Theory of transport in semiconductors
ECE 211A, Device Physics
Debdeep Jena (djena@engineering.ucsb.edu)
(May, 2002)

Introduction

Basic semiconductor physics

This is a brief recap of semiconductor physics that we will need for the transport theory that will follow. As we all know, a semiconductor has atoms arranged in a crystal lattice, with all atoms tetrahedrally bonded. Such a solid will allow only certain bands of allowed energies, separated by forbidden gaps. The gap separating the highest filled band (valence band) from the lowest unoccupied band (conduction band) is the well known bandgap of the semiconductor.

Semiconductors have to be doped to get sufficient carriers to do useful electronics. These dopants are typically substitutional impurity atoms that have energy levels very close to the conduction band edge (donors) or the valence band edge (acceptors). The electrons (in case of donors) and holes (acceptors) contributed by the dopants are bound to the parent atoms at very low temperatures. However, at room temperature, they have sufficient energy to escape the binding energy of the parent dopant and become free to wander around the whole crystal. How fast do these free mobile charges move? That is the question transport theory aims to answer. It is of fundamental importance, since it determines how fast your transistor can go, how well your laser will lase, and so on.

Electrons (or holes) that are free to wander around the crystal experience the Coulombic potential variation of the atoms in the crystal (see Figure 1). In a perfect crystal, this potential variation is periodic in nature. How is this problem solved? Bloch's theorem provides the vital link to the solution of the problem. It turns out that in a perfect crystal, you can remove the whole periodic crystal potential and lump it into the wonderful concept of the 'effective mass' of the electron. By doing so, you have freed the electron from the shackles of widely fluctuating Coulombic potentials and have reduced the problem to that of an electron in free space, albeit with a different mass [1]. Solving the problem for the electron taking the periodicity of the crystal into account yields the bandstructure of the semiconductor, which stores a wealth of information. Figure 2 shows the Brillouin Zone, the bandstructure, and the simplified bandstructure around the CB minima and VB maxima for GaN, which is the region of interest for most problems. It is important to note that the bandstructure is nothing but the allowed eigenvalues of the quantum mechanical problem for the perfect semiconductor. The k-values are quantized; the quantization is so fine that for all practical purposes, it is assumed continuous.

[1] This is the reason why you see that effective mass is not a scalar, but has dependence on the wavevector k. Roughly, since all directions are not the same in a crystal, neither should the periodic potential (and hence the effective mass in that direction). For our purpose though, we will be happy to deal with a direct gap semiconductor that has an isotropic effective mass at the Γ point of the conduction band.

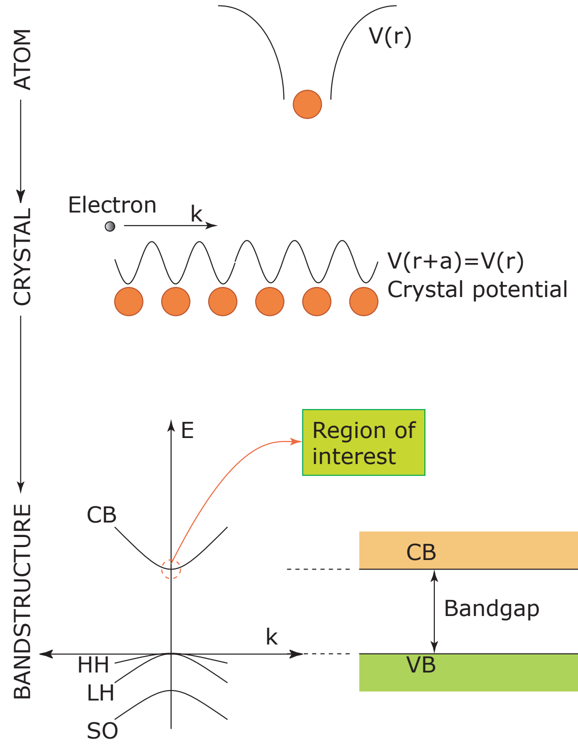

Figure 1: Evolution of bands and gaps in semiconductors

We now direct our attention towards the problem at hand - how to understand transport theory. So the model we start with is really the first problem of an undergraduate quantum mechanics class - you have a particle (electron, effective mass m^*) in a macroscopic crystal (say a GaN crystal). Let us assume the shape of the crystal to be a cube of side L [2]. Since the particle is confined in a box, we model it as an infinite quantum well in 3 dimensions. It is the good old particle in a box problem, which has well known solutions. The wavefunction of the electron will be

$$\psi(x, y, z) = \frac{1}{\sqrt{L^3}} e^{\iota(k_x x + k_y y + k_z z} = \frac{1}{\sqrt{L^3}} e^{\iota \mathbf{k} \cdot \mathbf{r}} \tag{1}$$

which when subjected to the boundary conditions [3] yields the limits on the wavevectors -

$$\mathbf{k} = (k_x, k_y, k_z) = \frac{2\pi}{L}(n_x, n_y, n_z) \tag{2}$$

where n_x, n_y, n_z are allowed only integral values (...,-1,0,1,...). The energy is $E(k) = \frac{\hbar^2 |\mathbf{k}|^2}{2m^*}$ and the momentum is $\mathbf{p} = \hbar \mathbf{k}$. The wavevector $\mathbf{k}$ for a free electron (not in a box) is related to its wavelength by the de Broglie relation $\lambda = \frac{h}{p}$; it is easy to see that it translates to $k = \frac{2\pi}{\lambda}$.

[2] L can be 1 cm; it is a macroscopic length, not microscopic!

[3] The periodic boundary condition requires $\psi(x + L, y, z) = \psi(x, y + L, z) = \psi(x, y, z + L) = \psi(x, y, z)$

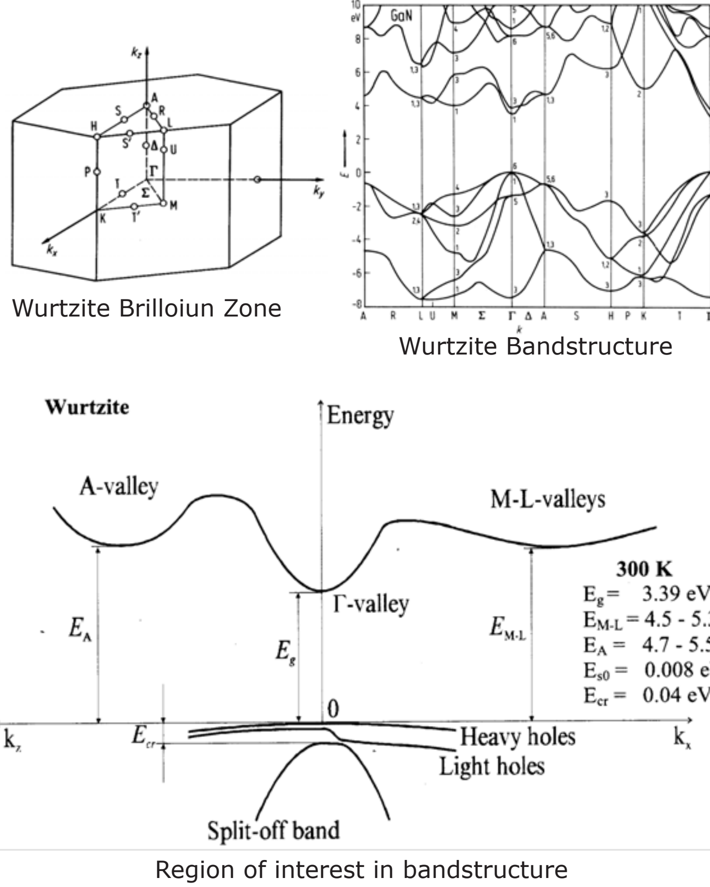

Figure 2: The Brillouin Zone of Wurtzite GaN crystal structure, The calculated bandstructure for different directions in the Brillouin Zone, and the Kane model of Bandstructure around the Γ point

When the electron is confined in a box, it can have only those wavelengths that can fit in the box; so, the wavelengths are 'quantized', according to $\mathbf{k}$. Electrons with wavevectors $k \to 0$ have very small momentum and are delocalised over the whole crystal, since their $\lambda \to \infty$. On the other hand, electrons with large wavevectors have really small wavelengths, and can be fitted into small microscopic areas. We will see how this is crucial in the phenomenon of screening, a very important phenomenon in transport and indeed in all of semiconductor physics.

