Classical Mechanics:
\[ \mathbf{F} = m\mathbf{a} \text{ from an unusual perspective} \]

Quantum Mechanics is the background of semiconductor physics: Semiconductors operate thanks to the existence of ‘energy gaps’ whose existence stems from the wavelike nature of electrons; many modern devices, such as junction (injection) lasers, quantum cascade lasers, resonant tunnel diodes principles of operation are based on basic quantum properties of electrons; gate leakage-current mechanisms of present Si CMOS threaten the success of future scaling and stem from quantum mechanical tunneling.... Thus we must revisit quantum mechanics. The needs for a nonclassical view of Nature will be reviewed in the next lecture. Here we want to start with a preamble, apparently disconnected: We want to see how Classical Mechanics can be reformulated in a very abstract way which will allow us to draw a close connection between the formalism of Classical Mechanics so reformulated and the mathematical foundations Quantum Mechanics.

- Define a physical system with \( N \) degrees of freedom via **canonical coordinates** \( \{q_i\} \) and **conjugate momenta** \( \{p_i\} \) (\( i = 1, N \)).
  
  **Examples:**
  
  1. \( q_1, q_2, q_3 = x, y, z \)
     \[ p_1, p_2, p_3 = p_x = mv_x, p_y = mv_y, p_z = mv_z \text{ for a single particle of mass } m \text{ in 3D.} \]
  2. \( q_1, q_2, q_3 = x_1, y_1, z_1 \)
     \[ q_4, q_5, q_6 = x_2, y_2, z_2 \]
     \[ p_1, p_2, p_3 = p_{x,1}, p_{y,1}, p_{z,1} \]
     \[ p_4, p_5, p_6 = p_{x,2}, p_{y,2}, p_{z,2} \text{ for 2 particles in 3D} \]
  3. \( q_1 = \phi \) and \( p_1 = I\omega_\phi = \dot{\phi} \text{ for a rigid body in 2D} \)

- Let’s define the **Poisson brackets** between two physical ‘observables’ \( A(q, p) \) and \( B(q, p) \) as follows:

\[
\{A, B\} = \sum_{i=1}^{N} \left( \frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \right) \quad (1)
\]
Note that \( \{ q_i, q_j \} = 0 \quad \{ p_i, p_j \} = 0 \quad \{ q_i, p_j \} = \delta_{ij} \).

- Let's define the **Hamiltonian** function(al):
  \[
  H(q, p) = T(q, p) + V(q, p)
  \]
  where \( T \) and \( V \) are, respectively, the kinetic and potential energy of the system.

**Examples:**

1. For \( N \) particles \((i = 1, N \text{ is now the particle index!})\) of mass \( m \) and charge \( e \) interacting via Coulomb force:

   \[
   H(q, p) = \sum_{i=1}^{N} \left( \frac{p_i^2}{2m} - \sum_{j \neq i}^{N} \frac{e^2}{4\pi\varepsilon_0 |q_i - q_j|^2} \right).
   \]  

2. For a free rigid body in 2D:

   \[
   H(\phi, L_\phi) = \frac{1}{2} I\omega^2 = \frac{L^2_\phi}{2I}.
   \]

- **Equations of motion:** It can be shown (starting from the Lagrangian formulation) the the time evolution of the dynamic variable \( A(q, p) \) is given by the equation:

  \[
  \frac{dA}{dt} = \{ A, H \} + \frac{\partial A}{\partial t}.
  \]

In particular (for the canonical coordinates):

1. For a particle in 1D:

   \[
   \frac{dq_i}{dt} = \{ q_i, H \} = \frac{\partial H}{\partial p_i}
   \]

   \[
   \frac{dp_i}{dt} = \{ p_i, H \} = -\frac{\partial H}{\partial q_i}
   \]
\[
\frac{dx}{dt} = \frac{\partial x}{\partial x} \frac{\partial H}{\partial p} - \frac{\partial x}{\partial p} \frac{\partial H}{\partial x} = \frac{\partial H}{\partial p} = \frac{\partial}{\partial p} \left( \frac{p^2}{2m} \right) = \frac{p}{m}
\]  

so that \( p = mv \), and:

\[
\frac{dp}{dt} = \frac{\partial p}{\partial x} \frac{\partial H}{\partial p} - \frac{\partial p}{\partial p} \frac{\partial H}{\partial x} = -\frac{\partial H}{\partial x} = -\frac{\partial V}{\partial x} = F
\]  

so that, using eq. above, \( ma = F \) ... finally!

- An important note on symmetry:
  
  If the Hamiltonian is invariant under a transformation, an associated conjugate momentum is a ‘constant of motion’ (that is, is conserved). Examples:

1. Hamiltonian independent of \( x \) (that is, invariant under translations along \( x \)) \( \rightarrow \) momentum \( p_x \) is conserved
2. Hamiltonian independent of \( p \) (that is, invariant under translations along \( p_x \) in momentum (Fourier) space) \( \rightarrow \) coordinate \( x \) is conserved (rigid-body, it does not move by translation)
3. Hamiltonian independent of \( \phi \) (that is, invariant under rotations) \( \rightarrow \) angular momentum \( L_{\phi} \) is conserved
4. Hamiltonian explicitly independent of \( t \) (that is, invariant under time-translations) \( \rightarrow \) total energy \( E \) is conserved
The crisis of Classical Mechanics

At the end of the XIX century Physics were entirely (and apparently successfully) based on Newtonian (classical) mechanics – CM – and the electromagnetic – EM – theory of Maxwell’s equations. The former had been exceptionally successful in explaining celestial phenomena occurring the Solar System, leading to calculations of the orbit of planets and moons which had an unprecedented high degree of agreement with observations (think of Le Verrirer’s prediction of the existence of Neptune or of Delaunay’s explanations of several features of the very complicated lunar orbit). EM had reached its climax in Maxwell’s mathematical proof of Faraday’s ‘hunch’ that light was nothing but a Hertzian wave. So, CM and EM seemed to be able to make us predict the future: As Laplace had stated, if a demon told us the position and velocity of all the particles in the Universe at a given time, nothing would prevent us from solving Newton’s and Maxwell’s equations up to the far distant future without any uncertainty about the outcome. It was the pinnacle of the mechanistic ideal of the Scientific Revolution which started in the XVII century. Yet, at the turn of the XX century there were several ‘small nagging’ issues which did not quite fit the accepted theories: The spectrum emitted by a black body, the nature of the photoelectric effect, the structure of the just-discovered atoms, and a few more peculiar experimental observations, such as Compton scattering between an electron and light. Let’s consider the first three, and see what sort of monumental changes had to be made to the neat and cozy picture of CM and EM Physics.

• **Black Body spectrum**: The ultraviolet catastrophe.
  – Classical electromagnetic theory and equipartition imply that the spectrum of a black-body should diverge at short wavelength. That is: if \( n(\nu) \) is density of modes with frequency in \( \nu, \nu + d\nu \), then \( n(\nu) \propto \nu^2 \) (Jeans’ law)
  – Experiments show, instead, \( n(\nu) \sim e^{-a\nu/(k_B T)} \)
  – Planck obtained the experimental spectrum assuming that:
    1. the energy in each mode proportional to the \( \nu \) of mode: \( E_\nu = h\nu \) with \( h \) a new (Planck) constant of nature
    2. energy is exchanged between waves and the walls of the black-body only via ‘chunks’ (‘quanta’) of size \( \Delta E_\nu = h\nu \)
• **The photoelectric effects**: Nothing works correctly...!
  – Hitting a solid with light, electrons should be emitted with energy proportional to the intensity of the light...
  – Instead, the electron energy grows with decreasing wavelength of the light. The light-intensity determines the number of emitted electrons
  – Einstein explained experiments with Planck’s quanta (1905, Nobel prize in 1921)

• **Structure and spectra of atoms**: why don’t electrons fall into the nucleus?
  – After Rutherford’s experiments, atoms were viewed as electrons orbiting the nucleus. But Maxwell equations predicted that the electrons should radiate their energy and fall into the nucleus.
  – Planck’s quanta can explain this as well, although more is needed to explain the spectra

• There’s more (Compton scattering, double-slit experiment,...). See any textbook on QM.

• Partial conclusion:
  Light behaves sometimes as wave (after all, Maxwell equations are valid in most cases), sometimes as a particle (photon)... Confusing...
Quick review of Quantum Mechanics

• **de Broglie's assumption** to explain atomic spectra (among other things):
  Why not assume that *all* forms of matter behave as particles or waves, depending on the situation? For a particle of mass $m$ the wavelength $\lambda$ of the associated ‘pilot wave’ will be give by:

$$p = mv = \frac{\hbar}{\lambda} \quad (9)$$

This may explain discrete electron ‘orbits' around the nucleus...

• **Schrödinger equation**: Explain structure of atom as a ‘boundary value' problem.
  Start from the classical Hamiltonian:

$$H = \frac{p^2}{2m} - \frac{e^2}{4\pi\epsilon_0 r}, \quad (10)$$

equate $H$ to $i\hbar\partial/\partial t$ (time translations are generated by energy) and replace $p$ with $-i\hbar\nabla$ (space translations are generated by momentum). Then solve the wave-equation for the ‘wavefunction' $\Psi(r, t)$:

$$H\Psi(r, t) = \left[ -\frac{\hbar^2\nabla^2}{2m} - \frac{e^2}{4\pi\epsilon_0 r} \right] \Psi(r, t) = i\hbar \frac{\partial\Psi(r, t)}{\partial t}. \quad (11)$$

If $H$ is time-independent, set $\Psi(r, t) = \psi(r)e^{-iEt/\hbar}$ with:

$$-\frac{\hbar^2}{2m}\nabla^2\psi(r) + V(r)\psi(r) = E\psi(r) \quad (12)$$

(time independent Schrödinger equation)

• What’s the wavefunction?
  **Copenhagen interpretation** (Bohr, Born):
$|\Psi(r, t)|^2 dr = \text{probability of finding an electron in a volume } dr \text{ around } r \text{ at time } t$. Of course:

$$\int \rho(r, t) \, dr = \int |\Psi(r, t)|^2 \, dr = 1$$

(13)

and from Schrödinger equation one can show that:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot S = 0 ,$$

(14)

where $S = \frac{i\hbar}{2m}[\Psi \nabla \Psi^* - \Psi^* \nabla \Psi]$ is the ‘probability density current’

- A simple example: An 1D electron in a constant potential $V(z) = V_0$:

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi(z)}{dz^2} = (E - V_0)\psi(z) .$$

(15)

Put $\psi(z) = e^{ikz}$ and find $k = \pm [2m(E - V_0)]^{1/2}/\hbar$ or $E = \hbar^2 k^2/(2m)$, so that $\hbar k$ is the $z$-component of the electron momentum.

- Another simple example: Particle in a box.

Potential of a ‘quantum well’:

$$V(z) = \begin{cases} \infty & \text{for } z \leq 0 \\ 0 & \text{for } 0 < z < L \\ \infty & \text{for } z \geq L \end{cases} .$$

(16)

Boundary conditions:

$$\psi(z) = 0 \text{ for } z = 0 \text{ and } z = L ,$$

(17)
since the electron cannot be in the ‘forbidden’ regions. Schrödinger equation:

\[-\frac{\hbar^2}{2m} \frac{d^2 \psi(z)}{dz^2} = E \psi(z) \text{ for } -L/2 < z < L/2, \tag{18}\]

with solutions \(\sin(k_n z)\) only for discrete values of \(k_n = \pi n/L:\)

\[\psi(z) = \left(\frac{2}{L}\right)^{1/2} \sin \left(\frac{n\pi z}{L}\right) \tag{19}\]

with \(E_n = \frac{n^2 \hbar^2 \pi^2}{2mL^2} \).

The energy depends on the boundary conditions at \(L\), like the note emitted by a vibrating violin string depends on where one places the finger.

- A tunneling problem: Potential of a 'barrier':

\[V(z) = \begin{cases} 
0 & \text{for } z \leq 0 \\
V > 0 & \text{for } 0 < z < L \\
0 & \text{for } z \geq L
\end{cases} \tag{20}\]

General solution for electron of energy \(E < V\) coming from the left, partially reflected back and partially transmitted:

\[\psi(z) = \begin{cases} 
A e^{ikz} + B e^{-ikz} & \text{for } z \leq 0 \\
Ce^{\kappa z} + De^{-\kappa z} & \text{for } 0 < z < L \\
Fe^{ikz} & \text{for } z \geq L
\end{cases} \tag{21}\]

with \(k = (2mE)^{1/2}/\hbar, \kappa = [2m(V - E)]^{1/2}/\hbar\).

Using continuity of \(\psi(z)\) and \(d\psi(z)/dz\) at \(z = 0\) and \(z = L\) one can find the coefficients \(A, B, C, D,\) and \(F\).

The ratio \(T = |F|^2/|A|^2\) is the ‘transmission coefficient’, the ratio \(R = |B|^2/|A|^2\) is the ‘reflection coefficient’. Note that, classically, one would expect \(T = 0\).
• A curiosity: Heisenberg matrices vs Schrödinger wavefunctions.
  – Only ‘expectation values’ of observables $A$ matter:
    \[
    \langle A \rangle = \langle \psi | A | \psi \rangle = \int \psi^*(r) A(r) \psi(r) \, dr
    \]  
    (22)

  – In Schrödinger ‘representation’ wavefunctions evolve in time, observables stay put...
    \[
    \langle A(t) \rangle_S = \langle \psi(t) | A | \psi(t) \rangle = \int \psi^*(r, t) A(r) \psi(r, t) \, dr
    \]  
    (23)

    For $H$ not explicitly dependent on time the solution of Schrödinger eq. is:
    \[
    |\psi(t)\rangle = e^{-iHt/\hbar} |\psi(0)\rangle
    \]  
    (24)

  – In Heisenberg’s ‘representation wavefunctions (or ‘states’) stay put, observables (‘operators’ which are matrices in infinite-dimension spaces) evolve with time with equations of motion which are just the Hamilton equations of classical mechanics, but with ‘commutators’ replacing the Poisson brackets: for an observable not explicitly dependent on time
    \[
    \frac{dA}{dt} = -\frac{i}{\hbar} [A, H],
    \]  
    (25)

    where $[A, B] = AB - BA$ (in general not zero, see rotations about two orthogonal axes, for example). Then:
    \[
    \langle A(t) \rangle_H = \langle \psi | A(t) | \psi \rangle = \int \psi^*(r) A(r, t) \psi(r) \, dr
    \]  
    (26)

    For $H$ not explicitly dependent on time the solution of Heisenberg eq. is:
    \[
    A(t) = e^{iHt/\hbar} A(0) e^{-iHt/\hbar}
    \]  
    (27)

  – It can be shown that the two representations are fully equivalent.
− **Canonical quantization**: Note how the Heisenberg equation of motion for the operator $A$ can be obtained from the classical Hamilton equation of motion by replacing the Poisson bracket $\{\ldots, \ldots\}$ with $-i/\hbar$ times the commutator $[\ldots, \ldots]$. 
• More on the Copenhagen interpretation.

1. To each physical system we associate a Hamiltonian and a linear (vector) space of ‘states’ (or wavefunctions) called ‘Hilbert space’.
2. Dynamic variables (observables such as energy, linear momentum angular momentum, etc.) are operators (like matrices) on this linear space.
3. If a state \( |\psi> \) is such that \( A |\psi> = a |\psi> \), (i.e., the state \( |\psi> \) is an eigenvector of \( A \) with eigenvalue \( a \)) a measurement of the observable \( A \) on the system in the state \( |\psi> \) will give the reading \( a \). For example, Schrödinger equation \( H |\psi> = E |\psi> \) says that the state \( |\psi> \) has energy \( E \).
4. If the system is in a linear superposition of eigenvectors \( |\psi> = c_1 |\psi_1> + c_2 |\psi_2> \), with \( A |\psi_1> = a_1 |\psi_1> \) and \( A |\psi_2> = a_2 |\psi_2> \) then a measurement of \( A \) will give:

\[
<\psi|A|\psi> = |c_1|^2 a_1 + |c_2|^2 a_2
\]

i.e. the result \( a_i \) will occur with probability \( |c_i|^2 \).

Note: We cannot predict with certainty which result we will obtain from the measurement. Even if we repeat the same measurements on many systems identically ‘prepared’, we shall not obtain always the same result. But after the measurement, the system will have ‘collapsed’ into one of the eigenstates. In a way, the measurement perturbs the system.

5. A system cannot be in an eigenstate of two non-commuting observables \( A \) and \( B \) (that is, \( [A, B] \neq 0 \)). In this case \( A \) and \( B \) are said to be ‘complementary’ observables. They cannot be measured simultaneously. More precisely,

\[
\Delta < A > \Delta < B > \neq 0 \tag{29}
\]

where \( \Delta < A > = [<\psi|A^2|\psi> - <\psi|A|\psi>^2]^{1/2} \).

In particular,

\[
[q_i, p_j] = i\hbar \delta_{ij} \rightarrow \Delta < q_i > \Delta < p_i > \geq \hbar / 2 \tag{30}
\]

(Heisenberg uncertainty principle).

Note: A system cannot be thought of having property \( A \) and property \( B \) at the same time, since only our act of measuring will determine which property the system will have...(!?)
• Perturbation theory I: Stationary (that is, time-independent).

- We must solve the problem

$$ (H_0 + \alpha H) |\phi_n> = W_n |\phi_n> $$

which is too complicated. However, $\alpha$ is small and we know how to solve $H_0|\psi_n> = E_n|\psi_n>$. 

- Then, one can show that the perturbed eigenvalues are:

$$ W_m = E_m + \alpha H_{mm} + \alpha^2 \sum_{n \neq m} \frac{|H_{nm}|^2}{E_m - E_n} + o(\alpha^3), $$

(32)

(where $H_{mn} = <\psi_n|H|\psi_m> = \int \psi_n^*(r)H\psi(r)dr$) and the perturbed eigenfunctions are:

$$ |\phi_m> = |\psi_m> + \alpha \sum_{n \neq m} \frac{H_{nm}}{E_m - E_n} |\psi_n> $$

(33)

Partial proof. Write $|\phi_n> = |\psi_n> + \alpha|\phi_n^{(1)}> + \alpha^2|\phi_n^{(2)}> + \ldots$ and $W_n = E_n + \alpha E_n^{(1)} + \alpha^2 E_n^{(2)} + \ldots$ and insert into the initial Schrödinger eq.:

$$ (H_0 + \alpha H) (|\psi_n> + \alpha|\phi_n^{(1)}> + \alpha^2|\phi_n^{(2)}> + \ldots) = (E_n + \alpha E_n^{(1)} + \alpha^2 E_n^{(2)} + \ldots) (|\psi_n> + \alpha|\phi_n^{(1)}> + \alpha^2|\phi_n^{(2)}> + \ldots) $$

(34)

Equating terms of the same order in $\alpha$ we get:

$$ H_0|\psi_n> = E_n|\psi_n> \text{ to order } \alpha^0; \text{ this is the unperturbed problem} $$

(35)

$$ H|\psi_n> + H_0|\phi_n^{(1)}> = E_n^{(1)}|\psi_n> + E_n|\phi_n^{(1)}> \text{ to order } \alpha. $$

(36)

Now put $|\phi_n^{(1)}> = \sum_i a_{ni} |\psi_i>$ so that this equation becomes:

$$ H|\psi_n> + H_0 \sum_i a_{ni} |\psi_i> = E_n^{(1)}|\psi_n> + E_n \sum_i a_{ni} |\psi_i> . $$

(37)
Now multiply by $<\psi_l|$ and integrate:

$$H_{ln} + E_l a_{nl} = E_n^{(1)} \delta_{ln} + E_n a_{nl}$$

(38)

For $l = n$ we get immediately $E_n^{(1)} = H_{nn}$. For $l \neq n$ we get

$$a_{nl} = \frac{H_{ln}}{E_n - E_l}$$

(39)

- **Perturbation theory II: Time-dependent.**
  - Now we must solve the following problem in which the perturbation term is time-dependent:

$$[H_0 + H(t)] |\Phi> = i\hbar \frac{\partial |\Phi>}{\partial t}$$

(40)

and we know that $H_0 |\psi_n> = E_n |\psi_n>$ and $|\Psi_n(t)> = e^{-iE_nt/\hbar} |\psi_n>.$

- Write $|\Phi(t)> = \sum a_n(t) |\Psi_n>$ so that $\partial |\Phi(t)> / \partial t = \sum_n [\dot{a}_n - (i/\hbar) E_n a_n] |\Psi_n>.$ Inserting into original equation we get:

$$[H_0 + H(t)] \sum_n a_n(t) |\Psi_n> = \sum_n [i\hbar \dot{a}_n + E_n a_n] |\Psi_n>,$$

