

Analog Integrated Circuits Design using the Inversion Coefficient

Preview Edition

Christian Enz (christian.enz@epfl.ch)

23.06.2026

Table of contents

Licensing	3
1 The Short-channel MOSFET Model	4
1.1 Introduction	4
1.2 Mobility reduction mechanisms	4
1.3 Velocity saturation	5
1.3.1 Velocity versus electric field models	5
1.3.2 Effect of velocity saturation on the drain current	7
1.3.3 Effect of velocity saturation on the transconductances	14
1.3.4 Effect of velocity saturation on the G_m/I_D ratio	16
1.4 Channel length modulation (CLM)	19
1.4.1 Channel length reduction	19
1.4.2 Impact of CLM on the drain current in saturation	20
1.4.3 Output conductance in saturation due to CLM	22
1.4.4 Experimental validation	23
1.5 Drain induced barrier lowering (DIBL)	26
1.5.1 Physical description	26
1.5.2 Modeling DIBL in the threshold or pinch-off voltage	27
1.5.3 Impact of DIBL on the drain current without VS	28
1.5.4 Impact of DIBL on the transconductances in saturation without VS	34
1.5.5 Output conductance in saturation due to DIBL without VS	36
1.5.6 Impact of DIBL on the transconductances in saturation including VS	37
1.6 Short-channel effects on thermal noise	39
1.6.1 Empirical model of the thermal noise excess factor for short-channel transistors	40
1.7 The simplified EKV model	41
1.7.1 Bulk technologies	42
1.7.2 FDSOI technologies	44
1.7.3 FinFET technologies	45
1.7.4 Comparison on 28 nm bulk, 22 nm FDSOI and 16 nm FinFET	47

Licensing

This document is licensed under the [Creative Commons License CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)



1 The Short-channel MOSFET Model

1.1 Introduction

In the previous Chapter, the channel was assumed to be long, allowing for a one-dimensional analysis using the gradual channel approximation. This Chapter is devoted to the effects appearing when reducing the length of the transistor to dimensions that get close to the depletion width. In such a situation, the one-dimensional approach is no more valid and a two-dimensional analysis is required. Furthermore, the terminal voltages are not scaled down proportionally to the device length reduction, the longitudinal electric field increases beyond a certain critical field above which the velocity starts to saturate. This *velocity saturation* (VS) effect is presented in Section 1.3 using different velocity-field relations. Another short-channel effect that strongly limits the performance (particularly the voltage gain) of analog circuits is the *channel length modulation* (CLM) described in Section 1.4. When the device length gets small the surface potential in the channel region is no more defined uniquely by the vertical field, but becomes influenced by the drain (or source) voltage. This effect is called the *drain induced barrier lowering* (DIBL) and is presented in Section 1.5. Short-channel effects such as VS and CLM not only impact the current, but also the thermal noise. A short-channel thermal noise model including the effects of VS is presented in Section 1.6. Finally, the EKV model in saturation and including VS is called the *simplified EKV model* or sEKV. It is presented in Section 1.7 where the model is compared to measurements made on transistors from various technologies. It is shown that the sEKV, although initially developed for bulk technologies, the sEKV model also works very well for advanced FDSOI and FinFET technologies.

1.2 Mobility reduction mechanisms

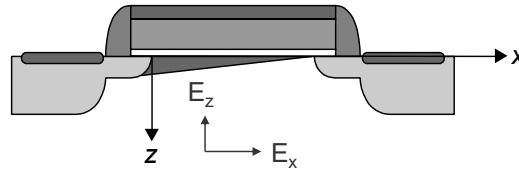


Figure 1.1: MOSFET cross-section showing vertical and longitudinal electric fields $\mathcal{E}_z$ and $\mathcal{E}_x$.

The previous Chapter presented the ideal MOS transistor having a wide and long channel and a constant mobility in the channel region. Of course in a more advanced model, the mobility can no more be considered as constant and various scattering mechanisms need to be accounted for. As illustrated in Figure 1.1, there are mainly two different mobility reduction mechanisms: the first is due to the vertical electric field $\mathcal{E}_z$, while the second is due to the longitudinal electric field $\mathcal{E}_x$ (along the channel direction).

The mobility reduction due to the vertical field $\mathcal{E}_z$ is caused by several scattering mechanisms, namely:

- *Coulomb scattering*: related to the interaction of mobile carrier with ionized impurity atoms (occurs at low field).
- *Phonon scattering*: related to the interaction with the lattice vibrations (occurs at medium field).
- *Surface roughness*: interaction between the mobile carrier and the Si-SiO₂ interface (occurs at high field).

The mobility reduction due to the longitudinal field $\mathcal{E}_x$ is caused by the saturation of the carrier drift velocity. While the mobility due to the vertical field is not restricted to short-channel devices but also affects long-channel devices in strong inversion where the vertical field is maximum. On the other hand, as the channel length gets smaller, the longitudinal field increases and the mobility degrades. The mobility reduction due to the longitudinal field $\mathcal{E}_x$ is therefore an important (if not the most important) short-channel effect.

Among all effects related to the reduction of the device geometry we can distinguish effects that are related to the reduction of the channel length leading to short-channel effects (SCE) which include:

- Velocity saturation (VS),
- Channel length modulation (CLM),
- Charge sharing,
- Drain induced barrier lowering (DIBL) and
- Reverse short-channel effect.

The reduction of the transistor width also leads to narrow-channel effects, whereas the reduction of the oxide thickness gives rise to the gate leakage current.

Among all the SCE listed above, we will consider the velocity saturation (VS), which is probably the most important SCE, the channel length modulation (CLM) and the drain induced barrier lowering (DIBL). They are discussed in the following sections.

1.3 Velocity saturation

1.3.1 Velocity versus electric field models

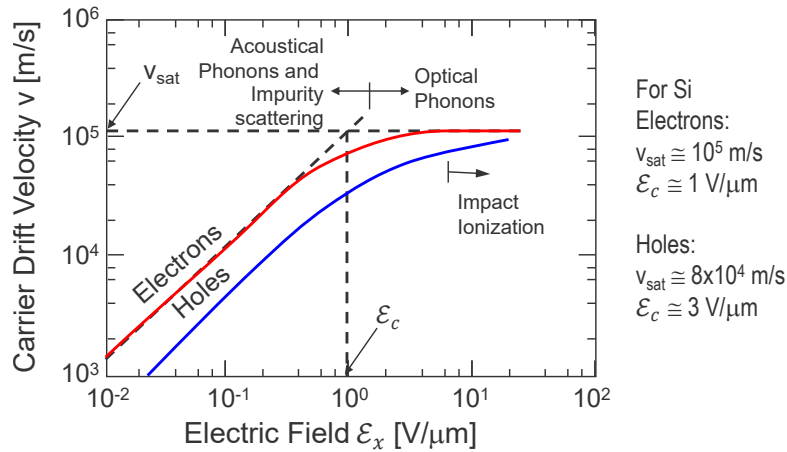


Figure 1.2: The carrier drift velocity versus the longitudinal electric field.

Table 1.1: Typical values of the saturated drift velocity and the critical field for bulk silicon at room temperature.

Type	Saturated velocity v_{sat}	Critical field $\mathcal{E}_c$
Electrons	10^5 m/s	$1 \text{ V}/\mu\text{m}$
Holes	$8 \times 10^4 \text{ m/s}$	$3 \text{ V}/\mu\text{m}$

An important effect that appears when the longitudinal electric field $\mathcal{E}_x$ within the device starts to become large is the saturation of the drift velocity. The drift velocity v_{drift} of electrons and holes in bulk silicon is plotted versus $\mathcal{E}_x$ in Figure 1.2. At low longitudinal electric field, the velocity is proportional

to the electric field with a proportionality factor equal to the mobility μ_z at low longitudinal field. When $\mathcal{E}_x$ becomes larger than the *critical field* $\mathcal{E}_c$, the velocity starts to saturate to a maximum value v_{sat} . The shape of the velocity-field relation in silicon is slightly different for electrons and holes. Typical values for $\mathcal{E}_c$ and v_{sat} at room temperature for electrons and holes are given in Table 1.1. The critical field $\mathcal{E}_c$ is related to the saturated drift velocity and the mobility at low longitudinal field by

$$\mathcal{E}_c \triangleq \frac{v_{sat}}{\mu_z}. \quad (1.1)$$

Note that the mobility at low longitudinal field μ_z actually depends on the vertical field $\mathcal{E}_z$. The larger the vertical field, the smaller μ_z and the larger the critical field, since v_{sat} is constant for a given type of silicon and a given doping.

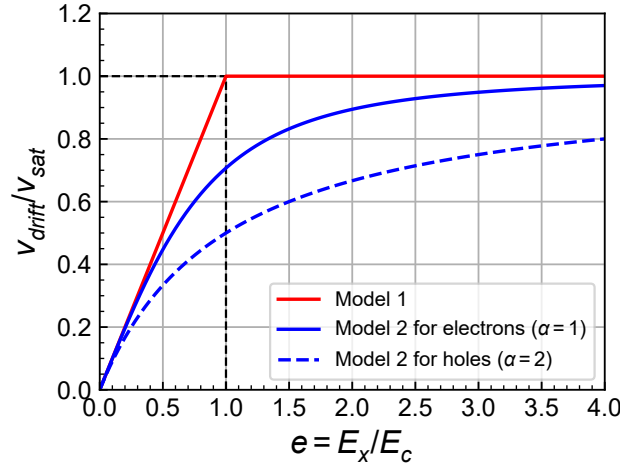


Figure 1.3: Various velocity-field models.

As shown in Figure 1.3, different velocity-field models can be considered to analyze the effect of velocity saturation on the drain current and on the transconductances.

The simplest model (Model 1) to describe the velocity-field dependence is to consider the velocity proportional to the longitudinal field with a constant slope up to the critical field, above which it stays constant and equal to the saturated velocity. This model is described by the piecewise linear model defined by

$$\nu(e) \triangleq \frac{v_{drift}}{v_{sat}} = \begin{cases} e & \text{for } e \leq 1 \\ 1 & \text{for } e > 1, \end{cases} \quad (1.2)$$

where e is the longitudinal electric field $\mathcal{E}_x$ normalized to the critical field $\mathcal{E}_c$

$$e \triangleq \frac{|\mathcal{E}_x|}{\mathcal{E}_c}. \quad (1.3)$$

The corresponding effective mobility μ_{eff} is then simply given by

$$\mu_{eff}(\mathcal{E}_x) \triangleq \frac{v_{drift}}{|\mathcal{E}_x|} = \begin{cases} \mu_z & \text{for } \mathcal{E}_x < \mathcal{E}_c \\ v_{sat}/|\mathcal{E}_x| & \text{for } \mathcal{E}_x \geq \mathcal{E}_c, \end{cases} \quad (1.4)$$

or in a normalized form

$$u(e) \triangleq \frac{\mu_{eff}}{\mu_z} = \begin{cases} 1 & \text{for } e < 1 \\ 1/e & \text{for } e \geq 1. \end{cases} \quad (1.5)$$

The main advantage of this model is obviously its simplicity, allowing to get a first understanding of the physical phenomenon. But on the other hand, it has discontinuous derivatives at $e = 1$ that may induce a bad behavior of the model.

A continuous model (Model 2) also accounting for the difference between the velocity-field characteristics of electrons and holes as illustrated in Figure 1.3 is defined in normalized form as

$$\nu(e) \triangleq \frac{v_{drift}}{v_{sat}} = \frac{e}{(1 + e^\alpha)^{\frac{1}{\alpha}}}. \quad (1.6)$$

where $\alpha = 2$ for electrons and $\alpha = 1$ for holes. Note that this continuous model has been used in many compact models with the approximation that α is equal to unity for both holes and electrons. The corresponding effective mobility is then given by

$$\mu_{eff}(\mathcal{E}_x) \triangleq \frac{v_{drift}}{|\mathcal{E}_x|} = \frac{\mu_z}{1 + |\mathcal{E}_x|/\mathcal{E}_c}, \quad (1.7)$$

which can also be written in normalized form as

$$u(e) \triangleq \frac{\mu_{eff}}{\mu_z} = \frac{1}{1 + e}. \quad (1.8)$$

Although the continuous velocity-field model 2 insures the continuity of the current and the output conductance versus the drain voltage, it requires the electric field to become infinity at the drain for the velocity and hence the current to saturate, which is not physical.

The two velocity-field models presented above are plotted in Figure 1.3. Note also that the shape of the velocity-field curve will strongly affect the current and output conductance versus drain voltage. We will now evaluate the impact of velocity saturation on the drain current and the transconductances. For simplicity we will be using Model 1.

1.3.2 Effect of velocity saturation on the drain current

1.3.2.1 Analysis in strong inversion

We will start evaluating the impact of velocity saturation in strong inversion considering that the current in weak inversion is carried mostly by diffusion and the surface potential gradient along the channel is close to zero and hence the longitudinal electric field is smaller than the critical field. We also will ignore the effect of the vertical field on the mobility and consider $\mu_z = \mu_0$ as constant. The drift component of the drain current is proportional to the velocity and can be written as [1]

$$I_D = W \cdot (-Q_i) \cdot v_{drift} = W \cdot (-Q_i) \cdot \mu_{eff} \cdot |\mathcal{E}_x| = W \cdot (-Q_i) \cdot \mu_{eff} \cdot \frac{d\Psi_s}{dx}, \quad (1.9)$$

or using the normalized variables defined above [1]

$$i_d = 2q_i \cdot \frac{\nu}{\lambda_c} = 2q_i \cdot \frac{u \cdot e}{\lambda_c} = u \cdot q_i \cdot \frac{d\psi_s}{d\xi}, \quad (1.10)$$

where the normalized variables i_d , q_i , ψ_s and ξ have their usual meaning. The velocity saturation parameter λ_c accounts for the velocity saturation effect and depends on the transistor length L according to [1]

$$\lambda_c \triangleq \frac{L_{sat}}{L} \quad (1.11)$$

where [1]

$$L_{sat} \triangleq \frac{2\mu_0 \cdot U_T}{v_{sat}} = \frac{2U_T}{\mathcal{E}_c} \quad (1.12)$$

The velocity saturation parameter λ_c corresponds to the parameter ζ defined in [2] [3] as

$$\zeta = \frac{\lambda_c}{2} = \frac{\mu_0 \cdot U_T}{L \cdot v_{sat}} = \frac{U_T}{L \cdot \mathcal{E}_c} = \frac{v_{diff}}{v_{sat}}, \quad (1.13)$$

where v_{diff} is the *diffusion velocity* defined as [2] [3]

$$v_{diff} = \frac{\mu_0 \cdot U_T}{L} = \frac{D_n}{L} \quad (1.14)$$

with $D_n = \mu_0 \cdot U_T$ the diffusivity coefficient for electron in silicon [4]. Parameter ζ can therefore be interpreted as the ratio of the diffusion velocity to the saturated velocity. λ_c and ζ are key parameters for short-channel devices that scale like $1/L$ and hence tend to zero for very long channel devices. Setting them to zero therefore corresponds to ignoring the effect of velocity saturation. Typical values of λ_c for different channel lengths are given in Table 1.2.

Table 1.2: Typical values of the velocity saturation parameter λ_c .

L	1 μm	0.5 μm	0.17 μm	0.1 μm
λ_c	0.05	0.1	0.3	0.5

Using the inversion charge linearization with respect to Ψ_s gives a relation between the surface potential gradient and the inversion charge gradient [1]

$$\frac{d\Psi_s}{dx} = \frac{1}{n \cdot C_{ox}} \cdot \frac{dQ_i}{dx}, \quad (1.15)$$

or in normalized form [1]

$$\frac{d\psi_s}{d\xi} = -2 \cdot \frac{dq_i}{d\xi}. \quad (1.16)$$

Replacing (1.16) into (1.10) results in [1]

$$i_d = -2u \cdot q_i \cdot \frac{dq_i}{d\xi}. \quad (1.17)$$

Compared to the constant mobility expression in strong inversion, the above expression includes now a field-dependent mobility term u which depends on the chosen velocity-field model.

The inversion charge density at the drain $-Q_{iD}$ (or its normalized form q_d) plays a particular role in the presence of velocity saturation. When the longitudinal field $\mathcal{E}_x$ becomes equal to the critical field $\mathcal{E}_c$ right at the drain, the drift velocity saturates right at the drain, $v_{drift} = v_{sat}$ at $x = L$ (or in normalized form $\nu = 1$ at $\xi = 1$). The drain current cannot increase anymore and is then limited to the saturation value $i_{d_{sat}}$ given by (1.10) evaluated at $\xi = 1$ with $\nu = 1$ and $q_i = q_{d_{sat}}$ [1]

$$i_{d_{sat}} \triangleq \frac{2}{\lambda_c} \cdot q_{d_{sat}} = \frac{q_{d_{sat}}}{\zeta}, \quad (1.18)$$

where $q_{d_{sat}}$ is the value of the inversion charge density at the drain which is required to sustain the drain current when the carrier velocity is saturated right at the drain. The main difference compared to the long-channel situation is that the charge density at the drain $-Q_{iD}$ (or its normalized form q_d) does not vanish to zero as it does at the onset of saturation for long-channel devices when $v_d = v_p$, but it has to be finite in order for the current to flow despite the saturated velocity. To account for this saturation of the drain charge, the normalized drain charge in strong inversion has to be modified according to

$$q_d = \begin{cases} \frac{v_p - v_d}{2} & \text{for } v_d < v_{d_{sat}} \\ q_{d_{sat}} & \text{for } v_d \geq v_{d_{sat}}, \end{cases} \quad (1.19)$$

where the *drain saturation voltage* $v_{d_{sat}}$ is the value of the drain voltage at which $\mathcal{E}_x = \mathcal{E}_c$ and hence the current and the drain charge saturate. It is obtained by replacing q_d and v_d in $q_d = (v_p - v_d)/2$ by $q_{d_{sat}}$ and $v_{d_{sat}}$ respectively, resulting in [1]

$$v_{d_{sat}} = v_p - 2q_{d_{sat}}. \quad (1.20)$$

As illustrated in Figure 1.6, due to velocity saturation, $v_{d_{sat}}$ is always smaller than the pinch-off voltage v_p corresponding to the saturation voltage when velocity saturation is not present.

