Chapter 5

The Transport Problem: Currents from Quantum Mechanics

5.1 Classical Drude model

$$m\frac{dv}{dt} = qE - \frac{mv}{\tau}, \tag{5.1}$$

steady state: $\frac{d}{dt}(...) \to 0$, yields

$$v = \frac{q\tau}{m}E = \mu E. \tag{5.2}$$

and current is

$$J = qnv = \frac{nq^2\tau}{m}E = \sigma E \implies \sigma_0 = \frac{nq^2\tau}{m}. \tag{5.3}$$

If the electric field was oscillating in time $E(t) = Ee^{i\omega t}$,

$$m\frac{dv}{dt} = qEe^{i\omega t} - \frac{mv}{\tau}, \tag{5.4}$$

assuming linear response $v(t) = v(0)e^{i\omega t}$, we get

$$\sigma(\omega) = \frac{\sigma_0}{1+i\omega\tau} = \underbrace{\frac{\sigma_0}{1+(\omega\tau)^2}}_{\mathrm{Re}(\sigma(\omega))} - i\underbrace{\frac{\omega\tau\sigma_0}{1+(\omega\tau)^2}}_{\mathrm{Im}(\sigma(\omega))}. \tag{5.5}$$

5.2 Quantum version

In semiconductor devices, we are often concerned with currents. A current is a measure of the flow of objects from one point in space to another. The flow of electric charges

constitutes an electric current, leading to the notion of electrical conductivity. In this chapter we develop the recipe to understand current flow from a quantum-mechanical viewpoint. Since the physical state of particles in quantum mechanics is represented by its wavefunction $\Psi(x,t)$, the current must be obtained from the wavefunction.

5.3 Probability current

Since $|\Psi(x,t)|^2 = \Psi^\star\Psi$ is the probability density, let's examine how it changes with time. We obtain

$$\frac{\partial|\Psi(x,t)|^2}{\partial t} = \Psi^\star\frac{\partial\Psi}{\partial t} + \frac{\partial\Psi^\star}{\partial t}\Psi, \tag{5.6}$$

where we use the time-dependent Schrodinger equation $i\hbar\partial\Psi/\partial t = (\hat{p}^2/2m + V)\Psi$ and its complex conjugate $-i\hbar\partial\Psi^\star/\partial t = (\hat{p}^2/2m + V)\Psi^\star$ to obtain

$$\frac{\partial|\Psi(x,t)|^2}{\partial t} = \Psi^\star\frac{(\hat{p}^2/2m + V)\Psi}{i\hbar} + \Psi\frac{(\hat{p}^2/2m + V)\Psi^\star}{-i\hbar}, \tag{5.7}$$

which simplifies to

$$\frac{\partial|\Psi(x,t)|^2}{\partial t} = \frac{1}{2mi\hbar}(\Psi^\star\hat{p}^2\Psi - \Psi\hat{p}^2\Psi^\star). \tag{5.8}$$

Since $\hat{p} = -i\hbar\nabla_{\mathbf{r}}$, we recognize the resulting equation

$$\frac{\partial|\Psi(x,t)|^2}{\partial t} = -\nabla_{\mathbf{r}}\cdot\big[\frac{1}{2m}(\Psi^\star\hat{p}\Psi - \Psi\hat{p}\Psi^\star)\big] \tag{5.9}$$

as the familiar 'continuity' equation in disguise. A continuity equation is of the form $\partial\rho/\partial t = -\nabla_{\mathbf{r}}\cdot\mathbf{j}$, where ρ is the particle 'density' and $\mathbf{j}$ is the current density. We read off the quantum mechanical current density as

$$\mathbf{j} = \frac{1}{2m}(\Psi^\star\hat{\mathbf{p}}\Psi - \Psi\hat{\mathbf{p}}\Psi^\star). \tag{5.10}$$

This equation provides us the required recipe for calculating the probability density flow, or current flow directly from the quantum mechanical wavefunctions of states. We make a few observations. If Ψ is real, $\mathbf{j} = 0$. Since Ψ has dimension of $1/\sqrt{Vol}$, the dimension

of $\mathbf{j}$ is per unit area per second. For 3D, volume is in m^3 and $\mathbf{j}$ is then in $1/(\mathrm{m}^2 \cdot \mathrm{s})$. For 2D $\mathbf{j}$ is in $1/(\mathrm{m} \cdot \mathrm{s})$, and it is simply $1/\mathrm{s}$ for 1D.

We also note that

$$\frac{d}{dt}(\int_{space} d^3r|\Psi|^2) = -\int_{space} d^3r \nabla \cdot \mathbf{j} = -\oint \mathbf{j} \cdot d\mathbf{S} = 0. \tag{5.11}$$

The conversion of the integral from volume to a closed surface uses Gauss' theorem. The value of the integral is zero because Ψ and consequently $\mathbf{j}$ goes to zero at infinity, and the equality must hold for all space. This equation is a statement of the indestructibility of the particle, which follows from $\int_{space} d^3r|\Psi|^2 = 1$. If the number of particles is *not* conserved, then one needs to add recombination ('annihilation') and generation ('creation') terms to the continuity equation. It then looks as $\partial\rho/\partial t = -\nabla \cdot \mathbf{j} + (G - R)$ where R and G are recombination and generation rates.

We also note that in the presence of a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$, the quantum-mechanical momentum operator $\hat{\mathbf{p}} \rightarrow \hat{\mathbf{p}} + q\mathbf{A}$ where q is the magnitude of the electron charge. This leads to an additional term in the expression of the current density

$$\mathbf{j} = \frac{1}{2m}(\Psi^\star \hat{\mathbf{p}}\Psi - \Psi\hat{\mathbf{p}}\Psi^\star) + \frac{q\mathbf{A}}{m}\Psi^\star\Psi. \tag{5.12}$$

The additional term depending on the magnetic vector potential $\mathbf{A}$ is useful to explain current flow in magnetic materials, magnetotransport properties, and superconductivity.

5.4 Charge current

Lets focus on determining the electric current. To account for the flow of *charge*, the probability density current $\mathbf{j}$ is modified simply to $\mathbf{J} = q\mathbf{j}$, where q is the charge (in Coulombs) of the charge particle. We assume these charge particles are electrons and $q = 1.6 \times 10^{-19}$ C and free mass $m_e = 9.1 \times 10^{-31}$ kg. In the absence of a magnetic field, the electric current density is then given by

$$\mathbf{J} = \frac{q}{2m_e}(\Psi^\star \hat{\mathbf{p}}\Psi - \Psi\hat{\mathbf{p}}\Psi^\star), \tag{5.13}$$

which is now in $\mathrm{A/m}^2$ for 3D, A/m for 2D, and A for 1D current flow, where A=C/s is the unit of current in Amperes. The current density is expressed in terms of the electron wavefunctions. We wish to make the expression more 'usable'.

Consider free electrons in 1D with periodic boundary conditions between $x = (0, L)$. The wavefunction for a state $|k\rangle$ of definite energy $E(k)$ is $\Psi_E(x,t) = (1/\sqrt{L})e^{ikx}e^{-iE(k)t/\hbar}$. In the QM expression for current, the time evolution portion is not affected by the momentum operator, and therefore factors to 1. It is another illustration of the virtues of working with states of definite energy. The current carried by state $|k\rangle$ is then obtained as $J(k) = I(k) = q\hbar k/m_e L$. The current density and current are the same in 1D. The current $I(k) = q\hbar k/m_e L = qv(k)/L$ connects to the classical notion of current carried by a particle with velocity $v(k) = \hbar k/m_e$ traversing a distance L. Another way to picture the same current is to split it as $I = q \times v(k) \times n$, where $n = 1/L$ is the 'volume density' of particles.

To find the total current carried by multiple k-states, we must sum the contribution from each state. We use the velocity picture since it carries over to wave packets and electrons in crystals. In the next section, we are going to prove the following very important result: that the velocity term that appears in the quantum mechanical expression for current for free electrons of wavefunction $\psi(x) = \frac{1}{\sqrt{L}}e^{ikx}$ as $v(k) = \frac{\hbar k}{m}$ simply changes to the 'group velocity' when the electron is put in a periodic crystal with Block eigenfunction $\psi(x) = \frac{1}{\sqrt{L}}e^{ikx}u(x)$. The group velocity is simply $\mathbf{v}_g(\mathbf{k}) = \nabla_{\mathbf{k}}E(\mathbf{k})/\hbar$, which is known when the electron bandstructure is known - *without recourse to the wavefunction.* So we trade the knowledge of wavefunction for the knowledge of the bandstructure $E(\mathbf{k})$, which is attractive because the bandstructure of semiconductors are experimentally measured and tabulated. We also generalize to any dimension d. The expression for the current then becomes

$$\mathbf{J}_d = \frac{q}{L^d}\sum_{\mathbf{k}} \mathbf{v}_g(\mathbf{k})f(\mathbf{k}), \tag{5.14}$$

where we have now included the Fermi-Dirac occupation probability of state $|\mathbf{k}\rangle$. We next split off the spin degeneracy $g_s = 2$, and allow a valley degeneracy g_v for each $\mathbf{k}$-state. Picture the current flow from a 'left' contact to a 'right' contact. Let us also allow for scattering, which are quantum-mechanical reflections in going from the left to the right contact. Let the quantum-mechanical transmission coefficient from the left to the right contact be $T(\mathbf{k})$. Then, the expression for the net current density flowing from the left to the right becomes

$$\mathbf{J}_d = \frac{qg_sg_v}{L^d}\sum_{\mathbf{k}} \mathbf{v}_g(\mathbf{k})T(\mathbf{k})[f_L(\mathbf{k}) - f_R(\mathbf{k})]. \tag{5.15}$$

The Fermi-Dirac function difference results because $f_L(\mathbf{k})[1-f_R(\mathbf{k})]-f_R(\mathbf{k})[1-f_L(\mathbf{k})] = f_L(\mathbf{k}) - f_R(\mathbf{k})$. The transmission coefficient $T(\mathbf{k})$ does not depend on the direction of current flow.

The sum over $\mathbf{k}$ states is typically always converted into an integral. The recipe for this step uses the fact that the allowed states in the $\mathbf{k}$-space are discrete 'boxes' of each side $2\pi/L$ and volume $(2\pi/L)^d$ in d dimensions. Then, the sum converts to an integral via

$$\sum_{\mathbf{k}}(...) = \int \frac{d^d\mathbf{k}}{(\frac{2\pi}{L})^d}(...). \tag{5.16}$$

Note the cancellation of the dependence on the macroscopic dimension L. The $\mathbf{k}$-coordinate system is chosen to be either cartesian, cylindrical, or spherical based on the specific problem. The expression for the current density is then

$$\mathbf{J}_d = \frac{qg_sg_v}{(2\pi)^d}\int d^d\mathbf{k} \times \mathbf{v}_g(\mathbf{k})T(\mathbf{k})[f_L(\mathbf{k}) - f_R(\mathbf{k})]. \tag{5.17}$$

The unit of current density is in A/m^{d-1} for current flow in d-dimensions. This expression for the current density is applicable in a wide range of situations ranging from ballistic transport, to scattering limited drift or diffusion, and to tunneling transport, and in multiple (1, 2, or 3) dimensions. The group velocity term $\mathbf{v}_g(\mathbf{k})$ is locked down by the electron band structure $E(\mathbf{k})$, and external forces modify the Fermi-Dirac functions. For example, the response of $f(\mathbf{k})$ to electric fields, or concentration gradients are tracked by a Boltzmann transport equation.