(41)

or

$$\sum_n a_n(t)(H_0 - E_n) |\Psi_n> + \sum_n [a_n H(t) - i\hbar \dot{a}_n] |\Psi_n> = 0 .$$

(42)

The first term vanishes. Multiplying by $<\Psi_l|$ and integrating (recalling that $<\Psi_l(t)|\Psi_n(t)> = \delta_{ln}$):

$$\sum_n [a_n(t) H_{ln}(t) e^{i(E_l - E_n)t/\hbar} - i\hbar \dot{a}_n \delta_{ln}] = 0 .$$

(43)

Put $\hbar \omega_{ln} = E_l - E_n$, so this equation becomes:

$$i\hbar \dot{a}_l = \sum_n a_n(t) H_{ln} e^{i\omega_{ln} t} .$$

(44)
Now assume 1. that at $t = 0$ the system was in the state $|\psi_j>$ (so that $a_j(0) = 1$, all other $a(0)$’s vanish) 2. the term $H$ is weak so that the coefficients $a_n$ in the rhs can be taken as constant. Then:

$$a_j(t) \approx 1 - \frac{i}{\hbar} \int_0^t H_{jj}(t) \, dt \rightarrow 1 - \frac{i}{\hbar} H_{jj} \ t \quad \text{for constant } H \text{ turned on at } t = 0 \quad (45)$$

$$a_k(t) \approx -\frac{i}{\hbar} \int_0^t H_{kj}(t) \, e^{i\omega_k t} \, dt \rightarrow \frac{H_{kj}}{\hbar \omega_k} (1 - e^{i\omega_k t}) \quad \text{for } k \neq j \ . \quad (46)$$

Now the probability of finding the system in the state $k \neq j$ at time $t$ will be $P_k(t) = |a_k(t)|^2$. The ‘transition rate’ (that is, probability per unit time) will be $P_k(t)/t$, which is:

$$\frac{1}{t} \left| \frac{H_{kj}}{\hbar \omega_k} (1 - e^{i\omega_k t}) \right|^2 = \left| \frac{2H_{kj}}{E_k - E_j} \right|^2 \frac{1}{t} \sin^2 \left( \frac{E_k - E_j}{2\hbar} t \right) = |H_{kj}|^2 \frac{4 \sin^2[(E_k - E_j)t/(2\hbar)]}{[E_k - E_j]^2 t} \ . \quad (47)$$

Now

$$\lim_{t \to \infty} \frac{4 \sin^2[(E_k - E_j)t/(2\hbar)]}{[E_k - E_j]^2 t} = \frac{2\pi}{\hbar} \delta(E_k - E_j) \ , \quad (48)$$

so that:

$$\text{transition rate } j \rightarrow k = \frac{2\pi}{\hbar} |H_{kj}|^2 \delta(E_k - E_j) \ , \quad (49)$$

which is the famous Fermi Golden Rule. For a harmonic perturbation, $H(t) = H(0)e^{-i\omega t}$, one gets a similar expression:

$$S(j \rightarrow k) = \frac{2\pi}{\hbar} |H_{kj}|^2 \delta(E_k - E_j - \omega) \ , \quad (50)$$
Crystals

- Definitions
  - **Crystals** are regular arrangements of atoms which look identical when viewed from 2 points

\[ r \quad \text{and} \quad r' = r + R_l, \tag{51} \]

where

\[ R_l = l_1 a + l_2 b + l_3 c \tag{52} \]

is a ‘translation operation’ with integers \( l_1, l_2, \) and \( l_3 \) and with ‘fundamental translation vectors’ \( a, b, \) and \( c. \) The vector \( R_l \) is also called a **lattice vector**.

- A **cell** is the parallelepiped formed by the fundamental translation vectors. Note that the repetition of a cell obtained by applying a translation operation fills the entire space.

- The **primitive** or **Wigner-Seitz cell** is the cell having the smallest possible volume, given by \( V_{\text{cell}} = |a \times b \cdot c|. \) It can generated by connecting all lattice points with straight lines, bisecting these lines with planes normal to the lines, and considering the volume enclosed by these planes.

- A **lattice** is the set of all points \( r' \) generated by translation operations. The symmetry properties of a lattice allow the existence of 14 types of **Bravais lattices** (those generated by Wigner-Seitz cells).

- A **basis** is the set of the coordinates of all atoms within a primitive cell. Logically the crystal structure is given by:

\[ \text{crystal structure} = \text{lattice} + \text{basis} \tag{53} \]

- An arbitrary plane will intersects the axes \( a, b, \) and \( c \) at distances \( (A, B, C) \) (in units of the lengths \( (a,b,c), \) the lattice constants). Take the reciprocal numbers \( (1/A, 1/B, 1/C) \) and find the smallest integers \( (l, m, n) \) having the same ratios as the reciprocals. The integers \( l, m, n \) are the **Miller indices** of the plane.
• Cubic and zinc-blend lattices
  - Most semiconductors of interest have a ‘tetragonal’ coordination (resulting from the \( sp^3 \) hybridization of the bonding orbitals). This yields a **face centered cubic (fcc) lattice**.
  - The WS cell of fcc lattices is generated by the fundamental vectors

\[
\mathbf{a} = \frac{a}{2}(\hat{x} + \hat{y}), \quad \mathbf{b} = \frac{a}{2}(\hat{y} + \hat{z}), \quad \mathbf{c} = \frac{a}{2}(\hat{z} + \hat{x}).
\]

(54)

• **Point Group**.
  There are 24 fundamental symmetry transformation which map the cubic lattice onto itself (the so-called \( T_d \) sub-group): All permutations of the 3 coordinates (6 operations) times sign-swapping two of them (4 operations).
  Accounting for inversions, there are 48 symmetry operations in all (the so-called \( O_h \) group).

• Translation invariance
  - Because of the translational symmetry of the crystal, any of its physical property, described for example by some function \( f(\mathbf{r}) \), must not change under translation:

\[
f(\mathbf{r} + \mathbf{R}_l) = f(\mathbf{r}),
\]

(55)

for any \( \mathbf{R}_l \).
  - The periodicity of \( f \) suggests that Fourier transforms may be useful:

\[
f(\mathbf{r}) = \sum_{\mathbf{K}_h} A_{\mathbf{K}_h} e^{i\mathbf{K}_h \cdot \mathbf{r}},
\]

(56)

where the Fourier coefficients \( A_{\mathbf{K}_h} \) are given by:

\[
A_{\mathbf{K}_h} = \frac{1}{\Omega} \int_{\Omega} f(\mathbf{r}) e^{-i\mathbf{K}_h \cdot \mathbf{r}} \, d\mathbf{r},
\]

(57)

The volume \( \Omega \) is the volume of the WS-cell and \( h \) is an integer labeling the wavevectors \( \mathbf{K} \).
– Now let’s use the periodicity:

\[ f(r + R_l) = \sum K_h A_{K_h} e^{iK_h \cdot (r + R_l)} = f(r) , \tag{58} \]

so we must have \( e^{iK_h \cdot R_l} = 1 \) or

\[ K_h \cdot R_l = 2\pi \text{ times an integer} \tag{59} \]

– Any vector of the form:

\[ K_h = h_1 A + h_2 B + h_3 C , \tag{60} \]

with

\[ A = \frac{2\pi}{\Omega} a \times b \quad B = \frac{2\pi}{\Omega} b \times c \quad C = \frac{2\pi}{\Omega} c \times a , \tag{61} \]

satisfies the conditions. The vectors \( A, B, \) and \( C \) are called reciprocal lattice vectors and they define the reciprocal lattice. They are usually labeled as \( G \)-vectors.

– The WS-cell of the reciprocal lattice is called the **Brillouin Zone (BZ)** of the lattice.

– fcc lattices have bcc reciprocal lattices (and vice-versa).

– **Bloch theorem:**

The wavefunction of electrons in the (periodic) crystal potential can be labeled by a vector-index \( k \) and satisfies the condition:

\[ \psi(k, r + R_l) = e^{ik \cdot R_l} \psi(k, r) , \tag{62} \]

which is equivalent to the condition:

\[ \psi(k, r) = e^{ik \cdot r} u_k(r) , \tag{63} \]

where \( u_k(r) \) is periodic:

\[ u_k(r + R_l) = u_k(r) . \tag{64} \]

Look at this expression as of a plane-wave (a free electron) modulating a periodic function describing the effect of the (periodic) ionic potential. However, since replacing \( k \) with \( k + G \) leaves the wavefunction unchanged, \( k \) is a **crystal momentum**, not necessarily directly related to the ‘real’ electron momentum.
Note that while \( \psi(\mathbf{r}) \) is not periodic, the only physically meaningful quantities we can derive from it are periodic. For example, the charge density, \( e|\psi(\mathbf{k}, \mathbf{r})|^2 \) is periodic:

\[
|\psi(\mathbf{k}, \mathbf{r} + \mathbf{R}_l)|^2 = \psi^*(\mathbf{k}, \mathbf{r} + \mathbf{R}_l) \psi(\mathbf{k}, \mathbf{r} + \mathbf{R}_l) = e^{-i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R}_l)} u^*_k(\mathbf{r} + \mathbf{R}_l) e^{i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R}_l)} u_k(\mathbf{r} + \mathbf{R}_l) = (65)
\]

\[
u^*_k(\mathbf{r}) u_k(\mathbf{r}) = e^{-i\mathbf{k} \cdot \mathbf{r}} u^*_k(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} u_k(\mathbf{r}) = |\psi(\mathbf{k}, \mathbf{r})|^2. \tag{66}
\]

**Bragg reflections:**

Let's discuss a little bit the meaning of the vector \( \mathbf{k} \). Let's Fourier transform the crystal potential:

\[
V_{lat}(\mathbf{r}) = \sum_G V_G e^{i\mathbf{G} \cdot \mathbf{r}}. \tag{67}
\]

Let's assume that a free electron (characterized by a plane wave of wavevector \( \mathbf{k} \)) enters the crystal and feels the lattice potential as a perturbation. The effect of this perturbation on the wavefunction \( |\mathbf{k} > = |\psi^*_k > \) above will be that of 'scattering' the wave to a new direction and wavelength, so to a new wavevector \( \mathbf{k}' \). The matrix element for this process is:

\[
<k'|V|k> = \sum_G V_G \delta_{\mathbf{k}-\mathbf{k}'-\mathbf{G}}, \tag{68}
\]

which vanishes unless \( \mathbf{k} + \mathbf{G} = \mathbf{k}' \). Assuming that the collision between electron and lattice is elastic (the recoil of the lattice, much much heavier, will be ignored), we can assume \( \mathbf{k} = \mathbf{k}' \), so that, squaring the equation above:

\[
G^2 = -2 \mathbf{k} \cdot \mathbf{G}, \tag{69}
\]

which is the condition for Bragg reflection. It also determines the edge of the BZ. Therefore, the electron wavevector \( \mathbf{k} \) can be treated as a 'quasi-momentum', but it is restricted to be inside the BZ: As soon as the electron wavelength approaches the BZ boundary, Bragg reflections will 'destroy' the wave.

- From the Bloch theorem and the condition for Bragg reflections, we also saw that waves with crystal momentum \( \hbar \mathbf{k} \) and \( \hbar \mathbf{k} + \hbar \mathbf{G} \) are equivalent. We can view this as due to the fact that the propagating wave
carries with it all the diffracted components $e^{iG \cdot r}$ as well. However, the energy of the wave will depend on which $G$ we are considering, as it will be seen when dealing with energy bands in solids.
Energy bands in crystals

• **Coupling atoms.**
  Read "intuitive approach" (example with Li atoms) in Colinge-Colinge text

• **The Kronig-Penney model.**
  1. Assume a periodic potential constituted by many wells of width $a$ and depth $U_0$ separated by barriers of width $b$. This is a ‘crystal’ with cell of size $a + b$.
     Let’s choose $x = -b$ and $x = a$ as cell boundaries, the well occupying the region $0 < x < a$, the barrier the region $-b < x < 0$.
  2. The Schrödinger equation we must solve takes the form:

\[
\frac{d^2 \psi_a}{dx^2} + \alpha^2 \psi_a = 0 \quad \text{for } 0 < x < a \tag{70}
\]

and

\[
\frac{d^2 \psi_b}{dx^2} + \beta^2 \psi_b = 0 \quad \text{for } -b < x < 0 \tag{71}
\]

with $\alpha = (2mE)^{1/2}/\hbar$ and

\[
\beta = \begin{cases} 
  i\beta_–; & \beta_– = [2m(U_0 - E)]^{1/2} \quad 0 < E < U_0 \\
  \beta_+; & \beta_+ = [2m(E - U_0)]^{1/2} \quad E > U_0 
\end{cases} \tag{72}
\]

3. It’s convenient to write the solutions as:

\[
\psi_a(x) = A_a \sin(\alpha x) + B_a \cos(\alpha x) \tag{73}
\]

\[
\psi_b(x) = A_b \sin(\beta x) + B_b \cos(\beta x) \tag{74}
\]

(later replacing $\sin$ and $\cos$ with $\sinh$ and $\cosh$ if necessary).
4. To find the coefficients $A_a$, $B_a$, $A_b$, and $B_b$ we have four boundary conditions:

\[
\begin{align*}
\text{continuity} & : \quad \left\{ \begin{array}{c}
\psi_a(0) = \psi_b(0) \\
\left. \frac{d\psi_a}{dx} \right|_0 = \left. \frac{d\psi_b}{dx} \right|_0
\end{array} \right. \\
\text{periodicity} & : \quad \left\{ \begin{array}{c}
\psi_a(a) = e^{ik(a+b)} \psi_b(-b) \\
\left. \frac{d\psi_a}{dx} \right|_a = e^{ik(a+b)} \left. \frac{d\psi_b}{dx} \right|_{-b}
\end{array} \right.
\]

which imply

\[
\begin{align*}
B_a &= B_b \\
\alpha A_a &= \beta A_b
\end{align*}
\]

\[
\begin{align*}
A_a \sin(\alpha a) + B_a \cos(\alpha a) &= e^{ik(a+b)} [-A_b \sin(\beta b) + B_b \cos(\beta b)] \\
\alpha A_a \cos(\alpha a) - \alpha B_a \sin(\alpha a) &= e^{ik(a+b)} [\beta A_b \cos(\beta b) + \beta B_b \sin(\beta b)]
\end{align*}
\]

Using the first two equations we can eliminate $A_b$ and $B_b$ from the last two:

\[
\begin{align*}
A_a [\sin(\alpha a) + (\alpha/\beta) e^{ik(a+b)} \sin(\beta b)] + B_a [\cos(\alpha a) - e^{ik(a+b)} \cos(\beta b)] &= 0 \\
A_a [\alpha \cos(\alpha a) - \alpha e^{ik(a+b)} \cos(\beta b)] + B_a [-\alpha \sin(\alpha a) - \beta e^{ik(a+b)} \sin(\beta b)] &= 0
\end{align*}
\]

5. We have reduced the problem to the solution of a linear homogeneous system in two unknowns. A non-zero solution exists only if the determinant of the matrix above vanishes. This simplifies to:

\[
-\frac{\alpha^2 + \beta^2}{2\alpha\beta} \sin(\alpha a) \sin(\beta b) + \cos(\alpha a) \cos(\beta b) = \cos[k(a+b)]
\]

6. This equation provides implicitly the ‘dispersion’ $E(k)$.

‘Energy gaps’ arise whenever the magnitude of the lhs exceeds unity.

The dispersion can be analyzed inside the BZ (i.e. for $-\pi/(a+b) < k \leq \pi/(a+b)$ since $a + b$ is the
fundamental translation ‘vector’ and \( G_n = \frac{2\pi n}{(a + b)} \) are the reciprocal lattice ‘vectors’. Alternatively, one can plot it in an ‘extended zone’ picture.

7. Some general considerations:
   1. The dispersion is quasi-parabolic near the zone center (electron- or hole-like).
   2. The dispersion is ‘flat’ (zero slope) at the zone-edge.
   3. The dispersion is more and more free-electron like at higher energies (why?).

• General formulation.
   – We want to solve the Schrödinger equation

\[
-\frac{\hbar^2 \nabla^2}{2m} \psi(r) + V_{\text{lat}}(r) \psi(r) = E \psi(r).
\] (84)

   – Since the lattice potential \( V_{\text{lat}}(r) \) is periodic, we use Bloch theorem and consider the wavefunction

\[
\psi(k, r) = e^{i k \cdot r} \sum_G u_{kG} e^{i G \cdot r}
\] (85)

   – Inserting into the Schrödinger equation:

\[
\frac{\hbar^2}{2m} \sum_G |k + G|^2 u_{kG} e^{i(k+G) \cdot r} + V_{\text{lat}}(r) \sum_G u_{kG} e^{i(k+G) \cdot r} = E(k) \sum_G u_{kG} e^{i(k+G) \cdot r}
\] (86)

   – Now use a ‘standard technique’ to solve equations in Fourier-transformed form: Multiply by \( e^{-i(k+G') \cdot r} \) and integrate over the whole volume of the crystal:

\[
\left[ \frac{\hbar^2}{2m} |k + G|^2 - E(k) \right] u_{kG} + \sum_{G'} V_{GG'} u_{kG'} = 0.
\] (87)
This linear homogeneous problem has nontrivial solutions only if the determinant of the coefficients vanishes:

\[
\text{Det} \left| \sum_{G'} \left\{ \left[ \frac{\hbar^2}{2m} |k + G|^2 - E(k) \right] \delta_{GG'} + V_{GG'} \right\} \right| = 0 .
\]  

(88)

- The term \( V_{GG'} \) can be simplified using its periodicity:

\[
V_{GG'} = \frac{1}{V} \int_V e^{-iG \cdot r} V_{\text{lat}}(r) e^{iG' \cdot r} \, dr ,
\]

assuming the wavefunctions have been properly normalized to the volume \( V \) of the crystal.

Now \( V_{\text{lat}} \) is the sum of the ionic potentials in the WS cell. If the ions in the WS cell are at \( \tau_\alpha (\alpha = 1, N_{\text{ions}}) \), then:

\[
V_{\text{lat}}(r) = \sum_{l,\alpha} \omega(r - R_l - \tau_\alpha) .
\]

(90)

Then, putting \( G'' = G - G' \):

\[
V_{GG'} = \frac{1}{V} \sum_{l,\alpha} \int_V e^{-iG'' \cdot r} \omega(r - R_l - \tau_\alpha) \, dr = \frac{1}{V} \sum_{l,\alpha} e^{-iG'' \cdot R_l} \int_V \omega(r - R_l - \tau_\alpha) e^{-iG'' \cdot (r - R_l)} \, dr .
\]

(91)

Now let’s put \( r' = r - R_l - \tau_\alpha \) as dummy integration variable:

\[
V_{GG'} = \frac{1}{V} \sum_{l,\alpha} e^{-iG'' \cdot R_l} \int_V \omega(r') e^{-iG'' \cdot (r' + \tau_\alpha)} \, dr' =
\]

\[
\sum_{\alpha} e^{-iG'' \cdot \tau_\alpha} \frac{1}{N \Omega} \sum_l e^{-iG'' \cdot R_l} \int_V \omega(r') e^{-iG'' \cdot r'} \, dr' .
\]

(93)
Since the ionic potential $\omega(r)$ is short-range, it decays very quickly at large distances and we can neglect the contribution to the integral coming from points $r'$ outside the WS cell:

$$\int_V \omega(r') e^{-iG'' \cdot r'} \, dr' \approx \int_{\Omega} \omega(r') e^{-iG'' \cdot r'} \, dr'.$$

(94)

Since $\sum_{l,\alpha} e^{-iG'' \cdot R_l} = N$, the number of cells in the volume $V$, we finally get:

$$V_{GG'} = \sum_{\alpha} e^{-i(G-G') \cdot \tau_{\alpha}} \frac{1}{\Omega} \int_{\Omega} \omega(r) e^{-i(G-G') \cdot r} \, dr = S(G-G') \omega_{G-G'}.$$  (95)

The factor $S(G)$ depends only on the location of the ions within the WS cell and it’s called ‘structure factor’. The factor $\omega_G$ is the Fourier transform of the atomic potential within the WS cell and it’s called ‘form factor’.

- A numerical solution of this problem can be obtained by considering a large-enough number, $N_G$, of $G$-vectors. The $N_G$ eigenvalues, $E_n(k)$ give the dispersion in $N_G$ bands, the $N_G$ eigenvectors the wavefunctions.
- The problem is: What is $V_{lat}$ (or, equivalently, $\omega(r)$)? Even for the 2-electron He atom, the self-consistent potential (modified by the electron charge which is modified by the potential...) requires approximations (Hartree-Fock).