Example 1: Matrix element evaluation
Evaluate $\langle \mathbf{k}' | \mathbf{k} \rangle$ for the free electron in the box of volume $V = L_x \cdot L_y \cdot L_z$.

$$|\mathbf{k}\rangle = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{3}$$

$$\langle \mathbf{k}'|\mathbf{k}\rangle = \frac{1}{V}\int d^3r e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} = \frac{1}{V}\int_0^{L_x} dx e^{i(k_x-k_x')x}\int_0^{L_y} dx e^{i(k_y-k_y')y}\int_0^{L_z} dz e^{i(k_z-k_z')z} \tag{4}$$

$$\langle \mathbf{k}'|\mathbf{k}\rangle = (\frac{e^{i(k_x-k_x')L_x}-1}{i(k_x-k_x')L_x})\cdot(\frac{e^{i(k_y-k_y')L_y}-1}{i(k_y-k_y')L_y})\cdot(\frac{e^{i(k_z-k_z')L_z}-1}{i(k_z-k_z')L_z}) \tag{5}$$

Now we use a very useful result that k_x, k_x' are quantized by the wave-fitting procedure as in Equation 2. So $(k_x - k_x')L_x = 2\pi n$ where n is an integer. The same holds for k_y, k_z as well. Using this, we see that unless $k_x = k_x', k_y = k_y', k_z = k_z'$ all hold true, the value of the integral would be zero. If they do hold, the value is ONE (prove it!). Thus, the result is

$$\langle \mathbf{k}'|\mathbf{k}\rangle = \delta_{k_x,k_x'}\delta_{k_y,k_y'}\delta_{k_z,k_z'} = \delta_{\mathbf{k},\mathbf{k}'} \tag{6}$$

Transport theory

If you apply an electric field and measure the drift velocity of electrons in a bulk semiconductor, you will see what is called the velocity-field characteristic. It is schematically shown in the left part of Figure 3. Drift velocity is proportional to the applied field for low fields; this is the 'ohmic' regime characterized by the electron mobility $\mu = v_d/F$. As you ramp up the field, the velocity will approach a saturation value - this is the 'hot-electron' regime of transport. If you run Hall measurements on mobile carriers in a bulk semiconductor and a two-dimensional electron gas at varying temperatures and measure the mobility, you will see behavior depicted in the right part of Figure 3.

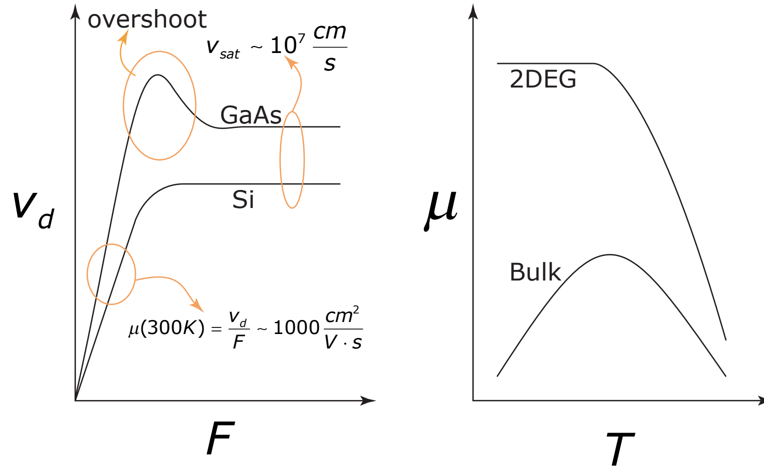

Figure 3: Velocity field curves (left) of Si and GaAs showing some definitions. Mobility versus temperature for generic semiconductor bulk and two-dimensional electron gas (2DEG). The figure is for motivating the study of transport theory

Most of our study in transport theory is geared towards understanding from a quantum-mechanical viewpoint the reasons for this kind of behavior (in Figure 3).

Setting up the Quantum Mechanical Problem

We begin with a completely general time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = (H_0 + H_1)\Psi(\mathbf{r},t) \tag{7}$$

We split the Hamiltonian into two parts -

$$H_0 = -\frac{\hbar^2}{2m^\star}\nabla^2 + E_C(\mathbf{r}) \tag{8}$$

The Hamiltonian which can be exactly solved. For a bulk semiconductor, H_0 is the crystal potential, which upon solution yields the bandstructure as the eigenvalues. We then concentrate on only the small region around the CB minima (bottom part, Figure 2).

The second part of the Hamiltonian is a small perturbation to the CB minimum introduced by impurities such as charged dopants, vacancies, or dislocations -

$$H_1 = Perturbation \tag{9}$$

Solving the exact part gives us the set of eigenfunctions $\Phi_m(\mathbf{r})$ and eigenvalues E_m corresponding to the Hamiltonian of the pure system, H_0)-

$$H_0\Phi_m(\mathbf{r}) = E_m\Phi_m(\mathbf{r}) \tag{10}$$

For the time dependent solution of the problem (with the scattering potentials), we expand the time dependent wavefunction in terms of the eigenfunctions -

$$\Psi(\mathbf{r},t) = \sum_m \psi_m(t)\Phi_m(\mathbf{r}) \tag{11}$$

The time-dependent equation (Equation 7) then becomes

$$i\hbar\frac{d}{dt}\psi_m = \sum_n \langle m|H|n\rangle\psi_n \tag{12}$$

Since the H_0 part has been solved exactly (Equation 10), we can use it in this equation to get the relation

$$\frac{d}{dt}\psi_m(t) + \frac{iE_m}{\hbar}\psi_m = \sum_n \frac{\langle m|H_1|n\rangle}{i\hbar}\psi_n \tag{13}$$

For a homogenous semiconductor, the solutions of the pure Hamiltonian are planewaves -

$$\Phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}}e^{i\mathbf{k}\cdot\mathbf{r}} \tag{14}$$

where $V = L_xL_yL_z$ is the sample volume. The quantization of wavevectors is given by $k_i = n_i\frac{2\pi}{L_i}, i = x, y, z$. The corresponding eigenvalues are given by

$$E_{\mathbf{k}} = E_C + \frac{\hbar^2|\mathbf{k}|^2}{2m^\star} \tag{15}$$

Fermi's Golden Rule

Evaluation of scattering rates requires us to use one of the most useful results of time dependent perturbation theory in quantum mechanics. Don't throw your hands up in despair, since it is not only simple, it is extremely useful too. Knowing it will help you figure out a lot of stuff in your research, no matter which field you are thinking about working in. The result is known by the glorified name of 'Fermi's Golden Rule'[4].

