ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/252413897

Notes on the Ballistic MOSFET

Article

CITATIONS
3

READS
291

1 author:

M.s. Lundstrom
Purdue University

534 PUBLICATIONS 18,355 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Scientific Software Development for nanoHUB View project

All content following this page was uploaded by M.s. Lundstrom on 03 July 2014.

The user has requested enhancement of the downloaded file.

Notes on the Ballistic MOSFET

Mark Lundstrom
Network for Computational Nanotechnology
and
Purdue University

1. **Introduction**
2. **The MOSFET as a bipolar transistor**
3. **Generic model for a nanotransistor**
4. **Application to a ballistic MOSFET**
5. **Discussion**
6. **Relation to traditional MOSFET theory**
7. **Summary**

1. Introduction

When analyzing semiconductor devices, the traditional approach is to assume that carriers scatter frequently from ionized impurities, phonons, surface roughness, etc. so that the average distance between scattering events (the so-called mean-free-path, λ) is much shorter than the device. When these conditions hold, we can describe carrier transport with drift-diffusion equations. The traditional derivation of the MOSFET I-V characteristic above threshold assumes that the drift current dominates [1]. For the subthreshold current, we usually assume that diffusion dominates [2]. Numerical simulation programs include both drift and diffusion under all bias conditions (e.g. MINIMOS [3]). As devices shrink, however, we should consider the possibility that the device dimensions become comparable to the mean-free-path for scattering. In the limit, $L << \lambda$, where the channel length is much shorter than the mean-free-path, we can ignore scattering completely. In this case, a MOSFET would operate more like a vacuum tube than like a conventional semiconductor device. In practice, scattering always occurs, but it is common now for the critical, current-limiting part of the device to be comparable in size to a mean-free-path. Modern devices, therefore, operate between the drift-diffusion and ballistic regimes. Drift-diffusion theory is no longer strictly valid, but it continues to provide insights into the operation of small semiconductor devices. A ballistic treatment is not strictly valid either, but it provides new insights that may prove useful as MOSFETs are scaled to their limits and as new devices are explored. The modern device engineer should be familiar with both approaches. In these notes, we develop a simple theory for the ballistic MOSFET.

2. The MOSFET as a Bipolar Transistor

Figure 1 is an energy band diagram of a MOSFET along the channel from the source to the drain with a large drain voltage applied. We see that there is an energy barrier between the source and the drain, and that the energy barrier prevents current flow. In a bipolar transistor, which has a similar energy band diagram, a positive base-emitter voltage lowers the height of the barrier and allows the current to increase exponentially. The MOSFET works the same way, except that the height of the barrier is modulated indirectly by a gate voltage. A positive gate voltage pushes the energy barrier down and allows current to flow. This analogy to the bipolar transistor is often

invoked below threshold, but it also applies above threshold [4]. The analogy to the bipolar transistor highlights the importance of the top of the energy barrier, which will play a central role in our theory of the ballistic MOSFET.

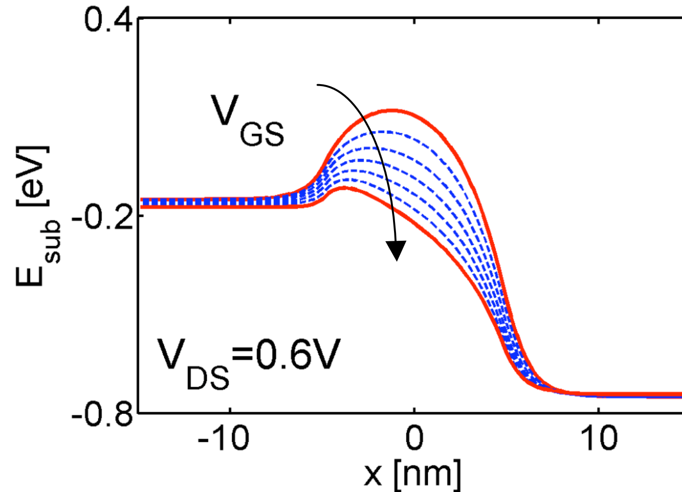

Fig. 1 Electron energy vs. position for an n-channel MOSFET. A high gate voltage is assumed, and the gate voltage is the parameter.

2. Generic Model for a Nanotransistor

Figure 2 shows a simple, generic model for a nanodevice [5]. The device consists of two contacts where strong scattering maintains thermal equilibrium. The left contact, the source, has a Fermi level of E_{F1} and the right contact, the drain, a Fermi level of $E_{F2} = E_{F1} - qV_D$. In equilibrium, these two Fermi levels are equal, but under bias they split. The two contacts are connected to the intrinsic device, which we describe by a density-of-states. A third terminal, the gate, is used to move the energy states in the device up and down through the self-consistent potential.

Figure 3 illustrates how we think about a ballistic transistor in terms of the model device of Fig. 2. We consider the region at the top of the barrier (we give it a length, $\mathcal{L}$) as the intrinsic device. The energy barriers between this region to the source and the drain are the "contacts" that allow carriers to enter and leave from the source and drain. The electronic states at the top of the barrier are filled and emptied from the source and drain, and these states are raised and lowered by the potential on the gate electrode. In the simplest case, we'll assume a parabolic $E(k)$ near the top of the barrier, and for a planar MOSFET, the two-dimensional density-of-states will be

$$D_{2D}(E) = W\mathcal{L}\frac{m^*}{\pi\hbar^2}\Theta(E - E_C), \tag{1}$$

where Θ is the unit step function, and E_C is the bottom of the first subband. (For simplicity in these notes, only a single subband is assumed.) For a nanowire transistor, we would use a one-dimensional density-of-states. Note that we have multiplied by W and $\mathcal{L}$ because we are

interested in the total number of states at energy, E, not the number of states per unit area. For a nanowire transistor, we would use a one-dimensional density-of-states. We could use any $E(k)$ and the corresponding density-of-states. The method is very general. For a molecular transistor, for example, it would not be possible to define an $E(k)$. For such a case, the density-of-states would simply refer to a set of broadened set of molecular levels.

