Quantum Physics of Electron Statistics, Ballistic Transport, and Photonics in Semiconductor Nanostructures

Debdeep Jena (djena@cornell.edu)
Departments of ECE and MSE, Cornell University

1. Introduction

This article discusses the quantum physics of electron and hole statistics in the bands of semiconductors, the quantum mechanical transport of the electron and hole states in the bands, and optical transitions between bands. The unique point of view presented here is a unified picture and single expressions for the carrier statistics, transport, and optical transitions for electrons and holes in nanostructures all dimensions - ranging from bulk $3d$, to $2d$ quantum wells, to $1d$ quantum wires. The focus in the early parts is on nanostructures of dimensions $d = 1, 2, 3$ that allow transport, and the $0d$ quantum dot case is discussed for photonics.

2. Electron Energies in Semiconductors

Electrons in free space have continuous values of allowed energies $E(k) = \frac{\hbar^2 k^2}{2m_e}$, where $\hbar = h/2\pi$ is the reduced Planck's constant, m_e the rest mass of an electron, and $k = 2\pi/\lambda$ is the wavevector. $\hbar k = \frac{h}{\lambda} = p$ is the momentum of the free electron by the de Broglie relation of wave-particle duality. A periodic potential $V(x + a) = V(x)$ in the real space of a crystal, when included in the Schrodinger equation, is found to split the continuous energy spectrum electron energies $E(k) = \frac{\hbar^2 k^2}{2m_e}$ into bands of energies $E_m(k)$, separated by energy gaps. The mth allowed energy band is labeled $E_m(k)$. The states of definite energy $E_m(k)$ have real-space wavefunctions $\psi_k(x) = e^{ikx} u_k(x)$, where $u_k(x + a) = u_k(k)$, called Bloch functions. Each band has exactly the same number of allowed electron states N as the number of unit cells in the real space. Each state of energy $E_m(k_n)$ in each band m is indexed by a unique $k_n = \frac{2\pi}{L} n$ with $n = 0, \pm 1, \pm 2$, where $L = Na$ is the macroscopic size of the crystal. The k_n lies in the first Brillouin zone, between $-G/2 \leq k \leq +G/2$, where $G = 2\pi/a$ is the reciprocal lattice vector. The Pauli exclusion principle allows each state to be occupied by $g_s = 2$ electrons of opposite spins. Thus, each band $E_m(k)$ can hold a maximum of $2N$ electrons. The bands have the property $E_m(k + G) = E_m(k)$, i.e., they are periodic in the $k-$space.

The k of the bands $E_m(k)$ have a different meaning from the free electron wavevector. $\hbar k$ is not the momentum of the electron of a unique energy $E_m(k)$. States of definite energy (or energy eigenstates) $E_m(k)$ are *not* states of definite momentum in a periodic crystal[1]. The Bloch states are a mixture of states of momenta $\hbar(k \pm G)$, which is why $\hbar k$ is referred to as the 'crystal momentum'.

The group velocity of state k in energy band $E(k)$ is $\mathbf{v}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k})$. In response to an external force $\mathbf{F}$, the crystal momentum changes according to $\mathbf{F} = \hbar \frac{d\mathbf{k}}{dt}$; the energies and velocities change to satisfy this relation.

The two statements above are the most remarkable results of the quantum mechanics of electrons in crystals, and are by no means obvious. The states at the very bottom *and the very top* of any band $E_m(k)$ must have zero group velocity. An empty band cannot carry a charge current because there are no electrons in them. A filled band *also* cannot carry a net charge current - but for a different reason. In a filled band, the currents carried by a filled state k has an exactly equal partner, but opposite in sign at $-k$, and the net current cancels and gives zero: $J_{filled} = \frac{q}{L^d} \sum_k f(k) v_g(k) = \frac{q}{L^d} \sum_k v_g(k) = 0$ for a filled band, where $f(k) = 1$ is the occupation probability of state k, q is the electron charge, and L^d the volume in d dimensions. An empty band can carry current only if electrons are put into it - by doping, electrostatically, optically, or thermally. A filled band can carry a net current only if electrons are *removed* from it - again, by doping, electrostatically, optically, or thermally. When electrons are removed from the top of a filled band, the net current is $J = \frac{q}{L^d} \sum f(k) v_g(k) = \frac{q}{L^d} \sum v_g(k) + \frac{(-q)}{L^d} \sum_k [1 - f(k)] v_g(k) = \frac{(-q)}{L^d} \sum_k [1 - f(k)] v_g(k)$, implying the current is

---

[1] A crude classical analogy is a ball moving at a constant velocity in a flat plane will maintain its momentum. But if it rolls in a periodic potential, its velocity increases and decreases periodically: its momentum is not fixed.

effectively carried by 'holes' or empty electron states, which behave in transport and electrostatically as positive charges $-q$, where q is the normal electron charge.

A crystal is an intrinsic semiconductor if there is a few eV bandgap between a completely filled valence band $E_v(k)$ and the next conduction band $E_c(k)$ that is completely empty. At the very bottom of the conduction band, we can expand the energies as the parabola $E_c(k) \approx E_c + \frac{\hbar^2 k^2}{2m_c^\star}$, where $E_c(0) = E_c$ is the band-edge, or lowest energy, and $m_c^\star$ is the conduction band edge effective mass. Similarly, for the top of the valence band is approximated as $E_v(k) = E_v - \frac{\hbar^2 k^2}{2m_v^\star}$, with $E_v(0) = E_v$ the valence band edge, and $m_v^\star$ the valence band effective mass. $E_c - E_v = E_g$ is the energy bandgap of the semiconductor. If the crystal momentum k at which the conduction band reaches the minimum and the valence band reaches its maximum are exactly the same, we have a 'direct bandgap' semiconductor, and if they are not, it is an 'indirect bandgap' semiconductor.

3. Semiconductor Statistics

We will build the 'Map of Semiconductor Physics' in Table 1. Some general results for semiconductor structures are now discussed, in which electrons and holes are allowed to move in d dimensions. $d = 3$ is the case in an ordinary 3D bulk semiconductor, $d = 2$ are 2D semiconductor membranes or heterostructure quantum wells, $d = 1$ are quantum wires, and $d = 0$ are quantum dots. The density of states of the conduction band in d-dimensions is given by

$$g_c^d(E) = \frac{g_s g_v}{(4\pi)^{\frac{d}{2}} \Gamma(\frac{d}{2})} \left(\frac{2m_c^\star}{\hbar^2}\right)^{\frac{d}{2}} (E - E_c)^{\frac{d-2}{2}}, \tag{1}$$

where g_s is the spin-degeneracy which is $= 2$ for conventional situations, and g_v is the valley degeneracy. The units are in $1/(eV \cdot cm^d)$. The valley degeneracy of the conduction band $g_v = 1$ for most direct-bandgap semiconductors in 3D, but is $g_v = 6$ for 3d Silicon, and $g_v = 4$ for 3d Germanium. $\Gamma(...)$ is the Gamma function, with $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$, and $\Gamma(n+1) = n\Gamma(n)$. For $d = 3$ the DOS increases as $\sqrt{E - E_c}$, for $d = 2$, the DOS is constant, and for $d = 1$, the DOS decreases as $1/\sqrt{E - E_c}$. The dependence on the conduction band edge effective mass is $(m_c^\star)^{\frac{d}{2}}$, which means a conduction band with a heavier effective mass has a higher DOS in *all* dimensions. Exactly in the same way, the valence band DOS has $m_v^\star$ for the mass, and the energy dependence on dimensions is $(E_v - E)^{\frac{d-2}{2}}$, the same as for the conduction band, except the argument is $E_v - E$ for obvious reasons at the top of the valence band.

At equilibrium, the Fermi level E_F determines the number of electrons in the conduction band n_d and the number of holes p_d in the valence band. On the other hand, at equilibrium, if we know either the number of electrons in the conduction band n_d or holes p_d in the valence band, the Fermi level is uniquely determined. There is a one-to-one correspondence between the densities and E_F at equilibrium. The relation is obtained by using the Fermi-Dirac occupation function with the density of states:

$$n_d = \int_{E_c}^{\infty} dE \cdot g_c^d(E) \cdot \frac{1}{1 + e^{\frac{E_c(k) - E_F}{k_b T}}} = N_c^d \cdot F_{\frac{d-2}{2}}\left(\frac{E_F - E_c}{k_b T}\right). \tag{2}$$

The prefactor $N_c^d = g_s g_v \left(\frac{2\pi m_c^\star k_b T}{h^2}\right)^{\frac{d}{2}}$ is called the effective band-edge DOS, and has units of $1/cm^d$ where d is the dimension. For typical semiconductors, $N_c^{1d} \sim 10^6/cm$, $N_c^{2d} \sim 10^{12}/cm$, and $N_c^{1d} \sim 10^{18}/cm^3$ at $T = 300$ K. The dimensionless factor $F_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^{\infty} du \frac{u^j}{1 + e^{u-\eta}}$ is the Fermi-Dirac integral of order j, the values of which may be obtained from tables, by calling functions in Matlab or Mathematica, or by direct integration. Two important limits of the Fermi-Dirac integral are $F_j(\eta) \approx e^\eta$ for $\eta << -1$, and $F_j(\eta) \approx \frac{\eta^{j+1}}{\Gamma(j+2)}$ for $\eta >> +1$. The dimensionless argument is $\eta = \frac{E_F - E_c}{k_b T}$, which is a measure of how far the Fermi-level E_F is from the band edge E_c, measured in units of the thermal energy $k_b T$. If for example, the Fermi level is at $E_F = E_c - 6k_b T$ for a 3D semiconductor with $N_c^{3d} = 10^{18}/cm^3$, $\eta = \frac{E_F - E_c}{k_b T} = -6$, and the electron density in the conduction band is $n_{3d} = N_c^{3d} \cdot F_{\frac{1}{2}}(\eta) \approx N_c^{3d} \cdot e^{-6} \approx \frac{10^{18}}{400} \approx 2.5 \times 10^{15}/cm^3$. On the other hand, if for the same semiconductor $E_F = E_c + 3k_b T$, $\eta = +3$, and the electron density is

$n_{3d} = N_c^{3d} \cdot F_{\frac{1}{2}}(\eta) \approx N_c^{3d} \cdot \frac{\eta^{\frac{1}{2}+1}}{\Gamma(\frac{1}{2}+2)} = 10^{18} \cdot \frac{3^{\frac{3}{2}}}{\frac{3\sqrt{\pi}}{4}} \sim 4 \times 10^{18}/\text{cm}^3$. Similar arguments carry over to find the statistics and densities of holes in the valence band.

In the absence of dopants and impurities, an intrinsic semiconductor is charge neutral. This implies if there are any electrons thermally ionized from the valence to the conduction band, there must be an equal number of holes in the valence band. This sets $n_d = p_d$, or $N_c^d F_{\frac{d-2}{2}}(\frac{E_F - E_c}{k_b T}) = N_v^d F_{\frac{d-2}{2}}(\frac{E_v - E_F}{k_b T})$, which is only possible for an unique E_F. This charge neutrality condition fixes the location of the Fermi level E_F. The product of the intrinsic carrier densities is

$$n_d \cdot p_d = N_c^d \cdot N_v^d \cdot F_{\frac{d-2}{2}}(\frac{E_F - E_c}{k_b T}) \cdot F_{\frac{d-2}{2}}(\frac{E_v - E_F}{k_b T}) \approx N_c^d \cdot N_v^d \cdot e^{-\frac{E_g}{k_b T}}, \tag{3}$$

where the approximation on the right holds only for non-degenerate carrier distributions, when $\frac{E_F - E_c}{k_b T} << -1$, and $\frac{E_v - E_F}{k_b T} << -1$. This is true when E_F is far inside the gap. In this case, the intrinsic electron or hole density in $d-$dimensions is obtained explicitly as $n_{di} = p_{di} = \sqrt{N_c^d N_v^d} \cdot e^{-\frac{E_g}{2k_b T}}$, and is exponentially dependent on the bandgap. For example, in 3d Silicon at $T = 300$ K, $n_i \approx 10^{10}/\text{cm}^3$.