We will derive Fermi's Golden Rule (FGR[5]) with the following problem in mind. We have an incident electron with wavevector along the z direction (Figure 4) in a uniform bulk semiconductor of dimensions $V = L_x L_y L_z$. The semiconductor is pure, except for a small impurity with a time-independent potential $U(r)$ located somewhere in the bulk. The incident beam wavefunction is given by e^{ikz} where $k = |\mathbf{k}|$.

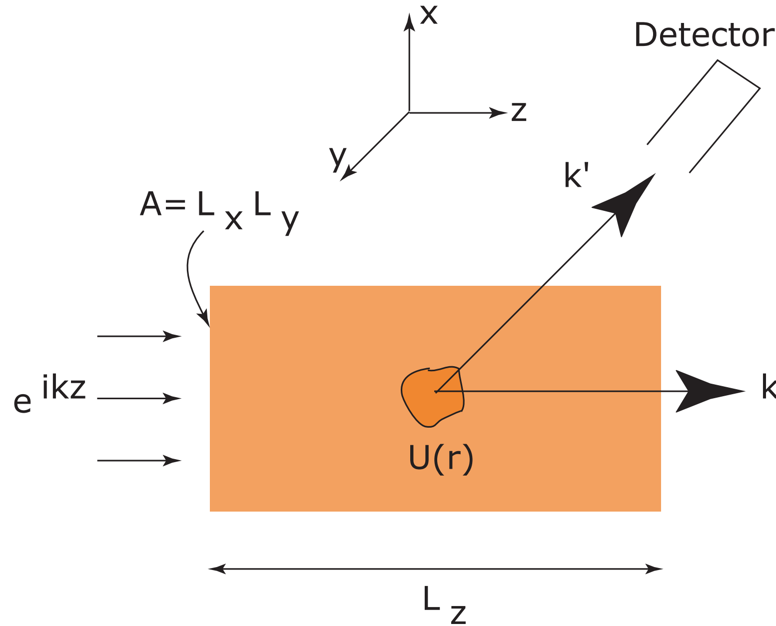

Figure 4: Figure for setting up the problem of scattering. An incident electron is scattered by a scattering potential $U(r)$ in an otherwise pure semiconductor. A detector is placed to find the angular dependence of scattering.

We have a detector at an angular position (θ, ϕ) and would like to find the probability $P_S(\theta, \phi)$ of finding the portion of the beam scattered by the potential $U(r)$ at that angular location. The result we derive is valid only in a weak scattering limit, i.e., the total scattering probability is much less than unity

$$P_S = \int_0^{\pi} d\theta \sin(\theta) \int_0^{2\pi} d\phi P(\theta, \phi) \ll 1 \tag{16}$$

[4]Golden is no exaggeration; however, refer to Kroemer's book, pg. for what he has to say about licensing the rule!

[5]I hate acronyms myself when I come across them. Please bear with me for this one three letter word!

In Equation 13, we saw how the time-coefficients of the solved wavefunction are linked to impurity potentials. A perturbation always causes transitions between allowed states of a exactly solved problem with known eigenfunctions and eigenvalues. In this problem, the exactly solved problem is again of the semiconductor bandstructure with allowed states indexed by the wavevectors $\langle \mathbf{k} \rangle$. The eigenfunction and eigenvalues are given by Equation 14 and Equation 15 respectively.

To evaluate the probability of a state in the eigenstate $\langle \mathbf{k} \rangle$ to scatter into any state $\langle \mathbf{k}' \rangle$, we have to find the amplitude of that state in the eigenfunction expansion and square it. The amplitudes are governed by Equation 13. WE now revert to the quantum number index $(m, n) \rightarrow (k, k')$ since the allowed states have been solved. Starting with the equation

$$\frac{d}{dt}\psi_{k'}(t) + \frac{iE_{k'}}{\hbar}\psi_{k'} = \sum_{k'' \neq k'} \frac{\langle k''|H_1|k'\rangle}{i\hbar}\psi_{k''} \tag{17}$$

The electron is incident with a wavevector along z at $t = 0$. So, the initial wavefunction is

$$\Psi(\mathbf{r}, t = 0) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{18}$$

From Equation 11, we get $\Psi(\mathbf{r}, t = 0) = \sum_{k'} \psi_{k'}(t = 0)\Phi_{k'}(\mathbf{r})$. So $\psi_{k'}(t = 0) = 1$ for $k' = k$ and 0 for all other k'. To find the amplitudes $\psi_{k'}(t)$ at all other times, we have to solve Equation 17. If the scattering potential is absent ($U(r) = 0$), then the equation is solved easily -

$$\psi^0_{k'}(t) = \psi^0_{k'}(0)e^{-i\frac{E_{\mathbf{k}'}t}{\hbar}} \tag{19}$$

We label the solution by '0' to emphasize that it is a zero order solution. From the condition 11, it is easy to see that only the amplitude of state $|\mathbf{k}'\rangle = |\mathbf{k}\rangle$ evolves with time, whereas the amplitude of all other states is 0. When there is a nonzero scattering potential, there are transitions between different eigenstates, so the amplitudes of $|\mathbf{k}'\rangle \neq |\mathbf{k}\rangle$ will not be zero anymore. However, the solution of Equation 17 is not simple anymore, and we have to take the route of successive approximations now. It is crucial to understand this part to appreciate the limitations of FGR.

We start by substituting the solution of the zero order solution on the RHS of Equation 17 and solve for the first order solution. This process is repeated for higher order solutions. However, if the scattering potential happens to be weak, it turns out that the first order solution is adequate and leads to Fermi's Golden Rule. Doing that, our time evolution equation becomes

$$\frac{d}{dt}\psi_{k'}(t) + \frac{iE_{k'}}{\hbar}\psi_{k'} = \frac{\langle k'|U(r)|k\rangle}{i\hbar}e^{-i\frac{E_{\mathbf{k}'}t}{\hbar}} \tag{20}$$

Subject to the initial condition Equation 19, the solution gives the amplitude of state $|\mathbf{k}'\rangle$ at time T

$$\psi_{k'}(T) = e^{-i\frac{E_{\mathbf{k}'}T}{\hbar}} \frac{\langle k'|U(r)|k\rangle}{i\hbar} \int_0^T e^{-i\frac{(E_{\mathbf{k}}-E_{\mathbf{k}'})T}{\hbar}} dt \tag{21}$$

Thus, the probability of finding an electron scattered into the final state $|\mathbf{k}'\rangle$ is the square of the amplitude of that state. That is exactly what we have determined in Equation 21. So,

$$P(\mathbf{k}',\mathbf{k}) = |\psi_{k'}(T)|^2 = \frac{|\langle\mathbf{k}'|U(r)|\mathbf{k}\rangle|^2}{\hbar^2} |\int_0^T e^{-i\frac{(E_{\mathbf{k}}-E_{\mathbf{k}'})T}{\hbar}} dt|^2 \tag{22}$$

