Chapter 22
MOS Devices

22.1 Introduction

The mathematical model of semiconductor devices, derived in Chap. 19, is applied here to the description of two fundamental devices of the insulated-gate type: the MIS capacitor, whose most important implementation is the MOS capacitor, and the IGFET, whose most important implementation is the MOSFET. Both devices can be realized starting from either a *p*-doped or an *n*-doped substrate; only the first realization is illustrated here: the extension of the theory to the other one is immediate. The analysis of the MOS capacitor is carried out using the simple example of a one-dimensional device in steady state, with the hypotheses of non-degeneracy and complete ionization, that lend themselves to an analytical treatment. Observing that in a steady-state condition the device is in equilibrium, the theory needs the solution of Poisson's equation only. From the solution of the latter, the device's capacitance is calculated, followed by a number of other important relations, that are useful in the subsequent treatment of the MOSFET. The theory of the MOSFET is then tackled in two dimensions and in steady-state conditions, first deriving a general expression for the channel current that holds in the case of a well-formed channel. The calculation is then completed by introducing the gradual-channel approximation: the differential conductances are derived first, followed by the expression of the drain current as a function of the applied voltages. A further simplification leads to the linear–parabolic model, which is widely used in the semiqualitative analyses of circuits. The complements address the solution of the Poisson equation in the channel when a non-equilibrium condition holds, to provide a formal proof of the relation between the surface and quasi-Fermi potentials used in the gradual-channel approximation; finally, a few phenomena that are not accounted for by the gradual-channel approximation are discussed, and the neglect of the dependence on position of the average carrier mobility is justified.

M. Rudan, *Physics of Semiconductor Devices,*
DOI 10.1007/978-1-4939-1151-6_22

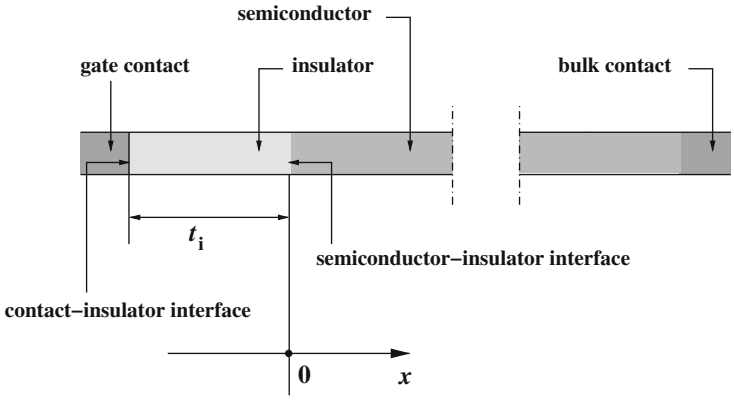

Fig. 22.1 Cross section of a metal–insulator–semiconductor capacitor. The thickness of the insulator layer is not realistic: in real devices the layer is much thinner than the contacts

22.2 Metal–Insulator–Semiconductor Capacitor

The Metal–Insulator–Semiconductor (MIS) capacitor is a fundamental device, that constitutes the basis for the field-effect transistors used in the fabrication of integrated circuits. The device has also extensively been used for studying the properties of semiconductor surfaces [103]. A one-dimensional version of it is sketched in Fig. 22.1: the structure is fabricated by depositing or thermally growing (Chap. 24) an insulator layer over a semiconductor substrate. The fabrication process must obtain an electrically clean interface; in fact, the number of localized electronic states at the interface must be kept to a minimum to avoid carrier trapping–detrapping processes. The contact deposited onto the insulator is called *gate contact*, the other one is called *bulk contact*.

In the standard silicon technology, the insulator is obtained by thermally growing silicon dioxide (Sect. 24.2). For this reason, the thickness of the insulator is indicated in the following with t_{ox} instead of the generic symbol t_i used in Fig. 22.1; by the same token, the insulator's permittivity is indicated with ε_{ox}, and the device is called MOS capacitor. In the last years, the progressive scaling down in the size of semiconductor devices has brought the insulator thickness to the range of nanometers. A smaller thickness has the advantage of providing a larger capacitance; however, it may eventually lead to dielectric breakdown and leakage by quantum tunneling. Silicon dioxide, which has been used as a gate insulator for decades, is being replaced in advanced devices with insulating layers made of materials having a larger permittivity (*high-k dielectrics*). Such layers are obtained by deposition (Sect. 24.5). Still with reference to the silicon technology, the conductive layers are made of metals, heavily-doped polycrystalline silicon, or metal silicides; here they will be indicated with the generic term "metal".

Like in the case of *p–n* junctions, the theory of the MOS capacitor is carried out with reference to a simplified structure, where the device is one dimensional

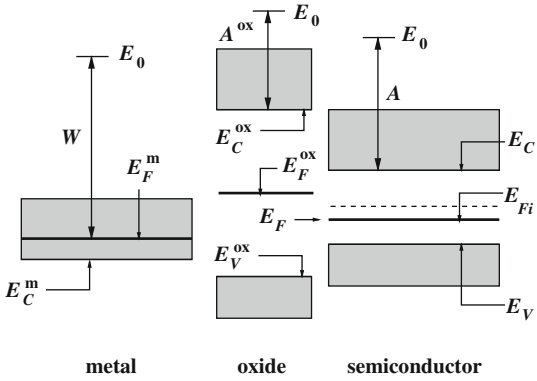

Fig. 22.2 The three materials forming the MOS capacitor shown separately. The symbols' meaning is illustrated in the text

and aligned with the x axis; in this case the semiconductor–insulator interface is a plane normal to x and, as shown in Fig. 22.1, its position is made to coincide with the reference's origin. A constant dopant concentration is present in the semiconductor region; to further simplify the analytical approach one also assumes that the conditions of non-degeneracy and complete ionization hold.

To describe the functioning of the device it is necessary to consider the fact that, in a region where the important electric phenomena occur, different materials are brought into an intimate contact. With reference to Fig. 22.2, the three materials (gate metal, oxide, and semiconductor) are initially considered separate from each other, and in the equilibrium condition. The left part of the figure shows the conduction band of the metal, with E_C^{m} the band's lower edge and E_F^{m} the metal's Fermi level. Due to the form of the Fermi-Dirac statistics, the probability that an electron's energy exceeds E_F^{m} is small; remembering the discussion of Sect. 7.2, the minimum energy necessary for an electron to transit from the metal into vacuum is the metal work function $W = E_0 - E_F^m$, with E_0 the vacuum level (left part of Fig. 22.2). For an insulator or a semiconductor, the electrons with maximum energy belong to states in the vicinity of the lower edge of the conduction band, E_C^{ox} (center) or E_C (right); in this case the minimum energy necessary for transiting into vacuum is the *electron affinity* $A^{\mathrm{ox}} = E_0 - E_C^{\mathrm{ox}}$ or $A = E_0 - E_C$, respectively. The semiconductor considered in the figure is uniformly doped of the p type, as shown by the fact that the Fermi level is below the intrinsic Fermi level E_{Fi} (Sect. 18.4.2). However, the analysis carried out here applies also to a semiconductor of the n type.

When the materials are brought into contact, they form a single, non-uniform system; as a consequence, in the equilibrium condition the Fermi levels must align with each other. On the other hand the vacuum levels must align as well, and the bands must adapt to compensate for a possible charge redistribution that occurs when the materials contact each other. The situation is similar to that represented in Fig. 21.3 for the p–n junction. The values of the parameters $W, A, \ldots$ selected for drawing Fig. 22.2 fulfill the relation

$$W - A = E_C - E_F. \tag{22.1}$$

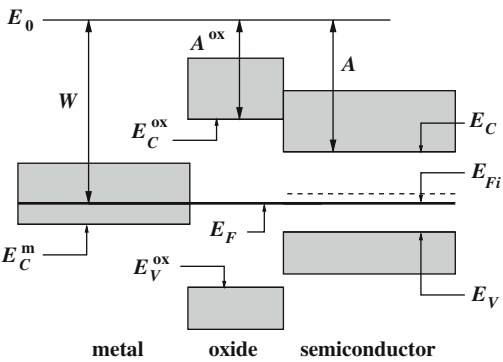

Fig. 22.3 The three materials forming the MOS capacitor after being brought into contact. The symbols' meaning is illustrated in the text

As a consequence, there is no need for the bands to modify their shape; as shown in Fig. 22.3, which represents the situation after the materials have been brought into contact, the bands do not deform. The condition where the semiconductor's bands are everywhere parallel to the Fermi level is indicated with *flat-band condition*.[1] It is important to remark that condition (22.1) seldom occurs in realistic cases; however, as shown below, the case $W - A \neq E_C - E_F$ is easily incorporated into the analysis.

When the bulk contact is considered, Figs. 22.2 and 22.3 must be completed by adding the band structure of the contact's material to the right of that of the semiconductor. Assuming that the gate and bulk contacts are made of the same material, the structure to be added is identical to that already present on the left part of the figures. In the interior of each material, due to spatial uniformity, the electric potential is piecewise constant, thus the electric field is zero.

Consider now the case where a voltage V_G is applied between the gate and bulk contacts; the voltage reference is such that $V_G > 0$ when the electric potential of the gate contact is larger than that of the bulk contact, and vice versa. In steady-state conditions, the insulator prevents a current from flowing through the device; therefore, during the transient consequent to the application of V_G, the electric charge adjusts itself to the new boundary conditions. At the end of the transient the device is again in an equilibrium condition, while the form of the bands is different from that of Fig. 22.3. Similarly, the electric potential in the oxide and semiconductor is not constant any longer; its form is found by solving the Poisson equation in each region.

22.2.1 Surface Potential

The solution of the Poisson equation in the semiconductor region follows the same pattern as for the *p*–*n* junction (Sect. 21.2.1). Here a uniformly *p*-doped region

[1] The form of (22.1) is general enough to hold for both *p*- and *n*-type semiconductors.

is considered; its extension along the x axis is large, so that, far away from the semiconductor–insulator interface, the semiconductor behaves as if it were isolated. This fixes the carrier-equilibrium concentrations in the bulk; the asymptotic value of the electric potential is set to zero, $\varphi(+\infty) = 0$ whence, remembering that the non-degeneracy and complete-ionization conditions hold, it is

$$p^{(0)} = p(+\infty) = p_{p0} \simeq N_A, \qquad n^{(0)} = n(+\infty) = n_{p0} \simeq \frac{n_i^2}{N_A}. \tag{22.2}$$

The Poisson equation in the semiconductor then reads

$$u'' = \frac{1}{L_A^2}\, A(u), \qquad A(u) = \frac{n_i^2}{N_A^2}\left[\exp(u) - 1\right] + 1 - \exp(-u), \tag{22.3}$$

with L_A the Debye length for the holes defined in (21.12). The normalized charge density $A(u)$ has the same sign as u (compare with Sect. 21.2.1). Multiplying by u' both sides of the first equation in (22.3), transforming its left hand side into $u''\,u' = (1/2)\left[(u')^2\right]'$, and integrating from $x \geq 0$ to $+\infty$, yields

$$\left(u'\right)^2 = \frac{2}{L_A^2}\, B(u), \qquad B(u) = \frac{n_i^2}{N_A^2}\left[\exp(u) - 1 - u\right] + u + \exp(-u) - 1. \tag{22.4}$$

Following the same reasoning as for (21.10), one finds that B is non negative and u monotonic. However, in contrast with the case of the p–n junction, where the sign of u' is positive due to the boundary condition at $x \to -\infty$, here u may either increase or decrease monotonically; in fact, the sign of u' is fixed by the boundary condition V_G, which in turn may be either positive or negative. In conclusion, one finds

$$u' = \pm\frac{\sqrt{2}}{L_A}\, F(u), \qquad F(u) = \sqrt{\frac{n_i^2}{N_A^2}\left[\exp(u) - 1 - u\right] + u + \exp(-u) - 1}, \tag{22.5}$$

where the sign must be found on a case-by-case basis. Separating (22.5), finally yields

$$\frac{\mathrm{d}u}{F(u)} = \pm\frac{\sqrt{2}}{L_A}\,\mathrm{d}x, \tag{22.6}$$

which must be solved numerically because it has no analytical solution.[2] Much information, however, is gained directly from (22.5), without the need of integrating

[2] The numerical evaluation from (22.6) of the inverse relation $x = x(u)$ is straightforward, though. Letting $\xi = \sqrt{2}\,x/L_A$, $\xi_0 = 0$, $u_0 = u_s$, $F_0 = F(u_0)$, it is $u_{k+1} = u_k \mp \delta u$, $F_{k+1} = F(u_{k+1})$, and $\xi_{k+1} = \xi_k + (1/F_k + 1/F_{k+1})\,\delta u/2$, with $k = 0, 1, \ldots$.