In the presence of dopants and impurities, at equilibrium a bulk semiconductor is *still* charge neutral. For a donor density N_D and donor state at energy E_D (meaning the ionization energy is $E_g - E_D$, using the valence band edge as the reference), and an acceptor density N_A and acceptor ionization energy E_A, the new charge neutrality condition for a d-dimensional semiconductor is $n_d + N_A^- = p_d + N_D^+$, which is written explicitly as

$$N_c^d F_{\frac{d-2}{2}}(\frac{E_F - E_c}{k_b T}) + \frac{N_A}{1 + 2e^{\frac{(E_A - E_F)}{k_b T}}} = N_v^d F_{\frac{d-2}{2}}(\frac{E_v - E_F}{k_b T}) + \frac{N_D}{1 + 2e^{\frac{-(E_D - E_F)}{k_b T}}}. \tag{4}$$

The solution to this equation gives the Fermi level E_F and the carrier density explicitly. The number of valence bands considered determines the prefactor of the exponent in the N_A^- term. It is 2 for a single valence band, and 4 when light hole and heavy hole bands are considered. The non Fermi-Dirac statistics of the ionized dopants occurs because of on-site Coulomb repulsion upon occupation of more than one electron at a dopant site.

4. Ballistic Transport in Semiconductors

Let the conduction band of a $d-$dimensional semiconductor have a fixed electron density[2] of n_d. If the electron states in this conduction band are connected to the electron states in two transparent (= ohmic) metal contacts, and a voltage V is applied across it, the resulting ballistic current is obtained in the following manner. The left electrode is called the source, with a Fermi level E_{Fs}, and the right electrode the drain with Fermi level E_{Fd}, where $E_{Fs} - E_{Fd} = qV$. The electron population of density n_d in the semiconductor conduction band at the 'source injection point' is split into right-going states and left going states. In ballistic transport, there is no scattering - meaning states moving to the right are in equilibrium with the source, and states moving to the left are in equilibrium with the drain. Because the net carrier density is not changed by the applied drain voltage, we have the relation

$$n_d = \frac{1}{2}N_c^d[F_{\frac{d-2}{2}}(\frac{E_{Fs} - E_c}{k_b T}) + F_{\frac{d-2}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]. \tag{5}$$

Solving this equation at the applied voltage V allows the unknown E_{Fs}, and consequently $E_{Fd} = E_{Fs} - qV$ to be uniquely determined. When $V = 0$, the left-going electron and right-going electron densities are identical, $E_{Fs} = E_{Fd} = E_F$, and the current is zero. For small voltages, E_{Fs} and E_{Fd} split nearly symmetrically around the zero-bias E_F. For large voltages, the argument $\frac{(E_{Fs} - qV) - E_c}{k_b T}$ of the second term which represents the left-going carriers at the source injection point becomes large and negative, meaning $F_{\frac{d-2}{2}}(\eta) \approx e^{\eta}$ becomes very small. In this case, almost all electron in the band are moving to the right, and almost none to

---

[2]The density may be chemically fixed by doping it with donors, or capacitively fixed by applying a voltage across a gate capacitor.

the left. Under this situation, E_{Fs} is in the band, but E_{Fd} enters the gap. For voltages larger than this voltage, E_{Fs} does not change, and the increase in voltage goes into lowering E_{Fd}.

Because of the imbalance of left- and right-going states when $V \neq 0$, a net current flows in the semiconductor band. To find the net quantum mechanical current, we sum the current due to individual $k-$states $J_c^d = q\frac{g_s g_v}{L^d}\sum_k f(k)\mathbf{v_g}(\mathbf{k}) \cdot \frac{\mathbf{F}}{F}$ where the component of the group velocity along the electric field at the source injection point is used. The sum may be written in the form

$$J_c^d = \frac{q}{L^d} \cdot g_s g_v \cdot [\sum_k f_R(k)\mathbf{v_g} \cdot \hat{z} - \sum_k f_L(k)\mathbf{v_g} \cdot \hat{z}], \tag{6}$$

where the dot product gives the projection of the group velocity along the direction of electric field, which is the direction of net current flow. $f_R(k)$ is the Fermi-Dirac distribution of the right-going electron states, characterized by the source Fermi level E_{Fs}, and $f_L(k)$ is the Fermi-Dirac distribution of the left-going electron states, characterized by the drain Fermi level $E_{Fd} = E_{Fs} - qV$. Using integrals of the form

$$\int_0^\infty dk \frac{k^d}{1 + e^{\frac{E_c + \frac{\hbar^2 k^2}{2m_c^\star} - E_F}{k_b T}}} = I(d,\eta) = \frac{1}{2} \cdot (\frac{2m_c^\star k_b T}{\hbar^2})^{\frac{d+1}{2}} \cdot \Gamma(\frac{d+1}{2}) \cdot F_{\frac{d-1}{2}}(\eta), \tag{7}$$

where $\eta = \frac{E_F - E_c}{k_b T}$, we get

$$J_c^d = \frac{q^2}{h} \cdot N_c^{d-1} \cdot (\frac{k_b T}{q}) \cdot [F_{\frac{d-1}{2}}(\frac{E_{Fs} - E_c}{k_b T}) - F_{\frac{d-1}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]. \tag{8}$$

See Table 1 for a summary for the ballistic current in various dimensions. The total electron density of Equation 5 is the *sum* of the left- and right-going electron densities, but the total current in Equation 8 the *difference* of the right and left going currents, as it should be. The current density in d-dimensions is proportional to the band-edge DOS of $d - 1$ dimensions, and the quantum of conductance $\frac{q^2}{h}$. When the voltage V is large, the left-going current becomes negligible, J_c^d becomes independent of the applied voltage V, and the net current *saturates*. The saturation of the current in a semiconductor - even in the quantum limit - is one of its very important characteristics that drives its useful applications in electronics.

If the electron density in the band is controlled not by the source and the drain, but capacitively by a third terminal called the gate, then n_d of Equation 5 can be dynamically changed; E_{Fs} as a result will be controlled by the gate voltage. This allows the current of Equation 8 flowing between the source and the drain to be controlled remotely by a third terminal, which does not draw any current itself. In other words, the resistance between the source and drain is 'transferred' to a third terminal, giving this device the name the Transferred-Resistor, or the Transistor. Because the control is exerted due to an electric field across a gate capacitor, this form of a transistor is called the Field-Effect Transistor, or the FET. The expressions above are for the ballistic FET. Scattering in the semiconductor will lower the net current that flows by mixing left and right going states and lowering the effective $E_{Fs} - E_{Fd}$ across the channel from its ballistic value of qV. But it will not change the fundamental qualitative picture described above in the quantum limit.

5. Ballistic Transport in Non-Uniform Potentials and Tunneling

Consider the situation when an electron in the conduction (or valence) band experiences a potential that varies in space according to $E_c(x)$. The non-uniformity may arise due to band bending due to electric fields, at junctions, or in quantized heterostructures such as quantum wells or dots. Because $|\psi(x)|^2 = |u_k(x)|^2$, the Bloch states extend over the whole crystal, and their probability densities are periodic. Thus, they are not well suited to model electrons in non-uniform potentials. By considering the linear combination of a small set of k states near the edge of the band, the effective mass approximation converts the Bloch picture of individual electron states $\psi_k(x) = e^{ikx}u_k(x)$ to a wavepacket $\phi(x) = \sum_k C_k \psi_k(x) \approx C(x)u_k(x)$, where $C(x)$ is a slowly varying function compared to the lattice constant. The resulting effective mass equation for a bandstructure $E_c(x,k) = E_c(x) + \frac{\hbar^2 k^2}{2m_c^\star}$, and any perturbation $W(x)$ in the band edge is then

4

$$[E_c(-i\frac{d}{dx}) + W(x)]C(x) \approx EC(x) \implies [-\frac{\hbar^2}{2m_c^\star}\frac{d2}{dx^2} + W(x)]C(x) \approx [E - E_c(x)]C(x), \tag{9}$$

from where the envelope function $C(x)$ can be determined. $C(x)$ should not be mistaken for the total wavefunction, which is $\phi(x) \approx C(x)u_k(x)$, but for much phenomena that occur for small energies within the same band, $u_k(x)$ will typically cancel. For a constant $W(x) = 0$ and a spatially constant $E_c(x) = E_c$, the envelope function $C(x) = \frac{1}{\sqrt{L}}e^{ikx}$ is a plane wave, where $k = \sqrt{\frac{2m_c^\star(E-E_c)}{\hbar^2}}$. But when the band edge $E_c(x)$ varies slowly for $W(x) = 0$, the solution of $C(x)$ for the equation on the right of 9 is given by the Wentzel-Kramers-Brillouin or WKB approximation:

$$\frac{d^2C(x)}{dx^2} \approx -\underbrace{\frac{2m_c^\star}{\hbar^2}[E - E_c(x)]}_{Q(x)}C(x) \implies C(x) \approx \frac{K}{Q(x)^{\frac{1}{4}}}e^{\pm\int_a^x du\sqrt{Q(u)}}, \tag{10}$$

where K is a normalization constant. If $E > E_c(x)$, $Q(x) = -k(x)^2 < 0$, and the exponential $e^{\pm\int du\sqrt{Q(u)}}$ in the WKB wavefunction is of the form $e^{\pm i\int dx\cdot k(x)}$, which is still oscillatory, like the plane wave situation. The electron density corresponding to the envelope function (*not* the total wavefunction[3]) is then given by the probability density $n(x) \approx |C(x)|^2 = |K|^2/|k(x)|$, from where it may be seen that for regions where $k(x)$ is large, $n(x)$ is small. The analogy to classical mechanics is the following: a particle spends less time in a region where its kinetic energy is large. The approximate current density $J \approx qnv \approx q(|K|^2/k(x)) \cdot (\hbar k(x)/m_c^\star) \approx q\hbar|K|^2/m_c^\star$ ensures the current carried by the wavepacket is continuous in space.

When an electron of energy E is incident on a barrier, $E < E_c(x)$, and $Q(x) = \kappa(x)^2 > 0$. In this case, the WKB envelope function $C(x) \approx \frac{K}{\sqrt{\kappa(x)}}e^{\pm\int dx\cdot\kappa(x)}$, which now is a exponentially decaying, or growing wavefunction amplitude, quite unlike the oscillatory function seen for $E > E_c(x)$. For propagation of the wavepacket from x_1 to x_2 the WKB transmission probability is

$$T_{WKB} \approx |\frac{C(x_2)}{C(x_1)}|^2 \approx e^{-2\int_{x_1}^{x_2} dx\cdot\kappa(x)} = e^{-2\int_{x_1}^{x_2} dx\cdot\sqrt{\frac{2m_c^\star}{\hbar^2}[E_c(x)-E]}}. \tag{11}$$

As an example, consider a spatially uniform potential barrier $V_0 - E$, and the the total barrier thickness to be $x_2 - x_1 = t_b$. Using them in Equation 11, we get $T_{WKB} \approx \exp[-\frac{t_b}{0.1\,\mathrm{nm}}\sqrt{\frac{m_c^\star}{m_e}\cdot\frac{V_0-E}{1\,\mathrm{eV}}}]$. For $V_0 - E = 1$ eV, $m_c^\star = m_e$ and $t_b = 3$ nm, the tunneling probability is $1/e^{30} \approx 10^{-13}$. If the barrier thickness decreases by 3 times to $t_b = 1$ nm, the tunneling probability increases substantially to $1/e^{10} \approx 5 \times 10^{-5}$. This trick is used in several semiconductor device applications such as in ohmic contacts, and in resonant tunneling diodes to boost the tunneling currents.

6. Scattering: Electron-Phonon, Electron-Defect, and Electron-Photon

The ballistic picture of transport assumes that the distribution functions $f(k)$ of the electrons in the bands $E_n(k)$ are at the mercy of the contacts to the electron states in the bands, and $f(k)$ shares the same Fermi-level as the contact(s) with which it can exchange electrons. The presence of the crystal, or the surroundings do not mix the electrons between different $k-$states against the wishes of the contacts. This is a good approximation in the smallest devices, but it is always possible for other physical processes to wrest part of the control of the occupation function of the electron states from the contacts. These scattering processes can be useful, or may create losses, and knowing their nature is a central tool to coax the desired function from the device.