We now will evaluate the impact of velocity saturation in strong inversion using the simplest velocity-field model, namely Model 1. The normalized drain current is obtained as before by integrating (1.10) from source to drain. For $v_d < v_{d_{sat}}$, the longitudinal field remains smaller than the critical field and according to the velocity-field Model 1 relation, there is no saturation of the carrier velocity. The effective mobility is then equal to $\mu_z = \mu_0$ (or in normalized form $u = 1$). The drain current for $v_d < v_{d_{sat}}$ and $q_d > q_{d_{sat}}$ is identical to what was obtained earlier for the long-channel transistor in strong inversion

$$i_d = i_f - i_r, \quad (1.21)$$

with

$$i_f = q_s^2, \quad (1.22)$$

$$i_r = q_d^2, \quad (1.23)$$

where q_s and q_d are the inversion charge densities at the source and drain, respectively. In strong inversion and assuming there is no velocity saturation at the source, they are related to the voltages according to

$$q_s \triangleq q_i(\xi = 0) = \frac{v_p - v_s}{2}, \quad (1.24)$$

$$q_d \triangleq q_i(\xi = 1) = \frac{v_p - v_d}{2}. \quad (1.25)$$

In the case of a long-channel transistor, the drain current saturates when $v_d = v_p$ and the drain charge becomes $q_d = 0$. For a short-channel transistor, when the drain voltage reaches the saturation voltage $v_d = v_{d_{sat}}$, the drain charge does not vanish to zero but saturates to $q_{d_{sat}}$. The reverse current i_r then also saturates to $i_{r_{sat}} = q_{d_{sat}}^2$. The source and drain charges for $v_d = v_{d_{sat}}$ are then given by

$$q_s = \frac{v_p - v_s}{2}, \quad (1.26)$$

$$q_{d_{sat}} = \frac{v_p - v_{d_{sat}}}{2}, \quad (1.27)$$

whereas the forward and reverse currents are given by

$$i_f = q_s^2, \quad (1.28)$$

$$i_{r_{sat}} = q_{d_{sat}}^2. \quad (1.29)$$

The drain current in saturation then becomes

$$i_{d_{sat}} = i_f - i_{r_{sat}} = q_s^2 - q_{d_{sat}}^2. \quad (1.30)$$

In order to express $q_{d_{sat}}$, $i_{r_{sat}}$, $i_{d_{sat}}$ and $v_{d_{sat}}$ in terms of q_s , we can solve (1.30), (1.18), (1.27), (1.28) and (1.29) for $q_{d_{sat}}$, i_f , i_r , $i_{d_{sat}}$ and $v_{d_{sat}}$, resulting in [1]

$$q_{d_{sat}} = \frac{1}{\lambda_c} \cdot \left[\sqrt{1 + (\lambda_c q_s)^2} - 1 \right] = \frac{\lambda_c q_s^2}{1 + \sqrt{1 + (\lambda_c q_s)^2}}, \quad (1.31)$$

$$i_f = q_s^2, \quad (1.32)$$

$$i_{r_{sat}} = q_{d_{sat}}^2 = \frac{1}{\lambda_c^2} \cdot \left[\sqrt{1 + (\lambda_c q_s)^2} - 1 \right]^2 = q_s^2 - \frac{2}{\lambda_c^2} \cdot \left[\sqrt{1 + (\lambda_c q_s)^2} - 1 \right], \quad (1.33)$$

$$i_{d_{sat}} = \frac{2}{\lambda_c} \cdot q_{d_{sat}} = i_f - i_{r_{sat}} = \frac{2}{\lambda_c^2} \cdot \left[\sqrt{1 + (\lambda_c q_s)^2} - 1 \right] = \frac{2q_s^2}{1 + \sqrt{1 + (\lambda_c q_s)^2}}, \quad (1.34)$$

$$v_{d_{sat}} = v_p - 2q_{d_{sat}} = v_p - \frac{2\lambda_c q_s^2}{1 + \sqrt{1 + (\lambda_c q_s)^2}}. \quad (1.35)$$

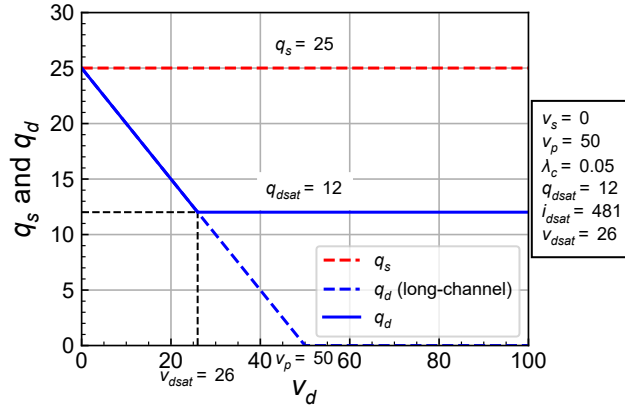


Figure 1.4: Effect of velocity saturation on the charges in strong inversion for Model 1.

The normalized charges q_s and q_d are plotted versus v_d in Figure 1.4 illustrating the saturation of q_d to q_{dsat} for $v_d > v_{dsat}$. While the source charge remains constant because it does not depend on v_d , the drain charge decreases linearly with v_d . For a long-channel transistor it decreases until it reaches zero at $v_d = v_p$. For a short-channel transistor it stops decreasing when v_d reaches v_{dsat} , point at which the longitudinal field $\mathcal{E}_x$ gets equal to the critical field $\mathcal{E}_c$ and the drift velocity starts to saturate

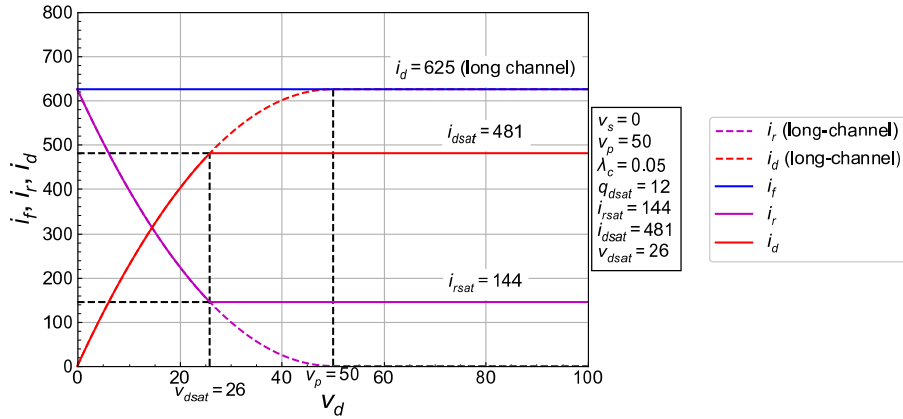


Figure 1.5: Effect of velocity saturation on the currents in strong inversion for Model 1.

Figure 1.5 shows the impact of velocity saturation on the currents. While the forward current i_f remains unchanged, because of velocity saturation, for $v_d > v_{dsat}$, the reverse current saturates to $i_{rsat} = q_{dsat}^2$ instead of zero as it would for a long-channel transistor. The drain current in saturation is therefore i_{rsat} lower than the long-channel current in saturation i_f . Note that, as expected, i_{rsat} tends to zero when there is no velocity saturation (i.e. $\lambda_c = 0$). Notice the discontinuity of the 1st-order derivative of i_d wrt v_d (which is simply the output conductance g_{ds}) at $v_d = v_{dsat}$ due to the discontinuous derivative of the drain charge.

The drain saturation voltage (1.35) is plotted versus the pinch-off voltage in Figure 1.6, which shows that v_{dsat} can be substantially smaller than v_p .

Figure 1.7 shows the forward, reverse and drain currents in saturation versus the pinch-off voltage. The blue curve corresponds to the normalized drain current i_d in saturation for a long-channel device for which velocity saturation can be ignored. In such a case the drain charge and the reverse current are both zero and the drain current i_d is simply equal to the forward current i_f . When velocity saturation is present, the drain charge q_d in saturation is no more zero but equal to q_{dsat} making the reverse current larger than zero as shown by the green curve in Figure 1.7. The drain current accounting for velocity saturation is then reduced by i_r resulting in the red curve in Figure 1.7. Notice also that the drain current being the difference between two quadratic functions of q_s and hence of $v_p - v_s$, it

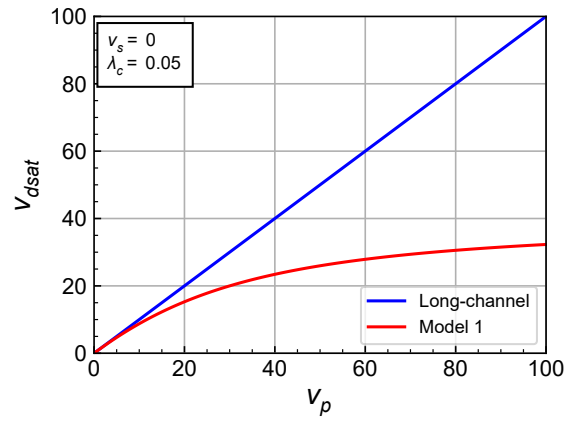


Figure 1.6: Effect of velocity saturation on the saturation voltage in strong inversion for Model 1.

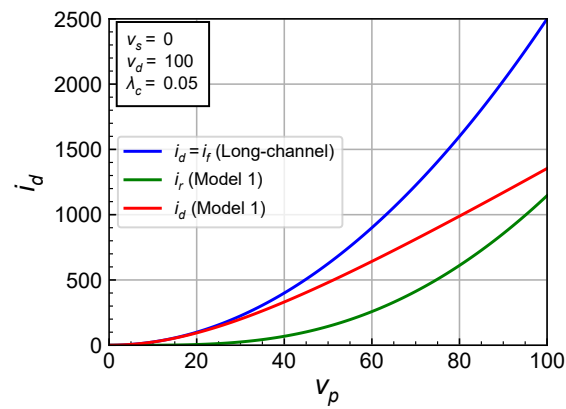


Figure 1.7: Drain current versus pinch-off voltage for a long- and short-channel transistor illustrating the effect of velocity saturation making the drain current a linear function of v_p .

becomes a linear function of v_p . Indeed, in very strong inversion under strong velocity saturation ($\lambda_c q_s \gg 1$), from (1.34) we get

$$q_{d_{sat}} \cong q_s, \quad (1.36)$$

$$i_{d_{sat}} \cong \frac{2q_s}{\lambda_c} = \frac{v_p - v_s}{\lambda_c}. \quad (1.37)$$

From (1.36) we observe that the inversion charge is constant along the channel and equal to q_s and from (1.37) that the current is no more proportional to q_s^2 but to q_s and hence to $v_p - v_s$.

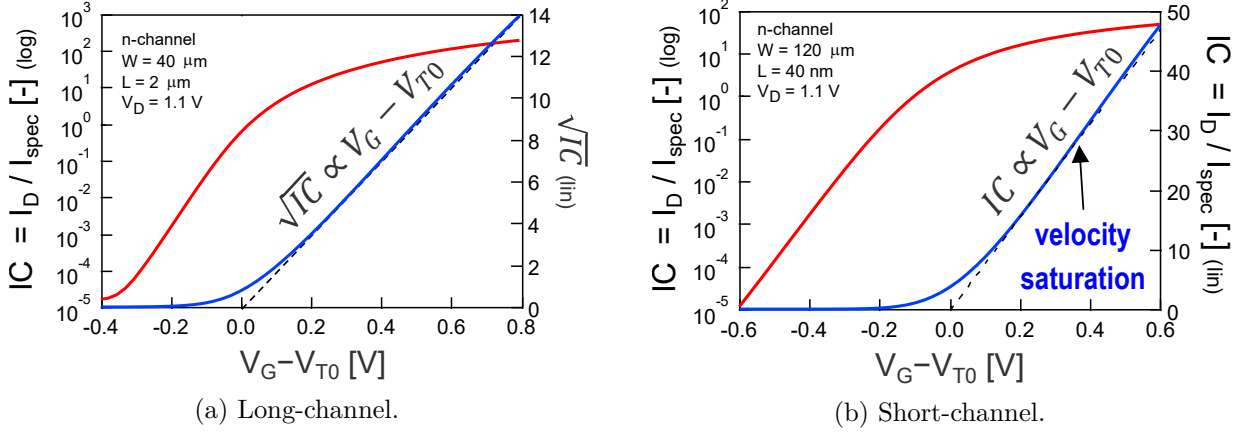


Figure 1.8: IC versus $V_G - V_{T0}$ measured on a long and short transistor from a 40 nm bulk CMOS process.

Figure 1.8 shows the normalized current $i_d = IC$ versus the overdrive voltage $V_G - V_{T0}$ measured on a long-channel and a minimum-length nMOS transistors from a 40 nm bulk CMOS process. The blue curve in Figure 1.8a corresponds to $\sqrt{IC}$ which, as expected, is proportional to the overdrive voltage $V_G - V_{T0}$ which validates the quadratic dependence for long-channel devices. The blue curve in Figure 1.8b corresponds to IC and shows to be proportional to the overdrive voltage $V_G - V_{T0}$. This is a clear sign of velocity saturation. Indeed, in strong inversion and for a minimum channel length, $\lambda_c q_s \gg 1$ and (1.34) simplifies to

$$i_{d_{sat}} \cong \frac{2q_s}{\lambda_c} \cong \frac{v_p - v_s}{\lambda_c}, \quad (1.38)$$

which, after denormalization, simplifies to

$$I_{D_{sat}} \cong n \cdot W \cdot C_{ox} \cdot v_{sat} \cdot (V_P - v_s) \cong W \cdot C_{ox} \cdot v_{sat} \cdot (V_G - V_{T0} - n \cdot v_s). \quad (1.39)$$

Equation (1.39) shows that when the channel is under strong velocity saturation conditions, the drain current in saturation does not depend on the channel length anymore and varies linearly instead of quadratically with respect to the overdrive voltage. This can be explained by the fact that in such condition on one hand the inversion charge becomes almost uniform along the channel from source to drain and equal to the value taken at the source and on the other hand that the carriers are moving at their maximum velocity which is constant along the channel. The charge moving from source to drain per unit time corresponding to the drain current is therefore constant and independent of the channel length L for a given W and $V_P - v_s$.

1.3.2.2 Analysis valid in all regions of inversion

The above analysis can be extended to also include the moderate and weak inversion regions [5] [6] [7] [8]. In order to do this we start with the drain current expression

$$I_D = \mu_{eff} \cdot Q_i \cdot \mathcal{E}_x + \mu_{eff} \cdot U_T \cdot \frac{dQ_i}{dx} \quad (1.40)$$

which can be written in normalized form as

$$i_d = 2q_i \cdot \frac{u \cdot e}{\lambda_c} - u \cdot \frac{dq_i}{d\xi}. \quad (1.41)$$

The inversion charge linearization equation (??) gives a relation between the longitudinal field and the inversion charge gradient

$$|\mathcal{E}_x| = \frac{d\Psi_s}{dx} = \frac{1}{nC_{ox}} \cdot \frac{dQ_i}{dx} \quad (1.42)$$

which can be written in normalized form as

$$e = -\lambda_c \cdot \frac{dq_i}{d\xi}. \quad (1.43)$$

Substituting (1.43) in (1.41) results in

$$i_d = -u \cdot (2q_i + 1) \cdot \frac{dq_i}{d\xi}. \quad (1.44)$$

To derive the model that is valid in all modes of inversion but includes the effect of velocity saturation, we use the simplest velocity-field model, namely Model 1. Integrating (1.44) from source to drain assuming first that there is no velocity saturation at the drain (i.e. $v_d < v_{d_{sat}}$ or $q_d > q_{d_{sat}}$ and hence $u = 1$ in (1.44)), we get the earlier long-channel result

$$i_d = q_s^2 - q_d^2 + q_s - q_d. \quad (1.45)$$

where $q_s^2 - q_d^2$ corresponds to the drift current, whereas $q_s - q_d$ corresponds to the diffusion current. For a long-channel device in saturation we have $q_d \ll q_s$ and the drain current in saturation is then simply given by

$$i_{d_{sat}} = q_s^2 + q_s. \quad (1.46)$$

For a short-channel device, velocity saturation needs to be accounted for in the drift part of the drain current in saturation by replacing q_d by $q_{d_{sat}}$. In weak inversion, the longitudinal electric field $\mathcal{E}_x$ usually stays smaller than the critical field $\mathcal{E}_c$ and the current should therefore not be affected by velocity saturation. Hence, the q_d term in the diffusion part of the current (1.45) should not be replaced by $q_{d_{sat}}$ and can simply be neglected in saturation. Therefore, as soon as velocity saturation starts to happen right at the drain, the current in saturation becomes

$$i_{d_{sat}} = q_s^2 - q_{d_{sat}}^2 + q_s. \quad (1.47)$$

i Note

In the original model proposed in [5] [6] [7] [8], the drain current in saturation $i_{d_{sat}}$ was actually affected by velocity saturation in weak inversion, because the q_d term of the diffusion current was also replaced by $q_{d_{sat}}$. However, the difference between this model and the original model presented in [5] [6] [7] [8] remains very small particularly for small values of λ_c .

Solving (1.47) and (1.18) for $q_{d_{sat}}$ and $i_{d_{sat}}$ results in

$$q_{d_{sat}} = \frac{1}{\lambda_c} \cdot \left[\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)} - 1 \right] = \frac{\lambda_c (q_s^2 + q_s)}{1 + \sqrt{1 + \lambda_c^2 (q_s^2 + q_s)}}, \quad (1.48)$$

$$i_f = q_s^2 + q_s, \quad (1.49)$$

$$i_{r_{sat}} = q_{d_{sat}}^2 = \frac{1}{\lambda_c^2} \cdot \left[\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)} - 1 \right]^2 = q_s^2 + q_s - \frac{2}{\lambda_c^2} \cdot \left[\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)} - 1 \right], \quad (1.50)$$

$$i_{d_{sat}} = \frac{2}{\lambda_c} \cdot q_{d_{sat}} = i_f - i_{r_{sat}} = \frac{2}{\lambda_c^2} \cdot \left[\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)} - 1 \right] = \frac{2(q_s^2 + q_s)}{1 + \sqrt{1 + \lambda_c^2 (q_s^2 + q_s)}}, \quad (1.51)$$

$$v_{d_{sat}} = v_p - 2q_{d_{sat}} = v_p - \frac{2}{\lambda_c} \cdot \left[\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)} - 1 \right]. \quad (1.52)$$

From (1.48) and (1.51), we see that for $\lambda_c = 0$ (no velocity saturation), we fall back to the long-channel equations with $q_{d_{sat}} = 0$, $i_{r_{sat}} = 0$ (saturation) and $i_{d_{sat}} = q_s^2 + q_s$. In strong inversion ($q_s \gg 1$) we get the earlier result (1.31) and (1.34) found for Model 1. In very strong inversion under velocity saturation ($\lambda_c q_s \gg 1$), we get the strong inversion asymptotes (1.37). As expected and contrary to the original model presented in [5] [6] [7] [8], in weak inversion we do not observe any impact of velocity saturation. Indeed, for $q_s \ll 1$, (1.48) and (1.51) reduce to

$$q_{d_{sat}} \cong \frac{\lambda_c}{2} \cdot q_s \ll 1, \quad (1.53)$$

$$i_{d_{sat}} \cong q_s. \quad (1.54)$$

1.3.3 Effect of velocity saturation on the transconductances

1.3.3.1 Analysis in strong inversion

The impact of velocity saturation on the transconductances in saturation can be derived from the above analysis by rewriting the normalized source transconductance in saturation as

$$g_{ms_{sat}} \triangleq -\frac{\partial i_{d_{sat}}}{\partial v_s} = -\frac{\partial i_{d_{sat}}}{\partial q_s} \cdot \frac{\partial q_s}{\partial v_s}. \quad (1.55)$$

We first will derive $g_{ms_{sat}}$ in strong inversion in which case

$$\frac{\partial q_s}{\partial v_s} = -\frac{1}{2}. \quad (1.56)$$

For Model 1, the derivative $\partial i_{d_{sat}} / \partial q_s$ is simply obtained by differentiating (1.34) resulting in

$$\frac{\partial i_{d_{sat}}}{\partial q_s} = \frac{2q_s}{\sqrt{1 + (\lambda_c q_s)^2}}. \quad (1.57)$$

The normalized source transconductance in strong inversion and saturation is then given by

$$g_{ms_{sat}} = \frac{q_s}{\sqrt{1 + (\lambda_c q_s)^2}}. \quad (1.58)$$

For circuit design it is more useful to have g_{ms} expressed in term of the inversion coefficient $IC = i_{d_{sat}}$. This can be done by solving (1.34) for q_s which results in

$$q_s = \sqrt{IC} \cdot \sqrt{1 + \left(\frac{\lambda_c}{2}\right)^2 IC}. \quad (1.59)$$

$g_{ms_{sat}}$ is then given by

$$g_{ms_{sat}} = \sqrt{IC} \cdot \frac{\sqrt{4 + \lambda_c^2 IC}}{2 + \lambda_c^2 IC}. \quad (1.60)$$

In very strong inversion ($q_s \gg 1$ or $IC \gg 1$), (1.58) and (1.60) reduce to

$$g_{ms_{sat}} \cong \frac{1}{\lambda_c} = \frac{L}{L_{sat}}. \quad (1.61)$$

Instead of increasing as $\sqrt{IC}$, as it does for long-channel devices, for short-channel transistors with heavy velocity saturation, the normalized source transconductance $g_{ms_{sat}}$ saturates to $1/\lambda_c$. Parameter λ_c therefore sets the maximum normalized source transconductance that can be achieved by a short-channel device for a given technology.

From (1.61) it seems that the transconductance scales like L , however this is not correct. Indeed, if we write the transconductance in denormalized form, we get

$$G_{m_{sat}} \cong \frac{G_{spec}}{\lambda_c} = n W C_{ox} v_{sat} \quad (1.62)$$

and hence

$$G_{m_{sat}} \cong \frac{G_{m_{sat}}}{n} = W C_{ox} v_{sat}. \quad (1.63)$$

From (1.62) and (1.63), we see that in strong inversion with heavy velocity saturation, the source and gate transconductances become actually independent of the transistor length L and independent of the current! They only depend on the transistor width W and the saturated drift velocity v_{sat} . This means that increasing the bias current does not help increasing the transconductance and the only way to increase $G_{m_{sat}}$ and $G_{m_{sat}}$ is to increase the transistor width.

