the usual thermal boundary resistance expression defined on the basis of eq. (5.91). In this limit, the thermal boundary resistance should clearly be zero but eq. (5.92) gives a finite value, even if one sets $\tau_{12} = 1$. This dilemma (Little, 1959) arises from the temperature definition used in eq. (5.91). So far, we have been careful in saying that T_e represents the temperature of the phonons coming toward the interface, and avoided discussing what are the real temperatures of medium 1 and medium 2 at the interface. This is because phonon interface transport is a highly nonequilibrium process and it is hard to define a temperature. Referring to figure 5.7(a), we see that on each side of the interface there are three groups of phonons. For example, on the side of medium 1, one group is the incoming phonons with a temperature T_{e1}, the other group is the reflected phonons leaving the interface, which has an energy distribution determined by the convolution of the incoming phonon at T_{e1} and the interface reflectivity. The third group comes from the transmission of side 2, with an energy distribution determined by the convolution of phonons at temperature T_{e2} and the interface transmissivity. The local phonon energy spectra at the interface are thus very different from that of the incoming phonons and cannot be represented by an equilibrium, or close to equilibrium, distribution with a single equilibrium temperature (Katerberg et al., 1977). However, if these phonons were to adiabatically approach an equilibrium, we could obtain the final equilibrium temperature of this adiabatic system and will call this temperature the *equivalent equilibrium temperature*. This equivalent equilibrium temperature is really just a measure of the local energy density rather than one that represents the spectral characteristics of the energy distribution, but it is consistent with the local equilibrium approximation used in heat transfer calculations such as the Fourier law. Figure 5.7(c) illustrates the difference between the incoming phonon temperature and the equivalent equilibrium temperature. It can be shown that the equivalent temperature can be related to the incoming phonon temperature by

$$T_1 = T_{e1} - (T_{e1} - T_{e2}) \int_0^1 \tau_{12}(\mu_1) d\mu_1 / 2 \tag{5.99}$$

$$T_2 = T_{e2} + (T_{e1} - T_{e2}) \int_0^1 \tau_{21}(\mu_2) d\mu_2 / 2 \tag{5.100}$$

where $\mu = \cos\theta$ is the directional cosine and we have assumed that transmissivity is independent of angle Φ. On the basis of this equivalent temperature and the consideration of the deviation of the phonon density of states from the Debye model, Chen and Zeng (2001) arrived at the following expression for the thermal boundary resistance:

$$\Re = \frac{T_1 - T_2}{q} = \frac{2[1 - (\int_0^1 \tau_{12}(\mu_1) d\mu_1 + \int_0^1 \tau_{21}(\mu_2) d\mu_2)/2]}{\int_0^1 [\int \tau_{12}(\mu_1)\, v_1 C_1(\omega) d\omega] \mu_1 d\mu_1} \tag{5.101}$$

Note that here we have dropped the subscript "e" because the temperatures are defined as a measure of the local energy density, not on the properties of the phonons coming toward the interface, as in eq. (5.91). Strictly speaking, the spectral specific heat should be evaluated at T_e but, under the assumption of a small temperature difference, it can be evaluated as the average of the equivalent equilibrium temperatures on the two sides. Equation (5.101) leads to zero thermal boundary resistance when the transmissivity from both sides is equal to one, that is, when no interface exists. This resolves the dilemma in eq. (5.92) that gives a nonzero thermal boundary resistance even when the transmissivity equals one (Little, 1959). When measuring thermal boundary resistance at low temperatures, it is possible to anchor temperature sensors so that T_{e1} and T_{e2} are measured, so that eq. (5.91) is a valid definition. Most high-temperature measurements of thermal boundary resistance, however, cannot determine T_e and the thermal boundary resistance values are obtained on the basis of the *equivalent equilibrium T*. Consequently, one should pay attention to using correct models to explain the experimental data.

The thermal boundary resistance discussed here exists even if the interface is perfect, as long as there exists phonon reflection at the interface. The order of magnitude of such thermal boundary resistance is, according to eq. (5.101), $\Re \sim 1/(Cv)$, where C is the volumetric specific heat and v is the phonon speed. Taking $C \approx 10^6$ J m^{-3} K^{-1} at room temperature and $v \approx 1000$ m s^{-1}, the thermal boundary resistance is then $\sim 10^{-9}$–10^{-8} K m^2 W^{-1}, consistent with experimental data for nearly perfect interfaces (Costescu et al., 2003). Less ideal interfaces have higher thermal boundary resistances (Stoner and Maris, 1993). Although the value of thermal boundary resistance for ideal interfaces seems exceedingly small, it becomes dominant for nanoscale systems with a large number of interfaces. For example, the thermal conductivity of superlattices in the direction perpendicular to the film plane is found to be dominated by the thermal boundary resistance. It should be pointed out, however, that the value of the thermal boundary resistance in a multilayer structure can differ from that of a single interface (Chen, 1998). In macroscopic structures, the thermal boundary resistance at the interfaces can be much larger because the two materials are not in perfect contact. We will not discuss these cases here.

5.3 Wave Propagation in Thin Films

In thin films, there are multiple interfaces. We should first emphasize that these thin films do not have to be an actual material. A thin vacuum space between two parallel plates can be considered a thin film. These interfaces will cause the reflection of the incident waves. The reflected waves can be superimposed on the incoming wave to cause interference effects that lead to the thickness dependence of reflectivity and transmissivity. One other new phenomenon that may occur in thin films is tunneling, which makes the total

reflection phenomenon that occurs at one interface disappear. These interference and tunneling processes can occur for photons, phonons, and electrons. In this section, we will first examine the interference phenomenon. The formulation established can also be applied to tunneling processes, which will be discussed in section 5.4.

5.3.1 Propagation of EM Waves

There are three ways to derive an expression for the radiative properties (reflectivity and transmissivity) of thin films: the field-tracing method, the resultant wave method, and the transfer matrix method, as explained in figure 5.9. The field-tracing method, figure 5.9(a), follows the trajectory of the wave and counts each reflection and transmission when the wave meets an interface (Born and Wolf, 1980), using the Fresnel reflection and transmission coefficients. This method is intuitive but cumbersome. Because all the forwarding waves in the same medium have the same exponential factor, we can sum them up into one wave with a undetermined amplitude and call this wave the resultant wave [figure 5.9(b)]. Similarly, all the backward propagating waves in the same medium can be summed into a resultant wave. There are then four resultant waves in the single layer thin film situation, one reflected, two inside the film (forward and backward), and one transmitted, as shown in figure 5.9(b). The amplitude of each resultant wave will be determined by applying the boundary conditions at the two interfaces. The transfer matrix method combines all the waves (both forward and backward) in each medium into one wave, and uses a matrix to relate the electric and magnetic fields between two different points inside a medium, as shown in figure 5.9(c). Because the tangential components of the electric and magnetic fields are continuous across the interface when no interface charge or interface current exists, the transfer matrix method can be easily extended to multilayers. We will therefore focus on the transfer matrix method.

Consider a TM wave, for example, the x-component of the electric field and the y-component of the magnetic field inside the film, as a function of location z:

$$E_x(z) = \cos\theta_2 E^+ e^{i\varphi(z)} + \cos\theta_2 E^- e^{-i\varphi(z)} \tag{5.102}$$

$$H_y(z) = \frac{n_2}{\mu c_0}[E^+ e^{i\varphi(z)} - E^- e^{i\varphi(z)}] \tag{5.103}$$

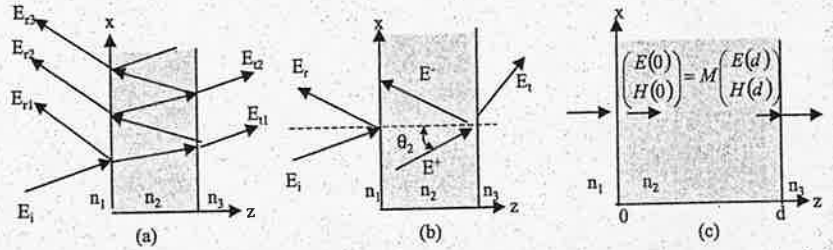


Figure 5.9 Three methods of treating reflection and transmission of electromagnetic fields through a thin film: (a) the field tracing method; (b) the resultant wave method; (c) the transfer matrix method.

where E^+ and E^- are the amplitudes of the resultant forward and backward propagating waves inside the film and

$$\varphi(z) = \frac{\omega n_2 z \cos\theta_2}{c_0} \tag{5.104}$$

where θ_2 is the angle formed between wavevector direction and z. Again, if n_2 is complex, this angle is also complex, and can be calculated according to the Snell law. In the above equations, we have dropped the terms $\exp(-i\omega t)$ and $\exp(-k_x x)$ because all terms have these factors and eventually cancel.

We want to relate the electric and magnetic fields at any location z inside the film to these fields at the interface $z = 0$. This can be realized by first taking $z = 0$ in eq. (5.102) and (5.103) and then eliminating E^+ and E^- in these equations,

$$E_x(z) = E_x(0)\cos\varphi(z) + ip_2 H(0)\sin\varphi(z) \tag{5.105}$$

$$H_y(z) = \frac{i}{p_2}E_x(0)\sin\varphi(z) + H_y(0)\cos\varphi(z) \tag{5.106}$$

where $p_2 = [\cos\theta_2/(n_2/\mu c_o)]$ is the *surface impedance* for a TM wave. The above equations can be written in matrix form

$$\begin{pmatrix} E_x(z) \\ H_y(z) \end{pmatrix} = \begin{pmatrix} \cos\varphi(z) & ip_2\sin\varphi(z) \\ \frac{i}{p_2}\sin\varphi(z) & \cos\varphi(z) \end{pmatrix} \begin{pmatrix} E_x(0) \\ H_y(0) \end{pmatrix} \tag{5.107}$$

Taking $z = d$ and inverting the above matrix, we get

$$\begin{pmatrix} E_x(0) \\ H_x(0) \end{pmatrix} = \begin{pmatrix} \cos\varphi_2 & -ip_2\sin\varphi_2 \\ -\frac{i}{p_2}\sin\varphi_2 & \cos\varphi_2 \end{pmatrix} \begin{pmatrix} E_x(d) \\ H_y(d) \end{pmatrix} = M \begin{pmatrix} E_x(d) \\ H_y(d) \end{pmatrix} \tag{5.108}$$

where $\varphi_2 = \varphi(d)$ and M is the second-order matrix in the above equation. We call M the interference matrix. It is easy to show that $|M| = 1$.

Equation (5.108) relates the electric and magnetic fields inside the film at $z = d$ to their values at the boundary $z = 0$. To find the reflectivity or transmissivity, we need to further relate them to the fields outside the film through the boundary conditions. For a boundary free of charge and current, eqs. (5.58) and (5.61) dictate that the electric and magnetic fields are continuous, which means that at $z = 0$,

$$E_x(0) = E_i\cos\theta_i + E_r\cos\theta_r = E_{ix} + E_{rx} \tag{5.109}$$

$$H_y(0) = \frac{n_1}{\mu c_0}(E_i - E_r) = \frac{1}{p_1}(E_{ix} - E_{rx}) \tag{5.110}$$

and at $z = d$, only the transmitted wave exists,

$$E_x(d) = E_t\cos\theta_t = E_{tx} \tag{5.111}$$

$$H_y(d) = \frac{n_3}{\mu c_0}E_t = \frac{1}{p_3}E_{tx} \tag{5.112}$$

where $p_1 = \cos\theta_i/(n_1/\mu c_0)$ and $p_3 = \cos\theta_t/(n_3/\mu c_0)$, and we have assumed that μ is the same for all layers because most materials are diamagnetic in the infrared to visible frequency range. We can again write the above equations in matrix form,

$$\begin{pmatrix} E_x(0) \\ H_y(0) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \frac{1}{p_1} & -\frac{1}{p_1} \end{pmatrix} \begin{pmatrix} E_{ix} \\ E_{rx} \end{pmatrix} \tag{5.113}$$

$$\begin{pmatrix} E_x(d) \\ H_y(d) \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{p_3} \end{pmatrix} E_{tx} \tag{5.114}$$

We now combine eqs. (5.113), (5.114), and (5.108), using the continuity of E_x and H_y at the interfaces, to get

$$\begin{pmatrix} 1 & 1 \\ \frac{1}{p_1} & -\frac{1}{p_1} \end{pmatrix} \begin{pmatrix} E_{ix} \\ E_{rx} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{p_3} \end{pmatrix} E_{tx} \tag{5.115}$$

where m_{ij} are the elements of the *interference matrix* M. Inverting the matrix of the left-hand side and multiplying out the three matrices, we obtain

$$\begin{pmatrix} E_{ix} \\ E_{rx} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (m_{11} + \frac{1}{p_3}m_{12}) + (m_{21} + \frac{1}{p_3}m_{22})p_1 \\ (m_{11} + \frac{1}{p_3}m_{12}) - (m_{21} + \frac{1}{p_3}m_{22})p_1 \end{pmatrix} E_{tx} \tag{5.116}$$

From the above matrix, we get the reflection and transmission coefficients through the film as

$$r = \frac{E_r}{E_i} = \frac{E_{rx}}{E_{ix}} = \frac{(m_{11} + \frac{1}{p_3}m_{12}) - (m_{21} + \frac{1}{p_3}m_{22})p_1}{(m_{11} + \frac{1}{p_3}m_{12}) + (m_{21} + \frac{1}{p_3}m_{22})p_1} \tag{5.117}$$

and

$$t = \frac{E_t}{E_i} = \frac{E_{ix}/\cos\theta_t}{E_{ix}/\cos\theta_i} = \frac{2c_{tm}}{(m_{11} + \frac{1}{p_3}m_{12}) + (m_{21} + \frac{1}{p_3}m_{22})p_1} \tag{5.118}$$

where $c_{tm} = \cos\theta_i/\cos\theta_t$. For a TE wave, the above expressions are still valid if p and c_{tm} are replaced by

$$p = -\frac{n\cos\theta}{\mu c_0} \text{ and } c_{te} = 1 \tag{5.119}$$

With the reflection and transmission coefficients known, we can calculate the reflectivity and transmissivity according to eqs. (5.75) and (5.76). For absorbing films, the above formulation is still valid by if n is replaced with the complex refractive index N.

The power of the matrix method can be best appreciated when dealing with multilayers of thin films. In this case, we can relate the electric and magnetic field inside the ith layer at both interfaces by the interference matrix M_i for that layer. Since the transverse components of the electric and magnetic fields are continuous at each interface that is free of net charge and current, the total interference matrix of the whole multilayer structure is

$$M = M_1 M_2 M_3 \ldots M_n \tag{5.120}$$

Thus, with such a simple substitution, all previous expressions for the single-layer film are still valid.

For a single layer of film, eqs. (5.117) and (5.118) can be written as

$$r = \frac{r_{12} + r_{23}e^{2i\varphi_2}}{1 + r_{12}r_{23}e^{2i\varphi_2}} \tag{5.121}$$

and

$$t = \frac{t_{12}t_{23}e^{i\varphi_2}}{1 + r_{12}r_{23}e^{2i\varphi_2}} \tag{5.122}$$

where r_{12}, r_{23} and t_{12}, t_{23} are the Fresnel reflection and transmission coefficients from medium 1 into medium 2 or from medium 2 to medium 3. The above formula is valid for both TM and TE waves.

On the basis of these expressions, we can calculate the reflectivity and transmissivity of the film. For a nonabsorbing film,

$$R = |r|^2 = \frac{r_{12}^2 + r_{23}^2 + 2r_{12}r_{23}\cos 2\varphi_2}{1 + 2r_{12}r_{23}\cos 2\varphi_2 + r_{12}^2 r_{23}^2} \tag{5.123}$$

$$\tau = \frac{n_3\cos\theta_t}{n_1\cos\theta_i}|t|^2 = \frac{(1 - r_{12}^2)(1 - r_{23}^2)}{1 + 2r_{12}r_{23}\cos 2\varphi_2 + r_{12}^2 r_{23}^2} \tag{5.124}$$

If the optical constants of any media are complex, we should use eq. (5.76) to calculate the transmissivity, and carry out complex number operation, $R = rr^*$ and $\tau = tt^*$.