Photons: Consider a semiconductor crystal on which photons or light is incident. The electric field of the light wave takes the form $E(x,t) = E_0[e^{i(kx-\omega t)} + e^{-i(kx-\omega t)}]/2$, where E_0 is the amplitude of the electric field, and the circular frequency ω and the photon wavevector $k = 2\pi/\lambda$ are related by $\omega = ck$. The electric

---

[3]The periodic part of the Bloch function always remains as a constant background oscillation, which does not change the arguments. $u_k(x)$ is similar to the individual compartments of a long train, and $C(x)$ is the train when looked from far - say from an airplane flying at a great height above the train, when the individual compartments are not seen. By looking at $C(x)$, we are looking at the transport of the entire train, or the entire electron wavepacket.

potential seen by electrons in the crystal will be of the form $W(x,t) = \int dx \cdot E(x,t)$. If the light wavelength is of the order of μm, which is much larger than the lattice constant and electron wavelengths in the crystal, the light wave may be approximated as $E(x,t) \approx E_0 e^{\pm i\omega t}$ (this is called the dipole approximation), and the perturbation seen by the electrons in the semiconductor is $W(x,t) = q \int dx \cdot E(x,t) \approx qE_0 x e^{\pm i\omega t} = W_0(x)e^{\pm i\omega t}$, where the perturbation is split into a spatial part $W_0(x)$ and a time-dependent part $e^{\pm i\omega t}$. This potential is experienced by the electrons, in addition to the perfect periodic potential $V_{per}(x)$ of the crystal, so the total potential seen by the electrons is $V_{per}(x) + W(x,t)$.

Phonons: Even when light is not incident on a semiconductor, the atoms in the crystal do not sit still, they vibrate around their equilibrium periodic positions. The amplitude of vibration increases as the heat energy, characterized by the lattice temperature T increases. Beyond a certain temperature, the crystal melts. This implies that the crystal is actually not even periodic! At room temperature, the vibrations from the mean positions are not too large, and can be broken down into quanta of orthogonal lattice wave modes called phonons. The displacement of an atom at location x at time t due to a single phonon is written as $u(x,t) = u_0 e^{i(qx-\omega t)}$. This phonon has an amplitude u_0, a wavevector $q = 2\pi/\lambda$, and frequency ω. In crystalline solids, ω and q have two relations: acoustic, and optical. For acoustic modes, $\omega = v_s q$ where v_s is the sound velocity, and for optical modes, $\omega \approx \omega_0$. The perturbation to the periodic crystal potential $V_{per}(x)$ due to the phonon is treated as a small time-dependent change due to the dilation or compression of the lattice, and a deformation potential D_c is used to measure the change in the electronic potential experienced by the electron: $W(x,t) = D_c \frac{du(x,t)}{dx}$ for a longitudinal acoustic wave. For such a wave, $W(x,t) = iqu_0 e^{i(qx-\omega t)} = W_0(x)e^{\pm i\omega t}$. Note that for both phonons and photons, the perturbation to the periodic potential is of the form $W_0(x)e^{\pm i\omega t}$, and is explicitly time-dependent.

Defects: Let us say that the crystal is sitting in the dark, and at a very low temperature. Even then, if there are defects in the crystal, they will introduce perturbations to the periodic potential. Let us say each defect has a potential $W_0(x)$, and there are identical defects at various locations $x_1, x_2, x_3, ...$ Then, the total potential seen by the electron is $V_{per}(x) + W_{tot}(x)$, where $W_{tot}(x) = W_0(x - x_1) + W_0(x - x_2) + W_0(x - x_3) + ...$ Note that the potential of these defects is time-independent, unlike those of phonons and photons. Such defect potentials may be considered to be time-dependent, but with $\omega = 0$.

Fermi's golden rule for Transitions: The electrons in the semiconductor crystal are now subject to these perturbation potentials of the form $W(x,t)$. The time-dependent Schrodinger equation, and the resulting scattering rate of states due to the perturbation is given by

$$i\hbar \frac{\partial \Psi}{\partial t} = [-\frac{\hbar^2}{2m_e} \frac{d^2}{dx^2} + V_{per}(x) + W(x,t)]\Psi \implies \frac{1}{\tau(k \rightarrow k')} \approx \frac{2\pi}{\hbar} \cdot |\langle k'|W_0(x)|k\rangle|^2 \cdot \delta[E(k) - E(k') \pm \hbar\omega].$$

(12)

Here, on the right is the scattering rate from the unperturbed Bloch state $\psi_k(x) = e^{ikx}u_k(x)$ to the unperturbed Bloch state $\psi_{k'}(x) = e^{ik'x}u_{k'}(x)$ by the perturbation $W(x,t) = W_0(x)e^{\pm i\omega t}$, given by the Fermi's golden rule. The scattering rate is in units of 1/sec. The matrix element $\langle k'|W_0(x)|k\rangle = \int dx \psi_{k'}^\star(x)W_0(x)\psi_k(x)$ consists only the spatial part, and the time-dependent part of the perturbation appears as the frequency ω inside the Dirac-delta function. In the matrix element, the entire Bloch functions $\psi_k(x) = e^{ikx}u_k(x)$ and $\psi_{k'}(x) = e^{ik'x}u_{k'}(x)$ appear. The spatial part $\langle k'|W_0(x)|k\rangle$ dictates whether a scattering transition from state $k \rightarrow k'$ is allowed or not, and if it is allowed, it determines its strength. The Dirac-delta function ensures energy conservation in the scattering process. For time-independent scattering potentials such as from defects, $\omega = 0$, and the energy of the electron before scattering $E(k)$ and after scattering $E(k')$ must be identical. This form of scattering is called elastic scattering. On the other hand, for time-dependent scattering potentials, the electron can gain or lose energy. In fact, Fermi's golden rule says more: because $E(k') = E(k) \mp \hbar\omega$, only those transitions are allowed in which the electron can absorb energy from, or emit energy into, phonon and photon modes in discrete *quanta* of energy $\hbar\omega$.

A typical order of magnitude of scattering time for a conduction to valence band (or interband) transition due to electron-photon interaction in a semiconductor is $\tau_{optical}^{interband} \sim$ ns. For transitions within the same band (or intraband transitions) due to phonons or photons, a typical time is $\tau_{phonon}^{intraband} \sim \tau_{phonon}^{intraband} \sim$ ps. Because scattering due to defects is typically elastic, it is intraband; the scattering times due to defects can vary widely depending on the defect densities, from $\tau_{defect} \sim$ ms in ultraclean defect-free semiconductors, to $\sim$ ps.

7. The Boltzmann Transport Equation

If we can determine the final distribution function $f(k)$ of the electron states considering transitions and scattering between all $k \leftrightarrow k'$, we can find how a semiconductor will respond to voltages, or heat or light. For example, the net charge current flowing in the semiconductor then will be $J = q \frac{g_s g_v}{L^d} \sum_k v_g(k) f(k)$. To include the variations of the distribution function in space x and time t, we write the occupation function as $f(x, k, t)$. Though the spatial location of a *Bloch state* of wavefunction $\psi_k(x) = e^{ikx} u_k(x)$ is ill defined because the state extends over the entire crystal, the spatial extent of a *wavepacket* $\phi(x) \approx C(x) u_k(k)$ is well defined by $\Delta x \sim 1/\Delta k$, where the wavepacket is constructed by a linear combination of Bloch states with wavevectors $k \pm \Delta k$. So when we write the distribution function as $f(x, k, t)$, we explicitly mean the occupation function of the wavepacket states.

Consider the distribution function $f(x, k, t)$ which is the occupation probability of an electron state k at spatial location x at time t. Let $S(k \rightarrow k')$ denote the scattering rate in 1/sec out of state k into other states k'. $S(k' \rightarrow k)$ then denotes the reverse, or rate of scattering in to state k from other states k'. These rates are obtained from the Fermi's golden rule result in Equation 12. It may be weighted if the physical process so demands; for example, for momentum scattering rate, $S_m(k \rightarrow k') = \frac{1}{\tau(k \rightarrow k')}[1 - \cos(\theta)]$, where $\frac{1}{\tau(k \rightarrow k')} = \frac{2\pi}{\hbar} \cdot |\langle k'|W_0(x)|k\rangle|^2 \cdot \delta[E(k) - E(k') \pm \hbar\omega]$, and θ is the angle between k and k'.

There is an interesting relation between the direction of a scattering event $S(k \rightarrow k')$ and its reverse $S(k' \rightarrow k)$. Single-particle quantum mechanics in the form of Fermi's golden rule does not distinguish between a rate and its reverse. But the quantum statistics of the states, a genuine many-particle effect, clearly dictates which processes are preferred. To see this, consider a situation when there are no electric fields, or no concentration gradients in the semiconductor, and $f(x, k, t)$ is the occupation function of an electron wavepacket. Because electrons are Fermions, if a state k is already filled, the Pauli exclusion principle kicks in, and it is not possible to scatter in to it. For an electron to scatter from state k' to k, it must be occupied. These conditions dictate that the rate of scattering in to a state k is proportional to the factor $f_{k'}(1 - f_k)$, leading to the following rate equation for Fermions:

$$\frac{df(x, k, t)}{dt} = \sum_{k'} [S(k' \rightarrow k) f_{k'}(1 - f_k) - S(k \rightarrow k') f_k(1 - f_{k'})]. \tag{13}$$

At equilibrium, $f(x, k, t)$ does not change with time, and the left hand of this equation is zero. But equilibrium is much more. If the total number of particles does not change with time, the rate of each microscopic scattering process going one way $k \rightarrow k'$ and its reverse, $k' \rightarrow k$ are also exactly equal. The occupation function under this equilibrium condition is the Fermi-Dirac distribution function for electrons, $f_k^0 = \frac{1}{1 + e^{\frac{E(k) - E_F}{k_b T}}}$. Here, E_F is the Fermi level, and T is the temperature of the heat bath with which the semiconductor electrons are in equilibrium[4]. Regardless of the temperature of the electrons in equilibrium, we must then have

$$S(k' \rightarrow k) f_{k'}^0 (1 - f_k^0) - S(k \rightarrow k') f_k^0 (1 - f_{k'}^0) = 0 \implies \frac{1 - f_{k'}^0}{f_{k'}^0} \cdot \frac{f_k^0}{1 - f_k^0} = \frac{S(k' \rightarrow k)}{S(k \rightarrow k')} = e^{\frac{E(k') - E(k)}{k_b T}}. \tag{14}$$

For elastic scattering, $E(k) = E(k')$, and $S(k \rightarrow k') = S(k' \rightarrow k)$. For emission or absorption of photons or phonons, $S_{abs} = S[E(k) \rightarrow E(k) + \hbar\omega] \propto N(\omega) = \frac{1}{e^{\frac{\hbar\omega}{k_b T}} - 1}$. Then, we obtain

$$\frac{S_{em}}{S_{abs}} = e^{\frac{\hbar\omega}{k_b T}} \implies S_{em} = e^{\frac{\hbar\omega}{k_b T}} S_{abs} \propto e^{\frac{\hbar\omega}{k_b T}} N(\omega) = 1 + \frac{1}{e^{\frac{\hbar\omega}{k_b T}} - 1} = 1 + N(\omega) \implies S_{em} \propto 1 + N(\omega), \tag{15}$$

---

[4]What exactly is this heat bath? The atoms of a semiconductor crystal sitting on a chip at room temperature are in thermal equilibrium with the chip, and the electrons in the semiconductor are in thermal equilibrium with the atoms of the crystal. Thus, the temperature of the chip makes it all the way to the temperature characterizing the Fermi-Dirac distribution of the electrons. If on the other hand, the electrons were interacting strongly with light, or photons from the sun in a solar cell, the temperature of the electrons will be somewhere between the temperature of the sun, and the temperature of the solar panel on which the semiconductor sits.