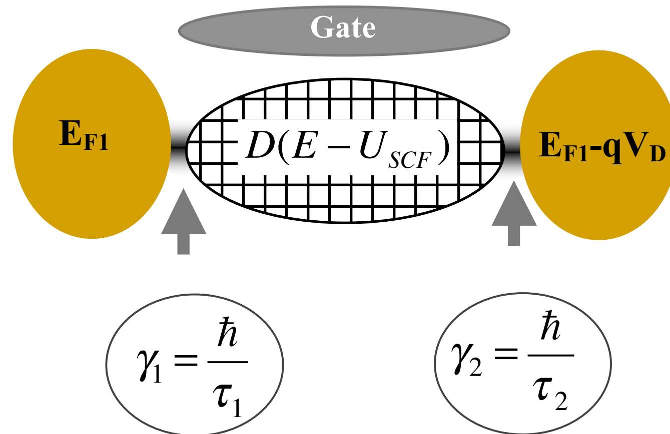

Fig. 2 Generic model for a nanodevice. (After [5])

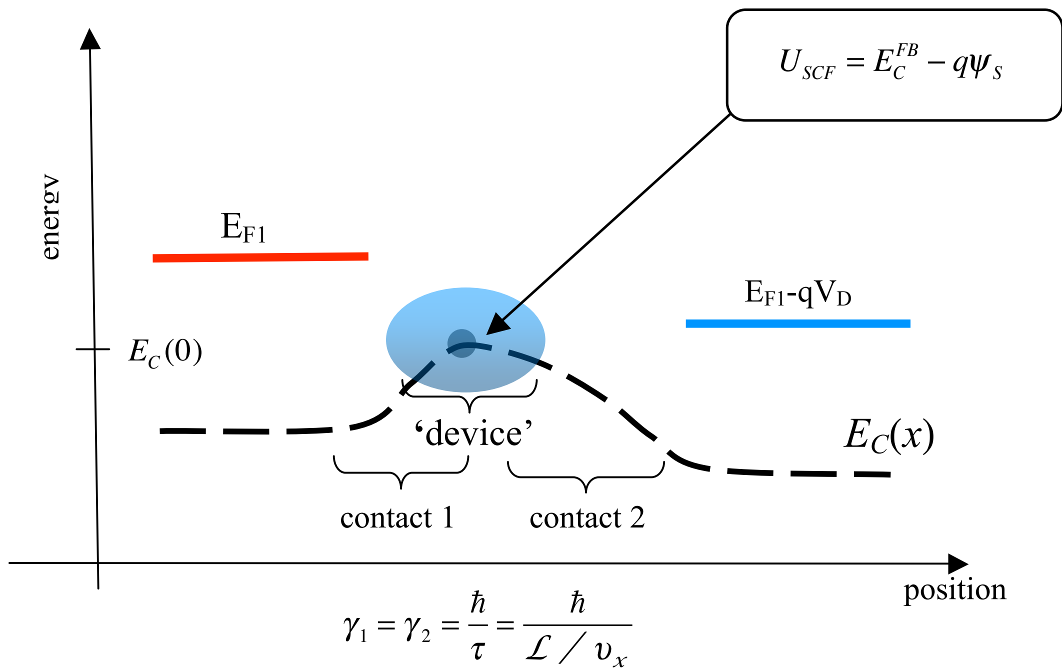

Fig. 3 Representation of a MOSFET in terms of the generic model.

Consider first the case where only the left contact is attached. Contact 1 would fill up the device according to the Fermi level of contact 1. We describe this filling up process with a simple rate equation,

$$\frac{dN(E)}{dt} = \frac{N_1^0(E) - N(E)}{\tau_1}, \tag{2}$$

where

$$N_1^0(E) = D(E - E_C) f_1(E) \tag{3}$$

is the equilibrium number of electrons in the device at energy, E, and τ_1 is a characteristic time that describes how fast carriers enter the device from contact 1 or leave the device through contact 1. Equation (2) states that if the number of electrons in the device is less than the number that contact would like to see in the device, then $N(E)$ increases with time as carriers from the source enter and fill up states in the device. If the number is greater than the equilibrium number, then $N(E)$ decreases with time as carriers leave the device through contact 1.

Similarly, if only the right contact is connected to the device, then contact 2 would try to fill up the device according to its Fermi level, and we describe this filling up process by

$$\frac{dN(E)}{dt} = \frac{N_2^0(E) - N(E)}{\tau_2}, \tag{4}$$

where

$$N_2^0(E) = D(E - E_C) f_2(E) \tag{5}$$

is the equilibrium number of electrons in the device at energy, E, and τ_2 is a characteristic time that describes how fast carriers enter the device from contact 2 or leave the device through contact 2. In general, both contacts are attached, so the rate equation for $N(E)$ becomes

$$\frac{dN(E)}{dt} = \frac{\left(N_1^0(E) - N(E)\right)}{\tau_1} + \frac{\left(N_2^0(E) - N(E)\right)}{\tau_2}. \tag{6}$$