1.3.3.2 Analysis valid in all regions of inversion

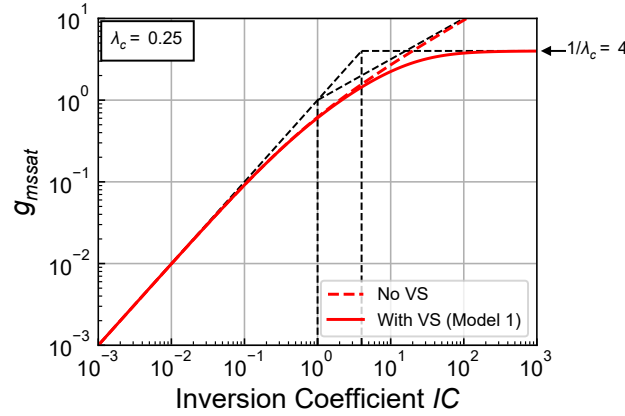


Figure 1.9: Effect of VS on the transconductance in saturation.

We now will derive an expression of the normalized source transconductance in saturation that is valid in all modes of inversion using the simplest velocity-field model, namely Model 1. To do this we first rewrite the source transconductance as

$$g_{m_{sat}} \triangleq -\frac{\partial i_{d_{sat}}}{\partial v_s} = \frac{\partial i_{d_{sat}}}{\partial q_s} \cdot \left(\frac{\partial v_s}{\partial q_s} \right)^{-1}. \quad (1.64)$$

Recalling the relation between the saturation voltage and the source charge which allows to write the source voltage in terms of the source charge at a given gate and pinch-off voltage as

$$v_s = v_p - 2q_s - \ln(q_s). \quad (1.65)$$

Differentiating (1.65) we get

$$\frac{\partial v_s}{\partial q_s} = -\frac{2q_s + 1}{q_s}. \quad (1.66)$$

For Model 1, the derivative $\partial i_{d_{sat}} / \partial q_s$ is simply obtained by differentiating (1.51) resulting in

$$\frac{\partial i_{d_{sat}}}{\partial q_s} = \frac{2q_s + 1}{\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)}} \quad (1.67)$$

The normalized source transconductance in saturation for Model 1 is then given by

$$g_{m_{sat}} = \frac{q_s}{\sqrt{1 + \lambda_c^2 (q_s^2 + q_s)}}, \quad (1.68)$$

which has the following asymptotes in weak and strong inversion

$$g_{msat} \cong \begin{cases} q_s & \text{in weak inversion } (q_s \ll 1) \\ \frac{q_s}{\sqrt{1+(\lambda_c q_s)^2}} \cong \frac{1}{\lambda_c} & \text{in strong inversion } (q_s \gg 1). \end{cases} \quad (1.69)$$

Without velocity saturation ($\lambda_c = 0$), (1.68) reduces to the long-channel expression $g_{msat} \cong q_s$. In strong inversion ($q_s \gg 1$) we get the earlier result (1.58) found in the strong inversion model. In very strong inversion under velocity saturation ($\lambda_c q_s \gg 1$), we get the strong inversion asymptote (1.61). In weak inversion, as expected velocity saturation has no impact as shown by (1.69) contrary to the original model presented in [5] [6] [7] [8] which shows some dependence on λ_c .

As above, we can also express g_{msat} in terms of the inversion coefficient $i_{dsat} = IC$ by solving (1.51) for q_s resulting in

$$q_s = \frac{1}{2} \left(\sqrt{4IC + 1 + (\lambda_c IC)^2} - 1 \right). \quad (1.70)$$

Replacing (1.70) in (1.68) results in

$$g_{msat} = \frac{\sqrt{4IC + 1 + (\lambda_c IC)^2} - 1}{2 + \lambda_c^2 IC} = \frac{2IC \left[1 + \left(\frac{\lambda_c}{2} \right)^2 IC \right]}{\left(1 + \frac{\lambda_c^2}{2} IC \right) \left(1 + \sqrt{4IC + 1 + (\lambda_c IC)^2} \right)}. \quad (1.71)$$

Again without velocity saturation ($\lambda_c = 0$), (1.71) reduces to the long-channel expressions

$$g_{msat} = \frac{\sqrt{4IC + 1} - 1}{2} = \frac{2IC}{1 + \sqrt{4IC + 1}}. \quad (1.72)$$

The asymptotes in weak and strong inversion are given by

$$g_{msat} \cong \begin{cases} IC & \text{in weak inversion } (IC \ll 1) \\ \sqrt{IC} \cdot \frac{\sqrt{4 + \lambda_c^2 IC}}{2 + \lambda_c^2 IC} \cong \frac{1}{\lambda_c} & \text{in strong inversion } (IC \gg 1). \end{cases} \quad (1.73)$$

We see that the source transconductance in weak inversion and in saturation for short-channel devices is not affected by velocity saturation and remains proportional to the inversion coefficient and hence to the drain current.

1.3.4 Effect of velocity saturation on the G_m/I_D ratio

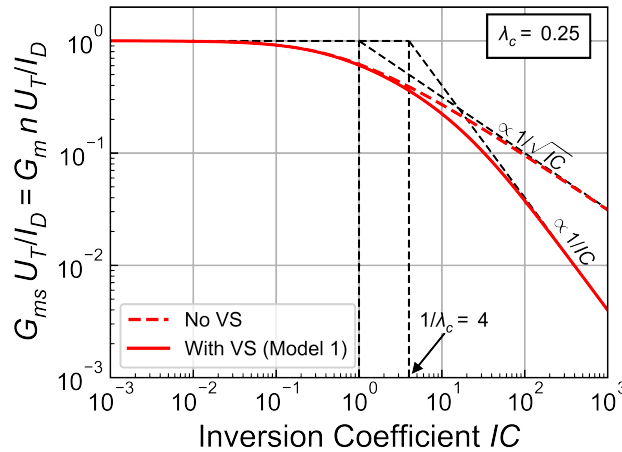


Figure 1.10: Effect of VS on the G_m/I_D ratio.

We can now easily derive an expression for the G_m/I_D FoM that accounts for velocity saturation and is valid from weak to strong inversion. The normalized G_m/I_D FoM is obtained by dividing (1.71) by IC resulting in

$$\frac{g_{ms}}{IC} = \frac{G_{ms} U_T}{I_D} = \frac{G_m n U_T}{I_D} = \frac{\sqrt{4IC + 1 + (\lambda_c IC)^2} - 1}{IC (2 + \lambda_c^2 IC)} = \begin{cases} 1 & \text{in WI and sat.} \\ \frac{1}{\lambda_c IC} & \text{in SI and sat.} \end{cases} \quad (1.74)$$

Equation (1.74) is plotted versus IC in red in Figure 1.10 and compared to the long-channel behavior corresponding to the dashed red curve. We clearly see that for a given current budget, the impact of velocity saturation in strong inversion is dramatic for low-power design since it reduces the current efficiency by a factor $\lambda_c \sqrt{IC}$. For example, for $\lambda_c = 0.5$ and $IC = 100$, the current efficiency is reduced by a factor 5, which is considerable. This means that the transistor needs to be five times larger to recover the same transconductance for a given current budget. On the other hand, the normalized G_m/I_D ratio remains invariant (i.e. $g_{ms}/IC = 1$) in weak inversion, irrespective of the channel length. This is simply due to the fact that in weak inversion the source transconductance remains proportional to the current and additionally velocity saturation has no effect on g_{ms} . As shown in Figure 1.10, the inversion coefficient for which the strong inversion asymptote of a short-channel device crosses the horizontal unity line is equal to $1/\lambda_c$. This is actually how the parameter λ_c can be extracted from measurements of a short-channel device.

We actually don't need to go to very deep-submicron CMOS technologies to see the effect of velocity saturation. Figure 1.11 shows the current efficiency measured on a minimum length nMOS transistor from a 130 nm bulk CMOS process and which already shows a loss of efficiency in strong inversion. It also illustrates how the λ_c parameter can be extracted from the intersection of the $1/IC$ asymptote with the horizontal unity line giving $1/\lambda_c = 3$ and hence $\lambda_c = 1/3$. Figure 1.11 also highlights the very good match between the simple expression (1.74) and the measured data.

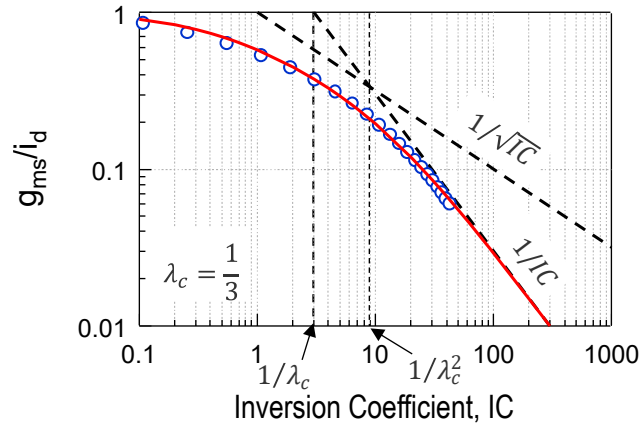


Figure 1.11: Effect of velocity saturation on the current efficiency measured on a 130 nm process.

The impact of velocity saturation is also illustrated in Figure 1.14 which shows the current efficiency measured on a long-channel (Figure 1.12a) and short-channel device (Figure 1.12b) from a 40 nm bulk CMOS process. It actually corresponds to the derivatives of the large-signal IC versus $V_G - V_{T0}$ characteristics shown in Figure 1.8. We clearly see the strong degradation introduced by velocity saturation on the current efficient of the short-channel transistor in Figure 1.12b. We also see a reasonable fit between the simple analytical expression of the normalized G_m/I_D and the measurements.

Figure 1.13 shows the current efficiency measured on different devices having the same width but three different lengths for the same 40 nm bulk CMOS technology. The blue curve corresponds to the long-channel asymptote ($L = 2 \mu m$), the green curve corresponds to a medium channel length ($L = 120 nm$) but already affected by velocity saturation and finally the red curve corresponds to the minimum channel length ($L = 40 nm$) which is strongly affected by velocity saturation. We observe

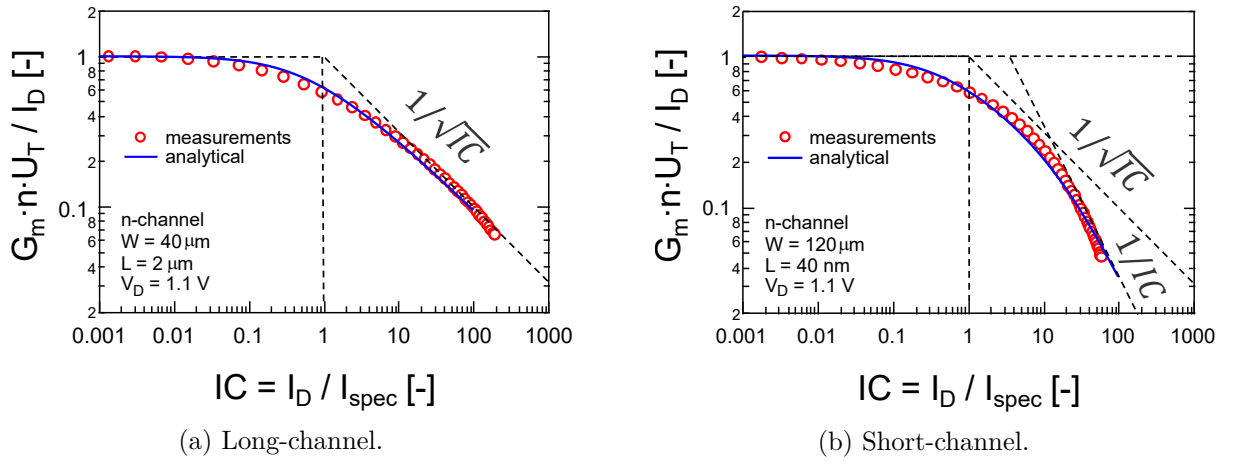


Figure 1.12: Effect of velocity saturation on the current efficiency measured on a 40 nm process.

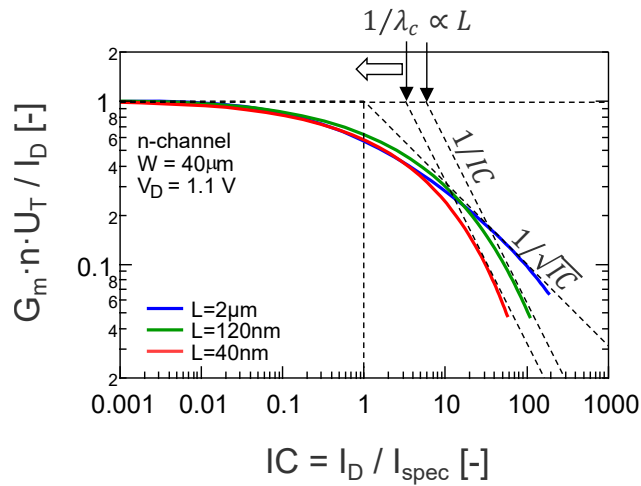


Figure 1.13: Scaling of the current efficiency characteristic with the transistor length L .

that the cross-point between the $1/IC$ asymptote which should scale like L indeed decreases and hence moves to the left. Additional measurements performed on more advanced bulk, FDSOI and FinFET technologies are presented in .

The model used in this Section, assumed that velocity saturation occurs when $v_d = v_{d_{sat}}$ and that the drain charge and current in saturation (for $v_d > v_{d_{sat}}$) remain constant. In reality this is not the case because of the effects of *channel length modulation* (CLM) and *drain induced barrier lowering* (DIBL) which are discussed in the next Sections.

1.4 Channel length modulation (CLM)

1.4.1 Channel length reduction

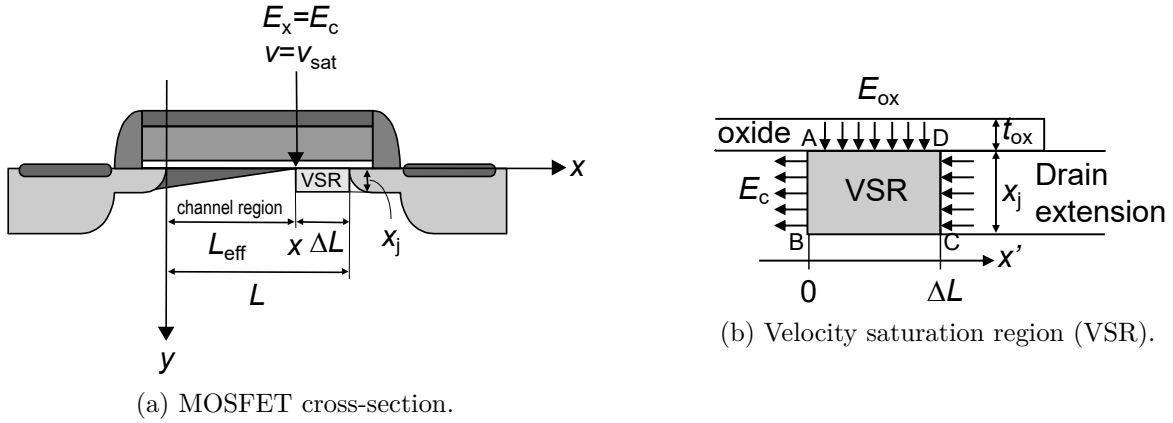


Figure 1.14: Channel length modulation.

In the previous Section, velocity saturation starts when the longitudinal field $\mathcal{E}_x$ becomes equal to the critical field $\mathcal{E}_c$ right at the drain limit of the channel. At this point, the drain voltage becomes equal to the saturation voltage $V_{D_{sat}}$. As the drain voltage gets larger than the saturation voltage $V_{D_{sat}}$, the saturation point moves away from the drain into the channel. As illustrated in Figure 1.14a, this is creating a *velocity saturated region* (VSR) in which the carriers are traveling at their maximum saturated velocity v_{sat} . Using a quasi-2D analysis applied to the VSR shown in Figure 1.14b, it can be shown that the channel length reduction ΔL is given by [9] [10] [11] [12] [13]

$$\Delta L \triangleq L - L_{eff} = \ell \cdot \ln(u + \sqrt{u^2 + 1}), \quad (1.75)$$

where u is defined as

$$u \triangleq \frac{V_{DS} - V_{D_{sat}}}{\ell \mathcal{E}_{sat}} = \frac{v_d - v_{d_{sat}}}{\ell \mathcal{E}_{sat}}. \quad (1.76)$$

$V_{D_{sat}} = V_{D_{sat}} - v_s$ is the drain-to-source saturation voltage accounting for velocity saturation. If we use Model 1 the normalized value of $V_{D_{sat}}$ is given by (1.35)

$$v_{d_{sat}} = v_p - \frac{2\lambda_c \cdot q_s^2}{1 + \sqrt{1 + (\lambda_c q_s)^2}}.$$

Parameter ℓ (which has length unit) is defined as [11] [12] [13]

$$\ell \triangleq \sqrt{\frac{\epsilon_{si} x_j}{C_{ox}}} \quad (1.77)$$

where x_j is the junction depth and C_{ox} the oxide capacitance per unit area. Parameter $\mathcal{E}_{sat}$ is defined as the longitudinal electric field at the pinch-off point which is actually equal to the critical field $\mathcal{E}_c$ defined in the velocity saturation model

$$\mathcal{E}_{sat} = \mathcal{E}_c = \frac{v_{sat}}{\mu_0} = \frac{2U_T}{L_{sat}}. \quad (1.78)$$

We can then rewrite (1.76) in terms of the normalized drain and saturation voltages according to

$$u = \frac{L_{sat}}{2\ell} \cdot (v_d - v_{d_{sat}}). \quad (1.79)$$

In the above model, L_{sat} appears in the VS parameter $\lambda_c = L_{sat}/L$ that is used for the computation of the drain current in saturation, but also in u to scale the voltage that is used for the calculation of ΔL . For $u \ll 1$, the channel length reduction simplifies to

$$\Delta L \cong \ell \cdot \ln(1 + u) = \ell \cdot \ln \left(1 + \frac{v_d - v_{d_{sat}}}{v_{dclm}} \right), \quad (1.80)$$

where $v_{dclm} = 2\ell/L_{sat}$ is considered an additional fitting parameter. This simpler model was initially proposed by Arora [12] [13] and used in the EKV v2.6 compact model [14]. We will use this model below to investigate the impact of CLM on the drain current in saturation.

1.4.2 Impact of CLM on the drain current in saturation

1.4.2.1 Analysis in strong inversion

CLM has obviously an impact on the drain current in saturation which can no more be considered as constant. Since the impact of VS is dominant in strong inversion, we first analyze the impact of CLM in strong inversion.

When accounting for VS, the transistor enters into saturation when the longitudinal field $\mathcal{E}_x$ becomes equal to the critical field $\mathcal{E}_c$ which happens when $v_d = V_{D_{sat}}$ (or in normalized form $v_d = v_{d_{sat}}$). At this point the channel length is still equal to L and the drain current is given by

$$I_{D_{sat0}} = I_{spec} \cdot i_{d_{sat0}} \quad (1.81)$$

with

$$I_{spec} \triangleq I_{spec\Box} \cdot \frac{W}{L} \quad (1.82)$$

and

$$i_{d_{sat0}} = \frac{2q_s^2}{1 + \sqrt{1 + (\lambda_{c0} q_s)^2}} \quad (1.83)$$

where the VS parameter λ_{c0} is defined as

$$\lambda_{c0} \triangleq \frac{L_{sat}}{L}. \quad (1.84)$$

The 0 in the subscript indicates that the operating point corresponds to the onset of saturation for which $v_d = v_{d_{sat}}$ and the length is still equal to L (or $\Delta v_d = v_d - v_{d_{sat}} = 0$ and $\Delta L = 0$).