5.5 Charge current in semiconductor crystals

In this section, we prove our assertion that the velocity term that appears for Bloch states in the expression for the current in quantum mechanics is the *group velocity* $\mathbf{v}_g(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}E(\mathbf{k})$. The definite energy wavefunctions of electrons in crystals are Bloch functions $\Psi(\mathbf{k},\mathbf{r},t) = \psi_E(\mathbf{k},\mathbf{r})e^{-iE(\mathbf{k})t/\hbar}$, where $\psi_E(\mathbf{k},\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{k},\mathbf{r})$ with $\mathbf{k}$ the Bloch wavevector in the reduced zone and $u(\mathbf{k},\mathbf{r}+\mathbf{a}) = u(\mathbf{k},\mathbf{r})$. $E(\mathbf{k})$ are the eigenvalues of the Bloch states, and constitute the electron bandstructure. To obtain the group velocity, we first derive the following useful identity for Bloch states:

$$\frac{m_e}{\hbar}\nabla_{\mathbf{k}}E(\mathbf{k}) = \int d^3\mathbf{r}\Psi^\star\hat{\mathbf{p}}\Psi = -\int d^3\mathbf{r}\Psi\hat{\mathbf{p}}\Psi^\star \tag{5.18}$$

The identity is evidently dimensionally correct since both sides have units of momentum. The middle and far right are complex conjugate relations of each other, whereas the left side is real. Note that the left side involves a gradient of $E(\mathbf{k})$ in $\mathbf{k}-$space, whereas the middle and right sides involve gradients of $\Psi, \Psi^\star$ in the $\mathbf{r}-$space.

To arrive at this identity, write the Schrodinger equation for the Bloch states as

$$-\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}}^2\psi_E(\mathbf{k},\mathbf{r})e^{-iE(\mathbf{k})t/\hbar} = [E(\mathbf{k})-V]\psi_E(\mathbf{k},\mathbf{r})e^{-iE(\mathbf{k})t/\hbar}. \tag{5.19}$$

The time evolution cancels since Bloch states are stationary states. We now take a gradient in the $\mathbf{k}-$space. Since $[\nabla_{\mathbf{r}}, \nabla_{\mathbf{k}}] = 0$, and $\nabla_k\psi_E = i\mathbf{r}\psi_E + e^{i\mathbf{k}\cdot\mathbf{r}}\nabla_{\mathbf{k}}u = i\mathbf{r}\psi_E + \mathbf{X}$ we have

$$-\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}}^2(i\mathbf{r}\psi_E+\mathbf{X}) = [E(\mathbf{k})-V](i\mathbf{r}\psi_E+\mathbf{X}) + [\nabla_{\mathbf{k}}E(\mathbf{k})]\psi_E. \tag{5.20}$$

We now use the identity for Bloch eigenstates:

$$\nabla_{\mathbf{r}}^2(\mathbf{r}\psi_E) = \mathbf{r}\nabla_{\mathbf{r}}^2\psi_E + 2\nabla_{\mathbf{r}}\psi_E. \tag{5.21}$$

in Equation 5.20 and rearrange to obtain

$$\underbrace{-\frac{\hbar^2}{2m_e}(i\mathbf{r}\nabla_{\mathbf{r}}^2\psi_E)}_{1} - \frac{i\hbar^2}{m_e}\nabla_{\mathbf{r}}\psi_E = \underbrace{[E(\mathbf{k})-V](i\mathbf{r}\psi_E)}_{1} + [E(\mathbf{k}) - \underbrace{(-\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}}^2+V)}_{\hat{H}}]\mathbf{X} + [\nabla_{\mathbf{k}}E(\mathbf{k})]\psi_E. \tag{5.22}$$

The terms indicated as '1' constitute the Schrodinger equation for Bloch eigenstates multiplied by $i\mathbf{r}$, and hence cancel. Multiplying by $\psi_E^\star$ on the left and integrating over all space, the second term on the right vanishes since $\psi_E^\star$ is an eigenstate. Then, we are left with

$$\nabla_{\mathbf{k}}E(\mathbf{k})\int d^3\mathbf{r}\psi_E^\star\psi_E = \nabla_{\mathbf{k}}E(\mathbf{k}) = -\frac{i\hbar^2}{m_e}\int d^3\mathbf{r}(\psi_E^\star\nabla_{\mathbf{r}}\psi_E), \tag{5.23}$$

which proves the identity in Equation 5.18 since $\hat{\mathbf{p}} = -i\hbar\nabla_{\mathbf{r}}$. The probability current density carried by a Bloch state is given by

$$\mathbf{j} = \frac{1}{2m_e}(\Psi^\star\hat{\mathbf{p}}\Psi - \Psi\hat{\mathbf{p}}\Psi^\star), \tag{5.24}$$

from which we obtain the *group velocity* of the Bloch state with help of the identity in Equation 5.18 as

$$\mathbf{v}_g(\mathbf{k}) = \frac{1}{2m_e}\int d^3\mathbf{r}(\Psi^\star\hat{\mathbf{p}}\Psi - \Psi\hat{\mathbf{p}}\Psi^\star) = \frac{\nabla_{\mathbf{r}}E(\mathbf{k})}{\hbar}. \tag{5.25}$$

The simplicity of the result $\mathbf{v}_g(\mathbf{k}) = \nabla_{\mathbf{r}}E(\mathbf{k})/\hbar$ is one of the central reasons why the $\mathbf{k}-$space is where one should investigate transport properties. This velocity is used in Equation 5.17 for evaluating the charge current carried by Bloch states in d-dimensions:

$$\mathbf{J}_d = \frac{qg_sg_v}{\hbar(2\pi)^d}\int d^d\mathbf{k}[\nabla_{\mathbf{k}}E(\mathbf{k})]T(\mathbf{k})[f_L(\mathbf{k}) - f_R(\mathbf{k})]. \tag{5.26}$$

Note that the group velocity term is a first-order derivative in k. Therefore, for the particular case of d =1 dimension, the current simplifies to

$$I = \frac{qg_sg_v}{\hbar(2\pi)}\int dE\cdot T(E)[f_L(E) - f_R(E)] \approx (g_sg_v\frac{q^2}{h})V \tag{5.27}$$

if we assume $T(E) = 1$ for perfect (ballistic) transmission and since $\int dE[f_L(E) - f_R(E)] = qV$ at $T \to 0$K. The ballistic conductance of a 1D mode is therefore given by $G_0 = g_sg_v \times q^2/h$ and does not depend on the exact bandstructure $E(\mathbf{k})$.

5.6 Energy (heat) current

The flow of an electron of energy $E(\mathbf{k})$ transports not just charge, but the associated energy as well. The energy current density is given by $\mathbf{J}_E(\mathbf{k}) = \sum_{\mathbf{k}} E(\mathbf{k})\cdot\mathbf{v}_g(\mathbf{k})/L^d$, and thus we obtain

$$\mathbf{J}_E = \frac{g_sg_v}{(2\pi)^d}\int d^d\mathbf{k}E(\mathbf{k})\mathbf{v}_g(\mathbf{k})T(\mathbf{k})[f_L(\mathbf{k}) - f_R(\mathbf{k})]. \tag{5.28}$$

The units are in Watts/m^{d-1} for transport in d-dimensions. Using the same approximations as earlier for a 1D conductor, we obtain the energy current carried by electrons to be $I_E \approx (g_sg_vq^2/h)\cdot V^2$ in Watts if the transport is ballistic ($T(\mathbf{k}) = 1$) at $T \to 0$K.

5.7 Any current

Based on the above recipes, spin currents, polarization currents, and other related quantities may be evaluated. We realize from the results derived in this chapter that the key to finding the current are the occupation functions $f(\mathbf{k})$ and their dependence on external voltages. This indeed is the driver of current for all cases. Determination of the occupation functions, and their dependence on external voltages is the topic of the next chapter (Chapter 6).

Chapter 6

The Concept of Equilibrium: Fermi-Dirac and Bose-Einstein

6.1 Introduction

In this chapter, we derive and discuss the Fermi-Dirac distribution function for fermions, and the Bose-Einstein distribution function for bosons. These functions provide us the statistical occupation number of quantum states for a system in thermodynamic equilibrium with a reservoir. The Fermi-Dirac distribution is central to finding the electron distribution over allowed energy or momentum values in various semiconductor devices. The Bose-Einstein distribution is central to finding the distribution of photons in the electromagnetic field, or phonons in semiconductor crystals. The two distributions together determine electron-phonon and electron-photon interactions. The importance of this chapter simply cannot be overemphasized! We discuss various properties of the distributions and limiting cases to gain familiarity. Then, we specifically map the concept of thermodynamic equilibrium to the fundamental semiconductor building blocks, such as the ohmic contact, Schottky contacts, the p-n junction, and a field-effect transistor (FET).

6.2 The physics of equilibrium

We begin by drawing upon a fundamental result from quantum statistical mechanics[1]. The most well-known result of statistical thermodynamics is the Boltzmann distribution. The result states the following: consider a *system* that in *thermal equilibrium* with a

[1]For a detailed derivation, see Thermal Physics by Kittel and Kroemer.

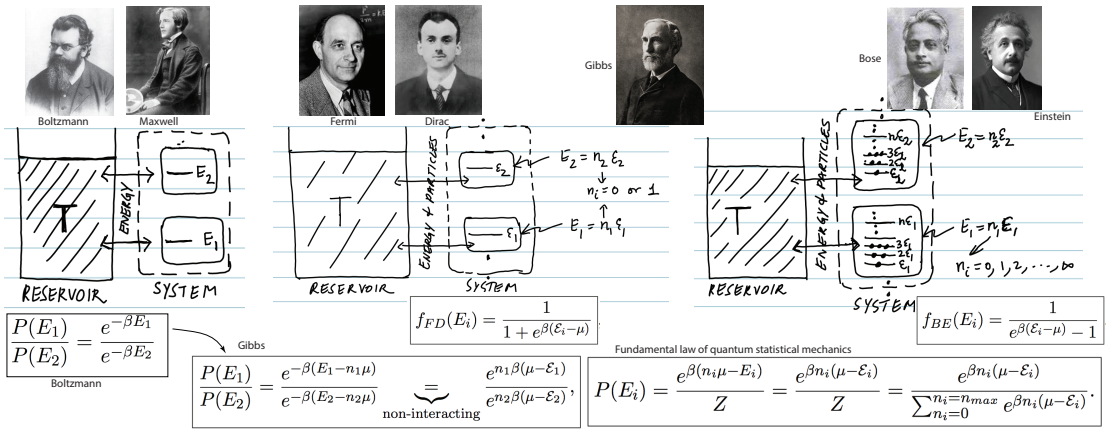

FIGURE 6.1: Illustration of the processes of thermodynamic equilibrium for the Boltzmann distribution, and the Gibbs partition function.

reservoir at temperature T. Each of the terms in italics have very specific meanings, which will be described shortly. Let E_1 and E_2 denote two energy states of the system. The Boltzmann result asserts that the probabilities of finding the system in these energies is related by

$$\boxed{\frac{P(E_1)}{P(E_2)} = \frac{e^{-\beta E_1}}{e^{-\beta E_2}},} \tag{6.1}$$

where $\beta = \frac{1}{kT}$, and k is the Boltzmann constant. Figure 6.1 illustrates the meanings of the terms in italics. The *reservoir* is a large source of particles and energy, characterized by a temperature T. It goes by the name *reservoir* because it can either take in, or give out any energy without changing its temperature T. As opposed to the reservoir, the *system* is much smaller, and can be found in energy states E_1, E_2, E_3, ... The statement that the system is in *thermal equilibrium* with the reservoir means that it can exchange energy with the reservoir, but *not particles*. Each energy state E_i is considered to be *individually* in thermal equilibrium with the reservoir. Only under this condition is the Boltzmann result in Equation 6.1 applicable. Since the temperature T is the measure of the energy which is being exchanged, the reservoir and the system share the same temperature upon reaching thermal equilibrium.