T can be seen as the time taken for the electron to traverse the length L_z; it is thus given by $T = \frac{L_z}{v} = \frac{m^\star V}{\hbar k A}$, where $A = L_x L_y$ is the cross sectional area. The probability can be recast in the form

$$P(\mathbf{k}',\mathbf{k}) = T\frac{|\langle\mathbf{k}'|U(r)|\mathbf{k}\rangle|^2}{\hbar^2} g(E_{\mathbf{k}'} - E_{\mathbf{k}}) \tag{23}$$

where

$$g(\varepsilon) = \frac{1}{T}|\frac{\sin(\varepsilon T/2\hbar)}{(\varepsilon T/2\hbar)}|^2 \tag{24}$$

Using a special property of the function $\int_{-\infty}^{+\infty} d\varepsilon g(\varepsilon) = 2\pi\hbar$, we see that in the limit of a long time $T \to \infty$, the function g becomes a delta function. Thus, $Lim_{T\to\infty} g(\varepsilon) = 2\pi\hbar\delta(\varepsilon)$. This allows us to write the probability as

$$P(\mathbf{k}',\mathbf{k}) = TS(\mathbf{k}',\mathbf{k}) \tag{25}$$

where

$$S(\mathbf{k}',\mathbf{k}) = \frac{2\pi}{\hbar}|\langle\mathbf{k}'|U(r)|\mathbf{k}\rangle|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}}) \tag{26}$$

This is the traditional way of stating the famous FGR for time-independent stationary potentials. Note that this is not yet a measurable probability, since the measured probability will be a sum over all $|\mathbf{k}'\rangle$ states. Thus, to find a measurable probability, we need to sum the probability over all such final states.

$$P_S = \sum_{\mathbf{k}'} P(\mathbf{k}',\mathbf{k}) \tag{27}$$

Using the recipe $\sum_{\mathbf{k}'}(...) \to \frac{V}{(2\pi)^3}\int d^3k'(...)$, we get

$$P_S = \frac{V}{(2\pi)^3}\int_0^\infty \int_0^\pi \int_0^{2\pi} k'^2 dk' \sin(\theta) d\theta d\phi P(\mathbf{k}',\mathbf{k}) \tag{28}$$

Remembering the problem we started out with, we see that the angular probability can now be evaluated. Comparing with Equation 16, we write

$$P(\theta,\phi) = \frac{V}{(2\pi)^3}\int_0^\infty dk' k'^2 P(\mathbf{k}',\mathbf{k}) \tag{29}$$

Using FGR for $P(\mathbf{k}',\mathbf{k})$, we get

$$P(\theta,\phi) = \frac{VT}{4\pi^2\hbar}\int dk' k'^2 |\langle\mathbf{k}'|U(r)|\mathbf{k}\rangle|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}}) \tag{30}$$

We now use two important techniques. We transform the variable of integration to the argument of the delta function ($k' \to E_{\mathbf{k}'}$) using the identity $dk' = \frac{dk'}{dE_{\mathbf{k}'}} dE_{\mathbf{k}'} = \frac{m^\star k'}{\hbar^2} dE_{\mathbf{k}'}$. We do this to take advantage of a very important property of a delta function -

$$\int_a^b d\varepsilon \delta(\varepsilon - \varepsilon_0) F(\varepsilon) = F(\varepsilon_0) \tag{31}$$

if $a < \varepsilon_0 < b$, and is 0 otherwise. Using this useful result, we get

$$P(\theta, \phi) = \frac{VT}{4\pi^2 \hbar} \frac{m^\star k}{\hbar^2} |\langle \mathbf{k}' | U(r) | \mathbf{k} \rangle|^2 \tag{32}$$

The delta function enforced energy conservation in the scattering process ($E_{\mathbf{k}'} = E_{\mathbf{k}}$), and as a result, also required us to conserve the magnitude of momentum ($k' = k$). Using the explicit form of the matrix element $\langle \mathbf{k}' | U(r) | \mathbf{k} \rangle = \frac{1}{V} \int d^3 U(\mathbf{r}) e^{i(\mathbf{k}' - \mathbf{k}) \cdot r}$, we write the probability in the form

$$P(\theta, \phi) = \frac{(m^\star)^2}{4\pi^2 \hbar^4 A} | \int d^3 U(\mathbf{r}) e^{i(\mathbf{k}' - \mathbf{k}) \cdot r} |^2 \tag{33}$$

The angular dependence is hidden in the $(\mathbf{k}' - \mathbf{k})$ term. We will encounter it when studying ionized impurity scattering that has a strong angular dependence.

This result can be recast in a visually more appealing form by rewriting $P(\theta, \phi)$ as $P(\theta, \phi) = \sigma(\theta, \phi)/A$, where

$$\sigma(\theta, \phi) = \frac{(m^\star)^2}{4\pi^2 \hbar^4} | \int d^3 U(\mathbf{r}) e^{i(\mathbf{k}' - \mathbf{k}) \cdot r} |^2 \tag{34}$$

is referred to as the 'scattering cross section'. The total scattering scattering cross section σ_S is evaluated by integrating over all angular contributions

$$\sigma_S = P_S A = \int_0^\pi d\theta \sin(\theta) \int_0^{2\pi} d\phi \sigma(\theta, \phi) \tag{35}$$

Scattering cross section lets us visualize the scatterer as a solid obstacle of cross sectional area σ_S, which will scatter an electron with the probability σ_S/A. The typical problem in transport is electrons moving in a semiconductor with a density N_I of scattering centers (or impurities. In time t, the total number of scatterers an electron encounters will be $N_I A v t$ where v is the velocity. Thus, the probability that the electron will be scattered in time t is

$$Pr(t) = \frac{t}{\tau} = \frac{t}{\frac{1}{\sigma_S N_I v}} \tag{36}$$

This relation holds only for short intervals ($t \ll \tau$). For large t, it is $Pr(t) = 1 - e^{-\frac{t}{\tau}}$ (prove it!). Thus using $T = \frac{L_z}{v} = \frac{m^\star V}{\hbar k A}$, $P(\mathbf{k}', \mathbf{k}) = TS(\mathbf{k}', \mathbf{k})$, and $P_S = \sum_{\mathbf{k}'} P(\mathbf{k}', \mathbf{k})$, the mean free time between two scattering events (collisions) can be written as

$$\frac{1}{\tau_c} = \sigma_S v N_I = \sum_{\mathbf{k}'} (N_I V) S(\mathbf{k}, \mathbf{k}') \tag{37}$$

This relation is easy to understand. $\sum_{\mathbf{k}'} S(\mathbf{k},\mathbf{k}')$ is the rate at which an electron of momentum $\mathbf{k}$ is scattered by a single impurity. Multiplying by the total number of impurities $N_I V$, the total scattering rate is obtained.

To find the scattering time that goes into determining mobility, we have to use the momentum scattering rate, τ_m. Experimental mobility is related to the momentum scattering time by $\mu = e\tau_m/m^\star$. The scattering rate we just derived is the inter-collision time.