Absorption of course is naturally *stimulated*, meaning if there are no photons or phonons to absorb, $S_{ab} = 0$. But we see that if an electron is emitting into bosonic modes of photons or phonons, $S_{em} \propto (1 + N(\omega))$, which means that even if $N(\omega) = 0$ and there are no photons or phonons present, the electron can still emit: this is called *spontaneous* emission. If there are photons or phonons present, the rate of emission is *enhanced*, this is *stimulated* emission. The possibility of stimulated emission was first conceived by Einstein, and has led to the realization of lasers. By considering a rate equation of a similar kind as Equation 13 but for Bosons, with f_k replaced by $N(\omega)$ for the equilibrium Bose-Einstein distribution, it may again be shown that $\frac{S(\hbar\omega' \to \hbar\omega)}{S(\hbar\omega \to \hbar\omega')} = e^{\frac{\hbar\omega' - \hbar\omega}{k_b T}}$, which is the same as the relation for Fermions.

Now if we turn on electric fields that exert a force F on the electrons, it will change k according to $F = \hbar \frac{dk}{dt}$, or create gradients in k space. We have also have electron concentration gradients in real space. In the presence of these gradients, the distribution function deviates from the equilibrium Fermi-Dirac function, $f_k \neq f_k^0$. If these processes of gradients are considered, Equation 13 must change to

$$f(x, k, t) = f(x - v \cdot dt, k - \frac{F}{\hbar} \cdot dt, t - dt) + dt \cdot \sum_{k'} [S(k' \to k) f_{k'} (1 - f_k) - S(k \to k') f_k (1 - f_{k'})], \quad (16)$$

which upon Taylor expansion of the first term on the right becomes

$$\frac{\partial f_k}{\partial t} + v \frac{\partial f_k}{\partial x} + \frac{F}{\hbar} \frac{\partial f_k}{\partial k} = \sum_{k'} [S(k' \to k) f_{k'} (1 - f_k) - S(k \to k') f_k (1 - f_{k'})]. \quad (17)$$

This is the Boltzmann transport equation. The mixing of states $k \leftrightarrow k'$ which was not allowed in our ballistic treatment earlier is now allowed, as seen explicitly in the right hand side of the equation. By solving this equation, we can obtain f_k, the distribution function for electrons, considering all imaginable processes. For example, in ballistic conditions, the RHS which captures scattering may be set to zero. To consider scattering, the right hand side must be included, and the equation solved for f_k. The Boltzmann equation also provides the time-dependent changes of f_k, meaning it captures the dynamical behavior of the electrons.

As an example, consider there are no concentration gradients $v \frac{\partial f_k}{\partial x} = 0$, and no electric fields, $\frac{F}{\hbar} \frac{\partial f_k}{\partial k} = 0$. Say the scattering sum on the RHS can be written as a small change from the equilibrium Fermi-Dirac distribution, $RHS \approx -\frac{f_k - f_k^0}{\tau}$ characterized by a scattering time τ. Then, the Boltzmann transport equation becomes

$$\frac{\partial f_k}{\partial t} \approx -\frac{f_k - f_k^0}{\tau} \implies f_k(t) \approx f_k^0 + (f_{in} - f_k^0) e^{-\frac{t}{\tau}}. \quad (18)$$

This example shows that if at time $t = 0$, the distribution function was a non-equilibrium value f_{in}, upon removal of all fields and gradients, the distribution function relaxes to the equilibrium Fermi-Dirac value exponentially with a characteristic time τ. The approximation that all scattering is captured by $RHS \approx -\frac{f_k - f_k^0}{\tau}$ is thus called the relaxation time approximation, meaning it can be characterized by one time constant τ. If on the other hand, an electric field is turned on, at steady state, $\frac{\partial f_k}{\partial t} = 0$, and the Boltzmann transport equation becomes

$$\frac{F}{\hbar} \frac{\partial f_k}{\partial k} \approx -\frac{f_k - f_k^0}{\tau} \implies f_k \approx f_k^0 - \frac{F}{\hbar} \tau \frac{\partial f_k}{\partial k} = f_k^0 - qE v_k \tau \frac{\partial f_k}{\partial E(k)} \implies f_k \approx f_k^0 + qE v_k \tau (-\frac{\partial f_k^0}{\partial E(k)}), \quad (19)$$

where we have used the group velocity of the wavepacket $v_k = \frac{1}{\hbar} \frac{\partial E(k)}{\partial k}$, and that the force on the electron is $F = qE$, were E is the electric field. We have further replaced $f_k \to f_k^0$ inside the derivative: this approximation holds for small electric fields. This solution of the Boltzmann transport equation is in a suggestive form because the equilibrium portion f_k^0 is split from the non-equilibrium portion $qE v_k \tau (-\frac{\partial f_k^0}{\partial E(k)})$. The non-equilibrium portion is driven by the electric field, and is the starting point for the evaluation of electrical conductivity, mobilities, and electric currents in the presence of scattering.

If the electric field was zero, but there was a concentration gradient, the steady state distribution function in the relaxation time approximation is

$$v_k \frac{\partial f_k}{\partial x} \approx -\frac{f_k - f_k^0}{\tau} \implies f_k \approx f_k^0 - v_k \tau \frac{\partial f_k^0}{\partial x}, \tag{20}$$

where we have again replaced the unknown $f_k \to f_k^0$ on the right for small gradients compared to equilibrium.

8. Flow of currents in the presence of scattering

The Boltzmann transport equation can be used to capture ballistic transport as a special case when there is no scattering. But it can also be solved to obtain the distribution function $f(k)$ in the presence of perturbations and scattering. For each external perturbation, it can provide a net current density. In response to an electric potential gradient $E = -\frac{dV(x)}{dx}$ due to an electric field, the charge drift current density is obtained from Equation 19 as

$$J_{drift} = q \frac{g_s g_v}{L^d} \sum_k v_k f_k \approx q \frac{g_s g_v}{L^d} \sum_k v_k [f_k^0 + q E v_k \tau_m(k)(-\frac{\partial f_k^0}{\partial E(k)})] = \frac{q^2 g_s g_v}{L^d} [\sum_k v_k^2 \tau_m(k)(-\frac{\partial f_k^0}{\partial E(k)})] E. \tag{21}$$

Here we have used $\sum_k v_k f_k^0 = 0$. For charge current, the momentum relaxation time $\tau_m(k)$ is obtained by first finding $1/\tau(k \to k')$ from the Fermi's golden rule considering individual scattering potentials $W(x, t)$. The momentum scattering rate is obtained by summing the rate $1/\tau(k \to k')$ over all final states k' with an angular weight for momentum scattering: $\frac{1}{\tau_m(k)} = \sum_{k'} \frac{1}{\tau(k \to k')}[1 - \cos \theta]$, where θ is the angle between k and k', or the angle by which the momentum is scattered. If the angle by which the electron is scattered is small, the net momentum in the direction of current is not affected much, and the conductivity should be high. But if the angle is large, say $\theta \approx \pi$, the conductivity is lowered significantly. This is captured by the $1 - \cos \theta$ weight heuristically. This factor can be derived exactly from the Boltzmann transport equation.

Relating this drift charge current to a Drude form $J_{drift} = \sigma E = q n \mu E$, where $n = \frac{g_s g_v}{L^d} \sum_k f_k$, we identify the electron drift mobility as

$$\mu = \frac{J_{drift}}{qnE} = \frac{\frac{q^2 g_s g_v}{L^d} \cdot [\sum_k v_k^2 \tau_m(k)(-\frac{\partial f_k^0}{\partial E(k)})] \cdot E}{q \cdot [\frac{g_s g_v}{L^d} \sum_k f_k] \cdot E} = q \cdot \frac{\sum_k v_k^2 \tau_m(k)(-\frac{\partial f_k^0}{\partial E(k)})}{\sum_k f_k^0}. \tag{22}$$

The numerator has the derivative of the Fermi-Dirac distribution with energy. This means if we have a degenerate Fermi gas when say $E_F > E_c$, for most of the electrons in the window $E_c \leq E(k) \leq E_F$, the term $-\frac{\partial f_k^0}{\partial E(k)} \approx 0$, and they cannot carry *net* currents. Their contribution to the ensemble mobility is low: this is Pauli blocking in action. Because $-\frac{\partial f_k^0}{\partial E(k)}$ peaks at the Fermi level, the contribution to the mobility will be dominated by states near the Fermi level $E \approx E_F$, or in other words, the conductivity is dominated by states near the Fermi level. If we consider the low temperature approximation $-\frac{\partial f_k^0}{\partial E(k)} \approx \delta(E - E_F)$ in the degenerate condition for parabolic bands, we can obtain $\mu \approx \frac{q \tau_m(k_F)}{m_c^*}$, where k_F is the Fermi wavevector. This form of the mobility is in the Drude form. For the more general situation, the summation over the k states must be performed.

In the presence of spatial gradients of the distribution function (or carrier concentration gradients), the current is obtained from Equation 20 with a momentum scattering time $\tau_m(k)$

$$J_{diff} = q \frac{g_s g_v}{L^d} \sum_k v_k [f_k^0 - v_k \tau_m(k) \frac{\partial f_k^0}{\partial x}] \approx -q \frac{g_s g_v}{L^d} \sum_k v_k^2 \tau_m(k) \frac{\partial f_k^0}{\partial x}. \tag{23}$$

With further approximations, the diffusion current may be shown to be of the form $J_{diff} = qD \frac{dn(x)}{dx}$, where $n(x)$ is the local electron concentration $\frac{1}{L^d} \sum_k f_k^0$, and D is a diffusion constant. Similarly, spin currents will

flow in response to spin density gradients $J_{spin} = D_s \frac{ds(x)}{dx}$ with a corresponding spin diffusion constant, and heat current in response to temperature gradients $J_{heat} = -\kappa_e \frac{dT(x)}{dx}$. The thermal diffusion constant κ_e is the electronic contribution to the thermal conductivity.

9. Explicit evaluation of scattering rates and mobility

We now succinctly outline the explicit calculation of the drift mobility due to scattering between effective mass wavepacket states centered at states $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$ within the *same band*, with corresponding time-dependent wavefunctions $\phi_{\mathbf{k}}(\mathbf{r}, t) \approx C_{\mathbf{k}}(\mathbf{r}, t)u_{\mathbf{k}}(\mathbf{r})$ and $\phi_{\mathbf{k}'}(\mathbf{r}, t) \approx C_{\mathbf{k}'}(\mathbf{r}, t)u_{\mathbf{k}'}(\mathbf{r})$. Consider a scattering potential of the form $W(\mathbf{r}, t) = W_0(\mathbf{r})e^{\pm i\omega t}$. The time-dependent effective mass equation for the wavepacket centered at $\mathbf{k}$ is

$$i\hbar \frac{\partial C_{\mathbf{k}}(\mathbf{r}, t)}{\partial t} = [E_c(-i\nabla) + W(\mathbf{r}, t)]C_{\mathbf{k}}(\mathbf{r}, t) = [-\frac{\hbar^2}{2m_c^\star}\nabla^2 + E_c(\mathbf{r}) + W(\mathbf{r}, t)]C_{\mathbf{k}}(\mathbf{r}, t). \tag{24}$$

The lattice periodic part $u_{\mathbf{k}}(\mathbf{r})$ cancels in this equation, just as in the time-independent version. In the absence of the perturbation, the envelope functions of the states are $C_{\mathbf{k}}(\mathbf{r})$ and $C_{\mathbf{k}'}(\mathbf{r})$. We can write the momentum scattering rate using Fermi's golden rule for the envelope function as

$$\frac{1}{\tau(\mathbf{k} \to \mathbf{k}')} = \frac{2\pi}{\hbar}|\langle C_{\mathbf{k}'}(\mathbf{r})|W_0(r)|C_{\mathbf{k}}(\mathbf{r})\rangle|^2\delta[E(\mathbf{k}') - (E(\mathbf{k}) \pm \hbar\omega)] \implies \frac{1}{\tau_m(\mathbf{k})} = \sum_{\mathbf{k}'}\frac{1}{\tau(\mathbf{k} \to \mathbf{k}')}(1 - \cos\theta), \tag{25}$$

where $\cos\theta = \frac{\mathbf{k}' \cdot \mathbf{k}}{|\mathbf{k}||\mathbf{k}'|}$. It is possible to directly use the envelope function in the golden rule for transitions *within the same band*, for small changes in $\mathbf{k}$, which ensures that the periodic part of the Bloch functions $u_{\mathbf{k}}(\mathbf{r})$ are the same. For such transitions, the time-dependent effective mass equation Equation 24 has an identical mathematical form as the exact time-dependent equation for the Bloch functions in Equation 12. Whenever this condition fails, one must resort back to using the functions $C_{\mathbf{k}}(\mathbf{r})u_k(\mathbf{r})$ and $C_{\mathbf{k}'}(\mathbf{r})u_{k'}(\mathbf{r})$, where the lattice-periodic parts are explicitly included. Such situations arise when there are transitions between bands, say from the conduction to the valence band, or from the Γ point with $|s\rangle$ orbital lattice-periodic states, to a $|p\rangle$ state in the same band, but at the other extrema, say near the Brillouin zone edge.