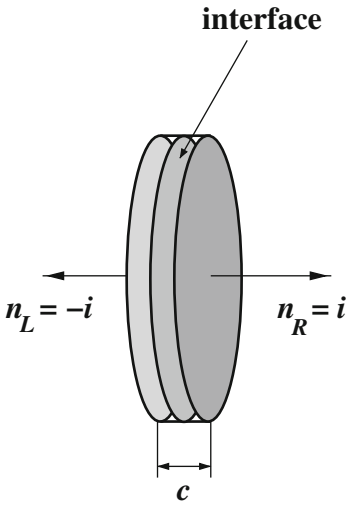

Fig. 22.4 The cylinder used to calculate the relation between electric displacement and charge per unit area across an interface

(22.6). To this purpose, one notes that the electric potential is continuous at the semiconductor–oxide interface; in fact, the normalized charge density in the semiconductor (22.3) has no charge layers in it, hence it can not contribute to a double charge layer at the interface. As a consequence, one can adopt the same symbol φ_s for the electric potential at $x = 0$, without distinguishing between the two sides of the interface; φ_s is called *surface potential*, whilst $u_s = q\varphi_s/(k_B\,T)$ is the *normalized surface potential*. In contrast, the electric field is discontinuous at the same interface; for this reason, one defines

$$u_s' = \lim_{x\to 0^+} \frac{\mathrm{d}u}{\mathrm{d}x}, \qquad E_s = -\frac{k_B\,T}{q}\,u_s', \qquad E_{\mathrm{ox}} = -\lim_{x\to 0^-} \frac{\mathrm{d}\varphi}{\mathrm{d}x}. \tag{22.7}$$

The relation between E_s and E_{ox} is found by considering a cylinder of thickness c placed across the semiconductor–oxide interface, such that the unit vector $\mathbf{n}_R$ normal to the right face is parallel to the unit vector $\mathbf{i}$ of the x axis, whereas the unit vector $\mathbf{n}_L$ normal to the left face is antiparallel to $\mathbf{i}$ (Fig. 22.4). Letting A_e be the common area of the two faces, the total charge within the cylinder is $A_e\,Q$, with Q the charge per unit area. Integrating $\mathrm{div}\mathbf{D} = \rho$ over the cylinder's volume and using (A.23) yields

$$A_e\,Q = \int_{A_e} \mathbf{D}\cdot\mathbf{n}\;\mathrm{d}A_e = A_e\,[D_L\,\mathbf{i}\cdot(-\mathbf{i}) + D_R\,\mathbf{i}\cdot\mathbf{i}] = A_e\,(D_R - D_L)\,, \tag{22.8}$$

with D_R (D_L) the electric displacement on the right (left) face. From $\mathbf{D} = \varepsilon\,\mathbf{E}$ one then finds $Q = \varepsilon_{\mathrm{sc}}\,E_R - \varepsilon_{\mathrm{ox}}\,E_L$. It has been shown above that there are no charge layers on the semiconductor's side; as for the oxide layer, in principle it should be free of charges, although some contaminants may be present (Sect. 24.1). Here it is assumed that the oxide is free of charge; in conclusion, letting the cylinder's thickness c go to zero, one obtains $Q \to 0$, whence, using the limits (22.7),

$$\varepsilon_{\mathrm{sc}}\,E_s = \varepsilon_{\mathrm{ox}}\,E_{\mathrm{ox}}. \tag{22.9}$$

To find $E_{\rm ox}$ one observes that in a one-dimensional medium free of charge the electric potential is linear, whence $E_{\rm ox}$ is given by the negative potential drop across the oxide divided by the oxide thickness $t_{\rm ox}$. To complete the analysis it is then necessary to consider the interface between oxide and gate metal.

In the interior of the metal the electric potential is uniform; its value with respect to the metal of the bulk contact is V_G. However, in the solution of the Poisson equation within the semiconductor, the asymptotic condition $\varphi(+\infty) = 0$ has been chosen, which holds inside the semiconductor region; the surface potential φ_s is referred to such a zero as well. It follows that V_G and φ_s are referred to two different zeros. Remembering the discussion carried out in Sect. 21.2.2, the difference between the external zero (namely, that within the bulk contact) and the internal zero (given by the asymptotic condition) is the built-in potential Φ_{mp} between the bulk contact and the p-type semiconductor; thus, the gate voltage referred to the internal zero is $V'_G = V_G - \Phi_{mp}$. Also, the electric potential is continuous across the interface between the oxide and the gate metal, because no double layer is present there. In contrast, as $E = 0$ within the metal while $E_{\rm ox} \neq 0$, the electric field is generally discontinuous; in fact, a charge layer of density

$$\rho_{\rm m} = Q_{\rm m}\, \delta(x + t_{\rm ox}^-), \tag{22.10}$$

with $Q_{\rm m}$ the charge per unit area of the metal, builds up at the gate–metal's surface. In conclusion, the electric field within the oxide reads

$$E_{\rm ox} = \frac{V'_G - \varphi_s}{t_{\rm ox}}. \tag{22.11}$$

22.2.2 Relation Between Surface Potential and Gate Voltage

Combining (22.5), (22.7), (22.9), and (22.11) one finds

$$C_{\rm ox}\,\left(V'_G - \varphi_s\right) = \mp\varepsilon_{\rm sc}\frac{k_B\,T}{q}\,\frac{\sqrt{2}}{L_A}\,F\big(\varphi_s\big), \qquad C_{\rm ox} = \frac{\varepsilon_{\rm ox}}{t_{\rm ox}}, \tag{22.12}$$

where $C_{\rm ox}$ is the *oxide capacitance per unit area*, and $F(\varphi_s)$ is obtained by replacing u with $q\varphi_s/(k_B\,T)$ in the second relation of (22.5). The left hand side of (22.12) is the charge per unit area $Q_{\rm m}$ in the gate metal. This is easily found by considering the same cylinder as above, this time placed across the metal–oxide interface. Integrating $\mathrm{div}\mathbf{D} = \rho$ over the cylinder's volume, using (22.10), and observing that $D_L = 0$ because the metal's interior is equipotential, yields

$$A_e\,Q_{\rm m} = \int_{A_e} \mathbf{D}\cdot\mathbf{n}\;\mathrm{d}A_e = A_e\,D_R = A_e\,C_{\rm ox}\,\left(V'_G - \varphi_s\right). \tag{22.13}$$

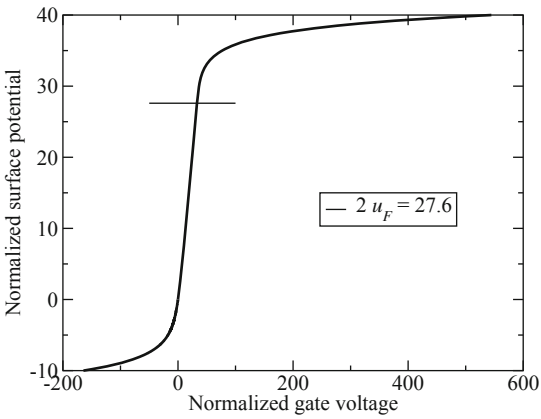

Fig. 22.5 Normalized surface potential u_s in an MOS capacitor with a p-type substrate ($N_A = 10^{16}$ cm^{-3}), as a function of the normalized gate voltage u'_G

Due to the global charge neutrality, the following relation holds between the charge per unit area in the gate metal, $Q_{\rm m}$, and that within the semiconductor, $Q_{\rm sc}$:

$$Q_{\rm m} + Q_{\rm sc} = 0, \qquad Q_{\rm sc} = \int_0^\infty q\big(p - n - N_A\big)\,{\rm d}x = -C_{\rm ox}\,\big(V'_G - \varphi_s\big)\,. \tag{22.14}$$

In conclusion, (22.12) provides the relation between surface potential and gate voltage. When $\varphi_s = 0$, the electric potential vanishes everywhere in the semiconductor, namely, $V'_G = 0$ corresponds to the flat-band condition. When $V'_G > 0$, the charge in the gate metal is positive; as a consequence, the left hand side of (22.12) is positive as well, whence $V'_G > \varphi_s$ and the positive sign must be chosen at the right hand side. The opposite happens when $V'_G < 0$. An example of the $\varphi_s = \varphi_s(V'_G)$ relation is given in Fig. 22.5, showing the normalized surface potential u_s in an MOS capacitor as a function of the normalized gate voltage $u'_G = qV'_G/(k_B\,T)$. The semiconductor's doping is of the p type with $N_A = 10^{16}$ cm^{-3}, corresponding to $2\,u_F = 2\,q\varphi_F/\big(k_B\,T\big) = \log\,(p_{p0}/n_{p0}) \simeq 27.6$.

