

Solid-State Electronics

EE230C

Class Notes for EE230C taught by Prof Sayeef Salahuddin
Scribed by Yen Kai Lin in 2018
Further Edited and Augmented by Pratik Brahma in 2023

August 28, 2023

Contents

1	General Considerations of Electronic Transport	5
1.1	Ohm's Law	5
1.2	Landauer's Formula	6
1.3	Broadening and Maximum Conductance	10
1.4	Electrostatics	11
2	Density of States	15
2.1	Schrodinger Equation	15
2.1.1	Time-Independent Schrodinger Equation	15
2.1.2	Formalism in Quantum Mechanics	17
2.1.3	Basis Transformation	19
2.2	Bloch's Theorem	21
2.3	Band Structure	23
2.3.1	One-Dimensional Monatomic Lattice	23
2.3.2	One-Dimensional Diatomic Lattice	25
2.3.3	Graphene	26
2.3.4	Graphene Nanotube	29
2.4	Density of States	31
2.5	Effective Mass Equation	32
2.5.1	Dimensionality	33
2.5.2	Maximum Conductance	34
2.6	Schrodinger's Equation and Finite Difference Method	35
2.6.1	Current Evaluation	36
2.6.2	Capacitance in State-of-The-Art Transistor	39
2.7	Slater-Koster Tight-Binding Method	41
3	Scattering	45
3.1	Free Electron	45
3.2	Electron Experiencing Potential	46

3.3	Master Equation	49
3.4	Relaxation Time	50
3.5	Phonons	51
3.5.1	Electron-Phonon Scattering	52
3.5.2	Simplification of Electron-Phonon Scattering Using Effective Mass Approximation	53
4	Semiclassical Transport	57
4.1	Boltzmann Transport Equation	57
4.2	Simplification of BTE	59
5	Transport in Small MOSFETs	61
5.1	Ballistic Transport	61
5.2	Transport with Scatterings (Quasi-Ballistic Transport)	62
5.3	Generalized Ballistic Transport in One Dimension	63
5.4	Generalized Quasi-Ballistic Transport	66
5.5	Anisotropic Effective Mass in Ballistic Transport	67
5.6	Issues of Scaling MOSFETs	68
5.7	Application of the Developed Current Model	70

Chapter 1

General Considerations of Electronic Transport

1.1 Ohm's Law

Ohm's law describes the relationship between the current (I) and applied voltage (V) or electric field (E) in a conductor and it can be expressed as

$$I = \frac{V}{R} \quad (1.1)$$

where R is the resistance and is a function of the size of conductor

$$R = \frac{\rho L}{A} = \frac{L}{A} \frac{1}{\sigma} \quad (1.2)$$

Hence, Ohm's law can be rewritten as

$$I = \frac{V \sigma A}{L} \Rightarrow J = \frac{I}{A} = V \frac{\sigma}{L} = \sigma E \quad (1.3)$$

σ is the conductivity and it's related to the mobility ($\mu = \frac{e\tau}{m^*}$).

$$\sigma = \frac{e^2 n \mu}{m^*} \quad (1.4)$$

From Equation (1.1) to (1.4), one may find that if L is very small or τ (scattering or collision time) approaches to ∞ (happens when the temperature is close to zero), then the conductivity or current (density) would become very huge without any bound. Therefore, the question is: in a very small device, how does the current look like? The experimental result¹ has shown that the conductance of a point contact is quantized at a low temperature, which seems contradict Ohm's law shown above. To explore this issue, Landauer's formula will be discussed in the next section.

¹Quantized Conductance of Point Contacts in a Two-Dimensional Electron Gas, B. J. van Wees et al., PRL, 1988

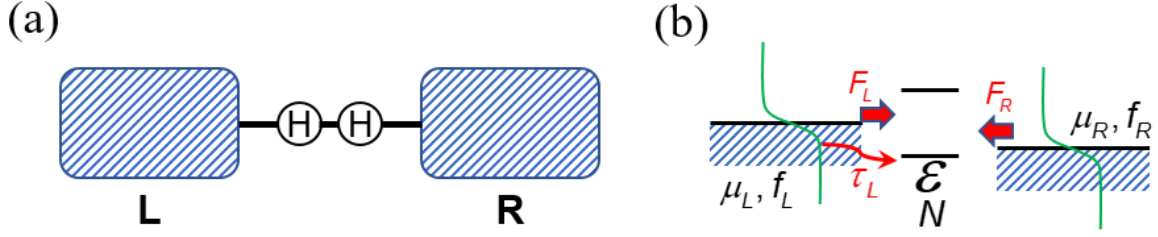


Figure 1.1: (a) Schematic diagram of the "hydrogen molecular" device with left (L) and right (R) contact. (b) Energy diagram when applying a bias $V = \frac{\mu_R - \mu_L}{e}$. $F_{L,R}$ and $f_{L,R}$ are the electron flux and Fermi function. The green curves indicate the probability distribution of finding an electron (Fermi distribution).

1.2 Landauer's Formula

Consider a device consisting of a hydrogen molecular (with two energy levels) connected between two contacts as shown in Fig. 1.1. When applying a bias V , the electrons are allowed to flow through the hydrogen molecular assuming no scattering. Note that in hydrogen molecular the Fermi function (distribution) is not applicable because the amount of electrons is not sufficient to be statistically meaningful². However, the contacts (like electron reservoirs) are assumed to be "large" enough to support the Fermi function. The equation of states (continuity equation) at a given energy ε can be written as

$$\frac{dN^\varepsilon}{dt} = F_L^\varepsilon + F_R^\varepsilon \quad (1.5)$$

where N is the occupation factor and the fluxes are

$$F_L^\varepsilon = \frac{f_L^\varepsilon - N^\varepsilon}{\tau_L^\varepsilon} \quad (1.6)$$

$$F_R^\varepsilon = \frac{f_R^\varepsilon - N^\varepsilon}{\tau_R^\varepsilon} \quad (1.7)$$

Hence, Equation (1.5) becomes³

$$\frac{dN}{dt} + N \left(\frac{1}{\tau_L} + \frac{1}{\tau_R} \right) = \frac{f_L}{\tau_L} + \frac{f_R}{\tau_R} \quad (1.8)$$

Now, we define⁴

$$\frac{1}{\tau_{L,R}} = \frac{\gamma_{L,R}}{\hbar} \quad (1.9)$$

²Ashley H. Carter, Classical and Statistical Thermodynamics.

³The superscript ε is omitted for simplicity.

⁴The physical meaning will be discussed later.

where γ has the unit of energy. Thus, we get

$$\hbar \frac{dN}{dt} + N(\gamma_L + \gamma_R) = \gamma_L f_L + \gamma_R f_R \quad (1.10)$$

At steady state, we have

$$N = \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} \quad (1.11)$$

Therefore, the current can be expressed as

$$\begin{aligned} I &= eF_L = e \frac{f_L - N}{\tau_L} \\ &= \frac{e}{\hbar} \gamma_L \left(f_L - \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} \right) \\ &= \frac{e}{\hbar} \frac{\gamma_L^\varepsilon \gamma_R^\varepsilon}{\gamma_L^\varepsilon + \gamma_R^\varepsilon} (f_L^\varepsilon - f_R^\varepsilon) \end{aligned} \quad (1.12)$$

From Equation (1.12), one may directly observe that the current only depends on the properties of the contacts. Now if there are multiple energy levels, the current becomes

$$I = \sum_{\varepsilon_i} \frac{e}{\hbar} \frac{\gamma_L^\varepsilon \gamma_R^\varepsilon}{\gamma_L^\varepsilon + \gamma_R^\varepsilon} (f_L^\varepsilon - f_R^\varepsilon) \times 2 \quad (1.13)$$

where 2 is coming from the spin degeneracy. Use the following trick

$$g(\varepsilon_i) = \int g(E) \delta(E - \varepsilon_i) \quad (1.14)$$

and the Equation (1.13) can be rewritten as

$$\begin{aligned} I &= \int dE \left(\frac{2e}{\hbar} \right) \frac{\gamma_L^E \gamma_R^E}{\gamma_L^E + \gamma_R^E} (f_L^E - f_R^E) \sum_{\varepsilon_i} \delta(E - \varepsilon_i) \\ &= \frac{2e}{\hbar} \int dE \times D(E) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} (f_L - f_R) \end{aligned} \quad (1.15)$$

where Equation (1.15) is **Landauer's formula** and $D(E)$ is the density of states (DOS).

$$D(E) = \sum_{\varepsilon_i} \delta(E - \varepsilon_i) \quad (1.16)$$

The general definition of DOS is the number of states per unit energy, so $D(E)$ also can be written as

$$D(E) = \frac{\partial N_E^{CS}}{\partial E} = \lim_{\Delta E \rightarrow 0} \frac{N_E^{CS}(E + \Delta E) - N_E^{CS}(E)}{\Delta E} \quad (1.17)$$

where N_E^{CS} is⁵

$$N_E^{CS} = \sum_{\varepsilon_i} \theta(E - \varepsilon_i) \quad (1.18)$$

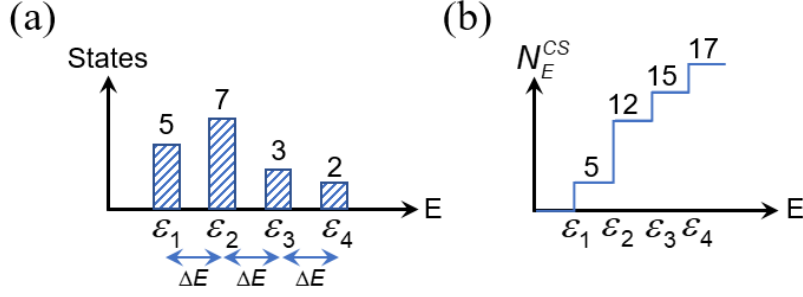


Figure 1.2: (a) Number of states distribution. (b) Cumulative sum of states.

where $\theta(E - \varepsilon_i)$ is the Heaviside step function with the implicit degeneracy as shown in Fig. 1.2. For example, at energy ε_1 there are 5 states [Fig. 1.2(a)]. As going from low energy to high energy, the cumulative sum of states increases from 5 to 17 [Fig. 1.2(b)]. Thus, the Equation (1.17) becomes

$$D(E) = \frac{\partial N^{CS}}{\partial E} = \sum_{\varepsilon_i} \delta(E - \varepsilon_i) \quad (1.19)$$

where $\delta(E - \varepsilon_i)$ is the delta function. After knowing the DOS, we obtain the occupation factor N

$$\begin{aligned} N &= \sum_{\varepsilon_i} \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} \\ &= \int dE \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} \sum_{\varepsilon_i} \delta(E - \varepsilon_i) \\ &= \int dE \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} D(E) \end{aligned} \quad (1.20)$$

Equation (1.20) is under the non-equilibrium condition. At the equilibrium condition (applied voltage $V = 0$), the Fermi functions and γ of both contacts are the same ($f_L = f_R = f_0$, $\gamma_L = \gamma_R$). Thus, the Equation (1.20) can be re-written as

$$N = \int_{-\infty}^{\infty} D(E) f_0(E) dE \quad (1.21)$$

The Equation (1.21) is commonly used to calculate the carrier concentration in semiconductors.

Now, we would like to figure out the maximum conductance based on Landauer's formula. Two assumptions are made in the following discussion: (1) the applied voltage V is very small and (2) the temperature is fairly low. At small V , the Taylor expansion with the first two terms can be applied to the Fermi functions.

$$f_L = f_0(E - \mu_L) \quad (1.22)$$

⁵Here is the cumulative sum (CS) of states at and below a given energy E .

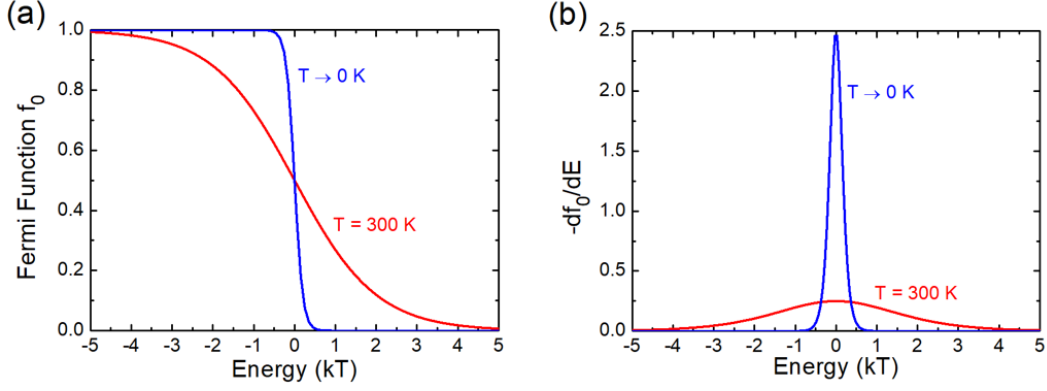


Figure 1.3: (a) Fermi functions and (b) their derivative at different temperatures. The Fermi level is located at $E = 0$.

$$\begin{aligned}
 f_R &= f_0(E - \mu_R) \\
 &= f_0(E + eV - \mu_L) \\
 &= f_0(E - \mu_L) + \left(\frac{\partial f_0}{\partial E} \right) eV
 \end{aligned} \tag{1.23}$$

Therefore, the Landauer's formula becomes

$$\begin{aligned}
 I &= \frac{2e}{\hbar} \int dE \times D(E) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} (f_L - f_R) \\
 &\approx \frac{2e^2}{\hbar} V \int dE \times D(E) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \left(\frac{-\partial f_0}{\partial E} \right)
 \end{aligned} \tag{1.24}$$

Hence, the conductance G is

$$G = \frac{\partial I}{\partial V} = \frac{2e^2}{\hbar} \int dE \times D(E) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \left(\frac{-\partial f_0}{\partial E} \right) \tag{1.25}$$

When the temperature approaches to zero, the derivative of the Fermi function with respect to the energy E is like a delta function (see Fig. 1.3)

$$\left(\frac{-\partial f_0}{\partial E} \right) \rightarrow \delta(E - \mu_L) \tag{1.26}$$

and the conductance G becomes

$$\begin{aligned}
 G &= \frac{2e^2}{\hbar} \int dE \times D(E) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \delta(E - \mu_L) \\
 &= \frac{2e^2}{\hbar} \left[D(E = \mu_L) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \Big|_{E=\mu_L} \right]
 \end{aligned} \tag{1.27}$$

To further examine the Equation (1.27), the concept of "energy broadening" should be introduced.

1.3 Broadening and Maximum Conductance

Due to the wave nature of the quantum objects, the uncertainty principle is inherent in the system we discussed. The uncertainty principle states that the complementary variables, such as position/momentum and energy/time, can not be determined simultaneously⁶.

$$\Delta E \Delta t \geq \frac{\hbar}{2} \quad (1.28)$$

The Δt can be interpreted as "lifetime" of electrons, implying that the ΔE is non-zero which is called energy "broadening." This effect has been seen in Equation (1.9) where γ can be interpreted as the energy broadening. When involving carriers exchange between energy levels (current flowing), these levels are not exact and have spectrum-like distribution. However, how the energy broadening looks like is hard to be experimentally determined, so here the Lorentzian equation⁷ is utilized to get analytic results

$$\delta(E - \varepsilon_i) = \frac{1}{2\pi} \frac{\gamma_L + \gamma_R}{(E - \varepsilon_i)^2 + \left(\frac{\gamma_L + \gamma_R}{2}\right)^2} \quad (1.29)$$

and the Equation (1.29) gives

$$\int_{-\infty}^{\infty} \delta(E - \varepsilon_i) dE = \frac{\gamma_L + \gamma_R}{2\pi} \frac{2}{\gamma_L + \gamma_R} \arctan \frac{E - \varepsilon_i}{\frac{\gamma_L + \gamma_R}{2}} \Big|_{-\infty}^{\infty} = 1 \quad (1.30)$$

Therefore, the Equation (1.27) can be re-written as

$$\begin{aligned} G &= \frac{2e^2}{\hbar} \sum_{\varepsilon_i} \delta(E - \varepsilon_i) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \Big|_{E=\mu_L} \\ &= \frac{2e^2}{\hbar} \sum_{\varepsilon_i} \frac{1}{2\pi} \frac{\gamma_L + \gamma_R}{(\mu_L - \varepsilon_i)^2 + \left(\frac{\gamma_L + \gamma_R}{2}\right)^2} \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \\ &= \frac{2e^2}{\hbar} \sum_{\varepsilon_i} \frac{4\gamma_L \gamma_R}{4(\mu_L - \varepsilon_i)^2 + (\gamma_L + \gamma_R)^2} \\ &\xrightarrow[\mu_L = \varepsilon_i]{\text{maximum}} \frac{2e^2}{\hbar} \sum_{\varepsilon_i} \frac{4\gamma_L \gamma_R}{(\gamma_L + \gamma_R)^2} \end{aligned} \quad (1.31)$$

From the inequality of arithmetic and geometric means⁸, the maximum conductance is

$$G^{max} = \frac{2e^2}{\hbar} \sum_{\varepsilon_i} 1 \equiv \frac{2e^2}{\hbar} M \quad (1.32)$$

where M is the number of modes. The Equation (1.32) indicates that the maximum conductance is actually limited by the material band structure and dimensionality. This is also the reason why the plateaus are observed in the conductance measurement at low temperature.

⁶D. J. Griffiths, Introduction to Quantum Mechanics, Chapter 3.

⁷We would like a function with (1) symmetry and (2) area under the curve = 1.

⁸ $\frac{a+b}{2} \geq \sqrt{ab} \Rightarrow \frac{4ab}{(a+b)^2} \leq 1$

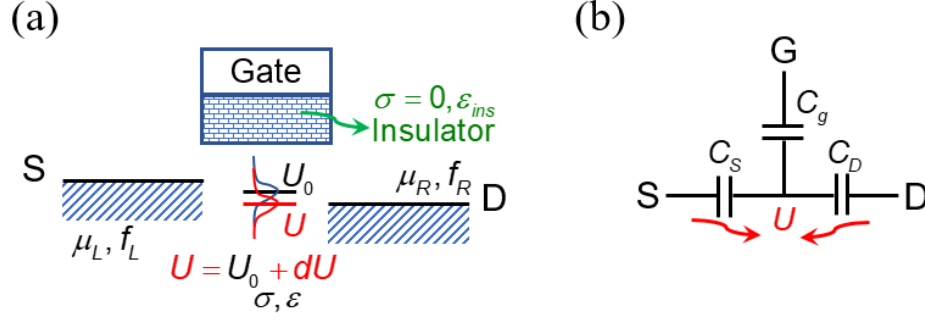


Figure 1.4: (a) Schematic diagram of the gated device. The energy level in the channel is broadened due to the current flow. (b) The equivalent circuit of the gated device. U and U_0 are the energy with and without biases.

1.4 Electrostatics

In the previous sections, the device with two terminals is discussed. Now, let's consider a three-terminal device (like MOSFET) with an insulator layer and a gate contact on the top of the channel as shown in Fig. 1.4. When the gate voltage is applied ($V_g \neq 0, V_S = V_D = 0$), the energy level in the channel is pulled down due to more electrons induced and the change in the electrostatic potential (ϕ) in the channel can be expressed as

$$d\phi = \frac{C_g}{C_g + C_S + C_D} dV_g \quad (1.33)$$

and the occupation factor with (N) and without (N_0) the bias are

$$N = \int_{-\infty}^{\infty} dE \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} D(E - U) \quad (1.34)$$

$$N_0 = \int_{-\infty}^{\infty} dE \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R} D(E - U_0) \quad (1.35)$$

$$\delta N \equiv N - N_0 \quad (1.36)$$

Similarly, if $V_S \neq 0, V_g = V_D = 0$, the potential change is

$$d\phi = \frac{C_S}{C_g + C_S + C_D} dV_S \quad (1.37)$$

and if $V_D \neq 0, V_g = V_S = 0$,

$$d\phi = \frac{C_D}{C_g + C_S + C_D} dV_D \quad (1.38)$$

Finally, if $V_g = V_S = V_D = 0$,

$$d\phi = \frac{dQ}{C_g + C_S + C_D} \quad (1.39)$$

Combining above equations, we obtain

$$d\phi = \alpha_g dV_g + \alpha_D dV_D + \alpha_S dV_S + \frac{dQ}{C_g + C_S + C_D} \quad (1.40)$$

where

$$\alpha_{g,S,D} = \frac{C_{g,S,D}}{C_g + C_S + C_D} \quad (1.41)$$

The Equation (1.40) also can be expressed using energy U

$$\begin{aligned} dU &= \alpha_g U_g + \alpha_D U_D + \alpha_S U_S - \frac{edQ}{C_g + C_S + C_D} \\ &= \alpha_g U_g + \alpha_D U_D + \alpha_S U_S - \frac{e^2 \delta N}{C_g + C_S + C_D} \end{aligned} \quad (1.42)$$

The Equation (1.42) tells us that when increasing the number of electrons (occupation) in the channel, there is an energy penalty which resists the changes. Solving Equation (1.34), (1.35), (1.36), and (1.42) by iterations, the energy U is obtained and the current can be immediately calculated using Landauer's formula

$$I = \frac{2e}{\hbar} \int dE \times D(E - U) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} (f_L - f_R) \quad (1.43)$$

where

$$D(E - U) \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \equiv T(E) \quad (1.44)$$

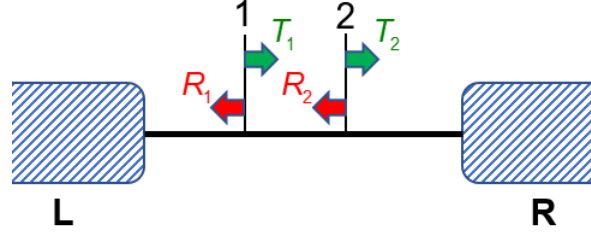
is the transmission probability due to scatterings. Assuming the applied drain-to-source voltage (V) is small and the energy broadening follows Equation (1.29), we have

$$\begin{aligned} I &= \frac{2e}{\hbar} \sum_{\varepsilon_i} \int dE \times \frac{1}{2\pi} \frac{\gamma_L + \gamma_R}{(E - \varepsilon_i)^2 + \left(\frac{\gamma_L + \gamma_R}{2}\right)^2} \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \left(\frac{-\partial f_0}{\partial E} \right) eV \\ &= \frac{2e^2}{h} V \sum_{\varepsilon_i} \underbrace{\int dE \times \frac{\gamma_L \gamma_R}{(E - \varepsilon_i)^2 + \left(\frac{\gamma_L + \gamma_R}{2}\right)^2} \left(\frac{-\partial f_0}{\partial E} \right)}_{\equiv \bar{T}_{\varepsilon_i}} \\ &= \frac{2e^2}{h} V \bar{M} \cdot \bar{T}_{\varepsilon_i} \end{aligned} \quad (1.45)$$

where \bar{M} is the average number of modes around the Fermi energy. The resistance is

$$R = \frac{\partial V}{\partial I} = \frac{h}{2e^2 \bar{M}} \cdot \frac{1}{\bar{T}_{\varepsilon_i}} \quad (1.46)$$

Thus, the question is: how to determine \bar{T}_{ε_i} ? Let's consider two scattering events in the channel as shown in Fig. 1.5. The total transmission T is

Figure 1.5: Two scattering events 1 and 2 with reflection R and transmission T .

$$\begin{aligned}
 T &= T_1 T_2 + T_1 R_2 R_1 T_2 + T_1 T_2 (R_2 R_1)^2 + \dots \\
 &= T_1 T_2 \left(1 + R_1 R_2 + (R_2 R_1)^2 + \dots \right) \\
 &= \frac{T_1 T_2}{1 - R_1 R_2} = \frac{T_1 T_2}{1 - (1 - T_1)(1 - T_2)} = \frac{T_1 T_2}{T_1 + T_2 - T_1 T_2} \\
 &\Rightarrow \frac{1}{T} = \frac{T_1 + T_2 - T_1 T_2}{T_1 T_2} \\
 &\Rightarrow \frac{1 - T}{T} = \frac{T_1 + T_2 - 2T_1 T_2}{T_1 T_2} \\
 &= \frac{1 - T_1}{T_1} + \frac{1 - T_2}{T_2}
 \end{aligned} \tag{1.47}$$

Therefore, if there are N scattering events and the channel length is L , the total transmission T probability is

$$\frac{1 - T}{T} = \sum_i^N \frac{1 - T_i}{T_i} \tag{1.48}$$

Assume the transmission probability of each scattering event is identical ($T_1 = T_2 = \dots = T_k$), we have

$$\begin{aligned}
 \frac{1 - T}{T} &= \frac{N(1 - T_k)}{T_k} \\
 \Rightarrow \frac{1}{T} &= \frac{N(1 - T_k) + T_k}{T_k}
 \end{aligned} \tag{1.49}$$

Define

$$\nu \equiv \frac{N}{L} \tag{1.50}$$

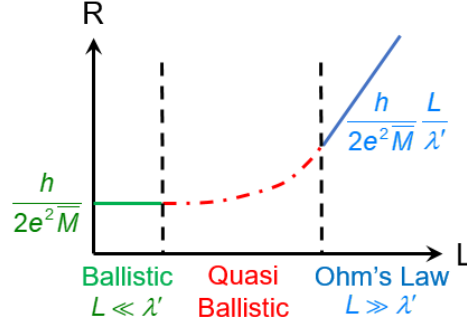


Figure 1.6: Resistance R versus channel length L with three regions: ballistic, quasi-ballistic, and Ohm's law.