The cosine function in eqs. (5.123) and (5.124) suggests that the reflectivity and transmissivity vary as a function of thickness, and when there is no absorption the variation is periodic, as shown in figure 5.10. This periodic variation in reflectivity and

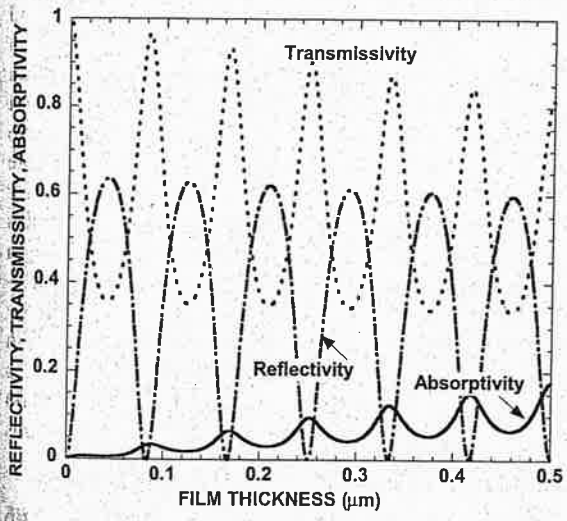


Figure 5.10 Reflectivity, transmissivity, and absorptivity of a thin film as a function of the film thickness, assuming vacuum on both sides.

transmissivity is the *interference phenomenon*, caused by the constructive or destructive superposition of the reflected and the incident waves. The maximum or minimum in the reflectivity can be found by setting $dR/d\varphi_2 = 0$, which leads to

$$\sin 2\varphi_2 = 0 \tag{5.125}$$

or

$$\frac{4\pi n_2 L \cos\theta_2}{\lambda_0} = m\pi \tag{5.126}$$

$$d = \frac{m\lambda_0}{4n_2 \cos\theta_2} \tag{5.127}$$

Under the above condition, eq. (5.123) becomes

$$R = \left(\frac{r_{12} - r_{23}}{1 - r_{12}r_{23}}\right)^2 = \left(\frac{n_1 n_3 - n_2^2}{n_1 n_3 + n_2^2}\right)^2 \quad (\text{for odd m} = 2l + 1) \tag{5.128}$$

$$R = \left(\frac{r_{12} + r_{23}}{1 + r_{12}r_{23}}\right)^2 = \left(\frac{n_1 - n_3}{n_1 + n_3}\right)^2 \quad (\text{for even } m = 2l) \tag{5.129}$$

where the first equality in the above two equations is valid for an arbitrary angle of incidence while the second is for normal incidence only. When the film thickness is $(2\ell + 1)\,\lambda_0/(4n_2 \cos\theta_2)$, the reflectivity R can be a maximum ($n_2 < n_3$) or a minimum ($n_2 < n_3$). Zero reflection occurs when the film has a refractive index $\sqrt{n_1 n_3}$ and its thickness satisfies eq. (5.127) for odd m. Such interference phenomena are the basis for *antireflection coatings*. When the film thickness is $\ell\lambda_0/(2n_2 \cos\theta_2)$, the reflectivity does not depend on the second layer.

The reflectivity and transmissivity of multilayer thin films can be calculated using the transfer matrix method. In practice, the reflectivity and transmissivity of multilayers can be controlled quite accurately with various thin-film deposition techniques and the possibility of controlling spectral and directional properties is large. One special example is the *Bragg reflector*, which is made from two alternating layers of thin films, figure 5.11(a). Each layer has a thickness equal to one-quarter of the light wavelength inside the film. Although, at one interface, the reflectivity between the two materials may be small, the coherent superposition of the reflected fields can create a reflectivity that is close to 100%. Such Bragg reflectors are used as coatings for mirrors that are highly reflective at a specific required wavelength, such as for lasers and X-rays. Figure 5.11(b) gives an example of the reflectivity of a quarter-wavelength mirror, similar to those used in special semiconductor laser structures called vertical-cavity surface-emitting lasers (Koyama et al., 1989; Walker, 1993). The reflectivity in certain spectral regions can reach 100%, meaning that no electromagnetic fields in that wavelength regime exist inside the reflector. These spectral regions, called stop bands, occur when the round-trip phase difference through one period (two layers) equals $2\ell\pi$, that is, when the forward and backward propagating fields inside the films cancel each other,

$$\frac{4\pi n_1 d_1 \cos\theta_1}{\lambda_0} + \frac{4\pi n_2 d_2 \cos\theta_2}{\lambda_0} = 2\ell\pi \tag{5.130}$$

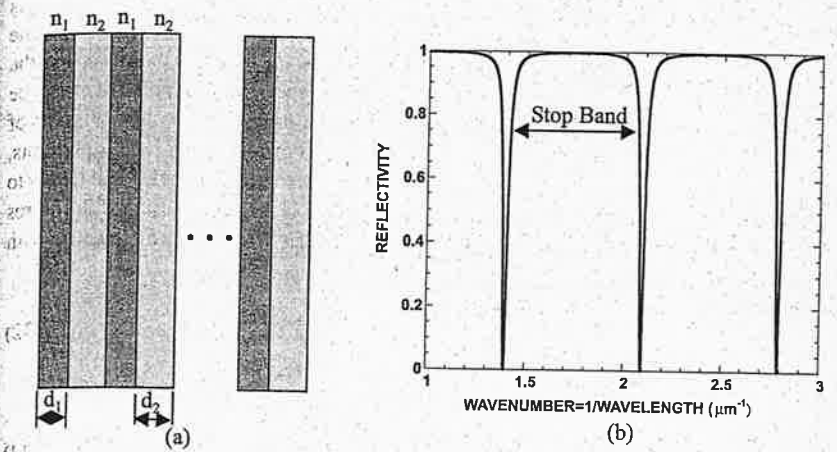


Figure 5.11 (a) A Bragg reflector is a periodic thin-film structure. (b) Calculated reflectivity of a Bragg reflector as a function of the incident photon wavelength for a reflector with refractive indices of 3 and 3.5 and a corresponding thickness of 417 Å and 352 Å for each layer.

where the subscripts 1 and 2 denote layer 1 and layer 2 respectively. Denoting $a = n_1 d_1 \cos\theta_1 + n_2 d_2 \cos\theta_2$ as the optical thickness of one period, the above equation can be written as (Knittl, 1976)

$$ka = \ell\pi \tag{5.131}$$

where $k(= 2\pi/\lambda_0)$ is the wavevector in vacuum. Equation (5.131) is identical to the condition of the electron bandgap formation discussed in chapter 3, which was obtained by solving the Schrödinger equation. We have said before that the formation of the electron bandgap is due to the cancellation of the electron waves inside the crystal. The discussion on the photon stop bands reinforces this picture. The similarities of these different waves, including electrons, photons, and phonons, have, in the past, been explored extensively to develop new concepts. For example, the *phonon interference filters* (Narayanamurti et al., 1979) and the electron minigaps (Esaki and Tsu, 1970), based on superlattices, benefited from the analogy of photon stop bands in interference filters. In return, it was exactly on the basis of the analogy of three-dimensional band structure in naturally existing crystals for electrons and phonons that the concept of three-dimensional *photonic crystals* was proposed (Yablonovitch, 1986), although one can also argue that this concept is an extension of the thin-film Bragg reflectors to three dimensions. Not only are these concepts very similar to each other; the mathematical techniques are also often interchangeable. For example, one popular approach for calculating the band structures of three-dimensional photonic crystals is based on a generalized transfer matrix method (Pendry, 1996).

5.3.2 Phonons and Acoustic Waves

In chapter 3, we considered phonon waves in a periodic lattice chain and discussed phonons in superlattices. The periodicity in naturally existing crystal lattices leads to

the representation of phonons in the first Brillouin zone. The periodicity of superlattices adds an additional restriction to the phonon wavevector and leads to the folded zone representation and the formation of phonon minibands [figure 3.30]. Similar to the photon stop bands, the phonon minigaps formed in the dispersion of superlattices can be thought of as stop bands generated by multiple reflections and coherent superposition of the lattice waves, as for photons in periodic structures. For long-wavelength phonons, that is, acoustic waves, one can also use the transfer matrix method as for optical waves to calculate the transmission of lattice waves through single-layer and multilayer structures (Nayfeh, 1995). The reflectivity r and transmissivity t of an SH wave through a film with thickness d can be calculated from the following matrix

$$\begin{pmatrix} 1 \\ r \end{pmatrix} = A_i^{-1} M A_t \begin{pmatrix} t \\ 0 \end{pmatrix} \tag{5.132}$$

where the interference matrix is similar to that of an electromagnetic wave

$$M = \begin{pmatrix} \cos\varphi_{T2} & i\sin\varphi_{T2}/Y_2 \\ iY_2\sin\varphi_{T2} & \cos\varphi_{T2} \end{pmatrix} \tag{5.133}$$

$$A_i = \begin{pmatrix} 1 & 1 \\ -Z_{Ti}\cos\theta_{Ti} & Z_{Ti}\cos\theta_{Ti} \end{pmatrix} \tag{5.134}$$

where $\varphi_{T2} = \omega d\cos\theta_2/v_{T2}$, $Y_2 = -Z_{T2}\cos\theta_2$, and A_t is obtained by replacing the subscript i in eq. (5.134) by t. The subscript T is used to represent properties of the transverse waves and, in this case, a transverse wave polarized perpendicular to the plane of incidence. The reflection and transmission coefficients are defined as

$$r = v_r(0)/v_i(0) \quad t = v_t(d)/v_i(0) \tag{5.135}$$

The matrix formulation for SH acoustic waves is clearly similar to that for optical waves. Multilayers can again be treated by simply replacing the interference matrix M with the product $M_1M_2\ldots M_{2n+1}$. The order of the matrices is the same as the sequence of the layers. For longitudinal waves (L) and vertically polarized transverse waves (SV) with the displacement polarized in the plane of incidence, the relationship between the incident, reflected, and transmitted wave velocity components of isotropic media is

$$\begin{pmatrix} v_{Ti}(0) \\ v_{Li}(0) \\ v_{Tr}(0) \\ v_{Lr}(0) \end{pmatrix} = B_i^{-1} M B_t \begin{pmatrix} v_{Tt}(d) \\ v_{Lt}(d) \\ 0 \\ 0 \end{pmatrix} \tag{5.136}$$

where v_{Ti} and v_{Li} are the amplitudes of the displacement velocities of the incident transverse and longitudinal waves, respectively, and subscripts r and t represent the reflected and transmitted waves, as usual. Matrix B_i is a 4×4 matrix given by

$$B_i = \begin{pmatrix} -\sin\theta_{Ti} & \cos\theta_{Li} & \sin\theta_{Ti} & -\cos\theta_{Li} \\ \cos\theta_{Ti} & \sin\theta_{Li} & \cos\theta_{Ti} & \sin\theta_{Li} \\ -\mu_1 k_{Ti}\sin 2\theta_{Ti} & (\lambda_1 + 2\mu_1\cos^2\theta_{Li})\,k_{Li} & -\mu_1 k_{Ti}\sin 2\theta_{Ti} & (\lambda_1 + 2\mu_1\cos^2\theta_{Li})\,k_{Li} \\ \mu_1 k_{Ti}\cos 2\theta_{Ti} & \mu_1 k_{Li}\sin 2\theta_{Li} & -\mu_1 k_{Ti}\cos 2\theta_{Ti} & -\mu_1 k_{Li}\sin 2\theta_{Li} \end{pmatrix} \tag{5.137}$$

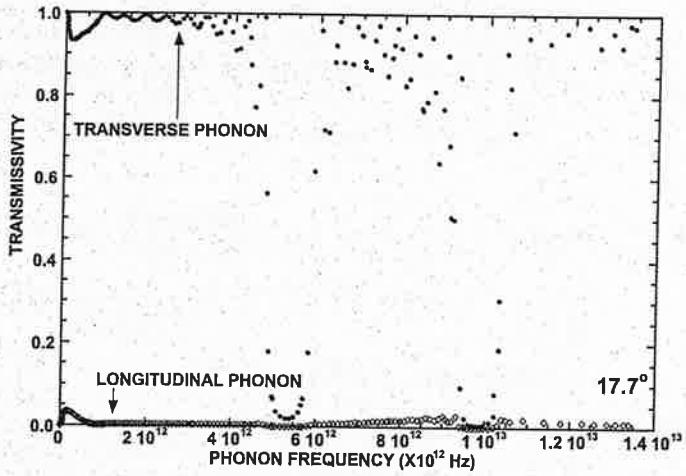


Figure 5.12 Transmissivity of a transverse acoustic wave polarized in the plane of incidence through a Si/Ge-like superlattice as a function of frequency with an incident angle of 17.7° (Chen, 1999).

In the above expressions, $k_i\,(=\omega/v_i)$ is the magnitude of the wavevector of the incident waves (SV or L, as distinguished by subscripts T and L). B_t is obtained by replacing subscript i with t, that is, from incident to transmitted waves. The interference matrix of the layer (with index 2) in eq. (5.136) is obtained from $M = B_2^{-1}N_2B_2$, where B_2 is obtained by replacing i in eq. (5.137) by 2, and N_2 is given by

$$N_2 = \begin{pmatrix} e^{i\varphi_{T2}} & 0 & 0 & 0 \\ 0 & e^{i\varphi_{L2}} & 0 & 0 \\ 0 & 0 & e^{-i\varphi_{T2}} & 0 \\ 0 & 0 & 0 & e^{-i\varphi_{L2}} \end{pmatrix} \tag{5.138}$$

The transfer matrix is 4×4 because, as shown in eq. (5.136), the longitudinal and transverse waves are coupled and the conversion between these two waves is possible at the interface. With eq. (5.136), the reflectivity and transmissivity for an incident field (either v_{Ti} or v_{Li}) can be calculated.

Figure 5.12 shows an example of phonon transmissivity through a Si/Ge-like superlattice obtained by the transfer matrix method (Chen, 1999), for a transverse wave polarized in the plane of incidence at an angle of incidence of 17.7°. The stop bands in transmissivity (zero transmissivity) correspond to the minigaps obtained from lattice dynamics simulation (figure 3.30) (Yang and Chen, 2001). The figure also shows that some transverse incident waves are converted into longitudinal waves.

5.3.3 Electron Waves

The study of electron wave propagation in layered media started with the investigation on superlattices (Esaki and Tsu, 1970). The most popular approach has been based on

solving the Schrödinger equation using the Kronig–Penney model. There are, however, also a few approaches based on the transfer matrix method for electron transport in superlattices (Tsu and Esaki, 1973; Huang and Wu, 1992). Because we have dealt extensively with the Kronig–Penney model in chapter 3 and described the transfer matrix method above, we will not present any details on the applications of the transfer matrix method to electron waves here.

5.4 Evanescent Waves and Tunneling

In section 5.2, we saw that when total internal reflection occurs for each of the three types of waves, an evanescent wave exists on the other side of the interface. The fields or wavefunctions of the evanescent wave decay exponentially from the interface. The time averaged net energy or particle flux carried by the evanescent wave is zero. However, if a third medium is brought close to the interface before the evanescent wave dies down completely, the evanescent wave can refract into this third medium. If this refracted wave is a propagating mode in the third medium, the evanescent wave becomes "revitalized" and a net energy or particle flux "tunnels" through the small region between the incident medium and the third medium. The reflection will no longer be total. In fact, one can even reach 100% transmission under appropriate conditions. The descriptions of these evanescent waves and the tunneling phenomena are based on the same mathematical expressions as we have obtained in the previous section.

5.4.1 Evanescent Waves

For electron reflection at a step potential, as shown, in figure 5.1, total reflection occurs when the electron energy E is smaller than the potential height U_0. In this case, eqs. (5.54)–(5.56) lead to

$$\Psi_t = \frac{2i|k_2|e^{-|k_2|z}}{k_1 + i|k_2|}\Psi_i \tag{5.139}$$

where

$$|k_2| = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}} \tag{5.140}$$

Thus the penetration depth of the evanescent wave, which we define as the depth at which the wavefunction decays to e^{-1} of its boundary value, is

$$\delta = \frac{1}{|k_2|} \tag{5.141}$$

Taking $U_0 - E = 1$ eV, we get $\delta \approx 2$ Å, which is a very short distance.

For an electromagnetic wave incident above the critical angle, the Snell law gives

$$\sin\theta_t = \frac{n_1 \sin\theta_i}{n_2} > 1 \tag{5.142}$$

and thus

$$\cos\theta_t = \sqrt{1 - \sin^2\theta_t} = i\sqrt{\left(\frac{n_1 \sin\theta_i}{n_2}\right)^2 - 1} = i|\cos\theta_t| \tag{5.143}$$

Substituting the above expression into eq. (5.72) and then into eq. (5.64) gives the evanescent electric field for a TM wave as

$$\frac{2n_1\cos\theta_i(\hat{\mathbf{x}}\cos\theta_t - \hat{\mathbf{z}}\sin\theta_t)}{n_2\cos\theta_i + n_1|\cos\theta_t|i}|\mathbf{E}_{i//}|\exp\left[-i\omega\left(t - \frac{n_2 x\sin\theta_t}{c_0}\right)\right]\exp\left(-z\frac{n_2\omega}{c_0}|\cos\theta_t|\right) \tag{5.144}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ are is the unit vectors along the x and z coordinate directions. The above equation demonstrates that the evanescent field decays exponentially with a penetration depth

$$\delta = \frac{\lambda_0}{2\pi n_2|\cos\theta_t|} \tag{5.145}$$

which is roughly of the same order as the wavelength inside the medium. Using eq. (5.144) and the corresponding expression for the magnetic field, it is also easy to show that the z-component time-averaged Poynting vector of the evanescent electromagnetic field is zero,

$$\langle \mathbf{S}_z \rangle = \frac{1}{2}\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*)_z = \frac{1}{2}\mathrm{Re}(E_x H_y)\hat{\mathbf{z}} = 0 \tag{5.146}$$

that is, no net energy flows across the interface. However, if the instantaneous Poynting vector is examined, it can be seen that there is instantaneous energy flow into and out of the second medium carried by the evanescent field. The net energy flow in and out averaged over time, however, is zero.

The above discussion shows the similarities between evanescent electron waves and electromagnetic waves. Evanescent acoustic waves can be similarly analyzed. We will not go into the details.