For transitions within the same band and for small changes in $\mathbf{k}$, consider the case of scattering by a *single* positively charged ionized impurity (say an ionized donor atom) in a semiconductor crystal. Fixing the origin of coordinates at the location of the impurity, the screened scattering potential is $W(r, t) = W_0(r)e^{\pm i\omega t} = W_0(r) = -\frac{q^2}{4\pi\epsilon_s r}e^{-\frac{r}{L_D}}$, where ϵ_s is the dielectric constant of the semiconductor, L_D is the Debye screening length, and $\omega = 0$, indicating the scattering potential is time-independent. The envelope functions are $C_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}}e^{i\mathbf{k}\cdot\mathbf{r}}$ and $C_{\mathbf{k}'}(\mathbf{r}) = \frac{1}{\sqrt{V}}e^{i\mathbf{k}'\cdot\mathbf{r}}$ where V is the macroscopic volume of the semiconductor crystal. The scattering matrix element is

$$\langle C_{\mathbf{k}'}(\mathbf{r})|W_0(r)|C_{\mathbf{k}}(\mathbf{r})\rangle = \int d^3r(\frac{1}{\sqrt{V}}e^{-i\mathbf{k}'\cdot\mathbf{r}}) \cdot (-\frac{q^2}{4\pi\epsilon_s r}e^{-\frac{r}{L_D}}) \cdot (\frac{1}{\sqrt{V}}e^{i\mathbf{k}\cdot\mathbf{r}}) = \frac{q^2}{\epsilon_s V}\frac{1}{\frac{1}{L_D^2} + |\mathbf{k} - \mathbf{k}'|^2}, \tag{26}$$

where the integral is evaluated in 3D spherical coordinates over the entire space by aligning the vector $\mathbf{k} - \mathbf{k}'$ at an angle θ to the $z-$axis. Note that the scattering matrix element between plane-wave like envelope states is equal to the Fourier transform of the scattering potential. Now the total momentum scattering rate of state $\mathbf{k}$ is

$$\frac{1}{\tau_m(\mathbf{k})} = \sum_{\mathbf{k}'}\frac{2\pi}{\hbar} \cdot |\frac{q^2}{\epsilon_s V}\frac{1}{\frac{1}{L_D^2} + |\mathbf{k} - \mathbf{k}'|^2}|^2 \cdot \delta[E(\mathbf{k}') - (E(\mathbf{k}) \pm \hbar\omega)] \cdot (1 - \cos\theta), \tag{27}$$

which upon summing over all 3D $\mathbf{k}'$ states, using the Dirac-delta function, and the considering the effect of a total of N_{imp} impurities results in

$$\frac{1}{\tau_m(\mathbf{k})} = \frac{g_s g_v q^4 m_c^\star}{8\pi\epsilon_s^2 V} \cdot \underbrace{\frac{N_{imp}}{V}}_{n_{imp}} \cdot \frac{1}{(\hbar k)^3} \cdot [\ln(1 + 4k^2 L_D^2) - \frac{4k^2 L_D^2}{1 + 4k^2 L_D^2}], \tag{28}$$

where $k = |\mathbf{k}|$ is the length of the wavevector, and $n_{imp} = N_{imp}/V$ is the volume density of the uncorrelated scatterers. The dimensions may be checked to be 1/sec. The term in the square brackets is a slowly varying function. The momentum scattering rate can now be used in Equation 22 to get the electron mobility explicitly.

For a parabolic bandstructure $E(k) = \frac{\hbar^2 k^2}{2m_c^\star}$ assuming a reference $E_c = 0$, the momentum scattering rate as a function of the electron kinetic energy is

$$\frac{1}{\tau_m(E)} = \frac{g_s g_v q^4 m_c^\star}{8\pi\epsilon_s^2} \cdot n_{imp} \cdot \frac{1}{(2m_c^\star E)^{\frac{3}{2}}} \cdot [\ln(1 + \frac{8m_c^\star E}{\hbar^2} L_D^2) - \frac{\frac{8m_c^\star E}{\hbar^2} L_D^2}{1 + \frac{8m_c^\star E}{\hbar^2} L_D^2}], \tag{29}$$

which indicates that the scattering rate reduces as the kinetic energy of electrons increases. For a non-degenerately doped semiconductor, using Maxwell-Boltzmann approximation of the equilibrium distribution function $f_{\mathbf{k}}^0 \approx e^{-\frac{E(\mathbf{k})}{k_b T}}$, the electron mobility due to ionized impurity scattering is obtained to be

$$\mu_{imp} \approx \frac{2^{\frac{7}{2}} (4\pi\epsilon_s)^2 (k_b T)^{\frac{3}{2}}}{\pi^{\frac{3}{2}} q^3 \sqrt{m_c^\star} n_{imp} F(\beta)} \sim \frac{T^{\frac{3}{2}}}{n_{imp}}, \tag{30}$$

where $F(\beta) = \ln[1 + \beta^2] + \frac{\beta^2}{1+\beta^2}$ is a slowly varying function, and $\beta = 2\sqrt{\frac{2m_c^\star (3k_b T)}{\hbar^2}}$. As the impurity density increases, the mobility decreases due to more frequent scattering. The mobility increases with temperature because an energetic electron is less perturbed by the scattering potential from its original path. The mobility limited by ionized impurity scattering is high for semiconductors with small effective masses, and for large dielectric constants.

10. Semiconductor Electron Energies for Photonics

Let the semiconductor bandgap be $E_c - E_v = E_g$, where E_c is the conduction band minimum, and E_v is the valence band maximum. Let the conduction band effective mass be $m_c^\star$ and the valence band effective mass $m_v^\star$, so that the conduction band states are given by $E_c(k) = E_c + \frac{\hbar^2 k^2}{2m_c^\star}$ and the valence band states are $E_v(k) = E_v - \frac{\hbar^2 k^2}{2m_v^\star}$. Now if there is an optical transition between the conduction band state E_2 and a valence band state E_1, the photon energy must be $h\nu = E_2 - E_1$ to ensure energy conservation. Because the photon in circularly polarized light carries an angular momentum $\pm\hbar$, the conduction band and the valence band should have a net angular momentum difference of $\hbar$, which they do - because the conduction band states derive from $|s\rangle$ orbitals, and the valence band states from $|p\rangle$ orbitals. Finally, to ensure momentum conservation, we must have $\hbar\mathbf{k}_c = \hbar\mathbf{k}_v + \hbar\mathbf{k}_\nu$, where $\hbar\mathbf{k}_\nu$ is the photon momentum. Since the electron states have $|\mathbf{k}_c| \approx |\mathbf{k}_v| \approx \frac{\pi}{a_0}$ where a_0 is a lattice constant, whereas the photon momentum $\hbar|\mathbf{k}_\nu| = \frac{h}{\lambda}$ is *much* smaller, we make the assumption $k_c = k_v = k$, i.e., the conduction and valence band states that talk to absorb or emit photons must have the same k. Pictorially, this is what we call 'vertical' transitions in the $E(k) - k$ diagram.

An energy state in the conduction band is then

$$E_2 = E_c + \frac{\hbar^2 k^2}{2m_c^\star}, \tag{31}$$

and one in the valence band is

$$E_1 = E_v - \frac{\hbar^2 k^2}{2m_v^\star}, \tag{32}$$

which leads to a photon energy

$$hv = E_2 - E_1 = (E_c - E_v) + \frac{\hbar^2}{2}(\frac{1}{m_c^\star} + \frac{1}{m_v^\star})k^2 = E_g + \frac{\hbar^2 k^2}{2m_r^\star}, \tag{33}$$

which is equivalent to

$$\frac{\hbar^2 k^2}{2m_r^\star} = hv - E_g. \tag{34}$$

Because the LHS is positive, the above calculation reinforces our intuitive observation that only photons of energy hv equal or larger than the energy bandgap E_g will interact with the electron states[5].

11. The Optical Joint Density of States $\rho_{jnt}(v)$

How many such electron k states are available per unit volume to emit, or absorb photons in the energy interval $[hv, h(v + dv)]$? This quantity is of central importance in the evaluation of electron-photon interactions, and is called the optical joint density of states (JDOS) of the semiconductor, denoted by $\rho_{jnt}(v)$. The number of states per unit volume is $\rho_{jnt}(v)d(hv)$. The optical JDOS $\rho_{jnt}(v)$ captures the property of light (through hv), and matter (through the semiconductor parameters the bandgap E_g, and band effective masses $m_c^\star$ and $m_v^\star$). The quantity $\rho_{jnt}(v)$ will also reflect if the optically active semiconductor electron states are unconfined and free to move in 3D bulk, or confined in quantum wells (2D), quantum-wires (1D), or quantum dots (0D). The quantum confined structures are routinely used in practical semiconductor LEDs, SOAs, modulators, and lasers.

To count the number of electron states in 3D, assume the electrons are confined in a cubic box of size $L_x = L_y = L_z = L$. Because an integer number of electron half-wavelengths must fit in the box, we get the condition $\mathbf{k} = (k_x, k_y, k_z) = \frac{\pi}{L}(n_x, n_y, n_z)$, where (n_x, n_y, n_z) is a triplet of integers each of whose values can be $0, 1, 2, ...$. This defines a 3D lattice in the $\mathbf{k} = (k_x, k_y, k_z)$ space in the first octant, where each point denotes an allowed electron state, occupying a volume $(\frac{\pi}{L})^3 = \frac{\pi^3}{V}$. Because the volume of the semiconductor cube is $V = L^3$ and we are interested in the JDOS $\rho_{jnt}(v)$ *per unit volume*, we also define the *total* JDOS as $D_{jnt}(v)$. The JDOS per unit volume is then related to the total JDOS by $\rho_{jnt}(v) = D_{jnt}(v)/V$.

We discuss the 3D optical JDOS first. The optical JDOS $\rho_v(k)$ in the $\mathbf{k}-$space counts the same states as in the energy space, meaning $\rho_k(k)dk = D_{jnt}(v)d(hv)$. Using equation 34 in the form $\frac{\hbar^2}{2m_r^\star}(k_x^2 + k_y^2 + k_z^2) = hv - E_g$, we obtain

$$\rho_k^{3D}(k)dk = 2 \cdot \frac{1}{8} \cdot \frac{4\pi k^2 dk}{\frac{\pi}{L_x}\frac{\pi}{L_y}\frac{\pi}{L_z}} = D_{jnt}(v)d(hv) \implies \boxed{\rho_{jnt}^{3D}(v) = \frac{1}{2\pi^2}(\frac{2m_r^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{hv - E_g}}, \tag{35}$$

where $\rho_{jnt}^{3D}(v) = D_{jnt}(v)/V$ was used. The units of $\rho_{jnt}^{3D}(v)$ are in $\frac{1}{\text{energy}\cdot\text{Volume}}$ as may be verified from equation 35. For semiconductors it is expressed in $\frac{1}{\text{eV}\cdot\text{cm}^3}$.