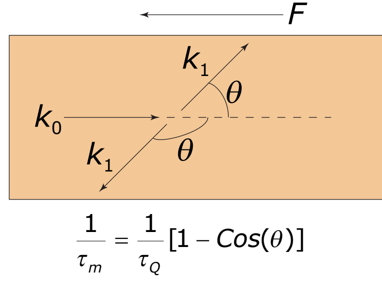

Figure 5: Figure to show the difference between the quantum lifetime τ_Q and the transport lifetime τ_m. If there is a small angle scattering, current is higher, and thus mobility is higher. For large angle scattering, current is lower, and mobility is lower. Thus, heuristically, we can say that the momentum lifetime and quantum lifetime (which is time between collisions) are related by $\frac{1}{\tau_m} = \frac{1}{\tau_Q}(1-\cos(\theta))$

Collision rate is $\frac{1}{\tau_Q}$ where τ_Q, the collision time determined by FGR is also called the quantum lifetime. From Figure 5, we note that some scattering events can change the angle of the $\mathbf{k}$ vector by very large amounts. Others have a weak effect. If there is small angle scattering, the current is higher, and hence the mobility is higher, so the momentum scattering *rate* is smaller. If the angle of scattering is large, momentum scattering rate will be smaller than the quantum scattering rate. This can be heuristically built into the formalism by writing

$$\frac{1}{\tau_m} = \frac{1}{\tau_Q}[1-\cos(\theta)] = \frac{1}{\tau_Q}[1-\frac{\mathbf{k_0}\cdot\mathbf{k_1}}{|k_0|^2}] \tag{38}$$

This result can be got by a rigorous analysis, so we believe in it and go ahead and use it to calculate mobility. Thus, momentum scattering rate is given by incorporating this small change to Equation 37 -

$$\frac{1}{\tau_c} = \sigma_S v N_I = \sum_{\mathbf{k}'}(N_I V)S(\mathbf{k},\mathbf{k}')(1-\cos(\theta)) \tag{39}$$

where $\cos(\theta) = \frac{\mathbf{k}\cdot\mathbf{k}'}{|\mathbf{k}|^2}$.

Example 2: An example of applying Fermi's Golden Rule

To get comfortable with Fermi's Golden rule we apply it to study scattering of a wavelike particle from a rectangular barrier of height $\triangle$ and thickness d.

Consider the problem of a wave incident on the rectangular barrier. The classical probability of reflection R is 1 for energy $E < \triangle$ and 0 for $E > \triangle$. Quantum mechanically, the behavior is more interesting. The reflection probability is given by

$$R = \frac{2R - 2R\cos(2k_1 d)}{1 + R^2 - 2R\cos(2k_2 d)} \tag{40}$$

where $R = |r|^2$ is the reflection probability for a single barrier, and $r = \frac{k_1 - k_0}{k_1 + k_0}$ is the reflection *amplitude* for a single barrier. Let us see how we can apply Fermi's Golden Rule to solve the same problem.

Fermi's Golden Rule will tell us the probability of scattering due to a small perturbation to the exactly solved Hamiltonian potential. Here, the exactly solved Hamiltonian is the free electron Hamiltonian. The solved wavefunctions are planewave states, given by

$$|\mathbf{k}\rangle = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{41}$$

If the particle is reflected from the barrier, we say that it has been scattered. FGR will tell us the probability of this event -

$$P_S = T\sum_{k'} S(k',k) = \sum_{\mathbf{k}'} \frac{2\pi T}{\hbar} |U(\mathbf{k}',\mathbf{k})|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}'}) \tag{42}$$

The barrier is in the z direction; the initial momentum vector of the electron points in the z direction - $k_z = k_0, k_x = k_y = 0$. So the matrix element of the barrier potential is given by

$$\langle \mathbf{k}'|\triangle|\mathbf{k}\rangle = \frac{\triangle}{V}\int_0^{L_x} dx e^{i(k_x - k'_x)x} \int_0^{L_y} dx e^{i(k_y - k'_y)y} \int_0^{d} dz e^{i(k_z - k'_z)z} \tag{43}$$

which, on using results of Example 1 evaluates to

$$\langle \mathbf{k}'|\triangle|\mathbf{k}\rangle = \frac{\triangle}{L_z}\delta_{k_x,k'_x}\delta_{k_y,k'_y}\left(\frac{e^{i(k_z - k'_z)d} - 1}{i(k_z - k'_z)}\right) \tag{44}$$

Thus, the scattered electrons have the same transverse momenta as the incident $(k_x, k_y) = (k'_x, k'_y)$. So, the sum over all scattered $\mathbf{k}'$ runs over only k'_z. Thus, we have

$$P_S = \sum_{k'_z} \frac{2\pi T}{\hbar} |U(\mathbf{k}',\mathbf{k})|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}'}) \tag{45}$$

where

$$|U(\mathbf{k}',\mathbf{k})|^2 = \frac{4\triangle^2}{L_z^2 (k - k'_z)^2}\sin^2\left(\frac{(k - k'_z)d}{2}\right) \tag{46}$$

Converting the sum over k'_z to an integral using the recipe $\sum_{k'_z}(...) \to \frac{L_z}{2\pi}\int dk'_z(...)$ and using the transformation $dk'_z = dE_{\mathbf{k}'}\frac{dk'_z}{dE_{k'}}$ we get

$$P_S = \frac{L_z T}{\hbar}\int dE_{\mathbf{k}'}\delta(E_{\mathbf{k}'} - E_{\mathbf{k}})\frac{dk'_z}{dE_{\mathbf{k}'}}|U_{\mathbf{k},\mathbf{k}'}|^2 \tag{47}$$

The delta function forces energy conservation; $\mathbf{k}' = \mathbf{k} \rightarrow k'_z = k_0$. Also, $\frac{dk'_z}{dE_{k'}} = \frac{m^\star}{\hbar^2 k'_z}$. Using this, the scattering probability becomes

$$P_S = \frac{m^\star T \triangle^2}{\hbar^3 k_0^3 L_z} \sin^2(k_0 d) \tag{48}$$

Using $T = \frac{L_z}{v_0} = \frac{m^\star L_z}{\hbar k_0}$, we get

$$P_S = (\frac{\triangle m^\star}{\hbar^2 k_0^2})^2 \sin^2(k_0 d) \tag{49}$$

This is the final scattering rate. It certainly does not look like the quantum mechanical result for probability of reflection as in Equation 38. The difference brings in focus the assumptions that are inherent in FGR. Consider the QM result. Let us see how the expression looks for *small* R. Then, neglecting R and R^2 in the D^r, the expression for reflection probability Equation 38 reduces to

$$R_B \simeq 2R - 2R\cos(2k_2 d) = 4R\sin^2(k_2 d) = \frac{4(k_2 - k_1)^2}{(k_2 + k_1)^2} \sin^2(k_2 d) \tag{50}$$

If R is small, $k_2 \simeq k_1 = k_0$, and the expression becomes

$$R_B \simeq (\frac{k_2 - k_1}{k_0})^2 \sin^2(k_0 d) \tag{51}$$

Since $k_2 = \sqrt{\frac{2m^*(E - E_C - \triangle)}{\hbar^2}}$ and $k_2 = \sqrt{\frac{2m^*(E - E_C)}{\hbar^2}}$, $k_2^2 - k_1^2 = \frac{2m^\star \triangle}{\hbar^2} \simeq 2k_0(k_2 - k_1)$. Using this, the result becomes

$$R_B \simeq (\frac{\triangle m^\star}{\hbar^2 k_0^2})^2 \sin^2(k_0 d) \tag{52}$$

This is exactly the result we got from the FGR calculation. So the assumption in FGR of a *weak* scattering potential is very important and should always be borne in mind when applying the rule.

FGR has widespread customers. A short list of it's clients other than semiconductor transport are optics, atomic processes, spectroscopy (Raman, NMR, etc) etc. Wherever there is a time-dependent transition between quantum mechanical states is involved, FGR rules.