5.4.2 Tunneling

Tunneling of the evanescent waves may occur if a third medium is brought close to the first interface such that the exponentially decaying evanescent wave has finite magnitude at the interface between the second medium and the third medium. If the wave refracted into the third medium is propagating, a net flow of particles or energy from the first to the third medium occurs. In figure 5.13, we illustrate the tunneling of electromagnetic and electron waves. The analysis of the tunneling process can be based on the same methods that are used for treating interference phenomena. For electromagnetic waves passing through one layer of a thin film, for example, the transmissivity is given by eq. (5.122). Substituting relations, eqs. (5.104) and (5.143), into eq. (5.122), we get

$$t = \frac{t_{12}t_{23}\exp\left[-\frac{2\pi n_2 d|\cos\theta_2|}{\lambda_0}\right]}{1 + r_{12}r_{23}\exp\left[-\frac{4\pi n_2 d|\cos\theta_2|}{\lambda_0}\right]} \tag{5.147}$$

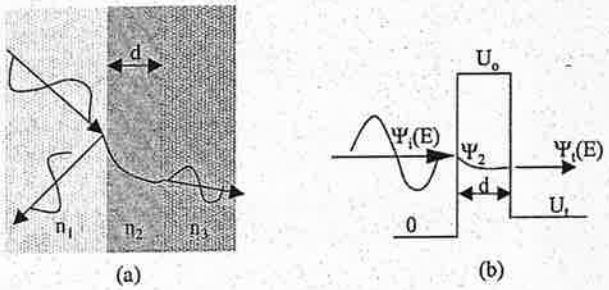


Figure 5.13 Tunneling of totally reflected waves, (a) for an electromagnetic wave and (b) for an electron wave.

where we have used θ_2 rather than θ_t to denote that the angle is for the wave inside the second medium, not medium 3. The transmissivity due to tunneling for a TM wave is

$$\tau = \frac{\mathrm{Re}(n_3 \cos\theta_t)}{n_1 \cos\theta_1} |t|^2 \tag{5.148}$$

If the refractive index n_3 is larger than n_2, there are certain incident angles of θ_i that allow a real solution for θ_t, while θ_2 is an imaginary angle. From the Snell law, this occurs when the incident angle falls in the range

$$\sin^{-1}\left(\frac{n_2}{n_1}\right) \le \theta_i \le \sin^{-1}\left(\frac{n_3}{n_1}\right) \tag{5.149}$$

When tunneling happens, it goes without saying that the reflectivity is no longer 100% as for the case of total reflection. If the medium through which the wave tunnels is nonabsorbing, energy or particle conservation gives $R = 1 - \tau$.

For electrons, we can solve the Schrödinger equation for a barrier structure as shown in figure 5.13. The solution follows closely the method we used in section 3.2.1, which also resembles the derivation of the transfer matrix method for electromagnetic waves. The tunneling transmissivity through a potential barrier of height U_0 and width d is (Cohen-Tannoudji et al., 1997)

$$\tau = \frac{4E(U_0 - E)}{4E(U_0 - E) + U_0^2 \sinh^2[\sqrt{2m(U_0 - E)}d/\hbar]} \tag{5.150}$$

When the argument of the hyperbolic sine function is large, the above expression can be approximated as

$$\tau \approx \frac{16E(U_0 - E)}{U_0^2} \exp[-2\sqrt{2m\,(U_0 - E)}d/\hbar] = \frac{16E\,(U_0 - E)}{U_0^2} e^{-2|k_2|d} \tag{5.151}$$

The same tunneling phenomenon can also occur for phonons. Figure 5.14 shows the transmissivity of an acoustic wave through a superlattice, calculated on the basis of the transfer matrix method as a function of frequency and angle of incidence (Chen, 1999). At a low angle of incidence, the transmission behaves as normal and has several stop

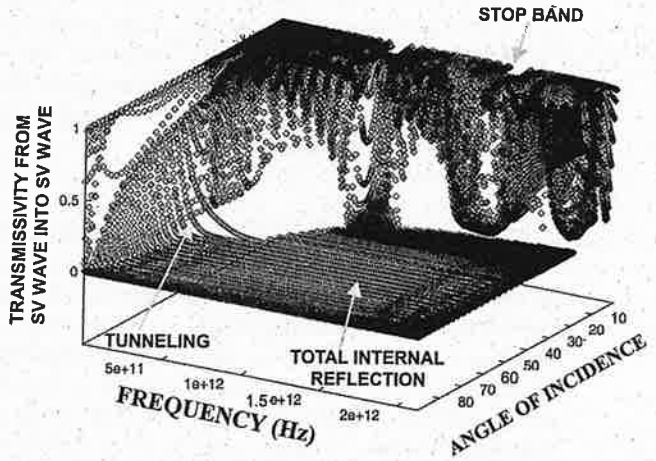


Figure 5.14 Phonon transmissivity through a Si/Ge-like superlattice, each layer 5 Å thick, showing the stop bands, the total reflection region, and the tunneling region (Chen, 1999).

bands. At a large angle of incidence, for which total reflection occurs, the transmissivity across the superlattice is not zero but decreases exponentially as the frequency increases, due to tunneling of acoustic waves.

Tunneling phenomena are the basis of several inventions that led to Nobel prizes, including the tunneling diode by Esaki (1958) and the scanning tunneling electron microscope (STM) (Binnig and Rhorer, 1982). The principle of an STM is shown in figure 5.15(a). A sharp tip is brought into close proximity with a conducting surface but without contacting the surface. Under an applied voltage, electrons tunnel through the vacuum gap and create a current in the loop. The current is extremely sensitive to the separation (sub-angstrom) between the tip and the surface, as one can easily

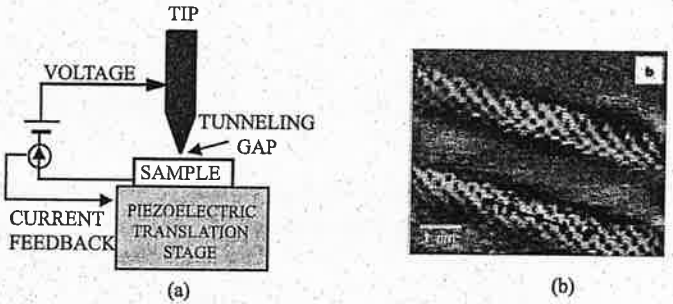


Figure 5.15 (a) The scanning tunneling microscope is based on the sub-angstrom level sensitivity of the tunneling current between a conducting tip to a conducting sample as a feedback to control the piezoelectric translation stage, which is also capable of sub-angstrom motion precision, to fly the tip over the sample and to obtain information on the topographical and electron structure of the sample surface. (b) STM image of two single-walled carbon nanotubes (Odom et al., 1998; courtesy of Nature Publishing Group).

see from eq. (5.151) since k_2 is of the order of ~ 1 Å^{-1}. As the tip is scanned over the sample, different regions have different potential barriers of different heights. By using the current as a feedback signal to control the tip–sample separation, one can map the electronic wavefunction surrounding individual atoms or the surface roughness. Figure 5.15(b) shows the STM images of two single-walled carbon nanotubes (Odom et al., 1998).

Since the invention of the STM, a host of other types of microscope have been invented, including the atomic force microscope (Binnig et al., 1986), photon scanning tunneling microscope (Reddick et al., 1989), scanning thermal microscope (Majumdar et al., 1995), and others. The photon scanning tunneling microscope is also based on the evanescent electromagnetic wave hovering above a surface. The atomic force microscope, however, is based on an even simpler principle: the effective spring constant between atoms can be quite large—much larger, for example, than that of a Si cantilever 3 μm (thickness) × 100 μm (length) × 10 μm (width). When such a cantilever is in contact with a sample via a sharp tip, the atoms of the sample will not be scratched off. Rather, the cantilever will be displaced. Angstrom-level displacement can be easily measured with either the STM (Binnig et al., 1986) or through laser deflection, making it possible to use such a device to measure the angstrom-level topography of surfaces, particularly for dielectric surfaces that cannot be characterized with an STM because the sample is nonconducting.

5.5 Energy Transfer in Nanostructures: Landauer Formalism

Knowing the transmissivity from one point to another in a system, one can easily estimate various fluxes associated with the carriers (charge, momentum, energy, etc.). Consider for example, the heat transfer between two reservoirs at temperatures T_1 and T_2, as shown in figure 5.16. The heat flux from reservoir 1 to 2 is

$$q_{1\to 2} = \sum_p \left[\frac{1}{V_1} \sum_{k_{x1}=-k_{\max}}^{k_{\max}} \sum_{k_{y1}=-k_{\max}}^{k_{\max}} \sum_{k_{z1}=0}^{k_{\max}} v_{z1} E \tau_{12} f(E, T_1) \right] \tag{5.152}$$

where E is the energy of one carrier and τ_{12} is the transmissivity from point 1 to point 2 for the carrier with energy E, v_{z1} is the velocity of the carrier, the index p represents summation over all the polarizations of the carriers, and the wavevector summation indices are over all values of k_x and k_y and positive values of k_z. Equation (5.152) is a recasting of eq. (5.86) and is valid for electrons and photons as well as phonons. We can similarly write the reverse heat flux from point 2 to point 1. The difference between these heat fluxes gives the net heat flux between point 1 and point 2, similar to eq. (5.88). The

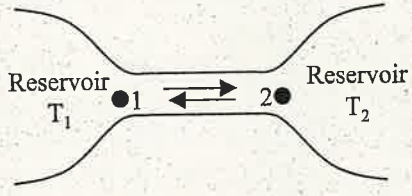


Figure 5.16 The Landauer formulation of the net (energy, charge, particle) flux between two points is based on the carrier transmissivity between the two points.

principle of detailed balance can once again be applied to obtain a relationship between the transmissivity from point 1 to point 2 and the reverse direction, as expressed in eq. (5.89). With such an approach, and when the difference between T_1 and T_2 is not large, the heat flux between point 1 and point 2 can be expressed as

$$q = (T_1 - T_2) \sum_p \int_{\Omega \geq 2\pi} \left[\int v_1 \cos\theta E \frac{\partial f(E,T)}{\partial T} \tau_{12}(E,\phi,\theta) D_1(E) dE/4\pi \right] d\Omega$$
$$= K \Delta T \tag{5.153}$$

where K is the thermal conductance with units of W m^{-1} K^{-1}, Ω is the solid angle defined in figure 3.27, and T is the average temperature. One can write down similar expressions for the current density and particle flux. These types of expressions for flux are called the *Landauer formalism*, which views the transport as a transmission process (Imry and Landauer, 1999).

The key for applying the Landauer formalism is the calculation of the transmissivity. When scattering exists, the calculation of the transmissivity is more difficult and the Landauer formalism is less useful. When no internal scattering exists, which is also called ballistic transport, the transmissivity can be calculated relatively easily, as we have done for a single interface and multilayers, and the Landauer formalism is very convenient to use.

The effects of interference and tunneling in thin films, and more generally in nanostructures, on the transport processes can be studied from the Landauer formalism, using appropriately calculated transmissivity between two points. As an example, we consider radiative heat transfer between two parallel plates, paying special attention to the case when the spacing between the plates is small. Quite a few studies have been devoted to radiative heat transfer across small gaps. In a series of studies, Tien and co-workers (Cravolho et al., 1967; Domoto et al., 1970) investigated the effects of tunneling and interference on radiative heat transfer between small vacuum gaps, which are used in low-temperature thermal insulation materials (Tien and Cunnington, 1973). More extensive experiments were performed by Hargreaves (1969). The approach championed by Tien and co-workers was equivalent to the Landauer formalism. Polder and van Hove (1971) established a direct approach that considered the emission processes based on Rytov's electromagnetic field fluctuation theory (Rytov et al., 1987; Narayanaswamy and Chen, 2004). Pendry (1999) provided a slightly different point of view on radiative heat transfer in small gaps, based on the Landauer formalism. Figures 5.17(a) and (b) show the modeling and experimental data for radiative heat transfer between small gaps. Due to tunneling, the radiation flux increases as the vacuum gap decreases and a radiative heat flux much higher than that between two blackbodies can be realized, as shown in figure 5.17(c). Some recent applications of the tunneling phenomena include the scanning tunneling microscope (Xu et al., 1994) and thermophotovoltaics (DiMatteo et al., 2001; Whale and Cravalho, 2002). When the two objects are identical, the maximum radiation heat transfer can be increased by n^2 times the blackbody radiation heat transfer between two surfaces through tunneling of the internally reflected wave. The n^2 limit is the blackbody emissive power inside an object, which can be derived following similar steps as we arrived at eq. (4.83). In addition to the tunneling of evanescent waves, the tunneling of surface waves that decay exponentially on both sides of the interface

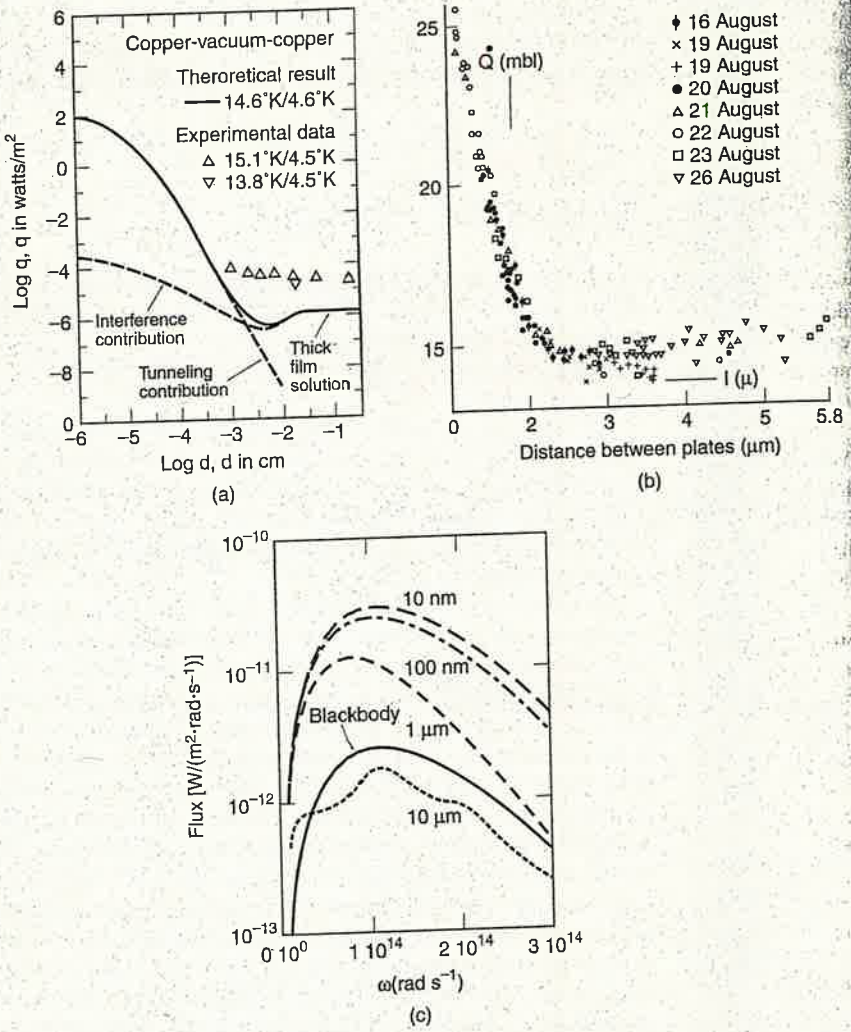


Figure 5.17 Size effects on radiation heat transfer between two parallel plates (a) at low temperature (Domoto et al., 1970; courtesy of ASME), (b) at room temperature (Hargreaves, 1969; courtesy of Elsevier). (c) Radiative heat flux as a function of frequency for radiation heat transfer between a plate at 300 K and another at 0 K, demonstrating that the heat flux can exceed the exchange between two blackbodies.

can also occur (Mulet et al., 2002; Narayanaswamy and Chen, 2003). The surface waves exist when the dielectric constants of the two sides of the interface are equal in magnitude but are of opposite signs (Raether, 1987). If one side of the interface is vacuum, the other side should have a negative dielectric constant close to unit, which can occur when the electrons or phonons are in resonance with electromagnetic waves

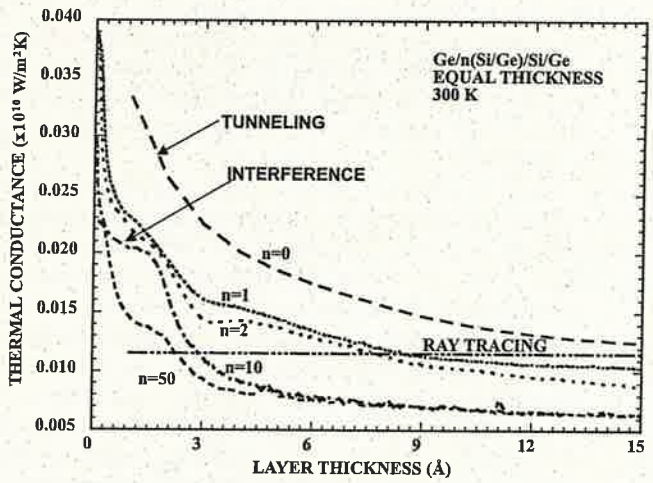


Figure 5.18 Thermal conductance of superlattice from transfer matrix calculation (Chen, 1999).

in the medium. The resonant modes of electromagnetic waves with optical phonons are called *phonon-polaritons* and those modes with electrons are called *plasmons*. Surface phonon-polaritons and surface plasmons have high energy density near the surface but decay rapidly away from the surface. Radiative heat flux through the tunneling of surface waves can significantly exceed the n^2 limit of evanescent waves.

The same interference and tunneling phenomena also affect the radiative properties of thin films grown on substrates. The emissivity of the surface will change with film thickness (Wong et al., 1995; Chen, 1996; Zhang et al., 2003). This affects, for example, the temperature of semiconductor wafers during thin film growth. The uncertainty in temperature measurement caused by the emissivity change is a significant factor in the design of semiconductor equipment used for rapid thermal processing (Nulman, 1989).