The Equation (1.49) becomes

$$\begin{aligned}
 \frac{1}{T} &= \frac{\nu L (1 - T_k) + T_k}{T_k} \\
 &= \frac{L + \frac{T_k}{\nu(1-T_k)}}{\frac{T_k}{\nu(1-T_k)}} \\
 &= \frac{L + T_k \lambda}{T_k \lambda} \\
 &= \frac{L + \lambda'}{\lambda'}
 \end{aligned} \tag{1.51}$$

where

$$\nu (1 - T_k) \equiv \frac{1}{\lambda} \tag{1.52}$$

and

$$\lambda' \equiv T_k \lambda \tag{1.53}$$

Here, λ and λ' are the mean free path, and $\lambda \approx \lambda'^9$. Therefore, the resistance becomes

$$R = \frac{\partial V}{\partial I} = \frac{h}{2e^2 M} \cdot \left(1 + \frac{L}{\lambda'}\right) \tag{1.54}$$

Fig. 1.6 shows the resistance versus the channel length. At $L \gg \lambda'$ limit, the Equation (1.54) reduces back to Ohm's law as mentioned in Section 1.1. At $L \ll \lambda'$ limit (ballistic limit without any scattering), the resistance is governed by $\frac{h}{2e^2 M}$ which sets the upper limit of the conductivity. Between above two regions, it's quasi-ballistic region where is the main focus of the discussion in next few chapters.

⁹For one-dimensional channel, $T_k \approx 1$. Otherwise, the electrons will be localized.

Chapter 2

Density of States

2.1 Schrodinger Equation

Although the classical physics was well developed in 1900s, the black body radiation implicitly indicated that something is missing in the classical physics. From classical physics, the more heat putting in an object would increase the kinetic energy of the object, and thus the energy of the emitted radiation should increase with the frequency which is known as "ultraviolet catastrophe." However, this contradicted with the experimental results that the radiation intensity tends to be zero at short wavelength (higher frequency) regime. To explain this black body radiation, the (implicit) emitted energy quantization and concept of photon (particle of light) were proposed by Planck in 1900¹. After that, physicists gradually noticed that every object exhibits both wave-particle duality and the energy states of a system are quantized. Finally, in 1925 Schrodinger wrote down an equation analogous to a wave equation: Schrodinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m_0} \frac{\partial^2}{\partial x^2} \Psi + V\Psi \quad (2.1)$$

where Ψ is the wavefunction of a particle, m_0 is the rest mass of the particle, and V is the potential energy experienced by a particle.

2.1.1 Time-Independent Schrodinger Equation

The time independent Schrodinger equation² predicts that the wavefunctions are able to form the standing waves which is called stationary states (also called "orbitals," such as atomic orbitals and molecular orbitals). In stationary states, the wavefunction can be divided into two parts using separation of variation method:

$$\Psi = \psi(x) \varphi(t) \quad (2.2)$$

¹Einstein explicitly assumed that the electromagnetic radiation is quantized in 1905.

²D. J. Griffiths, Introduction to Quantum Mechanics, Chapter 2.

where ψ is a function of x alone, and φ is a function of t alone. There, we can rewrite Equation (2.1)

$$i\hbar\psi\frac{\partial\varphi}{\partial t} = -\frac{\hbar^2}{2m_0}\frac{\partial^2\psi}{\partial x^2}\varphi + V\psi\varphi \quad (2.3)$$

Dividing through by $\psi\varphi$:

$$i\hbar\frac{1}{\varphi}\frac{\partial\varphi}{\partial t} = -\frac{\hbar^2}{2m_0}\frac{1}{\psi}\frac{\partial^2\psi}{\partial x^2} + V \quad (2.4)$$

Now, the left side and the right side are functions of t and x , respectively³. Thus, both sides are equal to a constant E . Then

$$i\hbar\frac{1}{\varphi}\frac{\partial\varphi}{\partial t} = E \quad (2.5)$$

and

$$-\frac{\hbar^2}{2m_0}\frac{1}{\psi}\frac{\partial^2\psi}{\partial x^2} + V = E \quad (2.6)$$

The general solution of Equation (2.5) is

$$\varphi(t) = e^{-iEt/\hbar} \quad (2.7)$$

and Equation (2.6) becomes

$$-\frac{\hbar^2}{2m_0}\frac{\partial^2\psi}{\partial x^2} + V\psi = E\psi$$

Or

$$\boxed{\left[-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial x^2} + V\right]\psi = E\psi} \quad (2.8)$$

The Equation (2.8) is called the **time-independent Schrodinger equation** and

$$-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial x^2} + V \equiv H \quad (2.9)$$

is called the **Hamiltonian** which is an operator. The Equation (2.8) implicitly states that solving the Schrodinger equation is an eigenvalue problem. The physical meaning of the constant E is actually the energy of the particle. Thus, the energy available in a quantum system can be determined by Equation (2.8). Taking a simplest example of a free electron ($V = 0$), we can immediately solve Schrodinger equation and get

$$\psi = Ae^{ikx} \quad (2.10)$$

and

$$E = \frac{\hbar^2 k^2}{2m_0} \quad (2.11)$$

³Here, V is only a function of x .

where k is the wavenumber. Note that the Equation (2.10) is a plane wave solution. If we apply the periodic boundary condition, we have

$$\begin{aligned}\psi(x=0) &= \psi(x=L) \\ 1 = e^{ikL} &\Rightarrow kL = 2\pi n \Rightarrow k = \frac{2\pi}{L}n\end{aligned}\tag{2.12}$$

and

$$E = \frac{\hbar^2}{2m_0} \frac{4\pi^2}{L^2} n^2\tag{2.13}$$

where n is an integer. Thus, for an one-dimensional free electron, the density of states is

$$\begin{aligned}D(E) &= \sum_n \delta(E - E_n) = \sum_n \delta\left(E - \frac{\hbar^2}{2m_0} \frac{4\pi^2}{L^2} n^2\right) \\ &= \int_{-\infty}^{\infty} \frac{dk}{\Delta k} \delta\left(E - \frac{\hbar^2 k^2}{2m_0}\right), \quad \frac{\hbar^2 k^2}{2m_0} \equiv z, \quad \Delta k = \frac{2\pi}{L} \\ &= \frac{Lm_0}{\pi\hbar} \frac{1}{\sqrt{2m_0}} \int_0^{\infty} dz \frac{1}{\sqrt{z}} \delta(E - z) \\ &= \frac{Lm_0}{\pi\hbar} \frac{1}{\sqrt{2m_0}} \frac{1}{\sqrt{E}}\end{aligned}\tag{2.14}$$

Note that the electron has up- and down-spin degeneracy, so the total DOS (per unit length per unit energy) is

$$\boxed{\frac{D(E)}{L} \equiv D(E) = \frac{2m_0}{\pi\hbar} \frac{1}{\sqrt{2m_0}} \frac{1}{\sqrt{E}}}\tag{2.15}$$

2.1.2 Formalism in Quantum Mechanics

The formalism in quantum mechanics is somehow different from the classical physics. There are two important constructs: wavefunctions and operators, which represent the state of a system and observables. In mathematical language (linear algebra), the wavefunctions are vectors and, the operators are the linear transformations. A vector in an N -dimensional space can be expressed using a specified orthogonal basis

$$|\psi\rangle \rightarrow \boldsymbol{\psi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}\tag{2.16}$$

and

$$\langle\psi| \rightarrow \boldsymbol{\psi}^\dagger = \begin{bmatrix} \phi_1^* & \phi_2^* & \dots & \phi_N^* \end{bmatrix}\tag{2.17}$$

The Equations (2.16) and (2.17) are the ket and bra, which form the **Dirac notation**. In general, the solution of Schrodinger equation is

$$|\psi\rangle = \sum_j^N c_j |\phi_j\rangle \quad (2.18)$$

and ϕ_j 's form a set of orthogonal basis in N -dimensional **Hilbert space** obeying

$$\int_{-\infty}^{\infty} \phi_i^*(x) \phi_j(x) dx = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (2.19)$$

where ϕ_i^* is the complex conjugate of ϕ_i and the Equation (2.19) is the inner product of two bases. The coefficients c_j 's can be obtained by

$$\boxed{\langle \phi_k | \psi \rangle = \int_{-\infty}^{\infty} \phi_k^* [c_1 \phi_1 + c_2 \phi_2 + \dots + c_k \phi_k + \dots] dx = c_k} \quad (2.20)$$

Using Dirac notation to rewrite Schrodinger equation, we have⁴

$$E|\psi\rangle = \mathbf{H}|\psi\rangle \quad (2.21)$$

If multiplying $\langle \phi_k |$ to the Equation (2.21),

$$\begin{aligned} \sum_j E c_j \langle \phi_k | \phi_j \rangle &= \sum_j \langle \phi_k | \mathbf{H} c_j | \phi_j \rangle \\ \Rightarrow E c_k &= \sum_j \langle \phi_k | \mathbf{H} | \phi_j \rangle c_j = \sum_j H_{k,j} c_j \end{aligned} \quad (2.22)$$

where

$$H_{k,j} \equiv \langle \phi_k | \mathbf{H} | \phi_j \rangle \quad (2.23)$$

Thus,

$$\begin{aligned} E \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} &= \begin{bmatrix} H_{11}c_1 + H_{12}c_2 + \dots \\ H_{21}c_1 + H_{22}c_2 + \dots \\ \vdots \end{bmatrix} = \mathbf{H} [c] \\ \Rightarrow \boxed{E [c] = \mathbf{H} [c]} & \quad (2.24) \end{aligned}$$

The Equation (2.24) explicitly indicates that solving the Schrodinger equation is mathematically an eigenvalue problem, where $[c]$ is the eigenvectors with the corresponding eigenvalues E .

⁴The bold font represents an operator here and later on. For example, the Hamiltonian is an operator and the energy E is a scalar.

2.1.3 Basis Transformation

As seen in Equation (2.18), there are multiple choices of the set of basis. The question is: how can we change the basis and what are the mathematical and physical meanings of such transformation? Let's consider

$$|\psi\rangle = \sum_j c_j |\phi_j\rangle = \sum_k d_k |u_k\rangle \quad (2.25)$$

$$\Rightarrow d_i = \langle u_i | \psi \rangle = \sum_j \langle u_i | \phi_j \rangle c_j \equiv \sum_j M_{ij} c_j \quad (2.26)$$

$$\Rightarrow \begin{bmatrix} d_1 \\ d_2 \\ \vdots \end{bmatrix} = [M] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} \quad (2.27)$$

where \mathbf{M} is the rotation matrix. Mathematically, the Equation (2.27) represents the rotation of a vector:

$$\mathbf{A}|x\rangle = |y\rangle \quad (2.28)$$

and

$$\langle y| = \langle x| \mathbf{A}^\dagger \quad (2.29)$$

such that

$$\begin{aligned} \langle y | \mathbf{A} | x \rangle &= \langle y | y \rangle \\ \Rightarrow \langle x | \mathbf{A}^\dagger \mathbf{A} | x \rangle &= \langle y | y \rangle = \langle x | x \rangle = 1 \end{aligned} \quad (2.30)$$

Therefore,

$$\mathbf{A}^\dagger \mathbf{A} = \mathbf{I} \quad (2.31)$$

where \mathbf{I} is the identity matrix and the Equation (2.31) is called unitary transformation. Now if we have a two-dimensional vector

$$|r\rangle = a|x\rangle + b|y\rangle \quad (2.32)$$

where

$$a = \langle x | r \rangle \quad \text{and} \quad b = \langle y | r \rangle$$

then

$$\begin{aligned} |r\rangle &= |x\rangle a + |y\rangle b = |x\rangle \langle x | r \rangle + |y\rangle \langle y | r \rangle \\ &= \underbrace{(|x\rangle \langle x| + |y\rangle \langle y|)}_{=1} |r\rangle \end{aligned} \quad (2.33)$$

More generally,

$$\sum_i |\phi_i\rangle \langle \phi_i| = 1 \quad (2.34)$$

The Equation (2.34) is called the identity operator. By using the identity operator, we obtain the new Hamiltonian in a new set of basis.

$$\begin{aligned}
 H_{kj} &= \langle u_k | \mathbf{H} | u_l \rangle \\
 &= \sum_{i,j} \langle u_k | \phi_i \rangle \langle \phi_i | \mathbf{H} | \phi_j \rangle \langle \phi_j | u_l \rangle \\
 &= \sum_{i,j} M_{ki} H_{ij} M_{jl} \\
 &\Rightarrow \mathbf{M}^\dagger \mathbf{H}^{\text{old}} \mathbf{M} = \mathbf{H}^{\text{new}}
 \end{aligned} \tag{2.35}$$

As mentioned earlier, solving the Schrodinger equation is a eigenvalue problem. Thus, the property of the Hamiltonian should be properly explored. Suppose we have the following eigenvalue problem

$$\mathbf{A}|x\rangle = \lambda|x\rangle \quad \text{and} \quad \langle x|\mathbf{A}^\dagger = \lambda^\dagger\langle x| \tag{2.36}$$

By multiplying $\langle x|$ and $|x\rangle$ from the left and right sides respectively to the Equation (2.36), we have

$$\langle x|\mathbf{A}|x\rangle = \lambda \quad \text{and} \quad \langle x|\mathbf{A}^\dagger|x\rangle = \lambda^\dagger \tag{2.37}$$

then

$$\begin{aligned}
 \langle x|\mathbf{A}|x\rangle - \langle x|\mathbf{A}^\dagger|x\rangle &= \langle x|\mathbf{A} - \mathbf{A}^\dagger|x\rangle = \lambda - \lambda^\dagger = 0 \quad \text{if} \quad \lambda \in \Re \\
 &\Rightarrow \mathbf{A} = \mathbf{A}^\dagger
 \end{aligned} \tag{2.38}$$

where \mathbf{A} is called the **Hermitian Matrix**. Because the energy E is real, **the Hamiltonian is also a Hermitian matrix**. From Copenhagen interpretation of quantum mechanics, the square of the coefficient c_i is the probability having the i th state and the total probability of all possible outcomes has to be one

$$\begin{aligned}
 \langle \psi | \psi \rangle &= \sum_{ij} c_i^* \langle \phi_i | c_j | \phi_j \rangle \\
 &= \sum_{ij} c_i^* c_j \langle \phi_i | \phi_j \rangle = \sum_i |c_i|^2 \\
 &= \text{Trace} \begin{bmatrix} |c_1|^2 & 0 & \dots \\ 0 & |c_2|^2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} = 1
 \end{aligned} \tag{2.39}$$

where each element on the digonal of the matrix is the probability at such state. The momentum operator \mathbf{p} is⁵

$$\mathbf{p} = -i\hbar \frac{\partial}{\partial x} \tag{2.40}$$

⁵D. J. Griffiths, Introduction to Quantum Mechanics, Chapter 1.

Thus, the expectation value of the momentum is⁶

$$\begin{aligned}\langle \mathbf{p} \rangle &= \langle \psi | \mathbf{p} | \psi \rangle = \sum_{ij} c_i^* c_j \underbrace{\langle \phi_i | \mathbf{p} | \phi_j \rangle}_{=\delta_{ij} p_{ij}} \\ &= \text{Trace} \left(\rho^{\text{eigenbasis}} p^{\text{eigenbasis}} \right)\end{aligned}\quad (2.41)$$

where $\rho^{\text{eigenbasis}}$ is the eigenvalue of the density matrix⁷, which is more generally defined as

$$\rho = \sum_i \rho_i |\phi_i\rangle \langle \phi_i| \quad (2.42)$$

where

$$\rho_i = \frac{n_i}{N} = |c_i|^2 = |\phi_i\rangle \langle \phi_i| \quad \text{and} \quad \sum_i \rho_i = 1 \quad (2.43)$$

n_i is the number of particles in the i th state and N is the total number of particles in a system. Check consistency of density matrix:

$$n = \int_{-\infty}^{\infty} \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2} D(E) dE \quad (2.44)$$

Assume $\gamma_1 = \gamma_2$ and $f_1 = f_2 = f_0(E - \mu)$, we have⁸

$$\begin{aligned}n &= \int_{-\infty}^{\infty} f_0(E - \mu) D(E) dE \\ &= \sum_{\varepsilon_i} \int_{-\infty}^{\infty} f_0(E - \mu) \delta(E - \varepsilon_i) dE \\ &= \sum_{\varepsilon_i} f_0(E - \mu) = \sum_i n_i = N \langle \psi | \psi \rangle = N\end{aligned}\quad (2.45)$$

2.2 Bloch's Theorem

If we want to describe the electron wavefunction in a solid, we need to consider the potential V experienced by the electrons. Let's consider an one-dimensional lattice as shown in Fig. 2.1. Because there is no way to distinguish the electron wavefunction at each site, the Schrodinger equations of electrons at each site can be written as

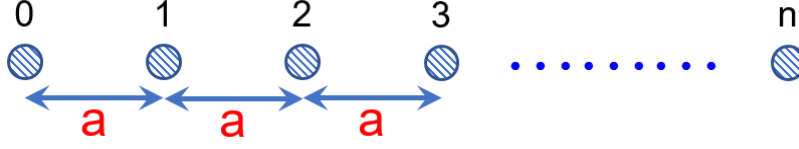
$$H|\psi(\vec{r} + \vec{a})\rangle = E|\psi(\vec{r} + \vec{a})\rangle \quad (2.46)$$

$$H|\psi(\vec{r} + 2\vec{a})\rangle = E|\psi(\vec{r} + 2\vec{a})\rangle \quad (2.47)$$

⁶ $\langle x | \mathbf{p} | x' \rangle = -i\hbar \delta'(x - x')$. See R. Shankar, Principles of Quantum Mechanics, second edition, page 64-64.

⁷It can be interpreted as the probability of having electrons in a given state.

⁸There is a confusion here. The unit of the density of states here is **per unit energy**. However, in many semiconductor textbooks, the density of states is in unit of **per unit volume (or area or length) per unit energy**. Thus, n here is the total number of electrons (carriers) instead of density.

