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.

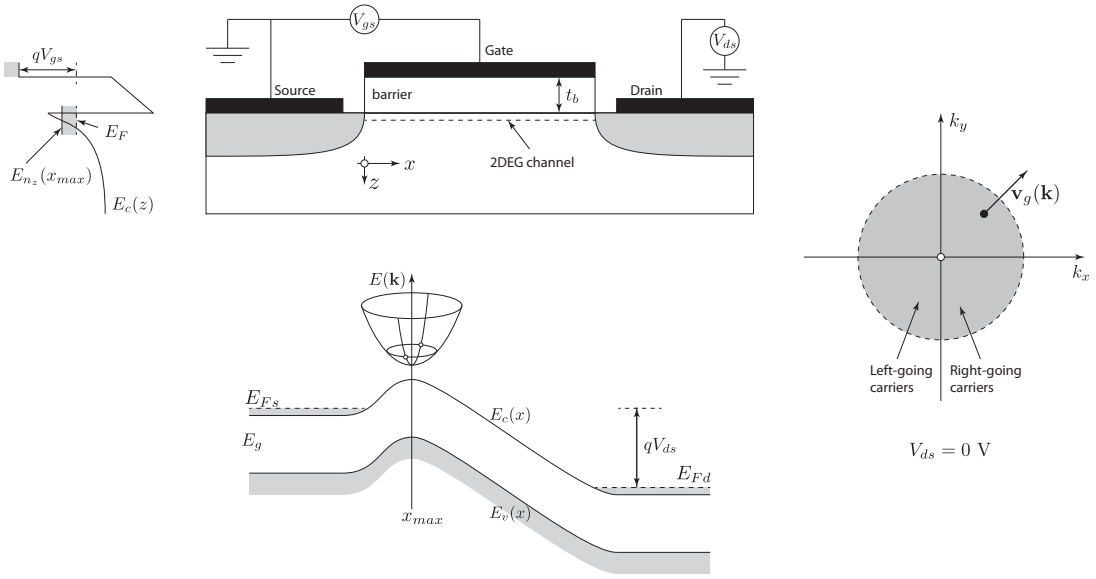

FIGURE 10.1: Field effect transistor, energy band diagram, and $\mathbf{k}-$space occupation of states.

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.

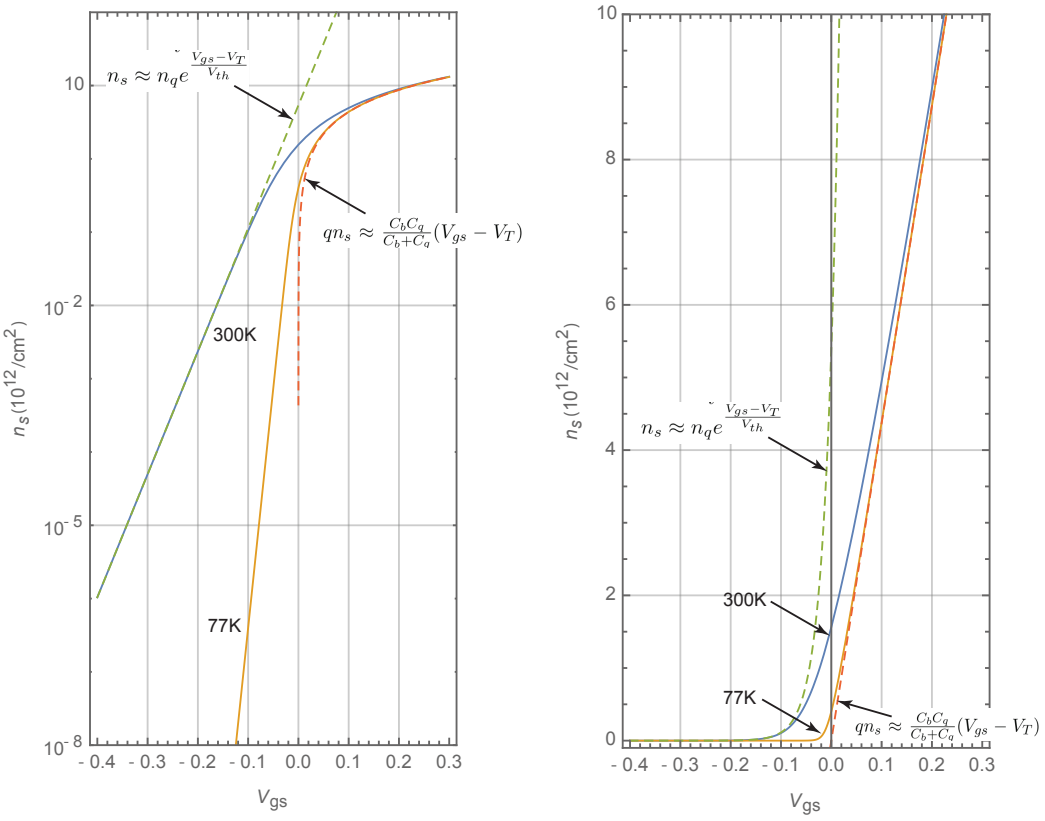

FIGURE 10.2: Illustrating the dependence of the 2DEG sheet density at the injection point on the gate voltage.

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.