Under steady-state, non-equilibrium conditions,

$$\frac{dN(E)}{dt} = 0, \tag{7}$$

and we can solve eqns. (6) and (7) for the steady-state number of electrons in the device as

$$N(E) = D_1(E) f_1(E) + D_2(E) f_2(E) \tag{8}$$

where

$$D_1(E) \equiv \frac{\tau_2}{\tau_1 + \tau_2} D\left(E - E_C\right) = \frac{\gamma_1}{\gamma_1 + \gamma_2} D\left(E - E_C\right) \tag{9a}$$

$$D_2(E) \equiv \frac{\tau_1}{\tau_1 + \tau_2} D\left(E - E_C\right) = \frac{\gamma_2}{\gamma_1 + \gamma_2} D\left(E - E_C\right). \tag{9b}$$

The quantity, D_1 is the density-of-states fillable by contact 1 and D_2 is the density-of-states fillable by contact 2. Note that when $\tau_2 \to \infty$, $D_2(E) \to 0$, and all of the states are filled according to the Fermi function in contact 1. Conversely, when $\tau_1 \to \infty$, $D_1(E) \to 0$, and all of

the states are filled according to the Fermi function in contact 2. When $\tau_1 = \tau_2$, one-half of the states are filled according to the Fermi function on contact 1 and one-half according to the Fermi function in contact 2. The parameters,

$$\gamma_1 \equiv \frac{\hbar}{\tau_1} \tag{10a}$$

$$\gamma_2 \equiv \frac{\hbar}{\tau_2}, \tag{10b}$$

with units of energy are alternative ways of describing the connections to the contacts. Physically, they represent the broadening of the energy levels in the device.

To understand why the density-of-states splits into two components examine Fig. 4, which shows a simple, parabolic $E(k)$ for the device. In the ballistic device, electrons flow in contact 1 and out from contacts 1 and 2. They also flow in from contact 2 and out from contacts 1 and 2. At the top of the barrier, electrons that came from the source have positive velocities and those that came from the drain have negative velocities. The positive velocity states ($+k$) can only be populated from the source, and the negative velocity states ($-k$) can only be populated from the drain. The source and drain each have one-half of the states available, and they populate their half according to their Fermi levels. The same approach applies to molecular levels in a molecular transistor, even though the concept of a bandstructure with an $E(k)$ does not apply.

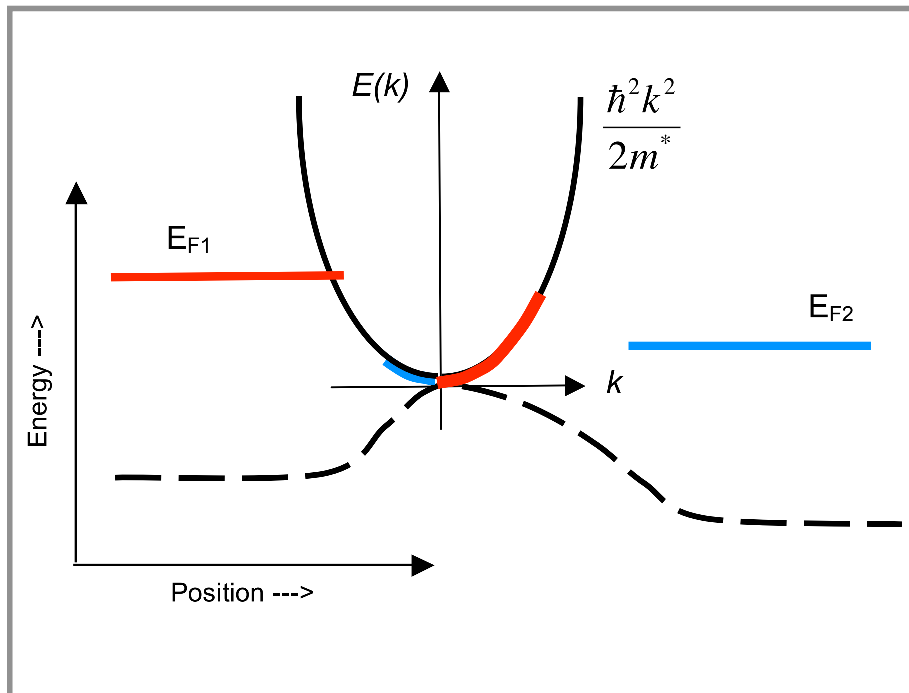

Fig. 4 Illustration of how states at the top of the barrier are filled according to the source and drain Fermi levels.

Finally, we integrate over energy to find the total number of electrons in the device as

$$\boxed{N = \int N(E)dE = \int \left[D_1(E) f_1(E) + D_2(E) f_2(E) \right] dE} \, . \tag{11}$$

Equation (11) is just like the conventional expression for computing carrier densities [1], except that in this case, we have two different Fermi levels and two different densities-of-states. Note that the non-equilibrium carrier densities are determined two different equilibrium Fermi levels. Under steady-state conditions, the number of electrons in the device is determined by a competition between what each of the two contacts would like it to be. To maintain this steady state, there is a continual flow of electrons in one contact and out of the other. The steady-state current is

$$I(E) = \frac{q\left(N_1^0(E) - N(E)\right)}{\tau_1} = \frac{q\left(N_2^0(E) - N(E)\right)}{\tau_2} \, . \tag{12}$$

Because we have solved for *N*, we can use eqn. (8) in (11) to find

$$I(E) = q\left\{ \frac{D(E)}{\tau_1 + \tau_2} \right\} \left[f_1(E) - f_2(E) \right] = \frac{2q}{h} M(E) \left[f_1(E) - f_2(E) \right], \tag{13}$$

where

$$M(E) \equiv \frac{h}{2\left(\tau_1 + \tau_2\right)} D(E) = \frac{h\gamma_1\gamma_2}{2\left(\gamma_1 + \gamma_2\right)} \frac{D(E)}{\hbar} \, . \tag{14}$$

The total current is obtained by summing the contributions from all energies,

$$\boxed{I = \int I(E)dE = \frac{2q}{h} \int M(E) \left[f_1(E) - f_2(E) \right] dE} \, . \tag{15}$$

Current flows when there is a difference of Fermi levels. The parameter, *M*, is particularly simple for a one-dimensional density-of-states. For such a case, $D(E) = 2\mathcal{L}/\pi\hbar\upsilon$ and $\tau_1 = \tau_2 = \mathcal{L}/\upsilon$, and we find $M = 1$. We interpret *M* as the number of propagating modes that carry current. Equation (15) is the well-known Landauer-Büttiker formula for mesoscopic devices, assuming a transmission, $T(E) = 1$ [6].