The $\varphi_s = \varphi_s\big(V'_G\big)$ relation lends itself to identifying different functioning regimes of the MOS capacitor. This identification can be carried out more accurately basing upon the values of the electron and hole concentrations at the semiconductor surface, $n_s = n(x = 0)$ and $p_s = p(x = 0)$. In the non-degenerate conditions considered here, the expressions of the surface concentrations read

$$n_s = n_{p0}\,\exp\big(u_s\big) = n_i\,\exp\big(u_s - u_F\big),\;\; p_s = p_{p0}\,\exp\big(-u_s\big) = n_i\,\exp\big(u_F - u_s\big). \tag{22.15}$$

Depending on the value of u_s, several functioning regimes are identified, which are listed in Table 22.1. The regimes' designations are given by comparing the carrier concentrations at the surface with the intrinsic and asymptotic ones. When $u_s < 0$ the majority-carrier surface concentration (holes, in the example used here) exceeds the asymptotic one; the regime is called *accumulation*. When $u_s = 0$, both

Table 22.1 MOS capacitor, *p* substrate—functioning regimes

Norm. surface potential	Surface concentrations	Designation
$u_s < 0$	$n_s < n_{p0} < n_i < p_{p0} < p_s$	Accumulation
$u_s = 0$	$n_s = n_{p0} < n_i < p_{p0} = p_s$	Flat band
$0 < u_s < u_F$	$n_{p0} < n_s < n_i < p_s < p_{p0}$	Depletion
$u_s = u_F$	$n_{p0} < n_s = n_i = p_s < p_{p0}$	Mid gap
$u_F < u_s < 2\,u_F$	$n_{p0} < p_s < n_i < n_s < p_{p0}$	Weak inversion
$u_s = 2\,u_F$	$n_{p0} = p_s < n_i < n_s = p_{p0}$	Threshold
$2\,u_F < u_s$	$p_s < n_{p0} < n_i < p_{p0} < n_s$	Strong inversion

majority- and minority-carrier concentrations equal the corresponding asymptotic concentrations everywhere, and the already-mentioned *flat-band condition* holds. For $0 < u_s < u_F$, the majority-carrier concentration is smaller than the asymptotic one, while the minority-carrier concentration is smaller than the intrinsic one. By continuity, the majority-carrier concentrations is smaller than the asymptotic one not only at the semiconductor's surface, but also in a finite region of width x_d, which is therefore depleted from carriers; for this reason, the condition $0 < u_s < u_F$ is called *depletion regime*, and x_d is called *depletion width*.[3] When $u_s = u_F$, both majority- and minority-carrier concentrations at the surface equal the intrinsic concentration; remembering that in an intrinsic semiconductor the Fermi level practically coincides with the gap's midpoint (Sect. 18.3), this regime is called *mid gap*. When $u_F < u_s < 2\,u_F$, the minority-carrier concentration at the surface exceeds that of the majority carriers; however, it is still lower than the asymptotic concentration of the majority carriers: the regime is called *weak inversion*. When $u_s = 2\,u_F$, the surface concentration of the minority carriers equals the asymptotic concentration of the majority carriers, and vice versa; the regime is called *threshold of the strong inversion*, or simply *threshold*. Finally, when $u_s > 2\,u_F$, the minority-carrier concentration at the surface exceeds the asymptotic concentration of the majority carriers, and the regime is called *strong inversion*. In Fig. 22.5 the normalized surface potential at threshold, $2\,u_F$, is marked by the horizontal bar; one notes that in the strong-inversion regime the surface potential rapidly saturates as the gate voltage increases.

The form of the electric potential and charge density is shown in Fig. 22.6 for the accumulation regime. The upper part of the figure shows the charge density, which is schematically represented by a negative charge layer at the metal–oxide interface and by the thicker, positive layer at the semiconductor oxide interface. The lower part of the figure shows the electric potential along with the band structure of the semiconductor; note that two different vertical axes are used, in such a way that energy increases upwards and the electric potential increases downwards. The

[3] The depletion width x_d is conceptually the same thing as the extension l_p of the space-charge region on the *p* side of a metallurgical junction (Sect. 21.2.2). A different symbol is used to avoid confusion in the analysis of the MOSFET (Sect. 22.6).

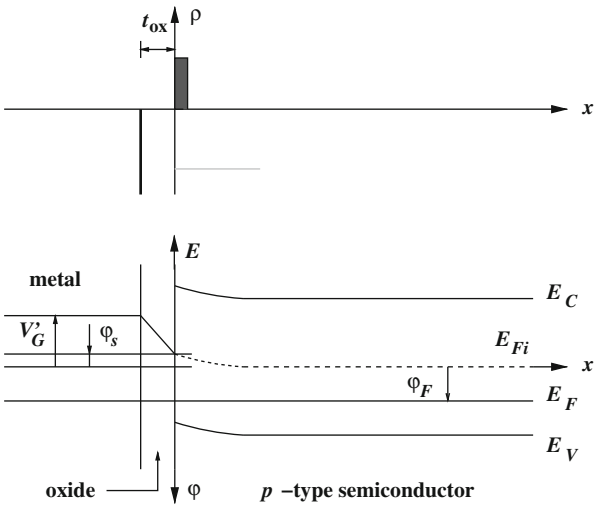

Fig. 22.6 Schematic representation of the charge density and electric potential in a *p*-substrate MOS capacitor in the accumulation regime

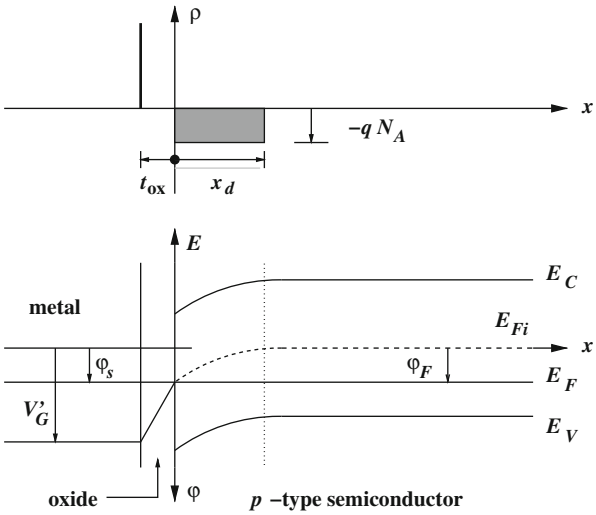

Fig. 22.7 Schematic representation of the charge density and electric potential in a *p*-substrate MOS capacitor in the mid-gap condition

zero of the electric potential coincides with the horizontal part of the dashed line (in fact, here it is $V'_G < \varphi_s < 0$). The mid-gap condition, $V'_G > 0$, $\varphi_s = \varphi_F$, is illustrated in Fig. 22.7, whose general description is similar to that of Fig. 22.6; here the charge layer on the gate metal is positive, and balances the negative charge of the semiconductor. Due to the depletion that occurs in the region adjacent to the semiconductor–oxide interface, the charge density is dominated by the contribution from the negative acceptor ions, $\rho \simeq -qN_A$. In the figure, the charge density of the semiconductor is schematically indicated by the shaded area, that corresponds to a charge per unit area equal to $-qN_A\,x_d$. Finally, Fig. 22.8 shows the form of the

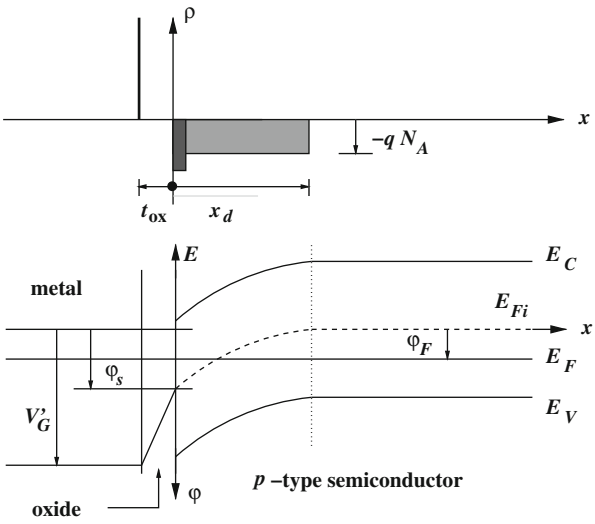

Fig. 22.8 Schematic representation of the charge density and electric potential in a *p*-substrate MOS capacitor at threshold

electric potential and charge density for the threshold condition, $V'_G > 0$, $\varphi_s = 2\,\varphi_F$. Again, the general description is similar to that of Figs. 22.6 and 22.7; here, there are two contributions to the negative charge of the semiconductor: the first one comes from the contribution of the negative acceptor ions, whose charge density is schematically indicated by the shaded area of width x_d (note that, due to the larger value of V_G, the depletion width is larger than that of the mid-gap condition). The second contribution to the semiconductor's charge is due to the electrons, whose concentrations at the interface or near it is sufficiently large to be significant; they form a negative layer, called *inversion layer*, whose width, albeit larger than that of the positive layer located at the metal–oxide interface,[4] is much smaller than x_d.

Numerical solutions of the semiconductor-device model show that, with the exception of the accumulation regime, the semiconductor region of a uniformly-doped MOS capacitor can be partitioned into a space-charge and a quasi-neutral region; the quasi-neutral region behaves as an isolated, uniform semiconductor, whereas in the volume of the space-charge[5] region the charge density is essentially dominated by the ionized dopants. In the threshold and strong-inversion regimes, the layer of mobile charges near the semiconductor–oxide interface gives a significant contribution, which must be accounted for; it can approximated as a charge layer at the interface. Considering the range $\varphi_s > 0$ only, namely, excluding the accumulation regime for

[4] The width of the region where the charged layer at the metal–oxide interface is significant is of the order of 1 nm. That of an inversion layer is of the order of 5 nm; an example is given below, with reference to Fig. 22.13.

[5] With reference to MOS devices, the space-charge region is also called *depleted region*.

the p-substrate MOS capacitor, the charge per unit area in the semiconductor is found to be

$$Q_{\rm sc} = \int_0^\infty \rho \, {\rm d}x \simeq \int_0^{x_d} \rho \, {\rm d}x \simeq -q \int_0^{x_d} \left(n + N_A\right) {\rm d}x = Q_i + Q_b, \qquad (22.16)$$

where the first approximation is due to neglecting the charge of the quasi-neutral region, the second one to neglecting the holes in the space-charge region. Quantities $Q_i, Q_b < 0$ are, respectively, the integral of $-qn$ and $-qN_A$; they are called *inversion charge per unit area* and *bulk charge per unit area*.

22.3 Capacitance of the MOS Structure

The capacitance per unit area of the MOS structure is given by[6]

$$C = \frac{{\rm d}Q_m}{{\rm d}V_G} = \frac{{\rm d}Q_m}{{\rm d}V'_G}. \qquad (22.17)$$

Combining (22.17) with (22.12) and (22.14) yields

$$\frac{1}{C} = \frac{{\rm d}V'_G}{{\rm d}Q_m} = \frac{{\rm d}\left(V'_G - \varphi_s\right) + {\rm d}\varphi_s}{{\rm d}Q_m} = \frac{1}{C_{\rm ox}} + \frac{{\rm d}\varphi_s}{{\rm d}\left(-Q_{\rm sc}\right)}. \qquad (22.18)$$

The above is recast in a more compact form by defining the *semiconductor capacitance per unit area*

$$C_{\rm sc} = -\frac{{\rm d}Q_{\rm sc}}{{\rm d}\varphi_s} = -\frac{q}{k_B\,T}\,\frac{{\rm d}Q_{\rm sc}}{{\rm d}u_s} = \pm\frac{\sqrt{2}\,\varepsilon_{\rm sc}}{L_A}\,\frac{{\rm d}F}{{\rm d}u_s} > 0, \qquad (22.19)$$

where the positive (negative) sign holds for $u_s > 0$ $\left(u_s < 0\right)$. In conclusion, the capacitance is the series of the oxide and semiconductor capacitances:

$$\frac{1}{C} = \frac{1}{C_{\rm ox}} + \frac{1}{C_{\rm sc}}. \qquad (22.20)$$

In (22.20) it is $C_{\rm ox} = {\rm const}$ while $C_{\rm sc}$ has a rather complicate dependence on u_s. However, basing on the second equation in (22.5), one may investigate the limiting cases of (22.20). For this, using $\exp\left(-2\,u_F\right) = n_i^2/N_A^2$, one finds for the asymptotic behavior of F in a p-substrate device,

$$F \simeq \exp\left(u_s/2 - u_F\right), \quad u_s \gg 1\,; \qquad F \simeq \exp\left(-u_s/2\right), \quad u_s \ll -1. \qquad (22.21)$$

[6] Like in the case of the depletion capacitance of the p–n junction (Sect. 21.4), definition $C = {\rm d}Q_{\rm m}/{\rm d}V_G$ is coherent with the choice of the voltage reference described in Sect. 22.2. The units of C are $[C] = {\rm F\,m^{-2}}$.

