Transport Theory #10: Quantum Transport

A. TUNNELING

According to classical transport theory, an electron can be located anywhere in space consistent with Newton's laws. Specifically, the sum of the single-particle kinetic energy plus potential energy must equal the total energy at each point in space and time, i.e., $U(x,t) = U_K(x,t) + U_P(x,t)$. In crystalline materials, semiclassical theory does not modify this sum rule, but constrains it through a specific relationship between U, U_K, and U_P: the band structure of the material. For example, in a spherical conduction band at wave vectors near $k = 0$, we know

$$U = U_c + \hbar^2 k^2 / 2m^* \tag{12.1}$$

where U_C is the potential energy of the conduction band relative to the zero of energy (often chosen as the valence band edge). As a consequence, neither the classical or semi-classical theories allow the transport of an electron in regions where $U < U_P$ since that would require a negative kinetic energy, or negative k^2 in the case of a spherical conduction band. And these peculiar requirements are "classically" and semi-classically forbidden.

A simple and common example of a classically forbidden scenario in solids is the transport of free electrons in the presence of a barrier, as shown in Fig. 12.1. Consistent with the above definitions, if the zero of potential energy is defined as the conduction-band edge (solid line), and the total energy U is represented as the horizontal dashed line, then a classically forbidden region occurs when its U_C exceeds U. The only way for an electron to move into or through this region is for its kinetic energy to go negative.

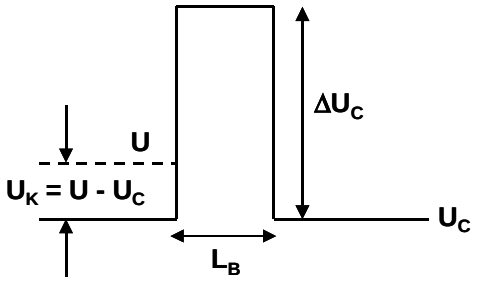

Fig. 12.1. Conduction band profile and physical parameters of a typical heterostructure barrier for an electron of kinetic energy U_K in a semiconductor.

In practice such barriers can occur a number of different ways, a common one being the adjoining of a semiconductor of band gap U_{G1} with a second semiconductor or insulator having much wider band gap U_{G2}. Some fraction of the band-gap difference appears as the barrier height.

Stationary-State Model of Tunneling

Quantum mechanics changes the interpretation of such a barrier in a radical way. By representing the electron as a probability (amplitude) wave function, *Schrödinger equation* will always allow some transmission or "leakage" through the barrier if it is thick and separates two semiclassically allowed regions. Then if the electron is incident to the barrier from one side, there will be a finite probability that it will leak or "tunnel" to the other side, and is calculated by the effective-mass *Schrödinger equation:*

$$\frac{-\hbar^2\partial^2 f}{2m^*\partial x^2}+V(x)f(x)=\frac{-\hbar}{j}\frac{\partial f}{\partial t}=Uf \tag{12.2}$$

$$\frac{\partial^2 f}{\partial x^2}+\frac{2m^*}{\hbar^2}\left[U-V(x)\right]f(x)=0 \tag{12.3}$$

where f is the envelope function and $\boldsymbol{k}$ is the crystal wave vector.

A very simple barrier type to analyze is the single-crystal heterobarrier consisting of a thin crystalline material sandwiched between two other crystalline materials of nearly the same lattice constant and atomic potential. In this case one can assume that the cell-periodic function is constant throughout the structure, and as proven in Chapter 11, deal only with the envelope function in solving the effective-mass *Schrödinger equation*. We merely "dress" it by the effective mass to account for the change in atomic potential.

A second simplifying assumption is that the electron interaction with the barrier is elastic so that its energy U_e is conserved during the tunneling process. Under this condition we can treat U_e as a constant across the structure, treat U_P as a constant within each region, and define U_p arbitrarily equal to zero in the incident and final regions so that in the barrier region $U_p = \Delta U_C$, the conduction band offset. This leads us to plausible solution to (12.3) in the left-hand region of

$$f_1=A_1exp\left[jk_1x\right]+A_2exp\left[-jk_1x\right] \qquad k_1=\sqrt{2m_1^*U_e}, \tag{12.4}$$

in the barrier region of,

$$f_2=B_1exp\left[jk_2x\right]+B_2exp\left[-jk_2x\right] \qquad k_2=\sqrt{2m_b^*\left(U_e-\Delta U_C\right)}, \tag{12.5}$$

in the right-hand region, and

$$f_3=C_1exp\left[jk_3x\right]+C_2exp\left[-jk_3x\right] \qquad k_3=\sqrt{2m_3^*U_e}. \tag{12.6}$$