3. Application to a Ballistic MOSFET

To evaluate the *I-V* characteristics of a MOSFET, we first compute the number of electrons in the device from eqn. (11) and then the current from (15). Assuming $\tau_1 = \tau_2$ and using eqn. (1) for the two-dimensional density-of-states in eqn. (11), we have

$$N = \frac{m^*}{2\pi\hbar^2} W\mathcal{L} \int_{E_C}^{\infty} [f_1(E) + f_2(E)] dE . \tag{16}$$

Consider the first integral,

$$\int_{E_C}^{\infty} [f_1(E)] dE = \int_{E_C}^{\infty} \frac{dE}{1 + e^{(E - E_{F1})/k_B T}} .$$

With a change of variables, $\eta = (E - E_C)/k_B T$, we find

$$\int_{E_C}^{\infty} [f_1(E)] dE = k_B T \int_0^{\infty} \frac{d\eta}{1 + e^{\eta - \eta_{F1}}} = k_B T \ln\left(1 + e^{\eta_{F1}}\right) \equiv k_B T \mathcal{F}_0(\eta_{F1}) ,$$

where $\eta_{F1} \equiv (E_{F1} - E_C)/k_B T$ and $\mathcal{F}_0(\eta_{F1})$ is the Fermi-Dirac integral of order zero. For future reference, recall that for a nondegenerate semiconductor ($\eta_F << 0$), $\mathcal{F}_j(\eta_F) \to e^{\eta_F}$ for any order, *j*, of the Fermi-Dirac integral. Note also, that $\mathcal{F}_0(\eta_{F1}) = \ln\left(1 + e^{\eta_{F1}}\right)$, but for other orders we need to numerically evaluate the integral [12]. Finally, returning to eqn. (16), we find the number of electrons in the device to be

$$N = \frac{N_{2D} W\mathcal{L}}{2} [\mathcal{F}_0(\eta_{F1}) + \mathcal{F}_0(\eta_{F2})] , \tag{17}$$

where

$$N_{2D} = \frac{m^* k_B T}{\pi\hbar^2} \tag{18}$$

is the so-called effective density-of-states per m^2.

The next step is to compute the current from eqn. (15). We begin with eqn. (14) and find

$$M(E) \equiv \frac{hD(E)}{2(\tau_1 + \tau_2)} = \frac{h}{4\tau} \frac{m^* W\mathcal{L}}{\pi\hbar^2} , \tag{19}$$

where we have assumed $\tau_1 = \tau_2 = \tau$. We evaluate τ from

$$\tau = \frac{\mathcal{L}}{\upsilon_x} = \frac{\mathcal{L}}{\upsilon\cos\theta} = \frac{\mathcal{L}}{\sqrt{2E/m^*}\cos\theta} \,. \tag{20}$$

We have assumed that each contact is perfect and that electrons enter or leave the device limited only by the velocity that the bandstructure imposes. Electrons are free to move in the *x-y* plane, and they enter the device with a spread of velocities, so we need to use the average velocity in the *x*-direction (or average cos θ). From

$$\langle\cos\theta\rangle = \int_{-\pi/2}^{+\pi/2} \frac{\cos\theta d\theta}{\pi} = \frac{2}{\pi}$$

and eqn. (20), we find

$$M(E) \equiv \frac{W\sqrt{2m^*E}}{\pi\hbar} \,. \tag{21}$$

Finally, using eqn. (21) in eqn. (15), we find

$$I_D = \frac{2q}{h}\int_{E_C}^{\infty} \frac{W\sqrt{2m^*E}}{\pi\hbar}\left[f_1(E) - f_2(E)\right]dE \,.$$

When we integrate $\sqrt{E}$ times a Fermi function, we get a Fermi-Dirac integral of order ½. After performing the integral and collecting up the constants, we can write the result as

$$I_D = \frac{qN_{2D}}{2} W\,\upsilon_T\left[\mathcal{F}_{1/2}(\eta_{F1}) - \mathcal{F}_{1/2}(\eta_{F2})\right], \tag{22}$$

where

$$\upsilon_T = \sqrt{\frac{2k_BT}{\pi m^*}} \tag{23}$$

is a thermal velocity (twice the so-called Richardson thermal velocity).

