Chapter 4
Silicon Nanowire Field-Effect Transistor

Dae Mann Kim, Bomsoo Kim and Rock-Hyun Baek

Abstract The field effect transistor was conceived in 1930s and was demonstrated in 1960s. Since then, MOSFET emerged as the mainstream driver for the digital information technology. Because of the simplicity of structure and low cost of fabrication, it lends to a large scale integration for the multifunctional system-on-chip (SOC) applications. Moreover, the device has been relentlessly downsized for higher performance and integration. The physical barriers involved in downscaling the device have prompted the development of process technologies. There has also been the development of device structures from 3D bulk to the gate-all-around nanowire. This chapter is addressed to the discussion of the silicon nanowire field effect transistor (SNWFET). The discussion is carried out in comparison and correlation with the well known theory of MOSFET. The similarities and differences between the two FETs are highlighted, thereby bringing out features unique to SNWFET. Also, an emphasis is placed upon the underlying device physics rather than the device modeling per se. The goal of this chapter is to provide a background by which to comprehend the theories being developed rapidly for SNWFETs.

Abbreviation

SOC	System-on-chip
SNWFET	Silicon nanowire field-effect transistor

D. M. Kim (✉) · B. Kim
Korea Institute for Advanced Study (KIAS), Seoul, Republic of Korea
e-mail: dmkim@kias.re.kr

B. Kim
e-mail: bomsoo@kias.re.kr

R.-H. Baek
SEMATECH, Albany, NY, USA
e-mail: rock-hyun.baek@sematech.org

D. M. Kim and Y.-H. Jeong (eds.), *Nanowire Field Effect Transistors: Principles and Applications*, DOI: 10.1007/978-1-4614-8124-9_4,

4.1 MOSFET

Overview: In this section the theory of MOSFET is compactly summarized to be used as the reference for discussing silicon nanowire field-effect transistor (SNWFET). MOSFET is a three-terminal, unipolar and normally off device and has been successfully scaled down to about 10 nm channel length. Also, the device has provided convenient platforms for a number of applications e.g., memory cells, sensors and solar cells, etc. Of the two types of MOSFETs, NMOS is singled out for discussion. The results obtained can readily be applied to PMOS by replacing the roles of electrons with those of holes.

NMOS I–V Behavior: Fig. 4.1 shows the cross section, consisting of the n^+ source and drain and the n^+ polysilicon gate, which is electrically insulated by SiO_2. The source and drain are separated by the p-type substrate, so that n^+-p and p-n^+ junctions are formed back to back. With the gate voltage off ($V_{GS} = 0$) and the drain voltage on ($V_{DS} > 0$), the p-n^+ junction at the drain end gets reverse biased, cutting off the current (off state), as detailed in Chap. 3.

However, with V_{GS} on at a value greater than the threshold voltage, i.e., $V_{GS} > V_{Tn}$, the channel is inverted. In which case, the n^+-p junction barrier at the source end is lowered and electrons are injected from the source into the channel and contribute to the drain current, I_D. Figure 4.2 shows the transistor I–V curves and transfer characteristics. Each $I_D - V_D$ curve divides into the triode and

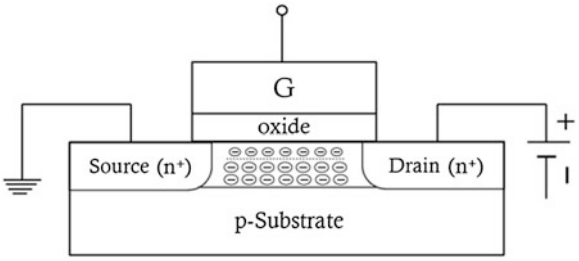

Fig. 4.1 The cross-sectional view of NMOS, consisting of the p-substrate, n^+ source, drain and gate electrodes

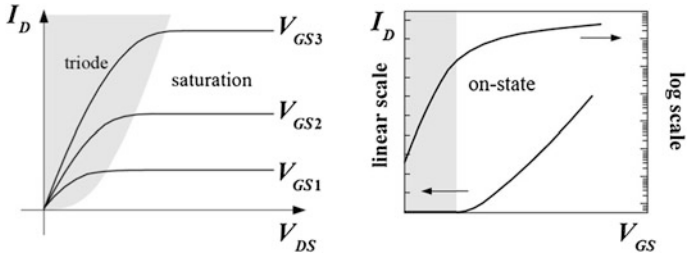

Fig. 4.2 The transistor I–V curves (*left*) and transfer characteristics (*right*) of the n-channel MOSFET

saturation regions. The ON-to-OFF current ratio, typically 10^6 or greater, is a parameter gauging the device as a switch.

The standard long channel I–V behavior is described by the SPICE model, level 1:

$$I_{\mathrm{DS}} = \frac{W}{L} C_{\mathrm{OX}} \mu_n (V_{\mathrm{GS}} - V_{\mathrm{Tn}} - \frac{1}{2} V_{\mathrm{DS}}) V_{\mathrm{DS}}, \quad 0 \le V_{\mathrm{DS}} \le V_{\mathrm{DSAT}} \equiv V_{\mathrm{GS}} - V_T \tag{4.1}$$

Here, μ_n is the electron mobility, V_{Tn} the threshold voltage, V_{GS} the gate-to-source voltage and the ratio between width and length of the channel, *W/L* is called the aspect ratio. The oxide capacitance per unit area is given in terms of the oxide permittivity, $\varepsilon_{\mathrm{OX}}$, and thickness, t_{OX}, as $C_{\mathrm{OX}} = \varepsilon_{\mathrm{OX}}/t_{\mathrm{OX}}$. The triode and saturation regions of I_{DS} are demarcated by $V_{\mathrm{DSAT}} (= V_{\mathrm{GS}} - V_{\mathrm{Tn}})$, at which the channel is pinched off.

Equivalently I_{DS} can be compacted into a simpler form as

$$I_{\mathrm{DS}} = Q_{\mathrm{nL}} v_D, \; Q_{\mathrm{nL}} \equiv W C_{\mathrm{OX}} (V_{\mathrm{GS}} - V_{\mathrm{Tn}} - V_{\mathrm{DS}}/2), \; v_D = \mu_n (V_D/L) \tag{4.2}$$

Here, $v_D = \mu_n (V_D/L)$ is the drift velocity of the electron, and Q_{nL} is the average line charge induced under the gate. In this representation, I_{DS} is shown to be contributed by Q_{nL} sweeping across the channel with v_D. In device saturation, where $V_{\mathrm{DSAT}} = V_{\mathrm{GS}} - V_{\mathrm{Tn}}$ (4.2) is reduced to

$$I_{\mathrm{DSAT}} = Q_{\mathrm{nSAT}} v_D, \; Q_{\mathrm{nSAT}} \equiv W C_{\mathrm{OX}} (V_{\mathrm{GS}} - V_{\mathrm{Tn}})/2 \tag{4.3}$$

with Q_{nSAT} denoting the average line charge in saturation.

4.1.1 Channel Inversion in NMOS

Consider the NMOS system as shown in Fig. 4.3 together with respective work functions. The work function of the semiconductor is the sum of the affinity factor, $q\chi$ and $E_C - E_F$ with $q\chi$ denoting the energy required to excite an electron from E_C to the vacuum level. When the three components are brought together in equilibrium contact, the Fermi level should line up and be flat, which necessitates the band bending, as discussed in Chap. 3.

Equilibrium Band Bending: The band bending occurs via the exchange of electrons between the n^+ gate and the p substrate. Specifically, the Fermi level, E_F, of the n^+ gate electrode is higher than that of the substrate; hence, electrons spill over from the gate to substrate. This leaves behind the positive charge sheet at the surface of the gate electrode, which in turn pushes holes in the p substrate away from the interface, thereby exposing acceptor ions uncompensated. Consequently, a dipolar space charge is formed, and the band bends downward as shown in Fig. 4.4. The total band bending is given by the work function difference between the gate electrode and the p substrate, and the bending occurs in both substrate and oxide.