For phonons in thin films, interference and tunneling phenomena may also affect heat conduction in extremely thin films such as superlattices with very short spatial periods. Chen (1999) evaluated thermal conductance in the limit of no scattering of thin films and superlattices, as shown in figure 5.18. Generally, when the film thickness is less than a few monolayers, tunneling can increase the conductance. Lattice dynamics simulations lead to similar conclusions (Tamura et al., 1999; Simkin and Mahan, 2000; Yang and Chen, 2001, 2003). So far, there have been some experimental data that suggest this phenomenon, but they are not very conclusive (Capinski et al., 1999; Venkatasubramanian, 2000).

Example 5.2 *Universal quantum thermal conductance*

Develop a model for the thermal conductance of a square nanowire between two thermal reservoirs, neglecting the internal scattering and assuming the phonon transmissivity for each allowable mode is one.

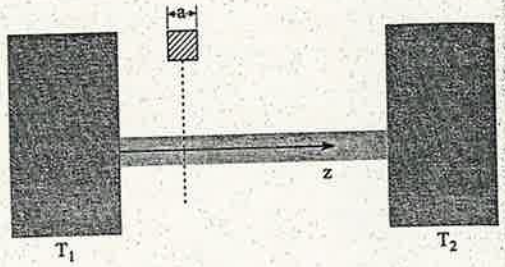


Figure E.5.2 Nanowire connecting two thermal reservoirs.

Solution: Consider a square nanowire as shown in figure E.5.2. In the cross-sectional direction, standing waves must be formed so that the allowable wavevectors in the x and y directions are

$$k_x = 2\pi\frac{m}{2a} = \frac{\pi m}{a}, k_y = \frac{n\pi}{a}(m, n = \pm1, \pm2, \dots)$$

We assume that the linear dispersion relation as used in the Debye model is still valid. The allowable modes inside the nanowire are

$$\omega = ck = c\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{a}\right)^2 + k_z^2} \tag{E5.2.1}$$

for each polarization. With a phonon transmissivity of one (which requires that the materials of the reservoirs and the nanowire are the same, and also a tapered joint between the wire and the reservoir, similar to that drawn in figure 5.16), we can use the Landauer formalism to express the heat transfer through the nanowire as

$$q_{12} = q_{1\to2} - q_{2\to1} = \frac{1}{2\pi}\sum_{m,n}\int v_z\hbar\omega\,[f(\omega, T_1) - f(\omega, T_2)]\,dk_z \tag{E5.2.2}$$

where the factor $(1/2\pi)$ arises from the conversion of the summation over the quantum state determined by k_z in eq. (5.152) into an integration over k_z, because the separation between two consecutive k_z is $2\pi/L$, where L is the length of the wire. We also used direct summation for the modes in the x and y directions because the separations between two consecutive wavevectors in these directions are large when a is small. We will see in the next section that the phonon velocity v_z used for energy transport calculations should be the group velocity

$$v_z = d\omega/dk_z \tag{E5.2.3}$$

Thus, eq. (E5.2.2) can be converted into integration over frequency,

$$q_{12} = \frac{1}{2\pi}\sum_{m,n}\int \hbar\omega\,[f(\omega, T_1) - f(\omega, T_2)]\,d\omega \tag{E5.2.4}$$

If we further assume that $T_1 - T_2$ is small, the thermal conductance of the nanowire, eq. (E5.2.2), can be expressed as

$$K = \frac{q_{12}}{T_1 - T_2} = \frac{3}{2\pi}\sum_{m,n}\int_{\omega_{mn}}^{\omega_{\max}} \hbar\omega\frac{df}{dT}d\omega \tag{E5.2.5}$$

where the factor of 3 represents the three phonon polarizations, $\omega_{\max}$ is the maximum phonon frequency, and

$$\omega_{mn} = c\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{a}\right)^2} \tag{E5.2.6}$$

For the first few quantized modes (m, n are small, so that ω_{mn} is small) and at low temperature, eq. (E5.2.5) can be simplified to

$$K = \frac{3\kappa_B^2 T}{h}\sum_{m,n}\int_{x_{mn}}^{\infty} \frac{x^2 e^x}{(e^x - 1)^2}dx \tag{E5.2.7}$$

where $x_{mn} = \hbar\omega_{mn}/\kappa_B T$. When m and n are small, such that the lower limit can be extended to zero, the integral value is $\pi^2/3$. In this limit, the thermal conductance of each mode is

$$K_1 = \frac{\pi^2\kappa_B^2 T}{3h} \tag{E5.2.8}$$

Comment. This thermal conductance expression does not depend on the material properties and thus is the same for all materials. Such universal conductance behavior also happens for electrons. Quantum size effects on thermal conductance have been observed experimentally (Angelescu et al., 1998; Schwab et al., 2000).

In closing this section, we would like to point out that the Landauer formalism, as expressed in eqs. (5.152) and (5.153), is based on the assumption that the carrier distributions going from point 1 to point 2 are determined by the equilibrium reservoir temperatures (and chemical potentials). In reality, the local carriers at point 1 are not at equilibrium because there are also carriers coming from point 2 with a characteristic temperature of T_2. So T_1 and T_2 actually do not represent the local temperatures at points 1 and 2. We have discussed this point carefully in connection with the treatment of the thermal boundary resistance as represented by figures 5.7(a)–(c). Whether one should use eq. (5.153) or the local equivalent equilibrium temperature (or chemical potential) depends on how experiments are conducted or how models are laid out. So far, most experiments are done in electron systems with the electrochemical potentials of the reservoirs measured, and thus the Landauer formulation is directly applicable. If one treats the transport inside the reservoirs concurrently with the ballistic transport between point 1 and point 2, the consistency of the definitions in each region must be considered carefully (Chen, 2003).

5.6 Transition to Particle Description

Our discussion in this chapter so far has centered on the wave nature of the energy carriers by considering their phase. From quantum mechanics, we know that energy carriers have both wave and particle characteristics. When the energy carriers are treated as waves, their particle characteristics are included, for example, by considering the energy of one phonon as $h\nu$. In macroscale, we often ignore the phase and treat the energy carriers as particles, either quantum particles such as photons or phonons, or simply as classical particles with no energy quantization. The question is then: when must we consider the energy carriers as waves, and when can we ignore the phase information and treat them simply as pure particles, either quantized particles or classical particles? We will attempt to answer these questions in this section.

5.6.1 Wave Packets and Group Velocity

In our previous discussion on energy propagation, such as eq. (5.152), we did not give much consideration to the meaning of the velocity. This velocity should represent the speed and direction of energy propagation and is usually the group velocity. To understand the *group velocity,* we first consider a plane wave traveling along the x-direction

$$\mathbf{A}e^{-i(\omega t - kx)} \tag{5.154}$$

Its phase velocity is [eq. (2.4)]

$$v_{p,x} = \frac{dx}{dt} = \frac{\omega}{k} \tag{5.155}$$

Is this velocity the speed of signal or energy propagation? Generally, the answer is no. We see that the plane wave represented by eq. (5.154) extends from minus infinity to plus infinity in both time and space. It has no start or finish, and does not represent any meaningful signal. In practice, a signal has a starting point and an ending point in time. Let's suppose that a harmonic signal at frequency ω_0 is generated during a time period $[0, t_0]$, as shown in figure 5.19(a). Such a finite-time harmonic signal can be decomposed through a Fourier series into the summation of true plane waves with time extending from minus infinity to plus infinity, as shown in figure 5.19(b). The frequencies of these plane waves are centered around ω_0 and their amplitudes decay as the frequency moves away from ω_0, as illustrated in figure 5.19(c). One can better understand these pictures by actually carrying out the Fourier expansion. Because each of the plane waves in such a series expansion is at a frequency slightly different from the central frequency ω_0, it also has a corresponding wavevector that is different from k_0, as determined by the dispersion relation between ω and k. The subsequent propagation of the signal can be obtained from tracing the spatial evolution of all these Fourier components as a function of time.

For simplicity, let's consider that the signal is an electromagnetic wave with the electric field points to y-direction. We pick only two Fourier components of equal amplitude and consider their superposition, one at frequency $\omega_0 - \Delta\omega/2$ and another at

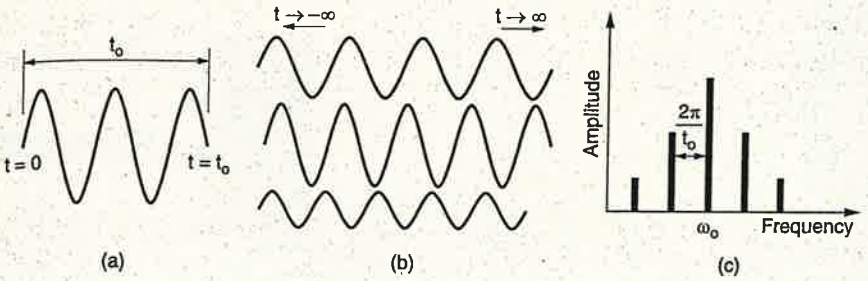


Figure 5.19 (a) A finite-time period signal generated between time period $(0, t_0)$ contains a spectrum of plane waves extending infinitely in time (b), with a power spectrum shown in (c). The propagation of these plane waves evolves into a wave packet.

frequency $\omega_0 + \Delta\omega/2$, propagating along the positive x-direction [Figure 5.20(a)]. The superposition of these two waves gives the electric field as

$$\begin{aligned} E_y(x,t) &= a\cos\left[\left(\omega_0 - \frac{\Delta\omega}{2}\right)t - \left(k_0 - \frac{\Delta k}{2}\right)x\right] \\ &\quad + a\cos\left[\left(\omega_0 + \frac{\Delta\omega}{2}\right)t - \left(k_0 - \frac{\Delta k}{2}\right)x\right] \\ &= 2a\cos(\Delta\omega t - \Delta k x)\cos(\omega_0 t - k_0 x) \end{aligned} \tag{5.156}$$

The above electric field is shown schematically in figure 5.20(b). There appear to be two waves: one is the carrier wave at central frequency ω_0, another is the modulation of the carrier wave by a wave at frequency $\Delta\omega$. If the frequency $\Delta\omega$ is much smaller than ω_0, we can calculate the Poynting vector time-averaged over a period much shorter than

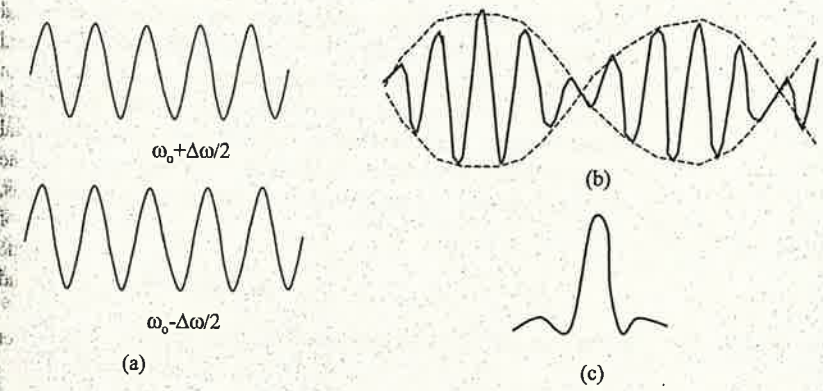


Figure 5.20 Example of the superposition of two plane waves (a) into wave packets (b). If there are many different frequency components, the superposition leads to a narrow wave packet (c).

$1/\Delta\omega$ but much longer than $1/\omega_0$ according to eq. (5.37), to obtain the average energy flux as

$$S(x,t) = \frac{4a^2 n}{c_0 \mu} \cos^2(t\,\Delta\omega - x\Delta k) \tag{5.157}$$

which is another wave propagating at the speed

$$v_{g,x} = \frac{\Delta\omega}{\Delta k} \tag{5.158}$$

This means that the energy is propagating at the speed of $v_{g,x}$ rather than the phase velocity. This $\mathbf{v}_g$ is called the group velocity. In the more general case of the existence of a spectrum of frequencies, the superposition of waves leads to a narrow *wave packet* as sketched in figure 5.20(c). The group velocity can be calculated from

$$\mathbf{v}_g = \nabla_{\mathbf{k}}\omega = \frac{\partial\omega}{\partial k_x}\hat{\mathbf{k}}_x + \frac{\partial\omega}{\partial k_y}\hat{\mathbf{k}}_x + \frac{\partial\omega}{\partial k_z}\hat{\mathbf{k}}_z \tag{5.159}$$

The above derivation is by no means rigorous, but the concept of wave packets and group velocity is generally applicable to all waves. In the following, we will discuss two points related to the group velocity. One is whether the group velocity is always the velocity of energy flow. The other is the difference between the momentum of a wave packet and the crystal momentum.

The group velocity is usually considered to be the velocity of energy propagation of all carriers. We have shown this point in the above example for electromagnetic waves. The recognition of the significance of the group velocity actually started with the study of sound by Rayleigh (1945). It should also be pointed out, however, that the group velocity does not always represent the energy velocity. We can appreciate this from the step between eqs. (5.156) and (5.157), where we assumed that $\Delta\omega$ is much smaller than ω_0. In the case of a very large variation in the dispersion relation, the group velocity no longer represents the energy velocity. In figure 5.21, we show the real and the imaginary parts of the refractive index of silver. There is a region, called anomalous dispersion, in which the refractive index changes rapidly with wavelength. In this region, the real part of the refractive index is less than 1 and the phase velocity is larger than the speed of light. The group velocity is also larger than the speed of light. This does not mean a violation of the principle of relativity, however, which says that the maximum speed cannot be larger than the speed of light in free space. In this region, if we have a signal starting at $t = 0$ as shown in figure 5.19(a), one cannot superimpose all spectra of the light and still obtain a nice wave packet as sketched in figure 5.20(c). For such situations, the superimposed wave is more extended in space and the velocity of the majority of the energy propagation, called the signal velocity (Brillouin, 1960), is still less than the speed of light. A detailed discussion of this anomalous region is given by Bohren and Huffman (1983).

The group velocity of electrons is the velocity at which the electron wave packet moves in free space and inside a crystal. The energy dispersion of a free electron is

$$E = \frac{(\hbar\mathbf{k})^2}{2m} \tag{5.160}$$

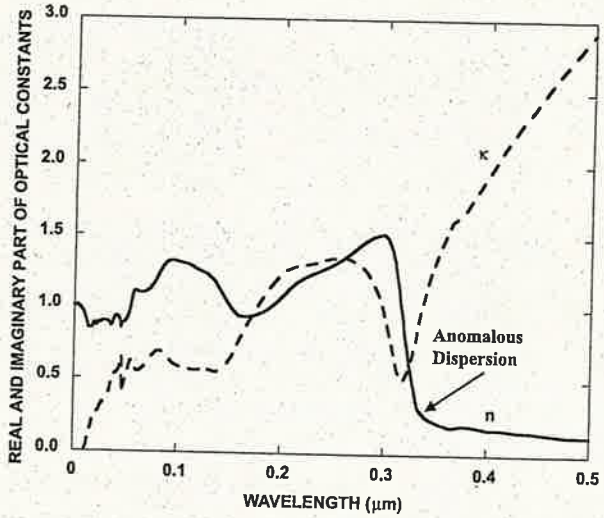


Figure 5.21 Refractive index of silver as a function of wavelength, showing the anomalous dispersion region in which both the phase velocity and the group velocity are larger than the speed of light. The signal velocity, however, is still smaller than the speed of light.

and thus the phase and group velocity are, respectively,

$$\mathbf{v} = \frac{E/\hbar}{\mathbf{k}} = \frac{\hbar\mathbf{k}}{2m} \text{ and } \mathbf{v}_g = \frac{\partial(E/\hbar)}{\partial\mathbf{k}} = \frac{\hbar\mathbf{k}}{m} \tag{5.161}$$

Clearly, the group velocity is consistent with the de Broglie relation $\mathbf{p} = \hbar\mathbf{k}$ and our classical relation $\mathbf{p} = m\mathbf{v}_g$, but not the phase velocity. When we deal with electron motion in crystals, however, $m\mathbf{v}_g$ does not normally equal $\hbar\mathbf{k}$, where $\mathbf{k}$ are the electron wavevectors determined from the von Karman boundary condition. We call $\hbar\mathbf{k}$ the crystal momentum and use it to satisfy the momentum conservation rules and to calculate the external force field, $\mathbf{F}_{\text{ext}}$

$$\mathbf{F}_{\text{ext}} = \frac{d(\hbar\mathbf{k})}{dt} \tag{5.162}$$

The reason for doing so is that the periodic potential also exerts another force on the electrons. When the crystal momentum is used, one can carry out the calculations as if electrons are not subject to the internal field of the crystal (Aschroft and Mermin, 1976; Slater, 1967). For such calculations, however, one still should use the group velocity as defined by eq. (5.159) as the actual speed of motion of the electrons, while using the crystal momentum for the external force and the momentum conservation rules. Similar arguments hold for phonons inside crystals.

5.6.2 Coherence and Transition to Particle Description

When should we consider the phase of the waves and when not? The answer to this question is fundamental for the transport of all these carriers and has been studied in

different disciplines. If the phase of the carriers must be considered, transport is coherent and the wave approaches illustrated in this chapter should be followed. In the other limit, the transport is incoherent. Transport in the incoherent regime will be treated in the following two chapters. In between the two limits is the partially coherent regime. Most engineering approaches for transport, built on diffusion equations, ignore the phase and treat carriers as incoherent particles. What are the conditions for these approaches to be valid?