Also, numerically one needs as many as $10^6$ $G$-vectors to account for the core of the ions.
A practical approach: **Empirical pseudopotentials.**

- The full ionic potential is too complicated. But do we need it all? After all, the core electrons are (to a large extent) unaffected by the presence of the crystal. And we care only about the bonding states describing the valence electrons and the excited states just ‘above’ (in energy) them.

- Since the core electrons feel the short-distance (i.e., short wavelength, large $G$) features of the ionic potential near the core, we can consider $\omega_G$ only for the first, smallest $G$ vectors. This will represent correctly the potential and the wavefunctions outside the ionic cores.

This approximated potential is called **pseudopotential** and the wavefunctions **pseudo-wavefunctions**. In principle, we should find form factors such that the wavefunction and its radial derivative at the core interface matches the ‘exact’ wavefunction and its derivative. In practice, one can treat the form factors $\omega_G$ as fitting parameters, calibrated to experimental data about the band structure.

- For semiconductors of the diamond structure (as Si, Ge, C) symmetry considerations imply that $\omega_G$ depends only on the magnitude $G$ of $G$. Only $G = \sqrt{3}$, $\sqrt{8}$, and $\sqrt{11}$ are required to obtain excellent fits to experimental data. For ionic compound semiconductors of the group-III and group-V elements (GaAs, InAs, GaP, etc.) require also $\omega_{\sqrt{4}}$ and the presence of two different ions in the cell renders the form factors complex numbers, so that 8 real numbers are required.

**Equations of motion for electrons in a crystal.**

- How do electrons move in these complicated energy bands when we add an external potential $V_{\text{ext}}(r,t)$ to the lattice potential?

- **Envelope approximation:** By the Bloch theorem, we know that

$$
\psi(k, r) = e^{i k \cdot r} \sum_G u_{kG} e^{i G \cdot r} \tag{96}
$$

It can be shown that if we stick to one band only, with dispersion $E(k)$, if the external potential is not too strong, and if we ignore changes to the Bloch component $u(r) = \sum_G u_{kG} e^{i G \cdot r}$, the wavefunction in a
small-enough neighborhood of $k_0$ can be approximated as

$$
\psi(k, r) = \psi(r) \sum_G u_{k_0 G} e^{iG \cdot r}
$$

(97)

and the ‘envelope’ $\psi(r)$ satisfies the wave equation

$$
[E(-i\nabla) + eV_{ext}(r)] \psi(r) = E \psi(r) .
$$

(98)

– Semiconductors are characterized by full valence bands, empty conduction bands, and a relatively small energy gap between the valence and conduction bands. ‘Doping’ results in the introduction of excess electron occupying the bottom of the conduction band, or the elimination of some electrons from the valence bands (which is viewed as the introduction of holes occupying the top of the valence band). Thus, the ‘action’ happens at the top of the valence band and/or at the bottom of the conduction band. At these extrema of the bands we can expand the dispersion as:

$$
E(k) = E(k_0) + \frac{1}{2}(k - k_0) \cdot \nabla_k^2 E(k) \biggl|_{k_0} \cdot (k - k_0) + ...
$$

(99)

(the first-order term in $\nabla E$ vanishes since we are at an extremum of the band). Thus, replacing the ‘momentum’ (measured from the band extremum $k_0$)

$$
k - k_0 \rightarrow -i\nabla
$$

(100)

thanks to the envelope approximation,

$$
E(-i\nabla) = E(k_0) + \frac{1}{2}(-i\nabla) \cdot \nabla_k^2 E(k) \biggl|_{k_0} \cdot (-i\nabla) + ... = E(k_0) - \frac{1}{2} \sum_{ij} \frac{\partial E}{\partial k_i \partial k_j} \biggl|_{k_0} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + ...
$$

(101)
- **Effective mass**: Let’s set $E(k_0) = 0$ by a suitable choice of the reference energy and let’s measure $k$ from $k_0$. Let’s ignore higher-order terms in the expansion above. Let’s define the inverse ‘effective mass’ tensor $1/m_{ij}^*$ via:

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \frac{\partial E}{\partial k_i \partial k_j} \bigg|_{k_0} \quad (102)$$

Then

$$E(k) \approx \sum_{ij} \frac{\hbar^2 k_i k_j}{2m_{ij}^*} \quad (103)$$

and the envelope wave equation becomes:

$$\left[ -\sum_{ij} \frac{\hbar^2}{2m_{ij}^*} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + eV_{ext}(r) \right] \psi(r) = E \psi(r), \quad (104)$$

which is beginning to look like a ‘conventional’ Schrödinger equation.

For GaAs, the effective mass in the CB is actually a scalar (since the first CB is formed by $s$-type orbitals), $m^* \approx 0.063 \, m_0$. For Si, the first CB has ellipsoidal equienergy surfaces, so that the effective mass tensor is diagonal along the [100] crystal direction:

$$m = \begin{pmatrix} m_L & 0 & 0 \\ 0 & m_T & 0 \\ 0 & 0 & m_T \end{pmatrix}, \quad (105)$$

where $m_L \approx 0.91 \, m_0$ and $m_T \approx 0.19 \, m_0$

- By the ‘correspondence principle’ of Quantum Mechanics we know that the center of motion of a wave packet can be obtained from classical mechanics by considering the quantum Hamiltonian and rendering ‘classical’ by replacing $-\hbar \nabla$ with the momentum $\hbar \mathbf{k}$. Thus, we can consider the classical Hamiltonian

$$H(p, r) = E(p/\hbar) + eV_{ext}(r). \quad (106)$$
From the Hamilton equations of motion:

\[ \mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{1}{\hbar} \nabla_k E(k), \]  

(107)

and

\[ \frac{dk}{dt} = \frac{e}{\hbar} \nabla V_{ext}(\mathbf{r}) = -e \mathbf{F}. \]  

(108)

– **Note**: We have been able to eliminate completely a very complicated potential, \( V_{lat} \), simply by replacing the free electron mass \( m_0 \) with an effective mass tensor. This is an extremely powerful results, but its limitations should be kept in mind.

– For **holes** the picture is the same: The top of the VB has a ‘negative’ effective mass. However, talking about ‘missing electrons’ (‘holes’) is more convenient and this is accomplished by considering these fictitious particles as positively charged (that is, moving in the opposite directions as electrons under an external field), and by ‘flipping’ all signs in the equations of motion, thus recovering a positive mass. Unfortunately, the structure of the valence bands is more complicated: The parabolic approximation breaks down very early, the effective mass is highly anisotropic, and three bands are energetically very close (degenerate at the center of the BZ, spin-orbit coupling reducing the degeneracy to two bands). Thus, one needs to consider always at two bands (so called ‘heavy’ and ‘light’ hole bands), often three (the ‘split-off’ band).
Doping.

- The technological importance of semiconductors stems from the fact that we can make them decent insulators or excellent conductors by adding ‘free carriers’ (electrons in the CB, holes in VB).
- This is achieved by introducing ‘substitutional’ impurities. In Si (a group IV element, with 4 electrons in the bonding \( sp^3 \) valence orbitals) the substitution of a group III element (‘acceptors’ such as B, Ga, In, having only 3 valence electrons) causes the absence of one electron in the valence band for each acceptor atom. Si so doped is said to be \( p \)-type, since current will be carried by holes, which can be viewed as positively charged ‘particles’. By ‘accepting’ an extra electron (that is, giving away the ‘hole’), acceptors become negatively charged ionic impurities. The substitution of a group V element (‘donors’ such as P, As, Sb, having 5 valence electrons) causes the presence of one extra electron in the CB for each donor impurity. Si so doped is said to be \( n \)-type, since negatively charged electrons carry the electric current. By ‘donating’ an extra electron acceptors become positively charged ionic impurities.
- The external potential caused by the substitutional ionized impurity is a ‘screened’ Coulomb potential: It is screened by:
  1. the valence electrons and by some lattice rearrangement, which is accounted for by the macroscopic dielectric constant \( \epsilon_s \) of the semiconductor
  2. the free carrier (holes in the VB, electrons in the CB)

Ignoring for now the second screening effect, the potential will be

\[
V_{dop}(r) = -\frac{e^2}{4\pi\epsilon_s r}
\]  

(109)

Thanks to the envelope approximation, we can treat this potential as an hydrogen-like problem. Assuming a scalar effective mass, the energy levels (measured from the band extremum) will be:

\[
E_n = -\frac{m^* e^4}{32\pi^3 n^2 h^2 \epsilon_s^2}.
\]  

(110)
Note how the magnitude of these energies is reduced by a factor $m^*\epsilon_0^2/(m_0\epsilon_s^2)$ with respect to the hydrogen atom. Since for most semiconductors $m^*/m_0 < 1$ and $\epsilon_0/\epsilon_s \sim 0.1$, the ionization energy is very small. Indeed dopants are almost always ionized. ‘Freezout’ (that is, the condition in which free carriers are trapped into impurity states and the semiconductor actually becomes an insulator) occurs at very low temperatures. Similarly, the Bohr radius of the $n$-th excited state,

$$a_n = -\frac{4\pi n^2 \hbar^2 \epsilon_s}{m^*e^2}, \quad (111)$$

is much larger than that of the H atom, being typically of the order of tens of nanometers. This justifies the use of the macroscopic dielectric constant.

• **Density of States.**
  - A free particle in a 3D box.

Let’s write the Schrödinger equation for a free electron in a cubic box of side $L$:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + k^2 \psi = 0, \quad (112)$$

with $k^2 = 2mE/\hbar^2$.

Let’s separate the problem: We look for a solution of the form $\psi(x, y, z) = \psi_x(x)\psi_y(y)\psi_z(z)$. Substituting this into the Schrödinger equation and diving by $\psi$, one gets:

$$\frac{1}{\psi_x} \frac{d^2 \psi_x}{dx^2} + \frac{1}{\psi_y} \frac{d^2 \psi_y}{dy^2} + \frac{1}{\psi_z} \frac{d^2 \psi_z}{dz^2} + k^2 = 0. \quad (113)$$

Since this is true for any $x, y, z$, we must have

$$\frac{1}{\psi_x} \frac{d^2 \psi_x}{dx^2} = \text{some constant} = -k_x^2. \quad (114)$$
Thus, accounting for the boundary condition (wavefunction must vanish at the wall of the cube):

$$\psi(x,y,z) = A \sin(k_xx) \sin(k_yy) \sin(k_zz)$$  \hspace{1cm} (115)

$$k_x^2 + k_y^2 + k_z^2 = k^2$$  \hspace{1cm} (116)

$$k_x = \frac{n_x \pi}{L}, \quad k_y = \frac{n_y \pi}{L}, \quad k_z = \frac{n_z \pi}{L},$$  \hspace{1cm} (117)

with $n_x, n_y, n_z = \pm 1, \pm 2, \pm 3, \ldots$

The number of solutions per unit volume of $k$-space will be (obviously?) $L^3 / \pi^3$. But since solutions differing only for an overall sign are identical and there are 8 possible permutations of this type, the number of allowed $k$-states per unit volume in $k$-space will be 8 times smaller. Finally, accounting for the electronic spin (ignored so far) the number of states will be twice as large. Thus, the density of $k$-states will be $L^3 / 4 \pi^3$.

- **Density of states at a given energy.**

  The density of $k$-states with energy between $E$ and $E + dE$ can be obtained by multiplying the density in $k$-space by the volume of the spherical shell with wavevector between $k$ and $k + dk$, that is

  $$4\pi k^2 \frac{L^3}{4\pi^3} \cdot dk.$$  \hspace{1cm} (118)

(Recall Jean’s law $n(\nu) \propto \nu^2$? Since for photons $2\pi \nu = ck$, where $c$ is the speed of light, the expression above is just Jean’s law for the black-body spectrum). Now use $E(k) = \hbar^2 k^2 / (2m^*)$, change variable and obtain the number of states per unit volume (recall that $L^3$ is the volume of the crystal) with energy between $E$ and $E + dE$, $\rho(E)$, as:

$$\rho(E) \ dE = \frac{m^*(2m^*E)^{1/2}}{\hbar^3 \pi^2} \ dE,$$  \hspace{1cm} (119)

where $E$ is measured from the band extremum.

- **A useful trick.**

  Suppose we must perform a sum over all $k$-states in $k$-space. Since the states are very dense, it makes sense to replace the sum with an integral (and it also makes our life easier, since integrals are in general easier
to do that sums). This can be done by noticing that summing over $k$-points in a given volume of $k$-space is equivalent to integrating over the same volume and dividing by the volume of each $k$-point, which is the inverse of the density. Thus

$$\sum_k \rightarrow 2 \frac{L^3}{(2\pi)^3} \int dk \text{ or } \frac{1}{V} \sum_k \rightarrow \frac{2}{(2\pi)^3} \int dk$$  \hspace{1cm} (120)

Another way to get the density of states (DOS). One can show that the density of states at energy $E$ can be obtained as:

$$\rho(E) = \frac{2}{(2\pi)^3} \int \delta[E - E(k)] \, dk$$  \hspace{1cm} (121)

keeping in mind the following general property of the delta-function:

If $x_i$ (for $i = 1, \ldots, N$) are the zeros of the function $f(x)$, then:

$$\int \delta[f(x)] g(x) \, dx = \sum_i \int \frac{1}{|f'(x_i)|} \delta(x - x_i) g(x) \, dx = \sum_i \frac{g(x_i)}{|f'(x_i)|}$$  \hspace{1cm} (122)

where we have used the notation $f'(x) = df(x)/dx$. This makes it easier to calculate the DOS for arbitrary dispersions $E(k)$.

**DOS for the Si CB:**

Consider the ellipsoid

$$E_{100}(k) = \frac{\hbar^2}{2} \left[ \frac{k_x^2}{m_L} + \frac{k_y^2 + k_z^2}{m_T} \right]$$  \hspace{1cm} (123)

with the longitudinal axis along the $k_x$ axis. Using the Herring-Vogt transformation

$$k_x^* = \left( \frac{m_0}{m_L} \right)^{1/2} k_x, \quad k_y^* = \left( \frac{m_0}{m_T} \right)^{1/2} k_y, \quad k_z^* = \left( \frac{m_0}{m_T} \right)^{1/2} k_z$$  \hspace{1cm} (124)
one can show that the DOS in the CB of Si is
\[ \rho(E) = \frac{m_d(2m_dE)^{1/2}}{\hbar^3\pi^{3/2}}. \] (125)

where the ‘density of states effective mass’ is \( m_d = (m_Lm_T^2)^{1/3} \).

- **The Fermi-Dirac distribution function.**
  - Let’s consider how a system of \( N_e \) ‘indistinguishable’ electrons at thermal equilibrium with the environment will be distributed over \( N \) energy levels, each with energy \( E_i \) and containing \( g_i \) (\( i = 1, \ldots, N \)) available states, by placing \( n_i \) electrons in each energy level \( i \). The total energy of the system is \( E_{tot} = \sum_{i=1}^{N} E_in_i \).
  - The number \( g_i \) is called the ‘degeneracy’ of the level \( i \).
  - Pauli’s principle tells us that we cannot put more than one electron in each state, so \( n_i \leq g_i \).
  - The number \( W_i \) of different ways in which \( n_i \) electrons can be placed into the \( g_i \) available states in the \( i^{th} \) level is:
    \[ W_i = \frac{g_i!}{(g_i - n_i)!n_i!}. \] (126)
    (Think of \( g_i \) bins in which we can put a ball or leave it empty, having \( n_i \) balls in total. We can put the first ball in any of the \( g_i \) bins, the second ball in any of the remaining \( g_i - 1 \) bins, etc., for a total of \( g_i! \) possible ways. But, since neither the balls nor the ‘voids’ can be distinguished, we must divide this number by the \( n_i! \) equivalent ways we can place the \( n_i \) balls into the bins and by the \( (g_i - n_i)! \) equivalent ways we can arrange the \( g_i - n_i \) voids.)
  - Thus, the total number of possible arrangements is just
    \[ W = \prod_i W_i = \prod_i \frac{g_i!}{(g_i - n_i)!n_i!}. \] (127)
  - The most likely distribution should be determined by maximizing \( W \) under the constraints of energy and particle conservation. It’s convenient to maximize \( \ln W \) instead. Let’s write
    \[ \ln W = \sum_i [\ln g_i! - \ln(g_i - n_i)! - \ln n_i!]. \] (128)
Let's use Stirling's approximation, valid for large $x$:

$$\ln x! \approx x \ln x - x,$$  \hspace{1cm} (129)

and write

$$\ln W \approx \sum_i \left[ g_i \ln g_i - g_i - (g_i - n_i) \ln(g_i - n_i) + (g_i - n_i) - n_i \ln n_i + n_i \right] =$$  \hspace{1cm} (130)

$$= \sum_i \left[ g_i \ln g_i - (g_i - n_i) \ln(g_i - n_i) - n_i \ln n_i \right].$$  \hspace{1cm} (131)

Then

$$d(\ln W) = \sum_i \frac{\partial \ln W}{\partial n_i} dn_i \approx \sum_i \left[ \ln(g_i - n_i) + 1 - \ln n_i - 1 \right] dn_i = \sum_i \ln(g_i/n_i - 1) \, dn_i.$$  \hspace{1cm} (132)

Our solution now comes from solving $d(\ln W) = 0$ subject to the conditions $\sum_i dn_i = 0$ (the total number of particles cannot change) and $\sum_i E_i dn_i = 0$ (the total energy should not change). This requires the use of the ‘method of Lagrange multipliers’: We multiply each equation describing a constraint by a constant (a ‘multiplier’) and add these equations to the original equation:

$$\sum_i \left[ \ln(g_i/n_i - 1) - \alpha - \beta E_i \right] dn_i = 0.$$  \hspace{1cm} (133)

Since this must be valid for each $i$, a solution will be given by:

$$\ln(g_i/n_i - 1) - \alpha - \beta E_i = 0$$  \hspace{1cm} for all $i$.  \hspace{1cm} (134)

Therefore, the density of electrons in the $i^{th}$ energy level $E_i$ will be:

$$f(E_i) = \frac{n_i}{g_i} = \frac{1}{1 + e^{\alpha + \beta E_i}}.$$  \hspace{1cm} (135)
For closely spaced levels we can go to the continuum. Moreover, more complex arguments of statistical mechanics imply \( \alpha = -E_F/(k_B T) \) and \( \beta = 1/(k_B T) \), where \( T \) is the temperature, \( k_B \) is Boltzmann constant, and \( E_F \) the Fermi energy. Thus:

\[
f(E) = \frac{1}{1 + e^{(E-E_F)/(k_B T)}}.
\] (136)

Here we should recall that the Fermi energy (or ‘Fermi level’, or the ‘chemical potential’ of the grand-canonical ensemble), the total electron density, and the temperature are related, as we’ll see below.

- Some useful approximations to the Fermi-Dirac distribution are the following:

\[
f(E) \approx \begin{cases} 
\theta(E - E_F) & \text{for } k_B T \ll E_F \\
 e^{(E-E_F)/(k_B T)} & \text{for } k_B T \gg E_F \text{ (Maxwell-Boltzmann distribution)} 
\end{cases}
\] , (137)

where \( \theta(x) \) is the Heavyside step function, \( \theta(x) = 1 \) for \( x < 0 \), \( \theta(x) = 0 \) otherwise.
• The electron (hole) density in the CB (VB).
  
  The total density of free electrons in the lowest-energy CB of a semiconductor will be given by:

\[
n = \frac{1}{V} \sum_k f[E(k)] = \frac{2}{(2\pi)^3} \int f[E(k)] \, dk.
\]

Going to spherical coordinates in \( k \)-space:

\[
n = \frac{1}{4\pi^3} \int_0^{2\pi} d\phi \int_0^\pi d\theta \, \sin \theta \int_0^\infty dk \, k^2 f[E(k)] = \frac{1}{\pi^2} \int_0^\infty dk \, k^2 f[E(k)].
\]

For an isotropic effective mass \( m^* \), \( E(k) = \hbar^2 k^2 / (2m^*) \), so that \( k = (2m^*E)^{1/2} / \hbar \) and \( dE = (\hbar^2/m^*)kdk \). So:

\[
n = \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \int_0^\infty dE \, E^{1/2} f(E) = \int_0^\infty dE \rho(E) f(E),
\]

recalling the definition of the DOS at energy \( E \) we have obtained earlier.

Note that the upper integration limit extends to \( \infty \). Of course, this is incorrect, strictly speaking, since we are limiting ourselves to the low-energy region of the lowest-energy CB. Yet, the Fermi-Dirac distribution decays so quickly (exponentially) at high energies, that the error we make is negligible.