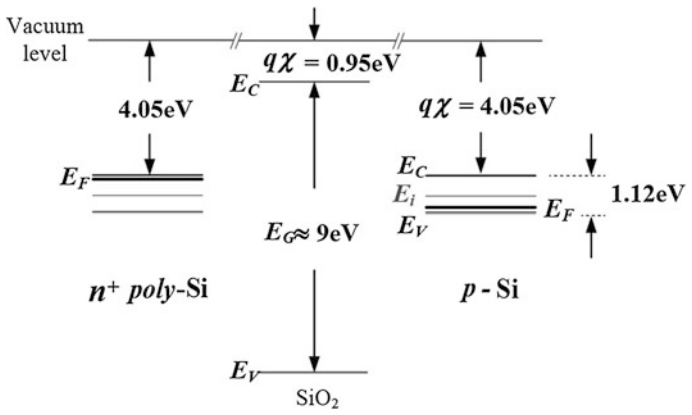

Fig. 4.3 NMOS system: n^+ poly-Si, SiO_2 and silicon p-substrate. Also shown are the affinity factor, $q\chi$ and Fermi levels

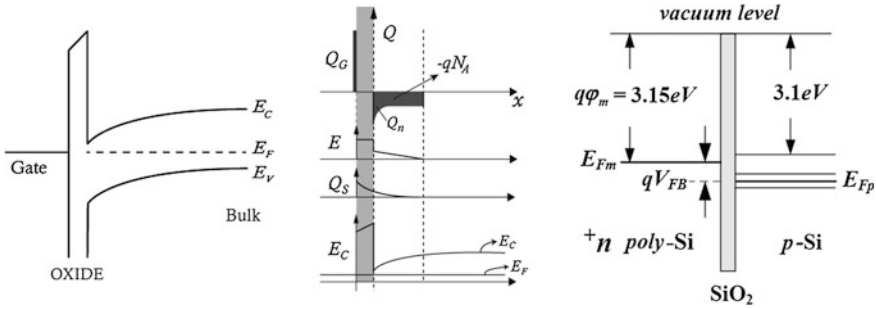

Fig. 4.4 NMOS system in equilibrium contact: The band bending (*left*) and space charge, field and potential underlying the bending (*middle*) are shown. Also sketched is the flat band voltage at which the band flattens out

Flat Band Voltage: The band bending is flattened out with the application of the flat band gate voltage, given by the difference between the work functions of the gate electrode and the p substrate:

$$\begin{aligned} qV_{\mathrm{FBn}} &\equiv q\chi + E_C - E_{\mathrm{Fn}} - (q\chi + E_C - E_{\mathrm{Fp}}) \\ &= E_{\mathrm{Fp}} - E_{\mathrm{Fn}} \end{aligned} \tag{4.4}$$

Since $E_{\mathrm{Fp}} < E_{\mathrm{Fn}}$ in this case $V_{\mathrm{FBn}} < 0$ and with V_{FBn} applied, the positive charge sheet in the gate electrode is annulled, and the space charge disappears together with the band bending.

Surface Charge: By using V_{FBn}, one can introduce the charging voltage, V'_G, which is dropped in both the oxide and the substrate,

$$V'_G \equiv V_G - V_{\mathrm{FBn}} = V_{\mathrm{OX}} + \varphi_S \tag{4.5}$$

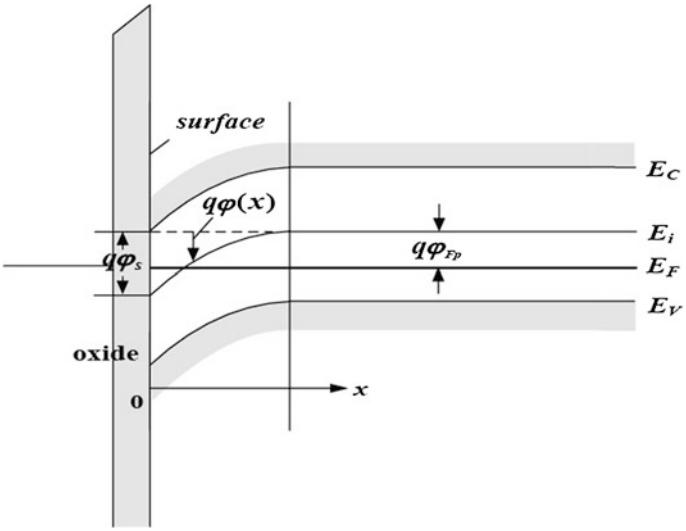

Fig. 4.5 The band bending, $q\varphi(x)$, in the n-channel MOSFET; $q\varphi_S$ and $q\varphi_{FP}$ denote the surface and Fermi potentials, respectively

Here, φ_S is the surface potential, i.e., $\varphi(x = 0)$. With $V'_G(>0)$ on, the band bends down, as indicated in Fig. 4.5. The resulting potential, φ, is found by solving the Poisson equation:

$$\frac{\mathrm{d}^2\varphi(x)}{\mathrm{d}x^2} = -\frac{\rho(x)}{\varepsilon_S},\ \rho(x) = q[(p_p(x) - N_A^- - n_p(x)] \tag{4.6a}$$

where the space charge, ρ, in the substrate is made up of the hole, acceptor ion and electron charges. In the p bulk the charge neutrality holds true, i.e., $p_{p0} = N_A^- + n_{p0}$, but near the surface n, p become x-dependent and is given by

$$p_p(x) = p_{p0}e^{-\beta\varphi(x)},\ n_p(x) = n_{p0}e^{\beta\varphi(x)},\ \beta \equiv q/k_BT \tag{4.6b}$$

(see Fig. 4.5). Hence, upon inserting (4.6b) into (4.6a) together with $N_A^- = p_{p0} - n_{po}$ there results,

$$\frac{\mathrm{d}^2\varphi(x)}{\mathrm{d}x^2} = -\frac{\rho(x)}{\varepsilon_S},\ \rho(x) = q\left[p_{p0}(e^{-\beta\varphi} - 1) - n_{p0}(e^{\beta\varphi} - 1)\right] \tag{4.7}$$

Note here that without the band bending, that is, for $\varphi = 0$ $\rho(x) = 0$, as it should. Also, Eq. (4.7) is strongly nonlinear and is difficult to solve. However, one can perform the first integration by recasting (4.7) by multiplying both sides by $d\varphi$ as

$$\int_0^{-\mathrm{E}} \mathrm{E}\mathrm{d}\mathrm{E} = -\frac{1}{\varepsilon_S}\int_0^{\varphi} \rho(\varphi)\mathrm{d}\varphi,\quad \mathrm{E} \equiv -\frac{\mathrm{d}\varphi}{\mathrm{d}x} \tag{4.8}$$

(problem 4.1). Here the integrations start from edge of the bulk substrate. Hence, upon inserting the expression for ρ in (4.7) into (4.8) and carrying out the integration, one finds

$$\mathrm{E}(x) \equiv -\frac{\mathrm{d}\varphi}{\mathrm{d}x} = \sqrt{2}\frac{k_B T}{q}\frac{1}{L_D}F(\beta\varphi, n_{po}/p_{p0}),\ L_D \equiv \left(\frac{k_B T \varepsilon_S}{q^2 p_{po}}\right)^{1/2} \tag{4.9a}$$

where L_D thus defined is the Debye length and the F function is given by

$$F(\beta\varphi) \equiv \left[(e^{-\beta\varphi} + \beta\varphi - 1) + e^{-2\beta\varphi_{\mathrm{Fp}}}(e^{\beta\varphi} - \beta\varphi - 1)\right]^{1/2},\ e^{-2\beta\varphi_{\mathrm{Fp}}} = n_{p0}/p_{p0} \tag{4.9b}$$

with φ_{Fp} denoting the hole Fermi potential in the p substrate.

The surface field, E_S at $x = 0$, can therefore be specified in terms of the surface potential, φ_S. Once E_S is found, the surface charge, Q_S is obtained from the well known boundary condition at the oxide interface, i.e.,

$$Q_S \equiv -\varepsilon_S \mathrm{E}_S(\varphi_s) = -\varepsilon_S\sqrt{2}\frac{k_B T}{q}\frac{1}{L_D}F(\beta\varphi_s) \tag{4.10}$$

Weak and Strong Inversion: Figure 4.6 shows Q_S as a function of φ_S and the space charge associated. For $V_G = V_{\mathrm{FB}}$, there is no band bending, that is, $\varphi_S = 0$, hence $Q_S = 0$. For $\varphi_S < 0$, the band bends up and the first term in (4.9b) becomes dominant, and holes as the majority carrier are accumulated at the surface. In the depletion region, in which $0 < \varphi_{\mathrm{S}} < \varphi_{\mathrm{Fp}}$, the band bends down, supported mainly by the acceptor ions and Q_S consists of the uncompensated ions. In the weak inversion region, $\varphi_{\mathrm{Fp}} < \varphi_S < 2\varphi_{\mathrm{Fp}}$, electrons begin to populate the interface region. For $\varphi_S \approx 2\varphi_{\mathrm{Fp}}$ $n_s \approx p_{p0}$, and from this point on, the increase of Q_S with increasing φ_S is primarily due to the electron charge induced, and the channel is thus inverted. In this regime, the electrons are concentrated practically at the surface and do not contribute significantly to the band bending, pinning φ_S approximately at a constant level. Therefore, the condition

$$\varphi_S = 2\varphi_{\mathrm{Fp}}$$

represents the onset of the strong inversion.