Equation (22) is eqn. (4) of ref. [10], but be careful about the 'roman F' and 'script F' definitions of the Fermi-Dirac integral! Natori uses

$$F_{1/2}(\eta) \equiv \int_0^{\infty} \frac{\varepsilon^{1/2}d\varepsilon}{1+\exp(\varepsilon-\eta)} \tag{24a}$$

but I use [12]

$$\mathcal{F}_{1/2}(\eta) \equiv \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\varepsilon^{1/2} d\varepsilon}{1+\exp(\varepsilon-\eta)} \tag{24b}$$

Let's pause and review what we have accomplished. We first derived an expression for the number of electrons in a ballistic MOSFET, eqn. (17),

$$N = \frac{N_{2D}}{2} W\mathcal{L}\left[\mathcal{F}_0(\eta_{F1}) + \mathcal{F}_0(\eta_{F2})\right],$$

and then we derived an expression for the drain current, eqn. (22),

$$I_D = \frac{qN_{2D}}{2} W \upsilon_T \left[\mathcal{F}_{1/2}(\eta_{F1}) - \mathcal{F}_{1/2}(\eta_{F2})\right].$$

To get the *I-V* characteristics of the device, we add ideal, MOS electrostatics. Above threshold,

$$C_{ox} W\mathcal{L}(V_{GS} - V_T) = qN, \tag{25}$$

which can be used to solve eqn. (17) for

$$\frac{qN_{2D}W}{2} = \frac{W C_{ox}(V_{GS} - V_T)}{\left[\mathcal{F}_0(\eta_{F1}) + \mathcal{F}_0(\eta_{F2})\right]}.$$

This expression can be used in eqn. (22) to find

$$I_D = W C_{ox} \upsilon_T \left(V_{GS} - V_T\right)\left[\frac{\mathcal{F}_{1/2}(\eta_{F1}) - \mathcal{F}_{1/2}(\eta_{F2})}{\mathcal{F}_0(\eta_{F1}) + \mathcal{F}_0(\eta_{F2})}\right],$$

or

$$\boxed{I_D = W C_{ox} \tilde{\upsilon}_T \left(V_{GS} - V_T\right)\left[\frac{1 - \dfrac{\mathcal{F}_{1/2}(\eta_{F1} - qV_D/k_BT)}{\mathcal{F}_{1/2}(\eta_{F1})}}{1 + \dfrac{\mathcal{F}_0(\eta_{F1} - qV_D/k_BT)}{\mathcal{F}_0(\eta_{F1})}}\right]}, \tag{26}$$

where

$$\boxed{\tilde{\upsilon}_T = \sqrt{\frac{2k_BT}{\pi m^*}} \frac{\mathcal{F}_{1/2}(\eta_{F1})}{\mathcal{F}_0(\eta_{F1})}} \tag{27}$$

is the so-called injection velocity, the average velocity at the top of the barrier. Equations (26) and (27) do not provide a complete description of the *I-V* characteristic of the MOSFET because they are expressed in terms of η_{F1}. The third equation allows us to find η_{F1} by solving,

$$\boxed{\mathcal{F}_0\left(\eta_{F1}\right)+\mathcal{F}_0\left(\eta_{F1}-qV_D/k_BT\right)=\frac{C_{ox}\left(V_{GS}-V_T\right)}{qN_{2D}/2}}\,. \tag{28}$$

Equations (26) – (28) describe a ballistic MOSFET under the assumption of ideal (1D) MOS electrostatics and above threshold operation.

Finally, note that we have assumed an $E(k)$ with a single spherical valley. For the silicon conduction band, for example, there are multiple ellipsoidal valleys. These results are readily extended for the silicon conduction band as described in refs. [9, 10, 13].

4. Subthreshold and 2D Electrostatics

Equations (17) and (22) describe the MOSFET under subthreshold as well as above threshold conditions, if we properly relate the surface potential to the gate voltage. The drain voltage can also affect the surface potential, and the resulting 2D electrostatics play an important role in the MOSFET. The surface potential controls the location of the top of the barrier according to

$$\eta_{F1}=\left(E_{F1}-E_C^{FB}+q\psi_S\right)\big/k_BT\ , \tag{29}$$

where E_C^{FB} is the location of the top of the barrier under flat band conditions, which we take as the reference for $\psi_S=0$. In general, we need a 2D simulation to determine ψ_S, but we can represent the effect of the terminal voltages on ψ_S by the equivalent circuit shown in Fig. 5. From this circuit, we find

$$\boxed{\psi_S=\frac{C_G}{C_\Sigma}V_G+\frac{C_G}{C_\Sigma}V_D+\frac{C_G}{C_\Sigma}V_S-\frac{q\left(N-N_0\right)}{C_\Sigma}}\,, \tag{30}$$

where

$$C_\Sigma=C_G+C_D+C_S \tag{31}$$

and N_0 is the carrier density at zero applied voltage and zero surface potential. For the ideal case, $C_D=C_S=0$, and the surface potential is completely controlled by the gate voltage.

Before we proceed, we should note that the charge in eqn. (30) is the mobile injected from the source and drain. In a bulk MOSFET, the channel is heavily doped, and we should include a depletion charge. For a uniformly doped substrate

$$Q_D=\sqrt{2q\kappa_S\varepsilon_0N_A\left(\psi_S+V_{SB}\right)}\,W\mathcal{L}\ ,$$

where V_{SB} is the reverse bias (if any) between the source and the body. In general, the depletion charge should be added to qN in eqn. (30), but we will not consider this complication in these notes.