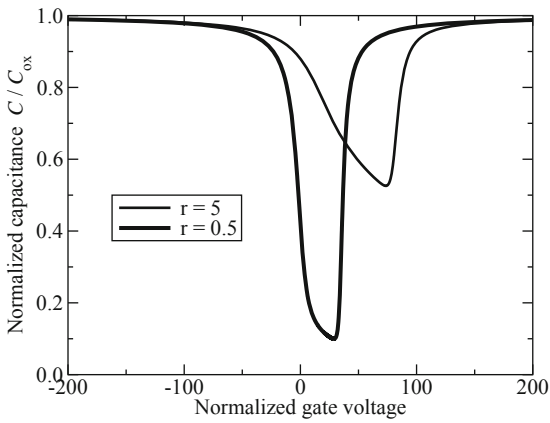

Fig. 22.9 Normalized capacitance $C/C_{\rm ox}$ as a function of the normalized gate voltage u'_G, in a p-substrate MOS capacitor with $N_A = 10^{16}$ cm^{-3}, for different values of $r = \varepsilon_{\rm sc}\, t_{\rm ox}/\left(\varepsilon_{\rm ox}\,\sqrt{2}\,L_A\right)$. The details of the calculations are in Problem 22.7.2

When, instead, it is $|u_s| \ll 1$, expanding the exponentials yields $\exp\left(\pm u_s\right) \simeq 1 \pm u_s + u_s^2/2$, whence, observing that $\exp\left(-2\,u_F\right) \ll 1$,

$$F^2 \simeq \frac{1}{2}\left[1 + \exp\left(-2\,u_F\right)\right] u_s^2 \simeq \frac{1}{2}\,u_s^2, \qquad F \simeq \pm\frac{u_s}{\sqrt{2}}. \tag{22.22}$$

Then, from (22.19) the asymptotic values of $C_{\rm sc}$ are found to be

$$C_{\rm sc} \simeq \frac{1}{\sqrt{2}}\,\frac{\varepsilon_{\rm sc}}{L_A}\,\exp\left(u_s/2 - u_F\right), \qquad u_s \gg 1, \tag{22.23}$$

$$C_{\rm sc} \simeq \frac{1}{\sqrt{2}}\,\frac{\varepsilon_{\rm sc}}{L_A}\,\exp\left(-u_s/2\right), \qquad u_s \ll -1. \tag{22.24}$$

Both limits correspond to the same asymptotic value of the capacitance per unit area,

$$C = \frac{C_{\rm ox}\,C_{\rm sc}}{C_{\rm ox} + C_{\rm sc}} \simeq C_{\rm ox}, \qquad |u_s| \gg 1. \tag{22.25}$$

Near the origin, instead, one finds

$$C_{\rm sc} \simeq \frac{\varepsilon_{\rm sc}}{L_A}, \qquad |u_s| \ll 1. \tag{22.26}$$

The limit of C for $u_s \to 0$ is called *flat-band capacitance per unit area*; from (22.26) one finds

$$C \simeq C_{FB} = \frac{C_{\rm ox}}{1 + C_{\rm ox}\,L_A/\varepsilon_{\rm sc}} < C_{\rm ox}, \qquad |u_s| \ll 1. \tag{22.27}$$

Examples of capacitance's calculations are shown in Fig. 22.9.

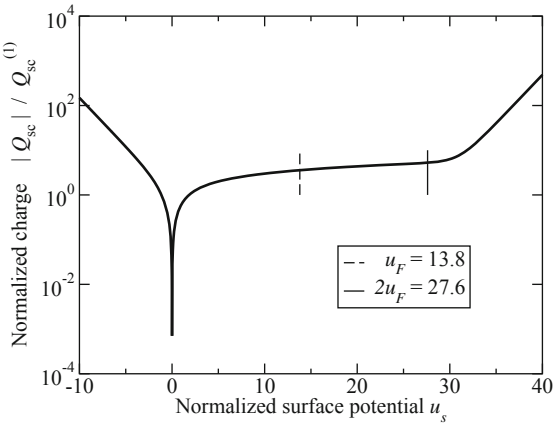

Fig. 22.10 Normalized charge per unit area as a function of the normalized surface potential, in a p-substrate MOS capacitor with $N_A = 10^{16}$ cm^{-3}

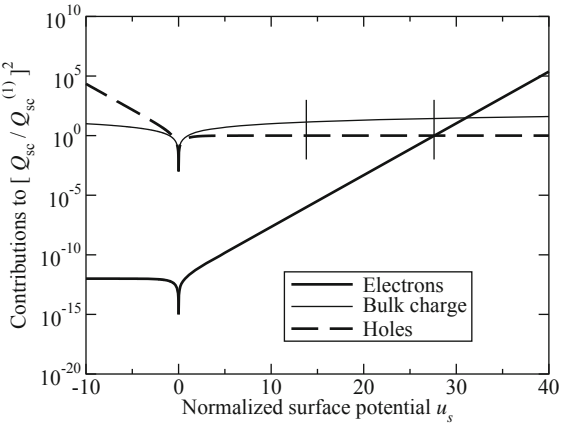

Fig. 22.11 Individual contributions of electrons, holes, and bulk charge to $F^2 = [Q_{sc}/Q_{sc}^{(1)}]^2$, as a function of the normalized surface potential u_s, in a p-substrate MOS capacitor with $N_A = 10^{16}$ cm^{-3}

22.4 Simplified Expression of the Inversion Charge

To the purpose of applying some results of the MOS capacitor's theory to the analysis of MOSFETs, it is convenient to determine a simplified form of the inversion layer's charge, that holds in all the functioning regimes with the exception of accumulation. For this, one starts from the expression of the semiconductor charge per unit area which, combining (22.12) and (22.13), reads

$$Q_{sc} = \pm Q_{sc}^{(1)}\ F\big(u_s\big), \qquad Q_{sc}^{(1)} = \varepsilon_{sc}\frac{k_B\,T}{q}\ \frac{\sqrt{2}}{L_A}, \tag{22.28}$$

where the negative (positive) sign must be chosen when $u_s > 0$ ($u_s < 0$), and $Q_{sc}^{(1)}$ is the value of Q_{sc} corresponding to $F = 1$. The relation $Q_{sc} = Q_{sc}\big(u_s\big)$ is shown in normalized units in Fig. 22.10. In turn, Fig. 22.11 shows, still in normalized form, the individual contributions of electrons, holes, and bulk charge to $F^2 = \big[Q_{sc}/Q_{sc}^{(1)}\big]^2$;

such contributions are, respectively, $\exp\left(-2u_F\right)\left[\exp\left(u_s\right)-1\right]$, $\exp\left(-u_s\right)-1$, and $\left[1-\exp\left(-2u_F\right)\right]u_s$. The contribution of holes dominates for $u_s < 0$, that of the bulk charge dominates for $0 < u_s < 2u_F$ and, finally, that of the electrons dominates for $u_s > 2u_F$.

When accumulation is excluded, in a p-substrate capacitor one must take $\varphi_s > 0$. The approximate dependence of F on the normalized potential is easily worked out from (22.5), whose limiting case in the depletion and weak-inversion regimes is

$$F \simeq \sqrt{u_s}, \qquad 0 < u_s < 2u_F. \tag{22.29}$$

Introducing (22.29) into (22.28) yields, for $0 < u_s < 2\,u_F$,

$$Q_{\rm sc} \simeq -\frac{2\,\varepsilon_{\rm sc}\,k_B\,T}{qL_A}\sqrt{\frac{q\varphi_s}{k_B\,T}} = -C_{\rm ox}\,\gamma\,\sqrt{\varphi_s}, \qquad \gamma = \frac{\sqrt{2\,\varepsilon_{\rm sc}\,qp_{p0}}}{C_{\rm ox}}, \tag{22.30}$$

where the expression (21.12) of the Debye length L_A has been used.[7] It is interesting to note that a relation identical to (22.30) is obtained using the full-depletion and ASCE approximations (Sect. 21.4); in fact, letting $\rho = -qN_A$ for $0 < x < x_d$ and $\rho = 0$ for $x > x_d$ (compare, e.g., with Figs. 22.7 and 22.8) yields a simplified form of the Poisson equation,

$$\varphi'' \simeq \frac{qN_A}{\varepsilon_{\rm sc}}, \qquad 0 < x < x_d. \tag{22.31}$$

The boundary conditions of (22.31) are obtained in the same manner as in the p–n junction, and read $\varphi\left(x_d\right) = 0$, $\varphi'\left(x_d\right) = 0$; the solution of (22.31) fulfilling the boundary conditions is $\varphi(x) = qN_A\left(x - x_d\right)^2/\left(2\,\varepsilon_{\rm sc}\right)$. Letting $x = 0$ in the above, yields a relation between the surface potential and the depletion width; in turn, in the full-depletion and ASCE approximations the bulk charge per unit area is $Q_b = -qN_A\,x_d$. In summary,

$$\varphi_s = \frac{qN_A}{2\,\varepsilon_{\rm sc}}\,x_d^2, \qquad Q_b = -qN_A\,x_d = -\sqrt{2\,\varepsilon_{\rm sc}\,qN_A\,\varphi_s}. \tag{22.32}$$

Observing that for $0 < u_s < 2\,u_F$ it is $Q_{\rm sc} \simeq Q_b$, and that $N_A \simeq p_{p0}$, one finds that the second relation in (22.32) coincides with the first one in (22.30). Combining $Q_{\rm sc} \simeq Q_b$ with (22.30) and with the general expression (22.14) yields $V'_G - \varphi_s = \gamma\,\sqrt{\varphi_s}$; from it, one finds a simplified $\varphi_s = \varphi_s\left(V'_G\right)$ relation, that holds in the depletion and weak-inversion conditions:[8]

$$\sqrt{\varphi_s} = \sqrt{V'_G + \left(\gamma/2\right)^2} - \gamma/2. \tag{22.33}$$

The contribution of electrons to the semiconductor charge per unit area, $Q_i = -q\int_0^{x_d} n\;{\rm d}x$, becomes relevant from the threshold condition on. Remembering the

[7] The units of γ are $[\gamma] = {\rm V}^{1/2}$.