Since $U_e - \Delta U_c < 0$ by the definition of a barrier, we see that k_2 is pure imaginary, so we define a real quantity $k_b = \sqrt{2m_b^*(\Delta U_C - U_e)}$ to get $f_2 = Be^{=\pm k_b x}$. Note that if we substitute (12.4), (12.5), and (12.6) back into (12.2), we find the time dependent part of the wave function is f $\propto exp\left[-jU_e t/\hbar\right]$ in all three regions. So when combined with the spatial dependence, equations (12.4) and (12.6) become the sum of two plane waves, the first (A_1 and C_1) representing waves moving to left, and the second (A_2 and C_2) representing waves moving to the right.

Equations (12.4), (12.5), and (12.6) generally represent an insoluble system because of the 6 unknown coefficients. But we can simplify the problem further with an approximation developed in Chapter 9 on semi-classical transport. We imagine that the complete wave function is a wave packet narrow enough in real space that the electron can be considered to be isolated to region 1 or region 3, or some combination of the two when it is tunneling through the barrier. We then consider the three regional functions (12.4), (12.5), and (12.6) as spatial Fourier components of the wave packet that must be consistent with any boundary conditions imposed on the wave packet as a whole. The similarity of our generic solutions to those of a one-dimensional electromagnetic scattering problem suggests the use of scattering boundary conditions at +/- infinity. A plane wave incoming from -∞ will scatter from the barrier, producing a reflected component toward -∞ and a transmitted component toward +∞.

If we set the incoming amplitude arbitrarily to unity, then

$$f_1 = exp\left[\, jk_1 x\right] \; + r\, exp\left[-jk_1 x\right] \tag{12.7}$$

$$f_2 = B_1 exp\left[k_b x\right] + B_2 exp\left[-k_b x\right] \tag{12.8}$$

$$f_3 = t\, exp\left[\, jk_3 x\right] \tag{12.9}$$

where *r* is the reflection amplitude and *t* is the transmission amplitude. These are related to reflection and transmission probabilities through the expressions $R = r^*r$ and $T = t^*t$ where the * denotes complex conjugation.

We have reduced the number of unknowns to four, and can now apply boundary conditions at the tunneling barrier itself to reduce this to two.

Aside on boundary conditions at barriers: probability current

One of the most important theorems in quantum mechanics is the conservation of probability. From elementary quantum mechanics, these come from the definition of the probability flux,

$$J \equiv \frac{\hbar}{2mj}\left(\psi^* \frac{\partial \psi}{\partial x} - \psi \frac{\partial \psi^*}{\partial x}\right) \tag{12.10}$$

When applied to the envelope function via the effective-mass theorem, this condition becomes

$$J \equiv \frac{\hbar}{2m^* j}\left(\frac{f^*}{m^*}\frac{\partial f}{\partial x} - \frac{f}{m^*}\frac{\partial f^*}{\partial x}\right). \tag{12.11}$$

A necessary and sufficient condition for the spatial continuity of J is that f and $(1/\mathrm{m}^*)df/dx$ are each continuous.

The continuity of f at $x = 0 \Rightarrow f_1(x{=}0) = f_2(x = 0)$, or from (12.7) and (12.8)

$$1 + r = B_1 + B_2 \tag{12.12}$$

Similarly, the continuity of f'/m^* at x = 0 $\frac{1}{m_1^*} f_1'(x=0) = \frac{1}{m_b^*} f_2'(x=0)$ or from (12.7) and (12.8)

$$\frac{jk_1}{m_1^*} - \frac{jk_1 r}{m_1^*} = \frac{B_1 k_b}{m_b^*} - \frac{B_2 k_b}{m_b^*} \tag{12.13}$$