When the drain voltage is increased above $V_{D_{sat}}$ (or $v_d > V_{D_{sat}}$), the pinch-off point (defined as the point along the channel at which $\mathcal{E}_x = \mathcal{E}_c$ or $V_{ch} = V_{D_{sat}}$) moves in the direction of the source reducing the channel length by ΔL . The drain current in saturation for $v_d > V_{D_{sat}}$ is then given by

$$I_{D_{sat}} = I_{spec\Box} \cdot \frac{W}{L_{eff}} \cdot i_{clm} \quad (1.85)$$

where L has been replaced by the effective length $L_{eff} \triangleq L - \Delta L$ in I_{spec} . The term i_{clm} accounts for the dependency of λ_{c0} on L in $i_{d_{sat0}}$ where L also needs to be replaced by L_{eff} . i_{clm} is defined as

$$i_{clm} = \frac{2q_s^2}{1 + \sqrt{1 + (\lambda_c q_s)^2}} \quad (1.86)$$

where the VS parameter λ_c is now defined as

$$\lambda_c \triangleq \frac{L_{sat}}{L_{eff}}. \quad (1.87)$$

We actually can keep the original definition of I_{spec} given by (1.82) by rewriting $I_{D_{sat}}$ as

$$\begin{aligned} I_{D_{sat}} &= I_{spec} \cdot \frac{W}{L_{eff}} \cdot i_{clm} = I_{spec} \cdot \frac{W}{L} \cdot \frac{L}{L_{eff}} \cdot i_{clm} \\ &= I_{spec} \cdot \frac{L}{L - \Delta L} \cdot i_{clm} = I_{spec} \cdot f_{clm} \cdot i_{clm} \end{aligned} \quad (1.88)$$

where f_{clm} is defined as

$$f_{clm} \triangleq \frac{L}{L_{eff}} = \frac{L}{L - \Delta L} = \frac{1}{1 - \Delta L/L}. \quad (1.89)$$

The VS saturation parameter λ_c can then also be written in terms of f_{clm} as

$$\lambda_c \triangleq \frac{L_{sat}}{L_{eff}} = \frac{L_{sat}}{L - \Delta L} = \frac{L_{sat}}{L} \cdot \frac{L}{L - \Delta L} = \frac{L_{sat}}{L} \cdot \frac{1}{1 - \Delta L/L} = \lambda_{c0} \cdot f_{clm}. \quad (1.90)$$

The normalized drain current in saturation $i_{d_{sat}}$ is then simply given by

$$i_{d_{sat}} \triangleq \frac{I_{D_{sat}}}{I_{spec}} = f_{clm} \cdot i_{clm}. \quad (1.91)$$

where I_{spec} is constant and given by (1.82).

CLM is impacting the drain current in saturation $i_{d_{sat}}$ in two ways. First, $i_{d_{sat}}$ increases due to the increase of f_{clm} . Second, the latter increase of f_{clm} also increases λ_c which reduces the drain current. This current decrease is captured by i_{clm} . There are therefore two antagonist effects induced by CLM that need to be carefully accounted for, particularly when deriving the output conductance.

In strong inversion ($q_s \gg 1$), and for strong VS ($\lambda_c q_s \gg 1$), the drain current at saturation is given by

$$i_{d_{sat0}} \cong \frac{2}{\lambda_{c0}} q_s. \quad (1.92)$$

Similarly, the i_{clm} term becomes

$$i_{clm} \cong \frac{2}{\lambda_c} q_s = \frac{2}{\lambda_{c0} f_{clm}} q_s. \quad (1.93)$$

The normalized drain current in saturation then becomes

$$i_{d_{sat}} = f_{clm} \cdot i_{clm} \cong f_{clm} \cdot \frac{2}{\lambda_{c0} f_{clm}} q_s = i_{d_{sat0}}. \quad (1.94)$$

In this case, CLM has no impact on the drain current because i_{clm} fully compensates f_{clm} and the drain current stays constant and equal to $i_{d_{sat0}}$.

This behavior is specific to short-channel devices. However, as it will be established below, for short-channel transistors, DIBL actually dominates CLM. We therefore need to focus on the impact of CLM on long-channel devices.

1.4.2.2 Long-channel expression in strong inversion

For long-channel transistors, we can ignore the impact of CLM on the i_{clm} term because λ_{c0} becomes very small and λ_c also remains small because f_{clm} does not increase much for long-channel devices. Therefore, for long-channel transistors, i_{clm} can be considered to be equal to the long-channel current $i_{clm} \cong q_s^2$. The drain current in strong inversion and in saturation for a long-channel transistor is then only impacted by f_{clm} according to

$$i_{d_{sat}} \cong f_{clm} \cdot q_s^2. \quad (1.95)$$

1.4.2.3 Linear dependence and empirical model

In the measurements presented below, we see that the output characteristic in saturation can be approximated by a linear function. Such a linear dependence of $i_{d_{sat}}$ on v_d can be obtained from the model of ΔL versus v_d described in (1.80) if parameter v_{dclm} and the maximum drain voltage are such that $\Delta v_d = v_d - v_{d_{sat}} \ll v_{dclm}$. Then $\Delta L/L$ becomes

$$\frac{\Delta L}{L} = \frac{\ell}{L} \cdot \ln \left(1 + \frac{v_d - v_{d_{sat}}}{v_{dclm}} \right) \cong \frac{\ell}{L} \cdot \frac{v_d - v_{d_{sat}}}{v_{dclm}} = \frac{L_{clm}}{L} \cdot (v_d - v_{d_{sat}}) \quad (1.96)$$

with $L_{clm} \triangleq \ell/v_{dclm}$. If we additionally can consider that $\ell < L$, then $\Delta L/L \ll 1$ and f_{clm} simplifies to

$$f_{clm} = \frac{1}{1 - \Delta L/L} \cong 1 + \frac{\Delta L}{L} \cong 1 + \frac{\ell}{L} \cdot \frac{v_d - v_{d_{sat}}}{v_{dclm}} = 1 + \frac{L_{clm}}{L} \cdot (v_d - v_{d_{sat}}). \quad (1.97)$$

Under the assumption that $\Delta L/L \ll 1$, the impact of f_{clm} on λ_c can be neglected and λ_c can be replaced by λ_{c0} which is independent of v_d . The drain current in saturation $i_{d_{sat}}$ then writes

$$i_{d_{sat}} \cong \left(1 + \frac{\Delta L}{L} \right) \cdot i_{d_{sat0}} \cong \left(1 + \frac{\ell}{L} \cdot \frac{v_d - v_{d_{sat}}}{v_{dclm}} \right) \cdot i_{d_{sat0}} = \left[1 + \frac{L_{clm}}{L} \cdot (v_d - v_{d_{sat}}) \right] \cdot i_{d_{sat0}}, \quad (1.98)$$

which has now the desired linear behavior with respect to v_d .

The drain current in saturation can therefore be expressed using an empirical model with a linear function having a slope equal to the output conductance G_{ds} and an Early voltage V_E (which is negative)

$$I_{D_{sat}} \cong G_{ds} \cdot (v_d - V_E) \quad (1.99)$$

or in normalized form

$$i_{d_{sat}} \cong g_{ds} \cdot (v_d - v_e). \quad (1.100)$$

g_{ds} and v_e are related to the parameters in (1.98) by

$$g_{ds} = \frac{L_{clm}}{L} \cdot i_{d_{sat0}}, \quad (1.101)$$

$$-v_e = \frac{i_{d_{sat0}}}{g_{ds}} - V_{D_{sat}}. \quad (1.102)$$

1.4.3 Output conductance in saturation due to CLM

The output conductance in saturation is defined as

$$G_{ds} = \left. \frac{\partial I_{D_{sat}}}{\partial V_{DS}} \right|_{V_G, v_s, v_d}. \quad (1.103)$$

or in normalized form

$$g_{ds} \triangleq \frac{G_{ds}}{G_{spec}} = \frac{\partial i_{d_{sat}}}{\partial v_{ds}} \quad (1.104)$$

with

$$G_{spec} \triangleq \frac{I_{spec}}{U_T}. \quad (1.105)$$

In this derivation we will consider that v_s is constant, therefore

$$g_{ds} = \frac{\partial i_{d_{sat}}}{\partial v_{ds}} = \frac{\partial i_{d_{sat}}}{\partial v_d}. \quad (1.106)$$

It can be shown that g_{ds} is actually proportional to the saturation current $i_{d_{sat}}$ and can be written as

$$g_{ds} = \frac{i_{d_{sat}}}{v_{clm}} \quad (1.107)$$

where v_{clm} is the CLM voltage given by

$$\frac{1}{v_{clm}} = \frac{g_{ds}}{i_{dsat}} = g_1 \cdot g_2, \quad (1.108)$$

where g_1 accounts for the dependence of f_{clm} to v_d , whereas g_2 is due to the change of λ_c with respect to v_d

$$g_1 \triangleq \frac{\ell}{L} \cdot \frac{f_{clm}}{v_{dclm} + v_d - v_{dsat}} = \frac{L_{clm}}{L} \cdot \frac{f_{clm}}{1 + \frac{v_d - v_{dsat}}{v_{dclm}}}, \quad (1.109)$$

$$g_2 \triangleq 1 - \frac{\lambda_c^2}{2} \cdot \frac{i_{clm}}{\sqrt{1 + (\lambda_c q_s)^2}}. \quad (1.110)$$

For a long-channel transistor, we can neglect the change of the drain current due to a change of λ_c and hence $g_2 \cong 1$, which leads to

$$\frac{1}{v_{clm}} \cong g_1 = \frac{\ell}{L} \cdot \frac{f_{clm}}{v_{dclm} + v_d - v_{dsat}} = \frac{L_{clm}}{L} \cdot \frac{f_{clm}}{1 + \frac{v_d - v_{dsat}}{v_{dclm}}}. \quad (1.111)$$

In the case of the linear model we can assume that $v_d - v_{dsat} \ll v_{dclm}$ for which

$$\frac{1}{v_{clm}} \cong \frac{L_{clm}}{L} \quad (1.112)$$

which scales as $1/L$ but is now bias independent. This means that for long-channel transistors, most of the bias dependence of g_{ds} due to CLM for a long-channel transistor is actually captured by i_{dsat} and v_{clm} can be considered as approximately bias independent and proportionnal to L .

1.4.4 Experimental validation

As will be discussed below, the output characteristic and output conductance for short-channel transistors are dominated by DIBL. For this reason we will focus on a long-channel transistor in order to validate the model derived above. Figure 1.15 shows the output characteristic and output conductance measured in strong inversion on a very wide and long nMOS transistor ($W = 3 \mu m$, $L = 1 \mu m$) from a 28 nm bulk CMOS technology. Despite the long channel, we see that the output conductance does not vanish to zero in saturation. The finite output conductance shown in Figure 1.15 is then clearly due to CLM.

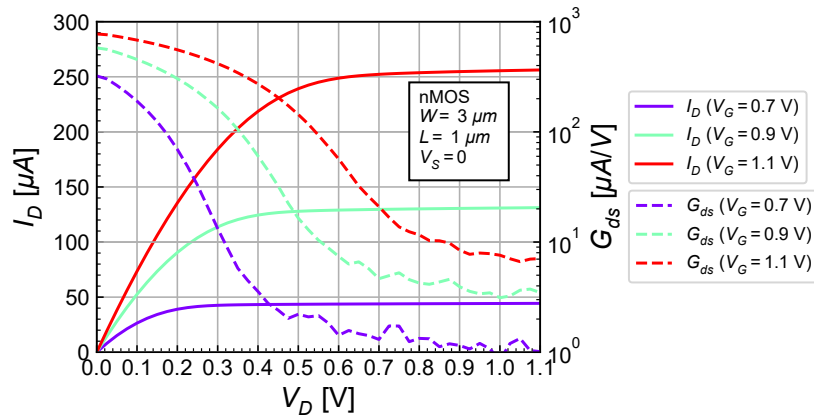


Figure 1.15: Output characteristic and output conductance measured on a long-channel nMOS transistor from a 28 nm bulk CMOS technology in strong inversion.

In order to perform the normalization and validation, the sEKV parameters have first been extracted from the I_D - V_G characteristics, resulting in the values presented in Table 1.3. We see that parameter λ_c is very small, justifying the long-channel approximation used above.

Table 1.3: sEKV parameters extracted for the long-channel nMOS transistor.

W	L	V_{T0}	n	I_{spec}	λ_c
$3 \mu m$	$1 \mu m$	0.427 V	1.176	$2.43 \mu A$	0.03

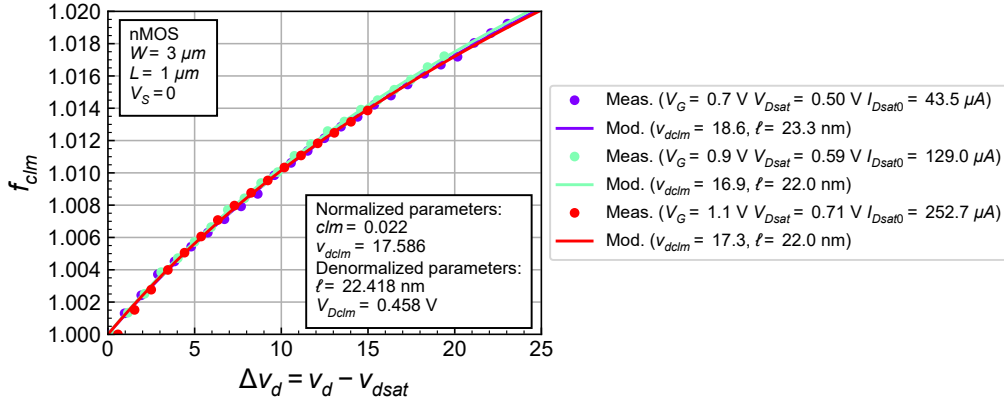


Figure 1.16: Fit of f_{clm} using (1.80).

Figure 1.16 shows the fit of the normalized measured output characteristic of this long-channel transistor for three different gate bias voltages. For long-channel the normalized current is about equal to f_{clm} which is calculated in Figure 1.16 using the channel length reduction expression given in (1.80). We see that a very good fit is obtained for values of the parameters $v_{dclm} = 17.7$ ($V_{DCLM} = 0.46 \text{ V}$) and $\ell = 17.6 \mu m$. We can also observe from Figure 1.16 that the output characteristic for a long-channel transistor is close to a linear dependence on Δv_d . We can therefore extract the parameters G_{ds} and V_E for each gate voltage resulting in the linear approximation shown in Figure 1.17. The parameters have been extracted using a linear regression using a saturation voltage slightly higher than the long-channel value V_P in order to have a better approximation of the output conductance in the saturation region.

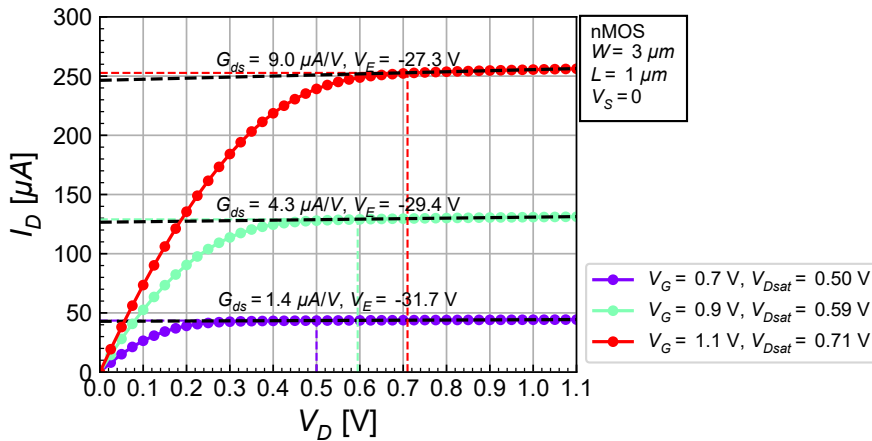


Figure 1.17: Fit of the output characteristic using the linear empirical model (1.99).

The empirical linear model can also be used in weak and moderate inversion as shown in Figure 1.18. From Figure 1.18, we observe that the parameter V_M used in the output conductance expression does not vary too much. Figure 1.19 shows the fit to the output conductance in saturation over all bias obtained using a single value of $V_M = 30 \text{ V}$. We can conclude that the simple model given by (1.107)

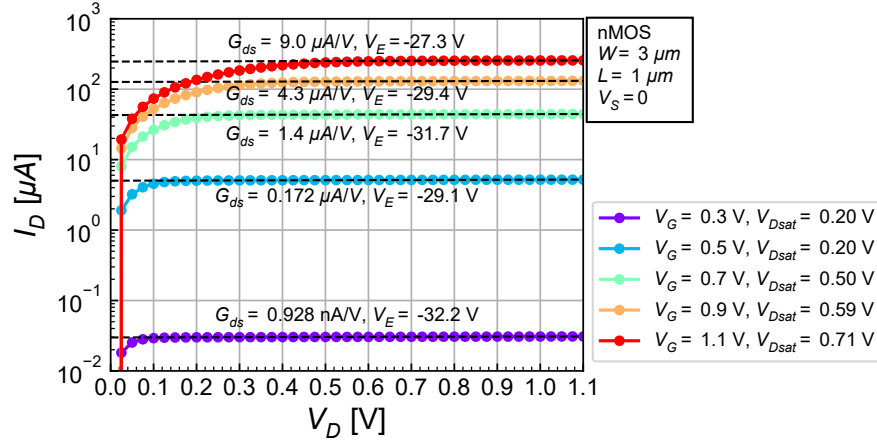


Figure 1.18: Output characteristic and output conductance in saturation due to CLM in all regions of inversion.

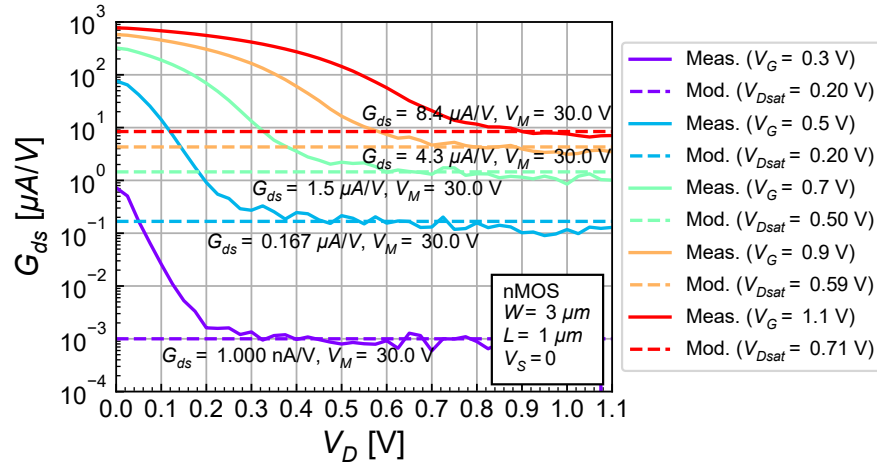


Figure 1.19: Output conductance in saturation due to CLM in all regions of inversion using the same value of parameter V_M .

provides a reasonable estimation of the output conductance in saturation due to CLM for long-channel transistors.

1.5 Drain induced barrier lowering (DIBL)

1.5.1 Physical description

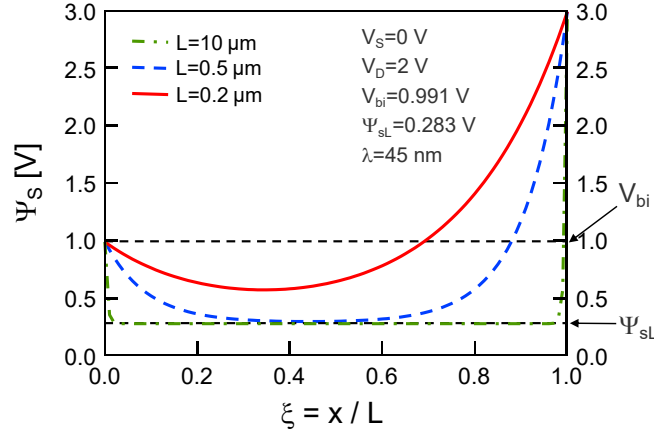


Figure 1.20: Drain induced barrier lowering (DIBL).