[8] The negative sign in front of the square root in (22.33) must be discarded.

discussion of Sect. 22.2.2, the electron charge is approximated as a charge layer at the interface, $-qn \simeq Q_i\,\delta(x^+)$; as a consequence, the space charge can be considered as entirely due to the ionized dopant atoms also when $\varphi_s > 2\,\varphi_F$, so that (22.31) and (22.32) still hold. From (22.16) one then finds the result sought, that is, a simplified form of the inversion layer's charge, that holds in the depletion, weak-inversion, and strong-inversion regimes:

$$Q_i = Q_{\rm sc} - Q_b = -C_{\rm ox}\left[\left(V'_G - \varphi_s\right) - \gamma\sqrt{\varphi_s}\right] < 0. \tag{22.34}$$

The theory worked out so far is based on the assumption that relation (22.1) holds between the gate metal's work function, the semiconductor's affinity, and the position of the semiconductor's Fermi level with respect to the band edges. Moreover, the gate oxide has been assumed free of charge. In fact, as the above conditions seldom occur, one must account for the effects of the difference $\Delta W = W - A - \left(E_C - E_F\right)$ and for the oxide charge. It can be shown that, if the oxide charge is fixed, the two additional effects produce a shift in V'_G with respect to the value $V'_G = V_G - \Phi_{mp}$ that holds for the ideal MOS capacitor with a *p*-type substrate [103, Sect. 9.4]. In fact, the combination of Φ_{mp}, ΔW, and of the oxide charge per unit area $Q_{\rm ox}$, provides an expression of the form $V'_G = V_G - V_{FB}$, where the constant $V_{FB} = V_{FB}(\Phi_{mp}, Q_{\rm ox}, \Delta W)$ is called *flat-band voltage*.[9]

22.4.1 Quantitative Relations in the MOS Capacitor

In the p-type silicon substrate considered so far it is $p_{p0} \simeq N_A = 10^{16}\ \text{cm}^{-3}$ and, at room temperature, $n_i \simeq 10^{10}\ \text{cm}^{-3}$. In turn, the asymptotic minority-carrier concentration is $n_{p0} = n_i^2/N_A \simeq 10^4\ \text{cm}^{-3}$; it follows $\exp\left(-2\,u_F\right) = n_{p0}/p_{p0} \simeq 10^{-12}$ and, as shown, e.g., in Fig. 22.5, $2\,u_F \simeq 27.6$. Using $k_B\,T/q \simeq 26$ mV then yields $2\,\varphi_F \simeq 0.72$ V. As $\varepsilon_{\rm sc} \simeq 11.7 \times 8.854 \times 10^{-14} = 1.036 \times 10^{-12}\ \text{F cm}^{-1}$, one finds

$$L_A = \sqrt{\frac{2\,\varepsilon_{\rm sc}\,k_B\,T}{q^2\,p_{p0}}} \simeq 5.8 \times 10^{-2}\ \ \mu\text{m}, \tag{22.35}$$

$$Q_{\rm sc}^{(1)} = \frac{2\,\varepsilon_{\rm sc}\,k_B\,T}{qL_A} = \sqrt{2\,\varepsilon_{\rm sc}\,k_B\,T\,p_{p0}} \simeq 9.3 \times 10^{-9}\ \ \text{C cm}^{-2}. \tag{22.36}$$

The relation $Q_{\rm sc}\left(V'_G\right)$ is found from $Q_{\rm sc} = -C_{\rm ox}\left[V'_G - \varphi\left(V'_G\right)\right]$, where $\varphi(V'_G)$ is obtained from (22.12). The result is shown in normalized form in Fig. 22.12.

[9] The designation is due to the fact that $V_G = V_{FB}$ establishes the flat-band condition in the semiconductor.

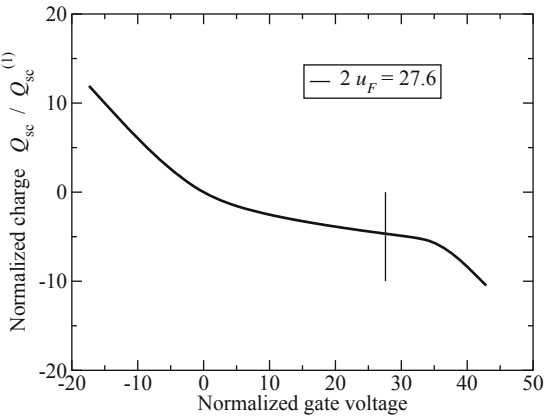

Fig. 22.12 Normalized semiconductor charge $Q_{sc}/Q_{sc}^{(1)}$ as a function of the normalized gate voltage u'_G, for a p-substrate MOS capacitor with $N_A = 10^{16}$ cm^{-3}

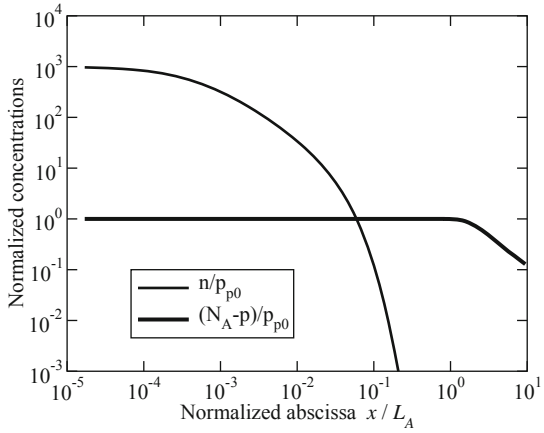

Fig. 22.13 Normalized concentrations n/p_{p0} and $(N_A - p)/p_{p0}$ as a function of position x/L_A, for a p-substrate MOS capacitor with $N_A = 10^{16}$ cm^{-3} in strong inversion $(u_s = 2.5\, u_F)$

Note that the results illustrated so far have been obtained without the need of integrating (22.6). The result of a numerical integration of (22.6) is shown in Fig. 22.13, where the dependence on position of n and $N_A - p$ is drawn for a p-substrate MOS capacitor with $N_A = 10^{16}$ cm^{-3} in the strong-inversion regime $(u_s = 2.5\, u_F)$. The term $(N_A - p)/p_{p0}$ is significant in a surface region of the semiconductor, whose thickness is several units of x/L_A. The term n/p_{p0} is much larger, but only in a much thinner region near the surface. If the width of the inversion layer is conventionally taken at the intersection between the two curves of Fig. 22.13, that occurs at $x/L_A \simeq 0.1$, one finds from (22.35) a width of about 5 nm.

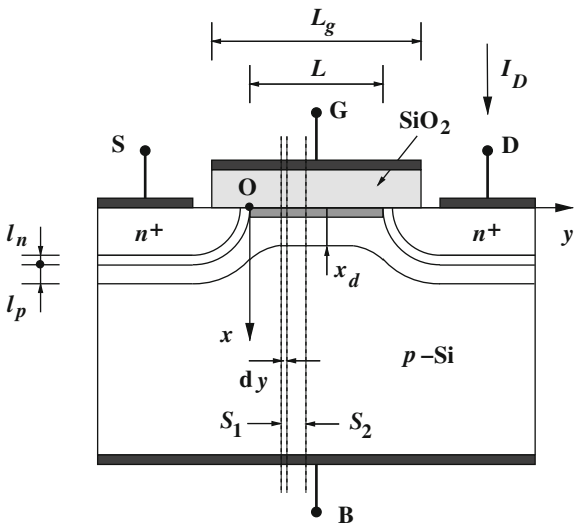

Fig. 22.14 Cross-section of an *n*-channel MOSFET. The *black areas* are the metal contacts

22.5 Insulated-Gate Field-Effect Transistor—MOSFET

The principle of the insulated-gate, field-effect transistor (IGFET) was demonstrated in the early 1930s [103, Chap. 10]. The first structures using a thermally-oxidized silicon layer were fabricated in 1960. The IGFET architecture using silicon dioxide as gate dielectric is more commonly called MOSFET. This device architecture, jointly with the continuous improvements in the silicon technology, made it possible the tremendous progress in the fabrication of integrated circuits during the last decades, and is the most common component in digital and analog circuits.

The electric current in a MOSFET is transported by one type of carriers only, electrons or holes; for this reason the device is called *unipolar*. In a *p*-substrate device, the carriers are the electrons that form the charge layer at the semiconductor–insulator interface; therefore, this type of transistor is called *n-channel* MOSFET. Conversely, in an *n*-substrate device the carriers are holes, and the transistor is called *p-channel* MOSFET. The schematic cross-section of an *n-channel* MOSFET is shown in Fig. 22.14. The starting point is a *p*-type silicon substrate, with an $N_A = \text{const}$ dopant concentration, onto which a layer of silicon dioxide is thermally grown and patterned; then, the gate contact (G) is deposited. The extension L_g of the gate metal in the horizontal direction is called *geometrical length* of the gate. The next step is the introduction of a heavy dose of an *n*-type dopant on the two sides of the gate. As shown in the figure, lateral diffusion (Sect. 23.7.3) makes the gate oxide to partially overlap the *n*-doped regions.[10]

[10] The *n*-type regions are typically obtained by ion implantation, whose lateral penetration is limited. However, ion implantation is followed by a thermal process (*annealing*), during which thermal diffusion takes place.

The metallizations of the n^+ regions provide two more contacts, called *source* (S) and *drain* (D); the bottom metal layer contacting the p-type substrate is indicated with *bulk* (B), and the term *channel* denotes the interfacial semiconductor region between the two junctions. To distinguish the applied voltages from one another, two letters are used; considering the bulk metallization as the reference contact, in an n-channel MOSFET a typical choice of the three independent voltages is $V_{GB} = V_G - V_B$, $V_{SB} = V_S - V_B$, and $V_{DB} = V_D - V_B$. As the standard MOSFET architecture is structurally symmetric, it is not possible to distinguish the source contact from the drain contact basing on geometry or dopant distribution: the distinction is to be based on the applied voltages; in fact, in the typical operating regime of the device the source–bulk and drain–bulk junctions are never forward biased, whence in an n-channel MOSFET it is $V_{SB} \geq 0$ and $V_{DB} \geq 0$. The drain contact is identified[11] by the condition $V_{DB} - V_{SB} = V_{DS} > 0$.

22.6 N-Channel MOSFET—Current-Voltage Characteristics

To work out the theory of the MOSFET, one introduces a reference whose x axis is normal to the semiconductor–insulator interface, while the y axis is parallel to it. The origin (O) is placed at the intersection of the source p–n junction and the interface (Fig. 22.14). The y coordinate corresponding to the intersection of the drain p–n junction and the interface is indicated with L; the latter is called *electric length* of the gate, or *channel length*. The device is considered uniform in the z direction; its width along such a direction is indicated with W.

Purpose of the analysis is to derive the steady-state characteristics, namely, the relations between the currents at the contacts and the applied voltages. To proceed, one assumes that the gate voltage V_{GB} is such that at all positions y along the channel the strong-inversion condition holds. Thus, a layer of electrons is present, indicated in Fig. 22.14 with the shaded area underneath the gate oxide; the term *well-formed channel* is used to denote this situation. The minimum gate voltage necessary for obtaining a well-formed channel will be identified later. In a steady-state condition there is no current through the gate contact because the gate insulator is in series to it; also, the current flowing through the bulk contact is negligibly small because the two junctions are reverse biased. Since the channel layer connects two heavily-doped regions of the n type, the application of a drain-source voltage $V_{DS} > 0$ gives rise to a current I_D that flows from the drain to the source contact (its reference is shown in Fig. 22.14); for a given V_{DS}, the drain current I_D is controlled by the amount of charge available in the channel, which is in turn controlled by the gate voltage V_{GB}. In other terms, the device is an electronic valve in which the gate–bulk port controls the current flowing into the drain–source port; moreover, in the steady-state

[11] In fact, in some types of logic circuits the source and drain contact may exchange their roles depending on the applied voltages.

condition the control port does not expend energy, because the gate current is zero. This, among other things, explains the success of the MOSFET concept.