Now if we let the system exchange energy *and particles* with the reservoir, as indicated in Figure 6.1, the Boltzmann relation needs to be generalized. A measure of the particle number is the chemical potential μ, which must also appear in addition to the temperature T in relations characterizing thermodynamic equilibrium between the system and the reservoir. This famous generalization was done by Gibbs, who gave the modified relation

$$\boxed{\frac{P(E_1)}{P(E_2)} = \frac{e^{-\beta(E_1 - n_1\mu)}}{e^{-\beta(E_2 - n_2\mu)}} \underbrace{=}_{\text{non-interacting}} \frac{e^{n_1\beta(\mu - \mathcal{E}_1)}}{e^{n_2\beta(\mu - \mathcal{E}_2)}},} \tag{6.2}$$

where μ is a common chemical potential of the reservoir+system, and n_i is the number of particles in the single-particle energy state $\mathcal{E}_i$. We are going to call a single particle energy eigenstate an *orbital*, drawing from the language of chemistry. Only if the particles considered are *non-interacting*, then the energy of the state is $E_i = n_i\mathcal{E}_i$ if there are n_i particles in orbital $|i\rangle$ of eigenvalue $\mathcal{E}_i$. If these conditions are met, then one defines a Gibbs-sum, or more popularly known as the grand partition function

$$Z = \sum_{\text{states}} \sum_{\text{n}} e^{\beta(n\mu - E_n)}. \tag{6.3}$$

The sum runs over all states of the system, and all number of particles allowed in each single-particle state. Note carefully what this means. For example, consider the situation when orbital $|3\rangle$ is in equilibrium with the reservoir. Since it is not interacting with the other orbitals (which are also separately in equilibrium with the reservoir), the partition function for the 'system' consisting of a variable number of particles in $|3\rangle$ is then $Z = \sum_{n_3=0}^{n_3=n_{max}} e^{\beta n_3(\mu-\mathcal{E}_3)}$. The 'system' here is the various occupation states of orbital $|3\rangle$.

When energy *and* particle exchange is allowed between the system and the reservoir, the fundamental law of equilibrium statistical mechanics may be stated as the following. Under thermodynamic equilibrium with a reservoir at temperature T, the absolute probability that the system will be found in the state $E_i = n_i\mathcal{E}_i$ with n_i particles in orbital $|i\rangle$ is

$$\boxed{P(E_i) = \frac{e^{\beta(n_i\mu - E_i)}}{Z} = \frac{e^{\beta n_i(\mu-\mathcal{E}_i)}}{Z} = \frac{e^{\beta n_i(\mu-\mathcal{E}_i)}}{\sum_{n_i=0}^{n_i=n_{max}} e^{\beta n_i(\mu-\mathcal{E}_i)}}.} \tag{6.4}$$

For sake of completeness and for future use, we generalize this result. We recognize that the allowed orbital energies $\mathcal{E}_i$ are the eigenvalues of the single-particle Hamiltonian $\hat{H}_0$ via $\hat{H}_0|i\rangle = \mathcal{E}_i|i\rangle$, the non-interacting many-particle Hamiltonian $\hat{H} = \sum \hat{H}_0$ gives $\hat{H}|n_1, n_2, ...n_i, ...\rangle = (\sum_i n_i\mathcal{E}_i)|n_1, n_2, ...n_i, ...\rangle$, and the number n_i of particles in the eigenstate (or orbital) $|i\rangle$ is $\hat{N}_i|n_1, n_2, ...n_i, ...\rangle = n_i|n_1, n_2, ...n_i, ...\rangle$, where $\hat{N}_i$ is occupation number operator for eigenstate $|i\rangle$, and $\hat{N} = \sum_i \hat{N}_i$. Then, the expectation value of any operator $\langle\hat{O}\rangle$ at thermodynamic equilibrium is

$$\boxed{\langle\hat{O}\rangle = \frac{\text{Tr}[\hat{O}e^{\beta(\mu\hat{N}-\hat{H})}]}{\text{Tr}[e^{\beta(\mu\hat{N}-\hat{H})}]},} \tag{6.5}$$

where Tr[...] stands for the Trace of the matrix or the operator. Note that the Hamiltonian matrix and the number operator are exponentiated. The Trace gives the sum of the diagonal elements, making Equation 6.5 equivalent to 6.4 in the diagonal representation.

But since the Trace is invariant between representations, Equation 6.5 also holds for non-diagonal conditions. Feynman[2] calls the fundamental results in Equation 6.4 (and 6.5) the "summit of statistical mechanics, and the entire subject either a slide-down from the summit, or a climb up to this result". We have not covered the climb-up, but since we will apply the result, let us slide down by applying it to derive the Fermi-Dirac and the Bose-Einstein distribution functions. We will use the version of Equation 6.5 in later chapters, and focus on Equation 6.4 for this chapter.

6.2.1 Fermi-Dirac Distribution

As we have discussed in Chapter 2, the number of Fermionic particles that can occupy an energy eigenstate $\mathcal{E}_i$ are $n_i = 0$ or 1 and *nothing else* because of the Pauli exclusion principle. Therefore, the partition function for the state of the system corresponding to energy $\mathcal{E}_i$ in thermodynamic equilibrium (in the Gibbs sense) with a reservoir of temperature T and chemical potential μ is simply

$$Z = \sum_{n_i=0}^{n_i=1} e^{\beta n_i(\mu-\mathcal{E}_i)} = e^0 + e^{\beta(\mu-\mathcal{E}_i)} = 1 + e^{\beta(\mu-\mathcal{E}_i)}, \tag{6.6}$$

and the probability that the system is in a state that has n_i particles in orbital $|i\rangle$ is simply $P(E_i) = e^{\beta(n_i\mu - E_i)}/Z$, where $E_i = n_i\mathcal{E}_i$ is the total energy of the orbital. Note that we are assuming that the particles that fill the orbital do not interact with each other. Then, the thermal average number of particles $\langle n_i\rangle$ in orbital $|i\rangle$ is given by $f(\mathcal{E}_i) = \langle n_i\rangle = \sum_i n_i P(E_i)$, which is

$$\langle n_i\rangle = f(\mathcal{E}_i) = \frac{0\cdot e^0 + 1\cdot e^{\beta(1\cdot\mu - 1\cdot\mathcal{E}_i)}}{1 + e^{\beta(\mu-\mathcal{E}_i)}} \implies \boxed{f_{FD}(\mathcal{E}_i) = \frac{1}{1 + e^{\beta(\mathcal{E}_i-\mu)}}}, \tag{6.7}$$

where the boxed equation is the Fermi-Dirac distribution. Note that it varies between 0 and 1, and is equal to $\frac{1}{2}$ when $\mathcal{E}_i = \mu$. We will discuss this further shortly.

6.2.2 Bose-Einstein Distribution

Unlike Fermions, there is no restriction on the number of Bosonic particles that can occupy an orbital $|i\rangle$. This means $n_i = 0, 1, ..., \infty$. Then, the partition function is

$$Z = \sum_{n_i=0}^{\infty} e^{\beta n_i(\mu-\mathcal{E}_i)} = \sum_{n_i=0}^{\infty} [e^{\beta(\mu-\mathcal{E}_i)}]^{n_i} = \frac{1}{1 - e^{\beta(\mu-\mathcal{E}_i)}}, \tag{6.8}$$

[2] *Statistical Mechanics*, by R. P. Feynman.

where the infinite sum is a geometric series $1+u+u^2+... = \frac{1}{1-u}$, valid for $u = e^{\beta(\mu-\mathcal{E}_i)} < 1$, or equivalently $\mu \leq \mathcal{E}_i$. The thermal average number of bosonic particles in orbital $|i\rangle$ is then

$$\langle n_i \rangle = f(\mathcal{E}_i) = \frac{0 \cdot u^0 + 1 \cdot u^1 + 2 \cdot u^2 + 3 \cdot u^3 + ...}{(1-u)^{-1}} \implies \boxed{f_{BE}(\mathcal{E}_i) = \frac{1}{e^{\beta(\mathcal{E}_i - \mu)} - 1}}, \quad (6.9)$$

where the boxed equation is the Bose-Einstein distribution. In arriving at the result, we used the relation $u\frac{d}{du}(\frac{1}{1-u}) = \frac{u}{(1-u)^2} = u + 2u^2 + 3u^3 + ...$, which is the sum that appears in the numerator, whereupon $\langle n_i \rangle = \frac{1}{u^{-1}-1}$. Note that for $\beta(\mathcal{E}_i - \mu) >> 1$, the Bose-Einstein distribution $f_{BE}(E_i) \to 0$. However, for $\beta(\mathcal{E}_i - \mu) << 1$, $f_{BE}(E_i) \approx \frac{1}{\beta(\mathcal{E}_i - \mu)}$ can increase without bound, which is surprisingly physical and indicates a *condensation* of all particles to the lowest energy orbitals. This phenomenon is related to Bose-Einstein condensation, a topic to be discussed further later in the book.

6.2.3 Discussion of the nature of the distribution functions

The key ideas and results in arriving at the distribution functions are summarized in Figure 6.1. In Figure 6.2, we plot the various distribution functions.

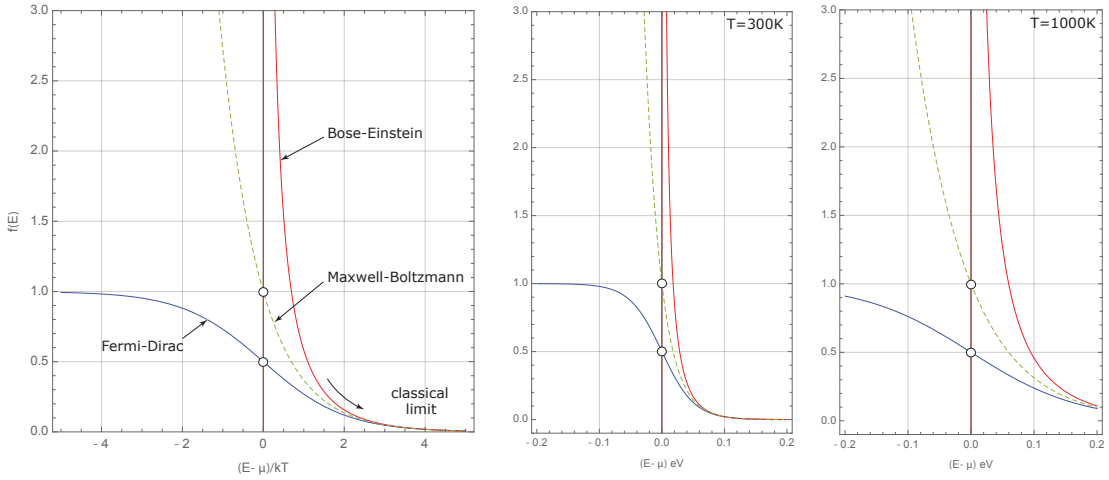

FIGURE 6.2: Illustration of the distribution functions and the effect of temperature.