The JDOS for quantum confined electron states is found in the same manner as for the 3D, except instead of a 3D k space, we have a 2D, or lower-dimensional k space. For example, for a quantum well laser, the electron and hole states are free to move in two dimensions (x, y) because of large dimensions $L_x = L_y = L$ and confined in the third z direction because of a heterostructure quantum well confining potential to a length L_z smaller than the electron de-Broglie wavelengths. This situation is realized if we create a structure in which $L_z << L$, where L_z is a few nm, whereas L is the size of the wafer - in a few 100s of μm or in mm.

For such a quantum-confined structure, Equation 34 now takes the form

$$\frac{\hbar^2}{2m_r^\star}(\underbrace{k_x^2 + k_y^2}_{k_\parallel^2} + k_z^2) = hv - E_g \implies \frac{\hbar^2}{2m_r^\star}k_\parallel^2 = hv - [E_g + \underbrace{\frac{\hbar^2}{2m_e^\star}(\frac{\pi}{L_z}n_z)^2}_{E_n} + \underbrace{\frac{\hbar^2}{2m_h^\star}(\frac{\pi}{L_z}n_z)^2}_{E_p}], \tag{36}$$

---

[5]We are neglecting excitonic and other multiparticle effects here and focusing on the strongest interband transitions.

where we identify the blue-shift in the minimum transition energy due to quantum confinement in the conduction and the valence bands $E_n + E_p$. This quantum-engineering of the effective bandgap is a very powerful tool for engineering light-matter interactions in semiconductor heterostructures.
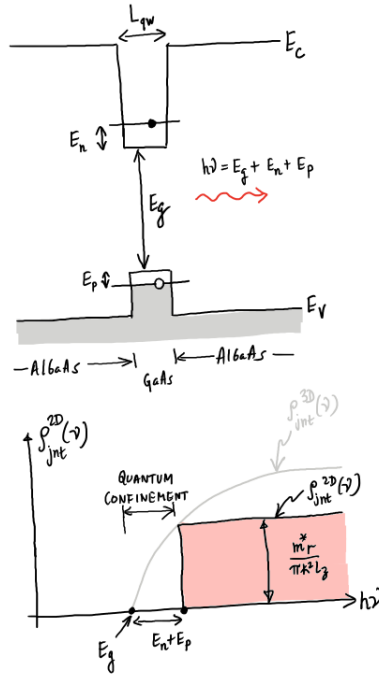

Figure 1: A Quantum Well Laser.

The allowed electron modes then form a dense 2D lattice in the $(k_x, k_y) = \frac{\pi}{L}(n_x, n_y)$ planes, separated by large $k_z = \frac{\pi}{L_z}n_z$ in the k_z direction. Then, the 'cylindrical' 4-fold symmetry around the k_z−axis of the allowed lattice of mode points in **k**−space has each mode occupying a volume $\frac{\pi}{L_x}\frac{\pi}{L_y}\frac{\pi}{L_z}$, and the JDOS is given by

$$\rho_k^{2D}(k)dk = 2 \cdot \frac{1}{4} \cdot \frac{(2\pi k_\parallel dk_\parallel)\frac{\pi}{L_z}}{\frac{\pi}{L_x}\frac{\pi}{L_y}\frac{\pi}{L_z}} = D_{jnt}(\nu)d(h\nu) \implies \boxed{\rho_{jnt}^{2D}(\nu) = \frac{m_r^\star}{\pi\hbar^2 L_z}\Theta[h\nu - (E_g + E_n + E_p)]}, \quad (37)$$

where we again used $\rho_{jnt}^{2D}(\nu) = D_{jnt}(\nu)/V$. The Heaviside unit-step function $\Theta[...]$ in the 2D optical JDOS yet again reinforces our intuition that the smallest photon energy that can interact with the quantum well electron states must have an energy $h\nu = E_g + E_n + E_p$ because quantum confinement has pushed the minimum energy in the conduction band up to $E_c + E_n$ and in the valence band down to $E_v - E_p$. The units of $\rho_{jnt}^{2D}(\nu)$ is still $\frac{1}{\text{eV} \cdot \text{cm}^3}$, the same as for 3D - this is verified from equation 37.

Continuing to lower dimensions, it is left to the reader to show that for electron states in 1D quantum wires confined in the (x, y) plane to (L_x, L_y) and free to move in the z−direction, the optical JDOS is

$$\boxed{\rho_{jnt}^{1D}(\nu) = \frac{2}{\pi L_x L_y}\sqrt{\frac{2m_r^\star}{\hbar^2}}\frac{1}{\sqrt{h\nu - (E_g + E_n + E_p)}}}, \quad (38)$$

where the quantum confinement energies are now $E_n = \frac{\hbar^2}{2m_e^\star}[(\frac{\pi}{L_x}n_x)^2 + (\frac{\pi}{L_y}n_y)^2]$ and $E_p = \frac{\hbar^2}{2m_h^\star}[(\frac{\pi}{L_x}n_x)^2 + (\frac{\pi}{L_y}n_y)^2]$.

For 0D quantum dots in which electrons are confined in all three directions to boxes of size (L_x, L_y, L_z), the structure mimics an atomic system, giving an optical JDOS

$$\rho_{jnt}^{0D}(\nu) = \frac{2}{L_x L_y L_z} \delta[h\nu - (E_g + E_n + E_p)],$$

(39)

where $\delta[...]$ is the Dirac-delta function highlighting that the electron energies are discrete and atomic-like, and do not form extended states or bands. The quantum confinement energies are now $E_n = \frac{\hbar^2}{2m_e^*}[(\frac{\pi}{L_x}n_x)^2 + (\frac{\pi}{L_y}n_y)^2 + (\frac{\pi}{L_z}n_z)^2]$ and $E_p = \frac{\hbar^2}{2m_h^*}[(\frac{\pi}{L_x}n_x)^2 + (\frac{\pi}{L_y}n_y)^2 + (\frac{\pi}{L_z}n_z)^2]$. It may be verified again that the units of $\rho_{jnt}^{0D}(\nu)$ is exactly the same as in other dimensions.

We will shortly see that the optical gain coefficient of a semiconductor in 1/cm units is given by $\gamma_0(\nu) = A \cdot [\frac{\lambda_0^2}{8\pi n^2}] \cdot [h\rho_{jnt}(\nu)] \cdot [f_c(E_2) - f_v(E_1)]$, where A is the Einstein coefficient of units 1/sec, λ_0 is the free-space photon wavelength, and n is the refractive index of the semiconductor. The gain coefficient is characteristically proportional to the optical joint density of states $\rho_{jnt}(\nu)$. Those conversant with atomic laser theory will immediately recognize the similarity to the gain coefficient $\gamma_0(\nu) = A \cdot [\frac{\lambda_0^2}{8\pi n^2}] \cdot [g(\nu)] \cdot [N_2 - N_1]$, where $g(\nu)$ is the lineshape function in units of *sec*, and $N_2 - N_1$ is the population inversion in 1/cm^3 units. The semiconductor counterpart of the lineshape function $h\rho_{jnt}(\nu)$ has units sec/cm^3, and the population inversion term motivates a discussion of the dimensionless occupation probabilities $f_c(E_2)$ and $f_v(E_1)$ of the allowed conduction and valence band electron energy states, to which we now turn.

12. Probability of Occupation of Electron States

Because electrons are Fermions, the Pauli exclusion principle forces them to follow the Fermi-Dirac distribution function in energy. If the semiconductor is in thermal equilibrium at a temperature T, the distribution of electrons in the semiconductor in energy is given by

$$f(E) = \frac{1}{1 + e^{\frac{E - E_F}{k_B T}}},$$

(40)

where the Fermi energy E_F is a single energy characterizing the number of electrons, and k_B is the Boltzmann constant. However, for semiconductor photonic devices, we will be interested in $p - n$ junction diodes under electrical bias conditions. In a $p - n$ junction, we will typically have the following situation: the conduction band electrons in the $n-$side of the junction will be connected to an ohmic metal contact lead, and the valence band electrons in the p-side of the diode to another ohmic metal contact lead. These two leads will then be connected to a battery of voltage V.

Application of a voltage V pulls the diode out of the state of equilibrium. We can now consider two *separate* electron distributions: those of electrons in the conduction band, and of electrons in the valence band. These two electron systems are in equilibrium amongst themselves, but they are out of equilibrium from each other. The electrons in the conduction band in the $n-$side of the diode are in equilibrium with the voltage of the battery terminal connected to them; this is modeled by assigning to them a *quasi-Fermi-level* F_n. Similarly the electron states in the valence band of the $p-$side of the diode are in equilibrium with the other terminal of the battery, sharing a quasi-Fermi-level F_p. The electron populations at the two terminals of the battery are held at an energy qV apart, meaning $F_n - F_p = qV$.

Because of the above definitions, the occupation probability of an electron state of energy E_2 in the conduction band is $f_c(E_2)$, and the occupation probability of an electron state of energy E_1 in the valence band is $f_v(E_1)$, where each is Fermi-Dirac function modified to account for the non-equilibrium physics:

$$f_c(E_2) = \frac{1}{1 + e^{\frac{E_2 - F_n}{k_B T}}} \quad \text{and} \quad f_v(E_1) = \frac{1}{1 + e^{\frac{E_1 - F_p}{k_B T}}}.$$

(41)

14

If we are interested in the 3D electron density in the conduction band, we can use the 3D conduction band density of states to write

$$n = \int_{E_c}^{\infty} dE \cdot \rho_c(E) f_c(E) = \int_{E_c}^{\infty} dE \cdot \frac{1}{2\pi^2} \left(\frac{2m_c}{\hbar^2}\right)^{\frac{3}{2}} \sqrt{E - E_c} \frac{1}{1 + e^{\frac{E - F_n}{kT}}}, \tag{42}$$

with similar expressions for electrons in the valence bands. For finding the probability of E_1 being an empty energy state in the valence band (or hole), we must use the probability that the electron state is *not occupied*, which is precisely $1 - f_v(E_1)$. Let us say that for photon emission, we need an electron occupying energy E_2 in the conduction band to transition into an empty state E_1 in the valence band such that $E_2 - E_1 = h\nu$. The probability for the electron states to satisfy the criteria is the product $f_c(E_2)[1 - f_v(E_1)]$.

13. Interband Absorption, and Spontaneous & Stimulated Emission

Einstein's re-derivation of the Planck blackbody radiation formula introduced the concepts of rates of spontaneous emission, stimulated emission, and absorption through the net rate equation

$$\frac{1}{d(h\nu)} \frac{d\Delta n}{dt} = - \underbrace{A\rho_{jnt}(\nu) f_c(E_2)[1 - f_v(E_1)]}_{=R_{sp}(\nu),\text{Spontaneous Emission}} - \underbrace{B\rho_{jnt}(\nu)\rho_v f_c(E_2)[1 - f_v(E_1)]}_{=R_{st}(\nu),\text{Stimulated Emission}} + \underbrace{B\rho_{jnt}(\nu)\rho_v f_v(E_1)[1 - f_c(E_2)]}_{=R_{ab}(\nu),\text{Absorption}},$$

$$\tag{43}$$

where the LHS tracks the rate of change of excess electron (or hole) density Δn per energy interval $d(h\nu)$. Electrons may be lost from this level due to spontaneous emission of photons of energy $h\nu = E_2 - E_1$ if the electron transitions to state E_1 in the valence band. This will happen at a rate $R_{sp}(\nu)$, as indicated in equation 43. The net spontaneous emission rate $R_{sp}(\nu)$ is made of the JDOS $\rho_{jnt}(\nu)$, which counts how many electron state pairs (E_2, E_1) are available per unit energy interval per unit volume satisfying $h\nu = E_2 - E_1$, and the Fermi factor $f_c(E_2)[1 - f_v(E_1)]$, which is the probability that the occupations of the states are just right for emitting a photon. The coefficient A, an effective spontaneous emission lifetime, at this stage may be considered something that can be obtained from experiments. The exact form of A may be calculated from time-dependent perturbation theory, and more accurately from quantum electrodynamics. It is a rate constant typically of the order $1/\text{ns}$ for most semiconductors.