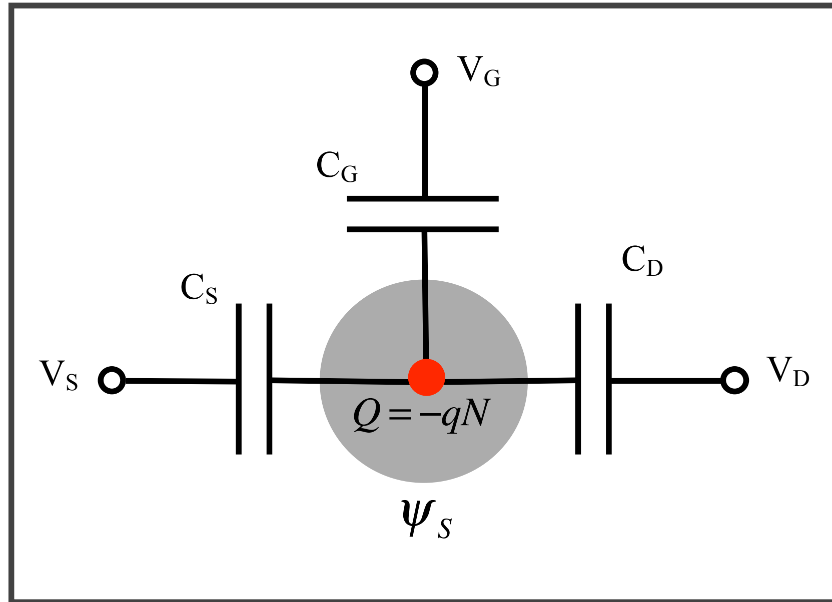

Fig. 5 A simple circuit to include the effect of the gate, drain, and source voltages on the surface potential.

Using eqn. (30), the procedure for computing the I-V characteristic of a MOSFET is as follows [7]:

1) Select a gate and drain voltage, V_G and V_D
2) Guess the surface potential, ψ_S from which we determine, η_{F1} from eqn. (29)
3) Evaluate N from eqn. (17)
4) Determine an improved guess for ψ_S by inserting the value of N from step 3) in eqn. (30)
5) Iterate between 3) and 4) until the surface potential converges to a fixed value
6) Evaluate the drain current from eqn. (22)
7) Go to 1) and repeat for another gate and drain voltage

To reduce this procedure to practice, there are few additional questions. First, how do we determine $\left(E_{F1} - E_C^{FB}\right)$, which we need to determine η_{F1} from eqn. (29)? This parameter just sets the source to channel barrier height when $V_G = V_D = 0$. It determines the off-current or, equivalently, the threshold voltage. In practice, we set this parameter to a large negative value (several k_BT/q in magnitude) to produce the desired off-current.

The next question is how to determine the three capacitors, C_G, C_D, and C_S. The first one is readily obtained

$$C_G = \frac{\kappa_{OX}\varepsilon_0}{T_{ox}} W \mathcal{L} .$$

We also know that under subthreshold conditions,

$$I_D \propto e^{q\psi_S / k_B T} ,$$

so

$$\left.\frac{\partial I_D}{\partial V_G}\right|_{V_D} \propto e^{q\psi_S / k_B T} \frac{\partial \psi_S}{\partial V_G} = \frac{C_G}{C_\Sigma} \frac{I_D}{k_B T / q} , \tag{32}$$

where we have used eqn. (30) to evaluate the derivative. Next, recall that according to standard MOSFET theory the subthreshold current varies as [2]

$$I_D \propto e^{qV_G / m k_B T} ,$$

so

$$\left.\frac{\partial I_D}{\partial V_G}\right|_{V_D} \propto e^{qV_G / m k_B T} \frac{q}{m k_B T} = \frac{1}{m} \frac{I_D}{k_B T / q} . \tag{33}$$

Comparing eqns. (32) and (33), we conclude that

$$\frac{1}{m} = \frac{C_G}{C_\Sigma} . \tag{34}$$

The parameter, m, is related to the subthreshold swing, S, of a transistor through [2]

$$S = 2.3 m k_B T / q . \tag{35}$$

Finally, from eqns. (34) and (35), we find

$$\boxed{\frac{C_G}{C_\Sigma} = \frac{2.3 k_B T / q}{S}} . \tag{36}$$

The parameter, S, which gives the gate voltage swing needed to increase the subthreshold current by a factor of 10, is frequently known experimentally or from a 2D Poisson solution, so it provides us with a convenient way to determine C_G/C_Σ.

An increase in drain voltage shifts the threshold voltage to a lower value. In a typical plot of $\log_{10} I_D$ vs. V_G at two different drain voltages, this is observed as a parallel shift of the subthreshold current vs. voltage plot. The shift is call DIBL (drain-induced barrier lowering) and is measured in V/V (voltage of shift in gate voltage at a constant I_D per volt of change in the drain voltage, Beginning with

$$I_D \propto e^{q\psi_S/k_BT},$$

we find

$$\left.\frac{\partial \log_{10} I_D}{\partial V_D}\right|_{V_G} \propto \frac{1}{2.3k_BT/q}\frac{\partial \psi_S}{\partial V_D} = \frac{1}{2.3k_BT/q}\frac{C_D}{C_\Sigma}.$$

For a given ΔV_D, the increase in $\log_{10} I_D$ is

$$\Delta \log_{10} I_D = \frac{\Delta V_D}{2.3k_BT/q}\frac{C_D}{C_\Sigma},$$

but we also know from standard MOSFET theory, that

$$\Delta \log_{10} I_D = \frac{\Delta V_G}{S}.$$

By equating these two expressions, we find

$$\boxed{\frac{C_D}{C_\Sigma} = \frac{2.3k_BT}{q}\frac{DIBL}{S}}, \tag{37}$$

which allows us to estimate C_D/C_Σ from S and DIBL, which are frequently known.