Answers to these questions are by no means straightforward and vary with the types of carrier. For photons, the scattering is less frequent and mostly elastic; consequently, the discussion of coherence has been based more or less on the spectral purity (Born and Wolf, 1980). For electrons, inelastic scattering is strong and thus the discussion of coherence is closely related to scattering. There is less research on phonon coherence. Consequently, we will first discuss photon coherence and then electron coherence, followed by some discussion on phonons.

5.6.2.1 Coherence of Electromagnetic Waves

From eq. (5.156), we infer that the spatial spread of the wave packets in figure 5.20 is $\Delta x\, \Delta k \sim 2\pi$, or, denoting Δx as ℓ_c,

$$l_c \approx \frac{c}{\Delta \nu} \tag{5.163}$$

For electromagnetic wave propagation, this length is called the *coherence length* (Born and Wolf, 1980), which is inversely proportional to the effective bandwidth of the waves in the system. A ray of electromagnetic waves contains a series of wave packets fired by individual emitters in the source, as shown in figure 5.22(a). Each wave packet has a coherence length given by eq. (5.163). However, no phase relations exist between the wave packets.

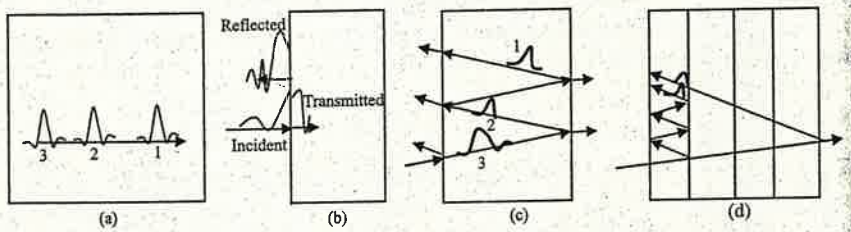


Figure 5.22 Traveling and interference of wave packets. (a) In a big domain, individual wave packets are uncorrelated and can be thought of as point particles. (b) At an interface, the tail of the wave packet and the reflected wave packet have a fixed relationship and thus can interfere with each other. (c) Inside a thick film, two wave packets can have transient interference but, since they do not have a fixed phase relationship, such transient interference can be anywhere inside the film. The end results are that no interference beats can be observed and thus geometrical optics should be used rather than wave optics. (d) In a periodic structure, however, the same wave packets are split many times at each interface and it is possible that the wave packets in a layer that are returned from other interfaces can overlap with the other wave packets inside the layer.

For blackbody thermal radiation, the energy uncertainty of the individual radiation emitters (atoms, electrons, or molecules) is of the order of $\kappa_B T$, due to the collision of the emitters with the reservoir, which also means that the effective bandwidth for thermal emission is $\kappa_B T/h$. Using this effective bandwidth and eq. (5.163), one can estimate that the coherence length is of the order of $hc/(\kappa_B T)$. A more detailed calculation of the coherence length leads to (Mehta, 1963)

$$\ell_c = 0.15\frac{hc}{\kappa_B T} \tag{5.164}$$

thermal length

This equation can be rewritten as

$$\ell_c T = 2167.8\ \mu\text{m K} \tag{5.165}$$

For reasons to be explained later, eq. (5.164) will also be called *thermal length*, reflecting the origin of this coherence length. Compared with Wien's displacement law, the coherence length of a blackbody radiation field is very close to the wavelength corresponding to the peak radiation intensity.

The coherence length, taken as a measure of the wave packet size, gives an indication of whether the phase information needs to be considered for transport processes or not. If the size of the transport domain is much larger than the wave packets or the coherence length, then the wave packets can be treated as point-wise particles traveling through the domain, as shown in figure 5.22(a). When a wave packet meets a perfect interface, however, it will be reflected and refracted. The reflected wave packet has a fixed phase relationship with the incoming one and can thus interfere with the incoming wave packet [figure 5.22(b)]. This is why we always use the Fresnel formula—the wave solution of the Maxwell equations—to calculate the reflectivity and transmissivity of a perfect interface. If multiple interfaces exist, as in a film, the reflected wave packet in the domain can encounter another incoming wave packet [figure 5.22(c)]. Although these two wave packets can create a transient interference when they overlap, their overlapping locations are not fixed because no fixed phase relationship exists between the two wave packets coming from a random thermal source. Because the number of random wave packets inside a ray is large, on average, we can ignore the phase relationship inside the domain and treat the wave packets as particles. Transport in this regime is called incoherent.

Consider now a thin film with two interfaces. If the size of a wave packet is small compared to the film thickness, transport is incoherent. In this regime, we can neglect the interference phenomenon discussed in section 5.3 and use energy superposition to calculate the reflectivity and transmissivity of the film. The energy superposition, also called ray tracing, is based on tracing the trajectory of photons and their intensity, rather than the electromagnetic fields, thus completely neglecting the phase information of the electromagnetic fields. The next two chapters will discuss such particle transport in more detail. The horizontal line in figure 5.23 gives the transmissivity of a non-absorbing thin film calculated on the basis of the energy superposition method, which is independent of the film thickness. In the other limit, when the wave packet is large compared to the film thickness, the same wave packet can overlap after experiencing multiple reflections; that is, the tail of one wave packet is still entering the film while the head has gone through multiple reflections. In such a case, consideration of the phase of the waves becomes

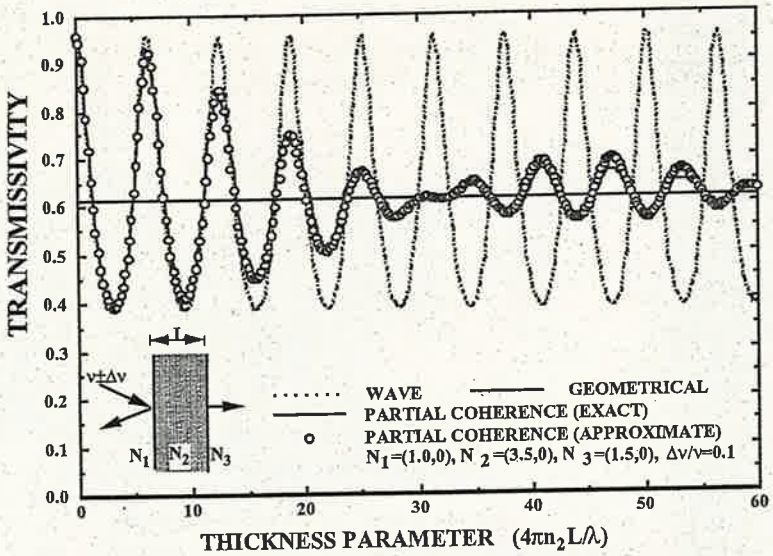


Figure 5.23 Transmissivity of a film subjected to polychromatic incident light as a function of the film thickness parameter calculated from different methods: wave approach, particle approach, and partial coherence theory. In the thin film limit, results from the partial coherence theory agree with wave approach. In the thick film limit, the theory agrees with the particle picture (Chen and Tien, 1992).

necessary and one can observe the interference phenomenon discussed in section 5.3 and shown in figure 5.23. If the number of overlapping reflections inside the film for the same wave packet is large enough, the solution obtained in section 5.3, which includes an infinite number of reflections for a single frequency (also called monochromatic), can be a good approximation, as shown in figure 5.23. In the intermediate case, when the wave packet is comparable to the film thickness, the same wave packet may overlap only partially inside the film or only within a few reflection cycles, and the transport falls into the partially coherent regime. In this regime, one can use the partial coherence theory for electromagnetic waves (Born and Wolf, 1980; Chen and Tien, 1992) to calculate the reflectivity and transmissivity. On the other hand, since a wave packet can be decomposed into the superposition of monochromatic waves, it is also permissible to calculate the reflectivity and transmissivity first for each frequency, using the wave formulation, and then to superimpose the results for all the frequencies to obtain the final results,

$$\bar{\tau} = \frac{J_t}{J_i} = \frac{\int_0^{\Delta\omega} \tau(\omega) J_i(\omega) d\omega}{\int_0^{\Delta\omega} J_i(\omega) d\omega} \text{ and } \bar{R} = \frac{J_t}{J_i} = \frac{\int_0^{\Delta\omega} R(\omega) J_i(\omega) d\omega}{\int_0^{\Delta\omega} J_i(\omega) d\omega} \tag{5.166}$$

where J_i is the incident photon Poynting vector and $\Delta\omega$ is the spectral width of the incident photon. Figure 5.23 shows results obtained from the above spectral averaging (marked as exact) and from a partial coherence formulation (marked as approximate) (Chen and Tien, 1992). The two approaches lead to the same results, which shows that when the film is thin, the single frequency formulation is approximately correct, and

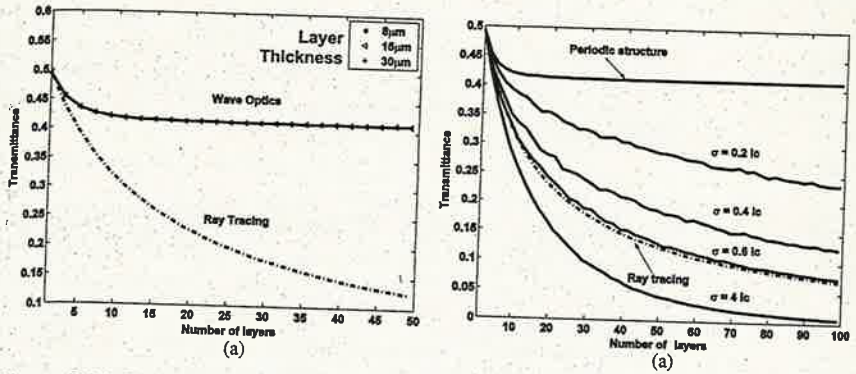


Figure 5.24 Directional and spectral averaged transmissivity of a Bragg reflector for a thermal radiation source at 1000 K. (a) Comparison between wave optics results and ray tracing results for perfect Bragg refectors at different period thickness, showing that for a thick period (thicker than the thermal length) the transmissivity does not depend on the period thickness, but the results from the two methods never agree with each other. (b) If the period varies randomly, the transmissivity decreases similarly to ray tracing but numerical values still do not always agree with those of the ray tracing method. Thus, random thickness variations do not justify the use of the particle picture. σ is standard deviation in thickness, and ℓ_c is the thermal wavelength given by eq. (5.165) (Hu et al., 2005).

when the film is thick, the particle picture should be used. For film thickness in between, in the partially coherent regime, either the spectral averaging method or a treatment based on partial coherence theory should be adopted.

The situation is more complicated if there are more interfaces. Consider a periodic structure such as a Bragg reflector, as shown in figure 5.22(d). A wave packet can experience multiple reflections in multiple layers and the coherence property of the wave packet can be altered. In a single-layer-thick film, for example, the multiple reflection of a narrow wave packet does not overlap with itself. In a multilayer structure, however, wave packets split at different interfaces have a chance of overlapping each other, as sketched in figure 5.22(d). Thus, in the case of Bragg reflectors, the spectral averaging of reflectivity and transmissivity based on wave optics and based on ray tracing lead to different results for a blackbody radiation source even with a period thickness much larger than the coherence length given by eq. (5.164). Figure 5.24(a) shows an example of transmissivity of a Bragg reflector for incident blackbody radiation at 1000 K, averaged over all frequencies and incident angles. The transmissivity, calculated from the wave method using the transfer matrix method, approaches a constant as the number of layers increases but is independent of the period thickness (for films with periods much larger than the blackbody coherence length). However, the results based on ray tracing continuously decrease with increasing number of periods. One can understand this phenomenon from the Kronig–Penney model for electron waves inside a periodic potential. In a periodic structure, the electron waves can extend over the whole structure, except in the bandgap region where no electrons exist at all. For these extended waves, the transmissivity is one even if the number of layers approaches infinity. On the other hand, if the phase of the waves is ignored, as in ray tracing, photons experience sequential reflection at every interface. As a result, the number of transmitted photons decrease

with increasing number of periods, as shown in figure 5.24(a). In another demonstration, it has been shown that the blackbody radiation passing through two pinholes is partially coherent even for a pinhole separation as large as a few centimeters, much longer than the coherence length of the blackbody radiation source (Mandel and Wolf, 1995; James and Wolf, 1991; Santarsiero and Gori, 1992).

The above discussion throws us into trouble. The coherence length is not a proper measure for neglecting the phase information of waves, as shown in the case of Bragg reflectors. In practice, however, the modeling of radiation transport through thick multilayers, such as windows, is often done with the ray tracing method. How we can justify the use of the particle picture? There are three possible justifications: (1) surface roughness; (2) nonparallel surfaces; and (3) thickness variations. All these factors create a certain randomness in the phase of the reflected and refracted waves. However, randomness in a structure does not necessarily lead to the particle picture.

As an example, we consider that the thickness of a Bragg reflector has a certain randomness. For this case, the transfer matrix method is still applicable. Lu et al. (2005) computed the transmissivity of Bragg reflectors with different level of randomness in the film thickness, as shown in figure 5.24(b). With randomness in the period thickness, the transmissivity does decrease with increasing number of periods as with ray tracing. However, the wave approach still does not agree with the ray tracing method. Depending on the degree of randomness, the transmissivity from wave optics can be either larger or smaller than the ray tracing results.

It turns out that the decreasing transmissivity in the case of random thickness variations is due to another phenomenon—the *wave localization*. When the phases have enough randomness, the superposition of waves can create complete cancellation of waves at certain frequencies for the one-dimensional superlattice considered here (Sheng, 1990). Figure 5.25 shows the transmissivity as a function of wavelength. It can be seen that the transmissivity of high-frequency waves is nearly zero. These waves do not propagate through the structure because of destructive interference among the waves. This phenomenon is called localization. In a three-dimensional structure, localization implies that waves are localized in some region and do not propagate. If a wave is localized, the transmissivity decreases exponentially with thickness of the structure. Thermal radiation, however, contains waves of many frequencies, not all of which can be localized. Thus the transmissivity does not exactly follow an exponential behavior. The phenomenon of localization was first studied for electron waves by Anderson (1958), and similar phenomena have been found for all kinds of waves, including photon and phonon waves (Sheng, 1990). The investigation of localization phenomena is still a very active area of research. It has been found that localization can easily occur in one-dimensional and two-dimensional structures, and when the number of modes is small. However, it is much more difficult to create localization in three-dimensional structures, particularly for electromagnetic waves (Garcia-Martin et al., 2000). The example of Bragg reflectors with random thickness variations is a simple one-dimensional case and thus it is relatively easy to observe the localization phenomenon, as the computed transmissivity in figures 5.24(b) and 5.25 shows. Surface roughness, however, makes the wave three-dimensional. In these situations, it is likely that the phase of the waves can be ignored and the particle picture can be used, provided the coherence length is smaller than the characteristic length of the surface randomness.

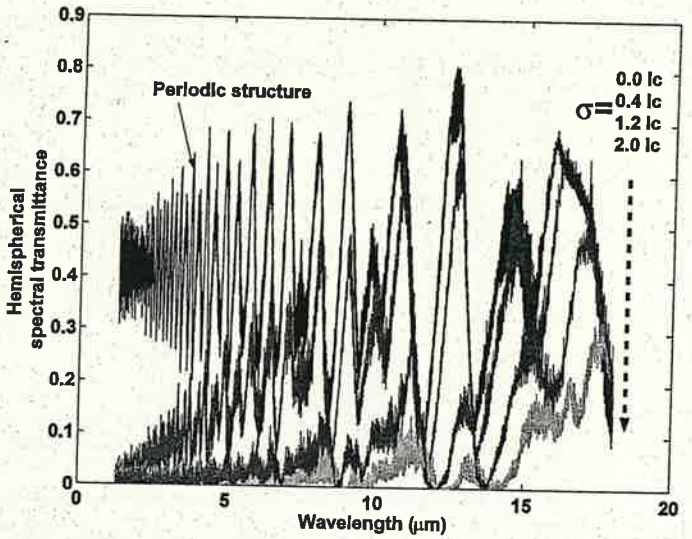


Figure 5.25 Photon transmissivity through a random Bragg reflection, showing that high-frequency photons are readily localized (Hu et al., 2005). The four curves have increasing thickness variations, σ, measured in terms of coherence length, as indicated by the dashed arrow.

5.6.2.2 Coherence of Electron Waves

The coherence of electron waves is an important issue for electron transport. One major difference between electrons and photons is the scattering mechanisms. Photon scattering is mostly elastic; that is, the wavelength and energy of the scattered photons are the same as that of the incoming photons and a fixed phase relationship between the incoming and scattered waves exists. The scattering of electrons can be *elastic*, in which the electrons merely change direction but have the same energy before and after the scattering, or *inelastic*, in which both the direction and the energy of the electrons are changed. The scattering of electrons by impurities and at the boundaries is elastic. The elastic scattering itself does not destroy the phase but the random locations of the impurities and the surface roughness may create enough randomness in the phase such that the particle approach is approximately valid. In other cases, the randomness can also create localization of the electron waves. On the other hand, inelastic scattering, such as electron–phonon scattering, randomizes the phases because the location and the phase of the electron–phonon scattering change all the time. These scattering events completely destroy the phase. If one uses the wave approach without proper consideration of the phase-destroying inelastic scattering processes, the results will be wrong.