Three limits are noteworthy:

1. **Low-temperature, degenerate limit.** At low \( T \), recalling that \( F(E) \) is unity for \( E < E_F \) but it vanishes for \( E > E_F \), we have the degenerate limit:

\[
n_{\text{low } T} \approx \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \int_0^{E_F} dE \, E^{1/2} = \frac{1}{3\pi^2} \left( \frac{2m^*E_F}{\hbar^2} \right)^{3/2} = \frac{k_F^3}{3\pi^2},
\]

where \( k_F = (2m^*E_F)^{1/2} / \hbar \) is the Fermi wavevector.
2. **High-temperature, non-degenerate limit.** On the contrary, at high temperature we have the non-degenerate limit \( f(E) \approx e^{-(E-E_F)/(k_B T)} \) and

\[
n_{\text{high } T} \approx \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} e^{E_F/(k_B T)} \int_0^\infty dE \ E^{1/2} e^{-E/(k_B T)} = \frac{1}{2\pi^2} \left( \frac{2m^* k_B T}{\hbar^2} \right)^{3/2} e^{E_F/(k_B T)} \int_0^\infty dx \ x^{1/2} e^{-x},
\]

having set \( x = E/(k_B T) \) in the last step. Now we should recall the definition of the ‘Gamma’ function:

\[
\Gamma(s) = \int_0^\infty dx \ x^{s-1} e^{-x}.
\]

Knowing that

\[
\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(s + 1) = s \Gamma(s),
\]

we get:

\[
n_{\text{high } T} \approx \frac{1}{4} \left( \frac{2m^* k_B T}{\pi \hbar^2} \right)^{3/2} e^{E_F/(k_B T)} = N_c e^{-(E_C-E_F)/(k_B T)}.
\]

In the case of the ‘ellipsoidal’ CB of Si we should replace the isotropic mass \( m^* \) with \( m_d = 6^{2/3}(m_L m_T^{2})^{1/3} \), the factor of 6 stemming from the 6-fold degeneracy of the CB minima. For holes we get, similarly,

\[
p_{\text{high } T} \approx \frac{1}{4} \left( \frac{2m^* k_B T}{\pi \hbar^2} \right)^{3/2} e^{-(E_G-E_F)/k_B T} = N_v e^{-(E_F-E_v)/k_B T},
\]

so that the product \( p n \) at equilibrium in the non-degenerate limit is

\[
pn = \frac{1}{2} \left( \frac{k_B T}{\pi \hbar^2} \right)^3 (m_d m^*)^{3/2} e^{-E_G/(k_B T)} = n_i^2,
\]
the square of the 'intrinsic carrier concentration', \( n_i \approx 1.4 \times 10^{10} \text{ cm}^{-3} \) in Si.

Other useful forms are:

\[
n \approx n_i e^{(E_F - E_i)/(k_BT)} , \quad p \approx n_i e^{(E_i - E_F)/(k_BT)} ,
\]

where \( E_i \) is the Fermi level in undoped (intrinsic) Si, which we shall derive shortly.

3. **General case.** In general, we have:

\[
n = \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \int_0^\infty dE \frac{E^{1/2}}{E(\mathbf{k})} f[E(\mathbf{k})] = \frac{1}{4} \left( \frac{2m^*k_BT}{\pi\hbar^2} \right)^{3/2} 2\sqrt{\pi} \int_0^\infty \frac{x^{1/2}}{1 + e^{x-\eta}} \, dx ,
\]

where, as before, we have changed the integration variable to \( x = E/(k_BT) \), we have set \( \eta = E_F/(k_BT) \), and we have multiplied and divided by \( 2/\sqrt{\pi} \) to obtain the same prefactor \( N_c \) defined above. The integral cannot be solved in closed form, but it is a well-know integral (appropriately called Fermi-Dirac integral of order 1/2 and labeled by \( F_{1/2}(\eta) \)). So, the electron density is usually written in terms of this integral as:

\[
n = \frac{1}{4} \left( \frac{2m^*k_BT}{\pi\hbar^2} \right)^{3/2} F_{1/2}(\eta) \approx N_c F_{1/2}(\eta) .
\]

– Fermi level in intrinsic Si: Set charge neutrality condition:

\[
n = p \rightarrow N_c e^{-(E_c-E_F)/(k_BT)} = N_v e^{-(E_F+E_v)/(k_BT)} ,
\]

so

\[
E_F = E_i = E_{mg} - \frac{k_BT}{2} \ln \left( \frac{N_c}{N_v} \right) ,
\]

which can be safely approximated with the 'mid-gap' \( E_{mg} = (E_v + E_c)/2 \).

– Fermi level in extrinsic (doped) Si: Set charge neutrality condition:

\[
n = p + N_{d^+} ,
\]
where $N_d^+$ is the density of ionized donors. But the probability that an electron will occupy the energy level $E_d$ associated to the H-like impurity potential will be

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

(155)

where $g_d$ is a degeneracy factor dependent on the microscopic structure of the impurity, typically taken as 1/2 since only one electron can occupy the impurity state while we have derived the Fermi function assuming 2 electrons with opposite spins per state. Therefore: $N_d^+ = N_d [1 - f(E_d)] \approx N_d$ and:

\[
n \approx p + N_d \rightarrow E_F \approx E_c - k_B T \ln \left( \frac{N_c}{N_d} \right) = E_i + k_B T \ln \left( \frac{N_d}{n_i} \right) .
\]

(156)
When we'll deal with scattering, we shall have to face the complication that an external perturbation will cause not only a perturbation to the single electron we are interested in, but also a global rearrangement of all other electrons in the crystal. This rearrangement, in turn, modifies the original perturbation. Therefore the total perturbing potential will be the sum of the original perturbation plus a ‘polarization’ potential due to the rearrangement of the charges.

Let’s assume that $\psi_{k,\mu}(\mathbf{r})$ ($\mu$ is a band index) are the wavefunction of the unperturbed crystal and suppose that we apply to a crystal an external perturbing potential dependent on both time and space. We can always Fourier transform such a function (both with respect to $\mathbf{r}$ as well as $t$), so we consider just one Fourier component (we will be able to add all components at the end):

$$V_{\text{ext}}(\mathbf{r}, t) = V_{\text{ext}}(\mathbf{q}, \omega) e^{i \mathbf{q} \cdot \mathbf{r}} e^{-i \omega t}. \quad (157)$$

Similarly, we shall consider the Fourier transform $V_{\text{pol}}(\mathbf{q}, \omega)$ of the potential due to the polarization charge $\delta \rho(\mathbf{q}, \omega)$. Our goal is to find the ‘dielectric function’ $\epsilon(\mathbf{q}, \omega)$ such that:

$$V_{\text{ext}}(\mathbf{q}, \omega) = V_{\text{tot}}(\mathbf{q}, \omega) - V_{\text{pol}}(\mathbf{q}, \omega) = \epsilon(\mathbf{q}, \omega) V_{\text{tot}}(\mathbf{q}, \omega), \quad (158)$$

or $V_{\text{tot}} = V_{\text{ext}}/\epsilon$ and so $\epsilon/\epsilon_0 \sim 1 - V_{\text{pol}}/V_{\text{tot}}$.

Now, in the ‘linear response approximation’ of a small external potential, the potential $V_{\text{ext}}$ will act as a perturbation to the wavefunctions. Let’s write the perturbed wavefunctions $\phi_{k,\mu}$ in terms of the unperturbed wavefunctions. Since we have only one perturbing wavevector $\mathbf{q}$ and we shall restrict our analysis to 2 bands (the lowest-energy conduction and the highest-energy valence bands), ignoring the possibility that many different wavevectors may interact (this is called the ‘Random Phase Approximation’ and it relies on the assumption that the phases of perturbations of different wavelengths will interfere destructively), we can write:

$$\phi_{k,\mu}(\mathbf{r}, t) = \psi_{k,\mu}(\mathbf{r}, t) + \delta \psi_{k,\mu}(\mathbf{r}, t). \quad (159)$$
The change $\delta \psi_{k,\mu}$ can be obtained from time-dependent perturbation theory. Assuming (for mathematical convenience) that the external potential has been turned on slowly with a ‘switching-on factor’ $e^{\eta t}$, this change can be expressed as:

$$\delta \psi_{k,\mu}(r, t) = -e \sum_{q\lambda} \frac{\langle k + q, \lambda|V_{tot}|k, \mu \rangle}{E_{\mu}(k) - E_{\lambda}(k + q) + \hbar \omega - i\hbar \eta} \psi_{k+q,\lambda}(r, t). \quad (160)$$

- The perturbed charge density, to first order, will be given by summing the contributions from all occupied states $k$ (described by an occupation number $p(k, \mu)$):

$$\delta \rho(r, t) = e^2 \sum_{k\mu} p(k, \mu) \left[ |\phi_{k,\mu}(r, t)|^2 - |\phi_{k,\mu}(r, t) - \delta \psi_{k,\mu}(r, t)|^2 \right] \quad (161)$$

or:

$$\delta \rho(r, t) = e^2 \sum_{k\mu} p(k, \mu) \sum_{qG\lambda} \psi_{k,\mu}^*(r, t) \frac{\langle k + q|V_{tot}|k \rangle \langle k + q + G, \lambda|e^{i\mathbf{q} \cdot \mathbf{r}}|k, \mu \rangle}{E_{\mu}(k) - E_{\lambda}(k + q) + \hbar \omega - i\hbar \eta} \psi_{k+q,\lambda}(r, t) + \cdots \quad (162)$$

Since $\langle k + q|V_{tot}|k \rangle = V_q \sum_G \langle k + q + G, \lambda|e^{i\mathbf{q} \cdot \mathbf{r}}|k, \mu \rangle$,

$$\delta \rho(r, t) = e^2 \sum_{qkG\mu\lambda} \frac{[p(k, \mu) - p(k + q, \lambda)] |\phi_{k+q+G,\lambda}(r, t)|^2}{E_{\mu}(k) - E_{\lambda}(k + q) + \hbar \omega - i\hbar \eta} V_q e^{-i\mathbf{q} \cdot \mathbf{r}}. \quad (163)$$

We see from this equation that, if we consider the Fourier transform

$$\delta \rho(r, t) = \sum_{\mathbf{q}} \delta \rho_{\mathbf{q}} e^{-i\mathbf{q} \cdot \mathbf{r} - i\omega t}, \quad (164)$$
we can identify the Fourier coefficients $\delta \rho_q$ with

$$\delta \rho_q = e^2 \sum_{kG_{\mu\lambda}} \frac{[p(k, \mu) - p(k + q, \lambda)] | < k + q + G, \lambda | e^{iq \cdot r} | k, \mu >|^2}{E_\mu(k) - E_\lambda(k + q) + \hbar \omega - i\hbar \eta} V_q.$$  \hfill (165)

- Now, from Poisson equation:

$$\nabla^2 V_{pol} = -\frac{\delta \rho}{\epsilon_0}$$ \hfill (166)

going to Fourier transforms we have

$$V_{polq} = \frac{\delta \rho_q}{q^2 \epsilon_0},$$ \hfill (167)

so that, finally, we see that

$$\epsilon(q, \omega) = \epsilon_0 \left\{ 1 - \frac{e^2}{q^2 \epsilon_0} \sum_{kG_{\mu\lambda}} \frac{[p(k, \mu) - p(k + q, \lambda)] | < k + q + G, \lambda | e^{iq \cdot r} | k, \mu >|^2}{E_\mu(k) - E_\lambda(k + q) + \hbar \omega - i\hbar \eta} \right\}.$$ \hfill (168)

- Some very useful limiting cases:
  1. **Static dielectric constant.**
     
     Let’s consider the crystal near its ground state near zero temperature. The occupation of a state $k$ in band $\mu$ will be the equilibrium (Fermi-Dirac) occupation at $T = 0$ plus a small perturbation representing free carriers:

     $$p(k, \mu) = f_{FD}(k, \mu) + f(k, \mu).$$ \hfill (169)

     Let’s consider now only the effect of the zero-$T$, equilibrium populations in the valence and conduction bands. Free carriers (which we are mostly interested on) will be considered separately below. Thus we have $f_{FD}(k, \mu) = 1$ if $\mu$ labels one of the filled valence bands, $f_{FD}(k, \mu) = 0$ for the empty conduction bands. Since the term $E_\mu(k) - E_\lambda(k + q)$ in the denominator is very large for states in bands $\mu$ and $\lambda$ respectively
very low and high in energy, we can restrict the sum only over the top VB and the bottom CB. So, $\mu = \nu$ and $\lambda = c$. Let’s use the ‘sum rule’

$$
\sum_{fin} (E_{fin} - E_{in}) | < fin| e^{i\mathbf{q}\cdot \mathbf{r}} | in > |^2 = \frac{\hbar^2 q^2}{2m^*}.
$$

(170)

Thus:

$$
\sum_{\mathbf{G}} < \mathbf{k} + \mathbf{q} + \mathbf{G}, c | e^{i\mathbf{q}\cdot \mathbf{r}} | \mathbf{k}, \nu >^2 = \frac{\hbar^2 q^2}{2m^* E_G}.
$$

(171)

In the denominator let’s approximate

$$
E_v(\mathbf{k}) - E_c(\mathbf{k} + \mathbf{q}) \approx -E_G,
$$

(172)

if $\mathbf{q}$ is small enough (thus, considering screening of long-wavelength perturbations). Note that $E_G$ is the ‘direct’ gap of the crystal averaged over the BZ. Then if we also assume that the frequency of the perturbation is small enough, $\hbar \omega < < E_G$ and we have:

$$
\frac{\epsilon(\mathbf{q} \to 0, \omega \to 0)}{\epsilon_0} \approx 1 + \frac{e^2 \hbar^2}{2m^* \epsilon_0 E_G^2} \sum_{\mathbf{k}} f_{FD}(\mathbf{k}, \nu) = 1 + \frac{e^2 \nu}{m^* \epsilon_0 E_G^2} = 1 + \frac{\hbar^2 \omega_P^2}{E_G^2} = \frac{\epsilon_s}{\epsilon_0},
$$

(173)

having defined the ‘VB Plasma frequency’ $\omega_P = (e^2 \nu / (m^* \epsilon_0))^{1/2}$.

Note how the static (bulk) dielectric constant scales with the inverse of the squared gap.

2. **Static screening by free carriers.**

In dealing with perturbations of the free carriers we shall almost always encounter scattering potentials which are spatially ‘slowly varying’. This means that $\mathbf{q} < < \mathbf{G}$ and we can set

$$
\sum_{\mathbf{G}} | < \mathbf{k} + \mathbf{q} + \mathbf{G}, c | e^{i\mathbf{q}\cdot \mathbf{r}} | \mathbf{k}, \nu > |^2 \approx 1.
$$

(174)
Therefore, considering only a single band (conduction for $n$-type semiconductors, valence for $p$-type), the dielectric response of the free carriers can be written as:

$$
\epsilon(q, \omega) = \epsilon_s + \frac{e^2}{q^2} \sum_k \frac{f(k) - f(k + q)}{E(k) - E(k + q) + \hbar \omega - i\hbar \eta}.
$$

(175)

For a static perturbation (like the potential of an ionized impurity), $\omega = 0$. Moreover let’s assume that the free carriers are in equilibrium, so that $f = f_{FD}$. Now, for long wavelengths we can use the expansions:

$$
E(k + q) - E(k) \approx q \cdot \nabla_k E(k),
$$

(176)

and

$$
f(k + q) - f(k) \approx q \cdot \nabla_k E(k) \frac{\partial f}{\partial E}.
$$

(177)

Then:

$$
\epsilon(q, \omega) = \epsilon_s + \frac{e^2}{q^2} \int dE \rho(E) \left( -\frac{\partial f_{FD}}{\partial E} \right) = \epsilon_s + \frac{e^2}{q^2} \int dE \rho(E) \left( \frac{\partial f_{FD}}{\partial E_F} \right) = \epsilon_s + \frac{e^2}{q^2} \left( \frac{\partial n}{\partial E_F} \right).
$$

(178)

Defining the ‘screening wavevector’ (or ‘screening parameter’ or ‘inverse screening length’)

$$
\beta^2 = \frac{e^2}{\epsilon_s} \left( \frac{\partial n}{\partial E_F} \right),
$$

(179)

we can write:

$$
\epsilon(q, \omega) = \epsilon_s \left( 1 + \frac{\beta^2}{q^2} \right).
$$

(180)
Note:

\[
\beta^2 \rightarrow \begin{cases} 
\frac{e^2 n}{\epsilon_s k_B T} & \text{nondegenerate limit, Debye-Hückel limit} \\
\frac{3e^2 n}{2\epsilon_s E_F} & \text{degenerate limit, Thomas-Fermi limit}
\end{cases}.
\]

(181)

Note that these expressions tend to \( \epsilon_s \) at very large \( q \). This means that free carriers do not screen perturbations which vary too fast spatially.

3. **Long-wavelength, dynamic screening by free carriers (Plasma oscillations).**

Now let's consider the opposite limit of nonzero \( \omega \) but in the limit of \( q \to 0 \). For \( \hbar \omega >> |E(k) - E(k + q)| \) for all \( k \) of interest, let's rewrite the RPA expression for the dielectric function of free carriers as:

\[
\epsilon(q, \omega) = \epsilon_s + \frac{e^2}{q^2} \sum_k \frac{2f(k) \left[ E(k) - E(k + q) \right]}{\hbar^2 \omega^2 - |E(k) - E(k + q)|^2}.
\]

(182)

Now let's expand the numerator in powers of \( q \):

\[
E(k) - E(k + q) \approx -\frac{1}{2} q \cdot \nabla_k^2 E \cdot q = -\frac{\hbar^2 q^2}{2m^*}
\]

(183)

(having recalled the definition of the effective mass) and retain only \( \hbar^2 \omega^2 \) in the denominator. Thus

\[
\epsilon(q \to 0, \omega) \approx \epsilon_s \left( 1 - \frac{1}{\omega^2} \frac{e^2 n}{\epsilon_s m^*} \right) = \epsilon_s \left( 1 - \frac{\omega_p^2}{\omega^2} \right),
\]

(184)

where \( \omega_p = [e^2 n/(\epsilon_s m^*)]^{1/2} \) is the ‘free-carrier plasma frequency’.

Note that this expression tends to \( \epsilon_s \) at very large \( \omega \). This means that free carriers do not screen perturbations which vary too fast in time.

4. **An important example.**

We have seen before that the Coulomb potential of an ionized impurity is screened by valence electrons simply
by replacing the vacuum dielectric constant $\epsilon_0$, with the static, bulk dielectric constant of the (intrinsic) semiconductor, $\epsilon_s$:

$$V_{imp}(r) = \frac{e}{4\pi\epsilon_s r}. \quad (185)$$

By going to Fourier transforms one can show that the potential screened also by free carriers will have the following ‘Yukawa-like’ form:

$$V_{imp}^{(screened)}(r) = \frac{ee^{-\beta r}}{4\pi\epsilon_s r}. \quad (186)$$
Lattice vibrations (phonons)

- Phonons are small lattice vibrations. As for all vibrations, they are better analyzed when decomposed into their ‘normal modes’ (the equivalent of Fourier components), which are plane waves describing a small displacement of each ion from its equilibrium position.

Like all waves in quantum mechanics (such as photons or electrons) they are ‘quantized’. However, since ions are heavy and large (at least when compared to electrons) the basic feature of lattice vibrations are well understood by studying them with classical mechanics. At the end, however, we must recall that the energy of phonons comes in ‘chunks’ (or ‘quanta’) $\hbar \omega_q$, multiple of the phonon frequency $\omega_q$ of a phonons with wavevector $\mathbf{q}$. Also, their occupation at thermal equilibrium is characterized by the Bose-Einstein distribution function:

$$f_{BE}(\mathbf{q}) = \frac{1}{e^{\hbar \omega_q/(k_BT)} - 1}, \quad (187)$$

which is for Bosons (particles or quasi-particles with integer spin) what the Fermi-Dirac distribution is for Fermions (particles or quasi-particles with half-integer spin).