Screening

In an insulator, atomic charges rearrange to screen external potentials. That is why we have the dielectric constant K, and the vacuum Coulomb potential $V(r) = \frac{e}{4\pi\epsilon_0 r}$ is screened to a value $V(r) = \frac{e}{4\pi\epsilon_0 K r}$. In a doped semiconductor, there are mobile carriers moving around and assist in screening further. The K gets changed. How? Let us see.

To understand the effect of screening, we start with an analogy. Consider a river with a rough bed. If there is a lot of water in the river (see Figure 5), a boat moving on the surface would be unaffected by the rough bed. The rough bed is 'screened' by water. However, if there is insufficient water, the rough bed shows up and impedes motion of the boat.

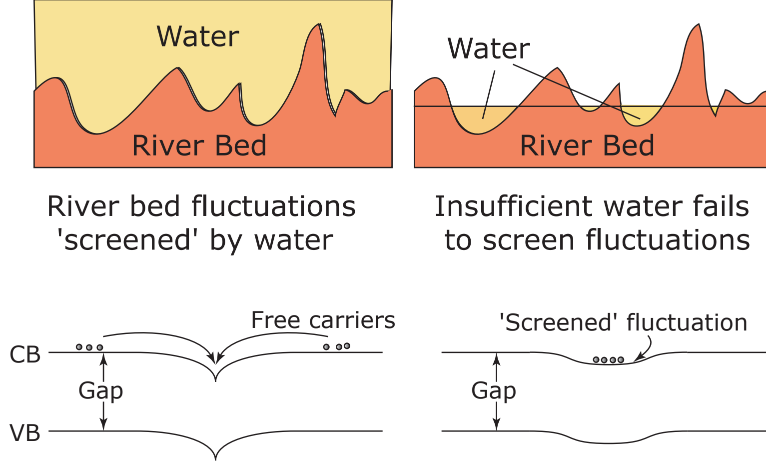

Figure 6: Figure to illustrate the phenomena of screening in semiconductors. The top figure gives an analogy of a river bed that is screened by water. The bottom figures show screening in action for a semcioductor.

Screening in semiconductors is very similar in behavior. Consider a band diagram shown in Figure 5. There is a impurity potential δV that causes a perturbation of the flat band diagram. This perturbation would be smoothed out by the flow of mobile carriers. We now treat the problem mathematically. The relation between the spatial carrier density variation and the conduction band edge is given by

$$n(\mathbf{r}) = N_C e^{-q\frac{E_C(\mathbf{r})-E_F}{k_B T}} \tag{53}$$

When there are no impurity potentials present, the conduction band edge is flat, $E_C(\mathbf{r}) = E_{C0}$. Then, the carrier density is uniform, and is given by $n_0 = N_C e^{-q\frac{E_{C0}-E_F}{k_B T}}$. Say the impurity potential causes a perturbation $E_C(\mathbf{r}) = E_{C0} + \delta V(\mathbf{r})$ in the conduction band edge. The carrier density then becomes

$$n(\mathbf{r}) = N_C e^{-q\frac{E_{C0}+\delta V(\mathbf{r})-E_F}{k_B T}} = n_0 e^{-q\frac{\delta V(\mathbf{r})}{k_B T}} \tag{54}$$

For *small* perturbations $\delta V(\mathbf{r}) \ll \frac{kT}{q}$, the exponential can be expanded using $e^x \simeq 1 + x$. Doing that, we get

$$n(\mathbf{r}) \simeq n_0 + \underbrace{\frac{q n_0}{k_B T}\delta V(\mathbf{r})}_{\delta n(\mathbf{r})} \tag{55}$$

For no extra charges in the semiconductor, the charge carriers are uniformly distributed and from charge neutrality,

$$\rho_u = q(p + N_D^+ - n - N_A^-) = 0 \tag{56}$$

holds. Let us consider just an $n - type$ semiconductor, where $p = N_A^- = 0$. The mobile charge changes to $n \rightarrow n_0 + \delta n(\mathbf{r}) = n(\mathbf{r})$ and the fixed ionized impurity occupations change to $N_D^+ \rightarrow N_{D0}^+ + \delta N_D^+(\mathbf{r}) = N_D^+(\mathbf{r})$. The $\delta(...)$ terms are in response to the perturbation $\delta V(\mathbf{r})$, which is screening in action. The screened potential can be got by solving Poisson's equation that relates charges to potentials -

$$\nabla^2(E_{C0} + \delta V(\mathbf{r})) = -\frac{\rho(\mathbf{r})}{\epsilon(0)} = -\frac{q(N_D^+ + \delta N_D^+(\mathbf{r})) - (n_0 + \delta n(\mathbf{r}))}{\epsilon(0)} \quad (57)$$

Flatband conditions take out the constant terms in the equation, and we are left with only the $\delta(...)$ terms -

$$\nabla^2 \delta V(\mathbf{r}) = -\frac{q(\delta N_D^+(\mathbf{r}) - \delta n(\mathbf{r}))}{\epsilon(0)} \quad (58)$$

Using our approximations $\delta n(\mathbf{r}) \simeq \frac{qn_0}{k_B T}\delta V(\mathbf{r})$ and $\delta N_D^+(\mathbf{r}) \simeq \frac{-qN_{D0}^+}{k_B T}\delta V(\mathbf{r})$, we get

$$\nabla^2 \delta V(\mathbf{r}) = \frac{q^2 n^\star}{\epsilon(0) k_B T}\delta V(\mathbf{r}) = \frac{1}{\lambda_D^2}\delta V(\mathbf{r}) \quad (59)$$

Where we have defined $\lambda_D = \sqrt{\frac{\epsilon(0)k_B T}{q^2 n^\star}}$, the Debye screening length. Here, $n^\star$ is an effective carrier concentration that is effective in screening. It is generally not equal to the free mobile carrier density. For a n-type sample the effective electron screening concentration is given by

$$n^\star = n + \frac{n(N_D - n)}{N_D} \quad (60)$$

For completely ionized non-degenerate carriers, it is a good approximation to assume the screening electron density to be doping density, i.e., $n^\star = N_D$.

Now we have a equation that we can apply to solve for any fluctuating potential $\delta V(\mathbf{r})$. For the special case of a spherically symmetric fluctuation (example, Coulombic potential around a charged donor), Equation 57 becomes

$$\frac{d^2}{dr^2}(r\delta V(r)) = \frac{r}{\lambda_D^2}\delta V(r) \quad (61)$$

The solution for this equation for the Coulombic potential $V(r) = \frac{q^2}{4\pi\epsilon_0 K r}$ is the screened potential $V(r) = \frac{q^2}{4\pi\epsilon_0 K r}e^{-r/\lambda_D}$. The Coulombic potential is thus 'screened' by the Yukawa potential which reduces the potential very quickly within a few Debye lengths. We will use this result in evaluating ionized impurity scattering rates later.

For two-dimensional electron gases, it turns out that screening is much weaker. There is a surprising result that the screening length is a constant independent of the 2DEG density for all practical purposes. The 2DEG screening length is called the Thomas-Fermi length, and is given by

$$\lambda_{TF} = \frac{2\pi\epsilon_0 K \hbar^2}{q^2 m^\star} = \frac{a_B^\star}{2} \quad (62)$$

where $a_B^\star$ is the effective Bohr radius in the semiconductor.