Continuity of f and f'/m^* at $x = L_b$ are also straightforward. From (12.8) and (12.9),

$$f_2(x = L_b) = f_3(x = L_b) \Rightarrow B_1 exp[k_b L_b] + B_2 exp[-k_b L_b]^{-k_b L_b} = t\, exp[jk_3 L_b]^{jk_3 L_b} \tag{12.14}$$

and,

$$\frac{1}{m_b^*} f_b'(x = L_b) = \frac{1}{m_3^*} f_3'(x = L_b) \Rightarrow \frac{B_2 k_b}{m_b^*} exp[k_b L_b] - \frac{B_2 k_b}{m_b^*} exp[-k_b L_b] = \frac{jk_3 t}{m_3^*} exp[jk_3 L_b] \tag{12.15}$$

Equations (12.12) thru (12.15) represent 4 equations in 4 unknowns: r, t, B_1, B_2. By clever manipulation of these expressions, one can derive the useful expression.

$$r = -\frac{exp\left[k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 + \left(\frac{k_3}{m_3^*}\right)^2 + \frac{jk_bk_1}{m_b^*m_1^*} - \frac{jk_bk_3}{m_b^*m_3^*}\right] + exp\left[-k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 + \left(\frac{k_3}{m_3^*}\right)^2 - \frac{jk_bk_1}{m_b^*m_1^*} + \frac{jk_bk_3}{m_b^*m_3^*}\right]}{\exp\left[k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 - \frac{k_1k_3}{m_1^*m_3^*} - \frac{jk_bk_1}{m_b^*m_1^*} - \frac{jk_bk_3}{m_b^*m_3^*}\right] + \exp\left[-k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 - \frac{k_1k_3}{m_1^*m_3^*} + \frac{jk_bk_1}{m_b^*m_1^*} + \frac{jk_bk_3}{m_b^*m_3^*}\right]}$$

(12.16)

Clearly (12.15) gets much simpler if $k_1 = k_3 = k_0$ and $m_l^* = m_3^* = m_0^*$. In this case

$$r = \frac{-\left[\left(\frac{k_b}{m_b^*}\right)^2 + \left(\frac{k_0}{m_0^*}\right)^2\right]\left[exp\left[k_b L_b\right] + exp\left[-k_b L_b\right]\right]}{exp\left[k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 - \left(\frac{k_0}{m_0^*}\right)^2 - \frac{2jk_bk_0}{m_b^*m_0^*}\right] + exp\left[-k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 - \left(\frac{k_0}{m_0^*}\right)^2 + \frac{2jk_bk_0}{m_b^*m_0^*}\right] - \left[\left(\frac{k_b}{m_b^*}\right)^2 + \left(\frac{k_0}{m_0^*}\right)^2 2\sinh k_b L_b\right]}$$

$$= \frac{}{exp\left[k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 - \left(\frac{k_0}{m_0^*}\right)^2 - \frac{2jk_bk_0}{m_b^*m_0^*}\right] + exp\left[-k_b L_b\right]\left[\left(\frac{k_b}{m_b^*}\right)^2 + \left(\frac{k_0}{m_0^*}\right)^2\right]}$$

(12.17)

And the reflection coefficient becomes

$$R \equiv rr^* = \frac{\left[\left(\frac{k_b}{m_b^*}\right)^2 + \left(\frac{k_0}{m_0^*}\right)^2\right] 4\sinh^2 k_b L_b}{\left|exp\left[k_b L_b\right]\left[\left(\frac{k_b}{m_b^*} - \frac{jk_0}{m_0^*}\right)^2\right] + exp\left[-k_b L_b\right]\left[\left(\frac{k_b}{m_b^*} + \frac{jk_0}{m_0^*}\right)^2\right]\right|^2}$$

(12.18)

For the special case of $m_b^* = m_1^* = m_3^*$, this reduces further to a simple enough form to be expressed directly in terms of barrier heights

$$R = \frac{\frac{\left(\Delta U_c\right)^2}{4U_e\left(\Delta U_c - U_e\right)}\sinh^2\left[\frac{\sqrt{2m^*\left(\Delta U_c - U_e\right)}L_b}{\hbar}\right]}{1 + \frac{\left(\Delta U_c\right)^2}{4U_e\left(\Delta U_c - U_e\right)}\sinh^2\left[\frac{\sqrt{2m^*\left(\Delta U_c - U_e\right)}L_b}{\hbar}\right]}$$

(12.19)

And then it is simple to show that by conservation of probability, $T = 1 - R$, and we can write

$$T = \frac{1}{1 + \frac{(\Delta U_c)^2}{4U_e(\Delta U_c - U_e)} \sinh^2\left[\sqrt{\frac{2m^*(\Delta U_c - U_e)L_b}{\hbar}}\right]} \tag{12.20}$$

Generally $T<<1$ unless $U_e \sim \Delta U_c$ (trivial case) or $L_b << \frac{\hbar}{\sqrt{2m^*(\Delta U_c - U_e)}}$, which generally corresponds to a very small barrier thickness.

