# 5

# EKV Model of the MOS Transistor

by Eric A. Vittoz and Christian C. Enz

## 5.1 Introduction and Definitions

The transistor model described in this chapter is the core of the EKV compact model that has been specially developed for low-voltage and/or low-current circuit design. It is a shortened description, focusing on weak and moderate inversion regimes that correspond to sub-threshold operation. More detailed derivations can be found in [45, 49, 106]. A very detailed presentation can be found in "Charge-Based MOS Transistor Modeling: The EKV Model for Low-Power and RF IC Design", by C. Enz and E. Vittoz, 2006, ©John Wiley & Sons Limited [52]. (Figures 1, 2, 4, 5, 6, 7, 10, 15, 16, 18 and 19 of this chapter are extracted from this book and reproduced with permission.)

The schematic cross-section of an N-channel MOS transistor with channel length $L$ and channel width $W$ is illustrated in Figure 5.1. In order to maintain

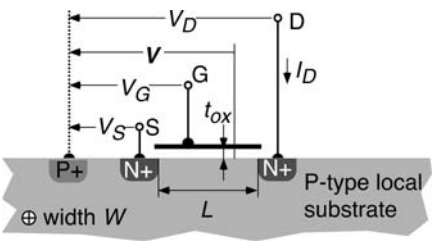



**Fig. 5.1.** Schematic cross-section of an N-channel MOS transistor, with definitions of voltages and current.

the intrinsic symmetry of the device, the source voltage $V_S$, gate voltage $V_G$ and drain voltage $V_D$ are all defined with respect to the local substrate. The

drain current $I_D$ is positive if it enters the drain. The *channel voltage* $V$ (quasi-Fermi potential of electrons in the channel) changes monotonically, from $V = V_S$ at the source end of the channel to $V = V_D$ at the drain end of the channel.

Another important voltage is the thermodynamic voltage

$$U_T = kT/q, \tag{5.1}$$

where $k$ is the Boltzmann constant and $q$ the elementary charge. Proportional to the absolute temperature $T$, it is a measure of the thermal energy of electrons. Its value is 25.8 mV at 300 K (or 27 °C) and it appears ubiquitously in MOS modeling equations.

The doping concentration of the substrate is assumed to have a constant value $N_b$ in the channel, and the gate oxide thickness $t_{ox}$ corresponds to a capacitance $C_{ox} = \epsilon_{ox}/t_{ox}$ per unit area.

Figure 5.2 shows the symbols that can be used in order to preserve the symmetry of the device. It also shows how the definition of positive voltages and current can be inverted so that the model developed for the N-channel transistor can be applied without any change to the P-channel device.

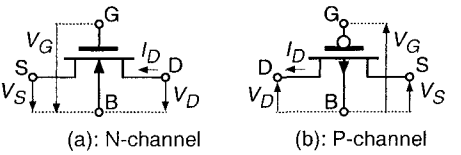



**Fig. 5.2.** Symbols for N-channel and P-channel devices with the respective definitions of positive voltages and current.

## 5.2 Density of Mobile Charge

### 5.2.1 Threshold Function

When a positive voltage is applied to the gate of the N-channel transistor, the holes in the channel are repelled away from the surface, thereby creating a depleted layer underneath the silicon surface and increasing its potential $\Psi_s$. This depleted layer is due to the remaining fixed ionized impurity atoms and is characterized by a negative charge density $Q_b$ (per unit area of channel) given by

$$Q_b = -\Gamma_b C_{ox} \sqrt{\Psi_s}, \tag{5.2}$$

where

$$\Gamma_b = \frac{\sqrt{2qN_b\epsilon_{si}}}{C_{ox}} \tag{5.3}$$

is the substrate modulation factor. This fixed charge $Q_b$ is useless since it cannot move to create a current. But the positive surface potential also attracts electrons to the surface, producing a mobile *inverted charge* of local density $Q_i$ that can carry a current. The thickness of this inversion charge layer is very small, therefore the voltage drop across it can be neglected (charge sheet approximation).

Using Gauss' law, the total charge density underneath the silicon surface is given by

$$Q_{si} = Q_b + Q_i = -C_{ox}(V_G - V_{FB} - \Psi_s), \tag{5.4}$$

where $V_{FB}$ is the flat-band voltage that includes the difference $\Phi_{ms}$ of extraction potentials between the gate and channel material and the effect of the fixed charge of density $Q_{fc}$ possibly trapped in the oxide and at its interface with the silicon:

$$V_{FB} = \Phi_{ms} - Q_{fc}/C_{ox}. \tag{5.5}$$

Combining (5.4) with (5.2) yields the density of inverted charge

$$Q_i = -C_{ox}(V_G - V_{FB} - \Psi_s - \Gamma_b\sqrt{\Psi_s}) = -C_{ox}(V_G - V_{TB}), \tag{5.6}$$

where

$$V_{TB} = V_{FB} + \Psi_s + \Gamma_b\sqrt{\Psi_s} \tag{5.7}$$

is a *threshold function*. This function of the surface potential $\Psi_s$ depends on the process through parameters $\Gamma_b$ and $V_{FB}$. It is represented in Figure 5.3(a) for a particular value of $\Gamma_b$. The figure also shows the value of $-Q_i/C_{ox}$ according to (5.6) and that of $-Q_b/C_{ox}$ for a particular value of the gate voltage $V_G$. This function is nonlinear due to the contribution of $Q_b$, and its slope $n > 1$

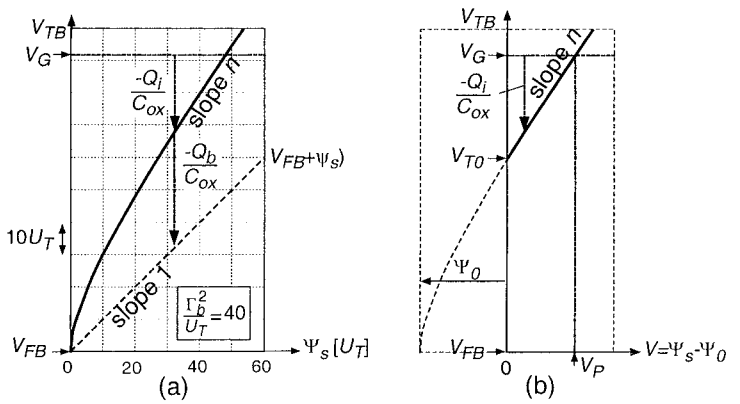



**Fig. 5.3.** Threshold function and inverted charge density: (a) as a function of the surface potential; (b) approximation in strong inversion.

is obtained by differentiation of (5.7):

$$n = \frac{\mathrm{d}V_{TB}}{\mathrm{d}\Psi_s} = 1 + \frac{\Gamma_b}{2\sqrt{\Psi_s}}. \tag{5.8}$$

Inspection of Figure 5.3(a) shows that, for a fixed value of the gate voltage $V_G$,

$$\frac{\mathrm{d}Q_i/C_{ox}}{\mathrm{d}\Psi_s} = n. \tag{5.9}$$

Now, the local value of inverted charge density $Q_i$ can only be obtained from Figure 5.3(a) if the local value of $\Psi_s$ is known. We shall first consider the case of strong inversion.

### 5.2.2 Approximation in Strong Inversion

It can be shown [107, 52] that the local inverted charge increases exponentially with $\Psi_s - V$ according to

$$Q_i \propto \exp \frac{\Psi_s - 2\Phi_F - V}{U_T}. \tag{5.10}$$

where $U_T = kT/q$ is the thermodynamic voltage, and $\Phi_F$ is the Fermi potential of the substrate, which depends on its doping concentration and on the intrinsic carrier concentration of silicon $n_i$ according to

$$\Phi_F = U_T \ln \frac{N_b}{n_i}. \tag{5.11}$$

Hence, as soon as $Q_i$ starts to dominate in strong inversion, the surface potential $\Psi_s$ only increases very slowly since the total charge $Q_{si}$ is limited by the limited field in the oxide. For this reason, the surface potential can be assumed to be independent of $V_G$ and to be given by [52]

$$\Psi_s = V + \Psi_0, \tag{5.12}$$

where

$$\Psi_0 = 2\Phi_F + \text{a few } U_T. \tag{5.13}$$

In this approximation, the threshold function $V_{TB}(V)$ is therefore identical to $V_{TB}(\Psi_s)$, but its vertical axis ($V = 0$) is shifted by $\Psi_0$, as illustrated in Figure 5.3(b). For $V = 0$, $V_{TB}$ takes the particular value $V_{T0}$ called the *equilibrium threshold voltage*, or for short *threshold voltage*. Its expression is obtained by replacing $\Psi_s$ by $\Psi_0$ in (5.7):

$$V_{T0} = V_{FB} + \Psi_0 + \Gamma_b\sqrt{\Psi_0}. \tag{5.14}$$

This *bias-independent* device parameter corresponds to the threshold voltage $V_T$ for $V_S = 0$ used in other models.