- **General formulation.**
  Let $\mathbf{R}_{l\gamma}(0)$ be the equilibrium position of the ion (labeled by the index $\gamma$ inside the unit cell $l$) and let $\delta \mathbf{R}_{l\gamma}$ its displacement from $\mathbf{R}_{l\gamma}(0)$. Let also $\mathbf{R}_l$ be the coordinate-origin of the cell $l$. In terms of the momentum coordinate of ion $\gamma$ in cell $l$, $\mathbf{P}_{l\gamma} = M_{\gamma} \dot{\mathbf{R}}_{l\gamma}$, where $M_{\gamma}$ is the mass of the ion, the ion Hamiltonian is:

$$H_{ion} = \frac{1}{2} \sum_{l\gamma} \left[ \frac{P_{l\gamma}^2}{M_{\gamma}} + \sum_{l',\gamma',}\sum_{ij} \frac{\partial^2 V_{\gamma'}^{(lat)} (\mathbf{R}_{l\gamma}(0) - \mathbf{R}_{l'\gamma'}(0))}{\partial R_{l\gamma i} \partial R_{l'\gamma' j}} \delta R_{l\gamma i} \delta R_{l'\gamma' j} \right], \quad (188)$$

where $V_{\gamma}^{(lat)}$ is the lattice potential of the ion $\gamma$. Note that here the indices $i$ and $j$ run over the spatial-coordinate indices $x, y, z$. Using Bloch theorem we can now expand $\delta \mathbf{R}_{l\gamma}$ in its spatial Fourier coefficients at
wavevector $\mathbf{q}$ and assume a time dependence $e^{i\omega t}$:

$$
\delta \mathbf{R}_{l\gamma} = \frac{1}{(NM)^{1/2}} \sum_{\mathbf{q}} \xi_{\mathbf{q}\gamma} e^{i\mathbf{q}\cdot\mathbf{R}_l} e^{i\omega t},
$$

where $M$ is the mass of the unit cell and $N$ is the number of cells in the crystal. It should be noted that the periodicity of the lattice implies that $\xi_{\mathbf{q}\gamma}$ does not depend on the cell index $l$, the entire dependence on the cell being absorbed by the phase $e^{i\mathbf{q}\cdot\mathbf{R}_l}$. We also define the “lattice Fourier transform” of the dynamic matrix as

$$
G_{ij\gamma\gamma'}(\mathbf{q}) = \sum_{l''} \hat{G}_{ij\gamma\gamma'}(l'') e^{i\mathbf{q}\cdot(\mathbf{R}_l - \mathbf{R}_{l''})}
$$

where

$$
\hat{G}_{ij\gamma\gamma'}(l'') = -\frac{\partial^2 V^{(lat)}(\mathbf{R}_{l\gamma}^{(0)} - \mathbf{R}_{l'\gamma'}^{(0)})}{\partial R_{\gamma i} \partial R_{\gamma' j}},
$$

where $\mathbf{R}_{l''} = \mathbf{R}_l - \mathbf{R}_{l''}$. Inserting the Fourier expansions for $\delta \mathbf{R}_{l\gamma}$ into the equations of motion determined by the Hamiltonian above, for each $\mathbf{q}$ we obtain the homogeneous linear system

$$
\sum_{j\gamma\gamma'} \left[ G_{ij\gamma\gamma'}(\mathbf{q}) - M \omega^2 \delta_{ij} \delta_{\gamma\gamma'} \right] \xi_{\mathbf{q}\gamma'j} = 0.
$$

The associated secular equation determines the eigenfrequencies $\omega_{\mathbf{q}\eta}$ of oscillations labeled by a “branch index” $\eta$ running over 3 (one LA and 2 TA) acoustic modes and over $3(N_{\text{ions}} - 1)$ optical modes, $N_{\text{ions}}$ being the number of ions in the primitive cell.
- A simple example: The linear chain.

Let’s consider the simpler case of a 1D chain of atoms of mass \( M \). We can drop the indices \( i, j \) (since we are in 1D) as well as the indices \( \gamma \) (since we consider only one atom in each 1cD cell). The ionic Hamiltonian becomes

\[
H_{\text{ion}} = \frac{1}{2} \sum_l \left[ \frac{P_l^2}{2M} + \sum_{l'} G_{ll'} \delta R_l \delta R_{l'} \right].
\]  

(193)

Note that:

\[ G_{l,l'} = G_{l-l',0} \] by translation symmetry

(194)

and

\[ \sum_l G_{0,l} = 0 \] since for equal shifts \( \delta R_l \) the crystal energy does not change.

(195)

The Hamilton equation for the ionic momentum will be:

\[
\dot{P}_l = -\frac{\partial H_{\text{ion}}}{\partial \delta R_l} = - \sum_l G_{ll'} \delta R_{l'},
\]

(196)

or

\[
M \frac{d^2 \delta R_l}{dt^2} = - \sum_l G_{ll'} \delta R_{l'},
\]

(197)

Let’s also make the simplifying assumption that each ion interacts only with its nearest neighbor (so that we can solve the problem ‘by hand’ and not numerically). Then:

\[
M \frac{d^2 \delta R_l}{dt^2} = G_{l,l} \delta R_l + G_{l,l+1} \delta R_{l+1} + G_{l,l+1} \delta R_{l-1}.
\]

(198)

By translation symmetry \( G_{l,l} = G_{0,0} \). Without loss of generality we can consider the case \( l = 0 \). So we can consider only the elements \( G_{0,0}, G_{0,\pm1}, \) and \( G_{\pm1,0} \). Using again the translation symmetry we have

\[
G_{0,1} = G_{-1,0},
\]

(199)
but since $G_{l'l'} = G_{l'l}$ (it's only a switch in the order of the derivatives), we also have

$$G_{0,1} = G_{1,0}, \tag{200}$$

so that $G_{0,1} = G_{1,0} = G_{-1,0} = G_{0,-1}$.

By the property $\sum_l G_{0,l} = 0$ above we have

$$G_{0,0} + G_{0,1} + G_{0,-1} = 0 \rightarrow G_{0,0} = -2G_{1,0}. \tag{201}$$

Let's simply call $G = G_{1,0}$ (the 'spring' constant of the inter-ionic force). Then the Hamilton equation of motion becomes simply:

$$M \frac{d^2 \delta R_l}{dt^2} = G \left[ -2 \delta R_l + \delta R_{l+1} + \delta R_{l-1} \right]. \tag{202}$$

Now let's set

$$\delta R_l = \xi q e^{iqa} e^{i\omega t}, \tag{203}$$

where $a$ is the equilibrium distance between the ions in the chain (our 'lattice constant'). Then:

$$- M \omega^2 = G \left[ 2 \cos(qa) - 2 \right], \tag{204}$$

becomes our trivial 'secular equation' from which we obtain the dispersion

$$\omega = 2 \left| \sin \left( \frac{qa}{2} \right) \right| \left( \frac{G}{M} \right)^{1/2}. \tag{205}$$

Note that the periodicity in $2\pi a$ for $q$ defines the BZ of the 1D chain.

- **Types of phonons.**
  As noted, in a 3D cell having $N_{ions}$ ions, there will be 3 acoustic phonons (no relative motion between ions,
the whole cell vibrates) and $3(N_{ions} - 3)$ optical phonons (characterized by relative displacement among ions in the same cell). The former are purely acoustic waves (like density waves) which effect electrons by modifying (linearly) the band structure with microscopic stress (they act like waves distorting the lattice constant as they travel). This effect is called ‘deformation potential’ interaction after Bardeen and Shockley. Optical phonons in polar materials (such as GaAs) carry an associated dipole field which scatters electrons in an obvious matter.

Acoustic and optical phonons can be ‘transverse’ or ‘longitudinal’, depending on whether the displacement $\delta R$ (also called ‘polarization’) is normal or parallel to the direction of travel $q$. 
Acoustic, transverse (TA)
\[ \delta \vec{R} \perp \vec{k} \]

Optical, transverse (TO)
\[ \delta \vec{R} \perp \vec{k} \]

Acoustic, longitudinal (LA)
\[ \delta \vec{R} \parallel \vec{k} \]

Optical, longitudinal (LO)
\[ \delta \vec{R} \parallel \vec{k} \]
A simple model for transport/conductivity: Drude’s theory.

- **Bloch oscillations in a perfect ballistic crystal.** Band theory predicts that at the edge of the BZ (with a single notable exception) $\nabla_k E(k) = 0$. So, moving across Brillouin Zones along a $k$-line we see a periodic function $E(k)$. For example, let’s assume the simple 1D dispersion for 1D crystal and see how an electron moves under the action of an external electric field $F$:

$$E(k) = E_0 \left[ 1 + \cos(ka + \pi) \right]. \tag{206}$$

The equations of motion imply:

$$v[k(t)] = \frac{1}{\hbar} \frac{dE}{dk} = -\frac{E_0}{\hbar} a \sin[k(t)a + \pi], \tag{207}$$

and

$$k(t) = k(0) - \frac{eFt}{\hbar}. \tag{208}$$

Thus

$$v(t) = \frac{E_0}{\hbar} a \sin(eaFt/\hbar + \pi). \tag{209}$$

The current, $j = ev$, oscillates in time under the action of a d.c. field. These oscillations, called ‘Bloch oscillations’, are not seen experimentally because of the effect of scattering: The frequency of these oscillations is larger than the scattering rate for electrons. In other words, collisions prevent electrons from exiting one BZ and entering the next without scattering. It follows that scattering is crucial in explaining the conductivity of solids.
Complications about scattering. There are three main reasons why a theory of current conduction in solids is complicated:

1. Scattering, as shown by the structure of Fermi Golden Rule, depends on the initial and final crystal momenta of the scattering carrier
2. Many electrons participate in the conduction, so that complicated 'many-body, collective' effects enter the picture
3. The electron (and hole) dispersion is given by a complicated band structure

Drude's model. Before diving into more rigorous formulation of electronic transport, let’s start with a simple model proposed by Drude. Let’s assume that when crossing a device an electron suffers so many collisions that it is enough to lump their collective effect into a 'frictional force', \( F_f = \frac{m v}{\tau} \) such that:

\[
\hbar \dot{k} = m \dot{v} = F_0 + F_f = -eF - \frac{m v}{\tau} ,
\]

(210)

where \( F \) is the external field and \( \tau \) is the relaxation time constant of the frictional force. To see this, note that if we set \( F = 0 \) and assume that at \( t = 0 \) the electron moves with velocity \( v_0 \), then from the equation of motion \( m \dot{v} = -\frac{mv}{\tau} \) we get \( v \propto \exp(-t/\tau) \), that is, the velocity decays to zero exponentially with time constant \( \tau \) if "damped" by the frictional force. The fact that we can assume the existence of a frictional force (bypassing all the difficulties related to the \( k - k' \) dependence of scattering) and that all electrons move at the same velocity \( v \) clearly simplifies the picture. The only justification for these drastic simplifications stems from the fact that Drude's theory can explain many low-energy, low-field, low-frequency transport phenomena. The current density \( j \) is given by

\[
j = -env .
\]

(211)

The equation of motion for \( v \) under the action of an electric field \( F \) and a magnetic field \( B \) will be:

\[
\hbar \dot{k} = m \dot{v} = F_0 + F_f = -e(F + v \times B) - \frac{m v}{\tau} .
\]

(212)

Let’s see 3 examples:

1. \( F \) independent of time, \( B = 0 \).
   Then the steady-state solution implies \( \dot{v} = 0 \) and \( mv = -eF \tau \). This corresponds to a dc current (no
Bloch oscillations) with:

\[ j = -env = \frac{e^2 \tau n}{m} F. \]  

(213)

The proportionality constant between \( j \) and field \( F \) is the **conductivity** \( \sigma \), while the proportionality constant between velocity and electric field is the **mobility** \( \mu \), so

\[ \mu = \frac{e\tau}{m}, \quad \sigma = en\mu. \]  

(214)

2. \( F \) is a low-frequency electric field and \( B = 0 \).

Let’s consider the Fourier component

\[ F = F_0 e^{i\omega t} + cc. \]  

(215)

For the equation of motion let’s try a solution of the form \( v = v_0 e^{i\omega t} + cc \). This gives:

\[ i\omega m v_0 e^{i\omega t} = -eF_0 e^{i\omega t} - \frac{mv_0}{\tau} e^{i\omega t}, \]  

(216)

so that:

\[ -v_0 = \frac{eF_0}{im\omega + \frac{m}{\tau}} = \frac{\mu F_0}{1 + i\omega\tau}. \]  

(217)

So, for small frequencies the velocity is just \( \mu F_0 \), as in a dc situation. But as soon as the frequency begins to approach the ‘relaxation frequency’ \( 1/\tau \), the semiconductor ceases to be only a resistor and it begins to present an inductive delay.

3. \( F \) independent of time, \( B \neq 0 \) and \( B = (0, 0, B_z) \).

We have:

\[ \frac{mv}{\tau} = -e(F + v \times B). \]  

(218)

The presence of the Lorentz force makes it impossible to have \( F \) and \( v \) parallel. Indeed, if we assume \( F = (0, 0, F_x) \) and \( v = (0, 0, v_x) \), for example, we run into troubles. Indeed we see that the Lorentz force causes a force in the \( y \) direction. Electrons flow along that direction and, since there are no current
channels along that direction, they pile up at the sample boundary, building an electric field also along the $y$ direction, to block – at steady state – the flow of more electrons. Thus, we may rewrite the equation of motion in the form:

$$\frac{m(v_x, 0, 0)}{\tau} = -e[(F_x, F_y, 0) - (0, -v_x B_z, 0)] ,$$

(219)

with solution

$$-v_x = \mu F_x ,$$

(220)

and

$$F_y = v_x B_z .$$

(221)

Since we can measure $F_x, F_y,$ and $B_z$, so we can derive $v_x$ and $\mu$. Since we know the current density $j_x = -env_x$, we can also determine the carrier concentration, $n$. This is called the ‘Hall effect’, $F_y$ the ‘Hall field’, and this experimental arrangement is one of the most commonly used method to extract the carrier mobility experimentally.

Our goal is to extend Drude theory to cases where it does not apply, considering in particular how to reassess the role of the relaxation time $\tau$, definitely over-simplified by the Drude model, and how to account for the complicated band-structure of the semiconductor. We shall also justify Drude’s theory in the cases in which it is valid and, in so doing, understand how to improve it. In order to put Drude’s theory on firmer ground, we shall have to look at the basic, microscopic physics of scattering processes in semiconductors, calculate ‘ensemble averages’ of the relaxation rates and derive simplified transport equations starting from more rigorous formulations of transport (namely, the Boltzmann Transport Equation). Our first next step is a look at some microscopic scattering rates.

- **Some important scattering rates using the Fermi Golden Rule.**

Here we consider the scattering rates for the most important collision processes in semiconductors: 1. with ionized (dopant) impurities, and 2. with phonons. Other collision processes are often crucial (scattering among carriers, impact ionization, scattering with rough interfaces between layers, etc.), but scattering with phonons and impurities are almost always important.

First, let’s consider the general formulation. Using Bloch theorem, let’s consider an electron in band $n$, described
by the wavefunction

\[ \Psi(k, n, r, t) = \frac{1}{\sqrt{V}} e^{i k \cdot r} u_k^{(n)}(r) e^{-i E_n(k)t/\hbar} = \left| kn \right>, \tag{222} \]

where \( u_k^{(n)}(r) \) is the periodic Bloch component and \( E_n(k) \) is the energy of the electron of crystal momentum \( \hbar k \) in band \( n \). \( \Psi \) is a solution of the Schrödinger equation for the lattice:

\[ H_{lat} \Psi = i\hbar \frac{\partial \Psi}{\partial t}, \tag{223} \]

where \( H_{lat} = \hbar^2 \nabla^2 / (2m) + V_{lat}(r) \). Let's add to \( H_{lat} \) a perturbation \( H' \) (the static Coulomb potential of the impurity or the oscillating – in space and time – scattering potential of a phonon). From the Fermi Golden Rule we know that the probability per unit time, \( P(k'n', kn) \) for the electron to make a transition to another state \( k' \) in band \( n' \) characterized by the wavefunction

\[ \Psi(k', n', r, t) = \frac{1}{\sqrt{V}} e^{i k' \cdot r} u_{k'}^{(n')} (r) e^{-i E_{n'}(k')t/\hbar} = \left| k'n' \right>, \tag{224} \]

is:

\[ P(k'n', kn) = \frac{2\pi}{\hbar} \left| < k'n' | H' | kn \right|^2 \delta \left[ E_{n'}(k') - E_n(k) \pm \hbar \omega \right], \tag{225} \]

where \( \omega \) is the energy exchanged (gained with the 'minus' sign, lost with the 'plus' sign) during the collision. The 'matrix element' \( \left| < k'n' | H' | kn \right| \) is:

\[ \left| < k'n' | H' | kn \right| = \frac{1}{V} \int_V dr \ e^{-i k' \cdot r} u_{k'}^{(n')} (r) H'(r) e^{i k \cdot r} u_k^{(n)} (r) = \tag{226} \]

\[ = \frac{1}{V} \sum_{G'G} u_{G'k'}^{(n')} u_G^{(n)} \int_V dr \ e^{i (G - G') \cdot r} e^{-i k' \cdot r} H'(r) e^{i k \cdot r}, \tag{227} \]
having used the Fourier transform of the Bloch components in the last step. Most of the perturbation Hamiltonians we shall consider have 'smooth' spatial behavior; that is: $H'(\mathbf{r})$ does not vary appreciably over a single cell. In this case the spatial integral will be significantly different from zero only when the short-wavelength oscillations $e^{i(G-G')\cdot \mathbf{r}}$ disappear, that is, when $G' = G$. Therefore we can write:

$$| < k'n'|H'|kn > | \approx \frac{1}{V} \sum_{G' \equiv G} u^{(n')}_{G'k'}^* u^{(n)}_{Gk} \int_V d\mathbf{r} \ e^{-i k' \cdot \mathbf{r}} H'(\mathbf{r}) \ e^{i k \cdot \mathbf{r}} = \mathcal{I}(k'k, n'n) < k'|H'|k >$$

(228)

where the term

$$\mathcal{I}(k'k, n'n) = \sum_{G} u^{(n')}_{Gk'}^* u^{(n)}_{Gk}$$

(229)

represents the 'overlap integral' between two Bloch functions. Limiting our analysis to a single band (so assuming $n = n'$), we shall ignore this term since when $k$ and $k'$ are not too distant, compared to the length of the $G$-vectors, this overlap factor is not too different from unity. Since we have limited our attention to a single band, we shall drop the band-index $n$ in the following. The matrix element

$$| < k'|H'|k > | = \frac{1}{V} \int_V d\mathbf{r} \ e^{-i k' \cdot \mathbf{r}} H'(\mathbf{r}) \ e^{i k \cdot \mathbf{r}}$$

(230)

is the matrix element of the perturbation Hamiltonian on 'envelope' plane waves, as we would have in the case of free electrons in vacuum.

The total scattering rate, $1/\tau(k)$ will be obtained by summing the transition rates for all final states $|k' >$:

$$\frac{1}{\tau(k)} = \frac{2\pi}{\hbar} \sum_{k'} | < k'|H'|k > |^2 \delta[E(k') - E(k) \pm \hbar \omega] .$$

(231)
1. **Scattering with ionized impurities.**

Let’s start by considering as perturbation Hamiltonian the Coulomb potential of an impurity located at \( r = 0 \):

\[
H'(r) = \frac{e^2}{4\pi\epsilon_0 r}.
\]  

(232)

This potential will be screened by both valence electrons and by free carriers. The effect of valence electrons is accounted for by simply replacing the vacuum dielectric constant \( \epsilon_0 \) with the static, long-wavelength dielectric constant of the semiconductor, \( \epsilon_s \). In order to account for the more complicated dielectric response of the free carriers, let’s note that the Coulomb potential does not change with time and that its effect will be mostly felt at large distances. Therefore we can employ the static, long-wavelength expression for \( \epsilon(q, \omega) \) (see Eq.(180) page 50):

\[
\epsilon(q, \omega = 0) = \epsilon_s \left( 1 + \frac{\beta^2}{q^2} \right).
\]  

(233)

So, let’s take the Fourier transform of the Coulomb potential. As done in the Homework Assignment 1, Problem 5, the derivation is a little ‘tricky’:

\[
H'_{q} = \frac{1}{V} \int_V dr \frac{e}{4\pi\epsilon_0 r} e^{iq\cdot r} = \frac{e^2}{2\epsilon_0 V} \int_0^R dr \int_0^\pi d\theta e^{iqr \cos \theta} = -\frac{e}{V\epsilon_0 q^2},
\]  

(234)

where the last step requires some ‘creativity’... Alternatively, we can start from Poisson equation for a charge \( e \) located at \( r = 0 \):

\[
\nabla^2 V(r) = -\frac{e}{\epsilon_0} \delta(r).
\]  

(235)

In order to take Fourier transforms, note that:

\[
\nabla^2 V(r) = \nabla^2 \sum_q V_q e^{iq\cdot r} = -\sum_q V_q q^2 e^{iq\cdot r},
\]  

(236)
and
\[ \delta(r) = \frac{1}{Vol} \sum_q e^{i\mathbf{q} \cdot \mathbf{r}}, \]  
(237)

where we have written \( Vol \) for the normalization volume to avoid confusion with the potential \( V \). Then, from Poisson equation and the two previous equations:
\[ V_{\mathbf{q}} = -\frac{e}{Vol \varepsilon_0 q^2}, \]
(238)
as before. The Hamiltonian screened by both valence and free electrons will be:
\[ H^{(s)}(\mathbf{r}) = \frac{\varepsilon_0}{\varepsilon_s} \sum_q \frac{H'_{\mathbf{q}}}{1 + \frac{\beta^2}{q^2}} = \frac{1}{Vol} \sum_q \frac{e^2}{\varepsilon_s} \frac{1}{q^2 + \beta^2} = \frac{e^2}{4\pi\varepsilon_s} e^{-\beta r}. \]
(239)

So, finally,
\[ | < \mathbf{k}' | H^{(s)} | \mathbf{k} > | = \frac{e^2}{Vol \varepsilon_s} \frac{1}{|\mathbf{k}' - \mathbf{k}|^2 + \beta^2}. \]
(240)

So, the scattering rate will be (using again \( V \) instead of \( Vol \)):
\[ \frac{1}{\tau_{imp}(\mathbf{k})} = \frac{2\pi}{\hbar} \frac{e^4}{V^2 \varepsilon_s^2} \sum_{\mathbf{k}'} \frac{1}{(|\mathbf{k}' - \mathbf{k}|^2 + \beta^2)^2} \delta[E(\mathbf{k}') - E(\mathbf{k})]. \]
(241)

Now let’s assume a semiconductor with isotropic effective mass \( m^* \) and ‘parabolic’ dispersion \( E(\mathbf{k}) = E(k) = \hbar^2 k^2/(2m^*) \). Let’s transform the sum over final wavevectors \( \mathbf{k}' \) to an integral:
\[ \frac{1}{\tau_{imp}(\mathbf{k})} = \frac{2\pi}{\hbar} \frac{e^4}{V \varepsilon_s^2} \frac{1}{(2\pi)^3} \int d\mathbf{k}' \frac{1}{(|\mathbf{k}' - \mathbf{k}|^2 + \beta^2)^2} \delta[E(\mathbf{k}') - E(\mathbf{k})]. \]
(242)
Now let’s consider polar coordinates with the $z$-axis along $k$ so that $|k' - k|^2 = k'^2 + k^2 - 2k'k \cos \theta$, let’s change the integration variable $k'$ to $E' = \hbar^2 k'^2 / (2m^*)$ and get:

$$\frac{1}{\tau_{imp}(E)} = \frac{e^4 m^* (2m^* E)^{1/2}}{2\pi \hbar^4 \epsilon_s^2 V} \int_0^\pi d\theta \sin \theta \frac{1}{\left[ \frac{4m^* E}{\hbar^2} (1 - \cos \theta) + \beta^2 \right]^2},$$

(243)

where $E = E(k)$, the initial energy. We have also used the fact that the scattering rate depends only on the initial energy (not on the direction of $k$).