Figure 2.1: An one-dimensional lattice with lattice constant of a .

and so on. However, above equations imply infinite degeneracy in a solid which is not physical. Therefore, between each wavefunction there should be a scalar multiplier. In other words,

$$|\psi(\vec{r} + 2\vec{a})\rangle = A|\psi(\vec{r} + \vec{a})\rangle \quad (2.48)$$

where A is a scalar. Note that

$$\begin{aligned} \langle\psi(\vec{r} + 2\vec{a})|\psi(\vec{r} + 2\vec{a})\rangle &= A^*A\langle\psi(\vec{r} + \vec{a})|\psi(\vec{r} + \vec{a})\rangle \\ \Rightarrow |A|^2 &= 1 \Rightarrow A = e^{i\phi} \end{aligned}$$

Thus,

$$|\psi(\vec{r} + n\vec{a})\rangle = e^{in\phi}|\psi(\vec{r})\rangle$$

or

$$\boxed{|\psi(\vec{r} + \vec{L})\rangle = e^{i\vec{k} \cdot \vec{L}}|\psi(\vec{r})\rangle} \quad (2.49)$$

where $L(= na)$ is the length of the one-dimensional lattice and $k(= \frac{2\pi}{a})$ is the wave number. The Equation (2.49) is called the **Bloch's Theorem**. In general, the Equation (2.49) can be further expressed, for example, in three-dimensional lattice, and the relation between the wave number and the lattice size follows the Fourier transformation.

$$u(\vec{r}) = \sum_{\vec{g}} A_{\vec{g}} e^{i\vec{g} \cdot \vec{r}} \quad (2.50)$$

where the space spanned by \vec{g} is called the **reciprocal space** [see Fig. 2.2(b)], and \vec{g} in three-dimensional lattice is

$$\begin{aligned} \vec{g} &= \frac{2\pi}{a}n_1\hat{x} + \frac{2\pi}{b}n_2\hat{y} + \frac{2\pi}{c}n_3\hat{z} \\ &= n_1\vec{b}_1 + n_2\vec{b}_2 + n_3\vec{b}_3 \end{aligned} \quad (2.51)$$

where a , b , and c are the lattice constant of x , y , and z directions, n_i 's are the integers, and \vec{b}_i 's are

$$\vec{b}_1 \equiv \frac{2\pi}{a}\hat{x}, \quad \vec{b}_2 \equiv \frac{2\pi}{b}\hat{y}, \quad \text{and} \quad \vec{b}_3 \equiv \frac{2\pi}{c}\hat{z} \quad (2.52)$$

The Equation (2.52) can be generalized

$$\begin{aligned}\vec{b}_1 &= 2\pi \frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)} \\ \vec{b}_2 &= 2\pi \frac{\vec{a}_3 \times \vec{a}_1}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)} \\ \vec{b}_3 &= 2\pi \frac{\vec{a}_1 \times \vec{a}_2}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)}\end{aligned}\quad (2.53)$$

where \vec{a}_i 's and \vec{b}_i 's are the basis of the real space and reciprocal space, and the denominator is the volume of the unit cell in the real space. Suppose we have a vector constructed by the basis of the real space

$$\vec{l} = m\vec{a}_1 + n\vec{a}_2 + o\vec{a}_3 \quad (2.54)$$

where m , n , and o are the integers. Then,

$$\begin{aligned}\vec{g} \cdot \vec{l} &= \left(n_1 \vec{b}_1 + n_2 \vec{b}_2 + n_3 \vec{b}_3 \right) \cdot \left(m\vec{a}_1 + n\vec{a}_2 + o\vec{a}_3 \right) \\ &= 2\pi (n_1 m + n_2 n + n_3 o) \equiv 2\pi N\end{aligned}\quad (2.55)$$

This implies that the wavefunction will repeat after a certain translation. If we define $\vec{k} = \vec{k}' + \vec{g}$, the Equation (2.49) becomes

$$\begin{aligned}\psi\left(\vec{r} + \vec{L}\right) &= e^{i\vec{k} \cdot \vec{L}} \psi\left(\vec{r}\right) = e^{i\left(\vec{k}' + \vec{g}\right) \cdot \vec{L}} \psi\left(\vec{r}\right) \\ &= \underbrace{e^{i(2\pi)N}}_{=1} e^{i\vec{k}' \cdot \vec{L}} \psi\left(\vec{r}\right)\end{aligned}\quad (2.56)$$

Since there is no way to distinguish k and k' , the wavefunction will repeat after shift k by $2\pi/a$. Within the region where the wavefunction does not repeat, it is called the **Brillouin zone** as shown in Fig. 2.2(a). A common used method to obtain the Brillouin zone is to draw the **Wigner Seitz cell** which bisects each nearest neighbor as shown in Fig. 2.2(b).

2.3 Band Structure

2.3.1 One-Dimensional Monatomic Lattice

Suppose we have an one-dimensional lattice shown in Fig. 2.1 and assume that the overlap of wavefunctions only presents between nearest neighbor atoms. The elements in Hamiltonian matrix are

$$H_{ii} = \langle \psi_i | \mathbf{H} | \psi_i \rangle = t_0 \quad (2.57)$$

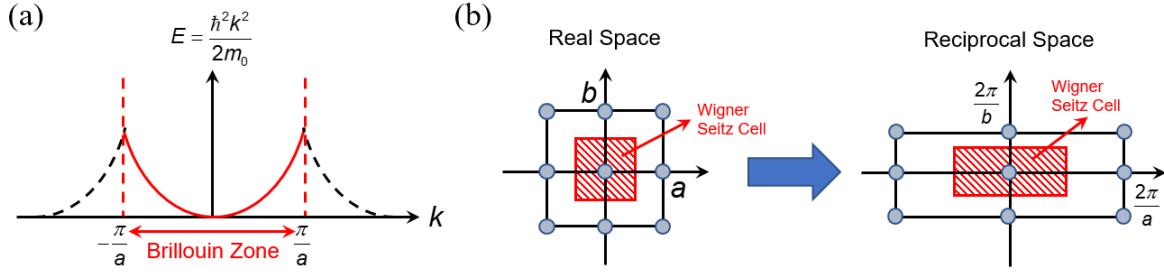


Figure 2.2: (a) Dispersion relation ($E - k$) of an one-dimensional lattice. (b) A two-dimensional lattice in real space and reciprocal space.

$$H_{i(i\pm 1)} = \langle \psi_i | \mathbf{H} | \psi_{i(i\pm 1)} \rangle = t_1 \quad (2.58)$$

The full matrix looks like

$$\mathbf{H} = \begin{bmatrix} t_0 & t_1 & 0 & 0 & 0 & \dots \\ t_1^* & t_0 & t_1 & 0 & 0 & \dots \\ 0 & t_1^* & t_0 & t_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (2.59)$$

The Schrodinger equation becomes⁹

$$\mathbf{H}\psi = \begin{bmatrix} t_0 & t_1 & 0 & 0 & 0 & \dots \\ t_1^* & t_0 & t_1 & 0 & 0 & \dots \\ 0 & t_1^* & t_0 & t_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{bmatrix} = E \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{bmatrix} \quad (2.60)$$

At the n th row:

$$t_0\psi_n + t_1\psi_{n+1} + t_1^*\psi_{n-1} = E\psi_n \quad (2.61)$$

and from Bloch's theorem

$$\psi_n = \psi_0 e^{ikna} \quad (2.62)$$

Thus, we have

$$E = t_0 + t_1 e^{ika} + t_1^* e^{-ika} \xrightarrow{\text{if } t_1 \text{ is real}} t_0 + 2t_1 \cos ka \quad (2.63)$$

If $ka \ll 1$ and $t_1 < 0$, the Equation (2.63) can be expanded by using Taylor series.

$$\begin{aligned} E &\simeq t_0 + 2t_1 \left[1 - \frac{(ka)^2}{2} + \dots \right] \\ &= t_0 + 2t_1 - (t_1 a^2) k^2 \\ &= t_0 - 2|t_1| + (|t_1| a^2) k^2 \end{aligned} \quad (2.64)$$

⁹Each ψ represents the wavefunction of an electron at each lattice site instead of the basis function.

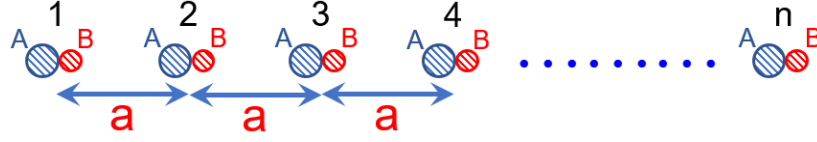


Figure 2.3: An one-dimensional diatomic (A and B) lattice with lattice constant of a .

Define the effective mass m^*

$$m^* = \frac{\hbar^2}{2a^2|t_1|} \quad (2.65)$$

The Equation (2.64) becomes

$$E = E' + \frac{\hbar^2}{2m^*} k^2 \quad (2.66)$$

where $E' \equiv t_0 - 2|t_1|$. When $|t_1|$ is large, the coupling between the nearest neighbor is strong so that the electron is "de-localized"¹⁰. The effective mass approximation is usually valid for the transport phenomenon in the commonly used semiconductors since the electrons involving the transport are near the bottom of the band (within 1eV). However, when solving the binding strength between the nearest neighbors, this approximation may not be good.

2.3.2 One-Dimensional Diatomic Lattice

The diatomic lattice contains two atoms in a lattice site as shown in Fig. 2.3. We assume that the wavefunction overlap only happens between the nearest neighbor atoms and atoms on site, i.e., A_1 and B_1 or B_1 and A_2 . Based on this assumption, the Hamiltonian can be expressed as

		1	1	2	2	3	3	
		A	B	A	B	A	B	
1	A	t_0	t_1	0	0	0	0	...
1	B	t_1^*	t_0	t_2	0	0	0	...
2	A	0	t_2^*	t_0	t_1	0	0	...
2	B	0	0	t_1^*	t_0	t_2	0	...
3	A	0	0	0	t_2^*	t_0	t_1	...
3	B	0	0	0	0	t_1^*	t_0	...
		\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

At the n th row¹¹, the Schrodinger equation gives

$$[H]_{n,n} [\psi_0] e^{ikna} + [H]_{n,n+1} [\psi_0] e^{ikna} e^{ika} + [H]_{n,n-1} [\psi_0] e^{ikna} e^{-ika} = E [\psi_0] e^{ikna} \quad (2.67)$$

¹⁰The electron can move freely through the lattice sites.

¹¹Each row represents each lattice site.

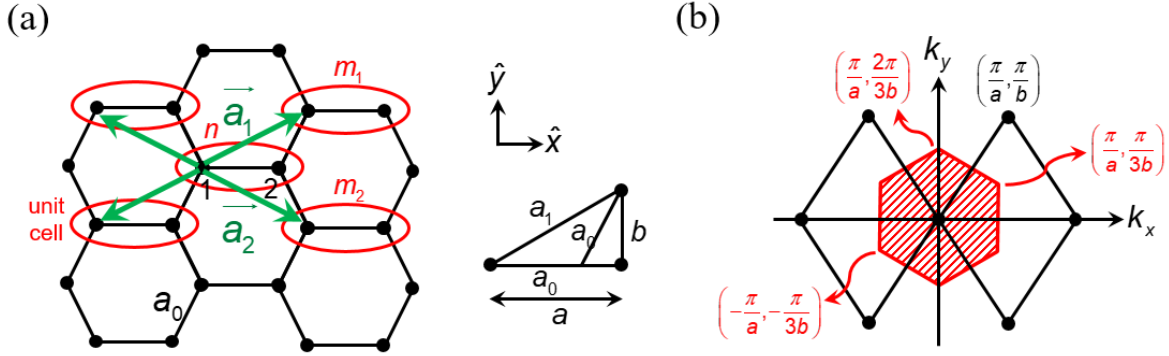


Figure 2.4: Graphene in (a) real space (lattice constant is a_0) and (b) reciprocal space. The green arrows in (a) are the translation vectors (\vec{a}_1 and \vec{a}_2) and two atoms circled by the red line form the unit cell. The red hexagon is the Wigner Seitz cell (Brillouin zone).

$$\begin{aligned} &\Rightarrow \begin{bmatrix} t_0 & t_1 + t_2 e^{ika} \\ t_1^* + t_2^* e^{-ika} & t_0 \end{bmatrix} \begin{bmatrix} \psi_{0A} \\ \psi_{0B} \end{bmatrix} = E \begin{bmatrix} \psi_{0A} \\ \psi_{0B} \end{bmatrix} \\ &\Rightarrow E = t_0 \pm \sqrt{|t_1|^2 + |t_2|^2 + 2\text{Re}(t_1 t_2 e^{ika})} \end{aligned} \quad (2.68)$$

The Equation (2.68) indicates that there are two bands in this system. If considering more interactions between more basis functions, more bands would be obtained. More generally, the Equation (2.67) can be expressed as

$$\sum_m [H]_n + [H]_{n,m} e^{i\vec{k} \cdot (\vec{r}_m - \vec{r})} \quad (2.69)$$

where the first and second terms are the Hamiltonian matrix on the lattice site and the Hamiltonian of the overlap between the nearest neighbors respectively, and m represents the nearest neighbor. The first term also can be absorbed into the second term and the Equation (2.69) becomes

$$\boxed{\sum_m [H]_{n,m} e^{i\vec{k} \cdot (\vec{r}_m - \vec{r})} \equiv \left[h(\vec{k}) \right]} \quad (2.70)$$

which is the **Fourier transformation** between the reciprocal and real spaces.

2.3.3 Graphene

Graphene consists of carbon atoms with two-dimensional hexagonal lattice as shown in Fig. 2.4. The coordinate of graphene in real space is defined in Fig. 2.4(a). Thus, the translation vectors in real space are

$$\vec{a}_1 = a\hat{x} + b\hat{y} \quad (2.71)$$

$$\vec{a}_2 = a\hat{x} - b\hat{y} \quad (2.72)$$

$$\vec{a}_3 = \hat{z} \quad (2.73)$$

and

$$a = a_0 + a_0 \sin \frac{\pi}{6} = \frac{3a_0}{2} \quad (2.74)$$

$$b = a_0 \cos \frac{\pi}{6} = \frac{\sqrt{3}}{2} a_0 \quad (2.75)$$

Then, the translation vectors in reciprocal space are

$$\vec{b}_1 = 2\pi \frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1 \cdot \vec{a}_2 \times \vec{a}_3} = \frac{\pi}{a} \hat{x} + \frac{\pi}{b} \hat{y} \quad (2.76)$$

$$\vec{b}_2 = \frac{\pi}{a} \hat{x} - \frac{\pi}{b} \hat{y} \quad (2.77)$$

Due to weaker coupling of p_z orbitals between nearest neighbors than that of $p_{x,y}$, the energy split is smaller (the energy states are closer to the Fermi level) so that the wavefunction overlap of p_z orbitals is considered. We also assume that only the wavefunction overlap of the nearest neighbors does matter. Thus, the Hamiltonian matrix of the n th unit cell [see Fig. 2.4(a)] can be expressed using the Equation (2.70)

$$[H]_{n,n} + [H]_{n,m_1} e^{i\vec{k} \cdot \vec{a}_1} + [H]_{n,m_2} e^{i\vec{k} \cdot \vec{a}_2} + [H]_{n,m_1} e^{-i\vec{k} \cdot \vec{a}_1} + [H]_{n,m_2} e^{-i\vec{k} \cdot \vec{a}_2} \quad (2.78)$$

where

$$[H]_{n,n} = \begin{bmatrix} t_0 & t \\ t & t_0 \end{bmatrix}, \quad [H]_{n,m_1} = \begin{bmatrix} 0 & 0 \\ t & 0 \end{bmatrix}, \quad [H]_{n,m_2} = \begin{bmatrix} 0 & 0 \\ t & 0 \end{bmatrix} \quad (2.79)$$

and

$$t_0 = \langle \psi_1 | \mathbf{H} | \psi_1 \rangle, \quad t = \langle \psi_1 | \mathbf{H} | \psi_2 \rangle, \quad t = t^* \quad (2.80)$$

Here, $\psi_{1,2}$ are the wavefunctions at the first (left side) and the second (right side) lattice sites in an unit cell. Therefore, the Equation (2.78) is

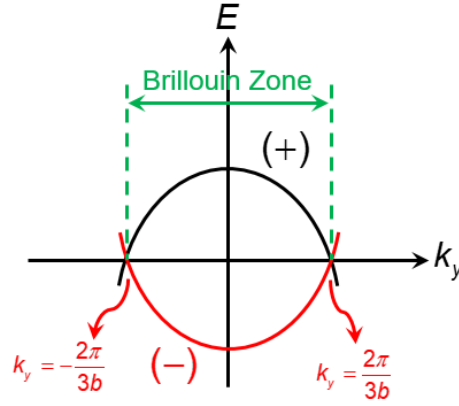
$$\begin{bmatrix} h(\vec{k}) \end{bmatrix} = \begin{bmatrix} t_0 & t + te^{-i\vec{k} \cdot \vec{a}_1} + te^{-i\vec{k} \cdot \vec{a}_2} \\ t + te^{i\vec{k} \cdot \vec{a}_1} + te^{i\vec{k} \cdot \vec{a}_2} & t_0 \end{bmatrix} \equiv \begin{bmatrix} t_0 & p^*(\vec{k}) \\ p(\vec{k}) & t_0 \end{bmatrix} \quad (2.81)$$

where

$$p(\vec{k}) = t + te^{i\vec{k} \cdot \vec{a}_1} + te^{i\vec{k} \cdot \vec{a}_2} = t \left[1 + 2e^{ik_x a} \cos k_y b \right] \quad (2.82)$$

The eigenvalue (energy) of the matrix (2.81) is

$$\begin{aligned} E(\vec{k}) &= t_0 \pm |p(\vec{k})| \\ &= t_0 \pm t \sqrt{1 + 4 \cos^2 k_y b + 4 \cos k_x a \cos k_y b} \end{aligned} \quad (2.83)$$

Figure 2.5: E - k diagram of graphene at $k_x = 0$.

The above equation describes the dispersion relationship (E - k diagram) of graphene. Let's now consider $k_x = 0$:

$$E = t_0 \pm t(1 + 2 \cos k_y b) \quad (2.84)$$

If $t_0 = 0$ is chosen, the E - k_y plot is shown in Fig 2.5. Note that $k_y = \pm \frac{2\pi}{3b}$ are the points located at the Brillouin zone boundary as shown in Fig. 2.4(b). These points are called "Dirac points." If $k_x \neq 0$, the E - k would look like a cone and the k_x and k_y satisfying $E = 0$ are $(\pm \frac{\pi}{a}, \pm \frac{\pi}{3b})$ where are at the zone boundary. Because the two bands are crossing at the Dirac points, the graphene behaves like a metal and the number of electrons near the Fermi level ($E_F = 0$) is $2 \times 2 \times 2 = 8^{12}$. Now, let's consider E near the Dirac point. The Taylor expansion of the Equation (2.83) assuming $t_0 = 0$ is

$$\begin{aligned} E(\vec{k}) &= \pm |p(\vec{k})| \\ &\approx \pm \left| \frac{\partial p}{\partial k_x} \Big|_{0, \pm \frac{2\pi}{3b}} (k_x - 0) + \frac{\partial p}{\partial k_y} \Big|_{0, \pm \frac{2\pi}{3b}} \left(k_y \pm \frac{2\pi}{3b} \right) \right| \\ &= \pm \left| -\frac{3}{2} a_0 t \left[i k_x + \left(k_y \mp \frac{2\pi}{3b} \right) \right] \right| \\ &= \pm \frac{3}{2} a_0 t \sqrt{k_x^2 + \left(k_y \mp \frac{2\pi}{3b} \right)^2} \\ &= \pm a t |k| \quad (\text{linear dispersion relationship}) \end{aligned} \quad (2.85)$$

which is different from the square dependence of k for the free electron. The definition of the effective mass $m^* = \frac{\hbar^2}{\frac{\partial^2 E}{\partial k^2}}$ at the Dirac points becomes physically meaningless where defining $p = \hbar k$ would be better. Since $p = 0$ at the Dirac points, there will be no current flow at

¹²2's are coming from two bands, up- and down-spin, and equivalent corner in the Brillouin zone.

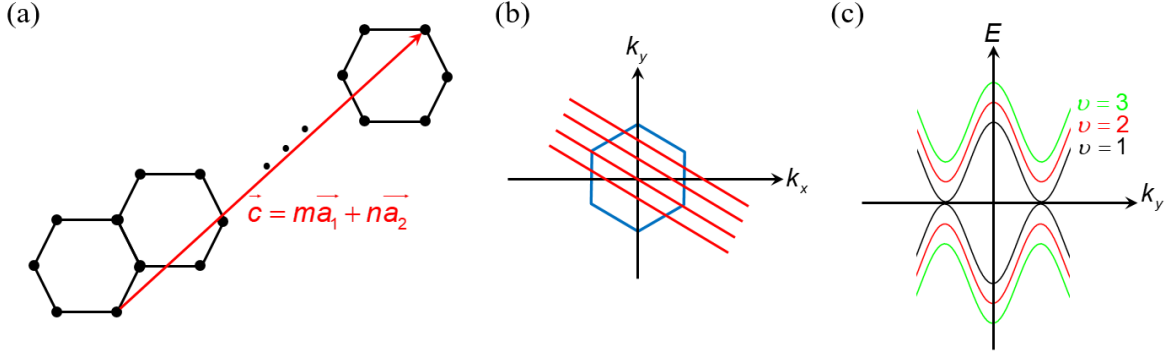


Figure 2.6: (a) Graphene with rolling vector \vec{c} . (b) Brillouin zone of graphene and possible solutions (red straight lines) of nanotube. (c) $E - k$ diagram of graphene nanotube, assuming $k_x = 0$.

that points. Around the Dirac point, the electron group velocity is

$$v = \frac{\partial \omega}{\partial k} = \frac{1}{\hbar} \frac{\partial E}{\partial k} = \frac{3}{2} \frac{a_0 t}{\hbar} = \text{constant} \quad (2.86)$$

where v is always written as v_F (Fermi velocity). Thus, the E - k relation of graphene also can be written in terms of v_F

$$E = \pm \hbar v_F |k| \quad (2.87)$$

where the electron behaves like a relativistic particle and cannot be accelerated.

2.3.4 Graphene Nanotube

Since the graphene behaves like a metal, engineering graphene to be semiconductor would be very important. One way to achieve that is to roll it into a tube as shown in Fig. 2.6. After rolling the graphene, we impose additional restriction to the lattice

$$\vec{k} \cdot \vec{c} = 2\pi\nu, \quad \nu \in Z \quad (2.88)$$

because the wavefunction will be a periodic function. The Equation (2.88) can be further written as

$$k_y = \frac{2\pi}{(m-n)b} \nu - \frac{(m+n)a}{(m-n)b} k_x \quad (2.89)$$

which represents straight lines shown in Fig. 2.6(b) depending on ν if m and n are known. Because the corners of the Brillouin zone [see Fig. 2.6(b)] correspond to the Dirac points, the graphene nanotube exhibits the metallic nature if the straight line crosses the corners. Therefore, whether the graphene nanotube behaves like a metal or a semiconductor would depend on how it is rolled (various combinations of m and n).

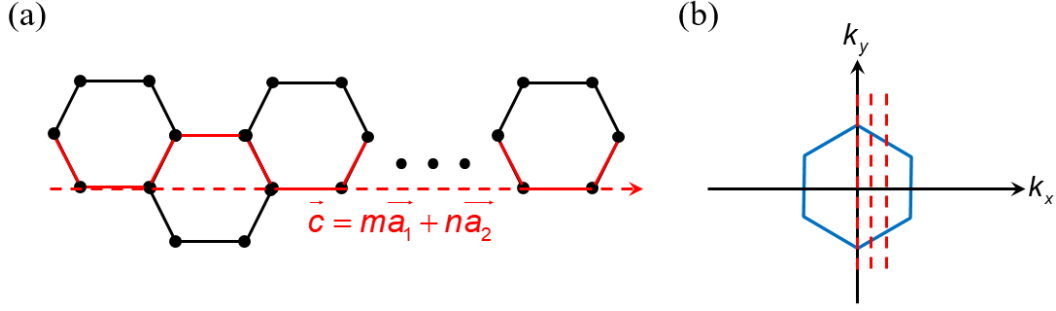


Figure 2.7: (a) Illustration of rolling an armchair nanotube. (b) Brillouin zone and possible solutions (straight lines) of armchair nanotube.