5. Discussion

Above threshold, Fermi-Dirac statistic must be used, but it is instructive to examine the final result for nondegenerate conditions where Fermi-Dirac integrals reduce to exponentials and eqn. (26) becomes

$$I_D = W C_{ox} \upsilon_T \left(V_{GS} - V_T\right)\left[\frac{1-e^{\eta_{F2}-\eta_{F1}}}{1+e^{\eta_{F2}-\eta_{F1}}}\right].$$

Since the drain voltage lowers the Fermi level of the drain by $-qV_{DS}$, we have $\eta_{F2} - \eta_{F1} = qV_{DS}/k_B T$, and the final result for the nondegenerate, ballistic MOSFET is

$$I_D = W C_{ox} \upsilon_T (V_{GS} - V_T) \left[\frac{1 - e^{-qV_{DS}/k_B T}}{1 + e^{-qV_{DS}/k_B T}} \right]. \tag{38}$$

It is instructive to examine eqn. (38) under low drain bias. If we expand the exponentials for small V_{DS}, and collect some constants, we find

$$I_D = \frac{W}{L} \mu_{Ball} C_{ox} (V_{GS} - V_T) V_{DS}, \tag{39}$$

$$\mu_{Ball} \equiv \frac{L(\upsilon_T/2)}{k_B T/q}. \tag{40}$$

Equation (39) is the conventional result, except that the effective mobility is replaced by a non-physical "ballistic mobility." The ballistic mobility is the expected expression for the mobility - except that the channel length replaces the mean-free-path. Since our model transistor assumes that scattering occurs in the source and in the drain, it is reasonable to call L the mean-free-path for scattering in a ballistic MOSFET.

For large V_{DS}, eqn. (38) becomes

$$I_D = W C_{ox} \upsilon_T (V_{GS} - V_T), \tag{41}$$

which shows that the saturated drain current increases linearly with $(V_{GS} - V_T)$ rather than quadratically as in a simple MOSFET models. We also see that the drain saturation voltage is a few $k_B T/q$ rather than $(V_{GS} - V_T)$.

It's worth pointing out that we can get eqn. (38) very simply. In the ballistic case,

$$Q(0) = q\left[n^+(E_F) + n^-(E_F - qV_D) \right] = Q^+(E_F) + Q^-(E_F - qV_D), \tag{42}$$

where we have assumed ideal, MOS electrostatics. We can also write the current as

$$I_D = -\left[Q^+ \upsilon^+ - Q^- \upsilon^- \right] = qn^+ \upsilon^+ \left[1 - \frac{n^- \upsilon^-}{n^+ \upsilon^+} \right]. \tag{43}$$

By writing eqn. (42) as

$$Q(0) = -qn^+ \left[1 + \frac{n^-}{n^+} \right], \tag{44}$$

solving for n^+, and inserting the result in eqn. (43), we find

$$I_D = -Q(0)\upsilon_T \frac{\left[1 - n^-/n^+\right]}{\left[1 + n^-/n^+\right]}, \tag{45}$$

where we have used the fact that $\upsilon^+ = \upsilon^- = \upsilon_T$ for a nondegenerate semiconductor. Finally, by assuming ideal, MOS electrostatics, $Q(0) \approx W C_{ox}(V_{GS} - V_T)$, and recognizing that for a nondegenerate semiconductor, $n^-/n^+ = e^{-qV_D/k_BT}$, we obtain eqn. (38).

The non-degenerate model is simple and gives some insight, but above threshold, degenerate statistics must be used. Degenerate carrier statistics raise the injection velocity, increase the saturation voltage, and cause the saturated drain current to vary as $(V_{GS} - V_T)^\alpha$, where $1 < \alpha < 1.5$. In summary, the ballistic MOSFET model predicts I-V characteristics that look much like those of a conventional MOSFET. In practice, we find that the magnitude of the current in a state-of-the-art MOSFET is roughly 50% of the corresponding ballistic MOSFET.

6. Relation to traditional MOSFET theory

The ballistic MOSFET is an idealization that does not occur in practice, but traditional MOSFET theory assumes channel lengths much longer than a mean-free-path for scattering, and that assumption has also become unrealistic. The model presented here can be modified to include scattering and to derive the traditional MOSFET model. Doing so by beginning from the ballistic perspective provides some new insights into MOSFET performance.

According to eqn. (15), to evaluate the current, we need to specify $M(E)$ and therefore the escape times, $\tau_1 + \tau_2$. In the presence of scattering, carriers will diffuse out of the device in a time

$$\tau_1 = \tau_2 = \frac{L^2}{2D_{eff}}, \tag{46}$$

where L is the channel length and D_{eff} is the effective diffusion coefficient of carriers. (Equation (46) is the same expression as for the transit time across the base of a bipolar transistor [1].) Using eqn. (46) in eqn. (19) for M, we find

$$M = \frac{W}{L}\frac{hD_{eff}}{2}\frac{m^*}{\pi\hbar^2}, \tag{47}$$

which can be inserted in eqn. (15) to find the drain current as

$$I_D = \frac{W}{L} q N_{2D} D_{eff}\left[\mathcal{F}_0(\eta_{F1}) - \mathcal{F}_0(\eta_{F2})\right]. \tag{48}$$

Next, we should relate N_{2D} to the carrier density in the channel, but we should not use eqn. (17). Equation (17) describes the filling of states in the ballistic case; positive *k*-states are filled from the source and negative *k*-states from the drain. In the presence of strong scattering, all of the *k*-states at the beginning of the channel are filled from the source – the $+k$ states by direct injection from the source and the $-k$ states by source-injected carriers that backscatter. It is better, therefore, to write

$$N = N_{2D} W \mathcal{L} \mathcal{F}_0(\eta_{F1}) = C_{ox}(V_{GS} - V_T)\frac{W\mathcal{L}}{q}, \tag{49}$$

where the final term assumes ideal (above-threshold) MOS electrostatics. By solving eqn. (49) for N_{2D} in terms of V_{GS} and inserting the result in eqn. (48), we find

$$I_D = \frac{W}{L} D_{eff} C_{ox}(V_{GS} - V_T)\left[1 - \frac{\mathcal{F}_0(\eta_{F2})}{\mathcal{F}_0(\eta_{F1})}\right]. \tag{50}$$

Equation (50) still looks much different than the traditional result, but if we assume non-degenerate carrier statistics (so $\mathcal{F}_0(\eta_F) \to e^{\eta_F}$), eqn. (50) becomes

$$I_D = \frac{W}{L} D_{eff} C_{ox}(V_{GS} - V_T)\left[1 - e^{-qV_D/k_BT}\right]. \tag{51}$$

For small drain to source voltage, eqn. (51) becomes

$$I_D = \frac{W}{L}\frac{D_{eff}}{(k_BT/q)} C_{ox}(V_{GS} - V_T)V_D = \frac{W}{L}\mu_{eff} C_{ox}(V_{GS} - V_T)V_D, \tag{52}$$

which is the traditional result for low drain voltages [1].