Due to the uniformity in the z direction, the electron and hole current densities have the form

$$\mathbf{J}_n = J_{nx}\,\mathbf{i} + J_{ny}\,\mathbf{j}, \qquad \mathbf{J}_p = J_{px}\,\mathbf{i} + J_{py}\,\mathbf{j}, \tag{22.37}$$

with $\mathbf{i}$, $\mathbf{j}$ the unit vectors of the x and y axes, respectively. On the other hand it is $J_{nx} = J_{px} = 0$ because no current can flow through the insulator, so only the y components J_{ny}, J_{py} are left in (22.37). In turn, the condition of a well-formed channel implies that the concentration of holes is negligibly small with respect to that of the electrons,[12] whence $|J_{ny}| \gg |J_{py}|$. It follows $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p \simeq \mathbf{J}_n = J_{ny}(x, y)\,\mathbf{j}$. The equality $\mathbf{J} = \mathbf{J}_n$ is the mathematical form of the MOSFET's property of being unipolar. It also entails $\mathrm{div}\mathbf{J} = \mathrm{div}\mathbf{J}_n$; therefore, remembering that in a steady-state condition it is $\mathrm{div}\mathbf{J} = 0$, it follows that $\mathrm{div}\mathbf{J}_n = 0$ as well.

Consider now two planes parallel to the x, z plane, placed at different positions y_1 and y_2 in the channel; their intersections with the x, y plane are respectively marked with S_1 and S_2 in Fig. 22.14. From the divergence theorem (A.23) and the property $\mathrm{div}\mathbf{J}_n = 0$ it follows[13]

$$\iint_2 \mathbf{J}_n \cdot \mathbf{j}\; \mathrm{d}x\; \mathrm{d}z - \iint_1 \mathbf{J}_n \cdot \mathbf{j}\; \mathrm{d}x\; \mathrm{d}z = 0\,; \tag{22.38}$$

as a consequence, the channel current

$$I = \int_{z=0}^{W} \int_{x=0}^{x_d} \mathbf{J}_n \cdot \mathbf{j}\; \mathrm{d}x\; \mathrm{d}z = W \int_0^{x_d} J_{ny}\; \mathrm{d}x \tag{22.39}$$

is independent of y. The last form of (22.39) derives from the uniformity in z. To express J_{ny} in (22.39) it is convenient to adopt the monomial form (19.141) of the electron current density, $\mathbf{J}_n = -q\mu_n\, n\, \mathrm{grad}\varphi_n$, with φ_n the electron quasi-Fermi potential. The components of $\mathbf{J}_n$ in monomial form read

$$J_{nx} = -q\mu_n\, n\, \frac{\partial \varphi_n}{\partial x}, \qquad J_{ny} = -q\mu_n\, n\, \frac{\partial \varphi_n}{\partial y}, \tag{22.40}$$

where $J_{nx} = 0$ as found above. In the channel it is $n \neq 0$, whence for J_{nx} to vanish it must be $\partial\varphi_n/\partial x = 0$; as a consequence, φ_n in the channel[14] depends on y only. In conclusion,

$$I = W\, \frac{\mathrm{d}\varphi_n}{\mathrm{d}y}\, Q_i(y), \qquad Q_i = \int_0^{x_d} -q\mu_n\, n\, \mathrm{d}x < 0, \tag{22.41}$$

---

[12] Compare with Fig. 22.13; the latter describes an equilibrium case, however the situation is similar to the one depicted here.

[13] In the integrals of (22.38) the upper limit of x is given by the depletion width $x_d(y)$ shown in Fig. 22.14. Compare also with Sect. 22.7.1.

[14] Far from the channel the semiconductor is practically in the equilibrium condition, whence $\varphi_n \to \varphi_F$ as x increases. However, in the bulk region where the dependence of φ_n on x is significant, the electron concentration is negligible; as a consequence, the integral in (22.39) is not affected.

where Q_i is the inversion-layer charge per unit area at position y in the channel. In the integral of (22.41) it is $n = n(x, y)$ and $\mu_n = \mu_n(x, y)$; defining the *effective electron mobility* as the average

$$\mu_e(y) = \frac{\int_0^{x_d} -q\mu_n n \, \mathrm{d}x}{\int_0^{x_d} -qn \, \mathrm{d}x} > 0, \tag{22.42}$$

yields

$$I = W \frac{\mathrm{d}\varphi_n}{\mathrm{d}y} \mu_e(y) \, Q_i(y). \tag{22.43}$$

In (22.43), μ_e and Q_i are positive- and negative-definite, respectively, and I, W are constant; it follows that $\mathrm{d}\varphi_n/\mathrm{d}y$ has always the same sign and, as a consequence, $\varphi_n(y)$ is invertible. Using the inverse function $y = y(\varphi_n)$ within μ_e and Q_i makes (22.43) separable; integrating the latter over the channel yields

$$\int_0^L I \, \mathrm{d}y = L \, I = W \int_{\varphi_n(0)}^{\varphi_n(L)} \mu_e(\varphi_n) \, Q_i(\varphi_n) \, \mathrm{d}\varphi_n. \tag{22.44}$$

In turn, the dependence of μ_e on y is weak,[15] whence

$$I = \frac{W}{L} \mu_e \int_{\varphi_n(0)}^{\varphi_n(L)} Q_i(\varphi_n) \, \mathrm{d}\varphi_n. \tag{22.45}$$

22.6.1 Gradual-Channel Approximation

In the derivation of (22.45) the condition of a well-formed channel has not been exploited yet; this condition makes a number of approximations possible, which are collectively indicated with the term *gradual-channel approximation*;[16] they lead to an expression of (22.45) in closed form. First, one uses the definition of surface potential which, in the two-dimensional analysis considered here, is given by $\varphi_s(y) = \varphi(x = 0, y)$; it is shown in Sect. 22.7.1 that the condition of a well-formed channel entails the relation

$$\varphi_s = \varphi_n + \varphi_F. \tag{22.46}$$

[15] This issue is discussed in Sect. 22.7.1.

[16] The gradual-channel approximation is not limited to the analysis of the MOSFET shown here. Indeed, it is a widely-used method to treat the Poisson equation in devices in which the geometrical configuration and applied voltages are such that the variation of the electric field in one direction is much weaker than those in the other two directions. Typically, the former direction is the longitudinal one (that is, along the channel), the other two the transversal ones. From the mathematical standpoint, the approximation amounts to eliminating a part of the Laplacian operator, so that the dependence on all variables but one becomes purely algebraic.

It follows that $\mathrm{d}\varphi_n/\mathrm{d}y$ in (22.43) can be replaced with $\mathrm{d}\varphi_s/\mathrm{d}y$, this showing that the transport in a well-formed channel is dominated by the drift term, $J_{ny} = -q\mu_n n\,\mathrm{d}\varphi_s/\mathrm{d}y = q\mu_n n E_{sy}$, with E_{sy} the y-component of the electric field at $x = 0$; using (22.46) one changes the variable from φ_n to φ_s in the integral of (22.45). The integration limits in terms of φ_s are found by the same reasoning leading to (22.46), and read (Sect. 22.7.1)

$$\varphi_s(0) = V_{SB} + 2\,\varphi_F, \qquad \varphi_s(L) = V_{DB} + 2\,\varphi_F. \tag{22.47}$$

In conclusion, (22.45) becomes

$$I = \frac{W}{L}\,\mu_e \int_{V_{SB}+2\,\varphi_F}^{V_{DB}+2\,\varphi_F} Q_i\big(\varphi_s\big)\ \mathrm{d}\varphi_s. \tag{22.48}$$

The next step of the gradual-channel approximation consists in determining the relation $Q_i(\varphi_s)$, for which the solution of the Poisson equation in two dimensions is necessary. As shown in Sect. 22.7.1, one can exploit the strong difference between the strengths of the electric-field components in the x and y directions, to give the equation a one-dimensional form in which the y coordinate acts as a parameter. This is equivalent to assimilating each elementary portion of the channel, like that marked with $\mathrm{d}y$ in Fig. 22.14, to a one-dimensional MOS capacitor whose surface potential has the local value $\varphi_s(y)$. The final step of the gradual-channel approximation is the adoption of the full-depletion and ASCE approximations (Sect. 21.4), so that the inversion-layer charge per unit area at position y in the channel is given by (22.34), namely, $Q_i = -C_{\mathrm{ox}}\left[\big(V'_{GB} - \varphi_s\big) - \gamma\sqrt{\varphi_s}\right] < 0$. Observing that $V_{DB} + 2\,\varphi_F > V_{SB} + 2\,\varphi_F$ while the integrand in (22.48) is negative, it follows that $I < 0$, whence $I_D = -I$ due to the reference chosen for the drain current (Fig. 22.14). In conclusion, (22.48) transforms into[17]

$$I_D = \beta \int_{V_{SB}+2\,\varphi_F}^{V_{DB}+2\,\varphi_F} \left[\big(V'_{GB} - \varphi_s\big) - \gamma\sqrt{\varphi_s}\right] \mathrm{d}\varphi_s, \qquad \beta = \frac{W}{L}\,\mu_e\,C_{\mathrm{ox}}. \tag{22.49}$$

22.6.2 *Differential Conductances and Drain Current*

The drain current's expression (22.49) of the n-channel MOSFET provides a relation of the form $I_D = I_D\big(V_{GB}, V_{DB}, V_{SB}\big)$, where $V_{GB} = V'_{GB} + V_{FB}$. In the integrated-circuit operation an important role is played by the *differential conductances* of the device, each of them defined as the partial derivative of I_D with respect to one of the applied voltages. In some cases the differential conductances can be found without the need of actually calculating the integral in (22.49); for this reason, here such conductances are calculated first. Prior to that, it is worth noting that in circuit

[17] The units of β are $[\beta] = \mathrm{A\ V^{-2}}$.

applications it is often preferred to use the source contact, instead of the bulk contact, as a voltage reference. The transformation from one reference to the other is easily obtained from

$$V_{DS} = V_{DB} - V_{SB} > 0, \qquad V_{GS} = V_{GB} - V_{SB}, \qquad V_{BS} = -V_{SB} \leq 0. \tag{22.50}$$

Then, the *drain conductance*[18] is defined as the derivative of I_D with respect to V_{DB}, at constant V_{SB} and V_{GB}; or, equivalently, as the derivative with respect to V_{DS}, at constant V_{BS} and V_{GS}:

$$g_D = \left(\frac{\partial I_D}{\partial V_{DB}}\right)_{V_{SB},V_{GB}} = \left(\frac{\partial I_D}{\partial V_{DS}}\right)_{V_{BS},V_{GS}}. \tag{22.51}$$

Remembering that the derivative of an integral with respect to the upper limit is the integrand calculated at such limit, from (22.49) one finds

$$g_D = \beta \left[\left(V'_{GB} - V_{DB} - 2\,\varphi_F\right) - \gamma\sqrt{V_{DB} + 2\,\varphi_F}\right]. \tag{22.52}$$

Using (22.47) and (22.34) yields

$$g_D = \beta \left[\left(V'_{GB} - \varphi_s(L)\right) - \gamma\sqrt{\varphi_s(L)}\right] = \frac{W}{L}\,\mu_e\,[-Q_i(L)], \tag{22.53}$$

namely, the output conductance is proportional to the inversion charge per unit area at the drain end of the channel. The quantity in brackets in (22.53) is non-negative by construction; its zero corresponds to the value of $\varphi_s(L)$ obtained from (22.33). Such a zero is indicated with φ_s^{sat} and is termed *saturation surface potential*. From (22.50), the *saturation voltage* in the bulk and source references is found to be

$$V_{DB}^{\text{sat}} = \varphi_s^{\text{sat}} - 2\,\varphi_F, \qquad V_{DS}^{\text{sat}} = V_{DB}^{\text{sat}} - V_{SB}, \tag{22.54}$$

respectively. Similarly, the current $I_D^{\text{sat}} = I_D\left(V_{DS}^{\text{sat}}\right)$ (which depends on V_{GS}) is called *saturation current*. If a value of V_{DS} larger than V_{DS}^{sat} is used, g_D becomes negative; this result is not physically sound and indicates that the gradual-channel approximation is not applicable in that voltage range.[19]

Still considering the drain conductance, it is also important to determine its limit for $V_{DB} \to V_{SB}$, or $V_{DS} \to 0$. Again, there is no need to calculate the integral in (22.49) which, in this limiting case, is the product of the integration interval V_{DS} times the integrand calculated in the lower integration limit; in turn, the derivative eliminates V_{DS}, whence

$$g_D(V_{DS} \to 0) = \beta \left[\left(V'_{GB} - V_{SB} - 2\,\varphi_F\right) - \gamma\sqrt{V_{SB} + 2\,\varphi_F}\right]. \tag{22.55}$$

[18] The drain conductance is also called *output conductance*; in this case it is indicated with g_o.