As shown by Figure 5.3(b), the slope $n$ for $V > 0$ can be considered constant and will be called the *slope factor*. Now in this approximation, for a

particular value of the gate voltage $V_G$, $Q_i = 0$ for a particular value $V_P$ of $V$ called the *pinch-off voltage*. Inspection of the figure shows that $V_P$ is related to $V_G$ by

$$V_P = \frac{V_G - V_{T0}}{n}, \tag{5.15}$$

and that $Q_i$ can be expressed as

$$-Q_i/C_{ox} = n(V_P - V), \tag{5.16}$$

where the slope factor $n$ is given by (5.8). It is usually convenient to evaluate it at $\Psi_s = \Psi_0 + V_P$, giving

$$n = 1 + \frac{\Gamma_b}{2\sqrt{\Psi_0 + V_P}}. \tag{5.17}$$

### 5.2.3 General Case

By differentiation of (5.10) we obtain

$$U_T \frac{\mathrm{d}Q_i}{Q_i} = \mathrm{d}\Psi_s - \mathrm{d}V. \tag{5.18}$$

Now, introducing (5.9) in (5.18) to eliminate $\mathrm{d}\Psi_s$ results in

$$\frac{\mathrm{d}V}{U_T} = \frac{\mathrm{d}Q_i}{nU_T C_{ox}} - \frac{\mathrm{d}Q_i}{Q_i}. \tag{5.19}$$

In should be pointed-out that the assumption of constant $n$ amounts to a *linearization* of the charge-potential relationship.

Further calculations can be simplified by normalizing voltage and charge according to

$$v = V/U_T \quad \text{and} \quad q_i = Q_i/Q_{spec}, \tag{5.20}$$

where

$$Q_{spec} = -2nC_{ox}U_T. \tag{5.21}$$

Equation (5.19) then becomes

$$-\mathrm{d}v = 2\mathrm{d}q_i + \mathrm{d}q_i/q_i. \tag{5.22}$$

Integrating both sides of this equation yields

$$\text{constant} - v = 2q_i + \ln q_i. \tag{5.23}$$

Now, in strong inversion, $\ln q_i \ll 2q_i$. The comparison with (5.16) after de-normalization shows that the constant in (5.23) is equal to $v_p = V_P/U_T$, hence

$$v_p - v = 2q_i + \ln q_i, \tag{5.24}$$

which is the general relationship between voltages and mobile inverted charge density. This relation is plotted in Figure 5.4 but, in the general case, it cannot be inverted to obtain the charge from the voltages.

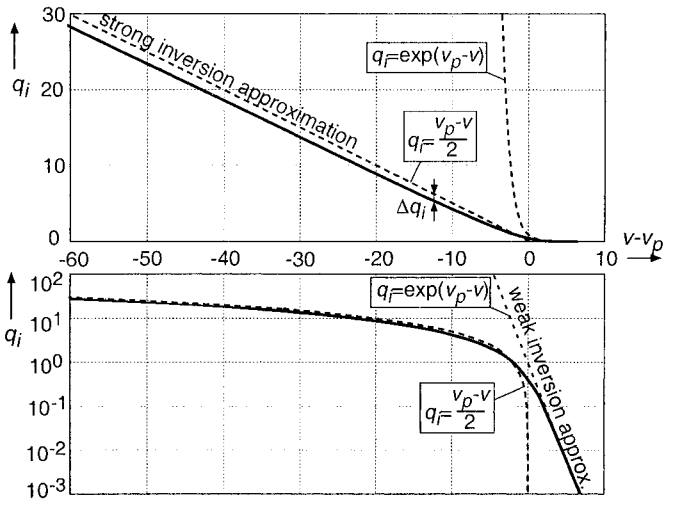



**Fig. 5.4.** Normalized inverted charge vs. normalized channel voltage.

### 5.2.4 Approximation in Weak Inversion

In weak inversion, $2q_i \ll |\ln q_i|$, therefore (5.24) can be approximated by

$$q_i = \exp(v_p - v) \quad \text{or} \quad -Q_i/C_{ox} = 2nU_T \exp\frac{V_P - V}{U_T}, \tag{5.25}$$

where the $V_P$ is the pinch-off voltage defined by (5.15) in the strong inversion approximation.

## 5.3 Drain Current and Modes of Operation

### 5.3.1 Charge-Current Relationship

The drain current $I_D$ is the sum of conduction and diffusion currents given by [108]

$$I_D = \mu W \left( \underbrace{-Q_i \frac{\mathrm{d}\Psi_s}{\mathrm{d}x}}_{\text{conduction}} \underbrace{+U_T \frac{\mathrm{d}Q_i}{\mathrm{d}x}}_{\text{diffusion}} \right), \tag{5.26}$$

where $\mu$ is the carrier mobility, and $x$ is the position along the channel starting from the source side. Now, by introducing (5.18), this relation becomes

$$I_D = \mu W(-Q_i)\frac{\mathrm{d}V}{\mathrm{d}x}. \tag{5.27}$$

Assuming constant mobility, integrating this equation along the channel gives

$$I_D = \beta \int_{V_S}^{V_D} \frac{-Q_i}{C_{ox}} \mathrm{d}V, \tag{5.28}$$

where

$$\beta = \mu C_{ox} W/L \tag{5.29}$$

is the transfer parameter of the transistor. Hence, the drain current is proportional to the integral from $V = V_S$ to $V = V_D$ of the $Q_i(V)$ function obtained in Section 5.2, as represented in Figure 5.5(a).

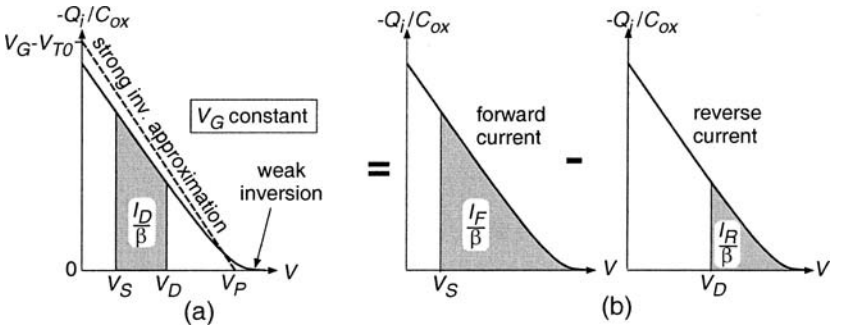



**Fig. 5.5.** (a) Drain current; (b) decomposition in forward and reverse components.

### 5.3.2 Forward and Reverse Components

Since $Q_i$ tends to zero for $V$ tending to infinity, the integral (5.28) can be rewritten as

$$I_D = \underbrace{\beta \int_{V_S}^{\infty} \frac{-Q_i}{C_{ox}} \mathrm{d}V}_{\text{forward current } I_F} - \underbrace{\beta \int_{V_D}^{\infty} \frac{-Q_i}{C_{ox}} \mathrm{d}V}_{\text{reverse current } I_R} = I_F - I_R. \tag{5.30}$$

Hence, as illustrated by Figure 5.5(b), the drain current can be expressed as the difference between a *forward current* $I_F$ and a *reverse current*$I_R$. $I_F$ depends on $V_G$ and $V_S$, but *not on* $V_D$, whereas $I_R$ depends on $V_G$ and $V_D$, but *not on* $V_S$. Furthermore, according to (5.30), $I_F(V_S) \equiv I_R(V_D)$ : $I_F$ and $I_R$ are indeed two values of the same function of $V$. Thus, the drain current is the *superposition* of *independent* and *symmetrical* effects of the source and drain voltages [109].