The second term in the RHS of equation 43 is the rate of *stimulated emission*, $R_{st}(\nu)$. The coefficient is B - which has different units from the A coefficient. Einstein showed that for the above picture of light-matter interaction to be consistent with Planck's blackbody radiation formula, the A and B coefficients are related by $\frac{A}{B} = \frac{8\pi n^3 \nu^3 h}{c^3}$. The JDOS and probability functions are identical for the stimulated emission and spontaneous emission rates. Other than the A and B coefficients, the two processes are distinguished by the term ρ_v, which is the energy density of the *photon field*, is in units of eV/cm^3. It is related to the *intensity* of the photon field I_ν by the relation $I_\nu = \rho_\nu \cdot (c/n)$, where c/n is the photon velocity in a medium of refractive index n. The intensity of the photon field is obtained from the Poyntying vector $\mathbf{I}_\nu = \mathbf{E}_\nu \times \mathbf{H}_\nu$ from the electric and magnetic fields, the magnitude of which is $I_\nu = \frac{E_\nu^2}{2\eta}$ in units of Watts/cm^2, or the rate of flow of energy per unit area. Here E_ν is the electric field amplitude, and η the electromagnetic wave impedance.

The spontaneous emission rate $R_{sp}(\nu)$ does not require photons to be present to induce emission, true to its name. The emitted photon goes into any of the allowed photon modes in the electromagnetic field: this is a random process and there is no correlation of which mode one spontaneously emitted photon goes with the next. The stimulated emission rate $R_{st}(\nu)$ and the absorption rate $R_{ab}(\nu)$ are both proportional to the photon field ρ_v implying they cannot occur in the absence of light. *The photon created by stimulated emission goes into exactly the same electromagnetic mode as the photon that stimulated it*. The emitted photon has the same phase, wavelength, polarization, and is coherent with the photon that stimulated it: it is a clone! The process of stimulated emission is responsible for gain and lasing. In stimulated emission, we start with one photon but end with two - implying *gain*.

Typically under optical amplification and lasing conditions, the spontaneous emission is a low-intensity background. Let us look specifically at the stimulated emission and absorption processes. The net emission rate or downward transitions is given by $R_{em}(\nu) = \underbrace{R_{st}(\nu)}_{\downarrow} - \underbrace{R_{ab}(\nu)}_{\uparrow}$:

$$R_{em}(\nu) = B\rho_{jnt}(\nu)\rho_\nu(f_c(E_2)[1 - f_c(E_1)] - f_v(E_1)[1 - f_c(E_1)]) = B\rho_{jnt}(\nu)\rho_\nu[f_c(E_2) - f_v(E_1)]. \qquad (44)$$

Because each such stimulated emission event increases the photon count by one, and the energy of the photon field by $h\nu$, we should track the energy flowing though a box of area S. The increase in the photon number in a volume $S \cdot dz$ results in an increase in the photon field intensity

$$[I_\nu(z + dz) - I_\nu(z)] \cdot S = R_{em}(\nu) \cdot h\nu \cdot S \cdot dz \implies \frac{dI_\nu(z)}{dz} = R_{em}(\nu)h\nu, \qquad (45)$$

which with the relation between the Einstein A and B coefficients, and the relation $I_\nu(z) = \rho_\nu \cdot c$ leads to

$$\frac{dI_\nu(z)}{dz} = \underbrace{A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h\rho_{jnt}(\nu) \cdot [f_c(E_2) - f_v(E_1)]}_{\gamma_0(\nu)} \cdot I_\nu(z) = \gamma_0(\nu)I_\nu(z), \qquad (46)$$

which indicates the semiconductor gain coefficient

$$\boxed{\gamma_0(\nu) = \underbrace{A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h\rho_{jnt}(\nu)}_{\text{depends on material \& } h\nu} \cdot \underbrace{[f_c(E_2) - f_v(E_1)]}_{\text{population inversion}}} \qquad (47)$$

If $\gamma_0(\nu) > 0$, the intensity $I_\nu(z)$ grows with distance according to $I_\nu(z) = I_\nu(0)e^{\gamma_0(\nu)z}$ and we have photon gain. If $\gamma_0(\nu) < 0$, photons are absorbed, and one should refer to this quantity as the absorption coefficient. The prefactor that depends on the semiconductor JDOS and the photon properties is alway positive. So in order to obtain gain, the Fermi difference function should meet the criteria

$$f_c(E_2) - f_v(E_1) = \frac{1}{1 + e^{\frac{E_2 - F_n}{k_B T}}} - \frac{1}{1 + e^{\frac{E_1 - F_p}{k_B T}}} > 0 \implies F_n - F_p > E_2 - E_1 \implies \boxed{F_n - F_p > h\nu > E_g}. \qquad (48)$$

The boxed condition $F_n - F_p > h\nu > E_g$ is referred to as the Bernard-Duraffourg population inversion criteria for semiconductors, or the condition to achieve optical gain. Since in a $p - n$ diode, $F_n - F_p = qV$, we judge from this condition that it is necessary to apply a voltage larger than the effective bandgap of the optically active region. For photon energies matching $F_n - F_p = h\nu$, the net gain coefficient $\gamma_0(\nu) = 0$, meaning the semiconductor medium is transparent to those photons. For photon energies exceeding this quasi-Fermi level split, the Fermi difference function is negative, leading to loss. Thus, the semiconductor gain spectrum has a bandwidth $\frac{E_g}{h} \leq \nu \leq \frac{F_n - F_p}{h}$, and the shape $\gamma_0(\nu)$ dictated by the product of the optical JDOS, and the Fermi difference function.

It is clear that at equilibrium, or for small levels of carrier injection when $F_n \approx F_p = E_F$, the factor $f_c(E_2) - f_v(E_1) \approx -1$, and we obtain the optical absorption coefficient of the semiconductor $\alpha_0(\nu) = A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h\rho_{jnt}(\nu)$, whereby the change of intensity of photons as they are absorbed $\frac{dI_\nu(z)}{dz} = -\alpha_0(\nu)I_\nu(z)$, leading to $I_\nu(z) = I_\nu(0)e^{-\gamma_0(\nu)z}$.

Returning to spontaneous emission, we obtain a spectrum

$$R_{sp}(\nu) = A\rho_{jnt}(\nu)f_c(E_2)[1 - f_v(E_1)] = A\rho_{jnt}(\nu)\frac{1}{1 + e^{\frac{E_2 - F_n}{k_B T}}} \cdot \frac{e^{\frac{E_1 - F_p}{k_B T}}}{1 + e^{\frac{E_1 - F_p}{k_B T}}} \approx A\rho_{jnt}(\nu)e^{-\frac{h\nu}{k_B T}}e^{\frac{F_n - F_p}{k_B T}} \qquad (49)$$

This condition is important for LEDs, because it indicates that by splitting the quasi-Fermi levels $F_n - F_p = qV$ with a voltage, we exponentially increase the spontaneous emission rate of photons. Of course these photons must be extracted before they get re-absorbed again, because though absorption rate is smaller, it is not negligible!