[19] In fact, beyond the saturation voltage the Poisson equation near the drain end of the channel can not be reduced any more to a one-dimensional equation where y is treated as a parameter (compare with Sect. 22.7.1).

Replacing V'_{GB} with $V_{GB} - V_{FB}$ and using (22.50) yields

$$g_D(V_{DS} \to 0) = \beta\ (V_{GS} - V_T)\,, \qquad V_T = V_{FB} + 2\,\varphi_F + \gamma\sqrt{2\,\varphi_F - V_{BS}} \tag{22.56}$$

where, remembering that the junctions are never forward biased, it is $V_{BS} \leq 0$. The $V_T = V_T\big(V_{BS}\big)$ voltage defined in (22.56) is called *threshold voltage*, and its dependence on V_{BS} is called *body effect*. Near $V_{DS} = 0$ the relation between I_D and V_{DS} is $I_D = \beta\ (V_{GS} - V_T)\ V_{DS}$: there, the current-voltage characteristics are approximated by straight lines whose slope, for a given V_{BS}, is prescribed by V_{GS}. At larger values of V_{DS} the limiting case (22.56) does not hold any longer: the slope of the $I_D = I_D(V_{DS})$ curves decreases, to eventually vanish when V_{DS} reaches $V_{DS}^{\rm sat}$. Considering that I_D is non-negative, the theory depicted above is applicable as long as $\beta\ (V_{GS} - V_T) \geq 0$; this observation allows one to better specify the condition of a well-formed channel, used at the beginning: from the formal standpoint the condition of a well-formed channel is $V_{GS} > V_T$.

The integration of (22.49) is straightforward and yields $I_D = I'_D - I''_D$, where I'_D is obtained by integrating $\beta\left(V'_{GB} - \varphi_s\right)$. In this calculation many terms cancel out, to yield the relatively simple expression

$$I'_D = \beta\left[(V_{GS} - V_{FB} - 2\,\varphi_F)\ V_{DS} - \frac{1}{2}\,V_{DS}^2\right]. \tag{22.57}$$

In turn, I''_D is obtained by integrating $-\beta\,\gamma\,\sqrt{\varphi_s}$, and reads

$$I''_D = \beta\,\frac{2}{3}\,\gamma\,\left[(V_{DS} + 2\,\varphi_F - V_{BS})^{3/2} - (2\,\varphi_F - V_{BS})^{3/2}\right]. \tag{22.58}$$

Comparisons with experiments show that the model $I_D = I'_D - I''_D$, where the two contributions are given by (22.57) and (22.58), provides a fair description of the drain current up to the saturation voltage. Beyond saturation, the model is not correct any longer: in fact, the terms with a negative sign within the expression of I_D give rise to a negative slope g_D; instead, the experiments show that for $V_{DS} > V_{DS}^{\rm sat}$ the current tends to saturate. For this reason, the analytical model is given a regional form: for a prescribed pair V_{GS}, V_{BS}, the regional model first separates the *on* condition $V_{GS} > V_T$ from the *off* condition $V_{GS} \leq V_T$. The *on* condition is further separated into the *linear region*[20] $0 < V_{DS} \leq V_{DS}^{\rm sat}$, where the drain current is described by the $I_D = I'_D - I''_D$ model worked out above, and the *saturation region* $V_{DS} > V_{DS}^{\rm sat}$, where the regional model lets $I_D = I_D^{\rm sat}$. Finally, in the *off* condition the model lets $I_D = 0$.

For a given bulk-source voltage V_{BS}, the $I_D = I_D\big(V_{DS}\big)$ curves corresponding to different values of V_{GS} are called *output characteristics*. Other types of characteristics

[20] The term *linear* originates from the behavior of the curves near the origin, shown by (22.56). The term is ascribed to the region up to $V_{DS}^{\rm sat}$ despite the fact that far away from the origin the curves are blatantly non linear.

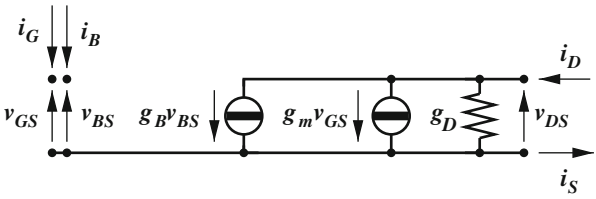

Fig. 22.15 Low frequency, small-signal circuit of an n-channel MOSFET

are also used to enrich the picture of the MOSFET's behavior: for instance, the *transfer characteristics* are the $I_D = I_D(V_{GS})$ curves drawn using V_{BS} as a parameter and letting $V_{DS} = \text{const}$, with a value of V_{DS} small enough to let the limiting case (22.56) hold.

Besides the drain conductance (22.51), two more differential conductances are defined in a MOSFET: the first one is the *transconductance* g_m, given by the derivative of I_D with respect to V_{GB}, at constant V_{DB} and V_{SB}; or, equivalently, as the derivative with respect to V_{GS}, at constant V_{DS} and V_{BS}. Observing that V_{GS} appears only in I'_D one finds

$$g_m = \left(\frac{\partial I_D}{\partial V_{GB}}\right)_{V_{SB},V_{DB}} = \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS},V_{BS}} = \beta\, V_{DS}. \tag{22.59}$$

The second one is the *bulk transconductance* g_B, defined as the derivative of I_D with respect to V_{BS} at constant V_{DS} and V_{GS}:

$$g_B = \left(\frac{\partial I_D}{\partial V_{BS}}\right)_{V_{DS},V_{GS}} = -\left(\frac{\partial I''_D}{\partial V_{BS}}\right)_{V_{DS}}. \tag{22.60}$$

The small-signal circuit of an n-channel MOSFET is shown in Fig. 22.15. Since the circuit is derived from the steady-state transport model, it holds at low frequencies only. The small-signal voltages are indicated with v_{DS}, v_{GS}, and v_{BS}. The gate and bulk contacts are left open because the corresponding currents are zero; as a consequence, $i_D = i_S$ is the only non-zero small-signal current of the circuit. Observing that

$$i_D = g_D\, v_{DS} + g_m\, v_{GS} + g_B\, v_{BS}, \tag{22.61}$$

the drain-source branch of the circuit is made of three parallel branches. One of them is represented as a resistor $1/g_D$ because the current flowing in it is controlled by the voltage v_{DS} applied to the same port; the other two branches are voltage-controlled generators because the current of each branch is controlled by the voltage applied to a different port.

It is worth adding that the body effect mentioned above is actually an inconvenience, because it introduces a complicate dependence on V_{BS} which must be accounted for during the circuit's design. The body effect is suppressed by letting

$V_{BS} = 0$: in a circuit's design, this is obtained by shorting the bulk and source contacts, which amounts to reducing the original four-contact device to a three-contact device.[21] This solution is adopted whenever the circuit's architecture allows for it.

22.6.2.1 Linear–Parabolic Model

In semiqualitative circuit analyses the whole term I''_D is neglected, this leading to a simplified model $I_D \simeq I'_D$, called *linear–parabolic model*. As the neglect of I''_D is equivalent to letting $\gamma \to 0$, it follows $g_B \simeq 0$ and, from the second relation in (22.56), the simplified threshold voltage reads

$$V_T \simeq V_{FB} + 2\,\varphi_F\,. \tag{22.62}$$

In turn, from $I_D \simeq I'_D = \beta\left[\left(V_{GS} - V_T\right) V_{DS} - V_{DS}^2/2\right]$ one finds for the drain conductance

$$g_D \simeq \beta\ \left(V_{GS} - V_{FB} - 2\,\varphi_F - V_{DS}\right) = \beta\ \left(V_{GS} - V_T - V_{DS}\right), \tag{22.63}$$

whence

$$V_{DS}^{\text{sat}} = V_{GS} - V_T\,. \tag{22.64}$$

The transconductance g_m is the same as in the general case. Note that the linear–parabolic expression of the drain current may be recast as $I_D \simeq \beta\left(V_{DS}^{\text{sat}}\ V_{DS} - V_{DS}^2/2\right)$, with V_{DS}^{sat} given by (22.64). As a consequence,

$$I_D^{\text{sat}} \simeq \beta\left[\left(V_{DS}^{\text{sat}}\right)^2 - \frac{1}{2}\left(V_{DS}^{\text{sat}}\right)^2\right] = \frac{1}{2}\,\beta\ \left(V_{GS} - V_T\right)^2\,. \tag{22.65}$$

The linear–parabolic model then yields for the saturation region

$$I_D = I_D^{\text{sat}}, \qquad g_D \simeq 0, \qquad g_m \simeq \beta\ \left(V_{GS} - V_T\right), \qquad g_B \simeq 0. \tag{22.66}$$

An example of the output characteristics of an n-type MOSFET obtained from the linear–parabolic model is given in Fig. 22.16, using $V_T = 1$ V, $\beta = 0.3$ A V^{-2}. The dashed curve represents (22.65).

22.7 Complements

22.7.1 Poisson's Equation in the MOSFET Channel

The derivation of the MOSFET's current carried out in Sect. 22.6 is based upon two integrals; the first one, (22.39), is calculated over a section of the channel at

[21] Note that letting $V_{BS} = 0$ also makes g_B to vanish.