The drain current was derived assuming the gradual channel approximation which states that the change in the electric field along the channel direction is much smaller than the change along the vertical direction. This assumption allows to reduce the Poisson equation to a one dimensional differential equation depending only on the vertical direction. As a consequence, the charge density in silicon Q_{si} only depended on the vertical electric field at the surface of the silicon (just below the oxide). This approximation is no more valid for short-channel transistors, and a more thorough analysis accounting for the variation of the longitudinal field $\mathcal{E}_x$ must be included in the calculation of Q_{si} . This analysis allows to model the influence of the source and drain voltages on the channel surface potential which tends to be larger than what would be obtained from the long-channel model using the gradual channel approximation. This effect is called *drain induced barrier lowering* (DIBL) because it basically corresponds to the drain (source) voltage(s) lowering the potential barrier or increasing the surface potential in the channel, resulting in a drain current that is larger than what is obtained from the long-channel approximation.

The DIBL effect is illustrated in Figure 1.20 which plots the surface potential along the channel versus the normalized distance $\xi = x/L$ for different transistor lengths and for $V_S = 0$, $V_D = 2$ V with a junction built-in voltage $V_{bi} \cong 1$ V. The corresponding surface potential obtained from the long-channel approximation is then $\Psi_{sl} = 0.283$ V. The green curve corresponds to a long-channel device ($L = 10$ μm). It starts at the built-in voltage V_{bi} of about 1 V at the source and then immediately drops to the long-channel value Ψ_{sl} . It then stays constant along the channel up to the drain where it jumps to $V_{bi} + V_D = 3$ V. For a shorter channel of $L = 0.5$ μm , we already see the influence of the source and particularly the drain voltage on the value of the surface potential close to the source and drain junctions. The surface potential reaches the long-channel only in the middle of the channel. For an even shorter channel ($L = 0.2$ μm), the surface potential now does not even reach the long-channel value anymore. The current is then significantly larger than the one that would be predicted by the long-channel approximation, particularly in weak and moderate inversion.

DIBL has to be accounted for in short-channel device and we will show in the next Section how it can be modeled as a change of the threshold or the pinch-off voltage.

1.5.2 Modeling DIBL in the threshold or pinch-off voltage

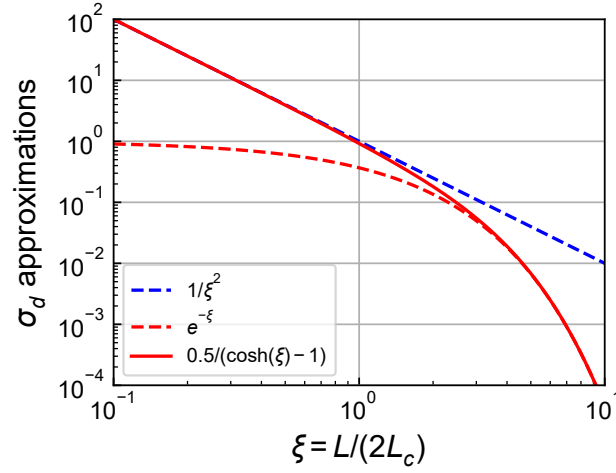


Figure 1.21: Scaling of DIBL with transistor length.

Since DIBL increases the current as the length is reduced for a given set of terminal voltages, it has traditionally been modeled in compact models as a reduction of the threshold voltage according to [15] [2] [3]

$$V_T \cong V_{T0} - \sigma_d \cdot (V_D + V_S) \quad (1.113)$$

where σ_d is a unitless parameter that has a strong dependence on the channel length according to [15]

$$\sigma_d = \frac{1}{2[\cosh(\xi) - 1]}, \quad (1.114)$$

where

$$\xi \triangleq \frac{L}{2L_c}. \quad (1.115)$$

L_c is the *characteristic length* defined as [15]

$$L_c \triangleq \sqrt{\frac{\epsilon_{si} t_d}{C_{ox}}}, \quad (1.116)$$

where t_d is an average value of the depletion thickness which is assumed to be constant. Note that for FDSOI, t_d would actually correspond to the silicon channel thickness. The scaling of σ_d with respect to $\xi \triangleq L/(2L_c)$ is plotted in Figure 1.21 together with the asymptotes given by

$$\sigma_d \cong \begin{cases} \frac{1}{\xi^2} = \left(\frac{2L_c}{L}\right)^2 & \text{for } \xi \ll 1 \\ e^{-\xi} & \text{for } \xi \gg 1. \end{cases} \quad (1.117)$$

For $L \ll 2L_c$ (or $\xi \ll 1$), σ_d scales as $1/L^2$, but for $2L_c \ll L$ ($\xi \gg 1$), σ_d decreases very fast. According to (1.113), V_T decreases as the drain voltage increases and therefore the drain current in saturation will increase with V_D instead of staying constant. Similarly to CLM, DIBL also introduces an output conductance. Because the output conductance due to CLM scales as $1/L$, for long-channel transistors, DIBL can be ignored so that the output conductance is mainly due to CLM. For this reason we will evaluate the impact of DIBL on the drain current and output conductance focusing on short-channel transistors.

In the EKV model, the DIBL effect can easily be included in the pinch-off voltage which is now redefined as [16] [17]

$$V'_P = V_P + \frac{\sigma_d}{n} \cdot (V_D + V_S) \quad (1.118)$$

or in normalized form

$$v'_p = v_p + \frac{\sigma_d}{n} \cdot (v_d + v_s) = v_p + \Delta v_p, \quad (1.119)$$

where Δv_p is the increase of the pinch-off voltage due to DIBL given by

$$\Delta v_p = \frac{\sigma_d}{n} \cdot (v_d + v_s). \quad (1.120)$$

In addition of depending on the gate voltage, the pinch-off voltage including DIBL v'_p now depends also on the drain and source voltages v_d and v_s . At a given gate and source voltages, if the drain voltage is increased, the pinch-off voltage v'_p also increases resulting in an increase of the drain current that can be observed from the linear to the saturation region.

Let's now analyze the impact of DIBL on the drain current first without considering the effect of VS.

1.5.3 Impact of DIBL on the drain current without VS

In this section we will investigate the impact of DIBL on the charges and currents in strong inversion. In a first step, for the sake of clarity and simplicity, we will ignore the effect of velocity saturation (VS). Of course this is not correct because DIBL affects short-channel transistors for which velocity saturation cannot be ignored. The combined effects of DIBL and VS are examined in a separate Notebook.

Since DIBL can be modelled as a reduction of the threshold voltage due to v_s and v_d , it basically will increase the drain current and shift the I_D - V_G transfer characteristic to the left. However it will introduce a non-zero output conductance in the I_D - V_D output characteristic. We will start analyzing the strong inversion case in the next Section.

1.5.3.1 Analysis in strong inversion without VS

We will start analyzing the impact of DIBL on the drain current in strong inversion and in the linear region, ignoring the effect of velocity saturation and CLM. The normalized drain current in strong inversion and in the linear region is given by

$$i_d = i_f - i_r. \quad (1.121)$$

where the forward and reverse currents are given by

$$i_f = q_s^2, \quad (1.122)$$

$$i_r = q_d^2. \quad (1.123)$$

The normalized source and drain charges in strong inversion including DIBL are now given by

$$q_s = \frac{v'_p - v_s}{2} = \frac{1}{2} \cdot \left[v_p - v_s + \frac{\sigma_d}{n} \cdot (v_d + v_s) \right] \quad (1.124)$$

$$= \frac{v_p - v_s}{2} + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} = q_{s,nodibl} + \frac{\Delta v_p}{2}, \quad (1.125)$$

$$q_d = \frac{v'_p - v_d}{2} = \frac{1}{2} \cdot \left[v_p - v_d + \frac{\sigma_d}{n} \cdot (v_d + v_s) \right] \quad (1.126)$$

$$= \frac{v_p - v_d}{2} + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} = q_{d,nodibl} + \frac{\Delta v_p}{2}. \quad (1.127)$$

where $q_{s,nodibl}$ and $q_{d,nodibl}$ are the source and drain charges without DIBL

$$q_{s,nodibl} = \frac{v_p - v_s}{2}, \quad (1.128)$$

$$q_{d,nodibl} = \frac{v_p - v_d}{2} \quad (1.129)$$

and Δv_p is given by (1.120). The normalized source and drain charges are increased by $\Delta v_p/2$ due to DIBL. This will obviously have an impact on the currents and transconductances. Notice that contrary to the long-channel case where q_s depends on $v_p - v_s$ and q_d on $v_p - v_d$, now both q_s and q_d depend on v_p , v_s and v_d .

The drain current in strong inversion can then be written as

$$\begin{aligned} i_d &= \left[v_p + \left(\frac{\sigma_d}{n} - \frac{1}{2} \right) v_d + \left(\frac{\sigma_d}{n} - \frac{1}{2} \right) v_s \right] \cdot \frac{v_d - v_s}{2} \\ &= \left[v_p - \frac{v_d + v_s}{2} + \frac{\sigma_d}{n} (v_d + v_s) \right] \cdot \frac{v_d - v_s}{2} \\ &= \left[v_p - \frac{v_d + v_s}{2} \right] \cdot \frac{v_d - v_s}{2} + \frac{\sigma_d}{n} \cdot \frac{v_d^2 - v_s^2}{2} \\ &= i_{d, no \text{ dibl}} + \frac{\sigma_d}{n} \cdot \frac{v_d^2 - v_s^2}{2}. \end{aligned} \quad (1.130)$$

where

$$i_{d, no \text{ dibl}} \triangleq \left[v_p - \frac{v_d + v_s}{2} \right] \cdot \frac{v_d - v_s}{2} \quad (1.131)$$

is the drain current in the linear region without DIBL. The drain current in the linear region including DIBL increases faster with v_d than without DIBL with the difference growing as v_d^2 .

If velocity saturation is ignored, saturation is reached when $q_d = 0$. We can find the saturation voltage $v_{d,sat}$ defined as the drain voltage for which $q_d = 0$ which is given by

$$v_{d,sat} = \frac{v_p + \frac{\sigma_d}{n} v_s}{1 - \frac{\sigma_d}{n}}. \quad (1.132)$$

For $v_s = 0$, the saturation voltage is $1/(1 - \frac{\sigma_d}{n})$ larger than v_p .

In saturation (i.e. for $v_{d,sat} < v_d$), $i_r = 0$ and the drain current is given by

$$\begin{aligned} i_{d,sat} &= i_f = q_s^2 = \left(\frac{v_p' - v_s}{2} \right)^2 = \frac{1}{4} \cdot \left[v_p - v_s + \frac{\sigma_d}{n} \cdot (v_d + v_s) \right]^2 \\ &= \frac{1}{4} \cdot (v_p - v_s)^2 + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} \cdot \left[v_p - v_s + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} \right] \\ &= i_{d,sat, nodibl} + \frac{\Delta v_p}{2} \cdot \left[v_p - v_s + \frac{\Delta v_p}{2} \right]. \end{aligned} \quad (1.133)$$

where

$$i_{d,sat, nodibl} \triangleq \left(\frac{v_p - v_s}{2} \right)^2 \quad (1.134)$$

is the drain current in saturation without DIBL. Because of DIBL, the drain current in saturation increases quadratically with v_d . Since usually $\sigma_d \ll 1$, $v_p - v_s \gg \sigma_d/n \cdot (v_d + v_s)/2$ and the drain current in saturation is approximately given by

$$i_{d,sat} \cong i_{d,sat, nodibl} + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} \cdot (v_p - v_s) = i_{d,sat, nodibl} + \frac{\Delta v_p}{2} \cdot (v_p - v_s), \quad (1.135)$$

which grows now linearly with v_d .

The normalized charges q_s and q_d are plotted versus v_d in Figure 1.22 for $v_s = 0$, $n = 1.3$ and $\sigma_d = 0.217$ for which the normalized saturation voltage is increased from $v_p = 50$ to $v_{d,sat} = 60$. We see that q_s increases with respect to v_d instead of staying constant as it would for the long-channel case without DIBL. On the other hand, q_d decreases until it reaches zero at the onset of saturation. We observe that, because of DIBL, the saturation voltage given by (1.132) is higher than v_p . Note that $\sigma_d = 0.217$ is already a fairly large value that has been used in Figure 1.22 to increase the saturation voltage by 10 U_T .

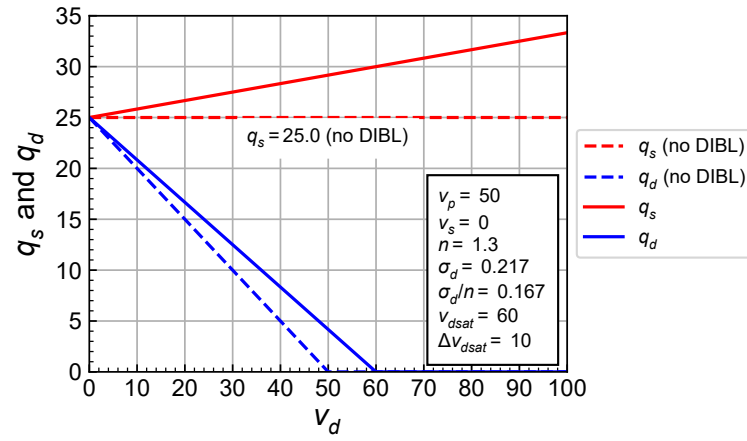


Figure 1.22: Impact of DIBL on the source and drain charges.

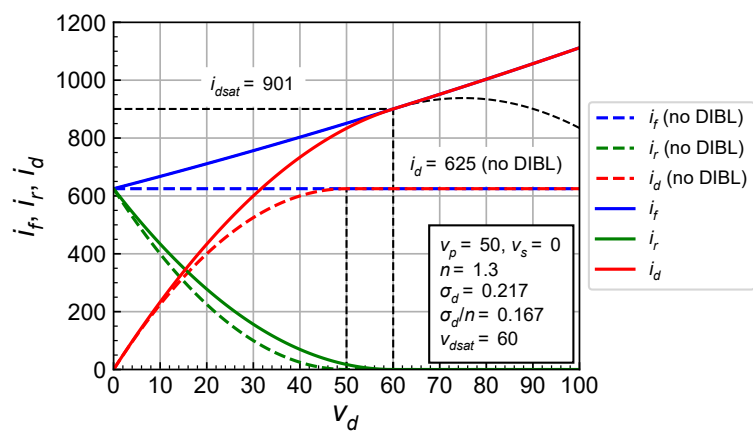


Figure 1.23: Impact of DIBL on the forward, reverse and drain currents in strong inversion.

The normalized forward, reverse and drain currents are plotted versus v_d in Figure 1.23 for the same value $\sigma_d = 0.217$. We see that the forward current increases with respect to v_d instead of remaining constant as it would in the long-channel case without DIBL. The reverse current is also slightly higher than the reverse current without DIBL. It decreases until reaching zero at $v_d = v_{d,sat}$. The drain current is larger than its value without DIBL and continues to increase in saturation for $v_{d,sat} < v_d$.

1.5.3.2 Analysis in weak inversion without VS

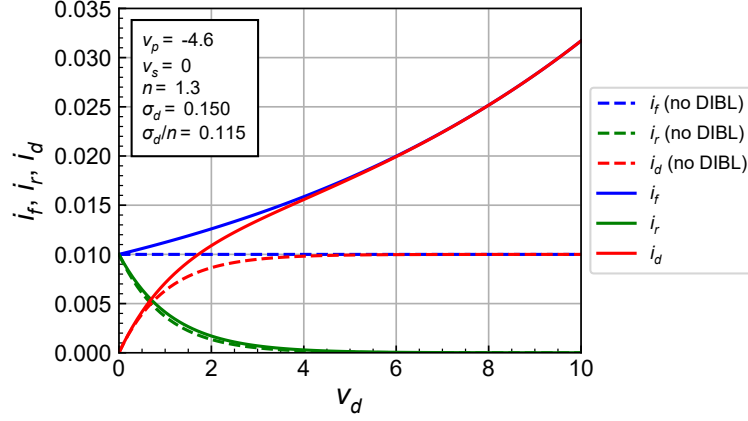


Figure 1.24: Impact of DIBL on the forward, reverse and drain currents in weak inversion.

The forward and reverse currents in weak inversion are given by

$$i_f = q_s, \quad (1.136)$$

$$i_r = q_d, \quad (1.137)$$

with the normalized source and drain charges q_s and q_d given by

$$q_s = e^{v'_p - v_s} = e^{v_p - v_s + \frac{\sigma_d}{n} \cdot (v_d + v_s)} = e^{\frac{\sigma_d}{n} \cdot (v_d + v_s)} \cdot e^{v_p - v_s} = e^{\Delta v_p} \cdot q_{s,nodibl}, \quad (1.138)$$

$$q_d = e^{v'_p - v_d} = e^{v_p - v_d + \frac{\sigma_d}{n} \cdot (v_d + v_s)} = e^{\frac{\sigma_d}{n} \cdot (v_d + v_s)} \cdot e^{v_p - v_d} = e^{\Delta v_p} \cdot q_{d,nodibl}, \quad (1.139)$$

where $q_{s,nodibl}$ and $q_{d,nodibl}$ are the source and drain charges without DIBL

$$q_{s,nodibl} = e^{v_p - v_s}, \quad (1.140)$$

$$q_{d,nodibl} = e^{v_p - v_d}. \quad (1.141)$$

DIBL has a significant effect on the source and drain charges since they are multiplied by $\exp[\Delta v_p]$.

The drain current in weak inversion is then given by

$$i_d = i_f - i_r = q_s - q_d = e^{\frac{\sigma_d}{n} \cdot (v_d + v_s)} \cdot [e^{v_p - v_s} - e^{v_p - v_d}] = e^{\Delta v_p} \cdot i_{d,nodibl} \quad (1.142)$$

The drain current in weak inversion is multiplied by $\exp[\Delta v_p]$. The normalized forward, reverse and drain currents are plotted versus v_d in Figure 1.24 in weak inversion for $\sigma_d = 0.15$. We see that the forward current i_f and drain current i_d increase in saturation because of DIBL. The impact of DIBL on the drain current ignoring VS is strong, significantly degrading the output characteristic for short-channel devices.

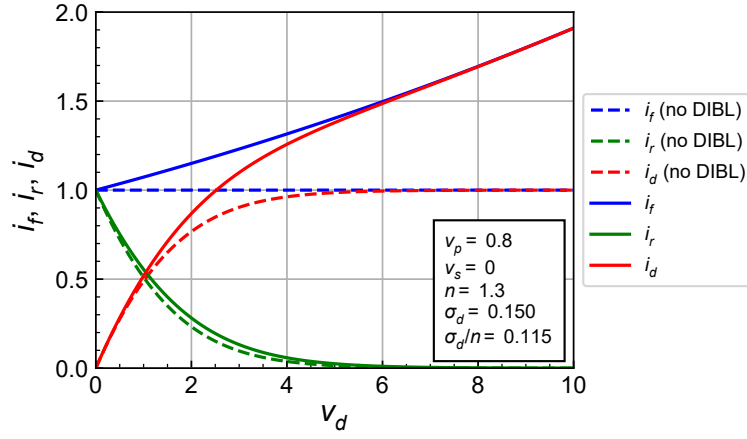


Figure 1.25: Impact of DIBL on the forward, reverse and drain currents in moderate inversion.

1.5.3.3 Analysis in all regions of inversion without VS

The normalized forward and reverse currents in all modes of inversion but ignoring VS are given by

$$i_f = q_s^2 + q_s, \quad (1.143)$$

$$i_r = q_d^2 + q_d. \quad (1.144)$$

The relation between charges and voltages are given by

$$v_p' - v_s = 2q_s + \ln(q_s), \quad (1.145)$$

$$v_p' - v_d = 2q_d + \ln(q_d), \quad (1.146)$$

with v_p' given by (1.119).

The normalized forward, reverse and drain currents are plotted versus v_d in Figure 1.25 in moderate inversion for $\sigma_d = 0.15$. We see that the forward current i_f and drain current i_d increase in saturation because of DIBL but the impact is less than in weak inversion for the same value of σ_d .

We see that DIBL can have a dramatic impact on the output characteristic in weak inversion where the current can increase exponentially with v_d . The above DIBL model valid in all regions of inversion is checked against measurements in the next Section.