	0 Dimension	1 Dimension	2 Dimensions	3 Dimensions	d Dimensions
Conduction Bandstructure	E_c	$E_c + \frac{\hbar^2 k_z^2}{2m_c^\star}$	$E_c + \frac{\hbar^2}{2m_c^\star}(k_x^2 + k_y^2)$	$E_c + \frac{\hbar^2}{2m_c^\star}(k_x^2 + k_y^2 + k_z^2)$	$E_c + \frac{\hbar^2}{2m_c^\star}(\sum_{i=1}^d k_i^2)$
Conduction Band DOS $g_c^d(E)$	$g_s g_v \delta(E - E_{cn})$	$\frac{g_s g_v}{2\pi}(\frac{2m_c^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{E - E_c}}$	$\frac{g_s g_v}{2\pi}(\frac{2m_c^\star}{\hbar^2})\theta(E - E_c)$	$\frac{g_s g_v}{(2\pi)^2}(\frac{2m_c^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{E - E_c}$	$\frac{g_s g_v}{(4\pi)^{\frac{d}{2}}\Gamma(\frac{d}{2})}(\frac{2m_c^\star}{\hbar^2})^{\frac{d}{2}}(E - E_c)^{\frac{d-2}{2}}$
Conduction Band-Edge Effective DOS N_c^d	$g_s g_v$	$g_s g_v(\frac{2\pi m_c^\star k_b T}{h^2})^{\frac{1}{2}}$	$g_s g_v(\frac{2\pi m_c^\star k_b T}{h^2})$	$g_s g_v(\frac{2\pi m_c^\star k_b T}{h^2})^{\frac{3}{2}}$	$g_s g_v(\frac{2\pi m_c^\star k_b T}{h^2})^{\frac{d}{2}}$
Electron Density n_d	–	$N_c^{1d}F_{-\frac{1}{2}}(\frac{E_F - E_c}{k_b T})$	$N_c^{2d}F_0(\frac{E_F - E_c}{k_b T})$	$N_c^{3d}F_{+\frac{1}{2}}(\frac{E_F - E_c}{k_b T})$	$N_c^d F_{\frac{d-2}{2}}(\frac{E_F - E_c}{k_b T})$
Source Fermi Level E_{Fs} under bias V	–	$n_{1d} = \frac{1}{2}N_c^{1d}[F_{-\frac{1}{2}}(\frac{E_{Fs} - E_c}{k_b T}) + F_{-\frac{1}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$	$n_{2d} = \frac{1}{2}N_c^{2d}[F_0(\frac{E_{Fs} - E_c}{k_b T}) + F_0(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$	$n_{3d} = \frac{1}{2}N_c^{3d}[F_{+\frac{1}{2}}(\frac{E_{Fs} - E_c}{k_b T}) + F_{+\frac{1}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$	$n_d = \frac{1}{2}N_c^d[F_{\frac{d-2}{2}}(\frac{E_{Fs} - E_c}{k_b T}) + F_{\frac{d-2}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$
Ballistic Electron Current Density J_d at voltage V	–	$\frac{q^2}{h} \cdot N_c^{0d} \cdot \frac{k_b T}{q} \cdot [F_0(\frac{E_{Fs} - E_c}{k_b T}) - F_0(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$	$\frac{q^2}{h} \cdot N_c^{1d} \cdot \frac{k_b T}{q} \cdot [F_{\frac{1}{2}}(\frac{E_{Fs} - E_c}{k_b T}) - F_{\frac{1}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$	$\frac{q^2}{h} \cdot N_c^{2d} \cdot \frac{k_b T}{q} \cdot [F_1(\frac{E_{Fs} - E_c}{k_b T}) - F_1(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$	$\frac{q^2}{h} \cdot N_c^{d-1} \cdot \frac{k_b T}{q} \cdot [F_{\frac{d-1}{2}}(\frac{E_{Fs} - E_c}{k_b T}) - F_{\frac{d-1}{2}}(\frac{(E_{Fs} - qV) - E_c}{k_b T})]$
Valence Bandstructure	E_v	$E_v - \frac{\hbar^2 k_z^2}{2m_v^\star}$	$E_v - \frac{\hbar^2}{2m_v^\star}(k_x^2 + k_y^2)$	$E_v - \frac{\hbar^2}{2m_v^\star}(k_x^2 + k_y^2 + k_z^2)$	$E_v - \frac{\hbar^2}{2m_v^\star}(\sum_{i=1}^d k_i^2)$
Valence Band DOS $g_v^d(E)$	$g_s g_v \delta(E - E_{vn})$	$\frac{g_s g_v}{2\pi}(\frac{2m_v^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{E_v - E}}$	$\frac{g_s g_v}{2\pi}(\frac{2m_v^\star}{\hbar^2})\theta(E_v - E)$	$\frac{g_s g_v}{(2\pi)^2}(\frac{2m_v^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{E_v - E}$	$\frac{g_s g_v}{(4\pi)^{\frac{d}{2}}\Gamma(\frac{d}{2})}(\frac{2m_v^\star}{\hbar^2})^{\frac{d}{2}}(E_v - E)^{\frac{d-2}{2}}$
Valence Band-Edge Effective DOS N_v^d	–	$g_s g_v(\frac{2\pi m_v^\star k_b T}{h^2})^{\frac{1}{2}}$	$g_s g_v(\frac{2\pi m_v^\star k_b T}{h^2})$	$g_s g_v(\frac{2\pi m_v^\star k_b T}{h^2})^{\frac{3}{2}}$	$g_s g_v(\frac{2\pi m_v^\star k_b T}{h^2})^{\frac{d}{2}}$
Hole Density p_d	–	$N_v^{1d}F_{-\frac{1}{2}}(\frac{E_v - E_F}{k_b T})$	$N_v^{2d}F_0(\frac{E_v - E_F}{k_b T})$	$N_v^{3d}F_{+\frac{1}{2}}(\frac{E_v - E_F}{k_b T})$	$N_v^d F_{\frac{d-2}{2}}(\frac{E_v - E_F}{k_b T})$
Source Fermi Level E_{Fs} under bias V	–	$p_{1d} = \frac{1}{2}N_v^{1d}[F_{-\frac{1}{2}}(\frac{E_v - E_{Fs}}{k_b T}) + F_{-\frac{1}{2}}(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$	$p_{2d} = \frac{1}{2}N_v^{2d}[F_0(\frac{E_v - E_{Fs}}{k_b T}) + F_0(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$	$p_{3d} = \frac{1}{2}N_v^{3d}[F_{\frac{1}{2}}(\frac{E_v - E_{Fs}}{k_b T}) + F_{\frac{1}{2}}(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$	$p_d = \frac{1}{2}N_v^d[F_{\frac{d-2}{2}}(\frac{E_v - E_{Fs}}{k_b T}) + F_{\frac{d-2}{2}}(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$
Ballistic Hole Current Density J_d	–	$\frac{q^2}{h} \cdot N_v^{0d} \cdot \frac{k_b T}{q} \cdot [F_0(\frac{E_v - E_{Fs}}{k_b T}) - F_0(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$	$\frac{q^2}{h} \cdot N_v^{1d} \cdot \frac{k_b T}{q} \cdot [F_{\frac{1}{2}}(\frac{E_v - E_{Fs}}{k_b T}) - F_{\frac{1}{2}}(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$	$\frac{q^2}{h} \cdot N_v^{2d} \cdot \frac{k_b T}{q} \cdot [F_1(\frac{E_v - E_{Fs}}{k_b T}) - F_1(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$	$\frac{q^2}{h} \cdot N_v^{d-1} \cdot \frac{k_b T}{q} \cdot [F_{\frac{d-1}{2}}(\frac{E_v - E_{Fs}}{k_b T}) - F_{\frac{d-1}{2}}(\frac{E_v - (E_{Fs} - qV)}{k_b T})]$
Electron-Hole-Photon Relation	$E_{cn} - E_{vn} = h\nu$	$E_g + \frac{\hbar^2 k_z^2}{2m_r^\star} = h\nu, (E_g = E_c - E_v, \frac{1}{m_r^\star} = \frac{1}{m_c^\star} + \frac{1}{m_v^\star})$	$E_g + \frac{\hbar^2}{2m_r^\star}(k_x^2 + k_y^2) = h\nu$	$E_g + \frac{\hbar^2}{2m_r^\star}(k_x^2 + k_y^2 + k_z^2) = h\nu$	$E_g + \frac{\hbar^2}{2m_r^\star}(\sum_{i=1}^d k_i^2) = h\nu$
Optical Joint DOS $\rho_{jnt}(h\nu)$ in $eV \cdot cm^{-3}$	$\frac{g_s g_v}{L_x L_y L_z}\delta[h\nu - (E_{cn} - E_{vn})]$	$\frac{g_s g_v}{2\pi L_y L_z}(\frac{2m_r^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{h\nu - E_g}}$	$\frac{g_s g_v}{2\pi L_z}(\frac{2m_r^\star}{\hbar^2})\theta(h\nu - E_g)$	$\frac{g_s g_v}{(2\pi)^2}(\frac{2m_r^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{h\nu - E_g}$	$\frac{g_s g_v}{(4\pi)^{\frac{d}{2}}L^{3-d}\Gamma(\frac{d}{2})}(\frac{2m_r^\star}{\hbar^2})^{\frac{d}{2}}(h\nu - E_g)^{\frac{d-2}{2}}$
Electron Density in the Optically Active Region	–	$N_c^{1d}F_{-\frac{1}{2}}(\frac{F_n - E_c}{k_b T})$	$N_c^{2d}F_0(\frac{F_n - E_c}{k_b T})$	$N_c^{3d}F_{+\frac{1}{2}}(\frac{F_n - E_c}{k_b T})$	$N_c^d F_{\frac{d-2}{2}}(\frac{F_n - E_c}{k_b T})$
Hole Density in the Optically Active Region	–	$N_v^{1d}F_{-\frac{1}{2}}(\frac{E_v - F_p}{k_b T})$	$N_v^{2d}F_0(\frac{E_v - F_p}{k_b T})$	$N_v^{3d}F_{+\frac{1}{2}}(\frac{E_v - F_p}{k_b T})$	$N_v^d F_{\frac{d-2}{2}}(\frac{E_v - F_p}{k_b T})$
Excess Electron-Hole Pair Density	–	$N_c^{1d}N_v^{1d}[F_{-\frac{1}{2}}(\frac{F_n - E_c}{k_b T})F_{-\frac{1}{2}}(\frac{E_v - F_p}{k_b T}) - F_{-\frac{1}{2}}^2(\frac{E_F - E_c}{k_b T})]$	$N_c^{2d}N_v^{2d}[F_0(\frac{F_n - E_c}{k_b T})F_0(\frac{E_v - F_p}{k_b T}) - F_0^2(\frac{E_F - E_c}{k_b T})]$	$N_c^{3d}N_v^{3d}[F_{+\frac{1}{2}}(\frac{F_n - E_c}{k_b T})F_{+\frac{1}{2}}(\frac{E_v - F_p}{k_b T}) - F_{+\frac{1}{2}}^2(\frac{E_F - E_c}{k_b T})]$	$N_c^d N_v^d[F_{\frac{d-2}{2}}(\frac{F_n - E_c}{k_b T})F_{\frac{d-2}{2}}(\frac{E_v - F_p}{k_b T}) - F_{\frac{d-2}{2}}^2(\frac{E_F - E_c}{k_b T})]$
Spontaneous Emission Spectrum $R_{sp}(\nu)$	–	$A \cdot \frac{g_s g_v}{2\pi L_y L_z}(\frac{2m_r^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{h\nu - E_g}} \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$	$A \cdot \frac{g_s g_v}{2\pi L_z}(\frac{2m_r^\star}{\hbar^2})\theta(h\nu - E_g) \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$	$A \cdot \frac{g_s g_v}{(2\pi)^2}(\frac{2m_r^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{h\nu - E_g} \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$	$A \cdot \rho_{jnt}(\nu) \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$
Stimulated Emission Spectrum $R_{st}(\nu)$	–	$B \cdot \rho_v \cdot \frac{g_s g_v}{2\pi L_y L_z}(\frac{2m_r^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{h\nu - E_g}} \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$	$B \cdot \rho_v \cdot \frac{g_s g_v}{2\pi L_z}(\frac{2m_r^\star}{\hbar^2})\theta(h\nu - E_g) \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$	$B \cdot \rho_v \cdot \frac{g_s g_v}{(2\pi)^2}(\frac{2m_r^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{h\nu - E_g} \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$	$B \cdot \rho_v \cdot \rho_{jnt}(\nu) \cdot f_c(E_2) \cdot [1 - f_v(E_1)]$
Absorption Spectrum $R_{ab}(\nu)$	–	$B \cdot \rho_v \cdot \frac{g_s g_v}{2\pi L_y L_z}(\frac{2m_r^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{h\nu - E_g}} \cdot f_v(E_1) \cdot [1 - f_c(E_2)]$	$B \cdot \rho_v \cdot \frac{g_s g_v}{2\pi L_z}(\frac{2m_r^\star}{\hbar^2})\theta(h\nu - E_g) \cdot f_v(E_1) \cdot [1 - f_c(E_2)]$	$B \cdot \rho_v \cdot \frac{g_s g_v}{(2\pi)^2}(\frac{2m_r^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{h\nu - E_g} \cdot f_v(E_1) \cdot [1 - f_c(E_2)]$	$B \cdot \rho_v \cdot \rho_{jnt}(\nu) \cdot f_v(E_1) \cdot [1 - f_c(E_2)]$
Photonic Gain Spectrum of Semiconductor $\gamma_0(\nu)$	–	$A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h \frac{g_s g_v}{2\pi L_y L_z}(\frac{2m_r^\star}{\hbar^2})^{\frac{1}{2}}\frac{1}{\sqrt{h\nu - E_g}} \cdot [f_c(E_2) - f_v(E_1)]$	$A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h \frac{g_s g_v}{2\pi L_z}(\frac{2m_r^\star}{\hbar^2})\theta(h\nu - E_g) \cdot [f_c(E_2) - f_v(E_1)]$	$A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h \frac{g_s g_v}{(2\pi)^2}(\frac{2m_r^\star}{\hbar^2})^{\frac{3}{2}}\sqrt{h\nu - E_g} \cdot [f_c(E_2) - f_v(E_1)]$	$A \cdot \frac{\lambda_0^2}{8\pi n^2} \cdot h\rho_{jnt}(\nu) \cdot [f_c(E_2) - f_v(E_1)]$

Table 1: Quantum Electronic, Photonic, and Statistical Properties of Conduction and Valence Band Electrons in Semiconductor Nanostructures.

- E_c is the band edge, and $m_c^\star$ the effective mass of the conduction band. E_v is the band edge, and $m_v^\star$ the effective mass of the valence band.
- For low dimensions, E_c and E_v, and the bandgap $E_g = E_c - E_v$ include the quantum confinement energies if present.
- h is Planck's constant, $\hbar = \frac{h}{2\pi}$, k_b is the Boltzmann constant, and q the electron charge.
- g_s is the spin degeneracy, and g_v the valley degeneracy.
- $F_j(\eta) = \frac{1}{\Gamma(j+1)}\int_0^\infty du \frac{u^j}{1 + e^{u - \eta}}$ is the Fermi-Dirac integral of order j, and $\Gamma(...)$ is the Gamma function.
- E_F is the Fermi level at equilibrium. E_{Fs} is the source quasi-Fermi level and E_{Fd} the drain quasi-Fermi level.
- Similarly, F_n is the conduction band quasi-Fermi level and F_p is the valence band quasi-Fermi level.
- $h\nu$ is the photon energy of frequency ν, and L_x, L_y, L_z are the dimensions of the semiconductor nanostructure.
- A and B are the Einstein A and B coefficients, $\lambda_0 = c/\nu$ the wavelength of the photon in vacuum, and n the refractive index of the semiconductor.
- $f_c(E_2) = 1/[1 + e^{\frac{E_2 - F_n}{k_b T}}]$ is the Fermi-Dirac occupation function of state $E_2 = E_c + \frac{\hbar^2 k^2}{2m_c^\star}$ in the conduction band.
- $f_v(E_1) = 1/[1 + e^{\frac{E_1 - F_p}{k_b T}}]$ is the Fermi-Dirac occupation function of state $E_1 = E_v - \frac{\hbar^2 k^2}{2m_v^\star}$ in the valence band.
- $E_2 - E_1 = h\nu = E_g + \frac{\hbar^2 k^2}{2m_r^\star}$ is the energy of the photon emitted when the electron transitions from $E_2 \to E_1$ radiatively.
- The Einstein A and B coefficients are related by $\frac{A}{B} = \frac{8\pi n^3 \nu^3 h}{c^3}$.
- The photon density is $\rho_\nu = I_\nu/(c/n)$ in eV/cm^3, c/n is the speed of light in a media of refractive index n, and $I_\nu = E_0^2/2\eta$ in W/cm^2 is the Poynting energy density with electric field amplitude E_0 and wave impedance η.