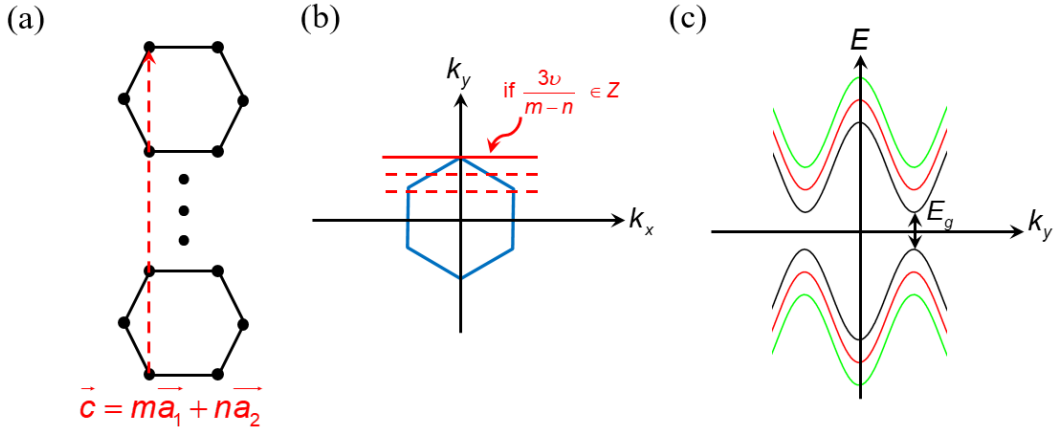


Figure 2.8: (a) Illustration of rolling a zigzag nanotube. (b) Brillouin zone and possible solutions (straight lines) of zigzag nanotube. (c) $E - k_y$ diagram when $(m - n)$ is not an integer multiple of 3.

Armchair Graphene Nanotube

If the graphene is rolled only in x -direction, it's called armchair nanotube as shown in Fig. 2.7. Because now the wavefunction is restricted in x -direction, the k_y is zero and thus the Equation (2.89) becomes

$$k_x = \frac{2\pi}{(m+n)a}\nu \quad (2.90)$$

Since $k_x = 0$ is a valid solution, the armchair nanotube is always metallic.

Zigzag Graphene Nanotube

If the graphene is rolled only in y -direction, it's called zigzag nanotube as shown in Fig. 2.8. Because now the wavefunction is restricted in y -direction, the k_x is zero and thus the Equation (2.89) becomes

$$k_y = \frac{2\pi}{3b} \frac{3\nu}{m-n} \quad (2.91)$$

If $\frac{3\nu}{m-n}$ is an integer, the zigzag nanotube will become metallic since k_y crosses the corner of Brillouin zone [see Fig. 2.8(b)]. If $\frac{3\nu}{m-n}$ is not an integer, the zigzag nanotube behaves like a semiconductor and there is a bandgap E_g as shown in Fig. 2.8(c). The bandgap can be evaluated using the Equation (2.85)

$$\begin{aligned}
 E_g &= 2 \times \frac{3}{2} a_0 t \sqrt{k_x^2 + \left(k_y \mp \frac{2\pi}{3b}\right)^2} \Big|_{k_x=0, k_y=\frac{2\pi}{3b} \frac{3\nu}{m-n}} \\
 &= 3a_0 t \frac{3}{m-n} \left(\nu \mp \frac{m-n}{3}\right) \frac{2\pi}{3b} \\
 &\xrightarrow{\text{minimum}} 3a_0 t \frac{3}{m-n} \frac{1}{3} \frac{2\pi}{3b} = \frac{2a_0 t \pi}{(m-n)b}
 \end{aligned} \tag{2.92}$$

If the diameter of the zigzag nanotube is d , the bandgap is

$$E_g = \frac{2a_0 t}{d} = \frac{0.8}{d \text{ (nm)}} \text{ eV} \tag{2.93}$$

where $\pi d = (m-n)b$. The Equation (2.93) is validated by the experimental results. However, if $d < 1$ nm, it is not accurate (but the trend is still correct) due to the presence of strain.

2.4 Density of States

From previous section, we have already obtained the $E - k$ relation of several lattices. From the developed band structures, we can therefore derive the density of states.

Graphene

From the definition of the density of states and $E - k$ relation of graphene [Equation (2.85)], we have

$$\begin{aligned}
 D(E) &= \sum_{\varepsilon_k} \delta(E - \varepsilon_k) \\
 &= \sum_{\varepsilon_k} \delta(E - at|k|) \\
 &= \int \frac{dk_x}{\frac{2\pi}{L_x}} \frac{dk_y}{\frac{2\pi}{L_y}} \delta(E - at|k|) \\
 &= \frac{A(\equiv L_x L_y)}{4\pi^2} \int 2\pi k dk \delta(E - at|k|)
 \end{aligned} \tag{2.94}$$

Considering spin- and valley-degeneracy (each contributes 2) and defining the density of states per unit area, we have

$$D(E) = \frac{2 \times 2}{2\pi} \int k \delta(E - at|k|) = \frac{2}{\pi a^2 t^2} |E| \tag{2.95}$$

which is different from the 2D DOS of the free electron

$$D(E) = \frac{m^*}{\pi \hbar^2} \quad (2.96)$$

Note that at $E = 0$ (Dirac point) the DOS is zero, indicating no current flowing around that energy level.

Zigzag Graphene Nanotube

The Equation (2.85) and (2.91) yield

$$\begin{aligned} E &= \pm at \sqrt{k_x^2 + \left[\frac{2\pi}{3b} \frac{3}{m-n} \left(\nu \pm \frac{m-n}{3} \right) \right]^2} \\ &= \pm at \sqrt{k_x^2 + \left[\frac{2}{d} \left(\nu \pm \frac{m-n}{3} \right) \right]^2} \\ &\equiv \pm at \sqrt{k_x^2 + B^2} \end{aligned} \quad (2.97)$$

The density of states per unit length with the spin- and valley-degeneracy is

$$\begin{aligned} D(E) &= 4 \times \sum_{k_x} \delta \left(E - at \sqrt{k_x^2 + B^2} \right) \\ &= 4 \times \frac{L_x}{2\pi} \int dk_x \delta \left(E - at \sqrt{k_x^2 + B^2} \right) \\ \Rightarrow \frac{D(E)}{L_x} &= \frac{2}{\pi} \int dk_x \delta \left(E - at \sqrt{k_x^2 + B^2} \right) \\ &= \frac{2}{\pi at} \frac{1}{\sqrt{E^2 - [2at \left(\nu - \frac{m-n}{3} \right)]^2} \frac{1}{d^2}} \end{aligned} \quad (2.98)$$

where d is the diameter of the nanotube. Fig. 2.9 shows the density of states of graphene and zigzag nanotube. Each rising in DOS of zigzag nanotube represents each sub-band that are created by the imposed boundary condition of rolling.

2.5 Effective Mass Equation

Suppose we have a simple cubic lattice with lattice constants of x -, y -, and z -direction of a , b , and c , and only one basis used per atom. The Equation (2.70) becomes

$$\begin{aligned} \left[h(\vec{k}) \right] &= E_0 + t_x e^{ik_x a} + t_y e^{ik_y b} + t_z e^{ik_z c} + \text{complex conjugate} \\ &= E_0 + 2t_x \cos k_x a + 2t_y \cos k_y b + 2t_z \cos k_z c \end{aligned} \quad (2.99)$$

where

$$\langle \psi_{x,y,z} | \mathbf{H} | \psi_{x,y,z} \rangle = t_{x,y,z} \quad (2.100)$$

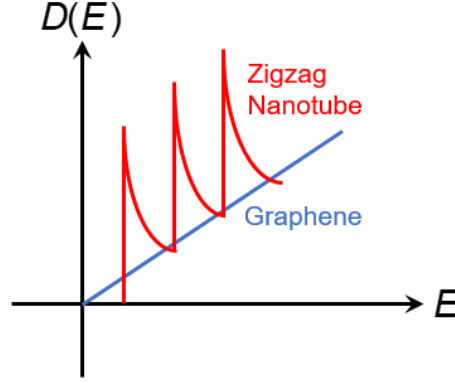


Figure 2.9: The density of states of graphene and zigzag nanotube.

By Taylor expansion, the Equation (2.99) can be rewritten as

$$\left[h(\vec{k}) \right] \approx E_0 + 2t_x + 2t_y + 2t_z - \frac{k_x^2 a^2}{2} t_x - \frac{k_y^2 b^2}{2} t_y - \frac{k_z^2 c^2}{2} t_z \quad (2.101)$$

If defining

$$\frac{\hbar^2}{2m_x^*} \equiv -\frac{a^2 t_x}{2}, \quad \frac{\hbar^2}{2m_y^*} \equiv -\frac{b^2 t_y}{2}, \quad \frac{\hbar^2}{2m_z^*} \equiv -\frac{c^2 t_z}{2} \quad (2.102)$$

then

$$\left[h(\vec{k}) \right] = E_{C0} + \frac{\hbar^2 k_x^2}{2m_x^*} + \frac{\hbar^2 k_y^2}{2m_y^*} + \frac{\hbar^2 k_z^2}{2m_z^*} \quad (2.103)$$

The effective mass approximation is good for the transport problem since it involves only a small range of E . Notice that the effective mass in the solid is generally anisotropic as seen in the Equation (2.103). However, for a free electron, the wavefunction can be separated into three direction by using separation of variables in three-dimensional Schrodinger equation, which gives an isotropic effective mass. Since in the solid the electron experiences various potential in each direction, the effective mass becomes anisotropic and it's captured by the calculated $E - k$ relation.

2.5.1 Dimensionality

Tow-Dimensional Solid

Suppose the z -direction is confined (x and y are not confined), the imposed boundary conditions are

$$\psi_z(z=0) = 0 \quad \text{and} \quad \psi_z(z=L_z) = 0 \quad (2.104)$$

The solution of wavefunction in z -direction is¹³

$$\psi_z(z) = A \sin k_z z \quad (2.105)$$

¹³The same as a particle in an one-dimensional infinite potential well.

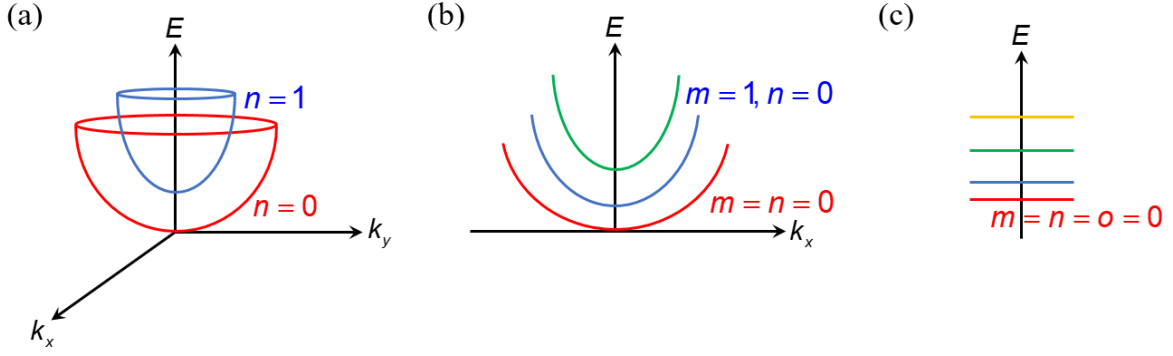


Figure 2.10: (a) Two-dimensional solid with one-directional being confined. (b) One-dimensional solid with two-directional being confined. (c) Zero-dimensional solid.

where $k_z = \frac{n\pi}{L_z}$, $n \in N$. Thus, the $E - k$ relation is

$$E = \frac{\hbar^2}{2m_z} \frac{\pi^2 n^2}{L_z^2} + \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y} \quad (2.106)$$

and is shown in Fig. 2.10(a) where the sub-bands form paraboloids with various n .

One-Dimensional Solid

Suppose the y - and z -direction are confined (x is not confined), similarly the $E - k$ relation is

$$E = \frac{\hbar^2}{2m_y} \frac{\pi^2 m^2}{L_y^2} + \frac{\hbar^2}{2m_z} \frac{\pi^2 n^2}{L_z^2} + \frac{\hbar^2 k_x^2}{2m_x} \quad (2.107)$$

and is shown in Fig. 2.10(b) where the sub-bands form parabolas with different combinations of m and n .

Zero-Dimensional Solid

If all three directions are confined, the $E - k$ relation becomes

$$E = \frac{\hbar^2}{2m_x} \frac{\pi^2 o^2}{L_x^2} + \frac{\hbar^2}{2m_y} \frac{\pi^2 m^2}{L_y^2} + \frac{\hbar^2}{2m_z} \frac{\pi^2 n^2}{L_z^2} \quad (2.108)$$

and is shown in Fig. 2.10(c) where the discrete energy levels form.

2.5.2 Maximum Conductance

In Chapter 1, we've discussed the maximum conductance based on the Landauer's equation

$$G^{max} = \frac{2e^2}{h} M \quad (2.109)$$

where the number of modes M represents the number of sub-bands in the system. If the Fermi level E_F of the system is known, then counting how many sub-bands below E_F results in M and thus the maximum conductance. Here, suppose we have an two-dimensional solid where the z -direction is confined. Up to E_F , we have n sub-bands (modes) which obeys

$$E_F = \frac{\hbar^2}{2m_z} \frac{\pi^2}{L_z^2} n^2 \quad \text{and} \quad n \approx M \quad (2.110)$$

Then,

$$M \approx \sqrt{\frac{2m_z E_F}{\pi^2 \hbar^2}} L_z^2 \quad (2.111)$$

The maximum integer is taken in the Equation (2.111). From the 2D density of states [see Equation (2.96)], the electron density is¹⁴

$$n_S = \frac{m_z}{\pi \hbar^2} E_F \quad (2.112)$$

Hence,

$$M \approx \sqrt{\frac{2}{\pi}} n_s L_z \quad (2.113)$$

The minimum resistance is

$$R_{min} = \frac{h}{2e^2 M} = \frac{h}{2e^2} \frac{1}{\sqrt{\frac{2}{\pi}} n_s} \frac{1}{L_z} \quad (2.114)$$

Therefore,

$$R_{min} \cdot L_z = \frac{h}{2e^2} \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{n_s}} \quad (2.115)$$

$$= 50 \, \Omega \mu\text{m}, \text{ for Silicon} \quad (2.116)$$

2.6 Schrodinger's Equation and Finite Difference Method

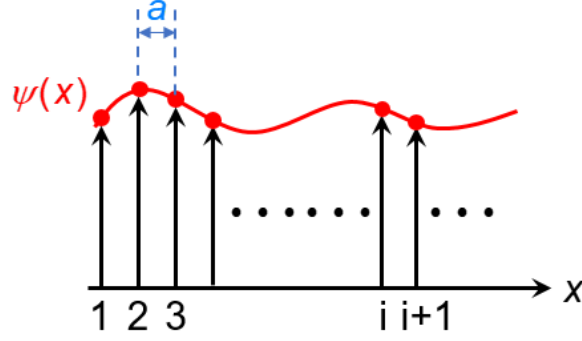
Since normally the Schrodinger's equation cannot be solved analytically, the numerical method should be always involved, where the commonly used method is called the finite difference method. The idea of this method is to sample the wavefunction discretely as shown in Fig 2.11. At the i th point, we have the first derivative of the wavefunction

$$\left. \frac{\partial \psi}{\partial x} \right|_i^> = \frac{\psi_{i+1} - \psi_i}{a} \quad (2.117)$$

and

$$\left. \frac{\partial \psi}{\partial x} \right|_i^< = \frac{\psi_i - \psi_{i-1}}{a} \quad (2.118)$$

¹⁴Assume $T \rightarrow 0$ such that below E_F the energy states are filled.

Figure 2.11: Sampling the wavefunction ψ with spacing a .

where $>$ and $<$ are denoted as the right and left sides. The second derivative of the wavefunction can be also obtained by

$$\frac{\partial^2 \psi}{\partial x^2} \Big|_i = \frac{\frac{\partial \psi}{\partial x} \Big|_i^> - \frac{\partial \psi}{\partial x} \Big|_i^<}{a} = \frac{\psi_{i+1} + \psi_{i-1} - 2\psi_i}{a^2} \quad (2.119)$$

Thus, the Schrodinger's equation at the i th point becomes

$$-\frac{\hbar^2}{2m_0 a^2} (\psi_{i+1} + \psi_{i-1} - 2\psi_i) + V_i \psi_i = E \psi_i \quad (2.120)$$

where V_i is the potential energy at the i th point. If defining $\frac{\hbar^2}{2m_0 a^2} \equiv t$, the Equation (2.120) becomes

$$(2t + V_i) \psi_i - t\psi_{i+1} - t\psi_{i-1} = E \psi_i \quad (2.121)$$

which can be generally written as a matrix form

$$\begin{bmatrix} 2t + V_1 & -t & 0 & 0 & \dots \\ -t & 2t + V_2 & -t & 0 & \dots \\ 0 & -t & 2t + V_3 & -t & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{bmatrix} = E \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{bmatrix} \quad (2.122)$$

Suppose V_i is a constant, then $E = V + 2t(1 - \cos ka)$. Otherwise, the Equation (2.122) should be solved numerically.

2.6.1 Current Evaluation

In Chapter 1, we've obtained the Landauer's equation for a very small device¹⁵.

$$I = \frac{2e}{\hbar} \int dE D(E - U) \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(E) - f_2(E)] \quad (2.123)$$

Now, suppose we have a three-terminal device like MOSFET and consider a channel with finite length and initial guess of potential with sampling as shown in Fig. 2.12. First of all,

¹⁵The channel is like a "dot."

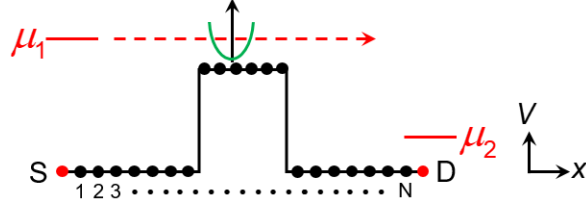


Figure 2.12: Initial guess of potential in MOSFET in real space.

the $E - k$ relation should be obtained for the given material based on the proposed method in previous section and thus the DOS. Since the current is the same at every point, the $E - k$ and DOS are only important at the top of barrier. Note that we cannot get assign $E - k$ relation to each sample point since we don't exactly simultaneously know the k and x due to the uncertainty principle. Now the question is: what is U in the Equation (2.123)? In general, the U can be determined by the carrier density from the Poisson's equation

$$\rho = -eN = \int \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2} D(E - U) dE \quad (2.124)$$

$$\nabla \cdot \vec{\varepsilon} = \frac{\rho}{\epsilon_r} \Rightarrow \frac{\partial^2 \phi}{\partial x^2} = -\frac{\rho}{\epsilon_r} \quad (2.125)$$

where ϕ is the electrostatic potential. To solve the Equation (2.125), the finite difference method is used.

$$\phi_{i+1} + \phi_{i-1} - 2\phi_i = -\frac{a^2 \rho_i}{\epsilon_i} \quad (2.126)$$

which can be expressed in the matrix form

$$\begin{bmatrix} 2 & -1 & 0 & \dots \\ -1 & 2 & -1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \end{bmatrix} = \frac{a^2}{\epsilon_r} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \end{bmatrix} \quad (2.127)$$

where a is the spacing. Here, the ρ_i can be determined by the density matrix $[\rho]$. Since the Hamiltonian \mathbf{H} is in the real space

$$\mathbf{H} = \begin{bmatrix} 2t + V_1 & -t & \dots \\ -t & 2t + V_2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \rightarrow \mathbf{H}(x, x') \quad (2.128)$$

but the eigenvalues and eigenvectors are in the k space, the basis transformation is needed to change the basis of $[\rho]$.

$$[\rho] \rightarrow \mathbf{M} \boldsymbol{\rho} \mathbf{M}^{-1} \rightarrow [\rho](x, x') \quad (2.129)$$

where the transformation between k and x spaces is the Fourier transformation. Since at the boundaries the ϕ_0 and ϕ_{N+1} are fixed, the Equation (2.127) can be rewritten as

$$\begin{bmatrix} 2 & -1 & 0 & \dots \\ -1 & 2 & -1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} = \begin{bmatrix} \frac{a^2}{\epsilon_r} \rho_{11} + \phi_0 \\ \frac{a^2}{\epsilon_r} \rho_{22} \\ \vdots \\ \frac{a^2}{\epsilon_r} \rho_{NN} + \phi_{N+1} \end{bmatrix} \quad (2.130)$$

where the left matrix is the diagonal of $[\rho]$. Note that the off-diagonal terms are omitted here since they do not matter for the transport problems. After solving the Poisson's equation to get the charge density and potential, the Hamiltonian is reconstructed and the new eigenvalues and eigenvectors are obtained. Then, the density matrix is used to get the charge density which is put back to the Poisson's equation. Once this iterations converge, the potential U is obtained and thus the current. Note that the above analysis is in **Quasi-Equilibrium** condition since in \mathbf{H} there is no information of $\mu_{1,2}$ and $\mu_{1,2}$ only exist in the density matrix $[\rho]$. If the system is far from the equilibrium, the **Non-Equilibrium Green's Function (NEGF)** should be involved¹⁶. However, we now don't know the $\gamma_{1,2}$ (coupling terms) in the Landauer's equation. If the contacts and device channel have the same material, e.g. n^+ Si and p-Si of nMOSFET, the γ would be a constant.

$$\gamma_1 = \gamma_2 = \gamma = \frac{\hbar}{\tau} \quad (2.131)$$

Thus, the current is

$$\begin{aligned} I &= \frac{2e}{\hbar} \int dE \frac{\gamma}{2} D(E - U) (f_1 - f_2) \\ &= \frac{2e}{\hbar} \int dE \frac{\hbar}{2\tau} D(E - U) (f_1 - f_2), \quad \tau = \frac{a}{v} \\ &= \frac{2e}{\hbar} \frac{1}{2} \frac{1}{a} \int v(E) D(E - U) (f_1 - f_2) dE \end{aligned} \quad (2.132)$$

Finally, we get the current density

$$\begin{aligned} J &= \frac{I}{A} = \frac{e}{Aa} \int v(E) D(E - U) (f_1 - f_2) dE \\ &= e \int v(E) D_V(E - U) (f_1 - f_2) dE \end{aligned} \quad (2.133)$$

where $V = Aa$, D_V is the DOS per unit volume per unit energy, and $v(E)$ is the velocity

$$v = \frac{1}{\hbar} \frac{\partial E}{\partial k} \quad (2.134)$$

Note that the Equation (2.133) is similar to the Drude model

$$J = env \quad (2.135)$$

¹⁶Out of the scope of EE230C.