### 5.3.3 General Current Expression

By introducing the normalized variable defined by (5.20), the forward or reverse components defined by (5.30) can be expressed in normalized form as

$$i_{f,r} = \frac{I_{F,R}}{I_{spec}} = \int_{v_{s,d}}^{\infty} q_i \mathrm{d}v, \tag{5.31}$$

where $v_{s,d}$ is the source or drain voltage normalized to $U_T$, and

$$I_{spec} = 2n\mu C_{ox}\frac{W}{L}U_T^2 = 2n\beta U_T^2 \tag{5.32}$$

is the *specific current* of the transistor.

Introducing (5.22) into (5.31) yields

$$i_{f,r} = \int_0^{q_{s,d}} (2q_i + 1)\mathrm{d}q_i = q_{s,d}^2 + q_{s,d}, \tag{5.33}$$

where $q_{s,d}$ is the value of the normalized charge density $q_i$ at the source end or at the drain end of the channel. Solving this equation for the charge gives

$$q_{s,d} = \frac{\sqrt{1 + 4i_{f,r}} - 1}{2}. \tag{5.34}$$

Using the voltage-charge relationship (5.24), we obtain finally

$$v_p - v_{s,d} = \sqrt{1 + 4i_{f,r}} + \ln\left(\sqrt{1 + 4i_{f,r}} - 1\right) - (1 + \ln 2) \tag{5.35}$$

This general expression of the current-voltage relationship is plotted in Figure 5.6 (curve a), by calculating the voltages from the current. This figure also shows the approximation in strong inversion (curve b for $i_{f,r} \gg 1$) and that in weak inversion (curve c for $i_{f,r} \ll 1$). Remembering that $i_d = i_f - i_r$ and $v_p = (v_g - v_{t0})/n$, (5.35) models the static transistor characteristics from weak to strong inversion with only 3 model parameters (besides $U_T$ used to normalize all voltages): the threshold voltage $V_{T0}$, the slope factor $n$ and the specific current $I_{spec}$ (used to normalized the currents) that includes the transfer parameter $\beta$ according to (5.32).

Now, since (5.35) cannot be inverted to calculate the current from the voltages, it can be approximated by [47, 48, 110]

$$i_{f,r} = \ln^2\left(1 + \exp\frac{v_p - v_{s,d}}{2}\right), \tag{5.36}$$

which is also plotted in Figure 5.6 (curve d).

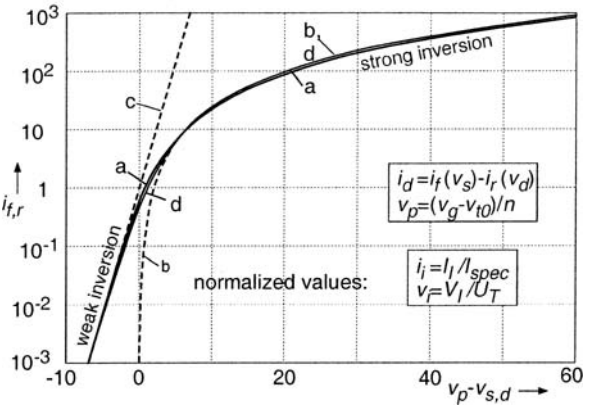



**Fig. 5.6.** Normalized forward or reverse current ; (a) from charge model (5.33); (b) strong inversion approximation; (c) weak inversion approximation; d) from interpolation formula (5.36).

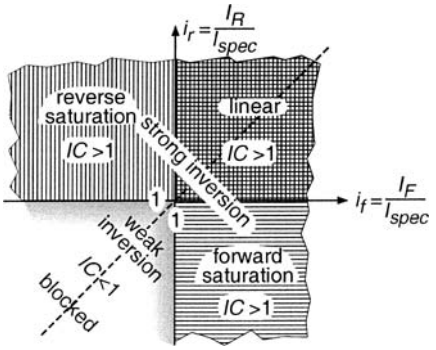



**Fig. 5.7.** Modes of operation of a MOS transistor.

### 5.3.4 Modes of Operation and Inversion Coefficient

The various possible modes of operation of a transistor depend on the values of $I_F$ and $I_R$. They can be described in the $(i_f, i_r)$ plane represented in Figure 5.7. Although weak and strong inversion are separated by a regime of moderate inversion (c.f. Figure 5.6), we shall simplify the discussion by assuming that $i_{f,r} = 1$ ($I_{F,R} = I_{spec}$) represents the limit between weak and strong inversion for each of the two components.

If $i_f > 1$ and $i_r > 1$, then both components are in strong inversion and the whole channel is strongly inverted. The transistor is said to be in *linear* mode.

If $i_f > 1$ but $i_r < 1$, the reverse component is negligible and the current does not increase anymore with the drain voltage: the transistor is still in strong inversion, but in *forward saturation*. If $i_r > 1$ but $i_f < 1$, the for-

ward component is negligible and the current does not increase anymore with the source voltage: the transistor is still in strong inversion, but in *reverse saturation* ($i_d < 0$).

If $i_f < 1$ and $i_r < 1$, then both components are in weak inversion, and the *whole channel is only weakly inverted.* The transistor is said to operate in *weak inversion.*

The global level of inversion of the transistor can be characterized by its *inversion coefficient* $IC$ defined by

$$IC = \max(i_f, i_r) \tag{5.37}$$

The transistor operates in weak inversion for $IC \ll 1$, in strong inversion for $IC \gg 1$, and in moderate inversion for $IC \cong 1$.

### 5.3.5 Output Characteristics and Saturation Voltage

In forward mode, $IC = i_f$. If the drain voltage is increased, the reverse current is progressively decreased until it becomes negligible. Therefore, the drain current $i_d = i_f - i_r$ increases until it saturates at the value $i_{dsat} = i_f$. The drain to source voltage $v_{ds} = v_d - v_s$ can be obtained by calculating $(v_p - v_s) - (v_p - v_d)$ from (5.35), with $i_r = IC(1 - i_d/i_f)$. This yields

$$v_{ds} = \sqrt{1+4IC} - \sqrt{1+4IC(1-i_d/i_f)} + \ln \frac{\sqrt{1+4IC}-1}{\sqrt{1+4IC(1-i_d/i_f)}-1}. \tag{5.38}$$

This expression is plotted in Figure 5.8 for several values of the inversion coefficient $IC$ ranging from weak inversion to strong inversion.

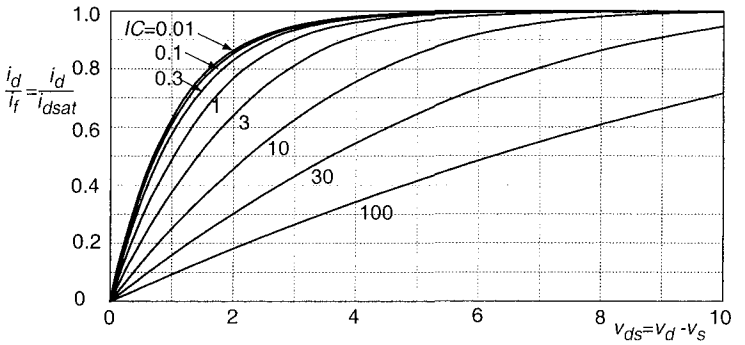



**Fig. 5.8.** Output characteristics for increasing values of inversion coefficient $IC$.

As can be seen, weak inversion ($IC \ll 1$) provides the minimum possible saturation voltage, since the drain current saturates for $V_{DS} \approx 5U_T$. This is why it is intrinsically associated with very *low voltage circuit* design. The

saturation voltage starts increasing in moderate inversion and tends to $V_P - V_S = 2U_T\sqrt{IC}$ in very strong inversion.