From the form of FGR, we see that the scattering potential $U(r)$ always appears as a matrix element

$$U(q) = \langle k+q|U(r)|k\rangle = \underbrace{\int d^3re^{-iqr}U(r)}_{FT(U(r))} \tag{63}$$

where $\mathbf{q} = \mathbf{k'} - \mathbf{k}$. It is easy to see that this is nothing but the Fourier transform of the potential $U(r)$. Fourier transform is an extremely useful technique to solve such problems. To start off, let me state a very useful result[6] - the 3D Fourier transform of the bare Coulomb potential $U(r) = \frac{e^2}{4\pi\epsilon_0 r}$ is given by

$$U(q) = \frac{e^2}{\epsilon_0 q^2} \tag{64}$$

For a semiconductor without free carriers, this is scaled by the static dielectric constant K - $U(q) = \frac{e^2}{\epsilon_0 K q^2}$. Note that it diverges as $q \to 0$ - this is indicative of the long-range nature of the Coulomb potential. We derive the Fourier transform for a screened Coulomb potential $U_{sc}(r) = \frac{e^2}{4\pi\epsilon_0 K r}e^{-r/\lambda_D}$ in the following.

$$U_{sc}(q) = \int d^3e^{i\mathbf{q}\cdot\mathbf{r}}U_{sc}(r) \tag{65}$$

Align $z - axis$ along $\mathbf{q}$.

$$U_{sc}(q) = \frac{e^2}{4\pi\epsilon^2 K}\int r^2\sin(\theta)drd\theta d\phi\frac{e^{-\frac{r}{\lambda_D}}}{r}e^{iqr\cos(\theta)} \tag{66}$$

Writing $1/\lambda_D = q_D$, and evaluating this integral gives us

$$U_{sc}(q) = \frac{e^2}{\epsilon_0 K(q^2+q_D^2)} = \frac{U(q)}{\epsilon(q)} \tag{67}$$

where $\epsilon(q) = 1 + \frac{q_D^2}{q^2}$ is a scaling factor that the bare Fourier transform can be divided by to get the screened Fourier transform. $q_D = 1/\lambda_D$ is the Debye-Huckel wavevector. Note that screening has removed the divergence as $q \to 0$. Thus screening removes the long-range components of the Coulomb potential.

For a 2DEG, the Fourier Transform scaling factor after screening is given by

$$\epsilon(q) = 1 + \frac{q_{TF}}{q} \tag{68}$$

where $q_{TF} = 1/\lambda_{TF}$. Screening is weaker for a 2DEG than in 3D[7]. So the Fourier transform of any screened Coulombic potential will be the Fourier transform of the bare potential divided by this factor.

[6] pg 350, John Davies

[7] John Davies, pg 352

1 Case studies of scattering in semiconductors

Low field mobility is calculated by doing the following -

1) Identify the defect (perturbation) and choose a suitable model for it's potential $U_i(r)$ for the ith type of impurity.
2) Calculate the momentum scattering rate for the particular type of impurity and sum it over all available final states $\mathbf{k}'$.
3) Calculate the total scattering rate using the Matheissen's rule - $\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}$.
4) Average the momentum scattering rate for drift velocity (Eqn 74).
5) Calculate the mobility using $\mu = e\tau_m/m^\star$.

We have gone through the whole formalism and have armed ourselves now with all the tools needed for doing the above steps. The following sections would be really short since we have done all the hard work earlier.

Case study - Ionized impurity scattering in bulk semiconductors and 2DEGs

Consider an ionized impurity in a bulk semiconductor. We find the quantum scattering time τ_Q for scattering from state $|\mathbf{k}\rangle$ to a general final state $|\mathbf{k}'\rangle$. The impurity may be an intentional donor or acceptor, or a charged vacancy state or any such impurity. Let the charge on the impurity be Z Coulombs. Equation 34, Equation 67 yield

$$\sigma(\theta,\phi) = \frac{4L_D^4}{a_0^2}\frac{1}{(1+qL_D)^2} \tag{69}$$

where $a_0 = \frac{4\pi\epsilon_0 K\hbar^2}{m^\star q^2}$, the effective Bohr radius. Since Coulombic scattering is elastic, $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ can be written as

$$q^2 = 2k^2[1-\cos(\theta)] \tag{70}$$

where $k = |\mathbf{k}| = |\mathbf{k}'|$. Using this, writing $\gamma = 2kL_D$ and integrating over all angles (θ,ϕ)[8], we get the total scattering cross section as

$$\sigma_S = \frac{\pi}{k^4 a_0^2}\frac{\gamma^4}{1+\gamma^2} \tag{71}$$

Thus, from Eq. 37, the quantum scattering rate is given by

$$\frac{1}{\tau_Q(k)} = N_I \cdot \frac{\hbar k}{m^\star} \cdot \frac{\pi}{k^4 a_0^2}\frac{\gamma^4}{1+\gamma^2} \tag{72}$$

The momentum relaxation rate can be got similarly by using Equation 39. The result is

$$\frac{1}{\tau_m(k)} = N_I \cdot \frac{\hbar k}{m^\star} \cdot \frac{2\pi}{k^4 a_0^2}[ln(1+\gamma^2) - \frac{\gamma^2}{1+\gamma^2}] \tag{73}$$

[8] From now on, you need to evaluate all integrals yourself and verify the results!

The total momentum scattering time is found by an averaging procedure. It is easy to convert $\tau_m(k) \rightarrow \tau_m(E)$, since the dispersion $E(k) = \frac{\hbar^2 k^2}{2m^\star}$ is known. For drift velocity along an applied field F_z (along the z axis), the averaging[9] procedure for a Fermi-Dirac distributed carrier population is given by

$$\langle \tau_m \rangle = -\frac{2}{3} \int_0^\infty \frac{\tau_m(E) \frac{\partial f_0}{\partial (E/k_B T)} (E/k_B T)^{3/2} d(E/k_B T)}{\int_0^\infty f_0 (E/k_B T)^{1/2} d(E/k_B T)} \tag{74}$$

Using this recipe, we get the mobility as

$$\mu_{imp} = \frac{e\langle \tau_m \rangle}{m^\star} = \frac{2^{7/2}(4\pi\epsilon_0 K)^2 (k_B T)^{3/2}}{\pi^{3/2} Z^2 e^3 (m^\star)^{1/2} N_I [ln(1+\gamma_0^2) - \frac{\gamma_0^2}{1+\gamma_0^2}]} \tag{75}$$

where Z is the total electron charges in the impurity and $\gamma_0 = 2\frac{m^\star}{\hbar}(\frac{2}{m^\star}3k_B T)^{1/2} L_D$. The dependence of this mobility on temperature T and impurity density N_I is seen to be

$$\mu_{imp} \propto \frac{T^{\frac{3}{2}}}{N_I} \tag{76}$$

In bulk semiconductors at low temperatures, ionized impurities set the mobility limits. This is the mobility dependence in Figure 3 at low temperatures (right curve, increasing mobility).

For 2DEGs, the phenomena of quantization and degeneracy of conducting electrons (all conducting electrons have $k \simeq k_F = \sqrt{2\pi n_{2D}}$) change the averaging procedure[10] and make the problem easy. Also, the remote location of impurities (modulation doping) introduces an exponential decay of the Coulomb potential. This results in much higher mobilities. An important fact is there is no temperature activation of carriers required for a 2DEG; the 2DEG density is independent of temperature. Thus, the ionized impurity limited mobility is *independent of temperature* for a 2DEG. This results in the 'flat' mobility saturation at low temperatures for 2DEGs (right plot, Figure 3).