MOS capacitor: Naturally, the channel is inverted by the gate voltage, V_G, and is due essentially to the capacitive coupling. To analyze the channel inversion in terms of the capacitive charging, consider the charging voltage divided in the gate oxide and the p substrate:

$$V_G' \equiv V_G - V_{\mathrm{FB}} = V_{\mathrm{OX}} + \varphi_S, \quad V_{\mathrm{OX}} \equiv \frac{|Q_S|}{C_{\mathrm{OX}}} \tag{4.11}$$

Since Q_S is a function of φ_S (see 4.10) φ_S is specified in terms of V_G or vice versa. The capacitor, C, connecting the gate electrode and the p substrate consists of the oxide (C_{OX}) and surface (C_S) capacitors connected in series, as shown in Fig. 4.7:

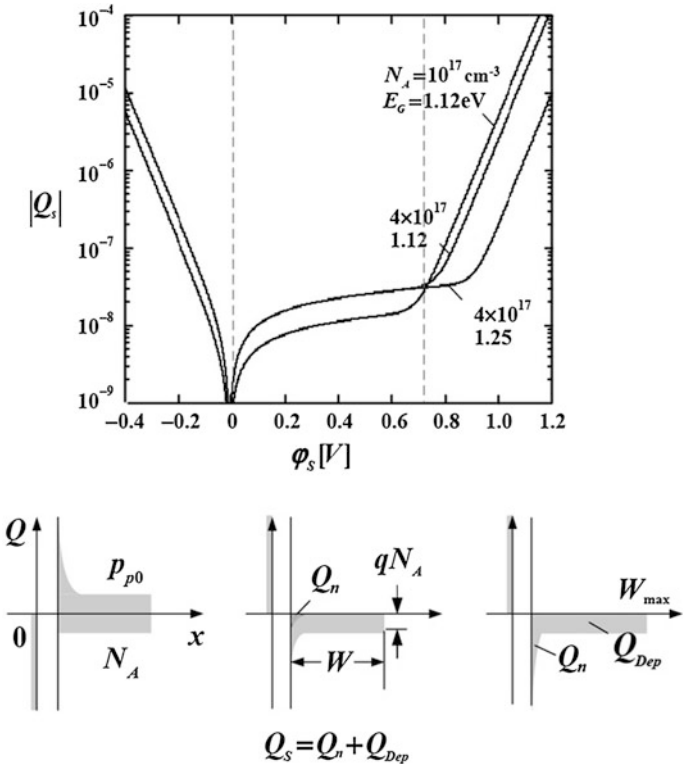

Fig. 4.6 The surface charge, Q_S, is shown versus the surface potential, φ_S, in NMOS with doping level and band gap as parameters. Also shown are the total surface charge, consisting of uncompensated acceptor ions and electrons induced under the gate electrode in accumulation, depletion and inversion regimes

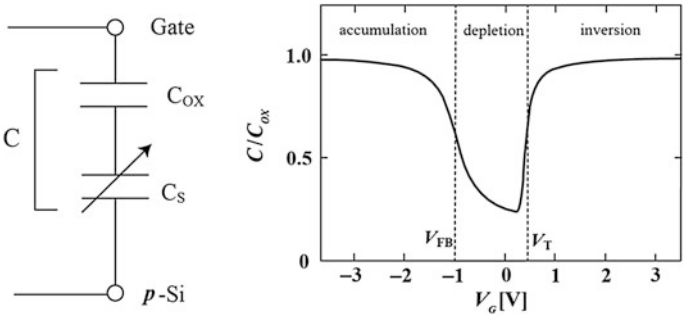

Fig. 4.7 The total capacitor of NMOS, consisting of oxide (C_{OX}) and the surface (C_S) capacitors connected in series (*left*). Also shown is the total capacitance in accumulation, depletion and inversion regions (*right*)

$$\frac{1}{C} = \frac{1}{C_{OX}} + \frac{1}{C_S},\ C_S \equiv \frac{\partial |Q_S|}{\partial \varphi_S} \tag{4.12}$$

Here, C_S accounts for the change in the surface charge, Q_S with φ_S. Since Q_S is a function of φ_S and since φ_S in turn depends on V_G (see 4.11), C_S is a variable capacitance, as shown in Fig. 4.7. In accumulation and inversion regions, $C_S \gg C_{OX}$, (see 4.9, 4.10), so that $C \approx C_{OX}$. In depletion and weak inversion regions, on the other hand, $C_S \ll C_{OX}$ and C is mainly determined by C_S. Also indicated in the figure is the location of the threshold voltage, V_T. Ideally, only the mobile electron charge should be induced by V_G, but the channel inversion requires the surface band bending, which is supported by the ionic charge. Hence, the ionic charge is inseparably coupled to the channel inversion.

4.1.2 I–V Modeling

ON Current: The lumped view of the drain current was discussed in Sect. 4.1 and is now analyzed, based on the channel inversion. The total surface charge consists of the electron (Q_n) and fixed ionic (Q_{DEP}) charges, i.e.,

$$Q_S \equiv -C_{OX} V_{OX} = Q_n + Q_{DEP} \tag{4.13}$$

and terminates the field lines emanating from the positive charge sheet on the gate electrode (see Fig. 4.4). Thus, the key to modeling I–V is to untangle Q_n from Q_{DEP}.

Now, given the substrate doping level, N_A, the depletion charge is given by $Q_{DEP} = -qN_A W_D$ with W_D denoting the depletion depth (see Fig. 4.4). Also, the surface potential, φ_S, is supported by Q_{DEP} and is given by $\varphi_S = qN_A W_D^2/2\varepsilon_S$, in the completely depleted approximation. Therefore, Q_{DEP} can be expressed in terms of φ_S as

$$Q_{DEP} = -qN_A W_D = -(2\varepsilon_S qN_A \varphi_S)^{1/2} \tag{4.14}$$

Hence, upon inserting (4.11) for V_{OX} and (4.14) for Q_{DEP} into (4.13), there results

$$Q_n = -C_{OX}(V_G - V_{FB} - \varphi_S - \gamma \varphi_S^{1/2});\ \gamma_n \equiv (2\varepsilon_S qN_A)^{1/2}/C_{OX} \tag{4.15}$$

Here, the constant, γ_n, is known as the body effect coefficient. Hence, Q_n at the onset of the strong inversion is obtained by replacing φ_S by $2\varphi_{Fp}$ as discussed:

$$Q_n = -C_{OX}(V_{GS} - V_{Tn}),\ V_{Tn} = V_{FB} + 2\varphi_{Fp} + \gamma_n (2\varphi_{Fp})^{1/2} \tag{4.16}$$

In this manner, the capacitive charging of Q_n is quantified.

When the drain voltage, V_{DS}, is turned on, it is distributed in the channel. Thus, at a channel position at y with the channel voltage, $V(y)$ $(0 \le V(y) \le V_D)$ the

condition of the channel inversion, $\varphi_S = 2\varphi_{FP}$ has to be generalized, incorporating the splitting of E_{Fn} and E_{Fp} caused by $V(y)$. However, for the simplicity of discussion, the role of $V(y)$ is taken simply to reduce the effective gate voltage at y and represent $Q_n(y)$ as

$$Q_n(y) = -C_{OX}(V_{GS} - V - V_{Tn}) \tag{4.17}$$

Transistor I–V: With the use of $Q_n(y)$ thus specified, the I–V behavior is derived next. In the channel element from y to $y + dy$, the channel voltage, V, drops by dV, which is given by

$$dV \equiv I_D dR,\ dR \equiv \frac{dy}{W\mu_n|Q_n|} \tag{4.18}$$

where the resistivity has been expressed in terms of Q_n. Therefore, one can recast (4.18) into two integrations, one involving y and the other V, i.e.,

$$\int_0^L I_D dy = \int_0^{V_D} dV \mu_n W|Q_n| \tag{4.19}$$

One can then perform the integrations in (4.19) by using (4.17) for Q_n, and the fact that I_D is constant throughout the channel, obtaining

$$I_D = \frac{W}{L}\mu_n C_{OX}(V_{GS} - V_{Tn} - \frac{1}{2}V_{DS})V_{DS},\ V_{DS} \le V_{GS} - V_{Tn} \tag{4.20a}$$

When the channel is pinched off at the drain end, $Q_n(L) \approx 0$ and the I–V relation given in (4.20a) ceases to be valid. The pinch-off voltage, V_{DSAT} is found by putting Q_n to zero in (4.17), i.e., $V_{DSAT} = V_{GS} - V_{Tn}$ and the saturation current at V_{DSAT} is therefore pinned at

$$I_{DSAT} = \frac{W}{2L}\mu_n C_{OX}(V_{GS} - V_{Tn})^2,\ V_{DSAT} \equiv V_G - V_{Tn} \tag{4.20b}$$

Equation (4.20) agrees with (1.1), and the parameters such as μ, V_{Tn} can further be refined to fit the measured data.

Subthreshold Current: The subthreshold current, I_{SUB}, bridges I_{OFF} and I_{ON} in the V_G range, $0 < V_G < V_T$, or in φ_S range, $0 < \varphi_S < 2\varphi_{Fp}$. In this regime, the second term of Q_S in (4.9b) is dominant and one can Taylor expand Q_S in (4.10) centered around the second term, obtaining

$$Q_S \equiv Q_{DEP} + Q_n \approx -(2qN_A\varepsilon_S\varphi_S)^{1/2}\left(1 + \frac{1}{2}\frac{e^{\beta(\varphi_S - 2\varphi_{Fp})}}{\beta\varphi_S}\right),\ \beta = q/k_BT \tag{4.21}$$

where the Debye length, L_D, in (4.9a) has been spelled out (problem 4.2). Evidently, the first term in (4.21) represents Q_{DEP}, while the second term denotes the surface charge of electron, Q_n.