We define the Fermi-Dirac *function* as

$$f_0(x) = \frac{1}{1 + e^{\beta x}} \quad (6.10)$$

which takes the argument $x = E - \mu$ to give us the Fermi-Dirac distribution

$$f_{FD}(E) = f_0(E - \mu) = \frac{1}{1 + e^{\beta(E-\mu)}}. \quad (6.11)$$

The distribution may be thought of a function of the energy E, or of the chemical potential μ. We use the compact notation $f_0 = f_0(E-\mu) = f_{FD}(E)$. The partial derivative with respect to energy is

$$\frac{\partial f_0}{\partial E} = -\frac{\partial f_0}{\partial \mu} = -\beta \cdot \frac{e^{\beta(E-\mu)}}{(1+e^{\beta(E-\mu)})^2} = -\beta \cdot f_0[1-f_0], \tag{6.12}$$

which can be rearranged to the form

$$\boxed{-\frac{\partial f_0}{\partial E} = +\frac{\partial f_0}{\partial \mu} = \frac{\beta}{4\cosh^2(\frac{\beta(E-\mu)}{2})}.} \tag{6.13}$$

The derivative of the Fermi-Dirac distribution evidently reaches its maximum value of $\frac{\beta}{4} = \frac{1}{4kT}$ at $E = \mu$. We have the identity $\int_{-\infty}^{+\infty} du \frac{\beta}{4\cosh^2[\frac{1}{2}\beta u]} = 1$, which indicates that in the limit of very low temperatures $\frac{1}{kT} = \beta \to \infty$, the derivative function should approach a Dirac-delta function in the energy argument, i.e.,

$$\boxed{\lim_{T\to 0}[-\frac{\partial f_0}{\partial E}] = \lim_{T\to 0}[+\frac{\partial f_0}{\partial \mu}] = \delta(E-\mu).} \tag{6.14}$$

This feature is illustrated in Figure 6.3.

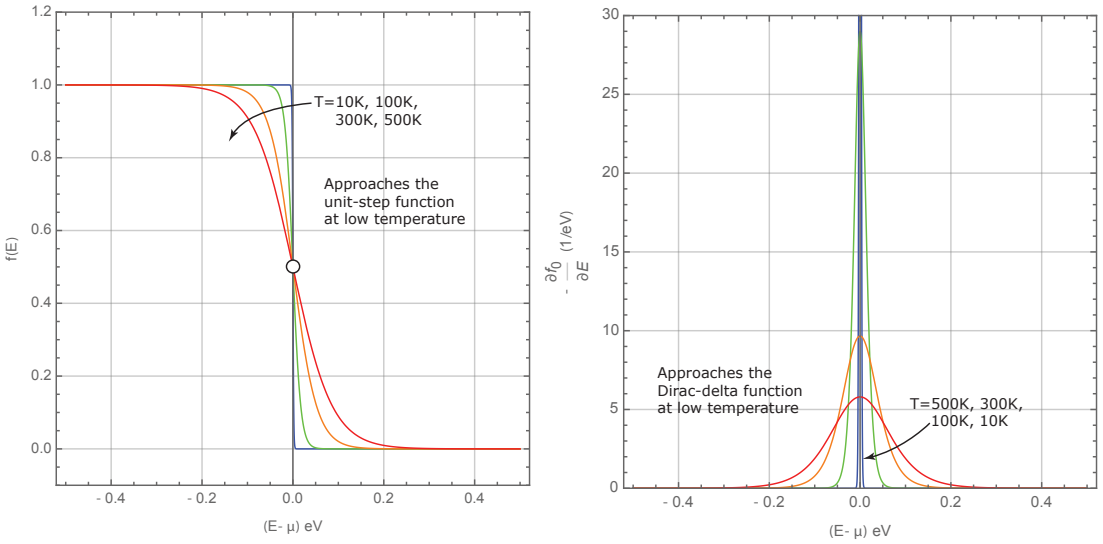

FIGURE 6.3: Illustration of the temperature dependence of the Fermi-Dirac distribution, and its derivative.

Now considering $f(u) = 1/(1+e^u)$ and $f(v) = 1/(1+e^v)$, we get the identity

$$\boxed{f(u) - f(v) = \underbrace{[f(u) + f(v) - 2f(u)f(v)]}_{\geq 0} \times \tanh(\frac{v-u}{2})} \tag{6.15}$$

Since $f(u), f(v) \leq 1$, the term in the square brackets is always positive. So the sign of the Fermi difference function is determined by the tanh(...) term. The Fermi difference function will make its appearance repeatedly when we study the optical and electronic transport properties of semiconductors and electronic and photonic devices.

The integral of the Fermi-Dirac function is

$$\boxed{\int_0^\infty dE f_0(E-\mu) = \int_0^\infty \frac{dE}{1+e^{\beta(E-\mu)}} = \frac{1}{\beta}\ln(1+e^{\beta\mu}),} \tag{6.16}$$

which leads to the very useful Fermi *difference* integral

$$\boxed{\int_0^\infty dE[f_0(E-\mu_1) - f_0(E-\mu_2)] = \frac{1}{\beta}\ln[\frac{1+e^{\beta\mu_1}}{1+e^{\beta\mu_2}}] = (\mu_1-\mu_2) + \frac{1}{\beta}\ln[\frac{1+e^{-\beta\mu_1}}{1+e^{-\beta\mu_2}}].} \tag{6.17}$$

If $\mu_1, \mu_2 >> kT$, the second term on the rightmost side is zero, and we obtain

$$\boxed{\int_0^\infty dE[f_0(\mu_1) - f_0(\mu_2)] \approx (\mu_1-\mu_2).} \tag{6.18}$$

That this relation is an identity is evident at $T \to 0$, or $\beta \to \infty$. The features of the Fermi difference function are illustrated in Figure 6.4. The integral at low temperatures is just the area under the dashed difference curve, which is rectangular and has a energy width of $\mu_2 - \mu_1$.

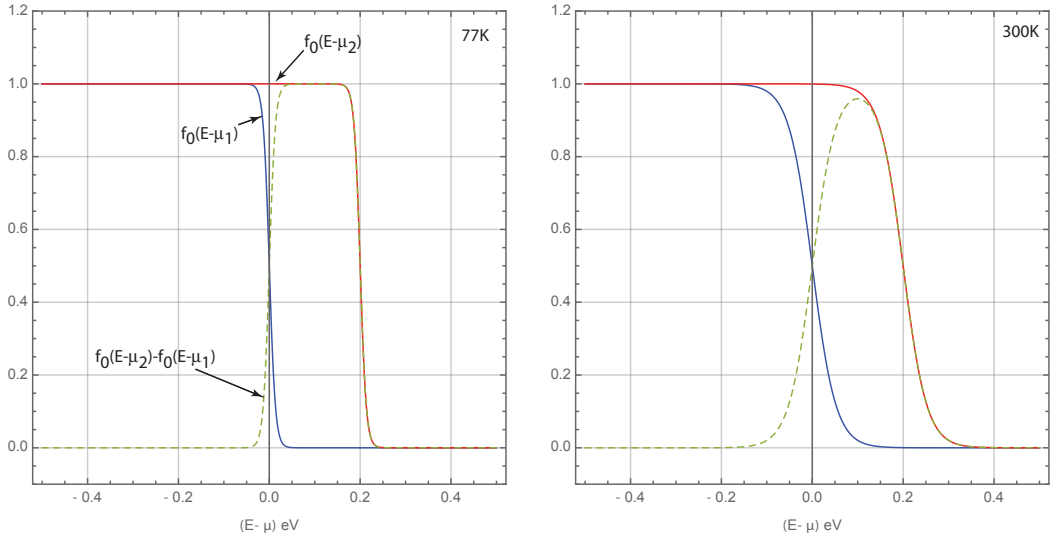

FIGURE 6.4: Illustration of the temperature dependence of the Fermi-difference distribution. The difference is a window between $\mu_2 - \mu_1$ that becomes increasingly rectangular as the temperature drops.

It is useful to define higher moment integrals of the Fermi-Dirac functions of the form

$$F_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^\infty du \frac{u^j}{1+e^{u-\eta}}. \tag{6.19}$$

The Fermi-Dirac integral is rendered dimensionless by scaling the chemical potential

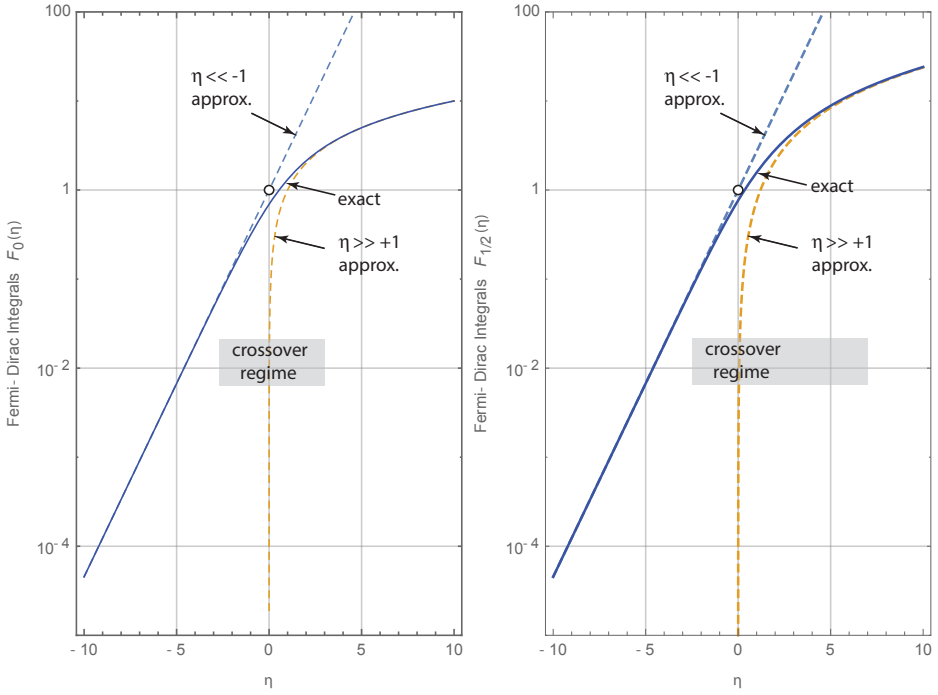

FIGURE 6.5: Fermi-Dirac integrals and their non-degenerate ($\eta << -1$) and degenerate ($\eta >> 1$) approximations, illustrating Equation 6.20.

$\eta = \beta\mu$, and the energy $u = \beta E$ by the thermal energy $kT = \frac{1}{\beta}$. Since we are integrating over u, the Fermi-Dirac integral $F_j(\eta)$ is a function of the chemical potential μ. The denominator is a normalizing Gamma function $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$ with the property $\Gamma(n+1) = n\Gamma(n)$, which means if n is an integer, $\Gamma(n) = (n-1)!$. A useful value of the Gamma function for a non-integer argument is $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. For $\eta << -1$, the exponential in the denominator is much larger than unity. An excellent approximation of the Fermi-Dirac integral then is $F_j(\eta) \approx e^\eta$, irrespective of the value of j. In the other extreme, when $\eta >> 1$, an excellent approximation is $F_j(\eta) \approx \frac{\eta^{j+1}}{\Gamma(j+2)}$. Due to the high importance of Fermi-Dirac integrals in semiconductor devices, we collect the results:

$$\boxed{F_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^\infty du \frac{u^j}{1+e^{u-\eta}}}, \quad \boxed{F_j(\eta) \underbrace{\approx}_{\eta<<-1} e^\eta}, \quad \boxed{F_j(\eta) \underbrace{\approx}_{\eta>>1} \frac{\eta^{j+1}}{\Gamma(j+2)}}. \tag{6.20}$$

From Equation 6.16, we have an exact analytical result for the Fermi-Dirac integral for $j = 0$: it is $F_0(\eta) = \ln(1+e^\eta)$. The validity of the approximations in Equation

6.20 are easily verified for this special case. No exact analytical expressions for other orders ($j \neq 0$) exist. The approximations in Equation 6.20 then assume increased importance for analytical evaluation of various physical quantities such as the mobile carrier densities in semiconductor bands, transport phenomena, and optical properties. The order j depends on the dimensionality of the problem. Figure 6.5 illustrates the cases of the Fermi-Dirac integrals and their approximations for the cases of $j = 0$ and $j = \frac{1}{2}$.