1.5.3.4 Experimental validation

The above model of DIBL is validated on a short nMOS transistor from a 28 nm bulk CMOS technology. Figure 1.26 shows the I_D - V_D characteristic measured in strong inversion. We can observe that the simple DIBL model can fit the measurements very well provided an additional scaling parameter i_{scale} is added. This additional parameter is needed to compensate for the fact that the sEKV parameter I_{spec} has been extracted at the highest drain voltage $V_D = 1.1$ V ($V_D = 42.5$). It also accounts for the current reduction in saturation due to VS which is not accounted for in this simple model. The corresponding output conductance are shown in Figure 1.27. We can observe that the model predicts an increase of g_{ds} wrt v_d which is actually not observed in the measured data. This is due to the fact that according to (1.133), the drain current in saturation increases with the square of v_d and hence the output conductance with v_d . A better model of the output conductance will be derived in the Notebook dedicated to the output conductance.

Figure 1.28 shows the I_D - V_D characteristic measured in weak inversion. We can see the very strong impact that DIBL can have on short transistors in weak inversion. Note that such an output characteristic is almost unusable for analog circuit. For this reason a non-minimum channel length

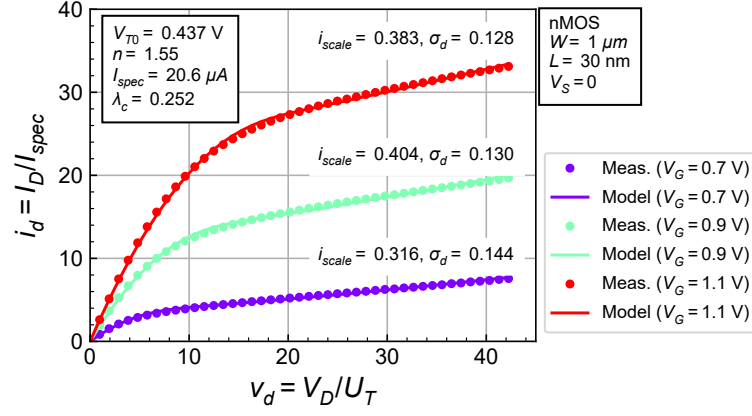


Figure 1.26: Validation of the DIBL output characteristic model in strong inversion (without accounting for VS) for a short nMOS transistor from a 28 nm bulk CMOS technology.

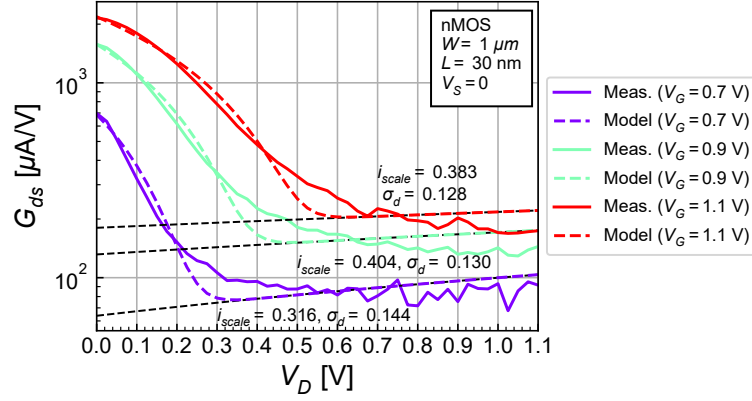


Figure 1.27: Validation of the DIBL output conductance model in strong inversion (without accounting for VS) for a short nMOS transistor from a 28 nm bulk CMOS technology.

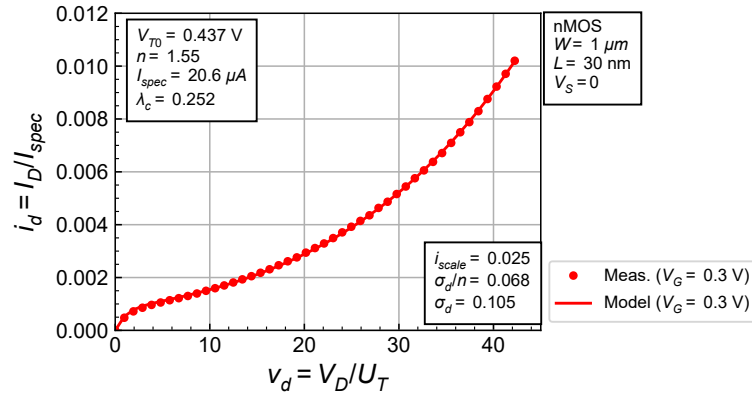


Figure 1.28: Validation of the DIBL output characteristic model in weak inversion (without accounting for VS) for a short nMOS transistor from a 28 nm bulk CMOS technology.

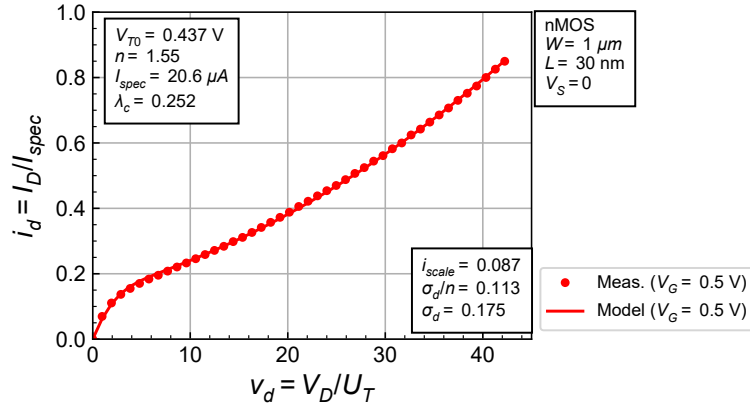


Figure 1.29: Validation of the DIBL output characteristic model in moderate inversion (without accounting for VS) for a short nMOS transistor from a 28 nm bulk CMOS technology.

should be chosen when biasing the transistor in weak inversion. Things improve slightly when moving to moderate inversion as shown in Figure 1.29. However, the output conductance is still huge and the minimum length device will perform poorly in an analog circuit when biased in moderate inversion. Again, this can be circumvented by choosing a non-minimum length

1.5.4 Impact of DIBL on the transconductances in saturation without VS

1.5.4.1 General expression valid in saturation from weak to strong inversion

The gate transconductance in saturation is given by

$$g_{m_{sat}} = \frac{\partial i_{d_{sat}}}{\partial v_g} = \frac{1}{n} \cdot \frac{\partial i_{d_{sat}}}{\partial v_p}. \quad (1.147)$$

Without VS, in saturation, $i_{d_{sat}} \cong i_f$ and hence

$$g_{m_{sat}} = \frac{1}{n} \cdot \frac{\partial i_f}{\partial v_p} = \frac{1}{n} \cdot \frac{\partial i_f}{\partial q_s} \cdot \left(\frac{\partial v_p}{\partial q_s} \right)^{-1}. \quad (1.148)$$

with

$$i_f = q_s^2 + q_s. \quad (1.149)$$

The normalized source charge q_s is related to v_p , v_s and v_d according to

$$v'_p - v_s = 2q_s + \ln(q_s) \quad (1.150)$$

with

$$v'_p = v_p + \Delta v_p. \quad (1.151)$$

where Δv_p is given by (1.120). The pinch-off voltage is then given by

$$v_p = v_s - \Delta v_p + 2q_s + \ln(q_s). \quad (1.152)$$

We therefore have

$$\frac{\partial i_f}{\partial q_s} = 2q_s + 1 \quad (1.153)$$

and

$$\frac{\partial v_p}{\partial q_s} = 2 + \frac{1}{q_s} = \frac{2q_s + 1}{q_s}. \quad (1.154)$$

The gate transconductance is then given by

$$g_{m_{sat}} = \frac{q_s}{n}. \quad (1.155)$$

The source transconductance is given by

$$g_{ms} = -\frac{\partial i_{d_{sat}}}{\partial v_s} \cong -\frac{\partial i_f}{\partial v_s} = -\frac{\partial i_f}{\partial q_s} \cdot \frac{\partial q_s}{\partial v_s} = -\frac{\partial i_f}{\partial q_s} \cdot \left(\frac{\partial v_s}{\partial q_s}\right)^{-1}. \quad (1.156)$$

The source voltage including DIBL is given by

$$v_s = \frac{1}{1 - \frac{\sigma_d}{n}} \cdot \left[v_p + \frac{\sigma_d}{n} v_d - 2q_s - \ln(q_s) \right] \quad (1.157)$$

from which we get

$$\frac{\partial v_s}{\partial q_s} = -\frac{1}{1 - \frac{\sigma_d}{n}} \cdot \frac{2q_s + 1}{q_s}. \quad (1.158)$$

The source transconductance in strong inversion is then given by

$$g_{ms} = \left(1 - \frac{\sigma_d}{n}\right) \cdot q_s. \quad (1.159)$$

The ratio g_{ms}/g_m in strong inversion and saturation is no more equal simply to n but

$$\frac{g_{ms}}{g_m} = n \left(1 - \frac{\sigma_d}{n}\right) = n - \sigma_d, \quad (1.160)$$

which is lower than n .

1.5.4.2 Analysis in strong inversion and saturation

We assume that $v_{d_{sat}} < v_d$ where the saturation voltage $v_{d_{sat}}$ is slightly larger than v_p according to (1.132). Using the general derivation, the gate transconductance in strong inversion and saturation without VS is then given by

$$g_{m_{sat}} = \frac{q_s}{n} = \frac{1}{n} \cdot \left[\frac{v_p - v_s}{2} + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} \right] = g_{m_{sat},nodibl} + \Delta g_{m_{sat},dibl}, \quad (1.161)$$

where $g_{m_{sat},nodibl}$ is the gate transconductance in strong inversion and saturation without DIBL

$$g_{m_{sat},nodibl} = \frac{1}{n} \cdot \frac{v_p - v_s}{2} \quad (1.162)$$

and $\Delta g_{m_{sat}}$ is the transconductance increase due to DIBL

$$\Delta g_{m_{sat},dibl} = \frac{\sigma_d}{n^2} \cdot \frac{v_d + v_s}{2} = \frac{\Delta v_p}{2n}. \quad (1.163)$$

The impact of DIBL on the gate transconductance is an increase $\Delta g_{m_{sat}} = \Delta v_p/(2n)$.

Using the above general derivation, the source transconductance in strong inversion and saturation is then given by

$$g_{ms} = \left(1 - \frac{\sigma_d}{n}\right) \cdot q_s = \frac{v_p - v_s}{2} + \frac{\sigma_d}{n} \cdot \left[\left(1 - \frac{\sigma_d}{n}\right) \frac{v_d + v_s}{2} - \frac{v_p - v_s}{2} \right] = g_{ms,nodibl} + \Delta g_{ms}, \quad (1.164)$$

where $g_{ms,nodibl}$ is the source transconductance without DIBL

$$g_{ms,nodibl} = \frac{v_p - v_s}{2} \quad (1.165)$$

and Δg_{ms} is the increase in the source transconductance due to DIBL

$$\Delta g_{ms} = \frac{\sigma_d}{n} \cdot \left[\left(1 - \frac{\sigma_d}{n}\right) \frac{v_d + v_s}{2} - \frac{v_p - v_s}{2} \right] = \frac{\sigma_d}{n} \cdot \left[\left(1 - \frac{\sigma_d}{n}\right) \cdot \frac{v_d - v_{d_{sat}}}{2} + v_s \right]. \quad (1.166)$$

The source transconductance in saturation including DIBL is increased by Δg_{ms} .

1.5.4.3 Analysis in weak inversion and saturation

Because of DIBL, the normalized source charge q_s , forward current i_f and drain current i_d are all multiplied by $\exp[\Delta v_p]$. Similarly, the gate transconductance in weak inversion and saturation is also multiplied by the factor

$$g_{m_{sat}} = \frac{q_s}{n} = \frac{i_f}{n} = e^{\Delta v_p} \cdot g_{m_{sat}, nodibl}, \quad (1.167)$$

where $g_{m_{sat}, nodibl}$ is the gate transconductance without DIBL.

The source transconductance in saturation including DIBL is multiplied by $(1 - \sigma_d/n) \cdot \exp[\Delta v_p]$.

$$g_{ms} = \left(1 - \frac{\sigma_d}{n}\right) \cdot e^{\Delta v_p} \cdot g_{ms, nodibl}. \quad (1.168)$$

1.5.5 Output conductance in saturation due to DIBL without VS

1.5.5.1 General expression

From Figure 1.23 we see that the drain current in saturation does no more remain constant but increases with v_d because of DIBL. This is due to the forward current i_f which now increases with v_d due to DIBL instead of remaining constant. The output conductance is hence no more zero, which obviously affects the self-gain G_m/G_{ds} . It is therefore important to be able to predict G_{ds} (or its normalized form g_{ds}) as accurately as possible.

Similarly to what was done in Section 1.4.3, we will consider v_s constant. The output conductance is then defined by (1.106). In saturation, the change of the forward current and hence of the drain current due to a change of the v_d voltage is due to DIBL through an increase of the pinch-off voltage v'_p . We can then write

$$g_{ds} = \frac{\partial i_{d_{sat}}}{\partial v_d} = \frac{\partial i_{d_{sat}}}{\partial v'_p} \cdot \frac{\partial v'_p}{\partial v_d}. \quad (1.169)$$

where v'_p is given by (1.119). The derivative $\partial v'_p / \partial v_d$ is therefore simply equal to

$$\frac{\partial v'_p}{\partial v_d} = \frac{\sigma_d}{n}. \quad (1.170)$$

On the other hand, the derivative $\partial i_{d_{sat}} / \partial v'_p$ is related to the transconductance according to

$$\frac{\partial i_{d_{sat}}}{\partial v'_p} = \frac{\partial i_{d_{sat}}}{\partial v_p} = \frac{\partial i_{d_{sat}}}{\partial v_g} \cdot \frac{\partial v_g}{\partial v_p} = n \cdot g_{m_{sat}}. \quad (1.171)$$

The output conductance in saturation is therefore proportional to the gate transconductance in saturation

$$g_{ds} = \sigma_d \cdot g_{m_{sat}}. \quad (1.172)$$

Note that the above derivation is also valid when including VS as long as the transconductance expression accounts for both DIBL and VS. Note also that the above derivation is mostly valid in strong inversion. Indeed, the reverse current of a long-channel transistor in weak inversion is a function of v_d also in saturation, but usually the later can be neglected for short-channel devices compared to the contribution of DIBL.

The transistor self-gain due to DIBL without accounting for CLM and VS is then simply

$$A_{dc} = \frac{G_m}{G_{ds}} = \frac{g_m}{g_{ds}} = \frac{1}{\sigma_d}. \quad (1.173)$$

From the general expression of the gate transconductance (1.155) accounting for DIBL (but not VS), we can write

$$g_{ds} = \frac{\sigma_d}{n} \cdot q_s. \quad (1.174)$$

The output conductance due to DIBL can then easily be calculated from the expression of the gate transconductance in saturation derived above.

Similarly to what is done for CLM, we can define a DIBL voltage as

$$\frac{1}{v_{dibl}} = \frac{g_{ds}}{i_{d_{sat}}} = \sigma_d \cdot \frac{g_{m_{sat}}}{i_{d_{sat}}} = \frac{\sigma_d/n}{q_s}. \quad (1.175)$$

We see that $1/v_{dibl}$ is proportional to the normalized G_m/I_D function in saturation which should include both DIBL and VS. Now, $i_{d_{sat}}$ is nothing else than the inversion coefficient IC and hence

$$\frac{1}{v_{dibl}} = \frac{g_{ds}}{i_{d_{sat}}} = \frac{\sigma_d}{n} \cdot \frac{g_{ms}}{IC}, \quad (1.176)$$

where g_{ms} is the normalized source transconductance including VS and DIBL.

To use the relation between the output conductance g_{ds} and the gate transconductance in saturation $g_{m_{sat}}$ given by (1.172), we would need to derive the expression of the gate transconductance in strong inversion and saturation including both DIBL and VS. This is done in the dedicated Notebook. Here we only will account for DIBL ignoring VS.

1.5.5.2 Analysis in strong inversion and saturation

From (1.161) we get the output conductance in strong inversion and saturation due to DIBL ignoring VS

$$g_{ds} = \sigma_d \cdot g_{m_{sat}} = \frac{\sigma_d}{n} \cdot \left[\frac{v_p - v_s}{2} + \frac{\sigma_d}{n} \cdot \frac{v_d + v_s}{2} \right] = \frac{\sigma_d}{n} \cdot \left[\frac{v_p - v_s}{2} + \frac{\Delta v_p}{2} \right]. \quad (1.177)$$

We see from (1.177) that the output conductance increases with $v_p - v_s$ and also depends on v_d . Usually $\sigma_d \ll 1$ and hence $v_p - v_s \gg \Delta v_p$. (1.177) can therefore be approximated by

$$g_{ds} \cong \frac{\sigma_d}{n} \cdot \frac{v_p - v_s}{2}. \quad (1.178)$$

1.5.5.3 Analysis in weak inversion and saturation

In weak inversion, the drain charge does not vanish to zero but decreases exponentially wrt v_d . This gives a finite output conductance even without DIBL. If we account for this and add the effect of DIBL we get

$$g_{ds} = e^{\Delta v_p} \cdot e^{v_p} \cdot \left[\frac{\sigma_d}{n} e^{-v_s} + \left(1 + \frac{\sigma_d}{n} \right) \cdot e^{-v_d} \right] \cong \frac{\sigma_d}{n} \cdot e^{\Delta v_p} \cdot e^{v_p - v_s}. \quad (1.179)$$

1.5.6 Impact of DIBL on the transconductances in saturation including VS

The derivation of the transconductances including both effects of DIBL and VS is tedious and is not presented here. The detailed analysis given in the Notebook dedicated to DIBL shows that a good and simple approximation of the normalized gate transconductance g_m versus the inversion coefficient that is valid in all regions of inversion is obtained by simply reusing the short-channel expression of g_m accounting for VS only and replacing λ_c by λ_{dibl}

$$g_{m_{sat}} = \frac{1}{n} \cdot \frac{\sqrt{4IC + 1 + (\lambda_{dibl} IC)^2} - 1}{2 + \lambda_{dibl}^2 IC}, \quad (1.180)$$

with

$$\frac{1}{\lambda_{dibl}} \triangleq \frac{1}{\lambda_c} + \frac{\Delta v_p}{2}. \quad (1.181)$$

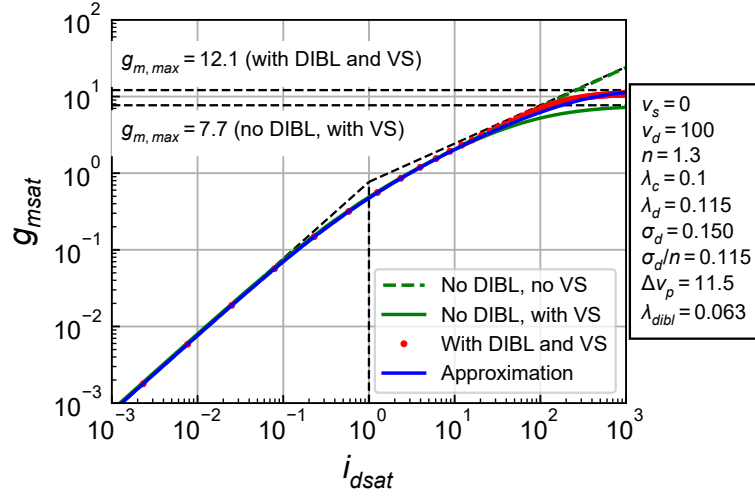


Figure 1.30: g_{msat} versus IC including DIBL and VS valid in all regions of inversion.

This approximation is plotted in Figure 1.30 by the blue line which is very close to the full expression corresponding to the red dots. For $IC \gg 1/\lambda_{dibl}$, (1.180) tends to $1/\lambda_{dibl}$ which is higher than $1/\lambda_c$ because of DIBL.