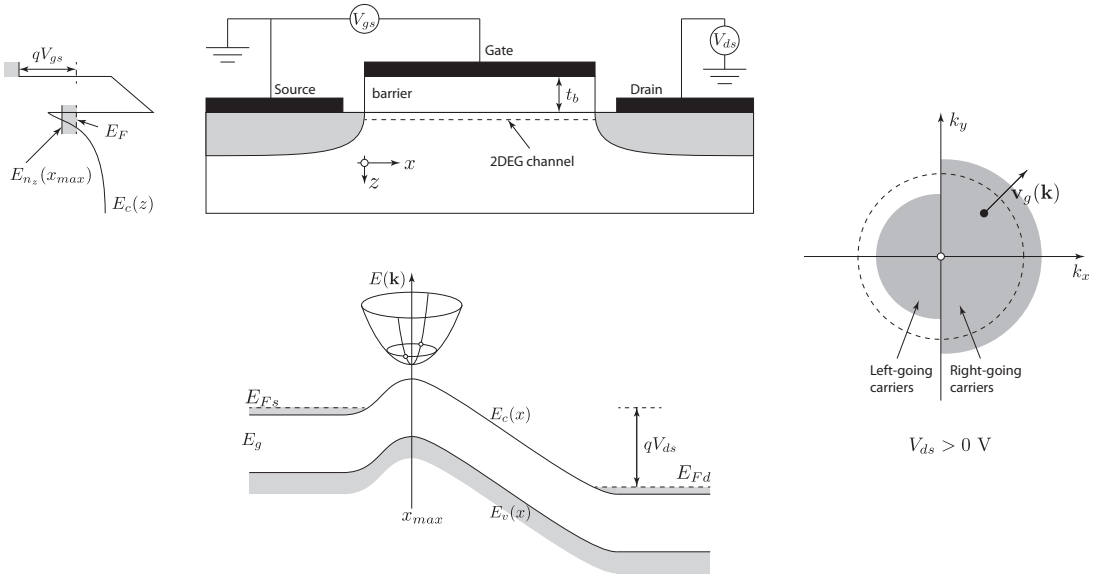

FIGURE 10.3: Field effect transistor, energy band diagram, and $\mathbf{k}-$space occupation of states.

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 $J_{2d} \rightarrow J_0^{2d} F_{1/2}(\ln[e^{2\eta_g} - 1])$.

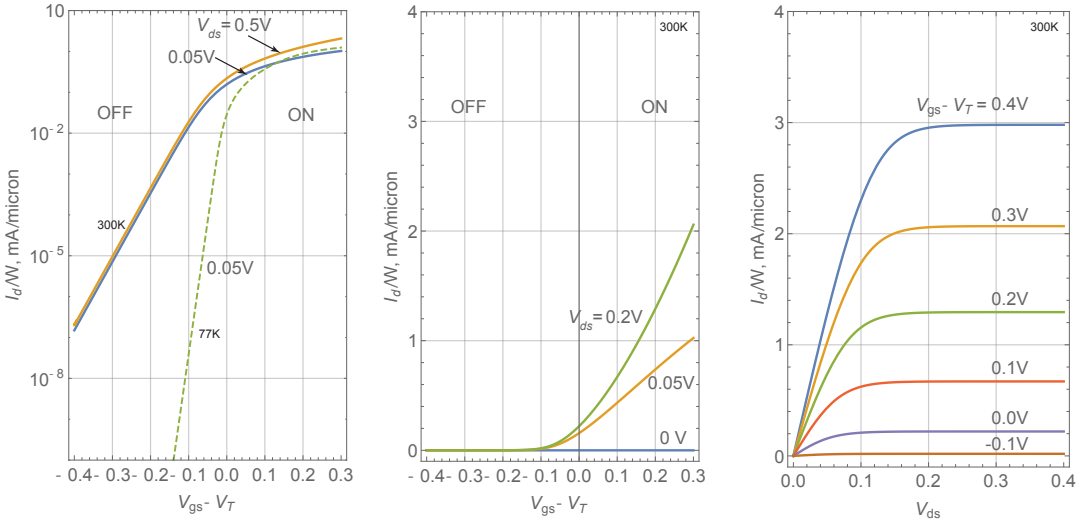

FIGURE 10.4: Ballistic Silicon FET. The device dimensions are $t_b = 1$ nm, $\epsilon_b = 10\epsilon_0$, and for Silicon, $m^\star = 0.2m_0$ and $g_v = 2.5$ are used.

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

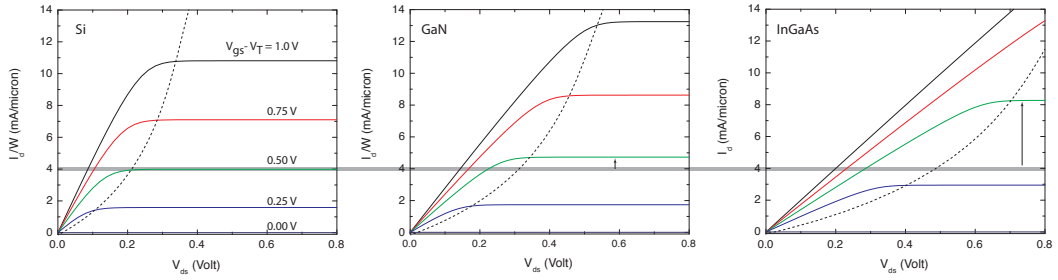

FIGURE 10.5: Ballistic FET characteristics at $T = 300$ K for Si, GaN, and $In_{0.53}Ga_{0.47}As$ channels.

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 $\sim 0.05\ \Omega-$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.