As clear from (4.21), Q_n is exponentially enhanced with increasing φ_S in this regime of the weak inversion. Also, in the presence of the substrate bias, V_B and channel voltages at the source and drain, V_S and V_{DS} and the expression of Q_n should incorporate the splitting of the quasi-Fermi levels. Thus, one can write

$$Q_n(0) = -qN_AL_D\left(\frac{1}{2\beta\varphi_{SS}}\right)^{1/2} e^{\beta(\varphi_{SS}-2\varphi_{Fp})},\ \varphi_{SS} \equiv \varphi_S - (V_S - V_B) \tag{4.22a}$$

at the source and

$$Q_n(L) = -qN_AL_D\left(\frac{1}{2\beta\varphi_{SD}}\right)^{1/2} e^{\beta(\varphi_{SD}-2\varphi_{Fp})},\ \varphi_{SD} \equiv \varphi_{SS} - V_{DS} \tag{4.22b}$$

at the drain. It is therefore clear from (4.22) that Q_n decreases exponentially from the source to the drain. Hence, I_{SUB} should be driven by diffusion, and one can write

$$\begin{aligned} |I_{SUB}| &\approx WD_n\frac{Q_n(0) - Q_n(L)}{L} \\ &\approx \frac{W}{L}D_nqN_AL_D\left(\frac{1}{2\beta\varphi_{SS}}\right)^{1/2} e^{\beta(\varphi_{SS}-2\varphi_{Fp})}(1 - e^{-\beta V_{DS}}) \end{aligned} \tag{4.23}$$

The subthreshold current thus derived bridges I_{ON} and I_{OFF}.

PMOS I–V Behavior: Naturally, the PMOS I–V modeling can be similarly carried out by interchanging the roles of electrons and holes. One can thus derive the hole surface charge, obtaining

$$Q_p = C_{OX}(|V_{GS}| - V_{Tp} - |V|) \tag{4.24a}$$

where the hole threshold voltage is given by

$$V_{Tp} \equiv V_{FB} + \varphi_{Fp} + \gamma_p\varphi_{Fp}^{1/2},\ \gamma_p \equiv (2\varepsilon_S qN_D)^{1/2}/C_{OX} \tag{4.24b}$$

(problem 4.3). Here, V_{FB} (>0) is the work function difference between the p^+ poly gate and n substrate, and the Fermi potential, φ_{Fn}, of the n substrate and the body coefficient, γ_p, are specified in terms of the donor concentration, N_D. Once Q_p is obtained, the I–V relation can be derived in a manner similar to NMOS.

Quantum Modifications: The channel inversion necessitates the band bending, downward in NMOS and upward in PMOS. As a result, the quantum wells are formed for electrons and holes, respectively, as shown in Fig. 4.8. This means that the inverted electrons and holes are 2D particles spatially confined in the direction normal to the surface, but moving freely on the interface plane. The resulting quantum mechanical modifications have to be taken into account.

First, the electrons and holes reside in the quantized sublevels or subbands, as sketched in Fig. 4.8. These sublevels are characterized by 2D density of states, as detailed in Chap. 1. Therefore, the statistics of the channel inversion should differ from what has been discussed for the 3D bulk MOSFET. Additionally, the

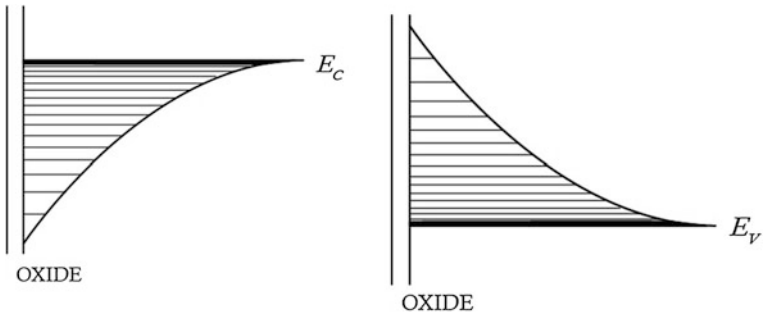

Fig. 4.8 The quantum well and the subbands of electrons (*left*) and holes (*right*), induced at the oxide interface by the respective gate voltages

probability density of the carrier wave function in each sublevel is peaked away from the oxide interface. This suggests that electrons or holes are inverted away from the interface, thereby increasing the effective oxide thickness and reducing the capacitance of the gate dielectric. Also, due to the discrete sublevels formed in the quantum well, the substrate band gap is in effect broadened. Consequently, the quantum modification reduces in essence the efficiency of channel inversion.

4.2 Silicon Nanowire Field Effect Transistor

Overview: The SNWFET holds up a promising potential as a driver of the nanoelectronics and is discussed in this section. In so doing, the well known theory of MOSFET is utilized both as the reference and the general background. In particular, the similarities and differences existing between SNWFET and MOSFET are highlighted, thereby bringing out the features unique to SNWFET. In addition, the ballistic transport operative in short channel SNWFETs is considered. For brevity, the discussion is confined to the n-channel SNWFETs. However, the results obtained can be readily applied to the p-channel FET by interchanging the roles of electrons and holes.

Also, the I–V modeling is focused on intrinsic SNWFET, for simplicity. An interesting feature of such FETs is that both n- and p-channel FETs can be fabricated with the use of the same nanowire by doping the source, drain and gate with donors and acceptors, respectively, as shown in Fig. 4.9. As a consequence, n^+-i and i-n^+ junctions and p^+-i and i-p^+ junctions are built in back to back, respectively. Hence, both types of SNWFETs are unipolar and normally off devices. Figure 4.10 shows a typical I–V and transfer characteristics from SNWFET with the channel length approximately 100 nm long. Clearly, the I–V behavior in such a long channel SNWFET is generally similar to that of MOSFET, which indicates that the physical principles underlying the operation are substantially same in both kinds of long channel FETs.

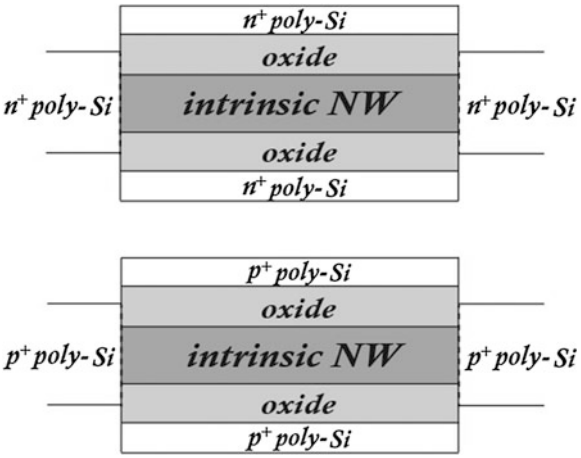

Fig. 4.9 The cross-sectional view of the *n*-type (*top*) and *p*-type (*bottom*) SNWFET

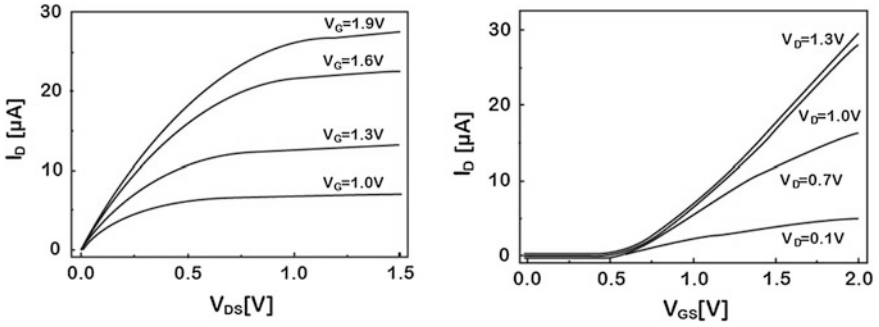

Fig. 4.10 The transistor I–V and transfer characteristics of silicon nanowire field-effect transistor with 100 nm channel length

4.2.1 The n-channel SNWFET

Equilibrium Contact: The equilibrium contact of the n^+ ploy-Si gate—SiO_2—intrinsic nanowire system is essentially the same as in NMOS, aside from the fact that the p substrate in NMOS is replaced by the intrinsic nanowire, as shown in Fig. 4.11. Nevertheless, the electrons are transferred from the gate electrode to the nanowire, again due to the difference in the Fermi levels. Inside the nanowire, the electrons reside in the sublevels therein and are not necessarily concentrated near the oxide interface as in the case of NMOS. This is because of the wave nature of electrons. Specifically, the probability density of electrons in each sublevel can be taken approximately uniform across the cross section of the nanowire, as was discussed in Chap. 1.