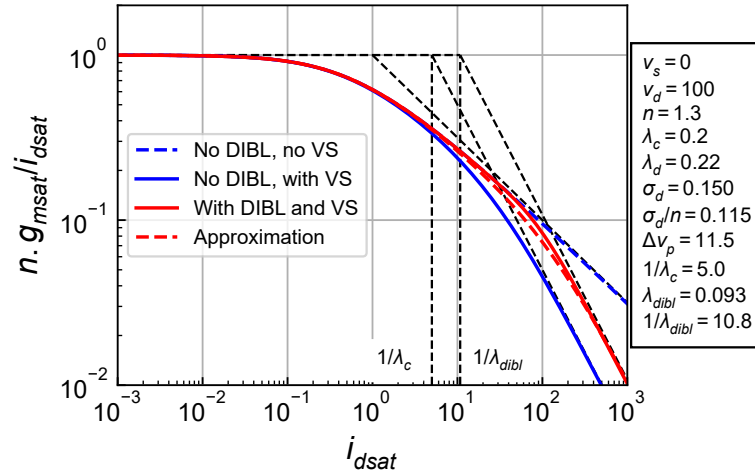


Figure 1.31: $n \cdot g_{msat}/IC$ versus IC including DIBL and VS valid in all regions of inversion.

Finally, the normalized G_m/I_D characteristic $n \cdot g_{msat}/IC$ is plotted versus IC in Figure 1.31. We see that the weak inversion asymptote remains at unity, while the strong inversion asymptote is shifted to higher inversion coefficient because of DIBL. The intersection of the strong inversion asymptote with the horizontal unity line is therefore shifted from $1/\lambda_c$ to $1/\lambda_{dibl}$ because of DIBL. We can reuse (1.180) to approximate the normalized G_m/I_D characteristic as

$$\frac{n \cdot g_{msat}}{IC} = \frac{\sqrt{4i_{dsat} + 1 + (\lambda_{dibl} IC)^2} - 1}{IC \cdot (2 + \lambda_{dibl}^2 IC)}, \quad (1.182)$$

which is plotted in Figure 1.31 as the dashed red line which is close to the red line corresponding to the full expression.

! Important

The fact that we can use the expression of the normalized gate transconductance g_m and of the normalized G_m/I_D including the effect of VS only and account for the additional DIBL effect by simply replacing λ_c by a slightly higher value λ_{dibl} explains why we get very good fits of the transconductances and the normalized G_m/I_D . The value extracted for λ_c actually corresponds to λ_{dibl} and hence also includes the effect of DIBL for short-channel transistors.

**Notebook: Output Conductance in Saturation due to CLM and DIBL**

This Notebook proposes a more accurate output conductance model that is valid in saturation from weak to strong inversion and covers short to long-channel transistors. It also tries to answer the question: what is the minimum transistor length for avoiding a strong degradation of the intrinsic gain [18] ?

1.6 Short-channel effects on thermal noise

Noise represents the ultimate lower limit of signal resolution in analog signal processing. For this reason it is key for the circuit designer to evaluate the noise of analog circuits whether at low-frequency or at RF. In CMOS circuits, the most important contributor to noise is the MOSFET. At RF, the noise in linear circuits is essentially due to the transistors white noise. At low-frequency the MOSFET noise is dominated by flicker noise. However, the latter can be strongly reduced using circuit techniques such as chopper stabilization, autozero or correlated double (or multiple) sampling which are presented in Chapter ?? [19] [20]. The remaining noise is then the white noise which is either thermal noise if the transistor is biased in strong inversion, or shot noise if the transistor is biased in weak inversion, or a mix if the transistor is biased in the moderate inversion region [1].

The transistor white noise is characterized by its transconductance G_m and the thermal noise excess factor γ_n , which is defined in ???. Since the transconductance also sets the gain and the bandwidth, it is essential for the designer to find the right G_m resulting from the trade-off between gain, bandwidth and noise. To this purpose the designer can use various figures-of-merit (FoMs) helping exploring the design space. The thermal noise excess factor γ_n is an essential FoM for the design of low-noise circuits since it evaluates how much noise is produced for a given transconductance. It can hence be used to derive the required transconductance needed for achieving a given input-referred noise or noise figure for a given thermal noise excess factor γ_n . The G_m/I_D is another FoM that evaluates how much transconductance you get for a given current. It can then be used to derive the required bias current after having chosen the most appropriate region of operation using the inversion coefficient [21] [8].

As shown in ??? that for a long-channel transistor the thermal noise excess factor γ_n is close to unity and shows only a slight bias dependence. However, in advanced technologies, short channel devices biased in strong inversion show a γ_n that can become larger than one. Since the input-referred noise resistance is proportional to γ_n/G_m , if γ_n increases, the transconductance has to be increased by the same amount in order to maintain the same input-referred noise resistance (or noise figure). This increase in G_m unavoidably results in an increase of the bias current and of the power consumption. It is therefore crucial for the designer to have a sufficiently accurate model of the thermal noise excess factor valid over a wide range of current (or inversion coefficient).

The long-channel thermal noise model derived in Chapter ?? assumed that the mobility was constant along the channel and that the channel length was sufficiently long so that velocity saturation and channel length modulation could be neglected. These assumptions are obviously not valid anymore for short-channel devices where velocity saturation and channel length modulation effects have to be accounted for. Mobility reduction due to the vertical field also greatly influences the thermal noise and has also to be included. Finally, the increase of the carrier temperature at high longitudinal field also

needs to be accounted for. The impact of all these effects on short-channel transistor is described in details in [22] [23] [24] [25]. In this section we will derive a simple empirical model of the thermal noise excess factor for short-channel transistors in terms of the most important design parameter, namely the inversion coefficient.

1.6.1 Empirical model of the thermal noise excess factor for short-channel transistors

The thermal noise excess factor γ_n of short-channel transistor can increase significantly because of various short-channel effects like VS and CLM [22] [23] [24] [25]. The behavior of γ_n with respect to the inversion coefficient can be modeled by a simple empirical model given by [26]

$$\gamma_{n_{sat}} \cong \gamma_{nwi} \cdot (1 + \alpha_n \cdot IC). \quad (1.183)$$

where α_n is used as a fitting parameter but can be related to the velocity saturation parameter λ_c according to

$$\alpha_n = \frac{n}{2} \cdot \lambda_c^2. \quad (1.184)$$

This empirical model has been validated on 5 different technologies from 240nm down to 40 nm. The empirical model (1.183) is compared to measured data in Figure 1.32 after extraction of parameters γ_{nwi} and α_n by curve fitting. We can extract the corresponding VS parameter according to

$$\lambda_c = \sqrt{\frac{2\alpha_n}{n}}. \quad (1.185)$$

The fitting parameters γ_{nwi} and α_n , the imposed slope factor n and the corresponding extracted value of λ_c are presented in Table 1.4. Figure 1.33 shows that α_n is actually scaling as $1/L$.

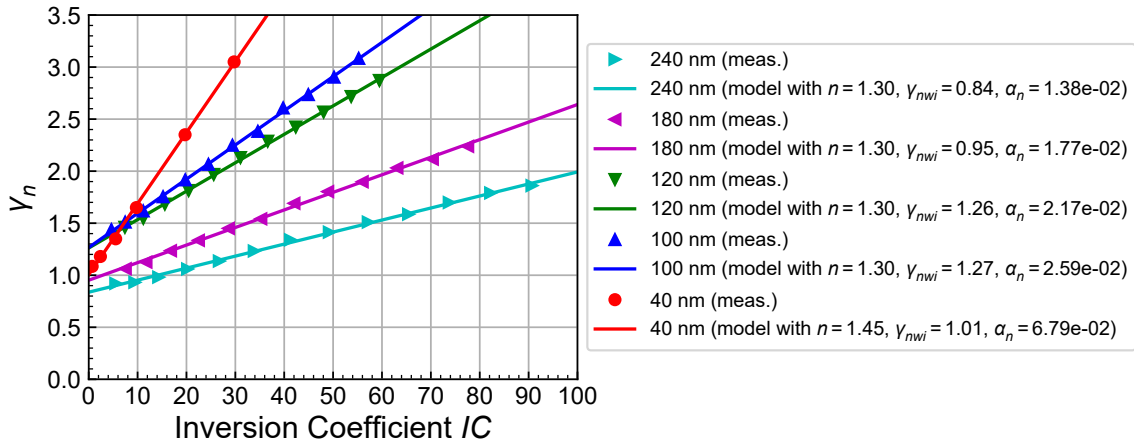
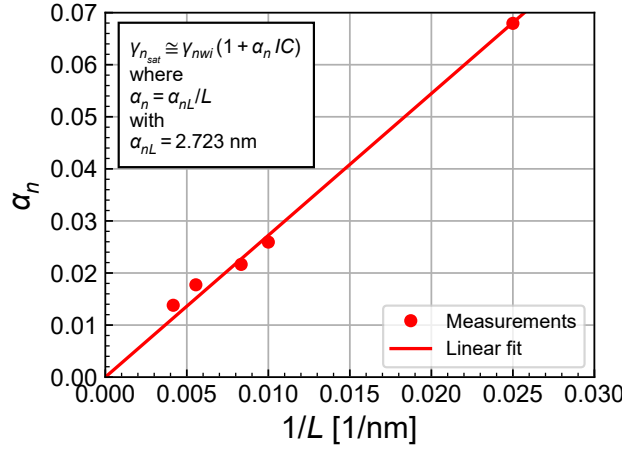


Figure 1.32: Experimental validation of the thermal noise excess factor empirical model.

Table 1.4: Fitting parameters used for thermal noise excess factor model used in Figure 1.32.

Techno	γ_{nwi}	α_n	n	λ_c
240nm	0.837034	0.0138037	1.3	0.135665
180nm	0.952055	0.0177346	1.3	0.153773
120nm	1.26136	0.0216531	1.3	0.169914
100nm	1.26643	0.0259371	1.3	0.185965
40nm	1.0053	0.0679447	1.45	0.306132

Figure 1.33: Scaling of the α_n parameter with $1/L$.

1.7 The simplified EKV model

The simplified EKV mode (sEKV) corresponds to the normalized drain current in saturation given by (1.51)

$$IC = \frac{2(q_s^2 + q_s)}{1 + \sqrt{1 + \lambda_c^2 (q_s^2 + q_s)}}, \quad (1.186)$$

which is completed with the voltage-charge relation giving the normalized saturation voltage $v_p - v_s$ in terms of the source charge q_s given by [1]}

$$v_p - v_s = 2q_s + \ln(q_s), \quad (1.187)$$

where

$$v_p = \frac{v_g - v_{t0}}{n}. \quad (1.188)$$

The inversion coefficient given by (1.186) can actually be solved to express q_s as a function of IC leading to

$$q_s = \frac{1}{2} \left(\sqrt{4IC + 1 + (\lambda_c IC)^2} - 1 \right). \quad (1.189)$$

which can then be replaced in (1.187) to obtain an expression of the saturation voltage $v_p - v_s$ in terms of IC . Unfortunately, the resulting equation cannot be inverted to express the inversion coefficient in terms of the saturation voltage $v_p - v_s$. This is actually not a problem when we use IC as the main design parameter from which we can then compute the saturation voltage using (1.189) and (1.187).

The simplified EKV model only requires four parameters to fit the I_D - V_G transfer characteristic, namely:

- the *slope factor* n ,
- the *specific current per square* $I_{spec\Box}$,
- the *threshold voltage* V_{T0} and
- the *velocity saturation parameter* λ_c (or L_{sat}).

Table 1.5: Typical parameter values for a 28-nm process.

Type	n	$I_{spec\Box}$ [nA]	V_{T0} [V]	L_{sat} [nm]
n-channel	1.1 - 1.5	850	0.4 - 0.55	15 - 25
p-channel	1.1 - 1.5	350	0.35 - 0.5	15 - 25

The methodology to extract these parameters from measured data is sketched in a specialized Notebook. Typical values for these parameters for a 28 nm bulk CMOS process are given in Table 1.5.

Note that the sEKV model does not include the effect of mobility reduction due to the vertical field (MRV). The reason is that because of voltage scaling the gate voltage has been reduced considerably making MRV less important. However, for older technologies it can be shown that it can actually be covered to a certain extent by the VS parameter λ_c .

Despite the very few number of parameters, the simplified EKV model can perfectly match data even from advanced CMOS processes as will be shown below.

1.7.1 Bulk technologies

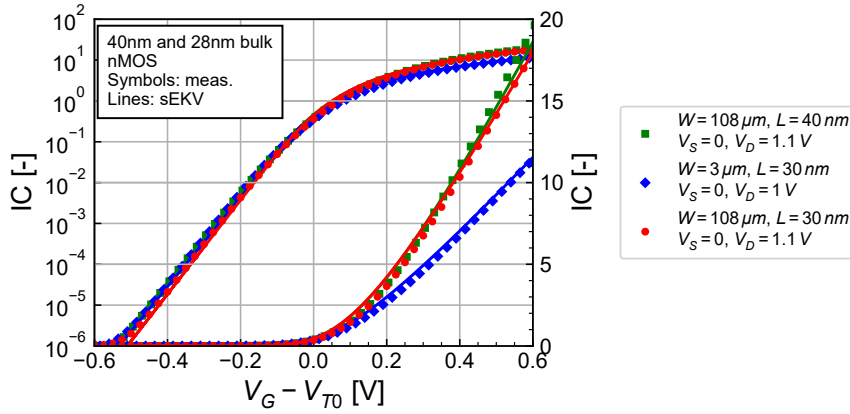


Figure 1.34: IC versus the overdrive voltage $V_G - V_{T0}$ measured in saturation on minimum length nMOS transistors from a 40 nm and two different 28 nm bulk CMOS processes [21].

The IC versus $V_G - V_{T0}$ transfer characteristics are plotted in Figure 1.34 and compared to measurements made on wide and minimal length transistors from three different bulk CMOS processes, namely a 40 nm and two different 28 nm processes. Although the drain current is measured by sweeping the gate voltage, the sEKV model is calculated from the measured current by first normalizing it to the specific current for each transistor to get the inversion coefficient, from which the overdrive voltage is computed using (1.189) and (1.187). Despite the very few number of parameters, the simple model fits the measurements very well over more than 6 decades of current. Note that the extraction of the parameters I_{spec} and L_{sat} is done for several different geometries (in particular different length) illustrating the rather good scalability of the simplified model. Notice that the measured points and analytical models of the $W = 108\mu m$, $L = 30nm$ (red circles) and $W = 108\mu m$, $L = 40nm$ (green squares) transistors almost fall on top of each other, indicating that the normalization process almost completely strips off the technology dependence. The difference with the $W = 3\mu m$, $L = 30nm$ (blue diamond) characteristic is due to a slightly larger value of λ_c . In other words, the four parameters almost fully characterize the technology at least for the transfer characteristics in saturation and in the regions of operation used for analog circuit design.

The most important small-signal parameter is without doubt the *gate transconductance* G_m . Remember that since in the EKV model the voltages are all referred to the bulk, we can define two other transconductances: the *source transconductance* $G_{ms} \triangleq -\partial I_D / \partial V_S$ and the *drain transconductance* $G_{md} \triangleq \partial I_D / \partial V_D$ [1]. Note that G_{md} should not be confused with the output conductance G_{ds} . In saturation $G_{md} = 0$ and $G_{ms} = n \cdot G_m$.

The *normalized source transconductance* in saturation G_{ms} is given by (1.71), which is repeated here for convenience

$$g_{ms} \triangleq \frac{G_{ms}}{G_{spec}} = \frac{n \cdot G_m}{G_{spec}} = \frac{\sqrt{4IC + 1 + (\lambda_c IC)^2} - 1}{2 + \lambda_c^2 IC},$$

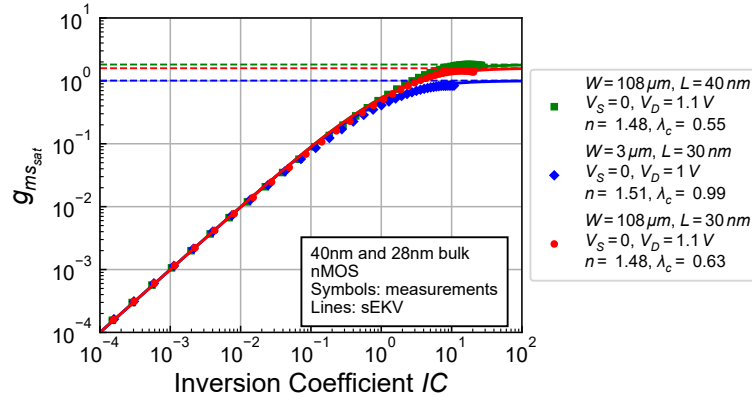


Figure 1.35: Normalized transconductance g_{ms} versus IC measured on minimum length nMOS transistors from a 40 nm and two different 28 nm bulk CMOS processes [21].

where $G_{spec} \triangleq I_{spec}/U_T = 2n\beta U_T$ with $\beta = \mu_0 C_{ox} W/L$. G_{ms} is plotted versus IC in Figure 1.35 and favorably compares to measurements obtained from the derivative of the characteristics shown in Figure 1.34 over a very wide range of bias (more than 4 decades of current). Note that for short-channel devices in strong inversion, the I_D - V_G transfer characteristic becomes a linear function of the gate voltage as illustrated in Figure 1.34 and hence the gate transconductance becomes independent of the drain current or the inversion coefficient and saturates to

$$g_{ms} \cong \frac{1}{\lambda_c} = \frac{L}{L_{sat}}.$$

The source transconductance also becomes independent of the length and only depends on W and v_{sat} according to (1.62)

$$G_{ms} \cong \frac{G_{spec}}{\lambda_c} = nWC_{ox}v_{sat}.$$

Instead of increasing as $\sqrt{IC}$, as it does for long-channel devices, for short-channel transistor with heavy velocity saturation, the normalized source transconductance G_{ms} saturates to $1/\lambda_c$. The inverse of the velocity saturation parameter λ_c is therefore a key parameter since it gives the maximum normalized transconductance that can be achieved for a short-channel device in a given technology.

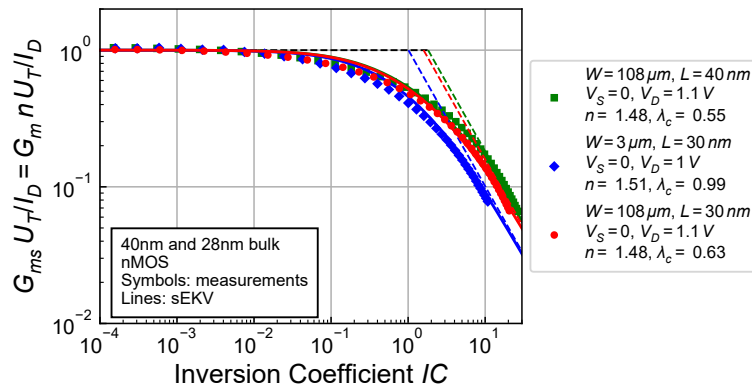


Figure 1.36: Normalized transconductance efficiency $G_m n U_T / I_D$ versus IC measured on minimum length nMOS transistors from a 40 nm and two different 28 nm bulk CMOS processes [21].

The *transconductance efficiency* G_m/I_D , sometimes also called the *current efficiency*, is one of the most important FoM for low-power analog circuit design. It is a measure of how much transconductance is produced for a given bias current and is a function of IC . The transconductance efficiency (or its inverse) appears in many expressions related to the optimization of analog circuits. In normalized

form, the transconductance efficiency is defined as the actual transconductance obtained at a given IC with respect to the maximum transconductance $G_m = I_D/(n U_T)$ reached in weak inversion and is given by (1.74) which is repeated here for convenience

$$\frac{g_{ms}}{IC} = \frac{G_{ms} U_T}{I_D} = \frac{G_m n U_T}{I_D} = \frac{\sqrt{4IC + 1 + (\lambda_c IC)^2} - 1}{IC (2 + \lambda_c^2 IC)} = \begin{cases} 1 & \text{in WI and sat.} \\ \frac{1}{\lambda_c IC} & \text{in SI and sat.} \end{cases} \quad (1.190)$$

The normalized transconductance efficiency given by (1.190) is compared to measurements in Figure 1.36 for the same devices shown in Figure 1.34 and Figure 1.35. Despite the normalized $G_m n U_T / I_D$ only requires one parameter (λ_c or L_{sat}), the model fits very well to the data over more than 5 decades of IC .