The wavevector $\hbar q = \hbar (k' - k)$ is the ‘momentum transfer’ and the polar angle $\theta$ is the ‘scattering angle’. A large momentum transfer corresponds to a large scattering angle. Note, however, that the momentum transfer $\hbar q$ enters the denominator of the integrand above with a high power (for small $\beta$ the integrand behaves like $q^{-4}$). Therefore, this scattering process is mainly a ‘forward scattering’ (or ‘small-angle scattering’) process: For small $\beta$ compared to $k$, (that is, for weak screening or high electron energy), electrons will not be deflected very much from their trajectory. On the contrary, at low electron energy or for strong screening (so that $k = k' << \beta$), the term $\cos \theta$ in the denominator does not play much of a role and scattering is almost ‘isotropic’.

The expression for $1/\tau_{imp}(E)$ we have just derived represents the rate for collisions with a single impurity in a volume $V$. If we assume that we have $N_I$ impurities per unit volume, we must replace $1/V$ above with $N_I$. Performing the integral we have, finally:

$$\frac{1}{\tau_{imp}(E)} = \frac{N_I e^4 m^* (2m^* E)^{1/2}}{\pi \hbar^4 \epsilon_s^2} \frac{1}{\beta^2 \left( \frac{8m^* E}{\hbar^2} + \beta^2 \right)}.$$

(244)

Note that for large initial energies the scattering rate decreases as $1/\tau_{imp}(E) \sim E^{-1/2}$. Indeed an electron of energy much larger than the depth of the ‘Coulomb well’ will be largely unaffected by the scattering potential. This is a general properties of Coulomb scattering. Note also that, since $\beta^2 \propto n$, where $n$ is the density of free carriers, as the density $n$ increases the scattering rate decreases. This is due to the fact that
more free carriers screen better the scattering potential. Finally, the scattering rate increases linearly with the
density of scattering centers $N_f$, as expected.

2. **Scattering with phonons.**

We have seen that phonons are oscillations of the ions in the lattice around their equilibrium positions. They
behave like sound waves traveling with wavevector $q$ and possessing energy $\hbar \omega_q$. The oscillating displacement
of an ion at the lattice site $l$ and position $\tau_\gamma$ in the cell was derived semiclassically above as:

$$
\delta R_{l\gamma} = \frac{1}{(NM)^{1/2}} \sum_q \xi_{q\gamma} e^{iq \cdot R_l} e^{i\omega_q t},
$$

(245)

where $M$ is the mass of the unit cell and $N$ is the number of cells in the crystal and $\xi_q$ is the ‘polarization’
(transverse or longitudinal) of the phonon.

Clearly, any distortion of the crystal will affect the electrons, so it is not surprising that phonons scatter
electrons. Two main differences between impurity scattering and phonon scattering are: 1. Phonon scattering
is not elastic: Electrons can absorb or emit phonons, thus losing or gaining energy. 2. Scattering with phonons
is intrinsically a quantum mechanical process. In particular, following a semiclassical, phenomenological
derivation outlined below, we can easily guess that phonon absorption must be proportional to the numbers of
phonons already present at the lattice temperature, given by the Bose-Einstein ‘occupation number’

$$
n_q = f_{BE}(q) = \frac{1}{e^{\hbar \omega_q/(k_BT)} - 1},
$$

(246)

but we cannot derive the result the phonon emission is proportional to $1 + n_q$, in analogy with the ‘stimulated
photon emission’ (discovered by Einstein) which is the basic idea behind the principles of operation of masers
and lasers. The derivation of this result can be followed only using a more advanced quantum mechanical
technique known as ‘second quantization’.

Bardeen and Shockley have proposed a phenomenological, semiclassical way of deriving the strength of the
scattering between electrons and acoustic phonons. When we compress or dilate a crystal, say by changing
the lattice constant $a \rightarrow a + u$, the band structure changes. In particular, the energy of the conduction-band
minimum moves by an amount

$$\Delta E_C = E_C(a + u) - E_C(a) \approx \frac{dE_C}{da} u.$$  \hfill (247)

In 3D, $\Delta E_C$ is proportional to the change of the volume $\Delta V$ which, for a lattice displacement $u$ is given by

$$\Delta E_C = V \frac{dE_C}{dV} \nabla \cdot u.$$  \hfill (248)

The constant $V \frac{dE_C}{dV}$ is called ‘deformation potential’ for the conduction band and denoted by $\Delta_{ac}$.

If we consider phonons of wavelength sufficiently long (that is, spanning many cells), we can approximate the effect of the density-wave associated with acoustic phonons (consisting of alternating crests/troughs of compression/dilation/compression/dilation,...) as having locally the same effect of a uniform, global compression or dilation of the crystal. So we can identify the uniform displacement $u$ above with the ionic displacement $\delta R_{l\gamma}$ and we can express the local shift of the CB-edge by the polarization $\xi_q$. This ‘fluctuating’ CB-edge is what causes the electrons to scatter. Skipping the derivation (which requires second-quantization techniques), the squared matrix element for electron/acoustic-phonon scattering has the form:

$$| < k' \mid H_{ac} \mid k > |^2 = \frac{\Delta_{ac}^2 \hbar \omega_q^2}{2V \rho c_s^2} \left( n_q + \frac{1}{2} \pm \frac{1}{2} \right),$$  \hfill (249)

where $\rho$ is the density of the crystal, $c_s$ is the sound velocity, the ‘plus’ sign refers to emissions, the ‘minus’ sign to absorption processes. Note that the frequency of acoustic phonons (in Si there are 2 transverse modes – TA – and a single longitudinal mode – LA) can be well approximated by a linear relation $\omega_q \approx c_s q$. Therefore the matrix element grows with momentum transfer and we see that scattering with acoustic phonons is a ‘large-angle scattering’, very effective in ‘randomizing’ the electron direction and momentum. This has a great effect on the mobility.

A similar expression holds for optical phonons simply replacing the frequency $\omega_q$ and deformation potential $\Delta_{ac}$ with their ‘optical’ counterparts $\omega_{op}$ and $\Delta_{op}$. However, in the literature usually one finds ‘optical
deformation potential’ usually denoted by \((D_t K)_{op}\) which is defined as \(\Delta_{op}(\omega_{op}/c_s)\), so that:

\[
| < k' | H_{op} | k > |^2 = \frac{(D_t K)^2_{op} \hbar}{2V \rho \omega_{op}} \left( n_{op} + \frac{1}{2} \pm \frac{1}{2} \right).
\]  

(250)

The calculation of the scattering rates for emission and absorption of optical phonons is relatively simple, since the frequency of optical phonons is roughly independent of \(q\), and so in the occupation number \(n_{op}\). Thus, proceeding as we did for impurity scattering:

\[
\frac{1}{\tau_{op}(E)} = \frac{(D_t K)^2_{op} m^*}{2^{1/2} \pi \hbar^3 \rho \omega_{op}} \left[ n_{op}(E + \hbar \omega_{op})^{1/2} + (1 + n_{op})(E - \hbar \omega_{op})^{1/2} \right],
\]  

(251)

where, of course, emission should be included only if \(E > \hbar \omega_{op}\).

For acoustic phonons things are more complicated, since the phonon energy depends on the phonon wavevector \(q\) and so does the Bose factor \(n_q\). A common approximation embraced to simplify the problem is to assume that the phonon energy is smaller that the thermal energy \(k_B T\) (at sufficiently low \(T\) and for sufficiently small \(q\)), so that \(n_q \approx k_B T/(\hbar \omega) = k_B T/(\hbar c_s q)\) and to ignore also the energy lost or gained in a collision, if the phonon energy is much smaller then the electron energy. These two approximations are called the ‘equipartition’ and the ‘elastic’ approximations, respectively. Having made these approximations one has:

\[
\frac{1}{\tau_{ac}(E)} = \frac{2^{1/2} \Delta_{ac}^2 m^*}{\pi \hbar^4 \rho c_s^2} E^{1/2}.
\]  

(252)
The graphs show the carrier-phonon scattering rate in Si, with separate plots for electrons and holes. The x-axis represents kinetic energy in eV, and the y-axis represents the scattering rate in units of $10^{14}$ sec$^{-1}$. Two temperatures, 300 K and 77 K, are compared for each type of carrier. The plots illustrate how the scattering rate changes with kinetic energy at these different temperatures.
The Boltzmann Transport Equation.

• Assumptions.

We are now ready to deal with the equation which describes the way free carriers (electrons in the CB, holes in the VB) move in a semiconductor under the action of an electric field.

In general, such an equation should be derived starting from the Schrödinger equation for a many particle (the free carriers) system, including the lattice potential as well as the perturbations which represent scattering. The problem is daunting and still largely unsolved. Instead, we shall make the following approximations:

1. Plane waves of the form $e^{i\mathbf{k} \cdot \mathbf{r}}$ (or the corresponding plane-wave ‘envelopes’ times a Bloch factor in a semiconductor) describe an electron with well-defined momentum but uniformly distributed in space. Instead, an electron localized in space must be described by a ‘wave packet’ which consists of a linear superposition of plane waves. Thus, as a consequence of Heisenberg principle, we must have some uncertainty about the electron position as well as about its momentum. We shall assume that the size of the device is much larger than the size of the electron wave-packet and that the energy scales of interest (mainly, the applied bias) is much larger than the uncertainty implied by the ‘spread’ of the electron momentum. Therefore, we can assume that the electron can be described by its well-defined position $\mathbf{r}$ and its well defined (crystal) momentum $\hbar \mathbf{k}$, as if were a classical particle.

2. The electric field and all other perturbations are weak enough so that we can use the envelope approximation to describe the motion of electrons (or holes) in a semiconductor.

3. All collision processes are independent, instantaneous, local in space, and are perturbations weak enough to justify the use of the first-order (in perturbation theory) Fermi Golden Rule.

4. As a consequence, all collision processes depend only on the local, instantaneous electron variables, without retaining memory of where and when the previous collision happened. This is the equivalent of Boltzmann’s assumption of ‘molecular chaos’ which is essential to arrive at an irreversible dynamics.

5. An unrelated assumption we shall need consists in assuming that the time-dependence of the system is slow enough to justify the use of electrostatics – rather than the full set of Maxwell equations of electrodynamics – and so the electric field will be determined only by the instantaneous distributions of the charge carriers via Poisson equation.
Formulation.

Let’s define the distribution function \( f(\mathbf{r}, \mathbf{k}, t) \) as a function in phase space \((\mathbf{r}, \mathbf{k})\) such that \( f(\mathbf{r}, \mathbf{k}, t)d\mathbf{r}dk \) is the number of particles which at time \( t \) are in a spatial volume \( d\mathbf{r} \) around the position \( \mathbf{r} \) and have (crystal) momentum in a volume \( d\hbar \mathbf{k} \) around the point \( \hbar \mathbf{k} \) in (crystal) momentum space.

The equation controlling the time evolution of the distribution function is:

\[
\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} = -\frac{d\mathbf{r}}{dt} \cdot \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t) - \frac{d\mathbf{k}}{dt} \cdot \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k}, t) + \sum_{\mathbf{k}'} \left[ P(\mathbf{k}, \mathbf{k}') f(\mathbf{r}, \mathbf{k}', t) - P(\mathbf{k}', \mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \right],
\]

which, using the equations of motion and introducing the electrostatic potential \( \phi(\mathbf{r}) \) such that \(-\nabla_{\mathbf{r}} \phi(\mathbf{r}) = \mathbf{F} \) (\( \mathbf{F} \) is the electric field), becomes:

\[
\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} = -\frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t) + e \nabla_{\mathbf{r}} \phi(\mathbf{r}, t) \cdot \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k}, t) + \sum_{\mathbf{k}'} \left[ P(\mathbf{k}, \mathbf{k}') f(\mathbf{r}, \mathbf{k}', t) - P(\mathbf{k}', \mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \right].
\]

This is the famous ‘Boltzmann transport equation’ (BTE). It’s a most difficult equation to deal with and several methods of solutions have been devised during its 150-year old history:

- **Direct solution:** The BTE is transformed into a gigantic linear problem by discretizing real-space and \( k \)-space. The method, however, requires enormous computing power, even by today’s standard, and can be used only in highly symmetric situations.
- **Iterative technique:** The BTE is transformed into a purely integral equation (Chamber’s path formulation) and this equation is solved by iterations. The process is slow, often not converging, and it has been used only in a limited range of applications.
- **Monte Carlo technique:** This method – first used in Los Alamos to simulate neutron transport for the A-bomb – is nowadays the most widely used when a full solution of the BTE is required. Particles are simulated along their paths – solution of the ballistic equations of motion – occasionally interrupted by collisions.
Expansions: The unknown $f$ is expanded into suitable basis functions (spherical harmonics, Legendre or Hermite polynomials) and a hierarchy of equations is obtained. The expansion – and so the hierarchy – is truncated at some suitable order. The resulting equations are partial differential equations (PDEs) which are solved numerically.

Moment methods: The BTE is multiplied by powers of $k$ (or of the velocity $v(k)$), the resulting equation is integrated over $k$-space to obtain PDEs for the ‘moments’ $Q_n(r, t) = \int \frac{d^3 k}{(2\pi)^3} k^n f(r, k, t)$. This method will be discussed below up to the third moment ($n = 0, 1, \text{and } 2$), since it leads to the drift-and-diffusion equations (DDE) which, with Poisson equation, constitute the ‘basic semiconductor equations’.

Linearization. The unknown is assumed to be of the form $f = f_{eq} + f_1$, where $f_{eq}$ is the Fermi Dirac equilibrium function, and $f_1$ is a small correction to leading order in the electric field. Clearly, this method is limited to near-equilibrium situations and it is manly used to derive an expression for the ohmic mobility of the charge carriers.
\[
\frac{\partial f(k, r, t)}{\partial t} = -\frac{d\mathbf{k}}{dt} \cdot \nabla_k f(k, r, t) - \frac{d\mathbf{r}}{dt} \cdot \nabla_r f(k, r, t) + \left( \frac{\partial f(k, r, t)}{\partial t} \right)_{\text{coll}}
\]

"drift"  "diffusion"  "in" - "out"
• **Method of moments.**

We now derive the equations for the first two moments of the BTE, thus obtaining the ‘drift-diffusion’ equations (DDE). **In the following we shall assume steady-state,** that is we assume no explicit time dependence of the distribution function, so that we assume \( \partial f / \partial t = 0 \) in the BTE.

Let’s define the ‘ensemble average’ of a dynamic variable \( Q(k) \) as

\[
\langle Q(r, t) \rangle = \frac{2}{n(r, t)} \int \frac{dk}{(2\pi)^3} Q(k) f(r, k, t).
\]  

(256)

As stated above, the \( n^{th} \) moment of the distribution function is the ensemble average of the dynamic variable \( Q(k) = k^n \), where the notation \( k^n \) denotes the tensor \( k_i k_j \ldots k_l \) of rank \( n \).

1. **First moment: Charge continuity.**

If \( Q(k) = 1 \), all we have to do is to simply integrate the BTE over \( k \)-space. Note that:

\[
\int \frac{dk}{(2\pi)^3} \frac{\partial f(r, k, t)}{\partial t} = \frac{\partial}{\partial t} \int \frac{dk}{(2\pi)^3} f(r, k, t) = \frac{\partial n}{\partial t},
\]  

(257)

where \( n \) is the carrier density. Also:

\[
\int \frac{dk}{(2\pi)^3} \frac{dr}{dt} \cdot \nabla_r f(r, k, t) = \nabla_r \cdot \int \frac{dk}{(2\pi)^3} v(k) f(r, k, t) = \frac{1}{e} \nabla_r \cdot j(r, t),
\]  

(258)

where the current density \( j(r, t) \) has been defined as

\[
j(r, t) = e \int \frac{dk}{(2\pi)^3} v(k) f(r, k, t).
\]  

(259)

Moreover:

\[
\int \frac{dk}{(2\pi)^3} \frac{dk}{dt} \cdot \nabla_k f(r, k, t) = \frac{dk}{dt} \cdot \int \frac{dk}{(2\pi)^3} \nabla_k f(r, k, t) = \frac{dk}{dt} \cdot \int_\Sigma f n d\sigma = 0,
\]  

(260)
(having used Green's identity) where the surface integral in \( \mathbf{k} \)-space is performed over a surface \( \Sigma \) infinitely far away and the distribution function is assumed to vanish at infinity. Finally,

\[
\int \frac{dk}{(2\pi)^3} \int \frac{dk'}{(2\pi)^3} \left[ P(\mathbf{k}, \mathbf{k'}) f(\mathbf{r}, \mathbf{k'}, t) - P(\mathbf{k'}, \mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \right] = 0 .
\]  

(261)

Therefore, by integrating the BTE over \( \mathbf{k} \)-space we get:

\[
\frac{\partial n}{\partial t} = \frac{1}{e} \nabla \cdot \mathbf{j} ,
\]

(262)

which expresses conservation of mass (or particle density). It is the continuity equation for charge (or density).