### 5.3.6 Weak Inversion Approximation

The components of the drain current in weak inversion can be obtained by integrating the charge density given by (5.25) according to (5.31). This yields

$$i_{f,r} = \exp(v_p - v_{s,d}) \quad \text{or} \quad I_{F,R} = I_{spec} \exp \frac{V_P - V_{S,D}}{U_T} \tag{5.39}$$

which is plotted as curve $c$ in Figure 5.6. It is only exact for $IC = \max(i_f, i_r) \ll 1$. Then by introducing the expression (5.15) of $V_P$:

$$I_D = I_{spec} \exp \frac{V_G - V_{T0}}{nU_T} \left( \exp \frac{-V_S}{U_T} - \exp \frac{-V_D}{U_T} \right). \tag{5.40}$$

The first term in the parenthesis belongs to the forward mode, the second term belongs to the reverse mode. The latter becomes negligible as soon as $V_D$ exceeds $V_S$ by a few $U_T$, as seen in Figure 5.8. The slope factor $n$ in the common exponential term represents the effect of the capacitive divider formed by the oxide capacitance $C_{ox}$ and the depletion capacitance $C_d$.

This equation can be rewritten by lumping the dependencies on $I_{spec}$ and $V_{T0}$

$$I_D = I_{D0} \exp \frac{V_G}{nU_T} \left( \exp \frac{-V_S}{U_T} - \exp \frac{-V_D}{U_T} \right), \tag{5.41}$$

where

$$I_{D0} = I_{spec} \exp \frac{-V_{T0}}{nU_T} \tag{5.42}$$

is the residual drain current in saturation for $V_G = V_S = 0$, or channel "leakage" current of CMOS digital circuits. This current increases exponentially when the threshold voltage is reduced.

## 5.4 Small-Signal Model

### 5.4.1 Transconductances

The DC small-signal equivalent circuit of the 4-terminal transistor is shown in Figure 5.9. Small variations of $V_S$, $V_D$ and $V_G$ produce small variations of the drain current proportional to the respective transconductances $G_{ms}$ (source transconductance), $G_{md}$ (drain transconductance) and $G_m$ (gate transconductance).

The drain current $I_D$ depends on $V_S$ through $I_F$ only and on $V_D$ through $I_R$ only. Hence, the source and drain transconductances are obtained by differentiating (5.35) and then inverting and de-normalizing the result. This gives

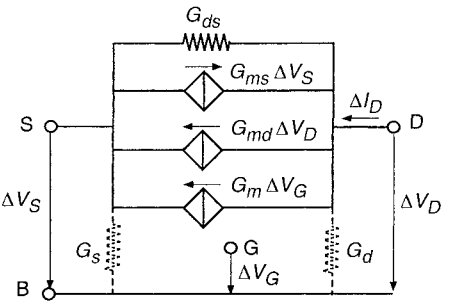



**Fig. 5.9.** DC small-signal equivalent circuit.

$$G_{ms,d} = \frac{I_{spec}}{2U_T}\left(\sqrt{1 + 4I_{F,R}/I_{spec}} - 1\right). \tag{5.43}$$

At small and large current, it tends to

$$G_{ms,d} = \frac{I_{F,R}}{U_T} \text{ (weak inversion)}, \tag{5.44}$$

$$G_{ms,d} = \frac{\sqrt{I_{F,R} \cdot I_{spec}}}{U_T} = \sqrt{2n\beta I_{F,R}} \text{ (strong inversion)}. \tag{5.45}$$

Now, since $I_{F,R}$ is a function of $V_P - V_{S,D}$, the gate transconductance is given by

$$G_m = \frac{\partial I_D}{\partial V_G} = \frac{\partial(I_F - I_R)}{\partial V_P} \cdot \frac{\partial V_P}{\partial V_G} = -\left(\frac{\partial I_F}{\partial V_S} - \frac{\partial I_R}{\partial V_D}\right)\frac{1}{n} = \frac{G_{ms} - G_{md}}{n}. \tag{5.46}$$

In (forward) saturation $I_R \ll I_F$, hence $G_{md} \ll G_{ms}$ and $G_m = G_{ms}/n$.

Since the transconductance always increases with the current, it is interesting to express the transconductance-to-current ratio. Dividing (5.43) by $I_{F,R}$ gives

$$\frac{G_{ms,d}}{I_{F,R}} = \frac{1}{U_T} \cdot \frac{2}{\sqrt{1 + 4I_{F,R}/I_{spec}} + 1}. \tag{5.47}$$

This result is plotted in Figure 5.10 for $G_{ms}$ (curve a). In weak inversion $G_{ms}/I_F$ reaches the maximum possible value $1/U_T$. It is reduced by about 40 % at $IC = 1$ and tends to $1/(U_T\sqrt{IC})$ in strong inversion.

Curve b in the same figure has been obtained by differentiating the approximation (5.36) of the drain current, resulting in

$$\frac{G_{ms,d}}{I_{F,R}} = \frac{1}{U_T} \cdot \frac{1 - \exp\left(-\sqrt{I_{F,R}/I_{spec}}\right)}{\sqrt{I_{F,R}/I_{spec}}}. \tag{5.48}$$

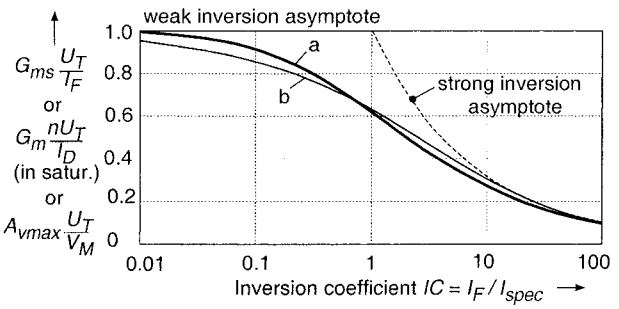



**Fig. 5.10.** Variation of the transconductance-to-current ratio with the inversion coefficient; (a) Charge-based model (5.47); (b) Approximation (5.48) obtained from (5.36).

### 5.4.2 Residual Conductance in Saturation and Maximum Voltage Gain

The conductance $G_{ds}$ included in the small-signal equivalent circuit of Figure 5.9 represents the residual output conductance in saturation that is due to channel length modulation by the drain voltage (in forward mode). This conductance limits the maximum possible voltage gain in common-gate configuration to

$$A_{vmax} = G_{ms}/G_{ds}, \tag{5.49}$$

whereas this maximum value is reduced by $n$ in common-source configuration.

In forward mode, $G_{ds}$ is approximately proportional to the saturation current $I_F$ according to

$$G_{ds} = I_F/V_M, \tag{5.50}$$

where $V_M$ is the *channel length modulation voltage*, proportional to the channel length. Therefore, Figure 5.10 also represents the variation of $A_{vmax}$ with the inversion coefficient, with a maximum value $V_M/U_T$ in weak inversion.

Conductances $G_d$ and $G_s$ included in the equivalent circuit of Figure 5.9 are the differential conductances of the reverse biased drain and source junctions. Their value is small but independent of the drain current. They may therefore become larger than $G_{ds}$ at very small current, thereby dominating the output conductance and limiting the voltage gain.

### 5.4.3 Small-Signal AC Model

For frequencies below $\mu U_T/L^2$, the dynamic behavior of the transistor can be modeled by means of lumped capacitors added to the small-signal model.

Intrinsic capacitors are due to the charge stored in the channel. Each of them is a bias-dependent fraction of the total gate oxide capacitor $WLC_{ox}$.

Extrinsic capacitors are those of the source and drain junctions, and the gate overlap capacitors to source and drain diffusions. Their values are essentially independent of the drain current.

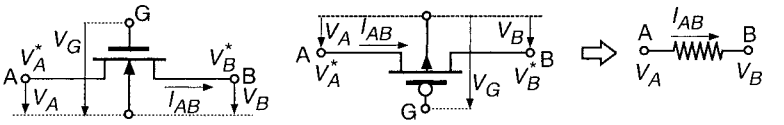
In *weak inversion*, most of the intrinsic capacitances are negligible if the channel is not very long. The only non-negligible value is the gate-to-substrate capacitance produced by the series connection of $C_{ox}$ and the depletion capacitance $C_d$. It is given by

$$C_{GB} = \frac{n-1}{n} \cdot WLC_{ox}, \tag{5.51}$$

and is smaller that the total gate oxide capacitance.