1.7.2 FDSOI technologies

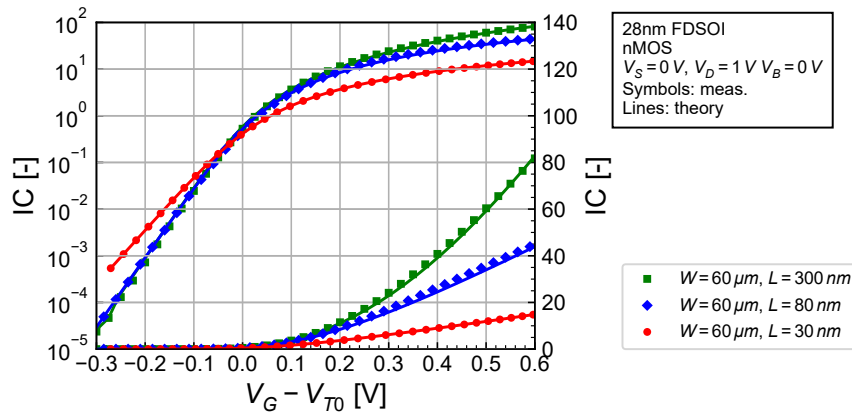


Figure 1.37: IC versus the overdrive voltage $V_G - V_{T0}$ in saturation measured on nMOS transistors with $L = 300 \text{ nm}$, 80 nm and 30 nm from a 28 nm FDSOI process [21].

The sEKV model was originally developed for the bulk MOS transistors. However, it can easily be extended to FDSOI process by making the threshold voltage depend on the back-gate voltage V_B [27]. The sEKV model has been successfully validated on a 28 nm FDSOI process [27]. After a proper parameter extraction, we get the I_D - V_G characteristic shown in Figure 1.37 demonstrating a good fit of the I_D - V_G characteristic in saturation for three different gate length ranging from $L = 300 \text{ nm}$ down to 80 nm and 30 nm .

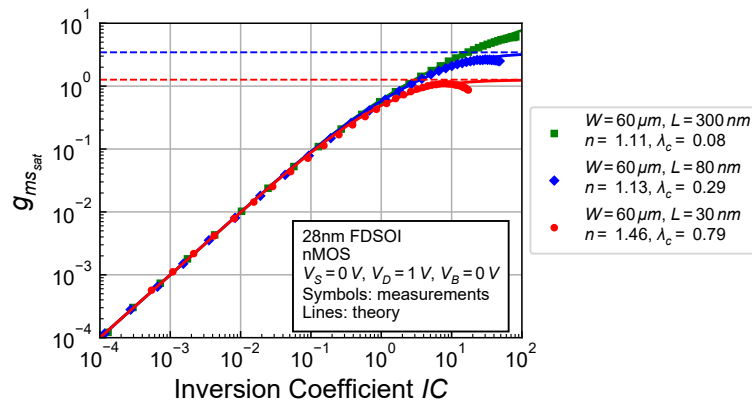


Figure 1.38: Normalized transconductance g_{ms} versus IC measured on nMOS transistors with $L = 300 \text{ nm}$, 80 nm and 30 nm from a 28 nm FDSOI process [21].

The normalized source transconductance g_{ms} is plotted versus IC in Figure 1.38 for the same devices than in Figure 1.37. We see that for the short-channel devices with $L = 80 \text{ nm}$ and $L = 30 \text{ nm}$, g_{ms} saturates to $1/\lambda_c$. On the other hand the g_{ms} of the long-channel transistor with $L = 300 \text{ nm}$ continues to increase as $\sqrt{IC}$. We also see that for the short devices, g_{ms} decreases at high IC . This is the effect of mobility reduction due to the vertical field which has not been accounted for in sEKV.

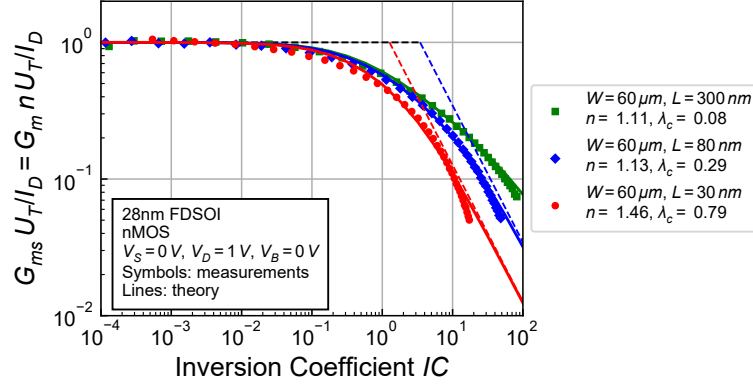


Figure 1.39: Normalized transconductance efficiency $G_m n U_T / I_D$ versus IC measured on nMOS transistors with $L = 300 \text{ nm}$, 80 nm and 30 nm from a 28 nm FDSOI process.

The normalized transconductance efficiency g_{ms}/IC is plotted versus IC for the same devices in Figure 1.39. We clearly see the strong impact of velocity saturation on the short-channel devices making g_{ms}/IC decrease sharply as $1/(\lambda_c IC)$ in strong inversion. We also see some deviations for the short-channel devices at high IC due to mobility reduction due to the vertical field.

1.7.3 FinFET technologies

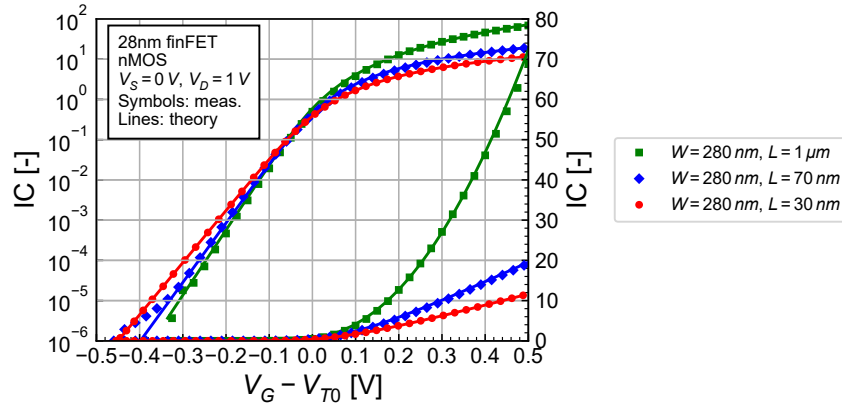


Figure 1.40: IC versus the overdrive voltage $V_G - V_{T0}$ in saturation measured on nMOS transistors with $L = 1 \mu\text{m}$, 70 nm and 30 nm from a 28 nm FinFET process [21].

Even though the FinFET transistor has a very different structure than the classical bulk transistor, the sEKV can also fit the I-V characteristics of FinFET transistors. This is illustrated in Figure 1.40 which shows the I_D - V_G characteristics of nMOS transistors for 3 different lengths ($L = 1 \mu\text{m}$, 70 nm and 30 nm) from a 28 nm FinFET process. We see a very good fit for all transistor lengths over more than 6 decades of current. There is a small deviation at low gate voltage for the 70 nm transistor probably due to additional leakage current which is not accounted for in sEKV.

Figure 1.41 shows the normalized source transconductance versus the inversion coefficient for the same devices. An almost perfect fit is obtained over 5 decades of transconductance.

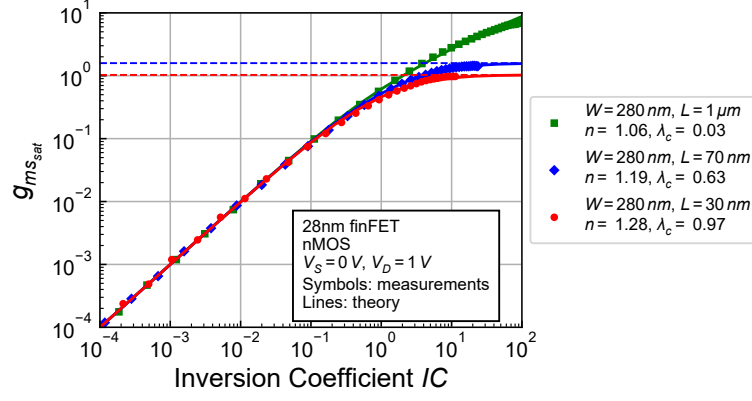


Figure 1.41: Normalized transconductance g_{ms} versus IC measured on nMOS transistors with $L = 1 \mu m$, $70 nm$ and $30 nm$ from a $28 nm$ FinFET process [21].

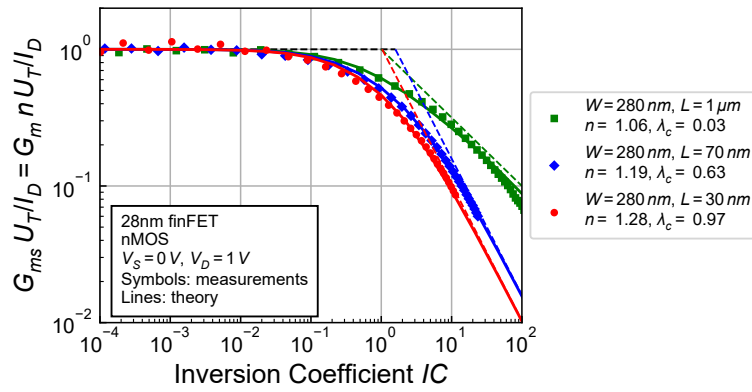


Figure 1.42: Normalized transconductance efficiency $G_m n U_T/I_D$ versus IC measured on nMOS transistors with $L = 1 \mu m$, $70 nm$ and $30 nm$ from a $28 nm$ FinFET process [21].

Finally, Figure 1.42 shows how the normalized G_m/I_D characteristic can be fitted versus the inversion coefficient over more than 5 decades with just a single parameter λ_c .

1.7.4 Comparison on 28 nm bulk, 22 nm FDSOI and 16 nm FinFET

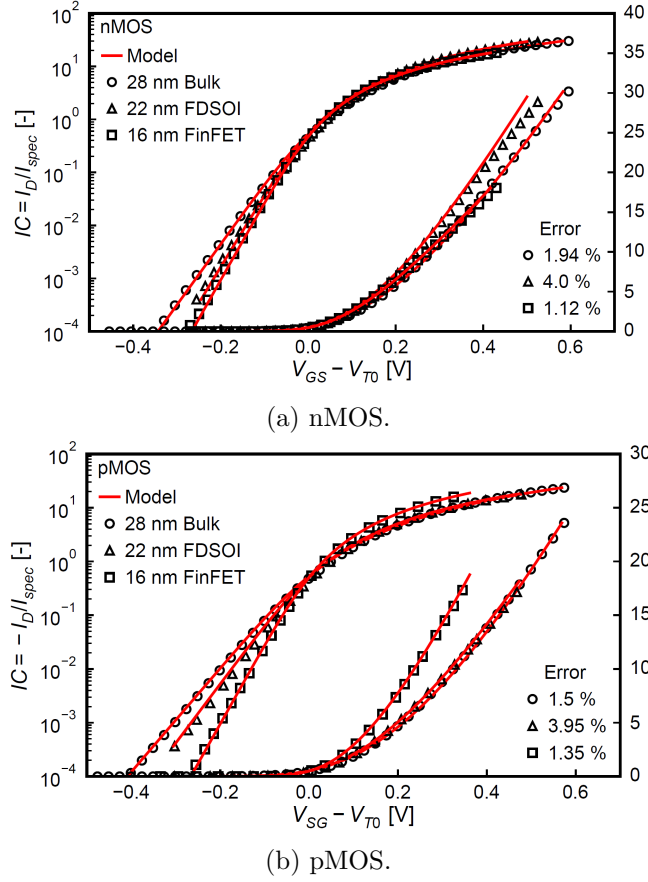


Figure 1.43: $IC-V_G-V_{T0}$ characteristic of transistors from 28 nm bulk (circles), 22 nm FDSOI (triangles) and 16 nm FinFET (squares).

The sEKV model has also been validated on n- and p-channel transistors from a 28 nm bulk, 22 nm FDSOI and 16 nm FinFET technologies. The I_D-V_G characteristic of the three technologies are compared in

Table 1.6: sEKV parameters used for the 28 nm bulk, 22 nm FDSOI and 16 nm FinFET.

Type	Technology	W [nm]	L [nm]	n	$I_{spec\Box}$ [nA]	λ_c	V_{T0} [V]
nMOS	28 nm bulk	3000	30	1.46	599	0.271	0.478
nMOS	22 nm FDSOI	1000	18	1.24	446	0.307	0.266
nMOS	16 nm FinFET	58	16	1.13	1290	0.475	0.366
pMOS	28 nm bulk	3000	30	1.74	446	0.253	0.501
pMOS	22 nm FDSOI	1000	18	1.52	468	0.36	0.312
pMOS	16 nm FinFET	58	16	1.12	995	0.331	0.431

- [1] C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, 1st ed. John Wiley, 2006.
- [2] C. Galup-Montoro and M. C. Schneider, *MOSFET Modeling for Circuit Analysis and Design*. World Scientific, 2007.

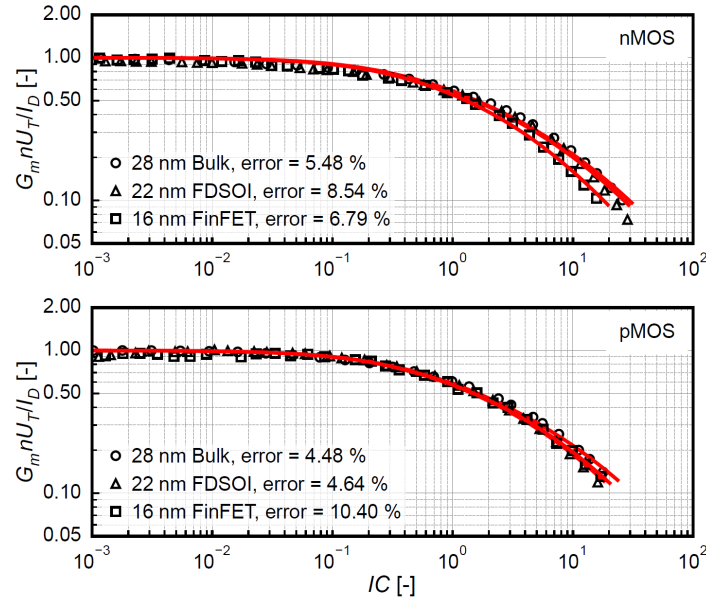


Figure 1.44: Normalized G_m/I_D versus IC of nMOS and pMOS transistors from 28 nm bulk (circles), 22 nm FDSOI (triangles) and 16 nm FinFET (squares).

- [3] M. C. Schneider and C. Galup-Montoro, *CMOS Analog Design Using All-Region MOSFET Modeling*. Cambridge University Press, 2011.
- [4] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. John Wiley, 2006. Available: https://ia601608.us.archive.org/5/items/PhysicsOfSemiconductorDevices_855/PhysicsOfSemiconductorDevices3rdEdition-S.M.SzeAndKwokK.Ng.pdf
- [5] A. Mangla, M. A. Chalkiadaki, F. Fadhuile, T. Taris, Y. Deval, and C. C. Enz, “Design Methodology for Ultra Low-power Analog Circuits Using Next Generation BSIM6 MOSFET Compact Model,” *Microelectronics Journal*, vol. 44, no. 7, pp. 570–575, 2013.
- [6] A. Mangla, “Modeling nanoscale quasi-ballistic MOS transistors,” PhD thesis, EPFL, 2014.
- [7] C. Enz and A. Pezzotta, “Nanoscale MOSFET Modeling for the Design of Low-power Analog and RF Circuits,” in *Proc. Of the int. Conf. On mixed design of integrated circuits and systems (MIXDES)*, 2016.
- [8] C. Enz, F. Chicco, and A. Pezzotta, “Nanoscale MOSFET Modeling: Part 2: Using the Inversion Coefficient as the Primary Design Parameter,” *IEEE Solid-State Circuits Magazine*, vol. 9, no. 4, pp. 73–81, 2017.
- [9] Y. A. El-Mansy and A. R. Boothroyd, “A simple two-dimensional model for IGFET operation in the saturation region,” *IEEE Transactions on Electron Devices*, vol. 24, no. 3, pp. 254–262, 1977.
- [10] P. K. Ko, R. S. Muller, and C. Hu, “A unified model for hot-electron currents in MOSFET’s,” in *International electron devices meeting (IEDM)*, 1981, pp. 600–603.
- [11] P. K. Ko, “Approaches to Scaling,” in *Advanced MOS device physics*, vol. 18, N. G. Einspruch and G. S. Gildenblat, Eds., Academic Press, 1989.
- [12] N. Arora, *MOSFET Models for VLSI Simulation - Theory and Practice*, 1st ed. Springer, 1993.
- [13] N. Arora, *MOSFET Modeling for VLSI Simulation - Theory and Practice*, 2nd ed. World Scientific, 2007.
- [14] M. Bucher, C. Lallement, C. Enz, F. Théodoloz, and F. Krummenacher, “The EPFL-EKV MOSFET Model Equations for Simulation.” https://github.com/chrisenz/EKV/blob/main/EKV2.6/docs/ekv_v26_rev2.pdf, 1998.
- [15] Z. H. Liu *et al.*, “Threshold voltage model for deep-submicrometer MOSFETs,” *IEEE Transactions on Electron Devices*, vol. 40, no. 1, pp. 86–95, 1993.
- [16] D. A. Pino-Monroy *et al.*, “Design-Oriented All-Regime All-Region 7-Parameter Short-Channel MOSFET Model Based on Inversion Charge,” *IEEE Access*, vol. 10, pp. 86270–86285, 2022.

- [17] D. A. Pino-Monroy *et al.*, “Design-oriented model for short-channel MOS transistors based on inversion charge,” in *IEEE 14th latin america symposium on circuits and systems (LASCAS)*, 2023, pp. 1–4.
- [18] C. Enz, H.-C. Han, C. Delignac, and T. Taris, “A comprehensive output conductance model valid in all regions of inversion,” in *2024 IEEE international symposium on circuits and systems (ISCAS)*, 2024, pp. 1–5. doi: [10.1109/ISCAS58744.2024.10558500](https://doi.org/10.1109/ISCAS58744.2024.10558500).
- [19] C. C. Enz and G. C. Temes, “Circuit Techniques for Reducing the Effects of Op-amp Imperfections: Autozeroing, Correlated Double Sampling, and Chopper Stabilization,” *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584–1614, 1996.
- [20] A. Boukhayma, A. Caizzzone, R. Capoccia, and C. Enz, “Design and Optimization of Low Power and Low Light Sensor: (Invited),” in *2020 IEEE custom integrated circuits conference (CICC)*, 2020, pp. 1–8.
- [21] C. Enz, F. Chicco, and A. Pezzotta, “Nanoscale MOSFET Modeling: Part 1: The Simplified EKV Model for the Design of Low-Power Analog Circuits,” *IEEE Solid-State Circuits Magazine*, vol. 9, no. 3, pp. 26–35, 2017.
- [22] A. S. Roy and C. C. Enz, “Compact modeling of thermal noise in the MOS transistor,” *IEEE Transactions on Electron Devices*, vol. 52, no. 4, pp. 611–614, 2005.
- [23] A. S. Roy, C. C. Enz, and J. Sallese, “Noise Modeling in Lateral Nonuniform MOSFET,” *IEEE Transactions on Electron Devices*, vol. 54, no. 8, pp. 1994–2001, 2007.
- [24] A. S. Roy, C. C. Enz, and J. Sallese, “Noise modeling methodologies in the presence of mobility degradation and their equivalence,” *IEEE Transactions on Electron Devices*, vol. 53, no. 2, pp. 348–355, 2006.
- [25] A. S. Roy, C. C. Enz, and J. M. Sallese, “Charge, current, and noise partitioning in MOSFET in the presence of mobility degradation,” *IEEE Electron Device Letters*, vol. 27, no. 8, pp. 674–677, 2006.
- [26] C. Enz, “Simple Expression of the Thermal Noise Excess Factor for LNA Design,” in *2022 29th IEEE international conference on electronics, circuits and systems (ICECS)*, 2022, pp. 1–4.
- [27] A. Pezzotta, F. Jazaeri, H. Bohuslavskyi, L. Hutin, and C. Enz, “A Design-oriented Charge-based Simplified Model for FDSOI MOSFETs,” in *2018 joint international EUROSIOI workshop and international conference on ultimate integration on silicon (EUROSIOI-ULIS)*, 2018, pp. 1–4.