Example: As an example, we consider $\Delta U_C \approx 0.3\ eV$ and $m^* \approx 0.1 \cdot m_0$, typical of many semiconductor barriers. The criterion for high transmission probability becomes $L_b <<$11 Å. While this is very small (≈ 4 monolayers in Si or GaAs), such barriers can in fact be grown by molecular beam epitaxy in the GaAs/AlGaAs materials system.

Current Density

With the stationary state model of tunneling, each electron contributes to electrical current through the barrier in proportion to its probability to tunnel. We can factor the problem into semiclassical and quantum mechanical components by writing

$$J_i = ev_i T_i \tag{12.21}$$

where the subscript i denotes the i^{th} longitudinal-energy interval, and we are only considering electrons tunneling from one side of the barrier to the opposite side.
The semiclassical component is the group velocity, which for a spherical energy band is just

$$v_i = (1/\hbar)\partial U / \partial k\big|_{U=U_i} \tag{12.22}$$

So the net electrical current is found by summing over the energy components

$$J_T \approx \sum J_U = 2e \sum_{k\perp} \int_0^\infty \frac{dk_\perp}{2\pi} \cdot \frac{v_U(\boldsymbol{k}) T(\boldsymbol{k})}{1 + exp\left[U(\boldsymbol{k}) - \mu / k_B T\right]} \cdot \tag{12.23}$$

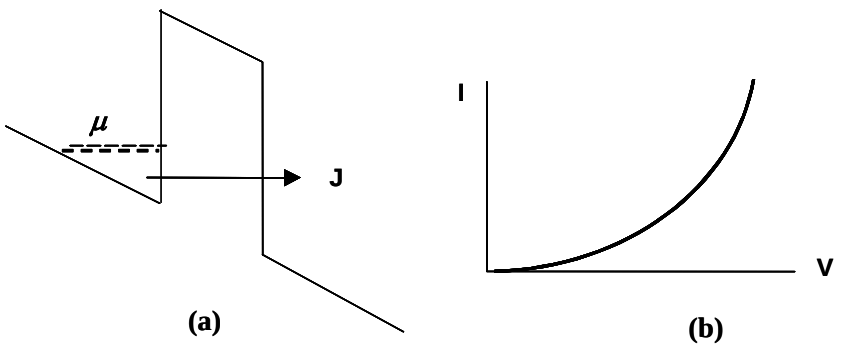

Fig. 12.2. Effect of electrical bias voltage on (a) potential profile of a single tunneling barrier, and (b) current-density-vs-voltage characteristic.

Now we need to sum over the perpendicular 2D plane, recalling the relationship for 2D electrons in the Sommerfeld-Fermi model:

$$2\sum_{k\perp}\frac{1}{1+exp\left[U(\boldsymbol{k})-\mu/k_BT\right]}\approx\frac{m^*k_BT}{\pi\hbar^2}\ln\left(1+exp\left[-U_{||}/k_BT\right]\right) \tag{12.24}$$

This justifies a separation of the total electron energy into longitudinal (parallel) and transverse (perpendicular) components, $U=U_{\parallel}+U_{\perp}$), so that

$$J_T=\frac{e}{2\pi\hbar}\int dU_{||}T\left(U_{||},V_B\right)\ln\left[1+exp\left[-U_{||}/k_BT\right]\right] \tag{12.25}$$

This is a famous expression usually credited to Tsu and Esaki (but derived earlier during the 1960s). In this model T increases with bias voltage V_b because it lowers the barrier electrostatically, as shown qualitatively in Fig. 12.2(a). This causes the current J_T to increase in kind, often superlinearly as shown qualitatively in Fig. 12.2(b).

Tunneling is usually very sensitive to the density-of-states in the barrier region. If we add more barriers separated by narrow "wells", we can engineer the density-of-states in the "tunnel region" to create different J vs V behavior. This is illustrated qualitatively in Fig. 12.3 for a double-barrier structure having one quasibound state U_q. The presence of this quasibound state creates a resonance in the T vs U curve, and an associated peak in the J vs V curve. This is called *resonant tunneling* and is the subject of the next section.