6.3 Meaning of equilibrium in semiconductor devices

Let us now consider a few semiconductor devices to develop a deeper understanding of the meaning of equilibrium in semiconductor devices. The first and simplest example is a 1D semiconductor (for example a carbon nanotube or a thin nanowire), which has *ohmic contacts* to two metal electrodes. The allowed energy eigenvalues in the semiconductor channel are those in the valence and conduction bands with band edge energies E_v, E_c, separated by a bandgap E_g, as indicated in Figure 6.6. Consider the 1D semiconductor to be doped n-type, with mobile electrons in the conduction band, and no mobile carriers in the valence band. Then the true meaning of an *ohmic contact* is the following: the electrons in the conduction band of the semiconductor are in thermodynamic equilibrium with the electrons in the metal contacts, in the Gibbs-sense. The conduction band states (or orbitals) in the semiconductor can freely exchange particles (electrons) and energy with the states or orbitals in the contacts, which is the reservoir. Connect this concept of Gibbs equilibrium in Figure 6.6 with the picture we used earlier in Figure 6.1. Note here we have *two reservoirs*. The particles in the left contact (reservoir) are in equilibrium with each other, and those in the right contact are in equilibrium with each other. When no external voltage is applied across them, the contacts are also in thermodynamic equilibrium with each other.

Inside the semiconductor connecting the contacts, there are particles that are moving to the right, and those moving to the left. Let us consider the situation where the left- and right-going carriers *do not mix*, i.e., there is no scattering of carriers. This is referred to as the *ballistic* case, and is approximately realized for very short semiconductor lengths. Consider the electrons moving to the right in the semiconductor. These electrons can only enter the semiconductor from the left contact. Then the electrons moving to the right are in thermodynamic equilibrium with the *left contact*. Similarly, carriers moving to the left in the semiconductor are in equilibrium with the *right contact*. Being in thermodynamic equilibrium in the Gibbs sense means the right-moving electron states share the same chemical potential μ and temperature T as the electrons in the left contact

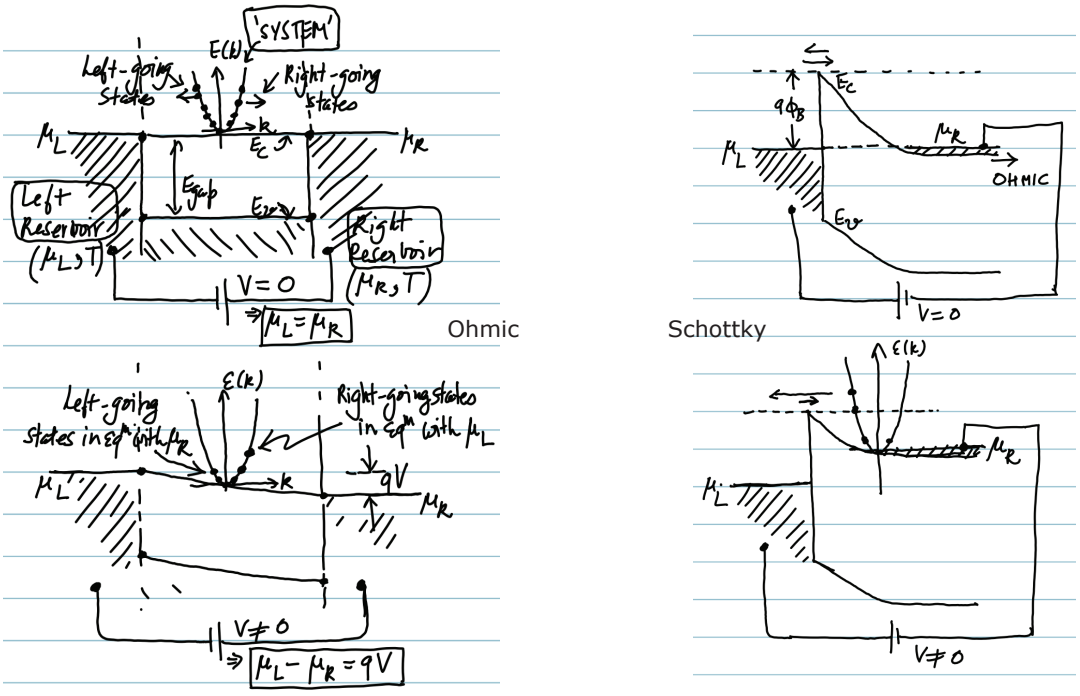

FIGURE 6.6: Illustration of the concept of equilibrium for Ohmic and Schottky contacts between metals and semiconductors.

metal. This is an extremely important consequence of thermodynamic equilibrium. Similarly the carriers in the semiconductor moving to the left could only have entered from the right contact, which keeps them in equilibrium with that contact and share μ and T. As long as the chemical potentials of the contacts are the same, the net current flow due the left and right moving carriers in the semiconductor exactly cancel, because they share the same μ.

When a voltage is applied between the contacts, the chemical potential of one contact is $\mu_L - \mu_R = qV$ larger than the other. This in turn breaks the delicate balance of left-and right-moving carriers inside the semiconductor. The imbalance of the left and right moving carriers as indicated in Figure 6.6 thus is the driver of an electric current through the semiconductor, completing the circuit. We will use this picture to calculate the current through a ballistic semiconductor channel in Chapter 7, and show that the conductance is quantized.

If the chemical potential potential of the metal lines up with energies in the bandgap of the semiconductor, a Schottky contact results, as indicated in Figure 6.6. The figure shows again a semiconductor in contact with two metals: the left contact is now Schottky, and the right contact is ohmic to the conduction band electrons in the semiconductor. Going back to our discussion of equilibrium in Section 6.1, we realize that the left-moving electrons in the semiconductor are in thermodynamic equilibrium with the right contact. But the right moving electrons in the semiconductor are *not* in thermodynamic

equilibrium with the left contact in the Gibbs sense, because there is a barrier between them that prevents free particle exchange. When a voltage is applied across the two metal contacts, the stronger imbalance of equilibrium between the left-and right-going carriers in the semiconductor cause a high asymmetry in the current flow as a function of the voltage. For the 'forward' bias condition shown, the left-moving carriers in the semiconductor that make it over the barrier to the metal are in equilibrium with the right contact. Since their chemical potential changes linearly with voltage, their concentration increases exponentially with voltage, causing a characteristic exponential turn-on of the diode. We will discuss this quantitatively in later chapters.

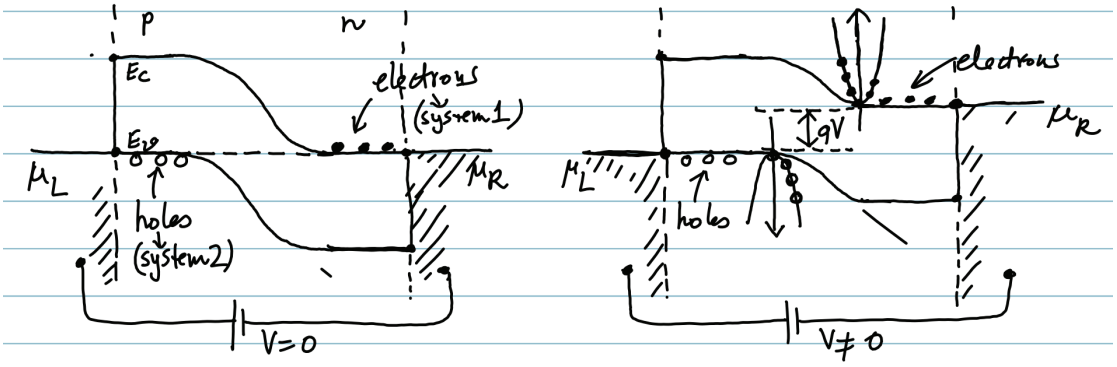

FIGURE 6.7: Illustration of the concept of equilibrium for p-n junctions.

Figure 6.7 shows a semiconductor p-n junction. Note the reservoirs are metals, but clearly we have chosen two *different* metals to form ohmic contacts individually to the p-side and the n-side of the semiconductor. An ohmic contact between a semiconductor and one metal electrode is possible for carriers in only one of the semiconductor bands, not both. This means with the proper choice of metals, we can form an ohmic contact to the conduction band of a n-type semiconductor for a n-type ohmic contact, and to the valence band of a p-type semiconductor for a p-type ohmic contact separately. So the holes in the valence band of the p-type semiconductor layer are in thermodynamic equilibrium with the p-ohmic metal (left), and the electrons in the n-type semiconductor layer are in thermodynamic equilibrium with the n-contact metal. Note now we also have two types of carriers - electrons in the conduction band, and holes in the valence band. When no voltage is applied, the holes in the p-side are in thermodynamic equilibrium with the electrons in the n-side - because they are in turn in equilibrium with their respective ohmic contact metal reservoirs. So they share a common chemical potential. However, when a voltage is applied, as indicated in Figure 6.7, the equilibrium is broken; the chemical potentials of the conduction band electrons in the n-side and valence band holes in the p-type now differ by $\mu_n - \mu_p = qV$. This again is responsible for current flow, as will be discussed in later chapters.

As a final example, consider a 3-terminal device, the field-effect transistor (FET) shown in Figure 6.8. The carriers in the inversion channel have ohmic contacts to the source and drain contacts, but a Schottky-type contact through an additional insulating barrier layer to the gate metal reservoir. So the carriers in the semiconductor channgel can be in thermal equilibrium with the carriers in the gate metal, but not in thermodynamic equilibrium in the Gibbs sense because the exchange of particles between the gate metal and the semiconductor channel is explicitly prohibited. The right-going carriers are again injected from the left contact, and the left-going carriers are injected from the right contact. But the carrier density in the semiconductor is controlled by the gate voltage capacitively. We will use this picture in Chapter 10 to discuss the FET in detail. The FET is the most commonly used semiconductor device today.