2. **Second moment: Current continuity.**

Following the recipe for getting moments, we could multiply the BTE by \( \mathbf{k} \) and integrate over \( \mathbf{k} \). Rather, we’ll follow a slightly different procedure in close analogy with the linearization approach, since it will allow us to define properly the carrier mobility. Let’s define implicitly the momentum relaxation time \( \tau_{p,i}(\mathbf{k}) \) along the direction \( i (= x, y, z) \) via the integral equation:

\[
\frac{1}{\tau_{p,i}(\mathbf{k})} = \int \frac{dk}{(2\pi)^3} \left[ 1 - \frac{v_i(\mathbf{k'})\tau_{p,i}(\mathbf{k'})}{v_i(\mathbf{k})\tau_{p,i}(\mathbf{k})} \right] P(\mathbf{k'}, \mathbf{k}) .
\]

(263)

A few comments about this ‘momentum (or velocity) relaxation time’: If the scattering process is elastic and the band-structure of the semiconductor is isotropic, then \( k = k' \), \( v_i(\mathbf{k}) = v_i(\mathbf{k'}) \), the relaxation time itself depends only on the magnitude of \( \mathbf{k} \), so that the term in square bracket becomes \( 1 - \cos \theta \), where \( \theta \) is the angle between \( \mathbf{k} \) and the \( x_i \)-axis. Thus the expression for \( \tau_{p,i} \) simplifies to:

\[
\frac{1}{\tau_{p,i}(\mathbf{k})} = \int \frac{dk}{(2\pi)^3} (1 - \cos \theta) P(\mathbf{k'}, \mathbf{k}) .
\]

(264)

This is the scattering rate ‘weighted’ by change of the \( i^{th} \) component of the velocity. For small-angle scattering (such as Coulomb scattering) small \( \theta \) dominate inside \( P(\mathbf{k'}, \mathbf{k}) \), the term \( (1 - \cos \theta) \) is small and
the effect of the scattering process is weak. In words, even if scattering is frequent, each collision changes
the \( i^{th} \) component of the velocity by such a small amount that the time required to relax the velocity to its
equilibrium value is very long. On the contrary, for large-angle scattering each collision is very effective. For
example, for optical-phonon scattering \( P(k', k) \) is isotropic (that is, \( P(k', k) \) is the same for every angle
between \( k' \) and \( k \), so \( P \) depends only on the magnitude of the wavevectors (\( i.e., P(k', k) = P(k', k) \)),
the term \( \cos \theta \) vanishes after integration over the polar angle and the momentum relaxation time becomes
identical to the scattering rate:

\[
\frac{1}{\tau_{p,i}(k)} = \frac{1}{\tau(k)} = \int \frac{dk}{(2\pi)^3} P(k', k) . \tag{265}
\]

Such an isotropic scattering process is thus called ‘randomizing’: A single collision is sufficient to relax the
momentum. In general, for anisotropic and inelastic processes the momentum relaxation time should be
calculated by solving the implicit definition above which is actually an integral equation. We shall not discuss
it here, but for some important processes – such as scattering with longitudinal optical phonons in GaAs –
extensive additional work is required to compute \( \tau_{pi} \).

Now let’s multiply the BTE by \( v_i(k) \tau_{p,i}(k) \) and integrate over \( k \). We get:

\[
j_i = e \sum_j \langle \tau_{p,i} \frac{\partial}{\partial x_j} (nv_iv_j) \rangle - \frac{e^2 n}{\hbar} \sum_j F_j \langle v_i \tau_{pi} \frac{\partial f}{\partial k_j} \rangle + \frac{\partial}{\partial t} \langle j_j \tau_{pi} \rangle . \tag{266}
\]

- **The drift-diffusion equation.**

The equation above presents an obvious problem: The quantity \( \langle nv_iv_j \rangle \) is a \textit{third} moment of the BTE. The
equations it obeys should be derived by taking the next moment of the BTE. In turn, as we shall see, these
equations (called ‘hydrodynamic equations’) contain a term which is a fourth moment, which in turn obeys
equations containing the fifth moment, and so on. In order to be able to consider only a finite number of
moments we need a ‘closure’, that is, a relation which, under realistic approximations, allow us to express an
\( n + 1^{st} \) moment in terms of an \( n^{th} \) moment.
In order to do that, let’s define two important quantities: The **mobility** tensor:

$$\mu_{ij} = -\frac{e}{\hbar} \langle \tau_{p,i} v_i \frac{\partial f}{\partial k_j} \rangle ,$$

and the **diffusion** tensor:

$$D_{ij} = \langle \tau_{p,i} v_i v_j \rangle .$$

(268)

Assuming that $f$ is a Boltzmann function at equilibrium, so that $f \approx f_{eq} = \exp\{ -[E(k) - E_F] / (k_B T) \}$, we have

$$\frac{\partial f_{eq}}{\partial k_j} = (\frac{\partial f_{eq}}{\partial E}) (\frac{\partial E}{\partial k_j}) = -\hbar v_j f_{eq} / (k_B T)$$

so that

$$D_{ij} = \frac{k_B T}{e} \mu_{ij} ,$$

(269)

which is called the ‘Einstein relation’. This is the desired ‘closure’ relating an ‘unwanted’ third moment to a second moment (the mobility tensor). (For degenerate semiconductors this relation can easily be extended.)

Let’s now simplify our notation assuming that the mobility tensor is actually a scalar quantity (a diagonal tensor with equal elements along the diagonal). Then, denoting by $\mu$ and $D$ the scalar mobility and diffusion constant, we have from the second-moment equation:

$$j = e \nabla (nD) + en\mu F ,$$

(270)

which says that the current density $j$ is the sum of a diffusion current, $e \nabla (Dn)$, and a ‘drift’ term driven by the electric field, $en\mu F$. A similar equation, which we shall write below, applies to hole transport as well. Note that, by the way we have derived the moments of the BTE, ignored higher-order moments, and assumed the Einstein closure, the DDE equations are valid in the semiclassical limit, in the limit of small electric fields and of carriers close to thermal equilibrium.

- **Ohmic mobility.**

  Note that by definition of mobility (along the $x$-axis, without loss of generality in the isotropic limit) we have:

$$\mu = -\frac{e}{\hbar} \langle \tau_{p,v_x} \frac{\partial f_{eq}}{\partial k_x} \rangle = -\frac{2e}{3nm^*} \int_0^\infty dE \rho(E) E \tau_p(E) \frac{\partial f_{eq}}{\partial E} =$$

(271)
In case of degenerate semiconductors this expression is trivially generalized to:

$$
\mu = \frac{2e}{3nm^*k_BT} \int_0^\infty dE \rho(E) E \tau_p(E) f_{eq}(E) \ .
$$

(272)

where now $f_{FD}$ is the equilibrium Fermi-Dirac distribution. This expression is known as the ‘Kubo-Greenwood’ equation.

A simpler expression is useful, since it is related to the simpler Drude model of conduction. Note that:

$$
\mu = -\frac{e}{\hbar} \langle \tau_p v_x \frac{\partial f}{\partial k_x} \rangle \approx \frac{e}{k_BT} \langle \tau_p v_x v_x \rangle \rightarrow \frac{e}{m^*} \langle \tau_p E \rangle.
$$

(274)

When discussing Drude’s models, we defined the mobility as $e\tau/m^*$. We can recognize the same expression above, provided we define Drude’s $\tau$ as the ‘energy weighted average’ of the momentum relaxation time $\langle \tau_p E \rangle/\langle E \rangle$.

As an example, let’s compute the electron mobility in intrinsic Si assuming only scattering with acoustic phonons. Since this type of scattering has been calculated before in the elastic approximation and it is isotropic, the momentum relaxation rate will be equal to the scattering rate which, we recall, is:

$$
\frac{1}{\tau_{ac}(E)} = \frac{1}{\tau_p(E)} = \frac{2^{1/2} \Delta_{ac}^2 m^* 3/2 k_BT}{\pi \hbar^4 \rho c_s^2} E^{1/2},
$$

(275)

where $m^*$ is the effective mass in a single valley of Si. Let’s also recall two more equations derived above which we need now: The electron density in the non-degenerate, high-$T$ limit is

$$
n = \frac{1}{4} \left( \frac{2m^*k_BT}{\pi \hbar^2} \right)^{3/2} e^{E_F/(k_BT)} = N_c e^{-(E_C-E_F)/(k_BT)},
$$

(276)
and, finally, the density of states is given by:

$$\rho(E) = \frac{m_d(2m_dE)^{1/2}}{\hbar^3\pi^2},$$

(277)

where \(m_d = 6^{2/3}(m_Lm_T^2)^{1/3}\) is the Si DOS effective mass. Putting it all together, we have

$$\mu = \frac{2e\hbar\rho c_s^2 m_d^{3/2} eE_F/(k_BT)}{3\pi m^*5/2 \Delta_{ac}^2 (k_BT)^2} \frac{\pi \hbar^2}{2m_d k_BT} \left( \frac{\pi \hbar^2}{2m_d k_BT} \right)^{3/2} \int_0^\infty dE E e^{-E/(k_BT)} = \int_0^\infty dE E e^{-E/(k_BT)} \quad \text{(278)}$$

$$= \frac{2^{3/2}\pi^{1/2} e\hbar^4 \rho c_s^2}{3m^*5/2 \Delta_{ac}^2 (k_BT)^{3/2}}. \quad \text{(279)}$$

Clearly, other scattering processes should be considered to get an accurate estimate for the electron mobility in intrinsic Si (mainly, phonon scattering between various valleys, the so-called ‘intervalley scattering’). Yet, the equations above already exhibits some correct trends, especially the \(T^{-3/2}\) dependence on temperature, which is indeed observed experimentally.

- **The drift-diffusion device equations.**

Let’s summarize what we have obtained so far by writing the set of equations which control the charge transport and the electrostatic potential in semiconductors. Let’s denote by \(\phi\) the electrostatic potential (so that \(V = -e\phi\) is the potential energy), \(n\) and \(p\) the electron and hole densities, by \(j_n\) and \(j_p\) the electron and hole current densities, and use similar subscripts for \(D\), \(\mu\), etc. Then, accounting also for generation, \(G\), and recombination, \(R\), processes which we shall discuss below:
\[
\frac{\partial n}{\partial t} = \frac{1}{e} \nabla \cdot j_n + G_n - R_n \quad \text{continuity equation for electrons}
\]
(280)

\[
\frac{\partial p}{\partial t} = -\frac{1}{e} \nabla \cdot j_p + G_p - R_p \quad \text{continuity equation for holes}
\]
(281)

\[
j_n = e\nabla(nD_n) + en\mu_n F \quad \text{drift and diffusion for electrons}
\]
(281)

\[
j_p = -e\nabla(pD_p) + ep\mu_p F \quad \text{drift and diffusion for holes}
\]
(281)

\[-\nabla[\epsilon_s \nabla \phi] = e[p - n + N_D - N_A] \quad \text{Poisson equation}
\]
(282)

**The third-moment: Hydrodynamic model and energy transport.**

For simplicity, let's adopt from now on the effective mass approximation, even though it is possible to work on the third-moment using an arbitrary band structure. Multiplying the BTE by the total 'kinetic energy tensor', 
\[E_{ij}(k) = \hbar^2 k_i k_j/(2m^*)\] and integrating over \(k\) we have:

\[
\frac{\partial}{\partial t} (n\langle E_{ij} \rangle) = -\sum_l \frac{\partial}{\partial x_l} (n\langle E_{ij} v_l \rangle) + \frac{en}{\hbar} \sum_l F_l \langle E_{ij} \frac{\partial f}{\partial k_l} \rangle - \frac{n\langle E_{ij} \rangle}{\hat{\tau}_{w,ij}},
\]
(283)

where the 'energy relaxation time' tensor \(\hat{\tau}_{w,ij}\) is defined as:

\[
\frac{1}{\hat{\tau}_{w,ij}} = \frac{2}{n\langle E_{ij} \rangle} \frac{\hbar^2}{2m^*} \int \frac{dk'}{(2\pi)^3} \int \frac{dk}{(2\pi)^3} (k_i k_j - k'_i k'_j) P(k', k) f(r, k, t).
\]
(284)

This complicated equation (actually: It's a tensor equation and so it represents 6 independent equations) is simplified by taking the trace, so obtaining an equation expressing conservation of kinetic energy 
\[E = \]
\[ \hbar^2 k^2 / (2m^*) \] = \frac{\partial}{\partial t} n\langle E \rangle = - \nabla \cdot (n\langle Ev \rangle) + enF \cdot \langle v \rangle - \frac{n\langle E \rangle}{\hat{\tau}_w} , \tag{285} \]

where the trace of the energy relaxation time is:

\[
\frac{1}{\hat{\tau}_w} = \frac{2}{n\langle E \rangle} \int \frac{dk'}{(2\pi)^3} \int \frac{dk}{(2\pi)^3} [E(k) - E(k')] P(k', k) f(r, k, t) . \tag{286} \]

The third-moment equation can be rewritten in the form:

\[
\frac{\partial}{\partial t} n\langle E \rangle = - \nabla \cdot S + j \cdot F - \frac{n\langle E \rangle}{\hat{\tau}_w} , \tag{287} \]

where the ‘energy flux’ \( S = n\langle Ev \rangle \) is the troublesome term involving fourth-moment terms (it is a product of three wavevectors \( k_i k_j k_l \)).

In this form the equation has a clear physical meaning: The kinetic energy of the carriers in a unit volume of real space, \( n\langle E \rangle \), grows with time because of the Joule-heating term \( j \cdot F \) (the acceleration caused by the external field), decreases because of energy losses to the lattice, \( n\langle E \rangle / \hat{\tau}_w \) and because of the flux of energy-density out of the unit volume, \( \nabla \cdot S \).

We now have to find a ‘closure’ (similar to the Einstein relation between \( \mu \) and \( D \) we found for the second moment equation). To do this, let’s decompose the kinetic energy \( \langle E \rangle \) into a ‘drift’ and a ‘thermal’ component:

\[
\langle E \rangle = \frac{1}{2} m^* v^2 = \frac{1}{2} m^* \langle v \rangle^2 + \frac{1}{2} m^* \langle v^2 - \langle v \rangle^2 \rangle = \frac{1}{2} m^* \langle v \rangle^2 + \frac{3}{2} k_B T_c , \tag{288} \]

where what should have been a carrier temperature tensor, \( T_{ij} = m^* \langle (v_i - \langle v_i \rangle)(v_j - \langle v_j \rangle) \rangle / (3k_B) \) has been converted to a scalar, \( T_c \), by taking its trace.

Thus, the energy-flux vector can be decomposed into a ‘drift’ term – related only to a second moment – and a
The ‘heat term’, \( Q \), still containing a third moment:

\[
S = \frac{1}{2} nm^* \langle v \rangle^2 \langle v \rangle + \frac{3}{2} k_B T_c \langle v \rangle + Q , \quad Q = \frac{1}{2} nm^* \langle (v - \langle v \rangle) (v^2 - \langle v \rangle^2) \rangle .
\] (289)

The ‘heat flux’ \( Q \) is finally related to a second-moment (the carrier temperature) via the phenomenological Fourier heat diffusion equation:

\[
Q = - \kappa \nabla T_c ,
\] (290)

where \( \kappa \) is the (electronic) thermal conductivity given by the Wiedermann-Franz law:

\[
\kappa = \left( \frac{5}{2} + c \right) \left( \frac{k_B T_c}{e} \right)^2 ne\mu T_c ,
\] (291)

where \( c \) is a constant whose value is still a subject of research.

- **Hot electrons.**
  Note that in the equations above \( T_c \) is the temperature of the carriers accelerated by the field, so it is in general higher than the lattice temperature \( T \). The relaxation term \( n \langle E \rangle / \hat{\tau}_w \) describes how fast the carrier-system attempts to reach thermal equilibrium with the lattice. For Boltzmann statistics and fields not too strong, one can show that

\[
\frac{\langle E \rangle}{\hat{\tau}_w} \approx \frac{\langle E \rangle - (3/2) k_B T}{\hat{\tau}_w} .
\] (292)

In words, if left alone the energy of the carrier-system decays exponentially to the thermal energy at the lattice temperature with decay constant \( \hat{\tau}_w \). If an external field drives the system, then, ignoring for simplicity the energy-flux term \( \nabla \cdot S \), and ignoring also the drift component of the energy (usually neglected, it matters only in quasi-ballistic transport in very short devices), so that \( \langle E \rangle \approx (3/2) k_B T_c \), at steady-state we have the ‘energy balance’:

\[
0 = \frac{\partial}{\partial t} \langle E \rangle \approx \frac{1}{n} j \cdot F - \frac{\langle E \rangle}{\hat{\tau}_w} \rightarrow \frac{2}{3k_B n} j \cdot F = \frac{T_c - T}{\hat{\tau}_w} ,
\] (293)

from which we see that electrons will always become ‘hot’ under the action of an electric field.

One important side-effect of carrier-heating is the degradation of the mobility and the existence of a ‘saturated
velocity'. We saw before that scattering with acoustic phonons yields an electron mobility in Si which decreases as \( T^{-3/2} \) as the temperature increases. Looking back at the derivation of this result, we see that a factor \( T^{-1} \) stems from the temperature dependence of the scattering rate, since the acoustic-phonon population grows as \( T \) (in the high-\( T \), equipartition limit) so that the frequency of the collisions grows with \( T \). Another factor \( T^{-1/2} \), instead, comes from the integration of the DOS and Boltzmann factors in the Kubo-Greenwood formula. This is related to the electron distribution, so that this \( T \) will become \( T_c \) in the hot electron regime. Therefore, we can see that the electron mobility decreases as

\[
\mu \approx \mu_{th} \left( \frac{T}{T_c} \right)^{1/2},
\]

where \( \mu_{th} \) is the Ohmic mobility (cold, near equilibrium electrons). Similarly, considering optical phonons we would obtain \( \tau_w \propto (T/T_c)^{1/2} \). Using now the energy-balance equation above,

\[
\frac{2}{3k_B n} j \cdot F \approx c (T_c - T) \left( \frac{T}{T_c} \right)^{1/2} \rightarrow \quad e\mu F^2 = c (T_c - T) \left( \frac{T}{T_c} \right)^{1/2} \rightarrow \quad T_c \approx T + c e\mu_{th} F^2,
\]

where \( c \) is some constant. Therefore, as the electric field grows, the hot-electron mobility will decay as:

\[
\mu \approx \mu_{th} \left( \frac{1}{1 + c e\mu_{th} F^2 / T} \right)^{1/2}.
\]

For the electron 'drift' velocity (the velocity along the direction of the applied field) we have:

\[
v_{drift} = \mu F \approx \mu_{th} F \left( \frac{1}{1 + c e\mu_{th} F^2 / T} \right)^{1/2}.
\]

For small \( F \) (Ohmic regime), \( v_{drift} \approx \mu_{th} F \), as usual. But at large \( F \) the drift velocity saturates, \( v_{drift} \rightarrow (c \mu_{th} T/e)^{1/2} = v_{sat} \). The physical reason behind the existence of a saturated velocity is that
electrons scatter more at higher energies, their drift velocity decreases as they are pushed at higher energy, and the mobility decreases.
Generation and recombination processes

Generation and recombination processes are scattering processes like those we've considered before (with phonons and ionized impurities), but exhibit some important characteristics on their own:

- The number of particles is not conserved. For example, absorption of a photon excites an electron from VB to CB, resulting in the creation of an electron-hole pair.
- Two bands (and, often, two bands and an impurity state) are always involved. So, the single-band approximation fails, and, often, the effective mass approximation fails.
- Large energies are involved. This is another reason for the failure of simple models for the band structure and for the effective mass or even envelope approximation. As a result, the calculation of the matrix elements is almost always a daunting task. Almost always one has to proceed empirically and phenomenologically.

We shall now discuss three main classes of generation/recombination processes and consider mainly their dependence on carrier density and temperature (if important):

1. **Radiative**: Absorption of light which creates electron-hole pairs, and recombination of electron-hole pairs via emission of photons.

2. **Auger**: An electron-hole pairs recombine and the excess energy is transferred to another free carrier. The opposite process (a free carrier transfer its energy to an electron in the VB, this generating an electron-hole pair) is the ‘inverse Auger process’. It is often called ‘impact ionization’ in the device community and literature.

3. **Shockley-Read-Hall (SRH)**: A trap (an impurity energy level deep in the gap) acts as a ‘catalyst’ for generation/recombination processes (usually non-radiative) of electron-hole pairs.
DIRECT TRANSITION

INDIRECT TRANSITION
(phonon or impurity assisted)
Radiative processes.
Radiative processes consist either in the absorption of light (a photon of energy $\hbar \omega > E_{\text{gap}}$, where $E_{\text{gap}}$ is the direct gap of the semiconductor), resulting in the generation of an electron-hole pair, or in the recombination of an electron-hole pair with the emission of a photon.