B. QUANTUM CONDUCTANCE

Another way to restrict the current flow is to make the current path so narrow that the quantum confinement energies are comparable to μ. The length scale for this to occur is the de-Broglie wavelength. The scenario is shown graphically in Fig. 12.8 where an electron "aperture" is drawn in an otherwise classically forbidden region. The electrical current is calculated using the diagram of Fig. 12.9. For each electron having one lateral state in both dimensions of the aperture, we find

$$J = \sum_k \frac{evT_\mu}{A} \cdot \left[\frac{1}{2} \frac{\Delta n}{\Delta \mu}(\mu) \cdot eV \right] \tag{12.39}$$

where the quantity $\Delta n/\Delta\mu = D_{1D}(\mu)$ is the one-dim density of states, and the summation is carried out over only one direction of current flow. We evaluate this by summing over wave vectors,

$$\sum_k \rightarrow \frac{4dk}{2\pi} \rightarrow \frac{1}{\pi\hbar v} \equiv D_{1D} \tag{12.40}$$

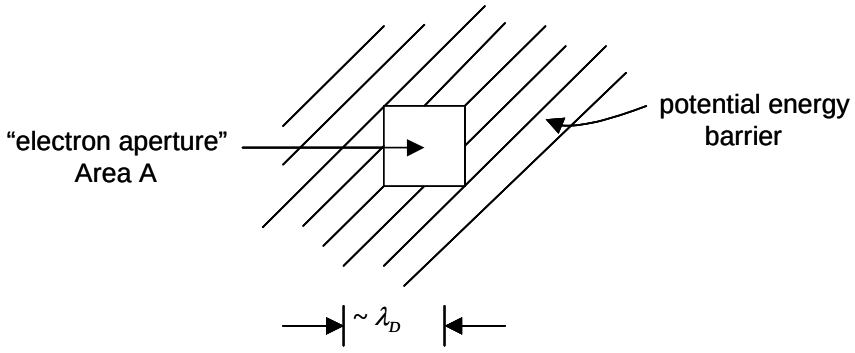

Fig. 12.8. Two-dimensional electron aperture used to define quantum conductance.

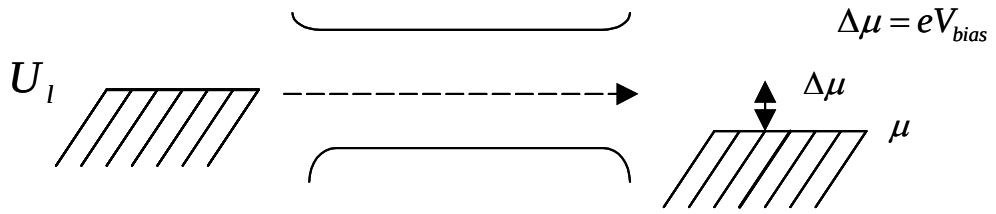

Fig. 12.9. Lateral view of electron transport through the two-dimensional aperture of Fig. 12.8.

This leads to

$$J = \frac{2e^2}{h} T(\mu) \frac{V}{A} \ , \tag{12.41}$$

a famous expression in quantum transport called the Landauer formula. It suggests a conductance of the form

$$G \equiv \frac{I}{V} = \frac{2e^2}{h} T(\mu) \tag{12.42}$$

In a more general analysis we would get the expression

$$G = \frac{2e^2}{h} \cdot nT \tag{12.43}$$

where *n* is the number of "channels" through the aperture for conduction. Independent of the number of channels, we have a universal quantum for conductance given by

$$\frac{e^2}{h} \approx \frac{1}{26\ K\Omega} = 3.9 \times 10^{-5} S \tag{12.44}$$