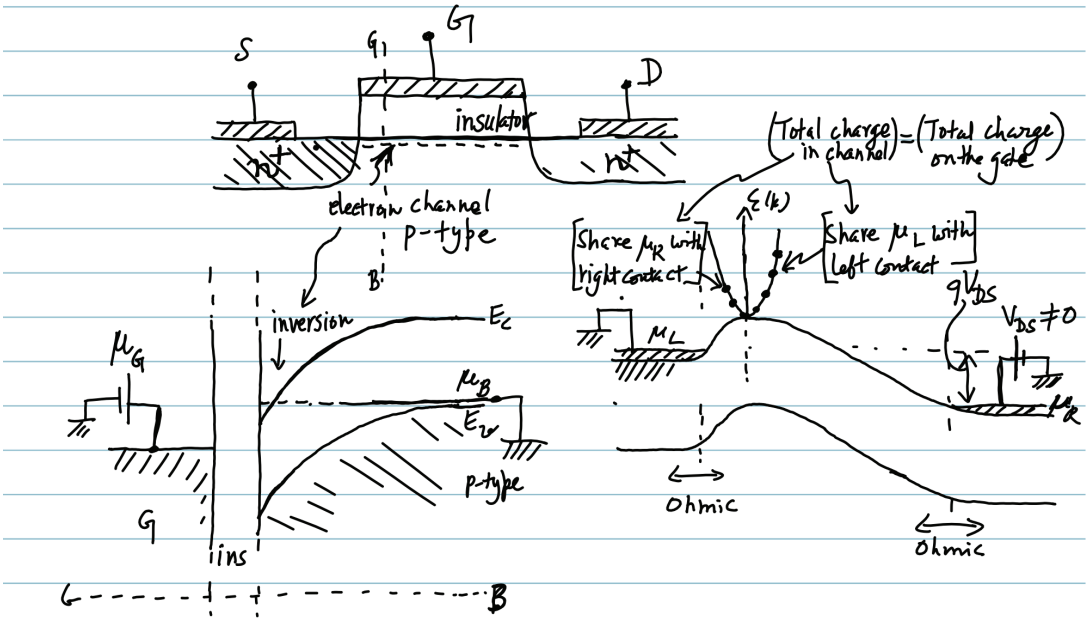

FIGURE 6.8: Illustration of the concept of equilibrium for a 3-terminal MOSFET device.

Chapter 10

Application II: The Ballistic Field-Effect Transistor

10.1 Introduction

In this chapter, we apply the formalism we have developed for charge currents to understand the output characteristics of a field-effect transistor. Specifically, we consider the situation when transport of electrons in the transistor occurs without scattering due to defects, i.e., ballistically from the source contact to the drain. The ballistic characteristics highlight various quantum limits of performance of a transistor. They guide material and geometry choices to extract the most of such devices. In this process we develop powerful insights into the inner workings of the remarkable device that powers the digital world.

10.2 The field-effect transistor

Figure 10.1 illustrates a typical field-effect transistor. A 2-dimensional electron gas (2DEG) at the surface of a semiconductor (or in a quantum well) is the conducting channel. It is separated from a gate metal by a barrier of thickness t_b and dielectric constant ϵ_b. The gate metal electrostatically controls the 2DEG density via the capacitance $C_b = \epsilon_b/t_b$. The source and the drain metals form low-resistance ohmic contacts to heavily doped regions indicated in gray. The FET width in the y-direction is W, which is much larger than the source-drain separation L and the barrier thickness t_b.

The 2DEG density at different points x of the channel from the source to the drain depends on the relative strength of the electrostatic control of the three contacts. We

assume that the source contact is grounded. V_{ds} is the drain potential and V_{gs} is the gate potential with respect to the source. When $V_{ds} = 0$ V, the 2DEG forms the lower plate of a parallel-plate capacitor with the gate metal. A threshold voltage V_T is necessary on the gate to create the 2DEG. Once created, the 2D charge density n_s in the 2DEG changes as $qn_s \approx C_g(V_{gs} - V_T)$, where $C_g = C_b C_q/(C_b + C_q)$, where C_q is a density-of-states or 'quantum' capacitance. Note that $qn_s \approx C_g(V_{gs} - V_T)$ is true only in the 'on-state' of the transistor, and will not give us the sub-threshold or off-state characteristics. The quantum capacitance arises because the density of states of the semiconductor band is lower than the metal: this forces a finite voltage drop in the semiconductor to hold charge. It may also be pictured as a finite spread of the 2DEG electrons, whose centroid is located away from the surface, adding an extra capacitance in series to the barrier capacitance. We will use the zero-temperature limit of $C_q \approx q^2 \times \rho_{2d}$ for our purposes here, where $\rho_{2d} = g_s g_v m^\star/2\pi\hbar^2$ is the DOS for each subband of the 2DEG. Since $V_{ds} = 0$ V, no *net* current flows from the source to the drain. However, when the 2DEG is present, the electrons are carrying current. The microscopic picture is best understood in the $\mathbf{k}-$space.

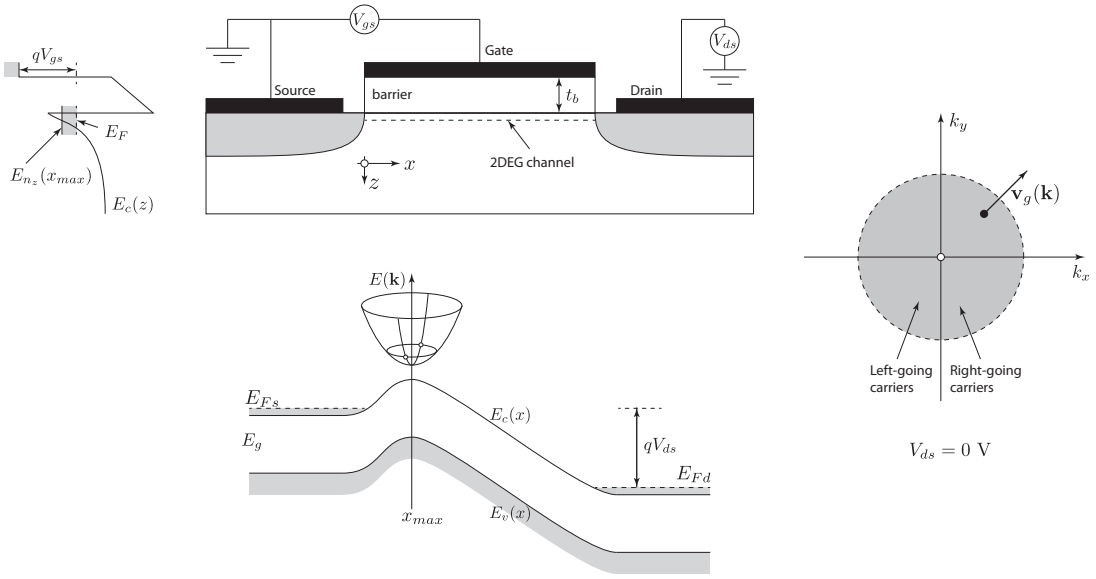

FIGURE 10.1: Field effect transistor, energy band diagram, and $\mathbf{k}-$space occupation of states.

The states of the first subband of the 2DEG are illustrated in the real-space energy band diagram and the occupation picture in $\mathbf{k}-$space in Figure 10.1. When $V_{gs} > V_T$, a quantum-well is created with the $z-$quantization resulting in a ground state energy E_{n_z}. The total energy of electrons in this 2DEG subband is given by

$$E(k_x, k_y) = E_c + E_{n_z} + \frac{\hbar^2(k_x^2 + k_y^2)}{2m^\star}, \tag{10.1}$$

where E_c is the conduction band edge energy at the interface, and $m^\star$ is the effective mass of the sub-bandstructure. We choose $E_c = 0$, and $m^\star$ to be isotropic. When $V_{ds} = 0$ V, the 2DEG electrons are in equilibrium with the source and drain. So the Fermi-level of the 2DEG electrons E_F is the same as the source and the drain. The band edge E_c and quantization energy E_{n_z} have to adjust to populate the channel with the charge dictated by the gate capacitor $qn_s = C_g(V_{gs} - V_T)$. The Fermi-Dirac distribution dictates the carrier distribution of the 2DEG in the $\mathbf{k}-$space. It is given by

$$f(k_x, k_y) = \frac{1}{1 + \exp\left[(\frac{\hbar^2}{2m^\star}(k_x^2 + k_y^2) - (E_F - E_{n_z}))/kT\right]} = \frac{1}{1 + \exp\left[\frac{\hbar^2(k_x^2+k_y^2)}{2m^\star kT} - \eta\right]}, \tag{10.2}$$

where we define $\eta = (E_F - E_{n_z})/kT$. Since the Fermi-level is controlled by the gate alone when $V_{ds} = 0$, we should be able to write η as a function of the gate voltage V_{gs}. The relation comes about by summing all occupied states in the $\mathbf{k}-$space:

$$C_g(V_{gs}-V_T) = q\underbrace{\frac{g_s g_v}{LW}\int \frac{dk_x}{\frac{2\pi}{L}}\frac{dk_y}{\frac{2\pi}{W}}\frac{1}{1+\exp\left[\frac{\hbar^2(k_x^2+k_y^2)}{2m^\star kT} - \eta\right]}}_{n_s} = q\frac{g_s g_v}{(2\pi)^2}\int_0^\infty\int_0^{2\pi}\frac{kdkd\theta}{1+\exp\left[\frac{\hbar^2 k^2}{2m^\star kT} - \eta\right]}. \tag{10.3}$$

We made the substitution $k_x = k\cos\theta$ and $k_y = k\sin\theta$. Pictorially, we are summing the states, or finding the 'area' of occupied states in the $\mathbf{k}-$space in Figure 10.1. At zero temperature, the shape is a circle with a sharp edge indicated by the dashed circle. At higher temperatures, the edge is diffuse, and the occupation probability drops exponentially as it is crossed. The spin-degeneracy of each state is g_s, and the semiconductor has g_v equivalent valleys, each with the same bandstructure.

The integral in Equation 10.3 is evaluated by first integrating out over θ which gives a factor 2π, and then making the substitution $u = \hbar^2 k^2/2m^\star kT$. Doing so with $V_{th} = kT/q$ yields

$$C_g(V_{gs} - V_T) = q\frac{g_s g_v m^\star kT}{2\pi\hbar^2}\underbrace{\int_0^\infty \frac{du}{1+\exp\left[u-\eta\right]}}_{F_0(\eta)} = C_q V_{th} F_0(\eta), \tag{10.4}$$

where we identify $C_q \approx q^2\rho_{2d} = q^2 g_s g_v m^\star/2\pi\hbar^2$ as the quantum capacitance, and the integral $F_0(\eta)$ as a special case of generalized Fermi-Dirac integrals of the form

$$F_j(\eta) = \int_0^\infty du \frac{u^j}{1+\exp\left[u-\eta\right]}, \tag{10.5}$$

with $j = 0$. The zeroth order Fermi-Dirac integral evaluates exactly to $F_0(\eta) = \ln[1 + \exp(\eta)]$. At this stage, it is useful to define $\eta_g = \frac{C_b}{C_b+C_q}(\frac{V_{gs}-V_T}{V_{th}})$. Thus the gate voltage V_{gs} tunes the Fermi level E_F of the 2DEG according to the relation

$$\eta = \frac{E_F - E_{n_z}}{kT} = \ln\left(e^{\eta_g} - 1\right). \tag{10.6}$$

For $V_{gs} - V_T >> V_{th}$, $\eta_g >> 1$, and we obtain $\eta \approx \eta_g$, implying $E_F - E_{n_z} \approx q(V_{gs} - V_T) \times C_b/(C_b + C_q)$. In other words, at a high gate overdrive voltage, the Fermi level changes approximately linearly with the gate voltage, as one would expect in a parallel plate capacitor. The capacitance factor is less than one, indicating a voltage division between the barrier and the channel. A part of the voltage must be spent to create the 2DEG since the density of states of the semiconductor conduction band is much smaller than a metal, as is apparent from the energy band diagram along the $z-$direction in Figure 10.1.