## 5.5 Transistor Operated As a Pseudo-Resistor

The appellations source and drain for the two diffused regions forming a transistor are purely functional (the source and drain functions being inverted if the voltage is inverted) and can normally not be identified in the structure itself. Let us therefore name them simply A and B to emphasize the symmetry of the device, as shown in Figure 5.11.




(a) N-channel pseudo-resistor (b) P-channel pseudo-resistor (c) Resistor

**Fig. 5.11.** Pseudo-resistors: (a) N-channel; (b) P-channel; (c) prototype resistor.

The expression (5.40) of the drain current in weak inversion then becomes

$$I_{AB} = \pm I_{spec} \exp \frac{V_G - V_{T0}}{nU_T} \left( \exp \frac{-V_A}{U_T} - \exp \frac{-V_B}{U_T} \right), \tag{5.52}$$

which is a linear relationship between the current and the exponential function of the voltages. The $-(+)$ sign applies to the N-channel (P-channel) transistor. Let us define *pseudo-voltages* [111, 112, 109] corresponding to $V_A$ and $V_B$ as

$$V_{A,B}^\star = \pm V_0 \exp \frac{-V_{A,B}}{U_T}, \tag{5.53}$$

and a *pseudo-conductance*

$$G^\star = \frac{1}{R^\star} = \frac{I_{spec}}{V_0} \exp \frac{V_G - V_{T0}}{nU_T}, \tag{5.54}$$

where $V_0$ is any positive voltage. Then (5.52) becomes

$$I_{AB} = G^\star(V_A^\star - V_B^\star) = (V_A^\star - V_B^\star)/R^\star, \tag{5.55}$$

which is a linear *pseudo-Ohm's law*. Hence, by similarity with a network of linear resistors, any network obtained by interconnecting transistors by their source and drain is *linear for currents* (and for pseudo-voltages, but *not* for voltages). In other words, at each node of such a network, the current splits linearly in the various branches [113].

Thus, any prototype made of real linear resistors may be converted to a pseudo-resistor network made of transistors only, by replacing each resistor by the source-drain port of a transistor. Moreover, each pseudo-resistance is controllable by the gate voltage of the transistor, according to (5.54). It must be noticed that the general principle is also valid in moderate and strong inversion [113, 111, 112, 109], but then the gate voltages must be identical for all transistors. Therefore, the possibility to control $R^\star$ by the gate voltage only exists in weak inversion ($I_F$ and $I_R \ll I_{spec}$).

It must be pointed out that no voltage should be applied or measured in such current-mode circuits. Therefore, the value of $V_0$ in (5.53) and (5.54) is irrelevant.

If one side of the transistor is saturated, the corresponding pseudo-voltage is zero. It is a *pseudo-ground* $0^\star$ that corresponds to a ground ($V = 0$) in the resistor prototype. Notice that, according to (5.53) pseudo-voltages for a N-channel (P-channel) are always negative (positive).

When used in weak inversion, the concept of pseudo-resistors is only degraded by channel-length modulation and short-channel effects. Examples of application will be given in Section 8.5.

## 5.6 Noise

### 5.6.1 Noise model

Noise is introduced in the model by adding two noise sources to the noiseless transistor as shown in Figure 5.12. The channel noise is modeled by a noise

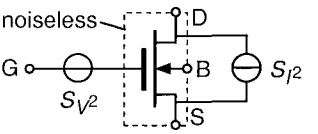



**Fig. 5.12.** Modeling the noise by two noise sources of power spectral densities $S_{I^2}$ and $S_{V^2}$.

current source of power spectral density $S_{I^2}$ (dimension $A^2/Hz$). The gate

interface noise is best modeled by a voltage noise source of power spectral density $S_{V^2}$ (dimension $V^2/Hz$).

### 5.6.2 Channel Noise

Channel noise is the most fundamental noise. In weak inversion, it is a shot noise [114] associated with the barrier that controls the amount of carrier diffusing in the channel. It can be shown that independent noise sources are associated with the source and drain barriers and that their respective spectral densities are $2qI_F$ and $2qI_R$ [52]. Hence, in weak inversion

$$S_{I^2} = 2q(I_F + I_R) = 2qI_F \left(1 + \exp\frac{-V_{DS}}{U_T}\right), \tag{5.56}$$

which is plotted in Figure 5.13.

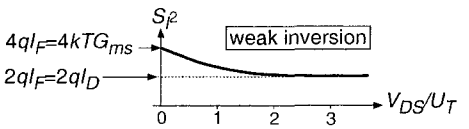



**Fig. 5.13.** Power spectral density of channel noise in weak inversion.

For $V_{DS} \gg U_T$ (saturated weak inversion), $I_D = I_F$ and

$$S_{I^2} = 2qI_D, \tag{5.57}$$

whereas for $V_{DS} = 0$, $I_D = I_F - I_R = 0$ and

$$S_{I^2} = 4qI_F = 4kT \cdot G_{ms}. \tag{5.58}$$

The spectral density is doubled when $V_{DS}$ is reduced from saturation to zero. For $V_{DS} = 0$, it is equal to that of the channel conductance $G_{ms} = G_{md}$. The latter result is also true for strong inversion, but $S_{I^2}$ is then reduced to 2/3 (instead of 1/2) when saturation is reached, for $V_{DS} \geq V_P$. Hence, by introducing expression (5.45) of the transconductance in strong inversion, in saturation ($I_D = I_F$)

$$S_{I^2} = 4kT \cdot \frac{2}{3} \cdot \frac{\sqrt{I_F I_{spec}}}{U_T} = 2qI_D \cdot \frac{4}{3\sqrt{IC}}, \tag{5.59}$$

which is decreased by increasing the inversion coefficient $IC$.

### 5.6.3 Interface Noise

The interface noise is due to a combination of carrier number fluctuation due to interface traps and surface mobility fluctuation. Its spectral density can be globally modeled by [110]

$$S_{V^2} = \frac{4kT\rho}{f \cdot WL} \tag{5.60}$$

in order to express its dependency on the inverse of both the frequency $f$ and the channel area $WL$. The parameter $\rho$ depends on the process and on the gate oxide capacitance ($\rho \propto C_{ox}^{-\alpha}$, with $1 < \alpha < 2$). Its value is somewhat dependent on the inversion coefficient, with a flat minimum around $IC = 1$ [52].

### 5.6.4 Total Noise

In saturation, the spectral density of the total *output* current noise for constant gate voltage is

$$S_{I_D^2} = S_{I^2} + G_m^2 S_{V^2}. \tag{5.61}$$

At a given current $I_D$, $S_{I_D^2}$ is *maximum* in weak inversion since both $S_{I^2}$ and $G_m$ are maximum. But the spectral density of the total *input* referred noise voltage for constant drain current is

$$S_{V_G^2} = S_{V^2} + \frac{S_{I^2}}{G_m^2} = S_{V^2} + \frac{4kT\gamma_n}{G_m} = 4kTR_n, \tag{5.62}$$

where $\gamma_n$ is the thermal noise excess factor, equal to $n/2$ in weak inversion and to $2n/3$ in strong inversion, and $R_n$ is the input referred *equivalent noise resistance*. At a given current $I_D$, $S_{V_G^2}$ is *minimum* in weak inversion since $G_m$ is maximum. The equivalent noise resistance is obtained by introducing (5.60) in (5.62):

$$R_n = \frac{\gamma_n}{G_m} + \frac{\rho}{f \cdot WL}. \tag{5.63}$$

## 5.7 Temperature effects

The dependence on temperature of the transistor characteristics can be modeled through that of the main parameters $V_{T0}$, $n$ and $\beta$ (and therefore also $I_{spec} = 2n\beta U_T^2$).