To first order in the electromagnetic field, the Hamiltonian describing the interaction of an electron with the electromagnetic field is, written in terms of the vector potential, $A(r, t) = A_0 e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)}$:

$$ H_{\text{phot}} = \frac{i e \hbar}{m c} \mathbf{A} \cdot \nabla . $$

The matrix element between two states in the valence and conduction bands thus has the form:

$$ \langle \mathbf{k}, c | H_{\text{phot}} | \mathbf{k}, v \rangle \approx \frac{i e \hbar}{m c} \int_V d\mathbf{r} \ e^{-i \mathbf{k}_c \cdot \mathbf{r}} u_{\mathbf{k}_c}^{\ast}(\mathbf{r}) e^{-\mathbf{q} \cdot \mathbf{r}} A_0 \cdot \nabla e^{i \mathbf{k}_v \cdot \mathbf{r}} u_{\mathbf{k}_v}(\mathbf{r}) . $$

If the wavelength of the photon is much longer than the electron wavelength, we can set $e^{-i \mathbf{q} \cdot \mathbf{r}} \approx 1$ and

$$ \langle \mathbf{k}, c | H_{\text{phot}} | \mathbf{k}, v \rangle \approx \frac{i e \hbar}{m c} A_0 \cdot \int_V d\mathbf{r} \ e^{-i \mathbf{k}_c \cdot \mathbf{r}} u_{\mathbf{k}_c}^{\ast}(\mathbf{r}) \nabla e^{i \mathbf{k}_v \cdot \mathbf{r}} u_{\mathbf{k}_v}(\mathbf{r}) . $$

This is called the ‘dipole approximation’. Now:

$$ \int_V d\mathbf{r} \ e^{-i \mathbf{k}_c \cdot \mathbf{r}} u_{\mathbf{k}_c}^{\ast}(\mathbf{r}) \nabla e^{i \mathbf{k}_v \cdot \mathbf{r}} u_{\mathbf{k}_v}^{\ast}(\mathbf{r}) = $$

$$ = \int_V d\mathbf{r} \ e^{i \mathbf{k}_v - \mathbf{k}_c) \cdot \mathbf{r}} u_{\mathbf{k}_c}^{\ast}(\mathbf{r}) \nabla u_{\mathbf{k}_v}(\mathbf{r}) + \int_V d\mathbf{r} \ e^{-i \mathbf{k}_c \cdot \mathbf{r}} u_{\mathbf{k}_c}^{\ast}(\mathbf{r}) u_{\mathbf{k}_v}(\mathbf{r}) \nabla e^{i \mathbf{k}_v \cdot \mathbf{r}} . $$

The Bloch functions $u_{\mathbf{k}_c}(\mathbf{r})$ and $u_{\mathbf{k}_v}(\mathbf{r})$ are orthogonal and the exponential is slowly varying. Therefore the second integral is usually neglected. Now notice that the product $u_{\mathbf{k}_c}^{\ast}(\mathbf{r}) \nabla u_{\mathbf{k}_v}(\mathbf{r})$ is the same in each unit
cell and it is multiplied by a slowly varying function which we can take outside the integral. Thus:

\[
<k, c | H_{phot} | k, v> \approx \frac{i e \hbar}{m c} A_0 \cdot \sum_{all cells} e^{i (k_v - k_c) \cdot r} \int_{\Omega} dr \, u_{k_c}^{c^*} (r) \nabla u_{k_v}^v (r) = \tag{303}
\]

\[
\frac{i e \hbar}{m c} A_0 \cdot \int_V dr \, e^{i (k_v - k_c) \cdot r} \frac{1}{\Omega} \int_{\Omega} dr \, u_{k_c}^{c^*} (r) \nabla u_{k_v}^v (r) . \tag{304}
\]

Now let’s define the ‘momentum matrix element’ between two Bloch functions:

\[
<c | p | v> = -\frac{i \hbar}{\Omega} \int_{\Omega} dr \, u_{k_c}^{c^*} (r) \nabla u_{k_v}^v (r) . \tag{305}
\]

Then:

\[
<k, c | H_{phot} | k, v> \approx -\frac{e}{m c} A_0 \cdot <c | p | v> \int_V dr \, e^{i (k_v - k_c) \cdot r} . \tag{306}
\]

From this we see that the optical matrix element vanishes unless \( k_c = k_v \). Only ‘vertical (or direct) transitions’ can happen.

Transitions which involve a change in (crystal) momentum can only happen at higher orders in perturbation theory, requiring extra momentum which must be provided by phonons, traps, etc.

- **Auger (and inverse Auger) processes.**
  
  An Auger process is a process in which an electron and a hole recombine transferring the excess energy to another free carrier (rather than to a photon, as in radiative processes). The inverse process, ‘impact ionization’, consists of an electron or hole with kinetic energy larger than \( E_G \) (the indirect gap), transferring its energy to an electron in the valence band, exciting it across the gap, thus generating an electron-hole pair.

  The matrix elements for these processes are rather complicated expressions involving the the Coulomb interactions between carriers. We shall only note that – as shown in the figures – the ‘impact ionization rate’ has a threshold at (or slightly above) \( E_G \), and it grows very fast with electron or hole energy. In device physics, in the spirit of the drift-diffusion approximation, usually we refer to an ‘ionization coefficient’ \( \alpha (F) \), function of the electric field. It expresses the number of pairs generated per unit length by a carrier at steady-state in a uniform electric field.
field $F$. It is usually approximated by the expression:

$$\alpha(F) = \frac{eF_{th}}{E_{th}} e^{-F_{th}/F}, \tag{307}$$

where $E_{th}$ is some empirical threshold energy and $F_{th}$ (with dimensions of an electric field) is an empirical constant. Given the high kinetic energy involved in this process, calculations of the inverse Auger process must go beyond the effective-mass, parabolic-band approximation and must be performed solving the full BTE, often using Monte Carlo techniques. Notice that phonon scattering is reduced at lower lattice temperatures. Therefore, carriers will get hotter and impact-ionize more easily at lower temperatures.
Shockley-Read-Hall (SRH) processes.

SRH process are processes in which a trap (that is: an ionic impurity with an electronic state whose energy lies in the band-gap) can capture a free electron in the CB and the trapped electron may later recombine with a hole in VB. The opposite process (a hole is captured by the trap and later recombines with an electron in the CB) can also occur.

The capture rate at which each free carrier is captured by one trap can be expressed in terms of the capture coefficients, $c_n$ and $c_p$, for electrons and holes, respectively, as:

\[
\text{electron capture rate per trap per electron} = n c_n = n v_{th}^{(n)} \sigma_n
\]

\[
\text{hole capture rate per trap per hole} = p c_p = p v_{th}^{(p)} \sigma_p
\]

where $v_{th} = (3k_B T/m^*)^{1/2}$ is the thermal velocity of the carriers and $\sigma$ the ‘capture cross section’ (in cm$^2$) of the trap.

The ‘emission rate’ is usually denoted by $e_n$ and $e_p$ and it depends exponentially on the energy $E_T$ of the trap state:

\[
e_{n,p} \propto e^{E_T/(k_B T)}.
\]

Generation and recombination rates.

Generation and recombination rates must be dealt with as additional ‘scattering’ terms in the collision term of the BTE. Note, however, that so far we have written this term in the non-degenerate limit. Now we must pay attention to Pauli’s exclusion principle and write (in the limit of continuous $k$, so converting $\sum k'$ to an integral):

\[
\left. \frac{\partial f}{\partial t} \right|_{GR} = \frac{V}{(2\pi)^3} \int dk' \left\{ P(k, k') f(r, k', t)[1 - f(r, k, t)] - P(k', k) f(r, k, t)[1 - f(r, k', t)] \right\}.
\]

Let’s simplify our lives by assuming that all functions depend only on the energy instead of on the $k$-vectors, and introduce the DOS in the calculations.
1. **Recombination rate.**

If we have free electrons in the CB, the density of full states in the energy interval \((E_c, E_c + dE_c)\) will be:

\[
\text{full states in the CB} = f_c(E_c) \rho_c(E_c) \, dE_c. \quad (311)
\]

Similarly, the density of empty states (that is: full hole states) in the energy interval \((E_v, E_v + dE_v)\) in the VB will be:

\[
\text{full states in the VB} = [1 - f_v(E_v)] \rho_v(E_v) \, dE_v. \quad (312)
\]

Therefore, denoting by \(E^0_v\) the top of the VB, the electron recombination term in the collisional term above will be:

\[
\int_{-\infty}^{E^0_v} dE_v \, f_c(E_c) \ [1 - f_v(E_v)] \rho_v(E_v) \ P(E_v, E_c). \quad (313)
\]

Let’s rewrite this term as follows: Note that the product of the electron and hole density can be written as:

\[
n \ p = \int_{E^0_c}^{\infty} dE_c \int_{-\infty}^{E^0_v} dE_v \ f_c(E_c) \ [1 - f_v(E_v)] \rho_v(E_v) \rho_c(E_c). \quad (314)
\]

Thus the recombination rate can be written as:

\[
R = n \ p \ \langle P(E_v, E_c) \rangle, \quad (315)
\]

where the ‘average recombination rate’ is defined as:

\[
\langle P(E_v, E_c) \rangle = \frac{\int_{E^0_c}^{\infty} dE_c \int_{-\infty}^{E^0_v} dE_v \ f_c(E_c) \ [1 - f_v(E_v)] \rho_v(E_v) \rho_c(E_c) \ P(E_v, E_c)}{\int_{E^0_c}^{\infty} dE_c \int_{-\infty}^{E^0_v} dE_v \ f_c(E_c) \ [1 - f_v(E_v)] \rho_v(E_v) \rho_c(E_c)}. \quad (316)
\]

We see that the Fermi levels in \(f_c\) and \(f_v\) cancel, so that this average does not depend on \(n\) or \(p\).
2. Generation rate.

Let’s now consider the opposite process, the generation of an electron-hole pair by the same physical mechanism. All we have to do is to exchange $E_c$ and $E_v$ in the previous discussion, so that:

$$G = \int_{E_c^0}^{\infty} dE_c \int_{-\infty}^{E_v^0} dE_v f_v(E_v) [1 - f_c(E_c)] \rho_v(E_v) \rho_c(E_c) P(E_c, E_v). \quad (317)$$

Usually one can neglect $f_c$ compared to 1 and assume $f_v(E_v) \approx 1$, so that

$$G \approx \int_{E_c^0}^{\infty} dE_c \int_{-\infty}^{E_v^0} dE_v \rho_v(E_v) \rho_c(E_c) P(E_c, E_v), \quad (318)$$

which does not depend on carrier density (unless we are dealing with Auger processes for which $P$ itself depends on the density!). At equilibrium $G = R$, so that, since $G$ does not depend on density, its value must be equal to its equilibrium value:

$$G = n_i^2 \langle P(E_v, E_c) \rangle. \quad (319)$$

3. SRH rates.

Let’s consider $N_{TT}$ traps, identical and non-interacting. At equilibrium we have $n_T = N_{TT} f_T$ full traps, where $f_T$ is the probability of having a full trap.

The recombination (capture) term in the collision piece of the BTE will be:

$$\sum_{traps} f_c(E_c)(1 - f_T)P(E_T, E_c) = N_{TT} f_c(E_c)(1 - f_T)P(E_T, E_c). \quad (320)$$

Thus, summing over all full electron states in CB:

$$R = N_{TT} \int_{E_c^0}^{\infty} dE_c f_c(E_c) (1 - f_T) \rho_c(E_c) P(E_T, E_c) = \quad (321)$$
\begin{align}
\langle P(E_T, E_c) \rangle &= \frac{\int_{E_c}^{\infty} dE_c f_c(E_c) \rho_c(E_c) P(E_T, E_c)}{\int_{E_c}^{\infty} dE_c f_c(E_c) \rho_c(E_c)}.
\end{align}

as before and having defined the ‘capture coefficient’ \( c_n = \langle P(E_T, E_c) \rangle \). We can finally rewrite \( R \) as

\begin{align}
R &= (N_{TT} - n_T)nc_n.
\end{align}

Similarly, we can write for the generation rate:

\begin{align}
G &= N_{TT} \int_{E_c}^{\infty} dE_c f_T [1 - f_c(E_c)] \rho_c(E_c) P(E_c, E_T),
\end{align}

and, assuming as usual \( f_c << 1 \),

\begin{align}
G &= n_T e_n,
\end{align}

where \( e_n \) is independent of density.

4. Rate equations.

Finally, we are ready to derive the SRH rate equations. The net generation/recombination rate will be:

\begin{align}
\text{Net GR rate} &= n_T e_n - (N_{TT} - n_T)nc_n = G_n - R_n.
\end{align}

Including also holes – so far neglected but obeying similar equations – the total change in the density of full traps will be:

\begin{align}
\frac{\partial n_T}{\partial t} &= R_n - G_n - R_p + G_p.
\end{align}
At steady state
\[ c_n n (N_{TT} - n_T) - n_T e_n = c_p p n_T - (N_{TT} - n_T) e_p . \] (329)

Solving this equation for \( n_T \) we get:
\[ n_T = \frac{N_{TT} (c_n n + e_p)}{c_n n + e_n + c_p p + e_p} . \] (330)

So, the net electron recombination at steady state will be:
\[ U_s = R_n - G_n = \frac{c_n c_p N_{TT} n p - e_n e_p N_{TT}}{c_n n + e_n + c_p p + e_p} . \] (331)

At equilibrium when the net rate vanishes,
\[ c_n c_p n p = e_n e_p \rightarrow \frac{e_n e_p}{c_n c_p} = n_i^2 , \] (332)

and, if we can assume that \( e_n, e_p, c_n, \) and \( c_p \) assume always the equilibrium values (not always true!), then:
\[ U_s = c_n c_p N_{TT} \frac{n p - n_i^2}{c_n n + e_n + c_p p + e_p} . \] (333)

Using the equilibrium carrier densities \( n_0 \) and \( p_0 \) in the denominator, we can define a constant \( \bar{c} \) as follows:
\[ \bar{c} = c_n c_p N_{TT} \frac{1}{c_n n_0 + e_n + c_p p_0 + e_p} , \] (334)

so that
\[ U_s = \bar{c} \left( n p - n_i^2 \right) . \] (335)
The constant $\tau_n = \bar{c}p_0$ is called the ‘electron lifetime’. Indeed, if we generate a small perturbation of the electron density, $n = n_0 + \delta n$, then $np_0 = (n_0 + \delta n)p_0 = n_0 p_0 + p_0 \delta n$ and

$$U_s = \bar{c} p_0 \delta n = \frac{\delta n}{\tau_n}. \quad (336)$$

Several alternative equivalent expressions can be found for the net recombination rate $U_s$. For example, noting that at equilibrium

$$c_n n (N_{TT} - n_T) = \sigma_n v_{th}^{(n)} N_{TT} \left[1 - f_T(E_T)\right] = e_n N_{TT} f_T(E_T), \quad (337)$$

that whenever the trap occupation is dictated by a Fermi function

$$f_T(E_T) = \frac{1}{1 + e^{(E_T-E_F)/(k_B T)}}, \quad (338)$$

and using the Boltzmann approximation $n = n_i \exp[(E_F - E_i)/(k_B T)]$, we get

$$e_n = v_{th}^{(n)} \sigma_n n_i e^{(E_T-E_i)/(k_B T)}, \quad (339)$$

and similarly for $e_p$, so that, assuming that $v_{th}^{(n)} \approx v_{th}^{(p)} = v_{th}$:

$$U_s = \frac{\sigma_n \sigma_p v_{th} N_{TT} (pn - n_i^2)}{\sigma_n [n + n_i e^{(E_T-E_i)/(k_B T)}] + \sigma_p [p + n_i e^{(E_i-E_T)/(k_B T)}]}. \quad (340)$$

This expression shows that traps whose energy is close to mid-gap $E_T \approx E_i$ are the most efficient recombination centers (since the denominator is minimized for $E_T = E_i$). Colinge and Colinge list the many possible simplifications one can produce from this expression.
1. **Decay of photoexcited carriers.**
   In $p$-type Si, light of energy larger than the gap excites electron-hole pairs uniformly distributed throughout the illuminated sample. Let $G$ be the generation rate due to the light. Let’s assume that the light source is weak enough so that the hole concentration, $p$, remains essentially the concentration $p_0$ in the absence of light. Thus the electric field, mainly due to the hole charge, vanishes ($F = 0$) and we have a uniform electron distribution, $\partial n/\partial x = 0$. The continuity equation (the first of Eqns. (280), page 83) now reads:

   \[
   \frac{\partial n}{\partial t} = G - U_s ,
   \]

   (341)

   since in a uniform situation the divergence of the current vanishes. Using

   \[
   U_s = \frac{\delta n}{\tau_n} = \frac{n - n_0}{\tau_n}
   \]

   (342)

   above, we see that the steady-state solution is

   \[
   n = n_0 + \tau_n G ,
   \]

   (343)

   where $n_0$ is the electron concentration in the sample in the absence of light. Let’s now turn off the light source at $t = 0$. Then:

   \[
   \frac{\partial n}{\partial t} = - \frac{n - n_0}{\tau_n} ,
   \]

   (344)

   The solution is:

   \[
   n = n_0 + \tau_n G e^{-t/\tau_n} .
   \]

   (345)

2. **Steady-state injection from one side.**
   Let’s consider now the steady-state injection of carriers from one side at $x = 0$ (as if we were illuminating the
sample from the left). The continuity equation (the first of Eqns. (280), page 83) now reads:

\[
\frac{\partial n}{\partial t} = \frac{1}{e} \frac{\partial j_n}{\partial x} - \frac{n - n_0}{\tau_n},
\]

(346)

while electron current continuity (the first of Eqns. (281), page 83) reads:

\[
j_n = e D_n \frac{\partial n}{\partial x},
\]

(347)

having taken the diffusion constant \( D_n \) outside the gradient, since it is independent of position. Combining these two equations we get a ‘diffusion equation’:

\[
\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} - \frac{n - n_0}{\tau_n},
\]

(348)

with boundary conditions \( n(x = 0) = n(0) \) (some constant, the steady-state electron concentration at the illuminated side) while \( n(x \to \infty) = n_0 \), the (minority) electron concentration in the sample in the absence of external generation. At steady state \( (\partial n/\partial t = 0) \) the solutions is:

\[
n(x) = n_0 + [n(0) - n_0] e^{-x/L_n},
\]

(349)

where \( L_n = (D_n \tau_n)^{1/2} \) is the ‘diffusion length’.

If now we extract the electrons from a contact at \( x = W \), then we must replace the boundary condition \( n(x \to \infty) = n_0 \) with \( n(W) = n_0 \) and the solution becomes:

\[
n(x) = n_0 + [n(0) - n_0] \frac{\sinh \left( \frac{W-x}{L_n} \right)}{\sinh(W/L_n)}.
\]

(350)
The current density at $W$ will be

$$j_n(W) = e D_n \frac{\partial n}{\partial x} \bigg|_W = -e [n(0) - n_0] \frac{D_n}{L_n} \frac{1}{\sinh(W/L_n)}.$$  \hspace{1cm} (351)

3. **Transient and steady-state diffusion.**

When shining a laser pulse on a sample, carriers are generated optically in a small region of the sample. Their transport after the short pulse is described by the following diffusion equation which can be derived from the DDEs (drift-diffusion equations, Eqns. (280) and (281) on page 83 coupled to the GR terms) as done above:

$$\frac{\partial n}{\partial t} = -n - n_0 \frac{\tau_n}{\mu_n F} + D_n \frac{\partial^2 n}{\partial x^2}.$$  \hspace{1cm} (352)

If no external field is applied to the sample, then the solution is:

$$n(x, t) = \frac{N}{(4\pi D_n t)^{1/2}} \exp \left( -\frac{x^2}{4D_n t} - \frac{t}{\tau_n} \right) + n_0,$$  \hspace{1cm} (353)

where $N$ is the total number of pairs per unit area generated by the pulse. If an external field is applied, then the solution is given by the same expression but with $x$ replaced by $(x + \mu_n F t)$: The whole Gaussian ‘packet’ of carriers simply shifts along the negative $x$-axis with velocity $\mu_n F$.

4. **Surface recombination.**

We have described above generation and recombination processes occurring in the bulk of the semiconductor, often mediated by traps states. Surfaces, however, usually present a higher density of defects – and so of trap states – because the termination of the crystal often leaves broken ionic bonds with electronic states whose energy may fall within the gap. The SRH processes mediated by these surface traps can be described in a way completely analogous to what we have done for bulk SRH states.

Let’s now consider our $p$-type sample and assume that at the left edge ($x = 0$) we have a density $N_{st}$ (per
unit area) surface traps with an associated ‘electron surface recombination velocity’

\[ S_n = \sigma_n v_{th}^{(n)} N_{st}. \]  (354)

This means that carriers reaching the left edge will recombine there, that is

\[ eD_n \frac{\partial n}{\partial x} = eS_n [n(0) - n_0]. \]  (355)

In words: The in-flux of electrons carried by the diffusion current at the left-edge must be equal to carrier out-flux due to their recombination. Then we expect that as we approach the left-edge of the sample the electron density will begin to drop from its bulk value \( n_0 \) as electrons disappear at the edge. The equation describing this situation is

\[ \frac{\partial n}{\partial t} = G - \frac{n - n_0}{\tau_n} + D_n \frac{\partial^2 n}{\partial x^2}, \]  (356)

whose solutions, with the boundary condition above, is:

\[ n(x) = n_0 + \tau_n G \left[ 1 - \frac{\tau_n S_n e^{-x/L_n}}{L_n + S_n} \right]. \]  (357)

For \( S_n \to 0 \) this expression tends to \( n_0 + \tau_n G \), result we have already obtained above. When the surface recombination velocity is very fast, then \( n(x) \to n_0 + \tau_N G [1 - \exp(-x/L_n)] \), so that the minority carrier concentration at the surface, \( n(0) \), approaches its equilibrium value as if the optical generation did not occur.