If we are interested in evaluating the sub-threshold characteristics of the ballistic FET, Equation 10.4 must be modified. It is evident that the RHS of this equation is always +ve, but when $V_{gs} < V_T$ in the sub-threshold, the LHS is -ve. To fix this problem, by looking at the energy band diagram in Figure 10.1 we rewrite the division of voltage drops as $qV_b + (E_F - E_{n_z}) = q(V_{gs} - V_T)$, where V_T now absorbs the surface barrier height, the conduction band offset between the barrier and the semiconductor, and the ground state quantization energy $(E_{n_z} - E_c)$. The term V_b is the voltage drop in the barrier given by $V_b = F_b t_b = (qn_s/\epsilon_b)t_b = qn_s/C_b$. The resulting relation between n_s and V_{gs} is then

$$\frac{q^2 n_s}{C_b} + kT\ln\left(e^{\frac{qn_s}{C_q V_{th}}} - 1\right) = q(V_{gs} - V_T) \implies \boxed{e^{\frac{qn_s}{C_b V_{th}}}\left(e^{\frac{qn_s}{C_q V_{th}}} - 1\right) = e^{\frac{V_{gs}-V_T}{V_{th}}}} \tag{10.7}$$

This is a transcendental equation, which must be numerically solved to obtain n_s as a function of V_{gs} to get the functional dependence $n_s(V_{gs})$. Note that since $n_s > 0$, both sides of the equation *always* remain +ve. As $V_{gs} - V_T$ becomes large and negative, $n_s \to 0$ exponentially but never reaches 0. This is the sub threshold characteristics of

the ballistic transistor. In Equation 10.7, two characteristic carrier densities appear: $n_b = C_b V_{th}/q$ and $n_q = C_q V_{th}/q$; the equation then reads $e^{\frac{n_s}{n_b}}(e^{\frac{n_s}{n_q}} - 1) = e^{\frac{V_{gs}-V_T}{V_{th}}}$. For $V_{gs} - V_T >> V_{th}$, the 1 in the bracket may be neglected, and $qn_s \approx \frac{C_b C_q}{C_b+C_q}(V_{gs} - V_T)$. On the other hand, when $V_{gs} - V_T << 0$, the RHS is small. Since $n_s > 0$, it must become very small. Expanding the exponentials and retaining the leading order, we obtain $n_s \approx n_q e^{\frac{V_{gs}-V_T}{V_{th}}}$. In the sub threshold regime, the carrier density at the source-injection point decreases *exponentially* with the gate voltage, and is responsible for the sharp switching of the device. Figure 10.2 illustrates this behavior. For the rest of the chapter, we focus on the on-state of the ballistic FET.

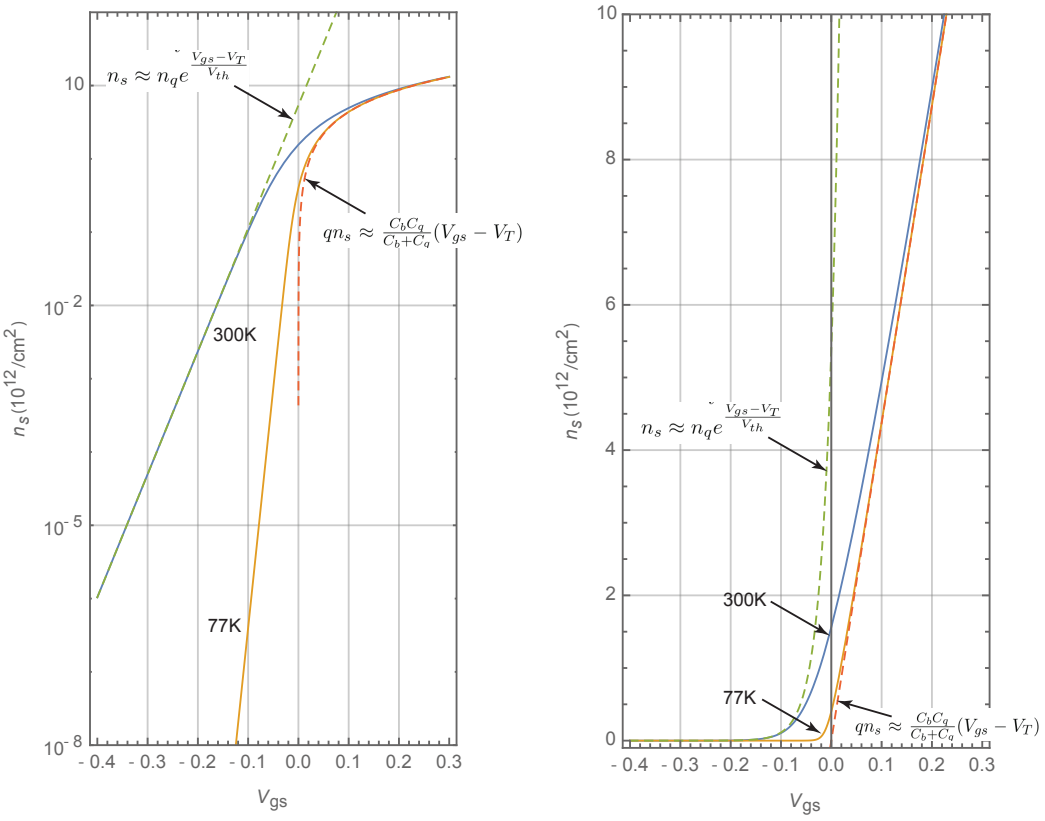

FIGURE 10.2: Illustrating the dependence of the 2DEG sheet density at the injection point on the gate voltage.

At this stage, it is instructive to find the right-going and left-going components of the net current at $V_{ds} = 0$ V, even though the net current is zero. We derived the general quantum-mechanical expression for current flowing in $d-$dimensions earlier as

$$\mathbf{J}_d = \frac{q g_s g_v}{(2\pi)^d} \int d^d\mathbf{k} \times \mathbf{v}_g(\mathbf{k}) f(\mathbf{k}), \tag{10.8}$$

where we assumed the transmission probability $T(\mathbf{k}) = 1$. For the 2DEG here, $d = 2$ and the group velocity of state $|\mathbf{k}\rangle$ is $\mathbf{v}_g(\mathbf{k}) = \hbar\mathbf{k}/m^\star$. From Figure 10.1, this velocity component points radially outwards from the origin in $\mathbf{k}-$space. Clearly evaluating

this integral will yield zero since there is a $|-\mathbf{k}\rangle$ state corresponding to each $|+\mathbf{k}\rangle$ state. So instead, we evaluate the current carried by electrons moving *only* in the $+k_x = |\mathbf{k}|\cos\theta = k\cos\theta$ direction. This is obtained from Eq. 10.8 by restricting the $\mathbf{k}-$space integral to the right half plane covered by $-\pi/2 \leq \theta \leq +\pi/2$ and using the velocity projected along the k_x axis $v_g = \hbar k\cos\theta/m^\star$ to obtain

$$J_{2d}^{\rightarrow} = \frac{qg_sg_v\hbar}{(2\pi)^2 m^\star}\int_{k=0}^{\infty}\int_{\theta=-\frac{\pi}{2}}^{+\frac{\pi}{2}}\frac{(k\cos\theta)kdkd\theta}{1+\exp\left[\frac{\hbar^2k^2}{2m^\star kT}-\eta\right]} = \underbrace{\frac{qg_sg_v\sqrt{2m^\star}(kT)^{\frac{3}{2}}}{2\pi^2\hbar^2}}_{J_0^{2d}}F_{\frac{1}{2}}(\eta), \qquad (10.9)$$

where $F_{1/2}(\eta)$ is the dimensionless Fermi-Dirac integral of order $j = 1/2$, and the prefactor J_0^{2d} has units of A/m or current per unit width. Since $J_{2d}^{\rightarrow} = J_{2d}^{\leftarrow} = J_0^{2d}F_{1/2}(\eta)$, the net current is zero. Another way to visualize this is to think of the right-going carriers as being created by injection into the 2DEG channel from the source, and thus the right-half carriers in $\mathbf{k}-$space are in equilibrium with the source. This statement is quantified by requiring $E_F^{\rightarrow} = E_{Fs}$. Similarly, the left-going carriers are injected from the drain contact, and are consequently in equilibrium with the drain $E_F^{\leftarrow} = E_{Fd}$. Since the source and the drain are at the same potential $E_{Fs} - E_{Fd} = qV_{ds} = 0$ V, the right going and left going carriers share a common Fermi level. Notice that we have defined two *quasi*-Fermi levels $E_F^{\rightarrow}$ and $E_F^{\leftarrow}$ and have thus split the carrier distribution into two types that can be in equilibrium amongst themselves, but out of equilibrium with each other. The current is zero at $V_{ds} = 0$ V due to the delicate balance between the left- and right-going current that exactly cancel each other.

This delicate balance is broken when a drain voltage is applied to the transistor.

10.3 Ballistic current-voltage characteristics

When a voltage V_{ds} is applied on the drain, the energy band diagram looks as indicated in Figure 10.1. Now the band edge $E_c(x)$ varies along the channel, with a maximum in the $x-y$ plane occurring at $x = x_{max}$, which is referred to as the 'top-of-the-barrier' (TOB) plane. The ground state of the quantum well $E_{n_z}(x)$ also varies along x depending upon the local vertical electric field, but has the fixed value $E_{n_z}(x_{max})$ at the TOB plane. Interestingly, there is no $x-$oriented electric field at x_{max}. The energy band diagram along the $z-$direction in the TOB plane is also indicated in Figure 10.1. Let's focus on this plane exclusively.

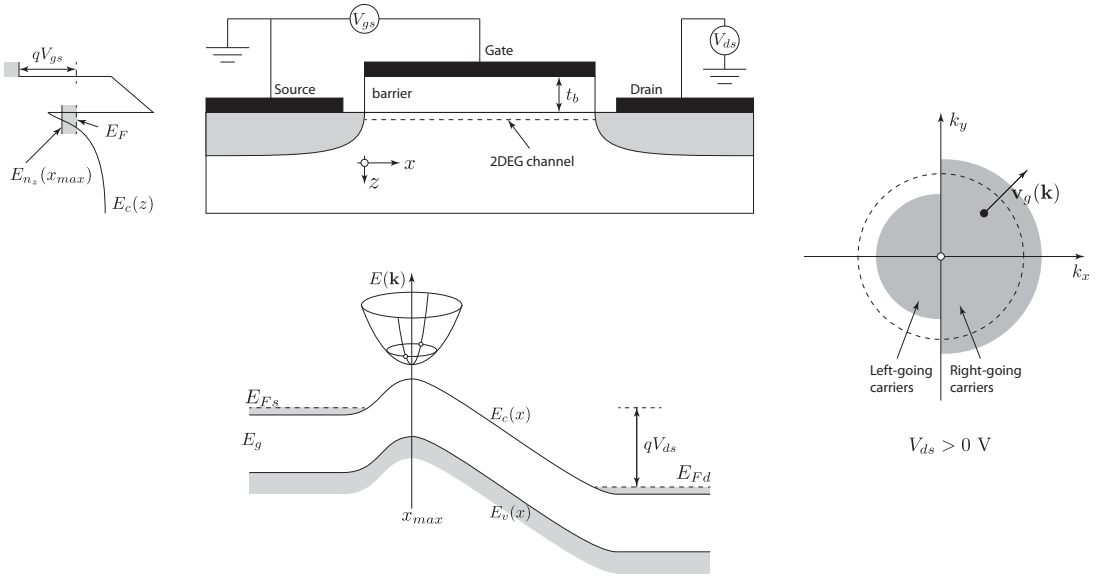

FIGURE 10.3: Field effect transistor, energy band diagram, and $\mathbf{k}-$space occupation of states.