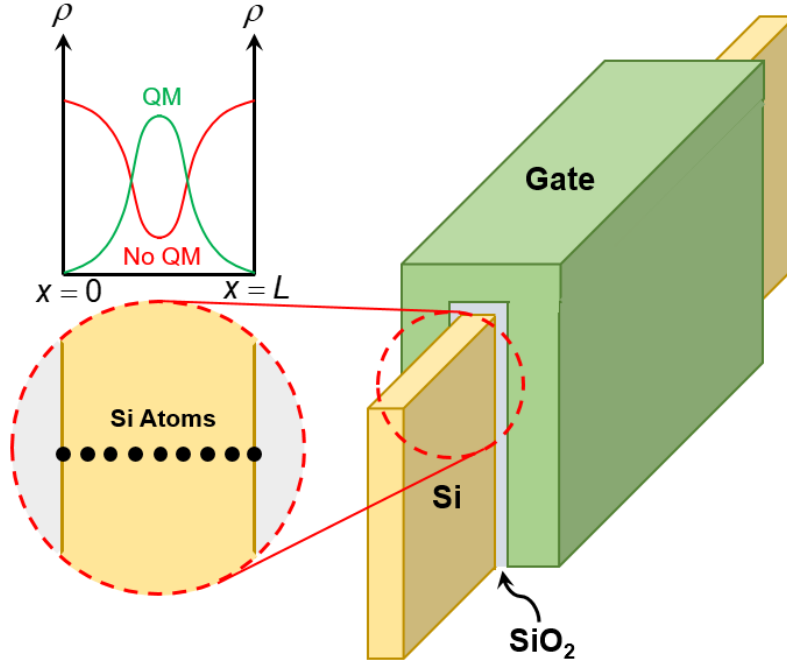


Figure 2.13: Schematic diagram of FinFET and the electron density profiles.

2.6.2 Capacitance in State-of-The-Art Transistor

FinFET is the state-of-the-art transistor that has been in the mass production. FinFET consists of a thin silicon fin wrapped by the oxide and metal as shown in Fig. 2.13. The charge distribution inside the silicon fin can be obtained by following the calculation procedure in the previous section. The Hamiltonian of silicon fin is constructed based on the chosen basis. Normally, only few bases is chosen to simplify and speed up the simulations for the transport problems in the commercialized TCAD tools. Furthermore, the convergence of such simulations is not guaranteed, so the grid for the silicon atom layers will be properly tuned. Once the electron density is determined, the capacitance can be easily obtained. Due to the charge centroid moves toward the center of fin (see Fig. 2.13), there is a charge layer separation (CLS) capacitance in series with the oxide capacitance.

$$\frac{1}{C} = \frac{1}{C_{\text{OX}}^{\text{SiO}_2}} + \frac{1}{C_{\text{CLS}}} = \frac{t_{\text{SiO}_2}}{\varepsilon_{\text{SiO}_2}} + \frac{t_{\text{centroid}}}{\varepsilon_{\text{Si}}} = \frac{1}{\varepsilon_{\text{SiO}_2}} \underbrace{\left(t_{\text{SiO}_2} + \frac{t_{\text{centroid}}}{3} \right)}_{\text{Effective Oxide Thickness (EOT)}} \quad (2.136)$$

We also can follow the definition of the differential capacitance

$$\begin{aligned} C &= \frac{\partial Q}{\partial V_g} = \frac{\partial Q}{\partial V_{\text{OX}} + \partial V_S} \\ \Rightarrow \frac{1}{C} &= \frac{\partial V_{\text{OX}}}{\partial Q} + \frac{\partial V_S}{\partial Q} = \frac{1}{C_{\text{OX}}} + \frac{\partial V_S}{\partial Q} \left(\equiv \frac{1}{C'} \right) \end{aligned} \quad (2.137)$$

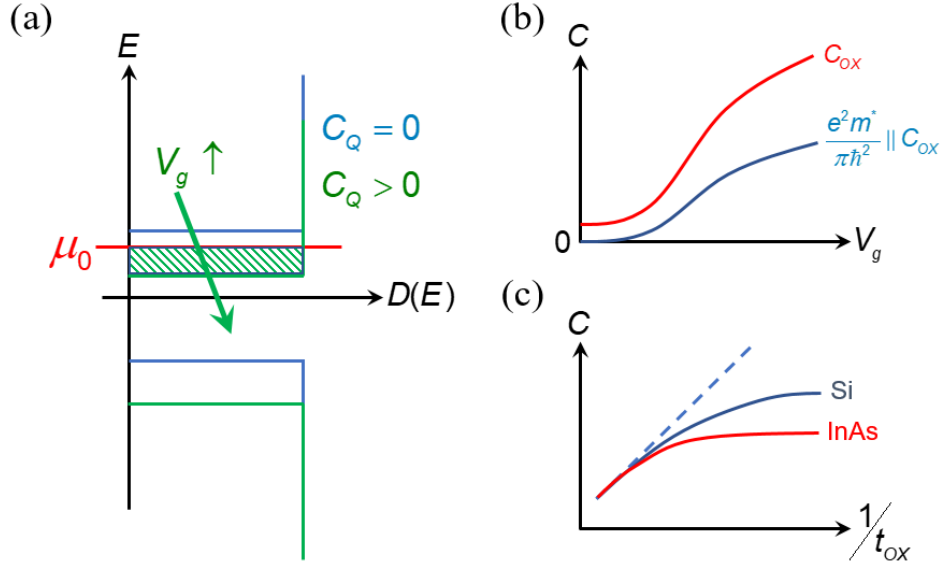


Figure 2.14: (a) The DOS of the silicon fin. (b) Capacitance versus V_g with and without the quantum capacitance. (c) Capacitance versus $\frac{1}{t_{ox}}$ for silicon and InAs

where V_S is the potential drop in the silicon fin and Q is the gate charge

$$\begin{aligned}
 Q &= e \int \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2} D(E - U) dE \\
 &= e \int f_0 D(E - U) dE, \quad \text{if } \gamma_1 = \gamma_2 \text{ and } f_1 = f_2 = f_0
 \end{aligned} \tag{2.138}$$

Then,

$$\begin{aligned}
 \frac{\partial Q}{\partial V_S} &= \frac{\partial Q}{\partial U} (-e) \\
 &= -e^2 \frac{\partial}{\partial U} \int f_0(E) D(E - U) dE \\
 &= -e^2 \frac{\partial}{\partial U} \int f_0(E' + U) D(E') dE' \\
 &= -e^2 \int dE' \left(-\frac{\partial f_0}{\partial U} \right) D(E') \\
 &= e^2 D|_{E=\mu}, \quad \text{at } T = 0K \\
 &\equiv C_Q
 \end{aligned} \tag{2.139}$$

where C_Q is called the quantum capacitance. Since the fin height and gate length is large enough, only the fin thickness is quantum mechanically confined which means that the silicon fin is actually a two-dimensional system. Therefore, the 2D DOS is considered as shown in Fig. 2.14(a) and the quantum capacitance is

$$C_Q = \frac{m^* e^2}{\pi \hbar^2} = \frac{\varepsilon m^* e^2}{(4\pi \varepsilon \hbar^2)^{\frac{1}{4}}} = \frac{\varepsilon}{\left(\frac{4\pi \varepsilon \hbar^2}{m^* e^2}\right)^{\frac{1}{4}}} = \frac{\varepsilon}{\frac{a_B}{4}} \quad (2.140)$$

where $a_B = \left(\frac{4\pi \varepsilon_0 \hbar^2}{m_0 e^2}\right) \varepsilon_r \frac{m_0}{m^*}$ is the Bohr radius of the material. The Bohr radius of silicon is about 3 nm, so the C_Q is $\frac{11.9}{0.75}$. If SiO₂ thickness is 1 nm, the ratio of C_Q to C_{OX} would be 4, which significantly degrades the gate capacitance as shown in Fig. 2.14(b). In III-V materials such as InAs, the effective mass is even smaller which further reduces C_Q and thus smaller capacitance (less charges). Since the mobility of the III-V materials is always high compared to the silicon, there may be a room for III-V materials to beat the silicon. However, the experiments show that the silicon has better performance against the III-V materials. Another benefit can be obtained from the III-V materials is the speed or the delay (τ)

$$\tau = \frac{CV}{I} = \frac{V}{\lambda I_{Si}} \lambda' C_{Si} \quad (2.141)$$

If $\lambda' < \lambda$, the materials could give better speed than the silicon. One of the possible candidates is the 2D material such as MoS₂. However, since the silicon FinFET already has a vertical nature, which significantly increases the number of transistors in a given area, it seems impossible to have the 2D material grown in the vertical direction. Therefore, the silicon at this stage may be the best choice.

2.7 Slater-Koster Tight-Binding Method

In the previous sections, only one basis function is considered at each atom side. However, few wavefunctions from different orbitals may involve. Thus, the contribution and interaction due to these orbitals will be included and called the **Slater-Koster Tight-Binding Method**¹⁷. The general equation of the tight-binding method has been given in the Equation (2.69).

$$\sum_i [H]_i + [H]_{i,j} e^{i\vec{k} \cdot (\vec{r}_j - \vec{r})}$$

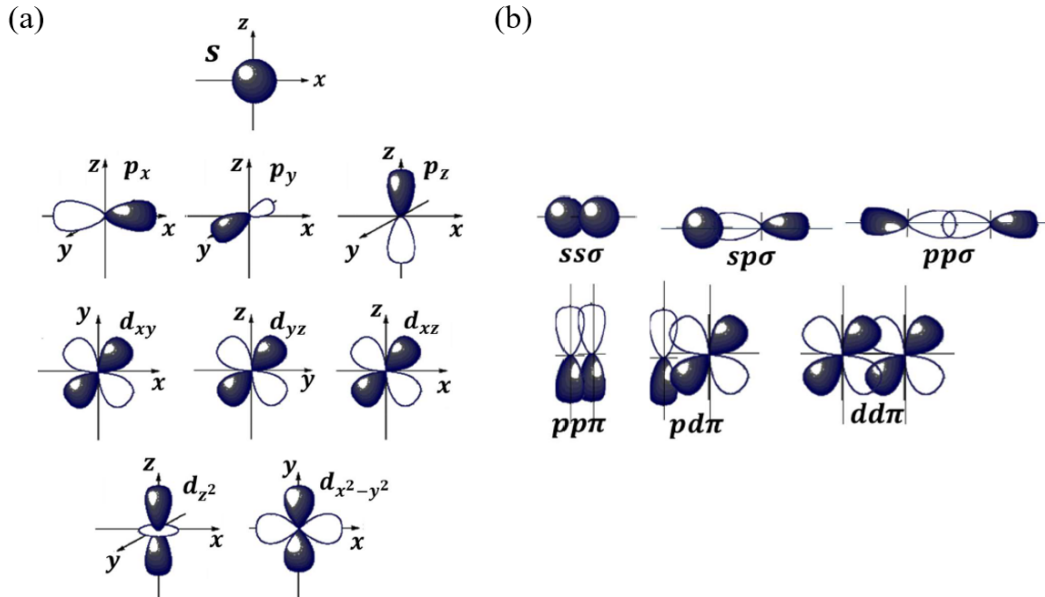
In $[H]_n$ and $[H]_{n,m}$, the overlap of different orbitals between each atom side should be considered. Solving from the Schrodinger equation for the atomic orbitals (Hydrogen-like orbitals), we have the wavefunction solution in the spherical coordinate

$$\psi_{n,l,m_l}(r, \theta, \phi) = e^{-\frac{r}{na_0}} \times f_{n,l}(r) \times Y_l^{m_l}(\theta, \phi) \quad (2.142)$$

where n , l ($0 \leq l \leq n-1$), and m_l ($-l \leq m_l \leq l$) are the quantum numbers, a_0 is the Bohr radius, the first two terms ($e^{-\frac{r}{na_0}} \times f_{n,l}(r)$) are only radius-dependent and is given by n (the principal quantum number), and $Y_l^{m_l}(\theta, \phi)$ is angular-dependent and gives s , p , d , and f orbitals (spherical harmonics) as shown in Fig. 2.15. These orbitals form a new set

l	m_l	$ Y_l^m(\theta, \phi) $	Atomic orbitals for $l = 0, 1, 2$
$l = 0$	$m_l = 0$	$s = 1/(2\sqrt{\pi})$	$s = 1/(2\sqrt{\pi})$
$l = 1$	$m_l = 1$	$p_{+1} = \sqrt{3}/(2\sqrt{2\pi}) \sin \theta e^{i\phi}$	$p_x = \frac{1}{2}(p_{+1} + p_{-1}) = \sqrt{3}/(2\sqrt{\pi}) \sin \theta \cos \phi$
	$m_l = 0$	$p_0 = \sqrt{3}/(2\sqrt{\pi}) \cos \theta$	$p_y = \frac{-i}{2}(p_{+1} - p_{-1}) = \sqrt{3}/(2\sqrt{\pi}) \sin \theta \sin \phi$
	$m_l = -1$	$p_{-1} = \sqrt{3}/(2\sqrt{2\pi}) \sin \theta e^{-i\phi}$	$p_z = p_0 = \sqrt{3}/(2\sqrt{\pi}) \cos \theta$
$l = 2$	$m_l = 2$	$d_{+2} = \sqrt{15}/(4\sqrt{2\pi}) \sin^2 \theta e^{2i\phi}$	$d_{x^2-y^2} = \frac{1}{2}(d_{+2} + d_{-2}) = \sqrt{15}/(4\sqrt{\pi}) (\sin^2 \theta \cos^2 \phi - \sin^2 \theta \sin^2 \phi)$
	$m_l = 1$	$d_{+1} = \sqrt{15}/(2\sqrt{2\pi}) \sin \theta \cos \theta e^{i\phi}$	$d_{xy} = \frac{-i}{2}(d_{+2} - d_{-2}) = \sqrt{15}/(2\sqrt{\pi}) \sin \theta \cos \phi \sin \theta \sin \phi$
	$m_l = 0$	$d_0 = \sqrt{5}/(4\sqrt{\pi}) (3\cos^2 \theta - 1)$	$d_{xz} = \frac{1}{2}(d_{+1} + d_{-1}) = \sqrt{15}/(2\sqrt{\pi}) \sin \theta \cos \phi \cos \theta$
	$m_l = -1$	$d_{-1} = \sqrt{15}/(2\sqrt{2\pi}) \sin \theta \cos \theta e^{-i\phi}$	$d_{yz} = \frac{-i}{2}(d_{+1} - d_{-1}) = \sqrt{15}/(2\sqrt{\pi}) \sin \theta \sin \phi \cos \theta$
	$m_l = -2$	$d_{-2} = \sqrt{15}/(4\sqrt{2\pi}) \sin^2 \theta e^{-2i\phi}$	$d_{z^2} = d_0 = \sqrt{5}/(4\sqrt{\pi}) (3\cos^2 \theta - 1)$

Relate the subscripts $\left\{ \begin{array}{l} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta \end{array} \right.$

Figure 2.15: The Spherical harmonics function at various l and m_l .Figure 2.16: (a) The shapes of orbitals at various l and m_l . (b) The schematic diagram of different types of bondings.

of the basis functions and the overlap between these functions are computed. There are 10 independent Slater-Koster parameters (bondings) as shown in Fig. 2.16: $ss\sigma$, $sp\sigma$, $sd\sigma$, $pp\sigma$, $pp\pi$, $pd\sigma$, $pd\pi$, $dd\sigma$, $dd\pi$, and $dd\delta$. Since $[H]_{i,j}$ in the Equation (2.69) is just the nearest neighbor integral with the Hamiltonian, integrating it using functions provided in Fig. 2.15 gives the final results of $[H]_{i,j}$ ¹⁸

$$H_{ij} = \langle \psi_i | \mathbf{H} | \psi_j \rangle \quad (2.143)$$

where $\psi_{i,j}$ are the wavefunctions of orbitals. For example, in Graphene, we only need to consider p_z , d_{yz} , and d_{zx} which form the Slater-Koster parameters of $pp\pi$, $pd\pi$, $dd\pi$, and $dd\delta$, so totally the size of the basis is 6 (3 orbitals per atom \times 2 atoms per basis).

¹⁷This lecture was taught by Dr. Shehrin Sayed.

¹⁸Please refer to the TABLE I in J. C. Slater and G. F. Koster, Phys. Rev., vol. 94, no. 6, pp. 1498-1524, June 1954.

Chapter 3

Scattering

3.1 Free Electron

Suppose there is an electron traveling with a momentum k without experiencing the potential energy ($U = 0$). At $t = 0$, the scattering happens and the momentum changes from k to k' as shown in Fig. 3.1. Thus,

$$\begin{aligned} |\psi(x, t)\rangle &= \sum_k c_k(t) |\phi_k(x)\rangle \\ &= c_k(t) |\phi_k(x)\rangle, \quad t = 0 \\ &= |\phi_k(x)\rangle = e^{ikx} \end{aligned} \tag{3.1}$$

where

$$c_k(0) = \begin{cases} 1 & \text{for momentum} = k \text{ since we already knew } k \\ 0 & \text{for every other momentum} \end{cases} \tag{3.2}$$

The Schrodinger equation tells

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} |\psi(x, t)\rangle &= -\frac{\hbar^2}{2m_0} \nabla^2 |\psi(x, t)\rangle \\ \Rightarrow \sum_k i\hbar \frac{\partial}{\partial t} c_k |\phi_k(x, t)\rangle &= \sum_k -\frac{\hbar^2}{2m_0} \nabla^2 c_k |\phi_k(x, t)\rangle \end{aligned}$$

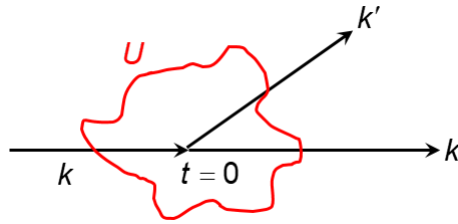


Figure 3.1: The electron with a momentum k and the scattering happens at $t = 0$.

$$\Rightarrow \sum_k |\phi_k(x, t)\rangle i\hbar \frac{\partial}{\partial t} c_k = \sum_k c_k \varepsilon_k |\phi_k(x, t)\rangle \quad (3.3)$$

$\langle \phi'_k(x, t) | :$

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} c_{k'} &= c_{k'} \varepsilon_{k'} \\ \Rightarrow \frac{\partial}{\partial t} c_{k'} &= -i\omega_{k'} c_{k'} \\ \Rightarrow \boxed{c_{k'}(t) = c_{k'}(0) e^{-i\omega_{k'} t}} \end{aligned} \quad (3.4)$$

3.2 Electron Experiencing Potential

If $U \neq 0$ and is a function of both position x and time t , the Schrodinger's equation becomes

$$i\hbar \frac{\partial}{\partial t} c_k(t) |\phi_k(x)\rangle = -\frac{\hbar^2}{2m_0} \nabla^2 c_k(t) |\phi_k(x)\rangle + U c_k(t) |\phi_k(x)\rangle \quad (3.5)$$

$\langle \phi'_k(x, t) | :$

$$i\hbar \frac{\partial}{\partial t} c_{k'} = c_{k'} \varepsilon_{k'} + \sum_k \langle \phi_{k'} | U | \phi_k \rangle c_k \quad (3.6)$$

If $\langle \phi_{k'} | U | \phi_k \rangle \equiv U_{k'k}(t)$ and $\frac{\varepsilon_{k'}}{\hbar} \equiv \omega_{k'}$, then

$$\boxed{\frac{\partial}{\partial t} c_{k'} + i\omega_{k'} c_{k'} = \frac{1}{i\hbar} \sum_k U_{k'k} c_k} \quad (3.7)$$

The general solution of the Equation (3.7) is

$$c_{k'}(T) = e^{-i\omega_{k'} T} \int_0^T dt \sum_k \frac{U_{k'k}}{i\hbar} c_k e^{i\omega_{k'} t} \quad (3.8)$$

Since the solution $c_{k'}$ contains itself c_k , the iteration is needed to get $c_{k'}$. Here, we adopt the c_k for the free electron Hamiltonian [Equation (3.4)] as a first-order approximation. Thus,

$$\boxed{c_{k'}(T) = e^{-i\omega_{k'} T} \int_0^T dt \sum_k \frac{U_{k'k}}{i\hbar} e^{i(\omega_{k'} - \omega_k) t}} \quad (3.9)$$

The scattering probability $P_{k'k}^T$ is

$$P_{k'k}^T = c_{k'} c_k = \frac{1}{\hbar^2} \left| \int_0^T \sum_k U_{k'k} e^{i(\omega_{k'} - \omega_k) t} dt \right|^2 \quad (3.10)$$

and the scattering rate $S_{k'k}$ is¹

$$\boxed{S_{k'k} = \frac{P_{k'k}^T}{T} = \frac{1}{T\hbar^2} \left| \int_0^T \sum_k U_{k'k} e^{i(\omega_{k'} - \omega_k) t} dt \right|^2} \quad (3.11)$$

¹The same results can be obtained using the **Perturbation Theory**. In fact, the first order approximation used in the Equation (3.9) is the perturbation theory itself.

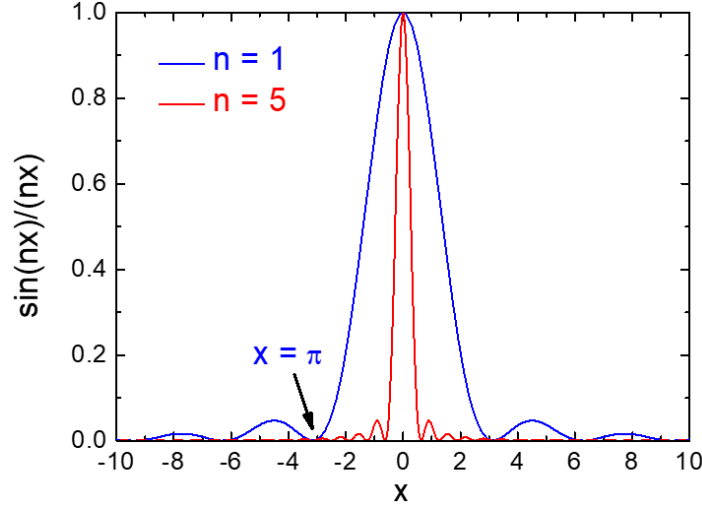


Figure 3.2: $\frac{\sin nx}{nx}$ versus x at $n = 1$ and 5 . When n is large, the function approaches to the delta function. In other words, if $\frac{\varepsilon T}{2\hbar} \gg \pi$, $\frac{\sin x}{x}$ becomes the delta function.