Essentially $V_{T0}$ and $n$ variation is due to $\Phi_F$ (Equation 5.11). It includes the direct effect of $U_T$ but also the variation of the intrinsic carrier according to [107]

$$n_i \propto T^{3/2} \exp \frac{-V_{gap}}{2U_T} \tag{5.64}$$

where $T$ is the absolute temperature and $V_{gap}$ the voltage corresponding to the band gap of silicon. By linearizing $V_{gap}(T)$ at some ambient temperature $T_0$, (5.64) becomes

$$n_i \propto T^{3/2} \exp \frac{-(V_{G0} - aT)}{2U_T} \propto T^{3/2} \exp \frac{-V_{G0}}{2U_T} \tag{5.65}$$

where $-a$ is the slope of the tangent to $V_{gap}(T)$ at $T = T_0$ and $V_{G0}$ is the value obtained by extrapolating this tangent to $T = 0$, called the *extrapolated band gap voltage*. Its value of about 1.2 V is only slightly dependent on $T_0$ due to the very small curvature of $V_{gap}(T)$. By neglecting the $T^{3/2}$ dependence with respect to the exponential dependence this gives finally

$$n_i = n_{i\infty} \cdot \exp \frac{-V_{G0}}{2U_T} \tag{5.66}$$

where $n_{i\infty}$ is a constant (extrapolated value of $n_i$ at $T = T_0$). Introducing this expression in (5.11) gives

$$\Phi_F = \frac{V_{G0}}{2} - U_T \ln \frac{n_{i\infty}}{N_b} \quad \text{and} \quad \frac{\mathrm{d}\Phi_F}{\mathrm{d}T} = -\frac{1}{T}\left(\frac{V_{G0}}{2} - \Phi_F\right). \tag{5.67}$$

Now, if the gate is highly doped, the flat-band voltage $V_{FB}$ defined by (5.5) includes $-\Phi_F$ (inside $\Phi_{ms}$), which makes it dependent on temperature. If we neglect the small difference in (5.13) between $\Psi_0$ and $2\Phi_F$, the temperature dependence of $V_{T0}$ obtained from (5.14) is

$$\frac{\mathrm{d}V_{T0}}{\mathrm{d}T} = \left(1 + \frac{\Gamma_b}{\sqrt{\Psi_0}}\right)\frac{\mathrm{d}\Phi_F}{\mathrm{d}T} = (2n_0 - 1)\frac{\mathrm{d}\Phi_F}{\mathrm{d}T} \tag{5.68}$$

where $n_0$ is the slope factor $n$ evaluated at $V = 0$ (or $\Psi_s = \Psi_0$). Then, by combination with (5.67):

$$\frac{\mathrm{d}V_{T0}}{\mathrm{d}T} = \frac{n_0 - 0.5}{T}(2\Phi_F - V_{G0}) < 0. \tag{5.69}$$

The temperature coefficient of $V_{T0}$ is always negative, with practical values ranging from $-2.5$ to $-1$ mV/°K.

If $n$ is evaluated at $V = V_P$ ($\Psi_s = \Psi_0 + V_P \cong 2\Phi_F + V_P$), its temperature coefficient is given by

$$\frac{\mathrm{d}n}{\mathrm{d}T} = \frac{\mathrm{d}}{\mathrm{d}\Phi_F}\left(1 + \frac{\Gamma_b}{2\sqrt{2\Phi_F + V_P}}\right)\frac{\mathrm{d}\Phi_F}{\mathrm{d}T} = \frac{n-1}{2T} \cdot \frac{V_{G0} - 2\Phi_F}{2\Phi_F + V_P} > 0. \tag{5.70}$$

The temperature coefficient of $n$ is always positive, with practical values below $10^{-3}$/°K ($< 0.1\%$/°K). It can therefore be neglected for $V_S = 0$. But for $V_S > 0$ it affects the effective threshold $V_{T0} + nV_S$, the temperature coefficient of which is improved by the opposite signs in (5.69) and (5.70).

The variation of $\beta$ with temperature is due to that of the mobility $\mu$ that can be approximated, in the range of ambient temperatures, by

$$\mu \propto T^{-\alpha}, \tag{5.71}$$

where $1.5 < \alpha < 3$ depends on the doping concentration $N_b$. Hence, from the expression (5.29) of $\beta$:

$$\frac{\mathrm{d}\beta/\beta}{\mathrm{d}T} = -\frac{\alpha}{T} < 0. \tag{5.72}$$

The temperature coefficient of $\beta$ is always negative, with practical values ranging from $-0.5$ to $-1\%/°\mathrm{K}$.

It is worth noticing that for $\alpha = 2$, the specific current $I_{spec}$ defined by (5.32) is independent of $T$ since the variation of $\mu$ compensates that of $U_T^2$.

## 5.8 Non-ideal effects

### 5.8.1 Mismatch

Two or more transistors of identical structures implemented on the same chip do not have exactly the same characteristics. This is due to small differences in their dimensions and/or to variations of physical parameters. These differences are reflected in the model parameters essentially as differences of $V_{T0}$, $n$ and $\beta$.

Assuming no temperature difference, the physical parameters possibly responsible for the mismatch of these three model parameters are identified in Table 5.1 on the basis of their respective equations.

**Table 5.1.** Physical parameters affecting the mismatch of the model parameters.

| Equ. | | $Q_{fc}$ | $N_b$ | $C_{ox}$ | $\mu$ | $W$ | $L$ |
|---|---|---|---|---|---|---|---|
| (5.14) | $\Delta V_{T0}$ | • | • | • | | | |
| (5.8) | $\Delta n$ | | • | • | | | |
| (5.29) | $\Delta\beta$ | | | • | • | • | • |

Systematic differences, due for example to gradients of physical parameters across the chip, can be eliminated by means of adequate layout techniques [115]. However, some random mismatch remains due to random fluctuations of the parameters.

It can be shown [116] that the standard deviation of the difference $\Delta P$ of average values of a parameter $P$ across two separate regions of area $WL$ is given by

$$\sigma(\Delta P) = \frac{A_P}{\sqrt{WL}}, \tag{5.73}$$

where $A_P$ is the area proportionality constant for parameter $P$.

The picture is somewhat different for the dimensions $W$ and $L$. Indeed, it can be shown [116, 52] that the standard deviation of their ratio $W/L$, that affects $\Delta\beta$, is given by

$$\sigma(\Delta(W/L)) = \frac{A_{WL}}{\sqrt{WL}} \cdot \sqrt{\frac{1}{W} + \frac{1}{L}}. \tag{5.74}$$

Hence, this contribution to $\beta$-mismatch is proportional to $(WL)^{-3/2}$ for constant $W/L$. It can therefore be made negligible with respect to those of $\Delta C_{ox}$ and $\Delta\mu$ by increasing $W$ and $L$.

As shown by Table 5.1, the mismatch of all three parameters may be correlated through $\Delta C_{ox}$, whereas $\Delta V_{T0}$ and $\Delta n$ may be further (positively) correlated through $\Delta N_b$.

Let us assume that $\Delta V_{T0}$ and $\Delta\beta$ are not correlated ($\Delta C_{ox}$ negligible), and that $\Delta n$ can be neglected. If two saturated transistors are biased at the same source and gate voltages (as in a current mirror), the standard deviation of their relative difference of drain currents is given by

$$\frac{\sigma(\Delta I_D)}{I_D} = \sqrt{\sigma^2(\Delta\beta) + \left[\frac{G_m}{I_D} \cdot \sigma(\Delta V_{T0})\right]^2}, \tag{5.75}$$

since a small $\Delta V_{T0}$ results in $\Delta I_D = -G_m \Delta V_{T0}$.

If, on the contrary, the two drain currents are imposed (or expected) to be equal, then the standard deviation of the difference of gate voltages required to compensate the mismatch of parameters (input offset voltage) is

$$\sigma(\Delta V_G) = \sqrt{\sigma^2(\Delta V_{T0}) + \left[\frac{I_D}{G_m} \cdot \frac{\sigma(\Delta\beta)}{\beta}\right]^2}. \tag{5.76}$$

The different weightings of $\sigma(\Delta\beta)$ and $\sigma(\Delta V_{T0})$ depend on $G_m/I_D$, which in turn depends on the inversion coefficient according to (5.47) [115]. The result is represented in Figure 5.14 for particular values of mismatch.