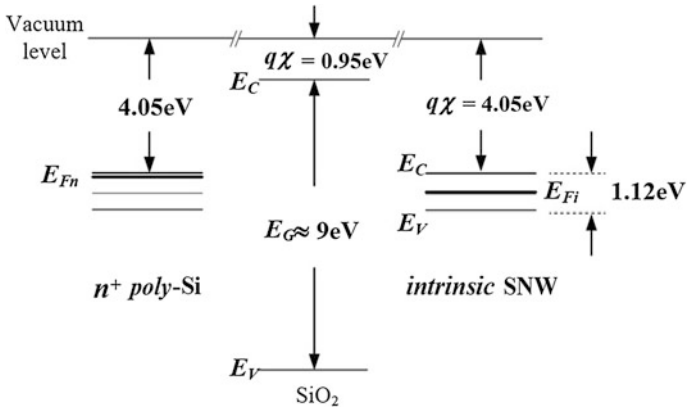

Fig. 4.11 n^+ poly-Si–SiO_2—intrisic Si nanowire system. Also shown are the affinity factors and the Fermi levels

Subband Spectra in Nanowire: The discussion of the channel inversion requires the quantum treatment of the electron states by solving the Schrödinger equation, coupled self-consistently with the Poisson equation in the effective mass approximation. Such an analysis has been carried out in the literature, and the electrons in the nanowire can be taken to move freely along the direction z of the wire, while confined in the x, y plane. The energy eigenequation for bound states is thus given by

$$\left[-\frac{\hbar^2}{2m_n}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)+V\right]u(x,y)=Eu(x,y) \tag{4.25}$$

where V represents the two-dimensional quantum well with a finite barrier height. The eigenequations and eigenvalues in the quantum wire with the rectangular cross section have been analyzed in detail in Chap. 1. The general features of the subbands are well represented by the simple analytical expression obtained for the infinite barrier height:

$$E_n=\sum\nolimits_j E_j n_j^2,\quad E_j=\hbar^2\pi^2/2m_n l_j^2;\quad j=x,\,y;\quad n_j=1,2,\ldots \tag{4.26}$$

Here, n_j is the quantum number, l_j the width of the rectangle in the j direction and E_j the jth sublevel energy

One can likewise analyze the subbands in a cylindrical nanowire with radius R by recasting (4.25) into the cylindrical coordinate frame, obtaining

$$\left[-\frac{\hbar^2}{2m_n}\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right)+V(r)\right]u(r,\varphi)=Eu(r,\varphi) \tag{4.27a}$$

where

$$V(r) = \begin{cases} 0 \text{ for } r \le R \\ V \text{ for } r > R \end{cases} \tag{4.27b}$$

As usual, one can look for the solution in the form

$$u(r, \varphi) \propto e^{\mathrm{in}\varphi} g(r) \tag{4.28}$$

in which case, (4.27a) reduces to the Bessel differential equation for $g(r)$ inside the well:

$$r^2 \frac{\mathrm{d}^2 g(r)}{\mathrm{d}r^2} + r \frac{\mathrm{d}g(r)}{\mathrm{d}r} + (k^2 r^2 - n^2) g(r) = 0,\ k^2 \equiv \frac{2mE}{\hbar^2} \tag{4.29a}$$

The solutions are therefore given in terms of the Bessel functions of the first (J_n) and second (Y_n) kinds. However, Y_n diverges for $r \to 0$ and should be discarded. Thus, the solution is given by

$$g(r) = C_1 J_n(kr),\ r \le R \tag{4.29b}$$

Outside the quantum well, there ensues the modified Bessel differential equation

$$r^2 \frac{\mathrm{d}^2 g(r)}{\mathrm{d}r^2} + r \frac{\mathrm{d}g(r)}{\mathrm{d}r} - (\kappa^2 r^2 + n^2) g(r) = 0,\ \kappa^2 \equiv \frac{2m}{\hbar^2}(V - E) \tag{4.30a}$$

The solutions are therefore given by the modified Bessel functions of the first (I_n) and second (K_n) kinds. However, I_n diverges for $r \to \infty$ and should be discarded and one can write

$$g(r) = C_2 K_n(\kappa r),\ r > R \tag{4.30b}$$

Once the energy eigenfunctions are found, the energy eigenvalues are obtained as usual by imposing the boundary conditions, namely that the eigenfunctions and its first derivatives be continuous at $r = R$. Here, again the general features of the sublevels are well represented by the analytic expressions valid for the infinite barrier height. In this limit, the eigenenergies are given by

$$E_n = E_0(3/2 + n + 2s)^2,\quad E_0 = \hbar^2 \pi^2 / 8 m_n R^2,\quad n, s = 0, 1, 2, \cdots \tag{4.31}$$

Here, n, s denotes the quantum numbers and R the radius of the cylindrical nanowire (problem 4.4).

Figure 4.12 shows the typical sublevel spectra which have been found numerically for the oxide barrier of 3.1 eV in intrinsic silicon nanowires with cross sections different in area and shape. Clearly, a few general trends of the sublevel spectra emerge from the figure. First, the spectra are different between the rectangular and square cross sections, particularly in lower lying sublevels, although the cross-sectional area is the same. This is due to the subband energy levels depending sensitively on the width in each direction of the rectangle see (4.26).

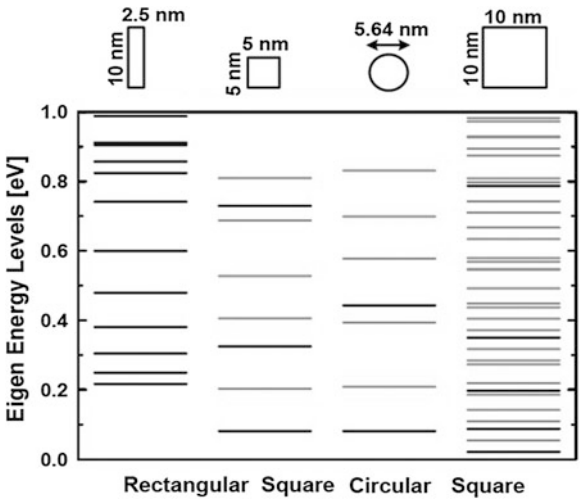

Fig. 4.12 The subband spectra in intrinsic silicon nanowire surrounded by the gate oxide for rectangular, square and circular cross sections

Also, the spectra in the square and circular cross sections with the same area are about the same, in general agreement with the results obtained in (4.26) and (4.31). More important, there is a significant difference in the sublevel spectra between the small- and large-area cross sections, as expected.

Surface Charge and 1D Density: Now that the sublevel spectra have been found, the surface charge of electrons induced is considered next. The first step for obtaining Q_n consists of finding the 1D density of electrons, n_{1D}, which is specified by

$$n_{1D}(\varphi) = \sum_{n=1}^{N} \int_{E_C+E_n}^{E_C+\Delta E_C} d\varepsilon\, g_{1D}(\varepsilon) F_n(\varepsilon)\,,\; g_{1D}(\varepsilon) = (\sqrt{2m_n}/\pi\hbar)/\varepsilon^{1/2} \tag{4.32a}$$

Here, g_{1D} is the 1D density of states [see (1.48)], N the total number of subbands in the quantum well and ΔE_C the conduction band width. The Fermi occupation factor for the electrons in the nth sublevel with energy, E_n is given by

$$F_n(E) = \frac{1}{1+\exp[(E - E_{Fi} - q\varphi)/k_B T]}\,,\; E = \varepsilon + E_C + E_n \tag{4.32b}$$

Here, E_{Fi} is the intrinsic Fermi level of the nanowire, and $q\varphi$ is the bulk band bending. The electron energy, ε, in the nth subband ranges from $E_C + E_n$ to $E_C + \Delta E_C$, and the difference, $E - E_{Fi}$ is reduced by the band bending, $q\varphi$.

The integration of (4.32a) can be carried out either numerically or analytically by expanding the Fermi function. Once $n_{1D}(\varphi)$ is found, the 3D density, n_{3D}, is obtained by dividing $n_{1D}(\varphi)$ with the cross-sectional area, A of the nanowire. The surface field can then be obtained by using the well known relation used in (4.10):

$$E(\varphi) = \sqrt{2}(q/\varepsilon_S)^{1/2}[N(\varphi)]^{1/2} \tag{4.33a}$$

where the integration over the space charge density in (4.8) for obtaining E in the analysis of MOSFET is equivalent to finding the excess electron density induced by the bulk band bending:

$$N(\varphi) = \int_0^{\varphi} [n_{3D}(\varphi) - n_{3D}(0)]\mathrm{d}\varphi, \quad n_{3D}(\varphi) = n_{1D}(\varphi)/A \tag{4.33b}$$

The surface charge, Q_n, can therefore be found in terms $\mathrm{E}(\varphi)$ by again using the well-known boundary condition, (4.10), i.e.,

$$Q_n(\varphi) \equiv -\varepsilon_S \mathrm{E}(\varphi) \tag{4.34}$$

In this manner, $Q_n(\varphi)$ is found in terms of the bulk band bending $q\varphi$ and the properties of the nanowire such as the shape and size of the cross section. Note that Q_n attained in the p substrate of NMOS is specified by the surface potential, $\varphi(x = 0)$. Whereas, Q_n in the nanowire is given in terms of the bulk band bending φ for the reasons discussed.