At $V_{ds} = 0$ V, there was a unique E_F at x_{max}, but the quasi-Fermi levels of the right-going carriers and left-going carriers are no longer the same, since $E_{Fs} - E_{Fd} = qV_{ds}$. Due to +ve drain voltage, it has become energetically unfavorable for the drain contact to inject left-going carriers. In the absence of any scattering in the channel, the right-going carriers are still in equilibrium with the source, and the left-going carriers are still in equilibrium with the drain. Thus, the current components now become $J_{2d}^{\rightarrow} = J_0^{2d} F_{1/2}(\eta_s)$ and $J_{2d}^{\leftarrow} = J_0^{2d} F_{1/2}(\eta_d)$. Here $\eta_s = [E_{Fs} - E_{n_z}(x_{max})]/kT$ and $\eta_d = [E_{Fd} - E_{n_z}(x_{max})]/kT = \eta_s - v_d$, where $v_d = qV_{ds}/kT$. The net current of the ballistic transistor is then given by $J_{2d} = J_{2d}^{\rightarrow} - J_{2d}^{\leftarrow}$ as

$$J_{2d} = \frac{qg_s g_v \sqrt{2m^\star}(kT)^{\frac{3}{2}}}{2\pi^2\hbar^2}[F_{\frac{1}{2}}(\eta_s) - F_{\frac{1}{2}}(\eta_s - v_d)] = J_0^{2d}[F_{\frac{1}{2}}(\eta_s) - F_{\frac{1}{2}}(\eta_s - v_d)]. \quad (10.10)$$

The first term is the right-going current carried by the larger gray half-circle in $\mathbf{k}-$apace in Figure 10.1, and the second term is the smaller left-going current carried by the left-going carriers. To evaluate the current, we need to find the dependence of η_s on the gate and drain voltages V_{gs} and V_{ds}.

When $V_{ds} = 0$ V, we found the relation between the unique η and V_{gs} in Eq. 10.6. How do we determine η_s when the carrier distribution looks as in Figure 10.1 with the asymmetric

left-and right-going occupation? Here we make the assumption that the net 2DEG density in the TOB plane at $x = x_{max}$ is *completely* controlled by the gate capacitance. This means the net 2DEG density in the TOB plane has not changed from the $V_{ds} = 0$ V case. Experimentally, this is possible when the transistor is electrostatically well designed, with negligible short-channel effects. Let us assume that such design has been achieved.

Then, just like for the current, we split the carrier distribution equation $C_g(V_{gs} - V_T) = C_q V_{th} F_0(\eta)$ from Equation 10.4 into the right-going and left-going carriers as

$$C_g(V_{gs} - V_T) = C_q V_{th} F_0(\eta) \rightarrow C_q V_{th}[\frac{F_0(\eta^{\rightarrow}) + F_0(\eta^{\leftarrow})}{2}]. \tag{10.11}$$

Identifying $\eta^{\rightarrow} = \eta_s$ and $\eta^{\leftarrow} = \eta_s - v_d$ and using $F_0(x) = \ln[1 + \exp(x)]$, we get the relation

$$\ln[(1 + e^{\eta_s})(1 + e^{\eta_s - v_d})] = \frac{2C_g}{C_q}(\frac{V_{gs} - V_T}{V_{th}}) = 2\eta_g = \ln[e^{2\eta_g}], \tag{10.12}$$

which is a quadratic equation in disguise. Solving for η_s yields

$$\eta_s = \ln[\sqrt{(1 + e^{v_d})^2 + 4e^{v_d}(e^{2\eta_g} - 1)} - (1 + e^{v_d})] - \ln[2], \tag{10.13}$$

which reduces to Equation 10.6 for $v_d = 0$. The expression for η_s with $J_{2d}(V_{gs}, V_{ds}) = J_0^{2d}[F_{\frac{1}{2}}(\eta_s) - F_{\frac{1}{2}}(\eta_s - v_d)]$ provides the complete on-state output characteristics of the ballistic FET at any temperature. Note that the expression depends on the values of Fermi-Dirac integrals of order $j = 1/2$. At $V_{ds} = 0$ V, the drain current is zero, as it should be.

Because of the use of Equation 10.11, just as in Equation 10.4, Equation 10.13 works only for the 'on-state' of the ballistic transistor. The advantage of this form is that the current can be calculated directly. However, if the off-state characteristics of the ballistic FET are desired, one must find the charge self consistently from Equation 10.7 which read $e^{\frac{qn_s}{C_b V_{th}}}(e^{\frac{qn_s}{C_q V_{th}}} - 1) = e^{\frac{V_{gs} - V_T}{V_{th}}}$ and gave us $n_s(V_{gs})$. Then, the expression to use for the entire 'on-state' and 'off-state' or sub-threshold behavior of the ballistic FET is simply

$$\eta_s = \ln[\sqrt{(1 + e^{v_d})^2 + 4e^{v_d}(e^{\frac{2n_s(V_{gs})}{n_q}} - 1)} - (1 + e^{v_d})] - \ln[2], \tag{10.14}$$

where we have simply replaced $\eta_g \rightarrow n_s(V_{gs})/n_q$ in Equation 10.13. Based on this general expression, we can evaluate the entire on-state and off-state characteristics of the ballistic FET.

10.4 Examples

The derived expression of the current of the ballistic FET does not depend on the gate length L. This is a consequence of ballistic transport. Figure 10.4 illustrates the entire output characteristics of a ballistic Silicon transistor. The left figure shows the 'transfer' characteristics in log scale, and the middle figure shows the same in linear scale. Note that Equation 10.14 must be used to obtain the on-off switching characteristics exhibited in this figure. Note that the switching is much steeper at a lower temperature, since the subthreshold slope is $\sim 60 \cdot (T/300)$ mV/decade. The right figure shows the drain current per unit width I_d/W as a function of the drain voltage V_{ds}. When V_{ds} is much larger than kT, $v_d >> 1$, and $\eta_s \rightarrow \ln[e^{2\eta_g} - 1]$. The current then becomes independent of V_{ds}, i.e., saturates to $J_{2d} \rightarrow J_0^{2d} F_{1/2}(\ln[e^{2\eta_g} - 1])$.

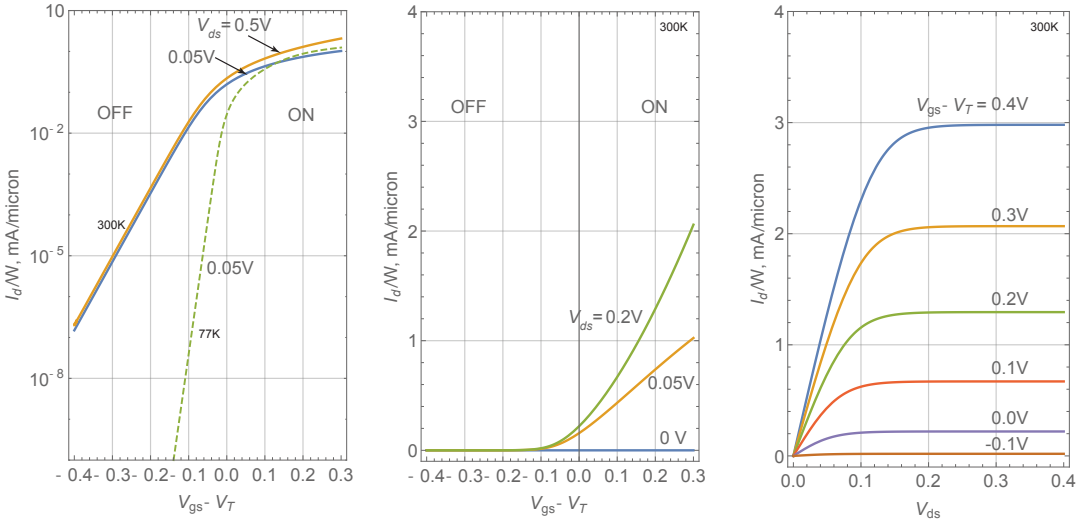

FIGURE 10.4: Ballistic Silicon FET. The device dimensions are $t_b = 1$ nm, $\epsilon_b = 10\epsilon_0$, and for Silicon, $m^\star = 0.2m_0$ and $g_v = 2.5$ are used.

The ballistic FET current expression in equation 10.10 is used to plot a few representative cases. The results at room temperature are shown in Figure 10.5. The barrier thickness for all three FETs is chosen to be $t_b = 2$ nm, of a dielectric constant of $\epsilon_b = 10\epsilon_0$. The channel materials chosen are Si, GaN, and $In_{0.53}Ga_{0.47}As$. For Si, an effective valley degeneracy of $g_v = 2.5$, and an effective mass $m^\star \approx 0.2m_0$ is used. For GaN, $g_v = 1$, and $m^\star \approx 0.2m_0$, and for $In_{0.53}Ga_{0.41}As$ $g_v = 1$, and $m^\star \approx 0.047m_0$ are used. Note

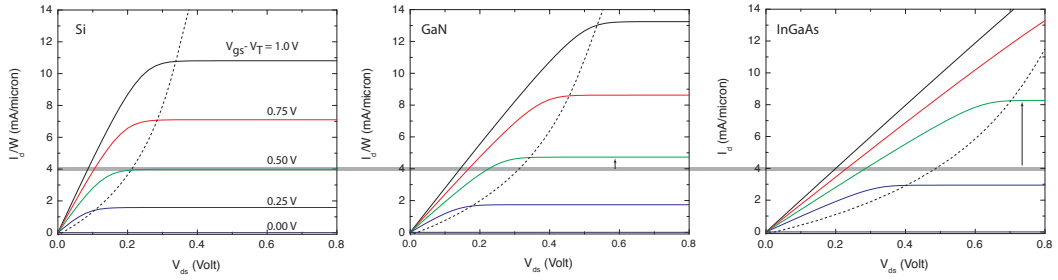

FIGURE 10.5: Ballistic FET characteristics at $T = 300$ K for Si, GaN, and $In_{0.53}Ga_{0.47}As$ channels.

that these are representative material parameters, for correlation with experiments, one must make accurate extraction of band parameters from the electronic bandstructures.

The current in Si channels is higher than GaN and $In_{0.53}Ga_{0.47}As$ channels at low V_{ds}, since it takes advantage of multiple valleys. At high drain bias voltages, the on-current is higher for low effective-mass materials for the same gate overdrive voltage $V_{gs} - V_T$. This boost is due to the higher velocity of carriers due to the low effective mass. For example, at $V_{gs} - V_T = 0.5$ V, the higher saturation currents in GaN and $In_{0.53}Ga_{0.47}As$ channels are shown by arrows in the Figure. However, it takes higher V_{ds} to attain current saturation.

Due to the ultra thin gate and high gate overdrive voltages, the on-currents predicted are rather high. Experimental highest on-current densities approach ~ 4 mA/micron for nanoscale GaN HEMTs, and lower for Si MOSFETs. The experimental currents are limited by source/drain ohmic contact resistances, and gate leakage. These effects have been neglected in the treatment of the ballistic FET.

However, it is remarkable that even for a ballistic FET with zero source and drain contact resistances and no scattering, the low-V_{ds} regime of the ballistic FET has linear $I_d - V_{ds}$ characteristics and looks like a resistor. One can extract effective on-resistances of the order of ~ 0.05 $\Omega-$mm from the linear regions. The origin of this resistance goes back to the limited number of $|\mathbf{k}\rangle$ states available for transport in the 2DEG channel.