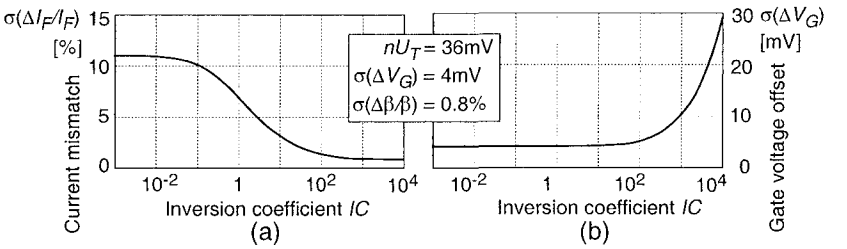



**Fig. 5.14.** Mismatch of (a) drain currents and (b) gate voltages for non-correlated $\sigma(\Delta V_{T0}) = 4mV$ and $\sigma(\Delta\beta)/\beta = 0.8\%$, and for $\sigma(\Delta n)$ negligible.

As can be seen, the matching of currents is very bad in weak inversion (where it tends to $\sigma(\Delta I_D)/I_D = \sigma(\Delta V_{T0})/(nU_T)$). It is improved by increasing $IC$ to reach $\sigma(\Delta\beta)/\beta$ in strong inversion.

On the contrary, the mismatch of gate voltages is limited to $\sigma(\Delta V_{T0})$ in weak inversion, but it is progressively degraded for $IC$ increasing.

In practice, $\Delta n$ is always negligible for $V_S = 0$. But for $V_S > 0$ it affects the effective threshold $V_{T0} + nV_S$, the matching of which is degraded by the positive correlation between $\sigma(\Delta V_{T0})$ and $\sigma(\Delta n)$ (through $\Delta C_{ox}$ and/or $\Delta N_b$).

The minimum possible fluctuation in $N_b$ is due to the limited number of impurities in the depletion volume $WLt_d$ (where $t_d$ is the thickness of the depleted layer). If this effect dominates, then $\sigma(\Delta V_{T0})$ and $\sigma(\Delta n)$ are fully correlated and can be calculated by assuming a Poisson's distribution of impurities, giving [52]

$$A_{VT0} = \sqrt{WL} \cdot \sigma(\Delta V_{T0}) = \frac{1}{C_{ox}} \sqrt[4]{q^3 \epsilon_{si} N_b \Phi_F}, \tag{5.77}$$

and

$$A_n = \sqrt{WL} \cdot \sigma(\Delta n) = \frac{A_{VT0}}{4\Phi_F} = \frac{1}{4C_{ox}} \sqrt[4]{\frac{q^3 \epsilon_{si} N_b}{\Phi_F^3}}. \tag{5.78}$$

Therefore

$$\sigma(V_{T0} + nV_S) = \sigma(V_{T0}) \left(1 + \frac{V_S}{4\Phi_F}\right). \tag{5.79}$$

As a consequence, the mismatch of gate voltages in weak inversion would be doubled for $V_S = 4\Phi_F = 1.2$ to 2 V.

### 5.8.2 Polysilicon Gate Depletion

In calculating the total charge density $Q_{si}$ by (5.4), we have implicitly assumed a constant potential $V_G$ throughout the thickness of the gate electrode, which would always be true if the gate material were metal. It is still true for a polysilicon gate, as long as the thickness of the layer of positive charge $Q_g$ (negative for a P-channel transistor) is so small that the voltage drop across it is negligible. If the gate is P-type (N-type for a P-channel), the layer is still very thin since it is formed by majority carriers. But if the gate is N-type (P-type for a N-channel), $Q_g$ is entirely produced by the depletion layer created at the lower face of the gate.

As long as the doping concentration $N_g$ of the gate is much larger than $N_b$, the voltage drop across this depletion layer is negligible compared to the surface potential. Now, while scaling down process dimensions, $N_b$ must be increased whereas $N_g$ cannot be increased proportionally. The voltage drop in the gate is therefore no longer negligible and $Q_i$ is no longer proportional to the difference between the gate voltage $V_G$ and the threshold function $V_{TB}(\Psi_s)$.

Hence, the diagram of Figure 5.3 no longer applies, and the slope factor of $Q_i(\Psi_s)/C_{ox}$ in (5.9) is no longer the same as that of $V_G(V_P)$ in (5.15). The single slope factor $n$ must be replaced by two distinct slope factors:

$$n_q = \frac{\mathrm{d}Q_i/C_{ox}}{\mathrm{d}\Psi_s} \quad \text{and} \quad n_v = \frac{\mathrm{d}V_G}{\mathrm{d}V_P}. \tag{5.80}$$

If the fixed charge density $Q_{fc}$ is negligible, these slope factors can be related to $n$ by [52]

$$n_q = n - \frac{1}{1 + (n-1)N_g/N_b} < n \tag{5.81}$$

and

$$n_v = n + N_b/N_g > n. \tag{5.82}$$

Moreover, the threshold voltage is increased to

$$V_{T0}^{+} = V_{T0} + \frac{N_b}{N_g} \cdot \Psi_0. \tag{5.83}$$

The situation is more complicated if $Q_{fc}$ is not negligible, since it can no longer be included in a constant flat-band voltage $V_{FB}$ [52].

The pinch-off voltage is now obtained by replacing $n$ by $n_v$ and $V_{T0}$ by $V_{T0}^{+}$ in (5.15). It is thus reduced, since $V_{T0}^{+} > V_{T0}$ and $n_v > n$. All normalized equations developed previously are applicable, provided $n$ is replaced by $n_q$ in the definitions of $Q_{spec}$ (5.21) and $I_{spec}$ (5.32). Both of them are thus reduced (in absolute value) since $n_q < n$.

The source and drain transconductances for a given current are reduced in strong inversion (5.45) but not in weak inversion (5.44). The gate transconductance is always reduced, since $n$ must be replaced by $n_v > n$ in (5.46).

### 5.8.3 Band Gap Widening

Since scaling down a process requires an increase of $N_b$, the electric field required at the surface to produce inversion is increased, requiring an increase of $C_{ox}$. As a consequence of this high field, the lowest allowed energy level for electrons in the conduction band is increased, which corresponds to a widening of the band gap.

It can be shown [52] that this band gap widening effect results in an increase of $\Psi_0$ with respect to its original value (5.13) by

$$\Delta\Psi_0 = A_{qm} \left[2q\epsilon_{si}N_b(\Psi_0 + V_P)\right]^{1/3}, \tag{5.84}$$

with $A_{qm} = 3.53\mathrm{Vm}^{4/3}A^{-2/3}s^{-2/3}$. As shown by Figure 5.15, this increase becomes significant for $N_b > 10^{17}\mathrm{cm}^{-3}$. The threshold voltage is therefore increased according to (5.14).

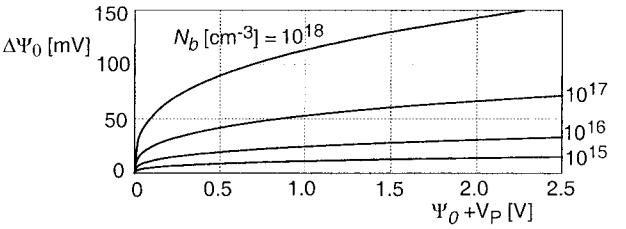



**Fig. 5.15.** Increase of inversion potential $\Psi_0$ due to band gap widening.

Furthermore, band gap widening reduces the specific current $I_{spec}$ by approximately $1+K_cC_{ox}$, with a value of $K_c$ between $10^{-2}$ and $2\times10^{-2}\mu m^2/fF$. Except in weak inversion for which it is irrelevant, this reduction becomes significant for $C_{ox} > 5$ fF/$\mu m^2$.

### 5.8.4 Gate Leakage

When $t_{ox}$ is reduced below 3nm, a gate-to-channel leakage current $I_G$ starts to appear even at low gate voltages as the result of direct tunneling of electrons through the gate oxide. This current increases exponentially with the reduction of $t_{ox}$ due to the increasing tunneling probability, which is also strongly dependent on the voltage across the oxide. $I_G$ is approximately proportional to the saturated drain current $I_F$ in weak inversion, but it increases faster than $I_F$ in strong inversion [52].