If $U_{k'k}$ is only a function of x

In practical cases, this approximation is valid since the impurities cannot move in the solids. The $U_{k'k}$ can be taken out from the integration in the Equation (3.11). Thus,

$$S_{k'k} = \frac{1}{T\hbar^2} \sum_k |U_{k'k}|^2 \left| \int_0^T e^{i\omega t} dt \right|^2 \quad (3.12)$$

where $\omega \equiv \omega_{k'} - \omega_k$. The integration in the Equation (3.12) can be evaluated analytically²

$$\begin{aligned} \left| \int_0^T e^{i\omega t} dt \right|^2 &= \left| \int_0^T (\cos \omega t + i \sin \omega t) dt \right|^2 \\ &= \frac{2}{\omega^2} (1 - \cos \omega T) = \frac{\sin \frac{\varepsilon T^2}{2\hbar}}{(\frac{\varepsilon T}{2\hbar})^2} T^2 \end{aligned} \quad (3.13)$$

where $\varepsilon \equiv \hbar\omega$. The function of $\frac{\sin x^2}{x^2}$ is shown in Fig. 3.2. When $T \gg \frac{2\pi\hbar}{\varepsilon}$ f-sec, $\frac{\sin \frac{\varepsilon T^2}{2\hbar}}{(\frac{\varepsilon T}{2\hbar})^2}$ can be approximated as a delta function. This approximation is valid in electronics since the scattering time is in order of p-sec. However, in optoelectronics, this approximation is not good because the scattering time as faster as f-sec is possible. To ensure that the integration is not changed after replacing $\frac{\sin \frac{\varepsilon T^2}{2\hbar}}{(\frac{\varepsilon T}{2\hbar})^2}$ with a delta function, the area under the function should be obtained

$$\int \frac{\sin \frac{\varepsilon T^2}{2\hbar}}{(\frac{\varepsilon T}{2\hbar})^2} d\varepsilon = \int \frac{\sin x^2}{x^2} dx \frac{2\hbar}{T} \quad (3.14)$$

²This is not a delta function since T is not infinity.

The above equation can be evaluated using the following trick

$$I(a) \equiv \int \frac{\sin ax^2}{x^2} dx \quad (3.15)$$

$$\frac{dI}{da} = 2 \int \frac{\sin ax}{x} dx = \pi \Rightarrow I(a) = a\pi \quad (3.16)$$

Then,

$$\int \frac{\sin x^2}{x^2} dx \frac{2\hbar}{T} = \frac{2\pi\hbar}{T} \quad (3.17)$$

Thus,

$$S_{k'k} = \frac{1}{T\hbar^2} \frac{2\pi\hbar}{T} \sum_k |U_{k'k}|^2 \delta(\varepsilon_{k'} - \varepsilon_k) T^2$$

Or

$$\boxed{S_{k'k} = \frac{2\pi}{\hbar} \sum_k |U_{k'k}|^2 \delta(\varepsilon_{k'} - \varepsilon_k)} \quad (3.18)$$

which is called the **Fermi Golden Rule**. From the above equation, if the potential is time-independent, the scattering will be elastic without losing or gaining any energy.

If $U_{k'k}$ is only a function of t

If the solid is illuminated by the light or is emitting the photons, the potential will be time-dependent. The potential $U_{k'k}$ can be generally expanded by the Fourier series

$$U_{k'k} = g(t) = \sum_{\omega} g_{\omega} e^{i\omega t} + \sum_{\omega} g_{-\omega} e^{-i\omega t} \quad (3.19)$$

Then,

$$\begin{aligned} S_{k'k} &= \frac{1}{T\hbar^2} \sum_{k,\omega} |g_{\omega}^{k'k}|^2 \left| \int_0^T e^{i(\omega_{k'} - \omega_k + \omega)t} dt \right|^2 + |g_{-\omega}^{k'k}|^2 \left| \int_0^T e^{i(\omega_{k'} - \omega_k - \omega)t} dt \right|^2 \\ &= \frac{2\pi}{\hbar} \sum_{k,\omega} |g_{\omega}^{k'k}|^2 \delta(\varepsilon_{k'} - \varepsilon_k + \hbar\omega) + \frac{2\pi}{\hbar} \sum_{k,\omega} |g_{-\omega}^{k'k}|^2 \delta(\varepsilon_{k'} - \varepsilon_k - \hbar\omega) \end{aligned} \quad (3.20)$$

The first term in the Equation (3.20) represents the emission process since

$$\varepsilon_{k'} = \varepsilon_k - \hbar\omega$$

and the second term represents the absorption process since

$$\varepsilon_{k'} = \varepsilon_k + \hbar\omega$$

Note that the cross term in the Equation (3.20) is ignored since physically the absorption and emission processes cannot happen simultaneously.

3.3 Master Equation

Since the scattering changes the momentum from k to k' , the distribution function of the electron also changes obeying the **Master Equation**

$$\frac{df_k}{dt} = \sum_{k'} S_{kk'} f_{k'} (1 - f_k) - \sum_{k'} S_{k'k} f_k (1 - f_{k'}) \quad (3.21)$$

where the first and second terms on the right side represent the scattering from k' to k and from k to k' . At equilibrium, $\frac{df_k}{dt} = 0$. Thus,

$$S_{kk'} f_{k'} (1 - f_k) = S_{k'k} f_k (1 - f_{k'})$$

or,

$$\frac{S_{kk'}}{S_{k'k}} = \frac{\frac{f_k}{1-f_k}}{\frac{f_{k'}}{1-f_{k'}}} = \frac{e^{-(\varepsilon_k - \mu)/k_B T}}{e^{-(\varepsilon_{k'} - \mu)/k_B T}} = e^{(\varepsilon_{k'} - \varepsilon_k)/k_B T} = e^{\hbar\omega/k_B T} \quad (3.22)$$

where $f_k = 1/[1 + e^{(\varepsilon_k - \mu)/k_B T}]$ and $\varepsilon_{k'} - \varepsilon_k \equiv \hbar\omega$. The Equation (3.22) indicates the restoration of the equilibrium. In other words, the probability of going from the higher energy state to the lower is proportional to $e^{\hbar\omega/k_B T}$. Recall the Equation (3.20). The emission and absorption have a similar relation

$$\frac{|g_\omega|^2}{|g_{-\omega}|^2} = e^{\hbar\omega/k_B T} \quad (3.23)$$

Since f_k is not a function of k' , it can be taken out from the summation. Thus, the Master equation becomes

$$\boxed{\frac{df_k}{dt} + \frac{f_k}{\tau} = \sum_{k'} S_{kk'} f_{k'} (1 - f_k)} \quad (3.24)$$

where

$$\frac{1}{\tau} = \sum_{k'} S_{k'k} (1 - f_{k'}) \quad (3.25)$$

is the lifetime (τ) of the electron due to the scattering. The general solution of the Equation (3.24) is

$$f_k = e^{-T/\tau} \int_0^T e^{t/\tau} \sum_{k'} S_{kk'} f_{k'} (1 - f_k) dt \quad (3.26)$$

Again, the Equation (3.26) needs to be solved numerically. Let's consider a simplified case: if an electron is coming with the momentum k , we have

$$\begin{cases} f_k = 1 & \text{for momentum} = k \text{ since we already knew } k \\ f_{k'} = 0 & \text{for every other momentum } k' \end{cases} \quad (3.27)$$

Thus,

$$1 - f_k \approx 0 \Rightarrow \frac{df_k}{dt} + \frac{f_k}{\tau} = 0 \Rightarrow f_k(t) = f_k(0) e^{-t/\tau} \quad (3.28)$$

3.4 Relaxation Time

The relaxation time is the time that a physical quantity \mathbf{A} is back to the thermal equilibrium and follows

$$\frac{d\langle \mathbf{A} \rangle}{dt} = \frac{\langle \mathbf{A} \rangle - \langle \mathbf{A}_0 \rangle}{\tau_A} \quad (3.29)$$

where τ_A is the relaxation time and

$$\langle \mathbf{A} \rangle = \langle \psi | \mathbf{A} | \psi \rangle = \text{Trace}(\mathbf{A}\rho) \quad (3.30)$$

Assume that there is no off-diagonal elements

$$\langle A \rangle_k = \langle \psi_k | A | \psi_k \rangle = f_k A_k \quad (3.31)$$

Thus,

$$\frac{d\langle \mathbf{A} \rangle}{dt} = \frac{d}{dt} \sum_k \langle A \rangle_k = \frac{d}{dt} \sum_k f_k A_k \quad (3.32)$$

Multiplying $\sum_k A_k$ to the Master equation, we have

$$\begin{aligned} \sum_k \frac{d}{dt} f_k A_k &= \sum_{kk'} \underbrace{S_{kk'} f_{k'} (1 - f_k) A_k}_{= S_{k'k} f_k (1 - f_{k'}) A_{k'}} - \sum_{kk'} S_{k'k} f_k (1 - f_{k'}) A_k \\ &= \sum_{kk'} S_{k'k} f_k (1 - f_{k'}) [A_{k'} - A_k] = \frac{\langle \mathbf{A} \rangle - \langle \mathbf{A}_0 \rangle}{\tau_A} \end{aligned} \quad (3.33)$$

Hence,

$$\begin{aligned} \frac{1}{\tau_A} &= \frac{\sum_{kk'} S_{k'k} f_k (1 - f_{k'}) [A_{k'} - A_k]}{\langle \mathbf{A} \rangle - \langle \mathbf{A}_0 \rangle} \\ &= \frac{\sum_{kk'} S_{k'k} f_k (1 - f_{k'}) [A_{k'} - A_k]}{\sum_k f_k A_k - \sum_k f_{k0} A_k} \\ &= \frac{\sum_{kk'} S_{k'k} f_k (1 - f_{k'})}{\sum_k f_k - f_{k0}} \left[1 - \frac{A_{k'}}{A_k} \right] \\ &= \sum_k \frac{1}{\tau_A}(k) \end{aligned} \quad (3.34)$$

where $\frac{1}{\tau_k}(k)$ is the k -dependent relaxation time

$$\boxed{\frac{1}{\tau_A}(k) = \sum_{k'} \frac{S_{k'k} f_k (1 - f_{k'})}{f_k - f_{k0}} \left[1 - \frac{A_{k'}}{A_k} \right]} \quad (3.35)$$

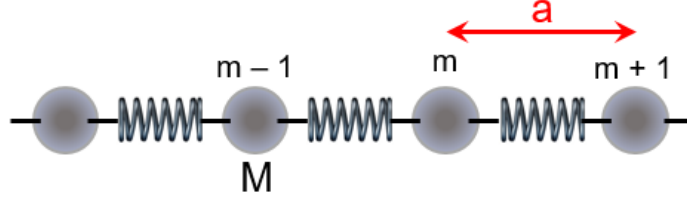


Figure 3.3: One dimensional lattice (lattice constant is a) with the springs connecting each atom with the mass M .

3.5 Phonons

The phonons are the quantized quasi-particles that represent the lattice vibration. The one-dimensional lattice can be modeled as the atoms connected to each other by the springs as shown in Fig. 3.3. The force equation says

$$F = M \frac{d^2 x_m}{dt^2} = K [(x_{m+1} - x_m) - (x_m - x_{m-1})] \quad (3.36)$$

where K is the spring constant and the x is the position of the atom. The solution of the Equation (3.36) is

$$\begin{aligned} \widetilde{x}_m &= \widetilde{x} e^{i(\beta m a - \omega t)} + \widetilde{x}^* e^{-i(\beta m a - \omega t)} \\ &= 2|\widetilde{x}| \cos(\beta m a - \omega t + \phi) \equiv 2|\widetilde{x}| \cos \theta \end{aligned} \quad (3.37)$$

where \widetilde{x} is the amplitude, β is the momentum, and ϕ is the phase. To determine the dispersion relation of phonons, we take the derivative to the Equation (3.37) with respect to time t twice and equate it to the Equation (3.36)

$$2|\widetilde{x}_m|(-\cos \theta)(-\omega)(-\omega) = -\frac{K}{M} 2|\widetilde{x}_m| 4 \cos \theta \sin^2 \frac{\beta a}{2} \quad (3.38)$$

Thus, the dispersion relation is

$$\boxed{\omega(\beta) = 2\sqrt{\frac{K}{M}} \sin \frac{\beta a}{2}} \quad (3.39)$$

When ω is close to zero, the Equation (3.39) can be expanded as a straight line. In other words,

$$\omega \propto \beta \quad \Rightarrow \quad \omega = c_s \beta \quad (3.40)$$

where c_s is the velocity of sound since the Equation (3.39) represents the acoustic phonon mode where all the atoms move in the same direction. For optical phonon mode, the frequency ω is nearly a constant say ω_0 and the two atoms move in the opposite direction. The kinetic energy of an atom in the acoustic mode at momentum β is

$$\frac{1}{2} M v^2 = 2M |\widetilde{x}_\beta|^2 \omega_\beta^2 \sin^2 (\vec{\beta} \cdot \vec{r} - \omega_\beta t) \quad (3.41)$$

If there are N atoms, the total energy will be

$$\begin{aligned} E^{\text{total}} &= 2M|\widetilde{x}_\beta|^2\omega_\beta^2N, \quad MN \equiv \rho V \\ &= 2\rho V|\widetilde{x}_\beta|^2\omega_\beta^2 \\ &= n_\beta\hbar\omega_\beta \end{aligned} \quad (3.42)$$

where ρ is the mass density and n_β is the number of phonons (fermion)

$$n_\beta = \frac{1}{e^{\frac{\hbar\omega_\beta}{k_B T}} - 1} \quad (3.43)$$

Hence,

$$|\widetilde{x}_\beta|^2 = \frac{\hbar n_\beta}{2\rho V\omega_\beta} \quad (3.44)$$

and

$$\widetilde{x}_\beta(\vec{r}, t) = \sqrt{\frac{\hbar}{2\rho V\omega_\beta}} \left[a_\beta e^{i(\vec{\beta} \cdot \vec{r} - \omega_\beta t)} + a_\beta^* e^{-i(\vec{\beta} \cdot \vec{r} - \omega_\beta t)} \right] \quad (3.45)$$

where

$$a_\beta^* a_\beta = n_\beta = \frac{1}{e^{\frac{\hbar\omega_\beta}{k_B T}} - 1} \quad (3.46)$$

3.5.1 Electron-Phonon Scattering

The electron-phonon scattering potential U_β^S can be approximated as

$$U_\beta^S(\vec{r}, t) = K_\beta \widetilde{x}_\beta(\vec{r}, t) \quad (3.47)$$

where K_β is the coupling constant related to the deformation of the lattice. For the acoustic mode, the movement of the atoms imposes the strain on the lattice, and thus

$$U_\beta^S = D_a S_\beta = D_a (\nabla \cdot \widetilde{x}_\beta) = i\beta D_a \widetilde{x}_\beta \equiv K_\beta \widetilde{x}_\beta \quad (3.48)$$

where D_a is the deformation potential and S_β is the strain. The Equation (3.47) is the result of the Taylor expansion of change in the potential. For the optical phonon mode, the frequency is nearly fixed. Thus,

$$K_\beta = D_o \quad (3.49)$$

Since we assume the phonon modes (ϕ_k) are the plane waves³, we have

$$\begin{aligned} U_{k'k}(t) &= \langle \phi_{k'} | U_\beta^S(\vec{r}, t) | \phi_k \rangle \\ &= K_\beta \sqrt{\frac{\hbar}{2\rho V\omega_\beta}} \left(\underbrace{\int a_\beta^* e^{-\vec{k}' \cdot \vec{r}} e^{-\vec{\beta} \cdot \vec{r}} e^{\vec{k} \cdot \vec{r}} e^{i\omega_\beta t}}_{\text{emission}} + \underbrace{\int a_\beta e^{-\vec{k}' \cdot \vec{r}} e^{\vec{\beta} \cdot \vec{r}} e^{\vec{k} \cdot \vec{r}} e^{-i\omega_\beta t}}_{\text{absorption}} \right) \end{aligned} \quad (3.50)$$

³S. Datta, Quantum Transport Atom to Transistor, 2005, p. 261.

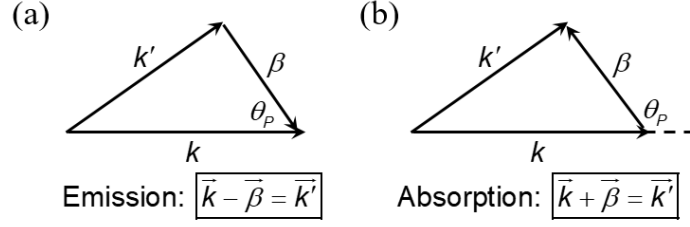


Figure 3.4: (a) Phonon emission and (b) absorption processes.

and

$$|U_{k'k}|^2 = \underbrace{|K_\beta|^2 \frac{\hbar a_\beta a_\beta^*}{2\rho V \omega_\beta} \delta(\vec{k} - \vec{k}' - \vec{\beta})}_{\text{emission}} + \underbrace{|K_\beta|^2 \frac{\hbar a_\beta^* a_\beta}{2\rho V \omega_\beta} \delta(\vec{k} - \vec{k}' + \vec{\beta})}_{\text{absorption}} \quad (3.51)$$

From the Equation (3.23), we have⁴

$$\frac{\text{emission} = a_\beta a_\beta^*}{\text{absorption} = a_\beta^* a_\beta = n_\beta} = e^{\hbar\omega/kT} \Rightarrow a_\beta^* a_\beta = n_\beta + 1 \quad (3.52)$$

Therefore, the scattering rate [the Equation (3.20)] becomes

$$S_{k'k} = \underbrace{\sum_{k,\beta} \frac{\pi(n_\beta + 1)}{\rho V \omega_\beta} |K_\beta|^2 \delta(\vec{k} - \vec{k}' - \vec{\beta}) \delta(\varepsilon_{k'} - \varepsilon_k + \hbar\omega_\beta)}_{\text{emission}} + \underbrace{\sum_{k,\beta} \frac{\pi n_\beta}{\rho V \omega_\beta} |K_\beta|^2 \delta(\vec{k} - \vec{k}' + \vec{\beta}) \delta(\varepsilon_{k'} - \varepsilon_k - \hbar\omega_\beta)}_{\text{absorption}} \quad (3.53)$$

Note that at $T = 0\text{K}$, the absorption $n_\beta = 0$ but the emission $n_\beta + 1 \neq 0$, indicating the spontaneous emission.

3.5.2 Simplification of Electron-Phonon Scattering Using Effective Mass Approximation

Since the Equation (3.53) involves the electronic band structure effect, it's difficult to solve in the real device geometry. Thus, the effective mass approximation is applied to simplify the problem. In the Equation (3.53), the momentum is conserved with the phonon momentum, so the momentum of the emission and absorption processes can be drawn as shown in Fig. 3.4. From the energy conservation of the emission process, we have

⁴Conceptually, $a_\beta a_\beta^*$ and $a_\beta^* a_\beta$ are the **lowering operator** and **raising operator** of the simple harmonic oscillator which satisfies $[a_\beta a_\beta^*, a_\beta^* a_\beta] = 1$ and thus $a_\beta^* a_\beta = n_\beta + 1$. See D. J. Griffiths, Introduction to Quantum Mechanics, 2005, p. 43-44.

$$\begin{aligned}
\varepsilon_{k'} - \varepsilon_k + \hbar\omega_\beta &= 0 \\
\Rightarrow \frac{\hbar^2}{2m} \left(k'^2 - k^2 + \frac{2m\omega_\beta}{\hbar} \right) &= 0, \quad k'^2 = k^2 + \beta^2 - 2k\beta \cos \theta_P \\
\Rightarrow \boxed{\cos \theta_P = \frac{m\omega_\beta}{\hbar k \beta} + \frac{\beta}{2k}} &\quad \text{For emission} \tag{3.54}
\end{aligned}$$

Similarly,

$$\boxed{\cos \theta_P = \frac{m\omega_\beta}{\hbar k \beta} - \frac{\beta}{2k}} \quad \text{For absorption} \tag{3.55}$$

Acoustic Phonon

For acoustic phonon, the dispersion relation can be simplified as [Equation (3.40)]

$$\omega_\beta = c_S \beta$$

Thus, the momentum β for the emission and absorption are

$$\beta^{\text{em}} = 2k \left(\cos \theta_P - \frac{mc_S}{\hbar k} \right) \tag{3.56}$$

$$\beta^{\text{ab}} = 2k \left(\frac{mc_S}{\hbar k} - \cos \theta_P \right) \tag{3.57}$$

Since $\cos \theta_P \leq 1$, for the emission to happen we have

$$\frac{mc_S}{\hbar k} < 1 \quad \Rightarrow \quad \frac{\hbar k}{m} > c_S \tag{3.58}$$

which is the Cherenkov condition saying that the velocity of the electron should be greater than the sound velocity to emit the acoustic phonon.

Optical Phonon

Since the frequency in the optical phonon mode is nearly fixed, we have

$$\cos \theta_P = \frac{m\omega_0}{\hbar k \beta^{\text{em}}} + \frac{\beta^{\text{em}}}{2k} \tag{3.59}$$

for the emission. Hence,

$$\beta^{\text{em}} = k \cos \theta_P \pm \sqrt{k^2 \cos^2 \theta_P - \frac{2m\omega_0}{\hbar}} \tag{3.60}$$

Since $\cos \theta_P \leq 1$, for the emission to happen we have

$$\frac{2m\omega_0}{\hbar} < k^2 \quad \Rightarrow \quad \frac{\hbar^2 k^2}{2m} > \hbar\omega_0 \tag{3.61}$$

indicating that the kinetic energy of the electron should be larger than that of the optical phonon for the emission.