For high drain voltages, eqn. (51) becomes

$$I_D = W\frac{D_{eff}}{\mathcal{L}} C_{ox}(V_{GS} - V_T) = W\upsilon_{diff} C_{ox}(V_{GS} - V_T). \tag{53}$$

Note that we have replaced the channel length, L, with an effective length, $\mathcal{L}$ and defined a diffusion velocity, $\upsilon_{diff} = D_{eff}/\mathcal{L}$, which is the average velocity at which carriers diffuse across the region, $\mathcal{L}$. The physical interpretation of eqn. (53) is that under high drain bias, there is short, low-field region at the beginning of the channel (the top of the barrier region as in Fig. 3) across which carries must diffuse. This bottleneck limits the current, because once they diffuse across the region, they are quickly swept across the high-field part of the channel [8].

Is the simple physical picture that emerges from our model consistent with traditional MOSFET theory? Physical arguments tell us that the length of the bottleneck at the beginning of the

channel is the distance over which the potential drops by k_BT/q [8]. (This result assumes a non-degenerate carrier gas, the required potential drop is larger for a degenerate gas [8].) We can write the length of this critical region as

$$\mathcal{L} = \frac{\mathcal{E}(0)}{\left(k_B T/q\right)}, \tag{54}$$

where $\mathcal{E}(0)$ is the electric field at the beginning of the channel. According to traditional MOS theory

$$\mathcal{E}(0) = \frac{\left(V_{GS} - V_T\right)}{2L}, \tag{55}$$

and the Einstein relation is

$$D_{eff} = \mu_{eff}\frac{k_B T}{q}. \tag{56}$$

By using eqns. (54) – (56) in (53), we find

$$I_D = \frac{W}{2L}\mu_{eff} C_{ox}\left(V_{GS} - V_T\right)^2, \tag{57}$$

which is the traditional result for the long channel MOSFET under high drain bias [1].

6. Summary

In these notes, I have outlined a simple theory for the ballistic MOSFET that extends the original treatment of the ballistic MOSFET of Natori [9, 10]. To test your understanding of the concepts discussed in these notes, you might try to extend the results to include two subbands, for a numerically tabulated $E(k)$, or for a nanowire transistor. The extension to ellipsoidal, multi-valley energy bands, specifically the conduction band of silicon, is discussed in [9, 10, 13]. For a more advance treatment of this subject, see ref. [7]. Reference [7] also includes a discussion of 'floating boundary conditions', which can be important in some cases. We also showed how the ballistic model can be modified to include scattering, and we showed that the results reduce to the traditional theory of the MOSFET. For a physical discussion of scattering in nanotransistors, see ref. [8]. We have discussed the ballistic limit and the diffusive limit. For a treatment of the general case, see [11].

References

[1] R.F. Pierret, *Semiconductor Fundamentals*, 2nd Ed., Prentice Hall, Englewood Cliff, NJ, USA, 2002.

[2] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge Univ. Press, Cambridge, UK, 1998.

[3] S. Selberherr, A. Schutz, and H.W. Potzl, “MINIMOS – A Two-Dimensional MOS Transistor Analyzer,” *IEEE Trans. Electron Dev*., **27**, pp. 1540-1550, 1980.

[4] E.O. Johnson, “The Insulated Gate Field Effect Transistor: A Bipolar Transistor in Disguise,” *RCA Review,* **34**, p. 80, 1973.

[5] S. Datta, *Quantum Transport: From Atom to Transistor,* Cambridge Univ. Press, Cambridge, UK, 2005.

[6] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge Univ. Press, Cambridge, UK, 1995.

[7] Anisur Rahman, Jing Guo, Supriyo Datta, and Mark Lundstrom, “Theory of Ballistic Nanotransistors,” *IEEE Trans. Electron. Dev*., and I*EEE Trans. on Nanotechnology*, joint special issue on Nanoelectronics, **50**, pp. 1853-1864, 2003.

[8] Mark Lundstrom and Zhibin Ren, “Essential Physics of Carrier Transport in Nanoscale MOSFETs,” *IEEE Trans. Electron Dev*., **49**, pp. 133-141, January, 2002.

[9] K. Natori, “Ballistic metal-oxide-semiconductor field effect transistor,” *J. Appl. Phys.*, **76**, pp. 4879-4890, 1994.

[10] K. Natori, “Scaling limit of the MOS transistor – A Ballistic MOSFET,” *IEICE Trans. Electron.*, **E84-C**, pp. 1029-1036, 2001.

[11] M.S. Lundstrom, “Elementary Scattering Theory of the Si MOSFET,” *IEEE Electron Dev. Letters*, **18**, pp. 361-363, 1997.

[12] J.S. Blakemore, *Semiconductor Statistics*, Dover, 1987.

[13] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M.S. Lundstrom, “Performance Limits of Silicon MOSFETS: A Theoretical Study,” *IEEE Trans. Electron Devices*, **47**, 232-240, 2000.

View publication stats