Thus, electron transport has used a different set of definitions from those of photon transport (Ferry and Goodnick, 1997). Three major length scales that are often used in describing electron transport are the *mean free path*, the *phase coherence length*, and the *thermal length*. The mean free path is a measure of the average distance between successive scattering events and is $\Lambda \approx v_F \tau$, where v_F is the electron velocity at the Fermi level, or the Fermi velocity, and τ is the relaxation time, that is, the average

time between successive collisions. The approximation sign is used because the Fermi velocity is only an approximation to the average electron velocity; this approximation works best in metals and in heavily doped semiconductors.* Not all the scattering events governing the mean free path are phase destroying. The inelastic scattering mean free path is $\Lambda_{in} = v_F \tau_\varphi$, where τ_φ is the inelastic relaxation time or phase-breaking time. The phase coherence length, also called the Thouless length, is defined as $\Lambda_\varphi = (a\tau_\varphi)^{1/2}$, where a is the electron diffusivity, which we will discuss more in the next chapter, and $a \approx v_F^2 \tau$. Because the relaxation time is used in the diffusivity, Λ_φ is slightly different from Λ_{in}. The use of diffusivity in the definition of phase coherence length implies that electrons may experience multiple elastic scattering, that is, diffusion, during the phase-breaking time. Typically, we have $\Lambda < \Lambda_\varphi < \Lambda_{in}$.

In addition to these length scales, there is also another length scale that is related to the thermal broadening of the energy levels of electrons. As in the discussion of the coherence length of blackbody radiation, the thermal broadening in energy is of the order of $\kappa_B T$. Thus, according to the Heisenburg uncertainty principle, the corresponding uncertainty in time is $h/(\kappa_B T)$. The thermal length is defined as $\Lambda_T = (a\hbar/\kappa_B T)^{1/2}$. Comparing this thermal length with the photon coherence length, eq. (5.164), the thermal length here is defined based on the diffusion transport, with the diffusion length given as $(a\tau_c)^{1/2}$, where τ_c is the characteristic time. The photon coherence length given by eq. (5.164) is based on the ballistic transport of photons of different energy spreading over $\kappa_B T$, with the transport length given by $v\tau_c$. Both lengths are a measure of the thermal spreading in the energy (wavelength) of the energy carriers, and thus are fundamentally similar concepts. This is the reason that we also call the coherence length given by eq. (5.164) the thermal length.

The phase coherence length Λ_φ and the thermal length Λ_T are usually used in judging whether transport is in the wave regime or the particle regime. If $\Lambda_T > \Lambda_\varphi$, the inelastic scattering is considered as the dominant phase-destroying process. Under this condition, if the structure characteristic length, such as the diameter of a nanowire or the width of a quantum well, is larger than Λ_φ, quantum states, as predicted by simple quantum well and quantum wire models in chapter 2, cannot be created because of the loss of phase correlation of the electron waves. The transport should be treated with the particle approach for such situations. If $\Lambda_\varphi > \Lambda_T$, thermal excitation is often considered as the dominant dephasing mechanism. Under this condition, if the structure characteristic length is much larger than Λ_T, it is often thought that the particle treatment leads to the same results as that of the wave approach. However, as pointed out before for photons, the wave and the particle approaches lead to the same results for simple geometries only. For periodic multilayer structures such as superlattices, the particle and the wave treatments do not lead to the same results, as explained in figures 5.22(c) and 5.24.

5.6.2.3 Coherence of Phonons

Phonon coherence from a transport point of view is the least considered one among electrons, photons, and phonons. The discussions on photons and electrons, however, shed some light on the coherence issues of phonons. Scattering of phonons by boundaries and impurities is elastic and thus not phase destroying. Consequently, these scattering centers can either lead to localization or make the particle picture a good approximation. Because the dominant phonon wavelength is typically very short, ~10–20 Å at room temperature (Chen, 1997), interface roughness of the same order or much larger exists at most material interfaces or boundaries, and thus the particle-based treatment is likely to be valid for most practical situations. Phonon–phonon scattering, which is dominant in most materials at room temperature—a topic we will discuss in the next chapter—is inelastic. The mean free path of such scattering processes can be long, however. For example, in silicon, the estimated mean free path is 2500–3000 Å (Chen, 1998; Ju and Goodson, 1999). For structures much larger than the phonon–phonon scattering mean free path, the phase of phonons must be ignored. For structures smaller than the mean free path, rough interface scattering most likely can justify the particle treatment because of the short phonon wavelength, as mentioned before. The particle approach is particularly useful when the detailed interface structures are not clear and thus exclude a full-scale wave treatment.

One can also estimate the thermal coherence length on a basis similar to that for photons and electrons (Chen, 1997). Using an energy spread of $\kappa_B T$, we obtain the phonon thermal wavelength, defined as $v/\Delta\nu$, as $\sim vh/(\kappa_B T)$. At room temperature, taking a typical value of $v \sim 5000\ \mathrm{m\,s^{-1}}$, we get a thermal length of 10 Å. Such a thermal length, although useful as an indicator of width of the wave packet due to thermal spreading, cannot be applied to periodic structures such as superlattices (Chen, 1999) in deciding whether wave effect can be neglected, as in the case of the application of eq. (5.165) to Bragg reflectors.

The thermal conductivity of superlattices is a good example to illustrate the coherence issues related to phonon transport. It has been experimentally observed that the thermal conductivities of superlattices are significantly reduced in comparison with values obtained from the Fourier heat conduction law using the bulk properties of each layer (Yao, 1987; Chen et al., 1994; Lee et al., 1997; Capinski et al., 1999). The mechanisms of the thermal conductivity reduction, however, have been under debate (Chen, 2001; Yang and Chen, 2003). One approach is based on treating phonons as particles, with the phonons in each individual layer having their bulk properties but experiencing incoherent interface scattering. The other approach is based on treating superlattices as a new crystal structure with a unit cell spanning over one period of the superlattice, that is, treating phonons as coherent waves extending over the whole structure. The particle approach assumes that interface scattering destroys the phase. Particle-based model can fit experimental data based on the assumption of how many phonons are diffusely scattered at the interfaces. The coherent phonon wave approach, based on a pure harmonic lattice dynamics model, assumes that the reduction in thermal conductivity is caused by the phonon spectrum change and the associated reduction in group velocity, as shown in figure 3.30. The ideal lattice dynamics model, however, cannot predict the same order of thermal conductivity reduction as is experimentally observed. Recently, it has been shown that a lattice dynamics model based on damped lattice waves, that is, lattice waves that do not extend over the whole superlattice but can exist in one layer or over a few periods, can capture the same trends as experimental data, as shown in figure 5.26 (Yang and Chen, 2003). Such damped waves are created by introducing a complex wavevector into eq. (3.43) (Simkin and Mahan, 2000). The imaginary part of the wavevector, for example, can be due to the loss of coherence resulting from diffuse interface scattering

*For non-degenerate semiconductors, that is, semiconductors with the chemical potential lying inside the bandgap, v_F should be replaced by the thermal velocity $v_t = (3\kappa_B T/m^*)^{1/2}$, where m^* is the electron effective mass.

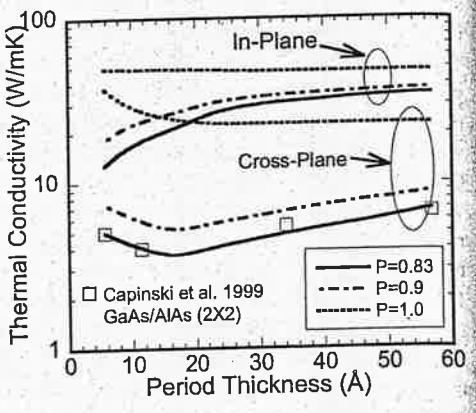


Figure 5.26 Thermal conductivity of superlattices obtained from a lattice dynamics model with damped lattice waves (Yang and Chen, 2003). The damping is determined by the interface specularity parameter p, representing the fraction of specularly scattered phonons. In the thin period limit, the results represent coherent transport whereas, in the thick period limit, the coherence is lost and the results represent the particle transport regime.

(Yang and Chen, 2003). In figure 5.26, the interface specularity parameter p represents the fraction of specularly scattered phonons, which are assumed to be coherent. The other phonons are diffusely scattered and are assumed to be incoherent. We first examine the case $p = 1$, that is, all phonons are specularly reflected and waves extend through the whole superlattice. In this case, the thermal conductivity is independent of the period thickness until the period is only 1–5 monolayers, in which regime the cross-plane thermal conductivity actually increases with decreasing film thickness. This recovery in thermal conductivity is due to phonon tunneling, as shown in figure 5.14. The fact that above about 10 Å the thermal conductivity does not change with thickness is related to thermal broadening. As we indicated earlier, the thermal length is also about 10 Å. However, in this case, the particle approach would lead to different results (Chen, 1999) and thus the wave and the particle approaches do not agree with each other. When p is less than one, the phonon waves are damped in the superlattice due to diffuse scattering. When the period thickness is large, the phonon waves are not coherent over many periods and thus the spectra calculated from lattice dynamics with imaginary wavevectors are close to those of bulk phonons. In this case, the thermal boundary resistances at interfaces dominate the thermal conductivity until the period becomes much larger than the mean free path in the bulk material, for which the thermal conductivity eventually approaches the predictions of the Fourier law (Chen, 1998). In the thin period limit, the superposition of coherent phonons extending over many periods leads to new phonon band structures and, correspondingly, wave phenomena such as stop bands, interference, and tunneling all contribute to the thermal conductivity behavior.

5.7 Summary of Chapter 5

This chapter discussed the wave picture of energy transport and the transition from the wave to the particle description. The purpose of section 5.1 was to familiarize the readers with various forms of waves including electromagnetic waves, acoustic waves, and material waves. The electromagnetic waves are governed by the Maxwell equations.

Acoustic waves, which are long-wavelength phonons, are described by the acoustic field equations or the Christoffel equation. Material waves are described by the Schrödinger equation. Solutions of the wave equations lead to the fields at each point as a function of time. The energy flux associated with each wave is usually a product of various fields, as given by the Poynting vector for electromagnetic and acoustic waves, and the particle flux expression for material waves. Although these waves are described by different governing equations, the key point is that all forms of waves share similar behavior, as is clearly demonstrated in the following four sections. Although the material presented in this chapter is diverse, some readers may be familiar with one or several forms of these waves and can understand other forms of waves by analogy.

At an interface, all waves experience the phenomena of reflection and refraction. A Snell-law type of relation governs the angles of incidence and transmission. The reflection and transmission coefficients, which are called Fresnel coefficients for electromagnetic waves, can be obtained by applying the appropriate boundary conditions for each type of wave. The expressions for these coefficients are quite similar among different types of waves. From the reflection and transmission coefficients, one can calculate the reflectivity and transmissivity of energy or particle flux. Several special cases for reflection and transmission of waves at one interface are of great importance. One example is total reflection, which occurs when the refractive index or the acoustic impedance of the medium at the incident side is larger than that at the transmission side for optical and acoustic waves, respectively, or the potential barrier is higher than the energy of the incident material waves. When total reflection occurs, an evanescent wave exists that extends into the second medium. The time-averaged energy or particle flux into the second medium carried by the evanescent wave is zero but the instantaneous field and energy are not zero. Thermal boundary resistance between two perfect solids is due to the reflection of phonons at the interface.

When multiple interfaces exist, superposition of waves due to reflection at multiple interfaces creates the familiar interference and tunneling phenomena in thin films. For multidimensional problems, which we did not discuss in this chapter here, the superposition of scattered waves leads to diffraction phenomena. We introduced the transfer matrix method for calculating the reflection and transmission coefficients of multilayers, which is valid for both interference and tunneling regimes. Interference gives the familiar oscillation of reflectivity and transmissivity of optical coatings as a function of the film thickness, and affects the thermal radiative properties of thin films and multilayers. In multilayer structures, particularly periodic structures, interference leads to the formation of stop-bands, which corresponds to the formation of gaps in the energy spectrum of electrons, phonons, and photons as discussed in chapter 3. Tunneling of evanescent waves that exist near the interface under appropriate conditions can occur when a third medium is brought close to the interface, before the evanescent wave significantly decays in the second medium, and when the third medium allows the propagation of the wave. The tunneling phenomenon is the basis of several recent inventions such as scanning tunneling microscopy for electrons and photons. It also occurs for acoustic waves and may affect heat conduction.

Given the transmissivity of heat carriers through two points of a system, we can calculate the net heat transfer (or other fluxes of interest) between the two points, using the Landauer formalism as manifested by eqs. (5.88) and (5.153). The Landauer formalism views transport as a transmission process. The net flux (energy or particle) between any

two points A and B is the difference between the corresponding flux transmitted from A to B and that from B to A. The principle of detailed balance can be used to write the final flux in terms of the properties of one side (or one point) only, together with the transmissivity.

Calculation of energy or particle transport under the wave picture is often tedious and requires mathematical manipulation of the field quantities. In section 5.6, we discussed under what conditions we can neglect the phase information and treat energy carriers as particles. First, we demonstrated that the superposition of monochromatic waves leads to wave packets that propagate at the group velocity rather than the phase velocity. This group velocity is normally the velocity at which energy is propagating, but in a highly dispersive medium the group velocity is not necessarily the energy propagation velocity. The width of these wave packets is the coherence length, which is inversely proportional to the inverse effective spectrum width (or energy spread) of the carriers. If the coherence length is long compared to the structural characteristic length, the wave picture should be used. In the opposite limit, however, we should be more careful. We can treat the transport as particles as long as the wave packets split from the same original one, for example, through reflection at an interface, do not overlap at the same place and the same time. This often happens when the structural size is large compared to that of the wave packets. However, in periodic structures, such as Bragg reflectors and superlattices, the wave packets reflected at different layers can merge and still overlap. Consequently, the particle approach and wave approach do not agree with each other. Elastic scattering, caused by inhomogeneities such as impurities and interface roughness, does not destroy the coherence of the waves. The random elastic scattering can potentially lead to two effects. One is localization, for which the waves are localized and do not propagate. Localization is generally easier to observe in low-dimensional structures than in three-dimensional structures. The other effect is that random scattering and the subsequent superposition of scattered waves usually leads to results that are close to those obtained from the particle treatment. When the exact locations of the scattering centers and surface topology are not known, which is usually the case, the particle treatment leads to better agreement with experimental results. Inelastic scattering completely destroys the phase. When the structure characteristic length is much larger than the inelastic scattering mean free path, or, for electrons, the phase coherence length, the particle treatment is mandatory.

5.8 Nomenclature for Chapter 5

a	electron diffusivity, $\mathrm{m\,s^{-2}}$
$\mathbf{A}$	amplitude and direction of field
$\mathbf{B}$	magnetic induction, $\mathrm{N\,s\,m^{-1}\,C^{-1}}$
c_0	speed of light in vacuum, $\mathrm{m\,s^{-1}}$
d	film thickness, m
$\mathbf{D}$	electric displacement, $\mathrm{C\,m^{-2}}$
E	allowed energy level, J
$\mathbf{E}$	electric field, $\mathrm{N\,C^{-1} = V\,m^{-1}}$
f	probability distribution function
$\mathbf{F}$	vector wave field; force, N
h	Planck constant, J s
$\hbar$	Planck constant divided by 2π, J s
$\mathbf{H}$	magnetic field, $\mathrm{C\,m^{-1}\,s^{-1} = A\,m^{-1}}$
i	imaginary number unit, $\sqrt{-1}$
$\mathbf{J}$	flux of particles, $\mathrm{s^{-1}\,m^{-2}}$
$\mathbf{J}_a$	acoustic wave power flux, $\mathrm{W\,m^{-2}}$
$\mathbf{J}_e$	current density, $\mathrm{C\,s^{-1}\,m^{-2}}$
$\mathbf{J}_s$	surface current density, $\mathrm{A\,m^{-1}}$
k	magnitude of wavevector, $\mathrm{m^{-1}}$
$\mathbf{k}$	wavevector, $\mathrm{m^{-1}}$
$\hat{\mathbf{k}}$	unit vector along wavevector direction
l	integer
ℓ_c	coherence length, m
m	mass, kg; integer
n	real part of complex refractive index N
N	complex refractive index or complex optical constant
p	surface impedance, $\mathrm{C^2\,J^{-1}\,s^{-1}}$
$\mathbf{P}$	polarization per unit volume, $\mathrm{C\,m^{-2}}$
q	magnitude of heat flux, $\mathrm{W\,m^{-2}}$
$\mathbf{q}$	heat flux vector, $\mathrm{W\,m^{-2}}$
Q	charge, C
r	reflection coefficient
$\mathbf{r}$	position vector
R	reflectivity
$\mathbf{S}$	Poynting vector, $\mathrm{W\,m^{-2}}$
$\overline{\overline{S}}$	strain tensor
t	time, s; or transmission coefficient
T	temperature, K
T_e	temperature of phonons coming toward interface, i.e., temperature of emitted phonons, K
$\mathbf{u}$	displacement, m
U_0	potential barrier height, J
v_F	Fermi velocity, $\mathrm{m\,s^{-1}}$
$\mathbf{v}$	velocity, $\mathrm{m\,s^{-1}}$

$\mathbf{v}_g$	group velocity, $\mathrm{m\,s^{-1}}$
Z	acoustic impedance, $\mathrm{kg\,m^{-2}\,s^{-1}}$
α	absorption coefficient, $\mathrm{m^{-1}}$
δ	skin depth, m
$\Delta\nu$	spectral width, $\mathrm{s^{-1}}$
ε	electric permittivity, $\mathrm{C^2\,N^{-1}\,m^{-2} = F\,m^{-1}}$
$\hat{\varepsilon}$	complex electric permittivity, $\mathrm{C^2\,N^{-1}\,m^{-2}}$
ε_0	electric permittivity of vacuum, $\mathrm{C^2\,N^{-1}\,m^{-2}}$
ε_r	dielectric constant
θ	angle, rad
κ	imaginary part of complex refractive index
κ_B	Boltzmann constant, $\mathrm{J\,K^{-1}}$
λ	wavelength, m
λ_L	Lamb constant, $\mathrm{N\,m^{-2}}$
Λ	mean free path, m
Λ_{in}	inelastic scattering mean free path, m
Λ_T	thermal wavelength, m
Λ_φ	phase coherence length, m
μ	magnetic permeability, $\mathrm{N\,s^2\,C^{-2}}$
μ_L	Lamb constant, $\mathrm{N\,m^{-2}}$
ν	frequency of phonons and photons, $\mathrm{s^{-1}}$
ξ	electric polarizibility
ρ	net charge density, $\mathrm{C\,m^{-3}}$
ρ_s	surface charge density, $\mathrm{C\,m^{-2}}$
σ_e	electrical conductivity, $\mathrm{\Omega^{-1}\,m^{-1}}$
τ	transmissivity
ϕ	azimuthal angle, rad
φ	phase factor
χ	electric susceptibility
Ψ	wavefunction
ω	angular frequency, rad.Hz
$\Re''$	specific thermal boundary resistance, $\mathrm{K\,m^2\,W^{-1}}$

Subscripts

0	vacuum
1, 2	medium 1 or medium 2
12	from medium 1 into medium 2
21	from medium 2 into medium 1
a	amplitude
c	complex
e	based on emitted phonon temperature
i	incident wave
L	longitudinal, or Lamb constant
r	reflected wave
s	surface
t	transmitted wave
T	transverse wave
x, *y*, *z*	Cartesian components

Superscripts

*	complex conjugate
=	second-order tensor
-	average

5.9 References

Anderson, P.W., 1958, "The Absence of Diffusion in Certain Random Lattices," *Physical Review*, vol. 109, pp. 1492–1505.