4.2.1.1 Channel Inversion

Capacitive Coupling: The channel inversion is next addressed to by using Q_n thus found. Consider the charging voltage divided into the gate oxide and the nanowire

$$V_G' \equiv V_G - V_{FB} = V_{OX} + \varphi, \; V_{OX} \equiv \frac{|Q_n(\varphi)|}{C_{OX}} \tag{4.35}$$

where V_{FB} is the flat band voltage associated with the n^+ gate and intrinsic nanowire, as depicted in Fig. 4.11. In intrinsic nanowire, there is no fixed ionic charge and therefore the surface charge, Q_S, consists solely of Q_n. Hence, the surface capacitance, C_S, should differ appreciably from that of NMOS. Nevertheless, the channel inversion via the capacitive coupling can be analyzed by using (4.35) with Q_n given in (4.34).

As in the case of NMOS, Q_n can be found explicitly as a function of V_G from (4.35). Figure 4.13 shows Q_n thus found plotted versus V_G for various kinds of the nanowire cross sections. Also shown in the figure is the $Q_n - V_G$ curve of an NMOS, for comparison. As expected, $Q_n - V_G$ curves in intrinsic nanowire do not exhibit the transition region demarcating the strong and weak inversion. Instead, Q_n in the nanowire increases exponentially in the small V_G regime. In this regime, the electron concentration is low, so that it requires relatively large bulk band bending, $q\varphi$, for the induced electrons to terminate the gate field lines. However, when V_G exceeds a certain value, the electron concentration attains such a level that any further increase in V_G and concomitant increase in the gate field lines can be compensated by electrons with relatively small changes in $q\varphi$. That is to say, $q\varphi$ is approximately pinned while supplying electrons sufficient to terminate the gate field lines. Hence, in this V_G range, Q_n should increase approximately in

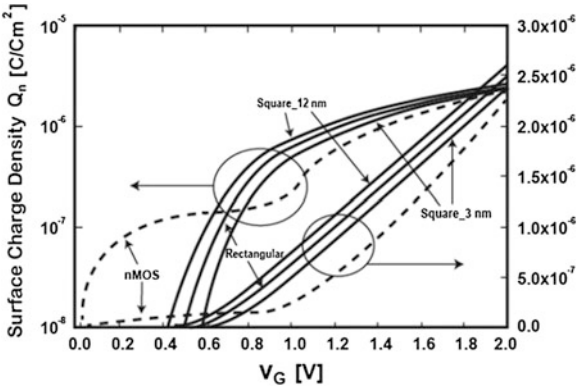

Fig. 4.13 The electron surface charge density is shown versus the gate voltage in nanowire with rectangular (3×12 nm) and square (3×3 nm, 12×12 nm) cross sections. Also plotted for comparison is the electron surface charge in NMOS with the substrate doping of $N_A = 10^{17}$ cm^{-3}

linear fashion with V_G just as in the case of NMOS after the onset of the strong inversion. Consequently, Q_n versus V_G curves naturally divide into the subthreshold and linear regimes as in the case of NMOS.

Moreover, it is clear from Fig. 4.13 that more surface charge, Q_n, is induced at given V_G in the larger cross section, and the threshold voltage therein is reduced. Here, V_T is defined as the value of V_G inducing a specified level of the drain current at given V_D, a procedure often used in the I–V characterization. According to this definition, V_T is simply the value of V_G which inverts a specified level of Q_n. The larger Q_n and smaller V_T with increasing cross-sectional area are consistent with the sublevel spectra shown in Fig. 4.12. As pointed out, the sublevels in larger cross section are more densely distributed at the energy level lower than those in smaller cross sections. Hence, more electrons should be induced for given V_G.

4.2.1.2 I–V Behavior in Long Channel *n*-type SNWFET

Overview: The modeling of I–V behavior in SNWFET has been extensively investigated from various standpoints. Understandably, the emphasis has been placed on the carrier scattering and transport. The carriers in the long channel are transported mainly by the drift-diffusion, while the ballistic transport is prevalent in the short channel. Thus, the carrier transport in FETs consists in general of the mixture of the drift–diffusion and ballistic transports. The I–V behavior in SNWFET is discussed from various standpoints in a few sections to follow, fusing together the two modes of the transport. In addition, the Landauer formulation of the ballistic transistor as applied to SNWFET is briefly touched upon. The discussion starts out with the long channel FET.

Naturally, a key quantity involved in the I–V modeling is the surface charge, Q_n, which is induced by V_G. As detailed in the preceding section, Q_n can be taken to increase linearly with the gate overdrive, and one can thus write

$$Q_n = C_{\text{eff}}[V_G - V_{\text{Tn}} - V] \tag{4.36}$$

where V is the channel voltage at y and C_{eff} the total capacitance of the nanowire. The capacitance, C_{eff}, is somewhat smaller than C_{OX} above V_T for the reasons as follows. As well known, C_{eff} consists of the oxide (C_{OX}) and surface (C_S) capacitances, connected in series, i.e., $1/C_{\text{eff}} = 1/C_{\text{OX}} + 1/C_S$. In NMOS, the electrons are induced practically at the oxide interface. Consequently, C_S is much larger than C_{OX}, and therefore C_T is practically identical to C_{OX} above V_T. In nanowire FET, however, the distribution of the induced electrons is nearly uniform across the cross section of the nanowire. This is due to the wave nature of electrons as pointed out earlier. Hence, the ratio of C_S with respect to C_{OX} is not as large as in NMOS; hence, C_{eff} is somewhat smaller than C_{OX}. However, the ratio C_S/C_{OX} is still large enough to keep C_{eff} practically constant at a level slightly below C_{OX}.

In addition, V_T in nanowire depends primarily on the geometry of the cross section, while in NMOS, V_T is determined mainly by the doping level of the substrate and the bandgap of the substrate. Obviously, this difference arises from the fact that in nanowire Q_n sensitively depends on the shape and size of the cross section, while in NMOS, Q_n is dictated by the doping level of the donor atoms and the band gap of the substrate. Moreover, in long channel nanowire FETs, the carriers are generally transported via the drift and diffusion. Hence, for modeling the I–V behavior in SNWFET, one can follow the same steps as used in NMOS, obtaining (4.20). In this case, the tools developed for SPICE model for fitting the I–V data can also be used.

4.2.1.3 I–V Behavior in Short Channel *n*-type SNWFET

Overview: An attractive feature of FETs is its scalability down to nano regimes. In such short channel FETs, the ballistic transport is prevalent. In ultrascaled MOSFETs, for example, the carrier transport has been taken up by the ballistic component by as much as 50 %. The ballistic efficiency in SNWFETs is believed to be comparable to or even higher than that of MOSFET. It is therefore important to consider the ballistic nanowire FETs.

Figure 4.14 shows the typical band diagram along the channel under bias. The maximum point of the band is located near the source end, the height of which is

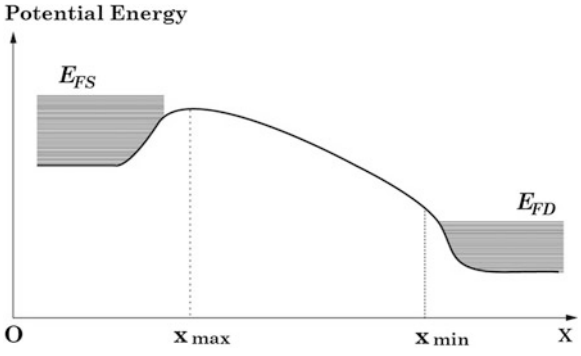

Fig. 4.14 The schematic view of the band bending in FETs is shown under the drain bias

determined by the n^+-i band bending at the source junction and also the voltages applied at the gate and the drain. The band diagram points specifically to the fact that the drain current is contributed by electrons injected from the n^+ source electrode and transported down the channel, subject, however, to the backscattering. Naturally, the height of the band maximum is reduced with increasing V_{GS}, so that more electrons are injected into the channel. In the following, the I–V behavior in short channel FETs is discussed from various standpoints, using the theories developed recently.