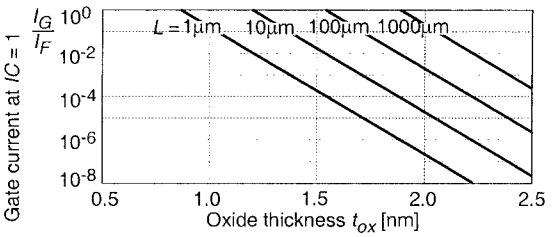



**Fig. 5.16.** Variation with oxide thickness of the relative gate current at $IC = 1$ (typical case).

Since $I_G \propto WL$ whereas $I_F \propto W/L$, the ratio $I_G/I_F$ increases with $L^2$, as shown in Figure 5.16 for a typical case calculated at $IC = 1$ (upper limit of weak inversion). As can be seen, for $t_{ox} > 2.5$nm, $I_G$ is negligible for most applications even for $W = 1$mm.

However, this is only the gate-to-channel leakage, which is limited by the limited number of electrons available in the channel. In practice, the total gate leakage in weak and moderate inversion might be dominated by tunneling in

the gate-source and gate-drain overlap regions, especially for small values of $L$.

### 5.8.5 Drain-Induced Barrier Lowering (DIBL)

Each of the the source-substrate and drain-substrate junctions creates a barrier of potential $\Phi_B$. For no applied voltage and in the flat-band situation ($\Psi_s = 0$) this barrier is given by

$$\Phi_B = U_T \ln \frac{N_{diff} N_b}{n_i^2}, \tag{5.85}$$

where $N_{diff}$ is the doping concentration of the source and drain diffusions. This barrier is increased for $V_{S,D} > 0$. When the surface potential $\Psi_s$ is increased above zero by increasing the gate voltage, the barrier height is decreased, allowing electrons to be injected in the channel. They can then diffuse from source to drain to produce the weak inversion current.

Now, the full potential transition across these barriers occurs along some length. For a long channel, this length is negligible, and the surface potential reaches the value imposed by the gate voltage $V_G$, as shown by Figure 5.17. But the same figure shows that, if the channel is too short, the two barriers invade the whole channel, and $\Psi_s$ can nowhere reach the value imposed by $V_G$. The maximum barrier height is reduced by an amount $\Delta\Psi_{smin}$, which will increase the drain current $I_D$. Moreover, $\Delta\Psi_{smin}$ increases with the drain voltage; $I_D$ therefore depends on $V_D$ even in saturation.

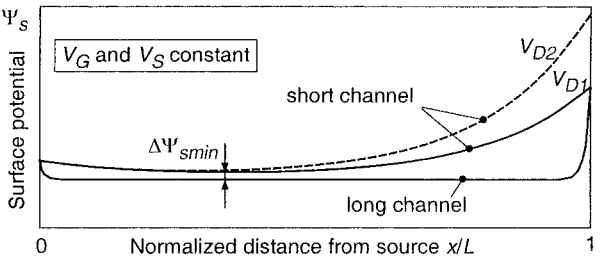



**Fig. 5.17.** Drain-induced barrier lowering (DIBL) by an amount $\Delta\Psi_{smin}$ in a short-channel transistor.

DIBL is mostly affecting weak inversion, since the current is then exponentially dependent on the surface potential. The reduction of the barrier height in weak inversion is given by [52]

$$\Delta\Psi_{smin} = 2\mathrm{e}^{-\lambda/2}\sqrt{(\Phi_B - \Psi_0 - V_P + V_S)(\Phi_B - \Psi_0 - V_P + V_D)}, \tag{5.86}$$

where

$$\lambda = \lambda_0 \left( \frac{\Psi_0}{\Psi_0 + V_P} \right)^{1/4} \quad \text{and} \quad \lambda_0 = L \sqrt{\frac{qN_b}{\epsilon_{si} \Gamma_b}} \Psi_0^{-1/4}. \tag{5.87}$$

It is exponentially dependent on $L$ through $\lambda_0$, whereas its dependence on $V_S$ and $V_D$ appears inside the square root. The current in weak inversion is obtained by modifying (5.39) to

$$I_{F,R} = I_{spec} \exp \frac{V_P - V_{S,D} + \Delta\Psi_{smin}}{U_T}. \tag{5.88}$$

The resulting drain current $I_F - I_R$ is shown in Figure 5.18 for $V_P = -5U_T$ (weak inversion) and $V_P = 0$ (moderate inversion). As can be seen, the current

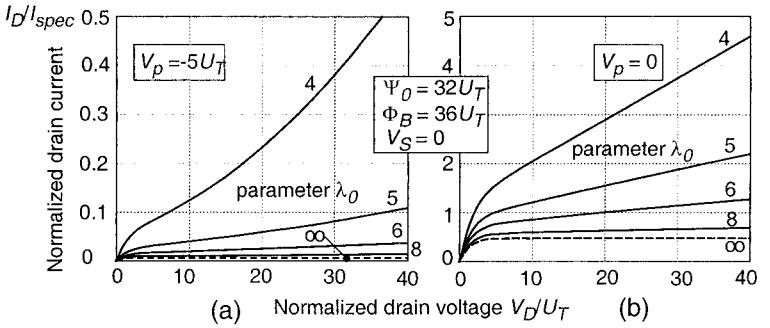



**Fig. 5.18.** Effect of DIBL on output characteristics: (a) example in weak inversion; (b) example in moderate inversion.

in weak inversion is dramatically increased for $\lambda_0 > 6$. The effect is still important in moderate inversion, but it is progressively reduced in strong inversion.

The effect of DIBL on transconductances in weak inversion can be obtained by differentiation of (5.88). The source transconductance becomes

$$G_{ms} = \frac{I_F}{U_T} \left( 1 - \frac{\mathrm{d}\Delta\Psi_{smin}}{\mathrm{d}V_S} \right) = \frac{I_F}{U_T} \left( 1 - \mathrm{e}^{-\lambda/2} \sqrt{\frac{(\Phi_B - \Psi_0 - V_P + V_D)}{(\Phi_B - \Psi_0 - V_P + V_S)}} \right). \tag{5.89}$$

It is reduced by a fraction that increases exponentially with $1/L$. The gate transconductance is more complicated to express, since $\lambda$ also depends on $V_G$ through $V_P$, but it is reduced by about the same fraction.

Most dramatic is the fact that $I_F$ also depends on $V_D$, which produces a residual drain transconductance in saturation $G_{mdsat}$ given by

$$G_{mdsat} = \frac{I_F}{U_T} \cdot \frac{\mathrm{d}\Delta\Psi_{smin}}{\mathrm{d}V_D} = \frac{I_F}{U_T} \cdot \mathrm{e}^{-\lambda/2} \sqrt{\frac{(\Phi_B - \Psi_0 - V_P + V_S)}{(\Phi_B - \Psi_0 - V_P + V_D)}}. \tag{5.90}$$

This effect overcomes the channel shortening effect discussed in Section 5.4.2. $G_{mdsat}$ replaces $G_{ds}$ in the expression (5.49) of the maximum possible voltage gain, which is drastically reduced, as shown in the example of Figure 5.19.

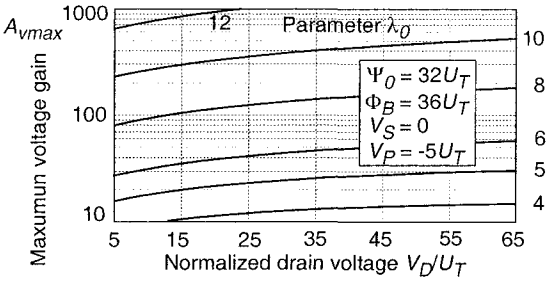



**Fig. 5.19.** Effect of DIBL on maximum voltage gain.

DIBL is the only important short-channel effect on a transistor operated in weak inversion. Indeed, velocity saturation is negligible, since the current is carried by diffusion, and the carriers do not reach a velocity approaching its saturation value. Therefore, a short channel does not increase noise, and the speed keeps increasing with $1/L^2$.