Angelescu, D.E., Cross, M.C., and Roukes, M.L., 1998, "Heat Transport in Mesoscopic Systems," *Superlattices and Microstructures*, vol. 23, pp. 673–689.

Ashcroft, N.W., and Mermin, N.D., 1976, *Solid State Physics*, Saunders College Publishing, Fort Worth, TX.

Auld, B.A., 1990, *Acoustic Fields and Waves in Solids, I, II*, 2nd ed., Krieger, FL.

Binnig, G., and Rohrer, H., 1982, "Scanning Tunneling Microscopy," *Helvetica Physica Acta*, vol. 55, pp. 726–735.

Binnig, G., Quate, C.F., and Gerber, C., 1986, "Atomic Force Microscope," *Physical Review Letters*, vol. 56, pp. 930–933.

Bohren, C.F., and Huffman, D.R., 1983, *Absorption and Scattering of Light by Small Particles*, Wiley, New York, pp. 235–239.

Born, M., and Wolf, E., 1980, *Principles of Optics*, 6th ed., Pergamon Press, Oxford, chapters 1 and 10.

Brillouin, L., 1960, *Wave Propagation and Group Velocity*, Academic Press, New York.

Capinski, W.S., Maris, H.J., Ruf, T., Cardona, M., Ploog, K., and Katzer, D.S., 1999, "Thermal-Conductivity Measurements of GaAs/AlAs Superlattices Using a Picosecond Optical Pump-and-Probe Technique," *Physical Review B*, vol. 59, pp. 8105–8113.

Chen, G., 1996, "Wave Effects on Radiative Transfer in Absorbing and Emitting Thin-Film Media," *Microscale Thermophysical Engineering*, vol. 1, pp. 215–224.

Chen, G., 1997, "Size and Interface Effects on Thermal Conductivity of Superlattices and Periodic Thin-Film Structures," *Journal of Heat Transfer*, vol. 119, pp. 220–229.

Chen, G., 1998, "Thermal Conductivity and Ballistic Phonon Transport in Cross-Plane Direction of Superlattices," *Physical Review B*, vol. 57, pp. 14958–14973.

Chen, G., 1999, "Phonon Wave Effects on Heat Conduction in Thin Films and Superlattices," *Journal of Heat Transfer*, vol. 121, pp. 945–953.

Chen, G., 2001, "Phonon Transport in Low-Dimensional Structures," *Semiconductors and Semimetals*, vol. 71, pp. 203–259.

Chen, G., 2003, "Diffusion–Transmission Interface Conditions," *Applied Physics Letters*, vol. 82, pp. 991–993.

Chen, G., and Tien, C.L., 1992, "Partial Coherence Theory of Thin Film Radiative Properties," *Journal of Heat Transfer*, vol. 114, pp. 636–643.

Chen, G., and Zeng, T., 2001, "Nonequilibrium Phonon and Electron Transport in Heterostructures and Superlattices," *Microscale Thermophysical Engineering*, vol. 5, pp. 71–88.

Chen, G., Tien, C.L., Wu, X., and Smith, J.S., 1994, "Measurement of Thermal Diffusivity of GaAs/AlGaAs Thin-Film Structures," *Journal of Heat Transfer*, vol. 116, pp. 325–331.

Cohen-Tannoudji, C., Diu, B., and Laloe, F., 1977, *Quantum Mechanics*, vol. I, Wiley, New York.

Costescu, R.M., Wall, M.A., and Cahill, D.G., 2003, "Thermal Conductance of Epitaxial Interfaces," *Physical Review B*, vol. 67, pp. 054302/1–5.

Cravalho, E.G., Tien, C.L., and Caren, R.P., 1967, "Effect of Small Spacing on Radiation Transfer between Two Dielectrics," *Journal of Heat Transfer*, vol. 89, pp. 351–358.

Dames, C., and Chen, G., 2004, "Theoretical Phonon Thermal Conductivity of Si/Ge Superlattice Nanowires," *Journal of Applied Physics*, vol. 95, pp. 682–693.

DiMatteo, R.S., Greiff, P., Finberg, S.L., Young-Waithe, K.A., Choy, H.K.H., and Masaki, M.M., 2001, "Enhanced Photogeneration of Carriers in a Semiconductor Via Coupling Across a Nonisothermal Nanoscale Vacuum Gap," *Applied Physics Letters*, vol. 79, pp. 1894–1896.

Domoto, G.A., Boehm, R.F., and Tien, C.L., 1970, "Experimental Investigation of Radiative Transfer between Metallic Surfaces at Cryogenic Temperatures," *Journal of Heat Transfer*, vol. 92, pp. 412–417.

Esaki, L., 1958, "New Phenomenon in Narrow Germanium p–n Junctions," *Physical Review*, vol. 109, pp. 603–604.

Esaki, L., and Tsu, R., 1970, "Superlattice and Negative Differential Conductivity in Semiconductors," *IBM Journal of Research and Development*, vol. 14, pp. 61–65.

Ferry, D.K., and Goodnick, S.M., 1997, *Transport in Nanostructures*, Cambridge University Press, Cambridge, UK.

Garcia-Martin, A., and Niero-Vesperinas, M., Sáenz, J.J., 2000, "Spatial Field Distributions in the Transition from Ballistic to Diffusive Transport in Randomly Corrugated Waveguides," *Physical Review Letters*, vol. 84, pp. 3578–3581.

Hargreaves, C.M., 1969, "Anomalous Radiative Transfer between Closely-Spaced Bodies," *Physics Letters*, vol. 30A, pp. 491–492.

Hu, L., Schmidt, A., Narayanaswamy, A., and Chen, G., 2005, "Effects of Periodic Structure on Coherence Properties of Blackbody Radiation," *Journal of Heat Transfer*, in press.

Huang, K.-M., and Wu, G.Y., 1992, "Transfer-Matrix Theory of the Energy Levels and Electron Tunneling in Heterostructures under an In-Plane Magnetic Field," *Physical Review B*, vol. 45, pp. 3461–3464.

Imry, Y., and Landauer, R., 1999, "Conductance Viewed as Transmission," *Review of Modern Physics*, vol. 71, pp. S306–S312.

James, D.F.V., and Wolf, E., 1991, "Spectral Changes Produced in Young's Interference Experiment," *Optics Communications*, vol. 81, pp. 150–154.

Ju, Y.S., and Goodson, K.E., 1999, "Phonon Scattering in Silicon Films of Thickness Below 100 nm," *Applied Physics Letters*, vol. 74, pp. 3005–3007.

Kapitza, P.L., 1941, "The Study of Heat Transfer in Helium II," *J. Phys.* (USSR) vol. 4, pp. 181–210; in *Collected Papers of P.L. Kapitza*, vol. 2, p. 581, ed. der Haar, D., 1965, Pergamon, Oxford.

Katerberg, J.A., Reynolds, C.L., Jr., and Anderson, A.C., 1977, "Calculations of the Thermal Boundary Resistance," *Physical Review B*, vol. 16, pp. 672–679.

Knittl, Z., 1976, *Optics of Thin Films*, Wiley, New York, chapter 4.

Koyama, F., Kinoshita, S., and Iga, K., 1989, "Room-Temperature Continuous Wave Lasing Characteristics of a GaAs Vertical-Cavity Surface-Emitting Laser," *Applied Physics Letters*, vol. 55, pp. 221–222.

Lee, S.M., Cahill, D.G., and Venkatasubramanian, R., 1997, "Thermal Conductivity of Si–Ge Superlattices," *Applied Physics Letters*, vol. 70, pp. 2957–2959.

Little, W.A., 1959, "The Transport of Heat between Dissimilar Solids at Low Temperatures," *Canadian Journal of Physics*, vol. 37, pp. 334–349.

Majumdar, A., Lai, J., Chandrachood, M., Nakabeppu, O., and others, 1995, "Thermal Imaging by Atomic Force Microscopy Using Thermocouple Cantilever Probes," *Review of Scientific Instruments*, vol. 66, pp. 3584–3592.

Mandel, L., and Wolf, E., 1995, *Optical Coherence and Quantum Optics*, Cambridge University Press, Cambridge, UK, pp. 307–318.

Mehta, C.L., 1963, "Coherence-Time and Effective Bandwidth of Blackbody Radiation," *Nuovo Cimento*, vol. 21, pp. 401–408.

Mulet, J.-P, Joulain, K.L., Carminati, R., and Greffet, J.-J., 2002, "Enhanced Radiative Heat Transfer at Nanometric Distances," *Microscale Thermophysical Engineering*, vol. 6, p. 209–222.

Narayanamurti, V., Stormer, J.L., Chin, M.A., Gossard, A.C., and Wiegmann, W., 1979, "Selective Transmission of High-Frequency Phonons by a Superlattice: the 'Dielectric' Phonon Filter," *Physical Review Letters*, vol. 43, pp. 2012–2016.

Narayanaswamy, A., and Chen, G., 2003, "Surface Modes for Near-Field Thermophotovoltaics," *Applied Physics Letters*, vol. 82, pp. 3544–3546.

Narayanaswamy, A., and Chen, G., 2004, "Direct Computation of Thermal Emission from Nanostructures," *Annual Review of Heat Transfer*, vol. 14, in press.

Nayfeh, A.H., 1995, *Wave Propagation in Layered Anisotropic Media*, Elsevier, Amsterdam.

Nulman, J., 1989, "Emissivity Issues in Pyrometric Temperature Monitoring for RTP Systems," *SPIE*, vol. 1189, *Rapid Thermal Processing*, pp. 72–82.

Odom, T.W., Huang, J., Kim, P., and Lieber, C.M., 1998, "Atomic Structure and Electronic Properties of Single-Walled Carbon Nanotubes," *Nature*, vol. 391, pp. 62–64.

Palik, E., 1985, *Handbook of Optical Constants of Solids*, Academic Press, San Diego, CA.

Pendry, J.B., 1996, "Calculating Photonic Band Structure," *Journal of Physics–Condensed Matter*, vol. 8, pp. 1085–1108.

Pendry, J.B., 1999, "Radiative Exchange of Heat between Nanostructures," *Journal of Physics—Condensed Matter*, vol. 11, pp. 6621–6633.

Pendry, J.B., 2000, "Negative Refraction Makes a Perfect Lens," *Physical Review Letters*, vol. 85, pp. 3966–3969.

Polder, D., and Van Hove, M., 1971, "Theory of Radiative Heat Transfer between Closely Spaced Bodies," *Physical Review B*, vol. 4, pp. 3303–3314.

Raether, H., 1987, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer-Verlag, Berlin.

Rayleigh, J.W.S., 1945, *The Theory of Sound*, Dover, New York.

Reddick, R.C., Warmack, R.J., and Ferrell, T.L., 1989, "New Form of Scanning Optical Microscopy," *Physical Review B*, vol. 39, pp. 767–770.

Rytov, S.M., Kravtsov, Y. A., and Tatarski, V.I., 1987, *Principles of Statistical Radiophysics*, vol. 3, Springer-Verlag, Berlin.

Santarsiero, M., and Gori, F., 1992, "Spectral Change in Young Interference Pattern," *Physics Letters A*, vol. 167, p. 123.

Schwab, K., Henriksen, E.A., Worlock, J.M., and Roukes, M.L., 2000, "Measurement of the Quantum of Thermal Conductance," *Nature*, vol. 404, pp. 974–977.

Shelby, R.A., Smith, D.R., and Schultz, S., 2001, "Experimental Verification of a Negative Index of Refraction," *Science*, vol. 292, pp. 77–79.

Sheng, P., ed., 1990, *Scattering and Localization of Classical Waves in Random Media*, World Scientific, Singapore.

Siegel, R., and Howell, J., 1972, *Thermal Radiation Heat Transfer*, 3rd ed., Hemisphere, Washington, DC, p. 928.

Simkin, M.V., and Mahan, G.D., 2000, "Minimum of Thermal Conductivity of Superlattices," *Physical Review Letters*, vol. 84, pp. 927–930.

Slater, J.C., 1967, *Quantum Theory of Molecules and Solids*, vol. 3, McGraw-Hill, New York, chapter 2 and appendix 1.

Stoner, R.J., and Maris, H.J., 1993, "Kapitza Conductance and Heat Flow between Solids at Temperatures from 50 to 300 K," *Physical Review B*, vol. 48, p. 16373.

Swartz, E.T., and Pohl, R.O., 1989, "Thermal Boundary Resistance," *Reviews of Modern Physics*, vol. 61, pp. 605–668.

Tamura, S., Tanaka, Y., and Maris, H.J., 1999, "Phonon Group Velocity and Thermal Conduction in Superlattices," *Physical Review B*, vol. 60, pp. 2627–2630.

Tien, C.L., and Cunnington, G.R., 1973, "Cryogenic Insulation Heat Transfer," *Advances in Heat Transfer*, vol. 9, pp. 349–417.

Tsu, R., and Esaki, L., 1973, "Tunneling in a Finite Superlattice," *Applied Physics Letters*, vol. 22, pp. 562–564.

Valanju, P.M., Walser, R.M., and Valanju, A.P., 2002, "Wave Refraction in Negative-Index Media: Always Positive and Very Inhomogeneous," *Physical Review Letters*, vol. 88, pp. 187401–1 to 187401 (1–4).

Venkatasubramanian, R., 2000, "Lattice Thermal Conductivity Reduction and Phonon Localization Phenomena in Superlattices," *Physical Review B*, vol. 61, pp. 3091–3097.

Veselago, V.G., 1968, "Electrodynamics of Substances with Simultaneously Negative Electrical and Magnetic Permeabilities," *Soviet Physics Usp.*, vol. 10, p. 509.

Walker, J.D., 1993, Vertical-Cavity Laser Diodes Fabricated by Phase-Locked Epitaxy, Ph.D. thesis, Berkeley, CA.

Whale, M.D., and Cravalho, E.G., 2000, "Modeling and Performance of Microscale Thermophotovoltaic Energy Conversion Devices," *IEEE Transactions on Energy Conversion*, vol. 17, pp. 130–142.

Wong, P.Y., Hess, C.K., and Miaoulis, I.N., 1995, "Coherence Thermal Radiation Effects on Temperature Dependent Emissivity of Thin-Film Structures on Optically Thick Substrates," *Optical Engineering*, vol. 34, pp. 1776–1781.

Xu, J.B., Lauger, K., Moller, R., Dransfeld, K., and Wilson, I.H., 1994, "Heat Transfer between Two Metallic Surfaces at Small Distances," *Journal of Applied Physics*, vol. 76, pp. 7209–7216.

Yablonovitch, E., 1986, "Inihibited Spontaneous Emission in Solid-State Physics and Electronics," *Physical Review Letters*, vol. 58, pp. 2059–2062.

Yang, B., and Chen, G., 2001, "Anisotropy of Heat Conduction in Superlattices," *Microscale Thermophysical Engineering*, vol. 5, pp. 107–116.

Yang, B., and Chen, G., 2003, "Partially Coherent Heat Conduction in Superlattices," *Physical Review B*, vol. 67, pp. 195311 (1–4).

Yao, T., 1987, "Thermal Properties of AlAs/GaAs Superlattices," *Applied Physics Letters*, vol. 51, pp. 1798–1780.

Zhang, Z.M., Fu, C.J., and Zhu, Q.Z., 2003, "Optical and Thermal Radiative Properties of Semiconductors Related to Micro/Nanotechnology." *Advances in Heat Transfer*, vol. 37, pp. 179–296.

5.10 Exercises

5.1 *Surface emissivity*. The refractive index of silicon at 0.63 μm is (3.882, 0.019). Calculate the surface reflectivity, transmissivity, and emissivity of a semi-infinite silicon wafer (a) at normal incidence, (b) at 30° angle of inidence, and (c) at 60° angle of incidence, for both TE and TM waves. Also, estimate the penetration depth for normal incidence.