1. *One-Flux Scattering Theory*:

To facilitate the discussion, the one-flux scattering theory by Lundstrom is introduced first, in which the saturated drain current is given by

$$I_{\mathrm{DSAT}} = Q_{\mathrm{nLS}} v_{\mathrm{eff}} \tag{4.37a}$$

Here, Q_{nLS} is the line charge induced at the source end and is given in terms of the total capacitance C_{eff} per unit area and the channel width, W_{eff}, of the nanowire as

$$Q_{\mathrm{nLS}} = C_{\mathrm{eff}} W_{\mathrm{eff}} (V_G - V_{\mathrm{Tn}}) \tag{4.37b}$$

And v_{eff} is the effective velocity with which the electrons are transported down the channel subject to the backscattering and is given by

$$v_{\mathrm{eff}} = v_{\mathrm{inj}} \eta,\ \eta \equiv \left(\frac{1 - r_c}{1 + r_c} \right) \tag{4.37c}$$

where r_c is the backscattering coefficient and v_{inj} is the injection velocity of electrons.

Evidently, the two representations of I_{DSAT} given in (4.37) for SNWFET and (4.3), for MOSFET, respectively, are of the same format and characterize I_{DSAT} in similar contexts. However, there are a few differences existing between the two. Note that Q_{nLS} in (4.3) is the average line charge under the gate electrode, given by the average gate overdrive, $(V_{\mathrm{GS}} - V_{\mathrm{Tn}})/2$. Whereas Q_{nLS} in (4.37b) is the line charge induced specifically at the source end of the channel. In ballistic FETs, the line charge at the source end is a key parameter dictating the drain current. Additionally, the velocity, v_D, appearing in (4.3) is the drift velocity as characterized by the small field mobility, μ_n, while v_{eff} in (4.37c) is dictated primarily by v_{inj}.

It is interesting that η appearing in (4.37c) also represents the ballistic efficiency. In the limit of small r_c at which $\eta \cong 1$, nearly all of the injected electrons traverse the channel via the ballistic transport with v_{inj}. By the same token, in the opposite limit of $\eta \cong 0$, the transport is dominated by the drift–diffusion instead. For the general case of r_c ranging from 0 to 1, η is essentially an indicator showing the degree of mixing between the two modes of transport.

Obviously r_c depends on the channel length, L, and therefore η is also dependent on L. The dependency is taken into account in the one-flux scattering theory as follows. In thermal equilibrium or in the limit of small longitudinal channel field the backscattering coefficient as a function of L is given by

$$r_{c0} = \frac{L}{L + \lambda} \tag{4.38}$$

where λ is the mean free path. With λ thus introduced, it follows from (4.37) and (4.38) that η gauges the mode of carrier transport at a given channel length, L. For $\lambda \ll L$, for example, $\eta \cong 0$; hence, carriers are driven by the drift and diffusion, as it should. For $\lambda \gg L$, on the other hand, $\eta \cong 1$ and the ballistic transport becomes prevalent, as expected.

Under the bias, the backscattering coefficient, r_c, can be treated in such a way that the two competing modes of transport are naturally fused to represent the effective carrier transport. Now, the coefficient, r_c, is given by

$$r_c = \frac{l}{l + \lambda} \tag{4.39}$$

In this representation, L has been replaced by the critical length, l, over which the electron gains the kinetic energy equal to the thermal energy, $k_B T$.

With the use of l introduced in this manner, the expression of I_{DSAT} given in (4.37a) is naturally generalized to incorporate both the drift–diffusion and ballistic transport and that as a function of L. This can be shown as follows. When V_{DS} is applied, the band bends down from the source to the drain, as sketched in Fig. 4.14, and the injected electrons roll down the potential hill in the channel. Once the electron surpasses the distance l, Price observation assures that the electron can be taken to proceed to the drain, irrespective of the backscattering. This is equivalent to the electron traversing the channel essentially via the ballistic transport.

Next, to be specific, the critical distance, l, is to be expressed in terms of the thermal energy, $k_B T$, and the longitudinal channel field, $\mathrm{E}(0^+)$, at the band maximum:

$$q\mathrm{E}(0^+)l \equiv k_B T \tag{4.40a}$$

Also the mean free path, λ, is specified approximately via the thermal velocity, v_T, and the mean collision time, τ_n, or equivalently the low field mobility, μ_n:

$$\lambda = v_T \tau_n = v_T (m_n \mu_n / q) \tag{4.40b}$$

where (2.25) has been used to replace τ_n by μ_n.
By inserting (4.40) and (4.39) into (4.37) one can easily obtain (problem 4.5),

$$I_{\mathrm{DSAT}} = Q_{\mathrm{nLS}} \frac{1}{\frac{1}{v_T} + \frac{1}{\mu_n \mathrm{E}(0^+)}} \tag{4.41}$$

In this representation of I_{DSAT}, both the drift–diffusion and ballistic transports are naturally fused in. Additionally, the relative importance of two competing modes of transport versus the channel length, L, is also accounted for. Thus, I_{DSAT} derived in (4.41) is applicable to a variety of short-channel FETs. It should be noted that since the longitudinal channel field scales with L, *i.e.,* $\mathrm{E}(0^+) \propto 1/L$, $\mu_n \mathrm{E}(0^+) \gg v_T$ in short channel FETs, and therefore the FET operates in ballistic mode with v_T serving as the saturation velocity. By the same token, in a long channel FET, $\mu_n \mathrm{E}(0^+) \ll v_T$, in which case the electrons traverse the channel via the usual drift velocity, v_D. In this latter limit, (4.41) converges to the SPICE I–V model given in (4.3).

However, the two line charges in (4.37b) and (4.3) are different in contents, as pointed out. In Eq. (4.3), Q_{nL} is the line charge averaged over the channel. The averaging procedure is required in the drift–diffusion formulation, since the transit time of electrons across the channel is long, so that the quasi-equilibrium settles in under the gate electrode. In this case, Q_n at the channel position, y, should be governed by the local gate overdrive. In the scattering theory formulation, on the other hand, I_{DS} is primarily determined by Q_{nLS} at the source end, sweeping across channel with a high ballistic efficiency. The two differing approaches should be duly taken into account for characterizing I_{DS} in SNWFETs with varying channel lengths.

2. *Apparent Mobility Model*:

The discussion of the apparent mobility model by Shur is in order at this point. According to the model, the total mobility consists of the usual low field mobility, μ_n, and the ballistic mobility, μ_{ball}, connected in series

$$\frac{1}{\mu} = \frac{1}{\mu_n} + \frac{1}{\mu_{\mathrm{ball}}},\ \mu_{\mathrm{ball}} = \kappa L \tag{4.42}$$

where μ_{ball} is taken commensurate with the channel length, L. The total mobility, μ, thus introduced is called the apparent mobility. One can then express I_{DSAT} with the use of μ as,

$$I_{\mathrm{DSAT}} = Q_{\mathrm{nLS}} \mu \mathrm{E}(0^+),\ \mu = \frac{\mu_n \mu_{\mathrm{ball}}}{\mu_n + \mu_{\mathrm{ball}}} \tag{4.43}$$

In the limit of short channel length, $\mu_{ball} \ll \mu_n$, and μ is reduced to μ_{ball}. In this case, the ballistic transport prevails with $\mu_{\mathrm{ball}} E(0^+)$ providing the saturation velocity. In the other limit, μ is reduced to μ_n, and I_{DSAT} is therefore driven by the drift–diffusion. It is therefore clear that the one-flux scattering theory and the apparent mobility model lead essentially to the same representation of I_{DSAT}.

3. *Landauer Formulation of Ballistic FET*:

The continued scaling down of the semiconductor devices has pushed the device dimensions into the mesoscopic regime in between the atomic and the microscopic regimes. In such short channel devices, the mean free path, typically 50 nm long, cannot be taken much shorter than the channel length. By the same token, the carrier relaxation and the coherence lengths that are closely linked to the

mean free path cannot be regarded much smaller than the channel length as well. Therefore, the wave nature of electrons and the mesoscopic scattering must be taken into consideration in modeling the I–V behavior. It thus behooves to consider the Landauer formulation for treating the ballistic FETs.

In the Landauer formulation, the drain current is taken due primarily to the net flux of electrons from the source to the drain via the ballistic transport. The drain current is thus given by

$$I_{\mathrm{DS}} = \frac{2q}{h}\sum_i \int_{E_i}^{E_u} dE[F(E, E_{\mathrm{FS}}) - F(E, E_{\mathrm{FD}})]T_i(E) \tag{4.44a}$$

Here the first term represents the electron flux from the source to the drain, while the second term denotes the flux in the reverse direction. Also, i is the summation index of the subbands with the energy level, E_i, in the well, and E_u is the upper limit of the integration given by the conduction band width, i.e., $E_u = E_C + \Delta E_C$. Naturally, the Fermi functions

$$F(E, E_{Fj}) = \frac{1}{1 + \exp\left(\frac{E - E_{Fj}}{k_B T}\right)},\ j = S,\ D \tag{4.44b}$$

are characterized by the Fermi levels at the source and drain ends, respectively. Under a drain bias, the two quasi-Fermi levels should split by the amount, qV_{DS}, i.e., $E_{\mathrm{FD}} = E_{\mathrm{FS}} - qV_{\mathrm{DS}}$, as detailed in Chap. 3. The factor T_i denotes the transport coefficient of electrons in the ith subband, and I_{DS} is thus contributed separately by electrons in each subband. In short channel FETs, one can put $T_i(E) \approx 1$ since the band bending in the channel is usually gradual, so that the backscattering therein is to be neglected.