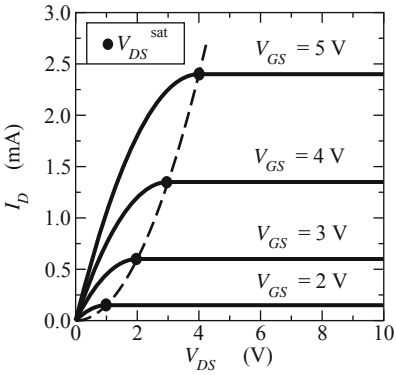

Fig. 22.16 Output characteristics of an *n*-type MOSFET obtained from the linear–parabolic model, with $V_T = 1\,\mathrm{V}$, $\beta = 0.3\,\mathrm{A\,V^{-2}}$. The *dashed curve* represents (22.65)

some position y, the second one, (22.44), is calculated along the channel from $y = 0$ to $y = L$. Apparently this procedure eliminates the need of solving the Poisson equation. In fact, the solution of the latter is deeply rooted in the relation (22.46), which is a fundamental point of the procedure itself, and in the choice of the integration limits (22.47), which are also related to (22.46).

The Poisson equation in a non-equilibrium condition is conveniently tackled by expressing the carrier concentrations in terms of the quasi-Fermi potentials φ_n and φ_p; the device considered here is the same *n*-channel MOSFET of Sect. 22.6. Using the normalized form $u_n = q\varphi_n/(k_B\,T)$ and $u_p = q\varphi_p/(k_B\,T)$, the concentrations read $n = n_i\,\exp(u - u_n)$ and $p = n_i\,\exp(u_p - u)$, respectively. Remembering that in the equilibrium limit it is $u_n, u_p \to u_F$, with u_F the normalized Fermi potential, it is useful to introduce the differences

$$\chi_n = u_n - u_F, \qquad \chi_p = u_p - u_F, \tag{22.67}$$

by which the concentrations take the form

$$n = n_{p0}\,\exp(u - \chi_n), \qquad p = p_{p0}\,\exp(\chi_p - u). \tag{22.68}$$

In the equilibrium limit it is $\chi_n, \chi_p \to 0$. Moreover, when a non-equilibrium condition holds, at any position y in the channel it is $\lim_{x\to\infty} \chi_n, \chi_p = 0$; in fact, as observed in Sect. 22.6, far from the channel the semiconductor is practically in the equilibrium condition, whence $\varphi_n \to \varphi_F$ as x increases. The same applies to φ_p. With these provisions, the charge density in the semiconductor reads $\rho = q\left[p_{p0}\,\exp(\chi_p - u) - n_{p0}\,\exp(u - \chi_n) - N_A\right]$, namely,

$$\rho = -qp_{p0}\,A, \qquad A = \frac{n_i^2}{N_A^2}\left[\exp(u - \chi_n) - 1\right] + 1 - \exp(\chi_p - u), \tag{22.69}$$

and the Poisson equation takes the form

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{1}{L_A^2}\,A, \tag{22.70}$$

with L_A the holes' Debye length defined in (21.12). One notes that (22.69), (22.70) are generalizations of (22.3). In the description of the MOSFET the accumulation condition is not considered, hence the holes' contribution $\exp\left(\chi_p - u\right)$ to A is negligible in the channel region; thus,

$$A \simeq \frac{n_i^2}{N_A^2}\left[\exp\left(u - \chi_n\right) - 1\right] + 1 \simeq \frac{n_i^2}{N_A^2}\exp\left(u - \chi_n\right) + 1 > 0, \tag{22.71}$$

where the term $n_i^2/N_A^2 = \exp\left(-2\,u_F\right)$ is negligible with respect to unity. Remembering that the quasi-Fermi potential in the channel does not depend on x, in (22.71) it is $u = u(x, y)$, $\chi_n = \chi_n(y)$, with $u(\infty, y) = 0$, $u'(\infty, y) = 0$ due to the charge neutrality of the bulk region; in fact, as shown by numerical solutions, both u and u' practically vanish when x reaches the value of the depletion width $x_d(y)$.

The Poisson equation is to be solved in two dimensions. If the condition of a well-formed channel holds, the components of the electric field along the x and y directions are quite different from each other. The x component at the semiconductor–oxide interface, E_{sx}, which is due to the voltage applied to the gate contact, is large because it maintains the strong-inversion condition of the surface. Moreover, the derivative $\partial E_x/\partial x = -\partial^2 u/\partial x^2$ is also large, because E_x changes from E_{sx} to zero in the short distance $x_d(y)$. In contrast, the y component of the electric field at the interface, E_{sy}, which is due to the voltage V_{DS} applied between the drain and source contacts, is small; in fact, V_{DS} is small in itself because the linear region only is considered, and the channel length L is larger than the insulator thickness. Moreover, numerical solutions show that for $0 < V_{DS} < V_{DS}^{\text{sat}}$ the dependence of both E_{sx} and E_{sy} on y is weak, which in particular makes $\partial E_y/\partial y = -\partial^2 u/\partial y^2$ also small at the semiconductor–insulator interface. In conclusion, one approximates (22.70) as

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \simeq \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} = \frac{1}{L_A^2}\,A. \tag{22.72}$$

The dependence of A on y remains, and y is treated as a parameter in the solution procedure. Due to the form of (22.72), the solution method is identical to that used in Sect. 22.2.1 to treat the equilibrium case; it yields

$$\left(\frac{qE_{sx}}{k_B\,T}\right)^2 = \frac{2}{L_A^2}\,F^2, \qquad F^2 = \exp\left(-\chi_n - 2\,u_F\right)\left[\exp\left(u_s\right) - 1\right] + u_s. \tag{22.73}$$

Remembering that the accumulation condition is excluded, here it is $u_s(y) \geq 0$; the flat-band condition $u_s = 0$ corresponds to $F = 0$. In the strong-inversion condition the contribution of the electron charge (proportional to $\exp\left(u_s\right) - 1$ in (22.73)) is dominant; for this to happen it is necessary that the exponent $u_s - \chi_n - 2\,u_F$ in (22.73) be positive; it follows that the threshold condition is identified by $u_s = \chi_n + 2\,u_F$. Remembering the definition (22.67) of χ_n one then finds

$$u_s = u_n + u_F, \qquad 0 \leq y \leq L, \tag{22.74}$$

that is, the normalized form of (22.46). Note that $u_s = \chi_n + 2\,u_F$ is coherent with the definition of the threshold condition at equilibrium (Table 22.1), which is obtained by letting $\chi_n = 0$. Also, specifying $u_s = \chi_n + 2\,u_F$ at the source and drain ends of the channel provides the integration limits (22.47) [103, Sect. 10.2].

The neglect of the variation in the y component of the electric field along the channel makes the general relation

$$Q_{\rm sc} = -C_{\rm ox}\,(V'_{GB} - \varphi_s) = -C_{\rm ox}\,\frac{k_B\,T}{q}\,(u'_{GB} - u_s) \tag{22.75}$$

still valid at each position y along the channel, with $Q_{\rm sc} < 0$ because $u_s > 0$ ($Q_{\rm sc} = 0$ in the flat-band condition $u_s = 0$). Also, when the inversion charge is approximated by a charge layer at $x = 0^+$, the volume charge is entirely due to the ionized dopants, whose contribution in the second relation of (22.73) is proportional to u_s like in the equilibrium case. Using the same relation $Q_i = Q_{\rm sc} - Q_b$ as in Sect. 22.4 one finally finds that the theory worked out in this section makes (22.34) applicable also in a non-equilibrium condition in which the channel is well formed.

22.7.2 Inversion-Layer Charge and Mobility Degradation

In the model (22.49) for the drain current worked out in the previous sections, the modulus of the inversion-layer charge Q_i decreases from source to drain due to the increase in φ_s along the channel (compare with (22.34)). Considering that the MOSFET current is carried by the inversion-layer charge, the vanishing of the latter occurring at the drain end of the channel when $V_{DS} \to V_{DS}^{\rm sat}$ may seem an oddity. However, it is important to remember that a number of approximations are necessary to reach the result expressed by (22.57) and (22.58); such approximations make the theory applicable only in the linear region and in the condition of a well-formed channel.

When $V_{DS} \to V_{DS}^{\rm sat}$, the vertical component of the electric field at the interface, E_{sx}, is made weaker by the interplay between the voltages applied to the gate electrode and to the nearby drain electrode; for this reason, at the drain end of the channel the flow lines of $\mathbf{J}_n$ do not keep close to the interface any longer, but spread into the substrate, this decreasing the carrier density. The phenomenon is better understood with the aid of Fig. 22.17, where the linear-parabolic model is used with, e.g., $V_T = 0.5$, $V_S = 0$, $V_{GS} = 1.5$, $V_{DS} = 2\,\text{V}$. As $V_{DS}^{\rm sat} = 1\,\text{V}$, the saturation condition $V_{DS} > V_{DS}^{\rm sat} = 1$ holds. Along the dash-dotted line enclosed in the right oval, the direction of the electric field is that shown in the vector diagram in the upper-right part of the figure; in particular, the vertical component of the field at the position marked by the vertical dashed line points upwards. As a consequence, the channel electrons are repelled downwards and the flow lines of the current density detach from the interface. At the source end of the channel, instead, the vertical component of the field points downwards. By continuity, a position within the channel exists where, in saturation, the vertical component of the field vanishes; such a position (not shown in the figure) is called *inversion point*. Also, the large component of the electric field along the y direction,

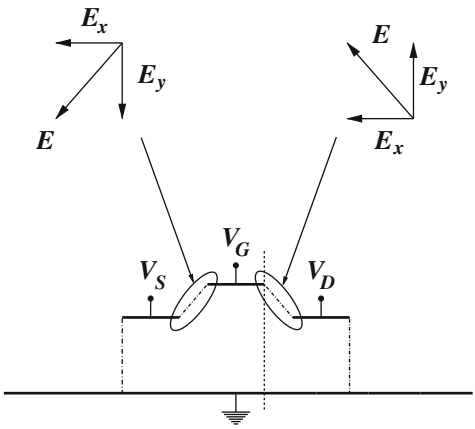

Fig. 22.17 Illustration of the electric field's components at the channel ends in the saturation condition

which exists within the space-charge region of the reverse-biased drain junction, makes the carriers' average velocity to increase at the drain end of the channel; as the total current is constant, a further decrease in the carrier concentration occurs. The two effects briefly illustrated above are not accounted for by the simplified model, and require a more elaborate approach in which the two-dimensional structure of the electric field is accounted for.

Another comment is useful, this time related to the *off* condition. When $V_{GS} \to V_T$, the current of a real device does not actually vanish; in fact, a current (called *subthreshold current*), due to the carriers present in the channel in the weak-inversion condition, flows for $V_{GS} < V_T$. Also in this case a more elaborate theory is necessary, showing that the subthreshold current vanishes exponentially as V_{GS} decreases [3].

It is worth concluding this section by commenting the simplification used in (22.45), where the effective electron mobility μ_e is assumed to be independent of the position y along the channel. The factors that affect mobility in a MOSFET are collisions with phonons, ionized impurities, and semiconductor–oxide interface; remembering the features of the macroscopic mobility models (Sect. 20.5), the electron mobility $\mu_n(x, y)$ is made to depend on the lattice temperature T, concentration of ionized impurities N_A, and x component of the electric field $E_x(x, y)$. The first two parameters, T and N_A, do not introduce a dependence on position because they are themselves constant. The x dependence of E_x is absorbed by the integral (22.42) that defines μ_e; it follows that the average mobility depends on y because E_x does. Such a dependence, in turn, is relatively weak in the strong-inversion condition as remarked in Sect. 22.7.1. Therefore, the dependence of μ_e on position is considered negligible.

Problems

22.1 Work out a method for drawing the curves of Fig. 22.9 without approximations.