5.2 *Inhomogeneous wave in an absorbing medium.* A plane wave in vacuum is reflected by a medium with a complex refractive index $N = n + i\kappa$ at an angle of incidence θ. Derive an expression for the electric and magnetic fields inside the medium. Show that the constant amplitude and constant phase surfaces of the wave do not coincide with each other. Such waves are called inhomogeneous waves. Derive an expression for the Poynting vector inside the medium.

5.3 *Heat generation distribution due to absorption.* A plane wave with an intensity of 10^4 W m^{-2} at 0.517 μm meets a gold surface at 30° of incidence. Determine the heat generation distribution inside the gold specimen. The refractive index of gold at 0.517 μm is $N = 0.608 + 2.12i$.

5.4 *Fresnel formula for TE wave.* Derive the Fresnel formula for a transverse electric wave incident onto a plane surface, that is, eqs. (5.73) and (5.74).

5.5 *Transmissivity into an absorbing medium.* If the medium is absorbing, one must be careful in writing down the Poynting vector. Examining eq. (5.76) and assuming that only n_2 is complex, $n_2 + i\kappa_2$, derive an expression for the transmissivity, using n_2 and κ_2 explicitly.

5.6 *Interference effects in thin films—Color of thin film.* Experienced workers in thin-film deposition can tell the film thickness from its color. At 0.5 μm, the refractive index of SiO_2 is $N = (1.46, 0)$ and that of silicon is $N = (4.14, 0.045)$. Calculate the reflectivity of a thin film of SiO_2 deposited on the silicon wafer for a film thickness between 500 Å and 2000 Å at normal incidence. Mark down a few colors you expect to see at normal incidence for a few film thickness values in the given range.

5.7 *Optical interference effects in thin films—Angle effects.* A substrate coated with a film may have different colors when looked at from different directions. At 0.5 μm, the refractive index of SiO_2 is $N = (1.46, 0)$ and that of silicon is $N = (4.14, 0.045)$. Calculate the reflectivity of a 500 Å SiO_2 film deposited on silicon wafer for the angles of incidence 0°, 30°, and 45°.

5.8 *Critical angle of incidence for optical waves.* For radiation going from a high refractive index medium into air, calculate the critical angle if the refractive index of the medium is (a) 1.4 and (b) 3.5.

5.9 *Acoustic wave reflection and transmission—SH wave.* For a transverse acoustic wave polarized in the direction perpendicular to the plane of incidence (an SH wave), calculate the reflectivity and transmissivity of the wave at an interface between two isotropic materials at the following angles of incidence: (a) normal, (b) 15°, and (c) 60°. The materials' properties are: material 1: $\rho_1 = 5.33 \times 10^3$ kg m^{-3}, $v_{T1} = 3900$ m s^{-1}; material 2: $\rho_2 = 2.33 \times 10^3$ kg m^{-3}, $v_{T2} = 6400$ m s^{-1}.

5.10 *Reflection of electron wave.* Calculate the reflectivity of a free electron with an energy of 1 eV propagating toward a potential barrier with the following barrier heights: (a) 0.2 eV, (b) 0.8 eV, (c) 1.5 eV.

5.11 *Thermal boundary resistance.* Estimate the thermal boundary resistance between two materials with the following properties on the basis of the diffuse interface scattering model: material 1: $v_1 = 3900$ m s^{-1}, $C_1 = 1.67 \times 10^6$ J m^{-3} K^{-1}; material 2: $v_2 = 6400$ m s^{-1}, $C_2 = 1.66 \times 10^6$ J m^{-3} K^{-1}. For a heat flux of 10^8 W m^{-2}, estimate the temperature drop occurring at the interface?

5.12 *Reflection of longitudinal acoustic wave.* A longitudinal acoustic wave is incident from medium 1 into medium 2. Derive an expression for the reflection and transmission coefficients of the excited longitudinal and transverse waves as a function of the angle of incidence. Both media are assumed to be isotropic and their properties are: $v_{T1} = 6400$ m s^{-1}, $v_{L1} = 8000$ m s^{-1}, $\rho_1 = 2.3 \times 10^3$ kg m^{-3}; $v_{T2} = 3900$ m s^{-1}, $v_{L2} = 5000$ m s^{-1}, $\rho_2 = 5.3 \times 10^3$ kg m^{-3}. Use Auld's book (1990) as a reference for solving this problem.

5.13 *Thermal boundary resistance at low temperature.* Thermal boundary resistance is a phenomenon that is important at low temperatures even for bulk materials and becomes important even at room temperature in nanostructures. Treating the transmissivity in eq. (5.92) as independent of angle and frequency, derive an expression for the proportionality coefficient in eq. (5.93) at low temperatures.

5.14 *Analogy of thermal boundary resistance for photons.* Reflection of carriers can be regarded as an additional resistance, as in the case of thermal boundary resistance. Photons can be reflected at an interface too, as we discussed in this chapter. Now we want to develop an analogy of thermal boundary resistance for photons by considering a partially reflecting and partially transmitting interface located between two parallel black walls maintained at temperatures T_1 and T_2. The transmissivity of the interface is τ_{12}. Derive an expression for the net radiation heat transfer exchange between the two walls, and a corresponding expression for the photon thermal boundary resistance at the interface. In radiation, however, we do not call such a phenomenon thermal boundary resistance.

5.15 *Interference in multilayer structures.* Two layers of thin films are grown on a silicon substrate. At the optical wavelength of 1 μm, the refractive index of silicon is (3.6,0). The refractive index of the layer grown directly on silicon is (2.4,0) and its thickness is 2000 Å. The refractive index of the subsequent layer is (1.3,0) and its thickness varies in the range of 0.1–1 μm. Calculate the reflectivity of the structure at the given wavelength, using the transfer matrix method, for normal incidence.

5.16 *Tunneling of electrons.* For a potential barrier of height 1.0 eV, plot the transmissivity of a free electron with an energy of 0.5 eV through the barrier for a barrier width ranging from 1 Å to 50 Å.

5.17 *Tunneling of photons.* A vacuum gap of 0.2 μm is formed between two glass substrates. Plot the transmissivity of light from one glass substrate into another as a function of angle of incidence for an incident TM wave with a wavelength at 0.5 μm. The refractive index of the glass is taken as 1.46. Compare the results with the situation if a thin film of glass, of 0.2 μm thick, is sandwiched within a vacuum.

5.18 *Landauer formula for phonon heat conduction.* A freestanding thin film of thickness d is suspended between two thermal reservoirs at temperatures T_1 and T_2. The dispersion can be approximated as

$$\omega = v\left[k_x^2 + k_y^2 + \left(\frac{n\pi}{d}\right)^2\right]^{1/2}$$

Assuming that the phonon transmissivity is one and neglecting scattering, derive an expression for the thermal conductance for heat conduction along the thin film plane (x-direction).

5.19 *Landauer formulation for electron thermal conduction.* A metallic square nanowire is placed between two thermal reservoirs at temperatures T_1 and T_2. Assume that electron transmissivity is equal to one. Derive an expression for the thermal conductivity of the nanowire contributed by the electrons.

5.20 *Coherence length of blackbody radiation.* Estimate the coherence length of a blackbody radiation source at 10 K and 300 K.

5.21 *Coherence length of laser radiation.* Estimate the coherence of a laser radiation with a central wavelength of 1.06 μm and a spectral width of 10 Å.

5.22 *Coherence properties of electrons.* At low temperatures, the Fermi velocity in a material is 2.76×10^5 m s^{-1}, the electron relaxation time is 3.8 ps (1 ps = 10^{-12} s), and the phase-breaking time is 18 ps. Calculate the mean free path and the phase coherence length of an electron.

5.23 *Phonon group velocity.* The phonon dispersion for a monatomic lattice chain is

$$\omega = 2\sqrt{\frac{K}{m}}\left|\sin\frac{ka}{2}\right|$$

Derive an expression of its group velocity. Prove that the group velocity at the zone boundary is zero.

5.24 *Difference between wave and particle approaches (project type).* In section 5.6 we stated that wave optics and geometrical optics do not lead to the same results for the radiative properties of periodic multilayer structures for blackbody radiation. Consider a periodic structure made of two alternating layers with refractive indices of (4,0) and (2,0), that is, nonabsorbing films. Blackbody radiation at 1000 K comes toward the periodic multilayer structure at normal incidence. Assuming both sides of the multilayer structure are vacuum, calculate the reflectivity and transmissivity averaged over the blackbody spectrum for the following cases, using wave optics and ray tracing:

(a) For each layer thickness of 1 μm, 10 μm, and 100 μm calculate the variation of reflectivity and transmissivity as a function of the number of periods in the structure. Compare the results for wave and ray tracing.

(b) For 10, 100, 1000 periods, calculate the average reflectivity and transmissivity as a function of the thickness of each layer, assuming all layers are of equal thickness, for the layer thickness range of 1 μm to 100 μm.

Geometrical optics can be obtained using the following recursive formula for the addition of every interface (Siegel and Howell, 1992, p. 928)

$$R_{n+m} = R_m + \frac{R_n T_m^2}{1 - R_m R_n} \qquad \tau_{m+n} = \frac{\tau_m \tau_n}{1 - R_m R_n}$$

where the subscript m refers to the total reflectivity and transmissivity of the first m interfaces (counted from the incident side) and n represents those of the subsequent n additional interfaces. For example, for one layer with two interfaces (the reflectivity and transmissivity at the first interface are R_1 and τ_1 and those at the second interface are R_2 and τ_2), the above formula becomes

$$R_{1+1} = R_1 + \frac{R_2 \tau_1^2}{1 - R_1 R_2} \qquad \tau_{1+1} = \frac{\tau_1 \tau_2}{1 - R_1 R_2}$$

Hint: one numerical problem with the transfer matrix method for thick film is that the exponential function may blow up. One must find ways to solve this problem for calculating thick films using the transfer matrix method.

6

Particle Description of Transport Processes: Classical Laws

We discussed in the previous chapter when we can ignore the coherence effects and treat heat carriers as individual particles without considering their phase information. In the next few chapters, we will describe how to deal with energy transfer under the particle picture. Most constitutive equations for macroscale transport processes, such as the Fourier law and the Newton shear stress laws, are obtained under such particle pictures. These equations are often formulated as laws summarized from experiments. In this chapter, we will see that most of the classical laws governing transport processes can be derived from a few fundamental principles.

In chapter 4, we studied systems at equilibrium and developed the equilibrium distribution functions (Fermi–Dirac, Bose–Einstein, and Boltzmann distributions). The distribution function for a quantum state at equilibrium is a function of the energy of the quantum state, the system temperature, and the chemical potential. When the system is not at equilibrium, these distribution functions are no longer applicable. Ideally, we would like to trace the trajectory of all the particles in the system, as in the molecular dynamics approach that we will discuss in chapter 10. This approach, however, is not realistic for most systems, because they have a large number of atoms or molecules. Thus, we resort to a statistical description of the particle trajectory.

In the statistical description we use nonequilibrium distribution functions, which depend not only on the energy and temperature of the system but also on positions and other variables. We will develop in this chapter the governing equations for the nonequilibrium distribution functions. In particular, we will rely on the Boltzmann equation, also called the Boltzmann transport equation. From the Boltzmann equation we will derive familiar constitutive equations such as the Fourier law, the Newton shear stress law, and the Ohm law. We will also demonstrate that conservation equations,

such as the Navier–Stokes equations for fluids and electrohydrodynamic equations for charged particles, can be obtained from the Boltzmann equation. Special attention will be paid to the approximations made in these derivations, which will be relaxed in the next chapter when we consider various classical size effects. A discussion is also presented in this chapter on thermal waves and their appropriate descriptions.

6.1 The Liouville Equation and the Boltzmann Equation

We discussed, in chapter 4, the probability distribution of an equilibrium system occupying a specific accessible quantum state. Because the system is at equilibrium, the probability distribution take a simple form. For example, the Boltzmann distribution depends only on the energy of the quantum state and on the system temperature. Transport occurs, however, only when the system is in a nonequilibrium state and consequently the equilibrium distribution can no longer describe the state of the system. Conceivably, to describe the state of such a nonequilibrium system, more information is needed. In this section, we will introduce nonequilibrium distribution functions that describe the states of systems and the governing equations for the evolution of the nonequilibrium distribution functions. We will start from the general Liouville equation, which is valid for all classical systems but is difficult to solve, and move on to the simpler Boltzmann equation that serves as the basis for our future analysis. We will also discuss the assumptions made in the Boltzmann equation and see, consequently, its limitations.

6.1.1 The Phase Space and Liouville's Equation

Consider a system with N particles, where each particle can be described by the generalized coordinate $\mathbf{r}$ and momentum $\mathbf{p}$. For example, the generalized coordinates of a diatomic molecule, $\mathbf{r}_1$, include the position (x_1, y_1, z_1), the vibrational coordinate (the separation between the two atoms, Δx_1), the rotational coordinates (polar and azimuthal angles, θ_1 and φ_1); likewise, the generalized momentum, $\mathbf{p}_1$, includes the translational momenta (mv_{x1}, mv_{y1}, mv_{z1}), the vibrational momentum proportional to the relative velocity of the two atoms ($md\Delta x_1/dt$), and the rotational momenta (angular momenta of rotation corresponding to θ and φ directions). We assume here that there are m degrees of freedom in space, that is, m generalized spatial coordinates, and m degrees of freedom in momentum for each particle. The number of the degree of freedom of the whole system is $2n = 2m \times N$. These $2n$ variables form a $2n$-dimensional space that is called a *phase space*. The system at any instant can be described as one point in such a space. The time evolution of the system, that is, the time history of all the particles in the system, traces one line in such a $2n$-dimensional *phase space*, which we will call the flow line as in fluid mechanics.

Now we consider an ensemble of systems—a collection of many systems satisfying the same macroscopic constraints—as we did in chapter 4. At time $t = 0$, each system in the ensemble is represented by a different point in the phase space, as shown in figure 6.1. From classical mechanics, we know that with a given initial condition the trajectory of the system is uniquely determined. Since the initial condition for each system differs from that of other systems in the ensemble, the traces of systems in

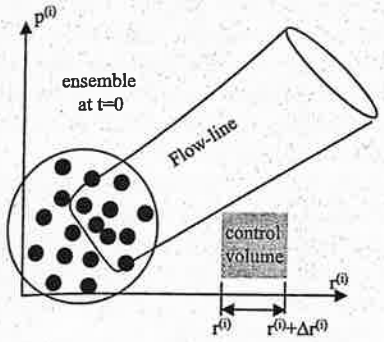


Figure 6.1 Phase space, and an ensemble in the phase space.

such an ensemble never intersect, so that the flow lines in phase space do not intersect each other.

The number of systems in an ensemble is usually very large, much larger than the number of the particles in one system. Because of the large number of systems in one ensemble, we can treat the points of the ensemble, each representing one microstate of the original macroscopic system, as forming a continuum in the phase space, just as we treat atoms or molecules in a macroscopic system as a continuous medium in real space. We define a particle density $f^{(N)}$ such that, surrounding any point $(\mathbf{r}^{(n)}, \mathbf{p}^{(n)})$ in the phase space, where $\mathbf{r}^{(n)} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = (r^{(1)}, r^{(2)}, r^{(3)}, \ldots, r^{(n)})$ includes all the space coordinates of N particles and similarly $\mathbf{p}^{(n)}$ represents all the momentum coordinates, the number of systems is

$$\text{No. of systems} = f^{(N)}(t, \mathbf{r}^{(n)}, \mathbf{p}^{(n)})\Delta\mathbf{r}^{(n)}\Delta\mathbf{p}^{(n)} \tag{6.1}$$

in a small volume of the phase space, $\Delta\mathbf{r}^{(n)}\Delta\mathbf{p}^{(n)}$, where $\Delta\mathbf{r}^{(n)} = \Delta\mathbf{r}_1\Delta\mathbf{r}_2\ldots\Delta r_N = \Delta r^{(1)}\Delta r^{(2)}\ldots\Delta r^{(n)}$ and $\Delta\mathbf{p}^{(n)} = \Delta\mathbf{p}_1\Delta\mathbf{p}_2\ldots\Delta\mathbf{p}_N = \Delta p^{(1)}\Delta p^{(2)}\ldots\Delta p^{(n)}$. We use superscript (n) to denote the generalized space and momentum coordinates, and superscript (N) to represent the N particles. The particles density in the phase space $f^{(N)}(t, \mathbf{p}^{(n)}, \mathbf{p}^{(n)})$ is called the N-particle distribution function, which represents the probability density of finding a particular system at a specific state defined by $\mathbf{r}^{(n)}$ and $\mathbf{p}^{(n)}$. If we assume that the ensemble is ergodic for all time, this distribution function also represents the probability of observing one system at a particular state $\mathbf{r}^{(n)}$ and $\mathbf{p}^{(n)}$ over a period of time (such a time period should be smaller than the characteristic time we use in tracing the trajectory, or the relaxation time that we will discuss later).

The time evolution of $f^{(N)}(t, \mathbf{r}^{(n)}, \mathbf{p}^{(n)})$ in the phase space is governed by the *Liouville equation*, which can be derived on the basis that the flow lines of systems in the ensemble do not intersect. Consider a tube formed by the traces of a set of points (a subset of systems in the ensemble) as shown in figure 6.1. Since the flow lines do not intersect, the points in the phase space are conserved. We want to derive an equation for the distribution function $f^{(N)}$ based on this conservation requirement. Recall that in fluid mechanics or heat transfer, we often use the control volume method rather than tracing the trajectory of individual fluid particles. We could do the same for the points in phase space and examine a small control volume in phase space, as shown in figure 6.1.