Lifetime of Electron-Phonon Scattering

If the condition $\vec{k} - \vec{k}' \pm \vec{\beta} = 0$ has been held, the Equation (3.53) can be further simplified using the effective mass approximation

$$\begin{aligned}
 S_{k'k}(k) = & \underbrace{\frac{\pi m}{\rho V \hbar^2} \sum_{\beta} \frac{|K_{\beta}|^2 (n_{\beta} + 1)}{\omega_{\beta} k \beta} \delta \left(\cos \theta_P - \frac{m \omega_{\beta}}{\hbar k \beta} - \frac{\beta}{2k} \right)}_{\text{emission}} \\
 & + \underbrace{\frac{\pi m}{\rho V \hbar^2} \sum_{\beta} \frac{|K_{\beta}|^2 n_{\beta}}{\omega_{\beta} k \beta} \delta \left(\cos \theta_P - \frac{m \omega_{\beta}}{\hbar k \beta} + \frac{\beta}{2k} \right)}_{\text{absorption}}
 \end{aligned} \quad (3.62)$$

Recall the Equations (3.25), (3.27), and (3.28), the lifetime of the electron-phonon scattering is⁵

$$\frac{1}{\tau(k)} = \sum_{\beta} S_{k'k}^{\text{em}} + \sum_{\beta} S_{k'k}^{\text{ab}} \quad (3.63)$$

and the summation over β is

$$\sum_{\beta} \rightarrow \int \frac{d\beta_x d\beta_y d\beta_z}{\frac{2\pi}{L_x} \frac{2\pi}{L_y} \frac{2\pi}{L_z}} = \frac{V}{8\pi^3} \int \beta^2 d\beta (2\pi) \int_{-1}^1 d\cos \theta_P \quad (3.64)$$

The lifetime of the acoustic phonon emission is

$$\begin{aligned}
 \frac{1}{\tau^{\text{em}}(k)} &= \frac{\pi m}{\rho V \hbar^2} \frac{V}{8\pi^3} 2\pi \int_{\beta_{\min}}^{\beta_{\max}} \frac{\beta^2 d\beta |K_{\beta}|^2 (n_{\beta} + 1)}{\omega_{\beta} k \beta} \int_{-1}^1 \delta \left(\cos \theta_P - \frac{m \omega_{\beta}}{\hbar k \beta} - \frac{\beta}{2k} \right) d\cos \theta_P \\
 &= \frac{m D_a^2}{4\pi \rho \hbar^2} \frac{1}{k c_S} \int_{\beta_{\min}}^{\beta_{\max}} \beta^2 d\beta (n_{\beta} + 1), \quad \omega_{\beta} = c_S \beta \quad \text{and} \quad |K_{\beta}|^2 = \beta^2 D_a^2
 \end{aligned} \quad (3.65)$$

For the acoustic phonon, $\hbar \omega_{\beta} \ll k_B T$, and thus $n_{\beta} \approx \frac{k_B T}{\hbar \omega_{\beta}} \gg 1$ and $n_{\beta} + 1 \approx n_{\beta}$. Furthermore, $\beta_{\max}^{\text{em}} \approx 2k$ and $\beta_{\min}^{\text{em}} \approx 0$. The Equation (3.65) becomes

$$\frac{1}{\tau^{\text{em}}(k)} = \frac{m D_a^2}{4\pi \rho \hbar^3} \frac{k_B T}{k c_S^2} \int_0^{2k} \beta d\beta = \frac{m D_a^2 k_B T}{2\pi \rho \hbar^3 c_S^2} k \quad (3.66)$$

Similarly, the lifetime of the absorption is

$$\frac{1}{\tau^{\text{ab}}(k)} = \frac{m D_a^2 k_B T}{2\pi \rho \hbar^3 c_S^2} k \quad (3.67)$$

Therefore, the total lifetime due to the electron-phonon scattering is

$$\begin{aligned}
 \frac{1}{\tau(k)} &= \frac{m D_a^2 k_B T}{\pi \rho \hbar^3 c_S^2} k, \quad k = \frac{\sqrt{2mE(k)}}{\hbar} \\
 &= \frac{(2m k_B T)^{3/2} D_a^2}{2\pi \rho \hbar^4 c_S^2} \sqrt{\frac{E(k)}{k_B T}}
 \end{aligned} \quad (3.68)$$

⁵Summing over k' is the same as summing over β .

For GaAs,

$$\tau = 7.6 \sqrt{\frac{k_B T}{E}} \text{ psec}$$

Chapter 4

Semiclassical Transport

4.1 Boltzmann Transport Equation

In Chapter 2, we've discussed how to calculate the current with the Hamiltonian of the material. However, the brute force of solving this complicated problem is numerically inefficient. Therefore, a semiclassical way, which includes the quantum mechanical correction, is desired. The Fermi distribution can be written as

$$f_k = \frac{1}{1 + e^{(E_k - \mu)/(k_B T)}} = \frac{1}{1 + e^{(\frac{P^2}{2m} - \mu)/(k_B T)}} \quad (4.1)$$

Since $p = \hbar k$, the f_k in general is shown in Fig. 4.1. The f_k can be further extended into space-, momentum-, and time-dependence, i.e.,

$$f_k = g(\vec{r}, \vec{p}, t) \quad (4.2)$$

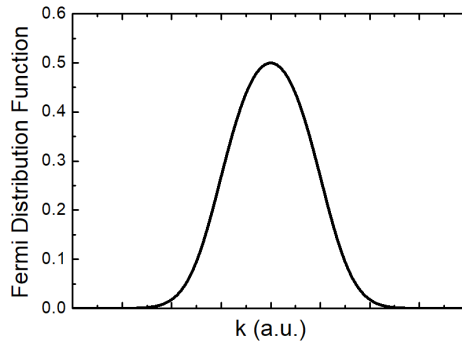


Figure 4.1: Fermi distribution function as a function of k .

where there are 7 dimensions (3 in space, 3 in momentum, and 1 in time). When applying the electric field \vec{E} , the Master equation says

$$\begin{aligned} \frac{df_k}{dt} &= \sum_{k'} S_{kk'} f_{k'} (1 - f_k) - \sum_{k'} S_{k'k} f_k (1 - f_{k'}) + \left. \frac{\partial f_k}{\partial t} \right|_{\vec{E}} \\ &\equiv \mathbf{S}^{\text{OP}} f_k + \left. \frac{\partial f_k}{\partial t} \right|_{\vec{E}} \end{aligned} \quad (4.3)$$

From the Newton's law, we have

$$\vec{F} = m \vec{a} = \frac{d\vec{p}}{dt} = -e\vec{E} \Rightarrow \Delta p_x = -eE_x \Delta t \quad (4.4)$$

and

$$\Delta x = v_x \Delta t \quad (4.5)$$

if the electric field is only in the x -direction. Assume that the electric field does not perturb the shape of the f after Δt .

$$f(\vec{r}, \vec{p}, t + \Delta t) = f(\vec{r} - v_x \Delta t, \vec{p} + eE_x \Delta t, t) \quad (4.6)$$

Apply the Taylor expansion to the Equation (4.6), we have

$$f_0 + \left. \frac{\partial f}{\partial t} \right|_{\vec{r}, \vec{p}, t} \Delta t = f_0 + \left. \frac{\partial f}{\partial r} \right|_{\vec{r}, \vec{p}, t} (-v_x \Delta t) + \left. \frac{\partial f}{\partial p} \right|_{\vec{r}, \vec{p}, t} eE_x \Delta t \quad (4.7)$$

If the scattering \mathbf{S}^{OP} is considered, then in general we have

$$\boxed{\frac{\partial f}{\partial t} = (\nabla_r f) \cdot (-\vec{v}) + (\nabla_p f) \cdot (e\vec{E}) + \mathbf{S}^{\text{OP}} f} \quad (4.8)$$

The above equation can be further simplified using the fact that

$$\nabla \cdot (f \vec{v}) = f \underbrace{(\nabla \cdot \vec{v})}_{=0} + \vec{v} \cdot (\nabla f) \quad (4.9)$$

where the first term at the right hand side is zero because there is no space-dependence in \vec{v} . Thus, the Equation (4.8) becomes

$$\boxed{\frac{\partial f}{\partial t} + \nabla \cdot (f \vec{v}) = e\vec{E} \cdot \nabla_p f + \mathbf{S}^{\text{OP}} f} \quad (4.10)$$

which is called the **Boltzmann Transport Equation (BTE)**.

4.2 Simplification of BTE

Suppose we have a measurable \mathbf{A}

$$\langle \mathbf{A} \rangle = \sum_k A_k f_k \quad (4.11)$$

Multiplying $\sum_k A_k$ to the BTE, we have

$$\frac{\partial \langle \mathbf{A} \rangle}{\partial t} + \nabla \cdot \sum_k A_k f_k v_k = \frac{e\vec{E}}{\hbar} \cdot \sum_k A_k \nabla_k f - \frac{\langle \mathbf{A} \rangle - \langle \mathbf{A} \rangle_0}{\langle \tau_A \rangle} \quad (4.12)$$

where $\langle \tau_A \rangle$ is the relaxation time as defined in the Equation (3.35), and we define

$$\vec{J}_A \equiv \sum_k A_k f_k v_k \quad (4.13)$$

$$\begin{aligned} G_A &\equiv \frac{e\vec{E}}{\hbar} \cdot \sum_k A_k \nabla_k f = \frac{e\vec{E}}{\hbar} \cdot \frac{V}{8\pi^3} \int_{-\infty}^{\infty} d^3k A_k \nabla_k f \\ &= \frac{e\vec{E}}{\hbar} \cdot \frac{V}{8\pi^3} \left[A_k \int_{-\infty}^{\infty} d^3k \nabla_k f - \int_{-\infty}^{\infty} \frac{dA_k}{dk} d^3k \int_{-\infty}^{\infty} d^3k \nabla_k f \right] \\ &= \frac{e\vec{E}}{\hbar} \cdot \frac{V}{8\pi^3} \left[f A_k \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} d^3k (f \nabla_k A_k) \right] \\ &= -\frac{e\vec{E}}{\hbar} \cdot \frac{V}{8\pi^3} \int_{-\infty}^{\infty} d^3k (f \nabla_k A_k) \\ &= -\frac{e\vec{E}}{\hbar} \sum_k f_k (\nabla_k A_k) \end{aligned} \quad (4.14)$$

$$R_A \equiv \frac{\langle \mathbf{A} \rangle - \langle \mathbf{A} \rangle_0}{\langle \tau_A \rangle} \quad (4.15)$$

If $\boxed{A_k = -e}$, we have

$$\frac{\partial \langle \mathbf{A} \rangle}{\partial t} = \frac{\partial}{\partial t} \sum_k f_k (-e) = -\frac{e \partial n}{\partial t} \quad (4.16)$$

$$\nabla \cdot \vec{J}_k = \nabla \cdot \sum_k (-e) f_k v_k = \nabla \cdot \vec{J} \quad (4.17)$$

$$G_A = -\frac{e\vec{E}}{\hbar} \sum_k f_k \nabla_k (-e) = 0 \quad (4.18)$$

$$\frac{1}{\langle \tau_k \rangle}(k) = \sum_{k'} \frac{S_{k'k} f_k (1 - f_{k'})}{f_k - f_{k0}} \left[1 - \frac{A_{k'}}{A_k} \right] = 0 \quad \text{since} \quad A_k = A_{k'} = -e \quad (4.19)$$

Therefore, we get the continuity equation from the BTE.

$$\boxed{e \frac{\partial n}{\partial t} = \nabla \cdot \vec{J}} \quad (4.20)$$

If $\boxed{A_k = -ev_k^x}$, we have

$$\frac{\partial \langle \mathbf{A} \rangle}{\partial t} = -e \frac{\partial}{\partial t} \left(\sum_k f_k v_k^x \right) = \frac{\partial J_x}{\partial t} \Rightarrow \langle \mathbf{A} \rangle = J_x \quad (4.21)$$

$$\nabla \cdot \vec{J}_k = \nabla \cdot \sum_k (-e) f_k (v_k^x)^2 \left(\frac{1}{2} m^* \right) \frac{2}{m^*} = -\frac{2e}{m^*} \nabla \cdot \langle \text{K.E.} \rangle = -\frac{2e}{m^*} \frac{\partial}{\partial x} \langle \text{K.E.} \rangle \quad (4.22)$$

$$G_A = -\frac{eE_x}{\hbar} \sum_k f_k \frac{\partial}{\partial k_x} (-ev_k^x) = \frac{e^2 E_x}{m^*} n \quad \text{since} \quad v_k^x = \frac{\hbar k_x}{m^*} \quad (4.23)$$

$$R_A = \frac{\langle \mathbf{A} \rangle - \langle \mathbf{A} \rangle_0}{\langle \tau_A \rangle} = \frac{J_x - J_0}{\langle \tau_m \rangle} = \frac{J_x}{\langle \tau_m \rangle} \quad (4.24)$$

where $\langle \tau_m \rangle$ is the momentum relaxation time and J_0 is the current density at equilibrium equal to zero. Therefore, the BTE becomes

$$\frac{\partial J_x}{\partial t} - \frac{2e}{m^*} \frac{\partial}{\partial x} \langle \text{K.E.} \rangle = \frac{e^2 E_x}{m^*} n - \frac{J_x}{\langle \tau_m \rangle} \quad (4.25)$$

At steady state, $\frac{\partial}{\partial t} = 0$. Thus,

$$\begin{aligned} \boxed{J_x} &= e \frac{e \langle \tau_m \rangle}{m^*} n E_x + \frac{2e}{m^*} \langle \tau_m \rangle \frac{\partial}{\partial x} \langle \text{K.E.} \rangle \\ &= \boxed{e \mu n E_x + 2\mu \langle U \rangle \frac{\partial n}{\partial x} + 2\mu n \frac{\partial}{\partial x} \langle U \rangle} \end{aligned} \quad (4.26)$$

where $\mu = \frac{e \langle \tau_m \rangle}{m^*}$ is the mobility and U is the average kinetic energy (K.E.) per electron

$$\langle \text{K.E.} \rangle = n \langle U \rangle$$

$$\frac{\partial}{\partial x} \langle \text{K.E.} \rangle = \langle U \rangle \frac{\partial n}{\partial x} + n \frac{\partial}{\partial x} \langle U \rangle \quad (4.27)$$

The third term at the right hand side of the Equation (4.26) represents the hot carrier effect happening in the high field region such as the drain-body junction of MOSFET. When the electric field E is small (near equilibrium), $\langle U \rangle = \frac{1}{2} k_B T$ and thus $\frac{\partial}{\partial x} \langle U \rangle = 0$. The Einstein relation also holds.

$$\frac{D}{\mu} = \frac{k_B T}{e} \quad (4.28)$$

where D is the diffusion constant. Therefore, the BTE becomes

$$\boxed{J_x = e \mu n E_x + \mu k_B T \frac{\partial n}{\partial x} = e \mu n E_x + e D \frac{\partial n}{\partial x}} \quad (4.29)$$

which is called the **drift-diffusion equation**. Note that for small device $\langle U \rangle \neq \frac{1}{2} k_B T$ and thus $\frac{\partial}{\partial x} \langle U \rangle \neq 0$ so that the drift-diffusion equation is not good.

Chapter 5

Transport in Small MOSFETs

5.1 Ballistic Transport

In a very small channel length MOSFET, the electron can travel through the device without any scatterings, which is called the ballistic transport. If the drain of MOSFET is biased at a high voltage V_{DS} , the band diagram would look like Fig. 5.1. The electron at the top of channel is the most important contribution to the current. This approach is called the **top-of-the-barrier model**. In general, the current at the top-of-barrier can be expressed as

$$I = (-e)W(n_S^+v_{inj}^+ - n_S^-v_{inj}^-) \quad (5.1)$$

where the $n_S^{+(-)}$ are the numbers of electrons that injects to the $+x$ - and $-x$ -directions, the $v_{inj}^{+(-)}$ are the injection velocities to the $+x$ - and $-x$ -directions, and W is the device width. The electron concentration Q_S at the top-of-barrier is related to the gate voltage V_G .

$$Q_S = C(V_G - V_{REF}) \quad (5.2)$$

where C is the gate capacitance and V_{REF} is the reference voltage that determines at which V_G there is no electron in the channel. Since the Fermi level of drain side is far away from the conduction band edge in the channel, the electron injection from the drain can be neglected

$$I \approx (-e)Wn_S^+v_{inj}^+ \quad (5.3)$$

and

$$n_S = n_S^+ + n_S^- \approx n_S^+ = \frac{C(V_G - V_{REF})}{-e} \quad (5.4)$$

Thus,

$$I = WC(V_G - V_{REF})v_{inj}^+ \quad (5.5)$$

which describes the **ballistic transport** in MOSFET.

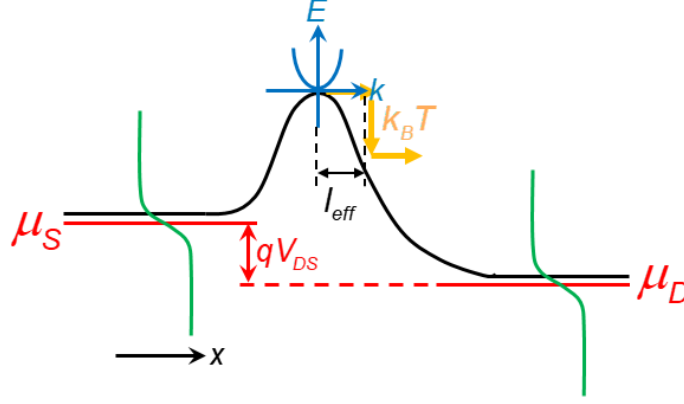


Figure 5.1: Band diagram of a small MOSFET.

5.2 Transport with Scatterings (Quasi-Ballistic Transport)

Since there are scatterings in the channel, there are some back-scattered electron n_S^- but they are only small portion of the $+x$ -direction injection

$$n_S^- = r n_S^+ \quad (5.6)$$

where r is the reflection coefficient. Note that the $n_S^{+(-)}$ are coming from the source side injection instead of the drain side as mentioned earlier. Since after scattering the electron only loses energy less or equal to $k_B T$, the back-scattered electron also has a velocity v_{inj}^- close to the v_{inj}^+ . Hence, the current becomes

$$\begin{aligned} I &= (-e) W n_S^+ v_{inj}^+ (1 - r), \quad n_S = n_S^+ + n_S^- = n_S^+ (1 + r) \\ &= (-e) n_S W v_{inj}^+ \frac{1 - r}{1 + r} W \\ &= W C (V_G - V_{REF}) v_{inj}^+ \frac{1 - r}{1 + r} \end{aligned} \quad (5.7)$$

Recall the transmission and reflection coefficients

$$T = \frac{\lambda}{\lambda + L}, \quad r = 1 - T = \frac{L}{\lambda + L} \quad (5.8)$$

The device length L should be replaced by l_{eff} as defined in Fig. 5.1 because the detailed simulations show that if the electron travels more than 1 to 2 times $\frac{k_B T}{q}$ down a potential drop, it is unlikely to re-emerge even if it does backscatter¹. Thus, the current becomes

$$I = W C (V_G - V_{REF}) v_{inj}^+ \frac{1}{1 + 2 \frac{l_{eff}}{\lambda}} \quad (5.9)$$

¹M. Lundstrom, Fundamentals of Carrier Transport, 2nd edition, p. 344.

Near equilibrium, the Einstein-Smoluchowski equation states that

$$\frac{2D}{\lambda} = v_T, \quad \frac{D}{\mu} = \frac{k_B T}{e} \quad (5.10)$$

where D is the diffusion coefficient, μ is the low-field mobility, and v_T is the thermal velocity. The electric field E near the top-of-barrier can be estimated using

$$E_{top} = \frac{\frac{k_B T}{e}}{l_{eff}} \Leftrightarrow l_{eff} = \frac{k_B T}{e E_{top}} \quad (5.11)$$

Hence,

$$\begin{aligned} I &= WC(V_G - V_{REF}) v_{inj}^+ \frac{1}{1 + \frac{2k_B T/e}{\lambda E_{top}}}, \quad \frac{2k_B T/e}{\lambda E_{top}} = \frac{2D}{\lambda} \frac{1}{\mu E_{top}} = \frac{v_T}{\mu E_{top}} \\ &= WC(V_G - V_{REF}) \frac{v_{inj}^+}{1 + \frac{v_T}{\mu E_{top}}} \end{aligned} \quad (5.12)$$

When the device is very long, E_{top} is very small and thus $v_T \gg \mu E_{top}$. Thus, the current is back to the standard MOSFET current equation

$$\begin{aligned} I &\approx W \mu C (V_G - V_{REF}) \left(\frac{v_{inj}^+}{v_T} \right) E_{top} \\ &\approx \frac{W}{L} \mu C (V_G - V_{REF}) \left(\frac{v_{inj}^+}{v_T} \right) V_{DS}, \quad V_{DS} \approx \frac{E_{top} L}{1} \end{aligned} \quad (5.13)$$

5.3 Generalized Ballistic Transport in One Dimension

The concept of the injection velocity has been introduced in the previous sections. The injection velocity can be interpreted as the average velocity that the electrons travel from the top-of-the-barrier.

$$\langle v \rangle = \frac{\sum_{k_x > 0} f v_{k_x}}{\sum_{k_x > 0} f} \quad (5.14)$$

where we define²

$$\begin{aligned} Z &\equiv \sum_{k_x > 0} f v_{k_x} = \frac{1}{2\pi} \int_0^\infty dk_x \frac{1}{1 + e^{\frac{E - \mu}{k_B T}}} \frac{\hbar k_x}{m^*}, \quad E = E_C + \frac{\hbar^2 k_x^2}{2m^*}, \quad \frac{\mu - E_C}{k_B T} \equiv \eta \\ &= \frac{1}{2\pi} \frac{\hbar}{m^*} \int_0^\infty k_x dk_x \frac{1}{1 + e^{\frac{\hbar^2 k_x^2}{2m^* k_B T} e^{-\eta}}}, \quad \frac{\hbar^2 k_x^2}{2m^* k_B T} \equiv x \\ &= \frac{1}{2\pi} \frac{\hbar}{m^*} \frac{m^* k_B T}{\hbar^2} \int_0^\infty dx \frac{1}{1 + e^{x - \eta}} \\ &= \frac{k_B T}{2\pi \hbar} \Gamma(1) \mathfrak{S}_0(\eta), \quad \mathfrak{S}_p(\eta) = \frac{1}{\Gamma(p+1)} \int_0^\infty dx \frac{x^p}{1 + e^{x - \eta}} \end{aligned} \quad (5.15)$$

²The pre-factor is $\frac{1}{2\pi}$ instead of $\frac{L}{2\pi}$ because we are calculating the electron "density."

and

$$\begin{aligned}
Z_2 &\equiv \sum_{k_x > 0} f = \frac{1}{2\pi} \int_0^\infty dk_x \frac{1}{1 + e^{\frac{\hbar^2 k_x^2}{2m^* k_B T} e^{-\eta}}} \\
&= \frac{1}{2\pi} \frac{m^* k_B T}{\hbar^2} \frac{\hbar}{\sqrt{2m^* k_B T}} \int_0^\infty dx \frac{x^{-1/2}}{1 + e^{x-\eta}} \\
&= \frac{1}{2\pi\hbar} \frac{\sqrt{m^* k_B T}}{\sqrt{2}} \sqrt{\pi} \mathfrak{S}_{-1/2}(\eta)
\end{aligned} \tag{5.16}$$