Other scattering mechanisms and where to read about them

At room temperature, the mobility in a reasonably pure semiconductor is almost always limited by phonon scattering. Phonon scattering is very important for operational devices and is treated well in Seeger's book (Semiconductor Physics). In fact, all scattering mechanisms are very well treated in this book. I highly recommend it if you want to learn more on transport and scattering. However, Seeger's book is an 'old' book, it does not include quantum structures like quantum wells, etc. For a newer treatment (though not as comprehensive as Seeger), John Davies' book (Physics of low dimensional semiconductors) is recommended. Another book which does a good and compact treatment of transport properties is Wolfe, Holonyak and Stillman's 'Physical properties of semiconductors'. For a comprehensive review of 2DEG transport properties, and every conceivable property of the 2DEG, refer to the 'Bible' of 2DEGs - a journal article that can be downloaded from the web. The article is

[9] Seeger, Semiconductor Physics, Pg 50

[10] John Davies, Pg 356

by Ando, Fowler and Stern; it appeared in Reviews of Modern Physics, (Rev. Mod. Phys. 54, 437-672 (1982)).

Alloy disorder scattering is important for transport in ternary (or even binary - SiGe) alloys. For 2DEGs, interface roughness scattering becomes severe at high 2DEG densities. Dislocation scattering is proving to be a hindrance for the III-V nitride semiconductors making it bigger than what they already are.

High field transport

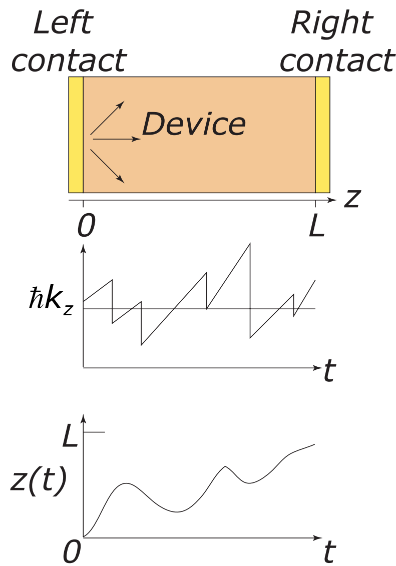

Figure 7: Figure depicting the actual transport process with time. The second shows how the z component of momentum increases linearly with time ($k_z(t) = \frac{F}{\hbar}t$) till it gets scattered to a random state. The third figure shows the progress of the particle in real space with time as it traverses the length of the device.

Consider an electron moving in the presence of impurity scattering with an increasing electric field. In the 'ohmic' regime we defined a mobility as the ratio of drift velocity and field. Now in the microscopic picture, the electron starts from zero velocity, accelerates by gaining energy from the field, and then scatters. This is the way it traverses the length of a semciconductor device. It is depicted in Figure 7.

After every scattering event, the electron starts gaining momentum (and hence energy) from the applied field. The energy it acquires before the next collision increases as the electric field increases. At a particular value of the electric field, the energy gained between two collisions becomes larger than the optical phonon energy $\hbar\omega_{op}$. When this happens, the probability of the electron losing all it's energy to the lattice by the emission of an optical phonon becomes large. The emission of a phonon results in no increase in current; it just heats the device. Any further increase of the field will cause an increase in the number of

optical phonons emitted, with no increase in the drift velocity. The drift velocity is said to be saturated.

Thus, a fair back of the envelope estimate of the saturated drift velocity can be made by equating the kinetic energy at the saturated drift velocity $KE = \frac{1}{2}m^\star v_d^2$ to the optical phonon energy in the semiconductor -

$$v_d^{sat} \sim \sqrt{\frac{2\hbar\omega_{op}}{3m^\star}} \tag{77}$$

For GaN, $\hbar\omega_{op} = 92meV, m^\star = 0.2m_0$, and the drift velocity estimate is $v_d^{sat} \sim 4 \times 10^7 cm/s$. This is a good 'order of magnitude' estimate. The true value is $v_d^{sat} \sim 2 \times 10^7 cm/s$. Thus, the velocity-field curve saturates; this is shown in Figure 3, left graph.

In direct gap semiconductors (GaAs, GaN, etc) there are two CB minimum in the $E - k$ diagram. Γ valley is the lowest and X valley is the second lowest valley. The X valley has a smaller curvature and hence a larger effective mass. At high fields, when high k^s are achieved between collisions, it becomes possible for electrons to transfer from the Γ tot eh X valley; the onset of this phenomena causes a region where the velocity *decreases* with increasing field. This is the origin of negative differential resistance (NDR) which is put to use to make Gunn oscillators which have very high frequency oscillations.

The velocity-field characteristics are of utmost importance in submicron devices. The reason is that for the same applied voltage, the field in a small length device will be large. Thus, saturation of drift velocity is achieved under fairly low voltages. With the miniaturization of transistors and decreasing gate lengths, it is very important for a device engineer to understand the hidden theories behind the velocity-field characteristics for innovation.

Transport Regimes

In a sufficiently pure semiconductors, electron wavelengths are sufficiently delocalized and can spread over large distances (large $\lambda \rightarrow$ small k). For such cases, transport occurs in the bands, this regime of transport is called band transport. Figure 8 shows schematically this form of transport. For low field regime, the transport proceeds by scattering from impurities. For high field regime, there are optical phonon emissions, but all transport is in the conduction (or valence) band.

If there is a lot of disorder in a semiconductor (say an polycrystalline semiconductor with grain boundaries), then there will be small localized pockets of carriers which will have to surmount the potential barriers between them to carry current. This process requires an activation energy, and the activation energy can be measured. This form of transport is very different in characteristic from band transport. The transport is called 'activated' or 'hopping' transport because the carriers hop from one localized state to the other. Theoretical treatment of transport in such disordered systems requires percolation theory and sophisticated techniques such as Green's functions, which I leave for you to explore.

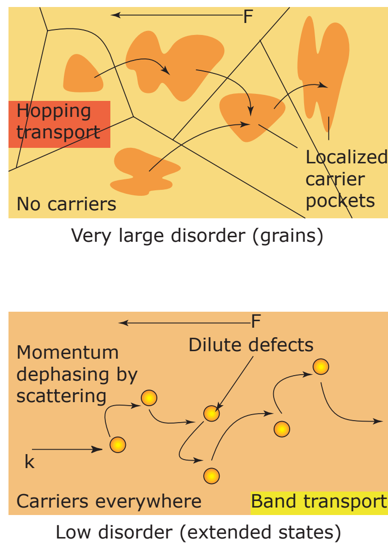

Figure 8: Band transport and Hopping regimes of transport. Band transport has been discussed in this work. Hopping transport requires activation to surmount the barriers between nearby localized pockets of carriers.

Current research, future directions

I list some topics of intense current interest. These are related to topics on transport theory covered in the class and you should be able to read papers related to these topics -

1) Transport of electrons in 2DEGs - The integral and fractional quantum Hall effects. It has already generated four Physics Nobel prizes.(Von Klitzing, Stormer, Tsui, Laughlin)
2) Transport in Disordered electronic systems - Metal-Insulator transistions in bulk and 2DEGs, Localization, etc. Has garnered at least 2 Physics Nobel prizes.(Mott, Anderson)
3) Ballistic transport through quantum point contacts, quantized conductance.
4) Transport in Quantum Wires, Dots.
5) Transport in Nanotubes and Nanowires.
6) Spin transport and spin scattering in dilute magnetic semiconductors.