C. QUANTUM BOLTZMANN TRANSPORT: HEURISTICS

\Beyond the Semiclassical Scattering Assumptions

One of the underlying assumptions of both the classical and semiclassical transport theories is the form of the carrier density-of-states in real space, and in velocity (or crystal-momentum) space. Specifically it is assumed that all points of ***r*** and of **v** or ***k***, are "accessible" to the carrier, meaning that they are: (i) physically allowed, and (ii) close-enough together that it makes sense mathematically to form the derivatives $\partial f / \partial \boldsymbol{r}$ and $\partial \boldsymbol{r} / \partial t$, or $\partial f / \partial \boldsymbol{v}$ and $\partial \boldsymbol{v} / \partial t$, or $\partial f / \partial \boldsymbol{k}$ and $\partial \boldsymbol{k} / \partial t$, that are so fundamental to classical or semi-classical Boltzmann transport. In this sense, we can understand the

primary difference between classical and semi-classical transport as just a restriction on the allowed values of energy U for each $\mathbf{v}$ or $\boldsymbol{k}$. Classically, all positive values of U are allowed for each $\mathbf{v}$ or $\boldsymbol{k}$. Semiclassically, the allowed values of U are restricted to those associated with the band structure of the solid. In both cases, all values of $\boldsymbol{k}$ are allowed, so that U is a continuous function of $\boldsymbol{k}$, and $\boldsymbol{k}$ is a continuous function of t.

Now suppose that some structural or chemical inhomogeneity exists in the solid such that not all values of $\boldsymbol{r}$ and of $\mathbf{v}$ or $\boldsymbol{k}$ are "accessible" to the carrier. This can come about several different ways, a common example being the two-phase solid consisting of a crystalline host material containing inclusions of a second type of material (not necessarily crystalline). The key assumptions about the inclusions are: (i) they are not accessible to the carriers in $\boldsymbol{r}$, and (ii) they occupy such a large enough fraction of the solid that the assumptions behind the quantum scattering theory in Chapter 9 are questionable. The key assumption in question is that the scattering can be modeled as the transition of an incident wave packet to a second, independent wave packet. In the case of "point-like" centers such as ionized impurities, the assumption is valid if the scattering centers are sparse enough that their spatial separation is much greater than the width of the typical wavepacket. In the case of "distributed" scattering such as that from acoustical or optical phonons, the assumption is valid if that the phonon-induced scattering transition rate is small compared to "ballistic" transition rate associated with drift. Qualitatively, the common criterion for both cases is that the scattering be "weak" compared to the inertial motion of the carrier in r and k space.

Quantum Transport with Point-Like Scattering- The Ioffe-Regel Criterion

The violation of the "weak-scattering" assumption is easiest to understand in the case of point-like scattering. We return to the fundamental issue of forming a wave packet in real space from Bloch functions in k space. Intuitively, we expect that the point-like scatterers will transform incident wave packets to independent wave packets if the wavepacket width ΔR is much less than distance between scatterers L, i.e.,

$$\Delta R << L \tag{12.45}$$

But the spatial extent of the wave packet is governed by Fourier analysis such that

$$\Delta R \, \Delta k > 1 \tag{12.46}$$

where Δk is the "bandwidth" in k space. Independent of whether we have an insulator, semiconductor, or metal, the "bandwidth" is generally less than the Fermi wavevector k_F, so that Δk has an upper limit

$$(\Delta k)_{max} \approx k_F \tag{12.47}$$

Substitution of (12.47) into (12.46) then allows us to define a minimum wavepacket size

$$(\Delta R)_{min} \approx 1/(\Delta k)_{max} \approx 1/k_F \tag{12.48}$$

It is worth emphasizing that this is the minimum size wave packet and that most carriers in the solid will have even larger wave packets than given by (12.48).

Substitution of (12.48) into (12.45) then yields an important criterion for having at least some of the carriers in the solid able to scatter as independent wave packets,

$$1/k_F << L, \text{ or } k_F L >> 1 \qquad (12.49)$$

This important result is called the *Ioffe-Regel criterion*, and is one of the most often-cited and confused "rules" in all of quantum transport theory. The confusion is that it represents the dividing line between classical and quantum transport. On the contrary, it just represents the condition for which some fraction – maybe a very small fraction – of the carriers can be transported by semi-classical means. The subsequent application of the semi-classical Boltzmann transport equation may or may not yield accurate results, but (12.49) defines the condition in which such an application is reasonable.

In the opposite limit of where $k_F L << 1$, the carriers can still be transported through the solid but by a highly disruptive process where the scattering of wave packets is occurring so often that there is no way to tell where the incident wave packet stops and the scattered wave packet begins. The transport in this case is called "percolation" in analogy to fluidic transport through an inhomogeneous medium such as sand.