Therefore, the average velocity is

$$\langle v \rangle = \sqrt{\frac{2k_B T}{\pi m^*}} \frac{\mathfrak{S}_0(\eta)}{\mathfrak{S}_{-1/2}(\eta)} \tag{5.17}$$

where

$$\sqrt{\frac{2k_B T}{\pi m^*}} \equiv v_T \tag{5.18}$$

is the thermal velocity. If the semiconductor is non-degenerate doped ($\frac{E_C - \mu}{k_B T} \geq 3k_B T \Leftrightarrow \eta \geq -3$), the Fermi integral can be approximated by

$$\int_0^\infty dx \frac{1}{1 + e^{x-\eta}} \approx \int_0^\infty dx e^{-x} e^\eta = e^\eta \tag{5.19}$$

Thus,

$$Z = \frac{k_B T}{2\pi\hbar} e^\eta \tag{5.20}$$

and

$$Z_2 = \frac{\sqrt{m^* k_B T}}{2\pi\hbar} \frac{1}{\sqrt{2}} e^\eta \sqrt{\pi} \tag{5.21}$$

where

$$\int_0^\infty x^{-1/2} e^{-x} = \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{5.22}$$

For non-degenerate semiconductor, the average velocity is

$$\langle v \rangle = \sqrt{\frac{2k_B T}{\pi m^*}} \approx 1.2 \times 10^7 \text{ cm/sec for silicon} \tag{5.23}$$

The ballistic current at the top-of-the barrier considering the injection from the source and drain is³

$$I = W(-e) \underbrace{2}_{\text{spin}} \left(\sum_{k_x > 0, k_y} f v_{k_x} - \sum_{k_x < 0, k_y} f v_{k_x} \right) \tag{5.24}$$

³In Section 5.1 and 5.2, the injection from the drain is ignored.

where the first term is coming from the source side while the second term comes from the drain side

$$\begin{aligned}
F^+ &\equiv \sum_{k_x > 0, k_y} f v_{k_x} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_y \int_0^{\infty} dk_x \frac{1}{1 + e^{\frac{E - \mu_S}{k_B T}}} \frac{\hbar k_x}{m^*}, \quad E = E_C + \frac{\hbar^2}{2m^*} \underbrace{(k_x^2 + k_y^2)}_{k^2} \\
&= \frac{1}{4\pi^2} \int_0^{\pi} d\theta \int_0^{\infty} k dk \frac{1}{1 + e^{\frac{\hbar^2 k^2}{2m^* k_B T}} e^{-\eta_S}} \frac{\hbar}{m^*} k \sin \theta \\
&= \frac{1}{4\pi^2} \frac{\hbar}{m^*} 2 \frac{m^* k_B T}{\hbar^2} \frac{\sqrt{2m^* k_B T}}{\hbar} \underbrace{\int_0^{\infty} dx \frac{x^{1/2}}{1 + e^{x - \eta_S}}}_{= e^{\eta_S} \Gamma(\frac{3}{2}) \approx e^{\eta_S} \frac{\sqrt{\pi}}{2}, \text{ if non-degenerate}} \\
&\approx \underbrace{\frac{1}{2}}_{\text{moving in } +x \text{ direction}} \underbrace{\left(\frac{m^* k_B T}{2\pi \hbar^2} \right)}_{\text{number of electrons within } k_B T} \underbrace{\left(\sqrt{\frac{2k_B T}{\pi m^*}} \right)}_{= v_T} e^{\eta_S} \tag{5.25}
\end{aligned}$$

and

$$F^- \approx \frac{1}{2} \left(\frac{m^* k_B T}{2\pi \hbar^2} \right) \left(\sqrt{\frac{2k_B T}{\pi m^*}} \right) e^{\eta_D} \tag{5.26}$$

Therefore, the Equation (5.24) becomes

$$\begin{aligned}
I &= (-2e)W \frac{1}{2} \frac{m^* k_B T}{2\pi \hbar^2} v_T (e^{\eta_S} - e^{\eta_D}) \\
&= (-e)W \frac{m^* k_B T}{2\pi \hbar^2} v_T e^{\eta_S} \left[1 - e^{-(\eta_S - \eta_D)} \right] \\
&= (-e)W \frac{m^* k_B T}{2\pi \hbar^2} v_T e^{\eta_S} \left[1 - e^{-\frac{eV_{DS}}{k_B T}} \right] \tag{5.27}
\end{aligned}$$

where

$$\eta_S - \eta_D = \frac{\mu_S - E_C}{k_B T} - \frac{\mu_D - E_C}{k_B T} = \frac{\mu_S - \mu_D}{k_B T} = \frac{eV_{DS}}{k_B T} \tag{5.28}$$

To link the Equation (5.27) to (5.1), we can calculate the charge density for $+x$ and $-x$ directions

$$\begin{aligned}
n_S^+ &= \underbrace{2}_{\text{spin}} \sum_{k_x > 0, k_y} f = 2 \frac{1}{4\pi^2} \int_0^{\pi} d\theta \int_0^{\infty} k dk \frac{1}{1 + e^{\frac{\hbar^2 k^2}{2m^* k_B T}} e^{-\eta_S}} \\
&= 2 \frac{1}{4\pi^2} \pi \frac{m^* k_B T}{\hbar^2} \int_0^{\infty} \frac{dx}{1 + e^{x - \eta_S}} = \left(\frac{m^* k_B T}{2\pi \hbar^2} \right) e^{\eta_S} \tag{5.29}
\end{aligned}$$

and

$$n_S^- = 2 \sum_{k_x < 0, k_y} f = \left(\frac{m^* k_B T}{2\pi \hbar^2} \right) e^{\eta_D} \tag{5.30}$$

The totla charge density n_S is

$$n_S = n_S^+ + n_S^- = \left(\frac{m^* k_B T}{2\pi \hbar^2} \right) (e^{\eta_S} + e^{\eta_D}) = \left(\frac{m^* k_B T}{2\pi \hbar^2} \right) e^{\eta_S} \left(1 + e^{-\frac{eV_{DS}}{k_B T}} \right) \tag{5.31}$$

Therefore, the Equation (5.27) becomes

$$\begin{aligned}
 I &= \underbrace{(-e)n_S}_{\equiv Q_S = C(V_G - V_{REF})} W v_T \left(\frac{1 - e^{-\frac{eV_{DS}}{k_B T}}}{1 + e^{-\frac{eV_{DS}}{k_B T}}} \right) \\
 &= WC(V_G - V_{REF}) v_T \underbrace{\left(\frac{1 - e^{-\frac{eV_{DS}}{k_B T}}}{1 + e^{-\frac{eV_{DS}}{k_B T}}} \right)}_{\equiv v_{inj}} \quad (5.32)
 \end{aligned}$$

As increasing V_{DS} , the injection velocity v_{inj} gradually approaches to the thermal velocity v_T which is similar to the velocity saturation concept in the traditional semiconductor device textbook. When $V_{DS} < \frac{k_B T}{e}$, $e^{-\frac{eV_{DS}}{k_B T}} \approx 1 - \frac{eV_{DS}}{k_B T}$. Thus, the Equation (5.32) becomes

$$\begin{aligned}
 I &\approx WC(V_G - V_{REF}) v_T \frac{V_{DS}}{2k_B T/e} \\
 &= \frac{W}{L} C(V_G - V_{REF}) \underbrace{\left(\frac{Lv_T}{2k_B T/e} \right)}_{\mu_B} V_{DS} \quad (5.33)
 \end{aligned}$$

where μ_B is the ballistic mobility⁴ since

$$v_T = \sqrt{\frac{2k_B T}{\pi m^*}} \Rightarrow 2k_B T = v_T^2 \pi m^* \quad (5.34)$$

$$\Rightarrow \frac{Lv_T}{2k_B T/e} = \frac{e}{m^*} \underbrace{\frac{L}{\pi v_T}}_{\tau} \equiv \frac{e}{m^*} \tau \equiv \mu_B \quad (5.35)$$

It is observed that the ballistic mobility is a linear function of L . As shortening L , the ballistic mobility decreases which is different from the prediction of scatterings in the channel. In other words, when L is large enough, the mobility should decrease with longer length since there are more scattering. Therefore, this linear relation provides us the insight of the ballistic transport happening.

5.4 Generalized Quasi-Ballistic Transport

In previous section, we've derived the ballistic current model. Now, let's consider the reflection in the injection from both source and drain, i.e., quasi-ballistic current model. The electron density at the top-of-the-barrier is

$$n_S = n_S^+(1 - r_S) + n_S^-(1 - r_D) + r_S n_S^+ = n_S^+ + n_S^-(1 - r_D) \quad (5.36)$$

⁴M. S. Shur, "Low Ballistic Mobility in Submicron HEMTs," IEEE Electron Device Letter, vol. 23, no. 9, pp. 511-513, Sept. 2002.

where $r_{S,D}$ are the reflection coefficients from the source and drain. Furthermore, the quasi-ballistic current can be expressed as

$$\begin{aligned}
I &= 2(-e)W \left[\sum_{k_x > 0, k_y} f v_{k_x} (1 - r_S) - \sum_{k_x < 0, k_y} f v_{k_x} (1 - r_D) \right] \\
&= 2(-e)W \left[(1 - r_S) \sum_{k_x > 0, k_y} f v_{k_x} - (1 - r_D) \sum_{k_x < 0, k_y} f v_{k_x} \right] \\
&= (-e)W \frac{m^* k_B T}{2\pi \hbar^2} v_T e^{\eta_S} \left[(1 - r_S) - (1 - r_D) e^{-\frac{eV_{DS}}{k_B T}} \right] \\
&= WC(V_G - V_{REF}) v_T \left[\frac{(1 - r_S) - (1 - r_D) e^{-\frac{eV_{DS}}{k_B T}}}{1 + (1 - r_D) e^{-\frac{eV_{DS}}{k_B T}}} \right] \\
&= WC(V_G - V_{REF}) v_T \frac{1 - r_S}{1 + r_S} \frac{1 - \frac{1 - r_D}{1 - r_S} e^{-\frac{eV_{DS}}{k_B T}}}{\frac{1}{1 + r_S} + \frac{1 - r_D}{1 + r_S} e^{-\frac{eV_{DS}}{k_B T}}} \quad (5.37)
\end{aligned}$$

where

$$n_S = \left(\frac{m^* k_B T}{2\pi \hbar^2} \right) e^{\eta_S} \left[1 + (1 - r_D) e^{-\frac{eV_{DS}}{k_B T}} \right] = C(V_G - V_{REF}) \quad (5.38)$$

Recall the Equation (5.8) and (5.9), the Equation (5.37) becomes

$$I = WC(V_G - V_{REF}) \frac{v_T}{1 + \frac{2l_{eff}}{\lambda}} \frac{1 - \frac{1 - r_D}{1 - r_S} e^{-\frac{eV_{DS}}{k_B T}}}{\frac{1}{1 + r_S} + \frac{1 - r_D}{1 + r_S} e^{-\frac{eV_{DS}}{k_B T}}} \quad (5.39)$$

Normally $r_{S,D}$ are close to zero since the electrons are not localized. In practical model fitting, $r_{S,D}$ are often the model parameters for fitting.

5.5 Anisotropic Effective Mass in Ballistic Transport

In previous section, the isotropic effective mass is assumed, i.e., $m_x = m_y = m^*$. Now, if $m_x \neq m_y$, the charge density (n_S) and electron flux (F) should be re-written by modifying the Fermi function.

$$f = \frac{1}{1 + e^{\frac{E - \mu_{S,D}}{k_B T}}} = \frac{1}{1 + e^{\left(\frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y} \right) / k_B T - \frac{(\mu_{S,D} - E_C)}{k_B T}}} \quad (5.40)$$

If defining

$$k'_x = \frac{k_x}{\sqrt{m_x}}, \quad k'_y = \frac{k_y}{\sqrt{m_y}}, \quad k'^2 = k_x'^2 + k_y'^2 \quad (5.41)$$

the similar integrations to the Equation (5.25), (5.26), (5.29), and (5.30), can be made assuming non-degenerated doping and give

$$n_S = \frac{\sqrt{m_x m_y} k_B T}{2\pi \hbar^2} e^{\eta_S} \left(1 + e^{-\frac{eV_{DS}}{k_B T}} \right) \quad (5.42)$$

and

$$F^{+,-} = \frac{1}{2} \frac{\sqrt{m_x m_y} k_B T}{2\pi \hbar^2} \sqrt{\frac{2k_B T}{\pi m_x}} e^{\eta_{S,D}} \quad (5.43)$$

Therefore, the ballistic current considering the anisotropic effective mass is

$$I = (-e)W \frac{\sqrt{m_x m_y} k_B T}{2\pi \hbar^2} \sqrt{\frac{2k_B T}{\pi m_x}} e^{\eta_S} \left(1 - e^{-\frac{eV_{DS}}{k_B T}} \right) \quad (5.44)$$

Notice that the current only depends on the effective mass in the direction perpendicular to the transport, which determines the electron density. Although the silicon has a high DOS close to metal, the gate dielectric cannot support such high electric field that achieves the high electron density. Now, if there are multiple subbands (assume z -direction is confined while x and y are not, i.e., two-dimensional electron gas), the total current is

$$I = \sum_{n=1}^N (-e)W \frac{\sqrt{m_x m_y} k_B T}{2\pi \hbar^2} \sqrt{\frac{2k_B T}{\pi m_x}} e^{\frac{\mu_S - E_C^n}{k_B T}} \left(1 - e^{-\frac{eV_{DS}}{k_B T}} \right) \quad (5.45)$$

where

$$E_C^n = E_{C0} + \frac{\hbar^2 n^2 \pi^2}{2m_z L_z^2} + \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y} \quad (5.46)$$

is the band edge of the n th subband, N is the total number of subbands, and L_z is the z -direction thickness. Note that the band-to-band scattering is ignored here to simplify the problem.

5.6 Issues of Scaling MOSFETs

Degraded Threshold Voltage and Subthreshold Slope⁵

In short channel MOSFETs, the electrostatic control from the drain becomes comparable to the gate control as shown in Fig. 5.2. Similar to the concept in Chapter 1, the potential at the top-of-the-barrier can be written as

$$\psi_S = \frac{C_{OX}}{C_{OX} + C_{dep} + C_D} V_{GS} + \frac{C_D}{C_{OX} + C_{dep} + C_D} V_{DS} \quad (5.47)$$

where C_{dep} is the depletion capacitance of the channel. From the Equation (5.47), C_D gradually increase with shortening the gate length, and thus the drain voltage starts to control the barrier. If V_{DS} is high, the threshold voltage (V_{TH}) is significantly degraded. This is called the **drain-induced barrier lowering (DIBL)**. Furthermore, since the channel potential (charges) is controlled by (shared with) the drain terminal (even the source terminal), the

⁵This lecture was taught by Prof. Jeffery Bokor. The topics of this section have been covered by EE230B. One may refer to: Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices, 2nd edition, 2009.

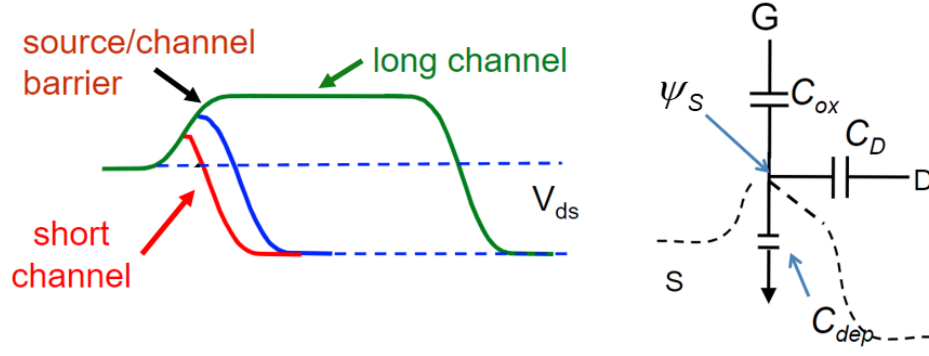


Figure 5.2: Illustration of channel potentials with various gate lengths and the equivalent circuit.

threshold voltage is also decreasing even when V_{DS} is low, which is known as V_{TH} **roll off**. The additional control from the drain terminal also deteriorates the subthreshold slope SS

$$SS \equiv \left[\frac{\partial \log I_{DS}}{\partial V_{GS}} \right]^{-1} = 60[\text{mV/dec}] \times \left(1 + \frac{C_{dep} + C_D}{C_{OX}} \right) \quad (5.48)$$

This is undesired since it significantly increases the OFF-state leakage current.

Gate Tunneling Leakage

Since the gate length is scaled to boost the current and allow more transistors in a give chip area, the gate oxide should be also scaled to improve the gate control as shown in the Equation (5.47). However, the thin oxide could lead to significant gate leakage due to the direct tunneling. Therefore, the high- κ dielectric was introduced in the 45-nm technology node to keep the gate capacitance but increase the physical oxide thickness. Nevertheless, there are still some difficulties of the high- κ dielectric: (1) chemical reactions between them and the silicon substrate and gate; (2) lower surface mobility than the Si/SiO₂ system; (3) too low a V_{TH} for P-channel MOSFET (as if there is positive charge in the high- κ dielectric); (4) long-term reliability; (5) a thin SiO₂ interfacial layer may be inserted between Si-substrate and high- κ film.

Mitigating Short Channel Effect of MOSFET

Since the channel potential is also controlled by C_{dep} and C_D , decreasing the depletion width by increasing the body doping is helpful but it may reduce the mobility due to the impurity scattering. Furthermore, using high- κ metal gate as mentioned earlier also overcomes the poly-depletion problem. To cut the leakage path underneath the channel due to lack of gate control, the shallow junction is adopted. However, it enlarges the series resistance and thus degrades the ON-state current. To improve the mobility, strained silicon is used in the channel. In addition to the mentioned techniques to improve the device performances of

the traditional MOSFET, new transistor structures [e.g. fully-depleted SOI (FDSOI) and FinFET] and new device operation mechanisms [e.g. tunnel FET (TFET) and negative capacitance FET (NCFET)] are proposed. The FDSOI can cut the leakage path and achieve low leakage, and the FinFET improves the gate control by increasing the gate capacitance. Furthermore, the TFET utilizes the tunneling mechanism which overcomes the problem from Boltzmann distribution, and the NCFET can achieve better subthreshold behavior since the channel potential can be boosted by the ferroelectric layer sandwiched by the metal gate and interfacial layer.

5.7 Application of the Developed Current Model

The ballistic and quasi-ballistic current models have been derived. In this section, how the developed model relates to the real measurement data will be examined. Recall the Equation (5.47). At low V_{GS} (let's say OFF-state), if the FDSOI or FinFET structure is adopted, the surface potential (ψ_S) in the channel is

$$\psi_S = \frac{C_{OX}}{C_{OX} + C_D} V_{GS} + \frac{C_D}{C_{OX} + C_D} V_{DS} \equiv \alpha_G V_{GS} + \alpha_D V_{DS} \quad (5.49)$$

where the contribution from the charge Q is relatively small. The quasi-ballistic current is

$$\begin{aligned} I^{OFF} &= (-e)W \frac{\sqrt{m_x m_y} k_B T}{2\pi \hbar^2} \sqrt{\frac{2k_B T}{\pi m_x}} e^{\frac{\mu_S}{k_B T}} e^{-\frac{E_C}{k_B T}} \left(1 - e^{-\frac{eV_{DS}}{k_B T}}\right) / \left(1 + \frac{2l_{eff}}{\lambda}\right), -E_C \rightarrow e\psi_S \\ &= (-e)W \frac{\sqrt{m_x m_y} k_B T}{2\pi \hbar^2} \sqrt{\frac{2k_B T}{\pi m_x}} e^{\frac{\mu_S}{k_B T}} e^{\frac{e\psi_S}{k_B T}} \left(1 - e^{-\frac{eV_{DS}}{k_B T}}\right) / \left(1 + \frac{2l_{eff}}{\lambda}\right) \\ &= A e^{\frac{\alpha_G V_{GS}}{k_B T/e}} e^{\frac{\alpha_D V_{DS}}{k_B T/e}} \left(1 - e^{-\frac{eV_{DS}}{k_B T}}\right) = B e^{\frac{\alpha_G V_{GS}}{k_B T/e}} \end{aligned} \quad (5.50)$$

Thus, the subthreshold slope is

$$\begin{aligned} SS &= \left[\frac{\partial \log I}{\partial V_{GS}} \right]^{-1} = \frac{2.3k_B T/e}{\alpha} = 60/\alpha_G [\text{mV/dec}] \\ &= 60 \times \left(1 + \frac{C_D}{C_{OX}}\right) [\text{mV/dec}] \equiv 60 \times m [\text{mV/dec}] \end{aligned} \quad (5.51)$$

where m is called the body factor. The DIBL is defined as

$$\text{DIBL} = \frac{V_{TH}^{low} - V_{TH}^{high}}{V_{DS}^{high} - V_{DS}^{low}} \quad [\text{mV/V}] \quad (5.52)$$

where the threshold voltage V_{TH} is defined using the constant current method

$$I_{TH} = A e^{\frac{e\psi_S^{high}}{k_B T}} = A e^{\frac{e\psi_S^{low}}{k_B T}} \frac{V_{DS}}{k_B T/e} \quad (5.53)$$

since

$$\left(1 - e^{-\frac{eV_{DS}}{k_B T}}\right) \approx \frac{eV_{DS}}{k_B T}, \quad \text{at low } V_{DS}$$

where "low" and "high" denote as low and high V_{DS} , and

$$\psi_S^{low} = \alpha_G V_{TH}^{low} + \alpha_D V_{DS}^{low} \quad (5.54)$$

and

$$\psi_S^{high} = \alpha_G V_{TH}^{high} + \alpha_D V_{DS}^{high} \quad (5.55)$$

Thus,

$$\text{DIBL} = \frac{\alpha_D}{\alpha_G} - \frac{1}{\alpha_G} \frac{\psi_S^{high} - \psi_S^{low}}{V_{DS}^{high} - V_{DS}^{low}} \quad (5.56)$$

However, from the Equation (5.53) we have

$$\psi_S^{high} - \psi_S^{low} = \frac{k_B T}{e} \left[\ln \frac{I_{TH}}{A} - \ln \frac{I_{TH}}{A} \frac{V_{DS}^{low}}{k_B T/e} \right] = \frac{k_B T}{e} \ln \frac{V_{DS}^{low}}{k_B T/e} \quad (5.57)$$

which is very small since V_{DS}^{low} is 50 mV⁶. Therefore,

$$\text{DIBL} = \frac{\alpha_D}{\alpha_G} = \frac{C_D}{C_{OX}} \quad (5.58)$$

which can be extracted from the subthreshold slope or the body factor.

⁶ $V_{DS} = 50$ mV is called the linear V_{DS} and is commonly used. If applying $V_{DS} = 26$ mV, the thermal noise could be an issue.