The gist of (4.44) can be seen by considering the case of small drain voltage. One can then Taylor expand the Fermi function at the drain, retaining only the first expansion term and write

$$F(E, E_{\mathrm{FS}}) - F(E, E_{\mathrm{FS}} - qV_{\mathrm{DS}}) \approx -\frac{\partial F(E, E_{\mathrm{FS}})}{\partial E} qV_{\mathrm{DS}} \tag{4.45a}$$

Thus, the difference between the two fluxes is shown to be commensurate to the first derivative of F, which is well approximated by the delta function,

$$\frac{\partial F(E, E_{\mathrm{FS}})}{\partial E} \approx \delta(E - E_{\mathrm{FS}}) \tag{4.45b}$$

Upon inserting (4.45) into (4.44) there results

$$I_{\mathrm{DS}} = G\sum_i g_i V_{\mathrm{DS}},\ G \equiv \frac{2q^2}{h} \tag{4.46}$$

In this manner, I_{DS} is naturally shown to be specified by the quantum conductance, G, and the sum of the separate contributions from all subbands, including the degeneracy, g_i, therein.

For an arbitrary V_{DS}, I_{DS} can also be specified as follows. For this purpose, one may first introduce a new variable of integration

$$\eta = E/k_BT$$

and compact the expression of I_{DS} in (4.44a) as

$$I_{DS} = G\left(\frac{k_BT}{q}\right)\tilde{M} \tag{4.47a}$$

where the form factor, $\tilde{M}$, reads as

$$\tilde{M} = \sum_i \int_{\eta_i}^{\eta_{max}} d\eta [\frac{1}{1+e^{(\eta-\eta_{FS})}} - \frac{1}{1+e^{(\eta-\eta_{FS}+qV_{DS}/k_BT)}}],\ \eta_{FS} = E_{FS}/k_BT \tag{4.47b}$$

where, $\eta_i = E_i/k_BT$ and $\eta_{max} = (E_C + \Delta E_C)/k_BT$. Obviously, to evaluate $\tilde{M}$, it is essential to find the relative location of E_{FS} with respect to say E_C under a given V_{GS}.

The difference, $E_C - E_{FS}$, varying as a function of V_{GS} is clearly illustrated in Fig. 4.15. In the figure are shown the band diagrams of the n^+ gate, gate oxide and intrinsic nanowire both in equilibrium and under a bias. In equilibrium, the band

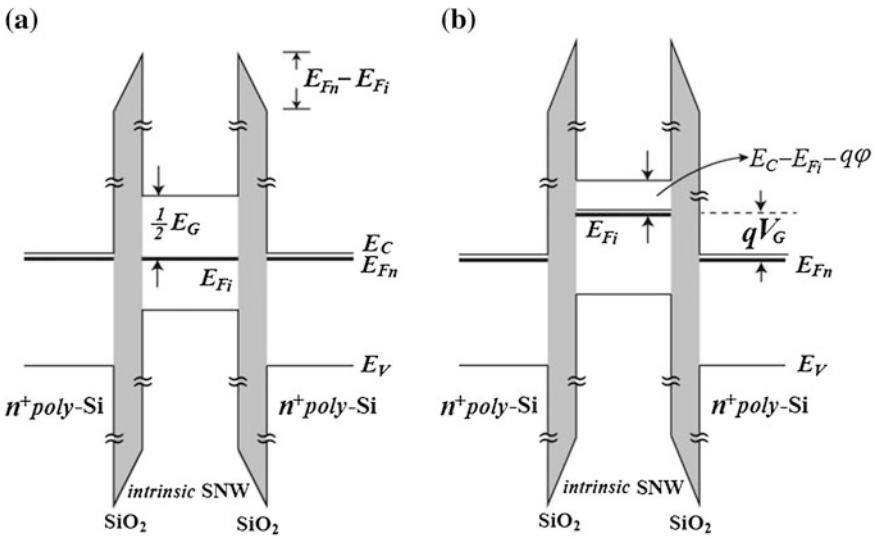

Fig. 4.15 The energy band diagram of n^+ poly gate, silicon dioxide and intrinsic silicon nanowire, in equilibrium (*left*) and under a positive gate bias (*right*). Here, $q\varphi$ denotes the bulk band bending induced by the gate voltage

bending should occur mainly in the gate oxide by an amount $E_{\mathrm{Fn}} - E_{\mathrm{Fi}}$ so that the Fermi level lines up and be flat (Fig. 4.15a). Under a gate bias, V_{GS}, the band in the n^+ gate electrode is further lowered and the two Fermi levels split (Fig. 4.15b). Then, V_{GS} is dropped in the gate oxide and also in the nanowire, inducing the bulk band bending, $q\varphi$, therein. As a result, $E_{\mathrm{Fn}} - E_{\mathrm{Fi}}$ is reduced from its equilibrium value by the amount $q\varphi$, and one can therefore write

$$E_C - E_{\mathrm{FS}} = E_C - E_{\mathrm{Fi}} - q\varphi \tag{4.48}$$

Now, $q\varphi$ is in turn specified in terms of V_{GS} by means of (4.35), as discussed in detail already. Thus, the difference, $E_C - E_{\mathrm{FS}}$ can be found explicitly as a function of V_{GS}, and one can therefore evaluate $\tilde{M}$ and find I_{DS} as a function of V_{GS} and V_{DS}. It has therefore become clear that the I–V modeling in Landauer formulation consists of solving the coupled equations, (4.47) and (4.35).

The resulting features of I–V curves in ballistic FETs can also be surmised on a general ground with the use of (4.35) and Fig. 4.13. For small V_{GS}, hence small φ, Q_n is at the low level, and most of the charging voltage is taken up by φ, the reasons of which have already been detailed in connection with Fig. 4.13. As a consequence, $E_C - E_{\mathrm{FS}}$ shrinks rapidly, enhancing I_{DS} exponentially. This small V_{GS} regime should therefore correspond to the subthreshold regime. Once φ exceeds a certain value, the growth of φ slows down, but I_{DS} still increases with V_{DS} for given V_{GS}, reproducing thereby the triode regime. Beyond a certain value of V_{DS}, the flux of electrons from the drain to the source becomes negligible, rendering I_{DS} insensitive to further increase of V_{DS} and the device enters the saturation mode.

In this chapter, the I–V behavior of SNWFET has been discussed in comparison with theory of MOSFET, thereby bringing out the features unique to SNWFET. The I–V behavior has also been discussed in long- and short-channel SNWFETs. The I–V modeling in short-channel FETs have been discussed from various standpoints by using the one-flux scattering theory, the apparent mobility model and the Landauer formulation.

Problems

4.1 (a) Multiply both side of the Poissson equation given in (4.7) and prove that the left hand side of (4.7) reduces to the left hand side of (4.8).

(b) Perform the integrations in (4.8) and derive (4.9).
(c) Use the result obtained for discussing the threshold voltages as a function of the doping level and the band gap.
(d) Repeat the corresponding analysis for PMOS.

4.2 Derive (4.21) by Taylor expanding Q_S given in (4.10).

4.3 The I–V modeling in *p*-type FETs can be done by following the similar steps as used in *n*-type FETs and also by replacing the roles of electrons with those of holes.

(a) Derive the surface charge and the threshold voltage of the PMOS given in (4.24).
(b) Derive I–V relation and compare the results with that of NMOS.

4.4 Starting from the energy eigenequation in the cylindrical nanowire given in (4.27) derive (4.29a) and (4.29b). Also derive the expression of the energy eigenvalue given in (4.31), which is valid for the infinite barrier potential.

4.5 (a) Derive the saturated drain current given in (4.41) by inserting the expressions of the critical distance, l and the mean free path, λ given in (4.40).

(b) Discuss the mode of carrier transport as a function of the channel length, L, in the silicon nanowire FET with the cross-sectional areas of 5×5 nm and 10×10 nm and for L ranging from 10 to 100 nm. You can use the results obtained for the case of the infinite barrier height for the simplicity of discussion.

Suggested Reading

1. Kim, D. M. (2010). *Introductory quantum mechanics for semiconductor nanotechnology*. USA: John Wiley-VCH.
2. Lundstrom, M. S. (2006). *Fundamentals of carrier transport* (2nd ed.,). Cambridge University Press.
3. Datta, S. (1995). *Electronic transport in mesoscopic structures*. Cambridge University Press.
4. Lundstrom, M. S., & Ren, Z. (2002). Essential physics of carrier transport in nanoscale MOSFETs. IEEE Transactions on Electron Devices, ED *49*, 133–141.
5. Lundstrom, M. S. (1997). Elementary scattering theory of the Si MOSFET. *IEEE Electron Device Letters, 18*, 361–363.
6. Natori, K. (2008). Compact modeling of ballistic nanowire MOSFETs. *IEEE Transactions Electron Devices,* ED *55*, 2877–2885.