Chapter 3

A MOS Transistor Model for Mixed Analog-digital Circuit Design and Simulation

Matthias Bucher¹, Christophe Lallement², François Krummenacher¹, Christian Enz³
¹Swiss Federal Institute of Technology, Lausanne (EPFL), Electronics Laboratory, ELB-Ecublens, CH-1015 Lausanne, Switzerland, Phone: +41 21 693 3975, Fax: +41 21 693 36 40, Email: matthias.bucher@epfl.ch
²ERM-PHASE/ENSPS, Parc d’Innovation, Bld. Sébastien Brant, F-67400 Illkirch, France, Phone: +33 388 63 95 07 Fax: +33 388 63 95 09, Email: lallem@erm1.u-strasbg.fr
³Swiss Center for Electronics and Microtechnology (CSEM), Jaquet Droz 1, CH-2007 Neuchâtel, Switzerland, Phone: +41 32 720 52 18, Fax: +41 32 720 57 42, Email: christian.enz@csem.ch

Abstract -- In the design cycle of complex integrated circuits, the compact device simulation models are the privileged vehicle of information between the foundry and the designer. Effective circuit design, particularly in the context of analog and mixed analog-digital circuits using silicon CMOS technology, requires a MOS transistor (MOST) circuit simulation model well adapted both to the technology and to the designer’s needs. The MOST model itself should also help portable design, since design-reuse becomes a major advantage in the fast development of new products. Clearly, the MOST model must be based on sound physical concepts, and be parameterized in such a way that it allows easy adaptation to very different CMOS technologies, and provides the designer with information on important parameters for design. This chapter describes an analytical, scalable compact MOST model, called ‘EKV’ MOST model, which is built on fundamental physical properties of the MOS transistor. Among the original concepts used in this model are the normalization of the channel current, and taking the substrate as a reference instead of the source. The basic long-channel model is formulated in
symmetric terms of the source-to-bulk and drain-to-bulk voltages. In particular, the transconductance-to-current ratio is accurately described for all levels of current from weak inversion through moderate and to strong inversion. This characteristic is almost invariant with respect to process parameters and technology scaling; therefore, the model can be adjusted to a large range of different technologies. Short-channel effects have been included in the model for the simulation of deep submicron technologies. A full charge-based dynamic model as well as the thermal noise model are derived within the same approach. The continuity of the model characteristics is based on the use of a single equation, enhancing circuit convergence. The relative simplicity of the model and its low number of parameters also ease the process of parameter extraction, for which an original method is proposed. This MOST model is used in the context of low-voltage, low-current analog and analog-digital circuit design using deep submicron technologies. A version of this model based on the same fundamental concepts, is also available as a public-domain model in various commercially available simulators.

1. INTRODUCTION

The continuing decrease of the supply voltage to reduce the power consumption of digital circuits strongly affects the design of the analog part of a mixed analog/digital IC. As a consequence of the supply voltage reduction, the operating points of many MOS transistors forming the analog circuits move towards the region of moderate inversion. Low-voltage design of CMOS circuits, under supply voltages as low as 1V or below, typically requires operation in moderate inversion. During the design process, the operating points are very often set in terms of available drain current and targeted transconductances. Once the current and transconductances are chosen, the transconductance-to-current ratio is defined and the corresponding operating point can be fixed in terms of aspect ratio (or inversion coefficient [1]). Since the circuit performances directly depend on the devices’ transconductances, a good control and prediction of the transconductances and the corresponding operating points is crucial, even if they fall in the moderate inversion region. This can easily be done by using the transconductance-to-current ratio modeling approach described in this chapter.

Most of the MOS transistor (MOST) models [2] currently available in the public domain have been developed starting from the large-signal strong inversion operation and then extended to weak inversion by different means which not always ensure the continuity and/or the accuracy of the characteristic in the moderate inversion. The transconductances may then be
wrongly estimated which can result in serious design problems. The
\( g_{\text{on}} I_D \) approach, which is the basis of the EKV MOST model, starts from
the small-signal model, from which the large-signal static, dynamic as well
as thermal noise models are derived by integration. This intrinsically ensures
the continuity (including higher-order derivatives) of all the model
characteristics.

The basic model features used in this approach and contrasting from
those used in many other MOST models, can be summarized as follows:
- **bulk reference** for all voltages instead of source reference
- the *pinch-off voltage* \( V_p \) and *slope factor* \( n \), are the principal model
  internal variables, and are both function mainly of the gate voltage
- **use of normalized current** as another model internal variable of central
  importance
- non-regional approach, using single equations for all operating regimes
- correct behavior in the asymptotic regions of *weak* and *strong inversion*
  operation
- correct prediction of all transconductances in *moderate inversion*
- symmetrical forward and reverse operation
- low number of parameters
- *hierarchical model structure* allowing to formulate simple hand
  calculation expressions.

The use of current normalization has numerous advantages. It is
particularly useful for ratio-based design techniques as described in [3]. The
normalized drain current is also the principal model variable used in this
modeling approach.

The present contribution shows this approach to be valid over many
generations of MOS technologies. New features extending the model to
application with deep submicron technologies are presented. The normalized
transconductance to current ratio characteristic, which represents the
foundation of the model, is presented in Section 2, and is used to derive the
ideal long-channel static model. The latter allows simple hand-calculation
expressions to be formulated [1], which are useful in general design practice
[3] as well as in educational contexts [4]. Vertical field dependent mobility
and non-uniform channel doping effects are included in the long-channel
model. In Section 3, effects related to short and narrow device geometries
are introduced. In particular, simple models for the reverse short-channel
effect (RSCE) and the bias dependent series resistances are presented. In
Section 4 a dynamic model for the node charges as well as a thermal noise
model are developed, using the normalized current as main variable. Some
aspects of the model for computer simulation are discussed in Section 5. A
complete parameter extraction method from DC measurements is presented, demonstrating the scalability of the model for submicron CMOS technologies.

2. THE LONG-CHANNEL MODEL

2.1 Transconductance-to-current ratio

Unlike many other MOST modeling approaches, the voltages are all referred to the local substrate instead of the source, to preserve the structural symmetry of the MOS device also in the model. Two transconductances can be defined, namely the gate transconductance,

\[ g_{mg} \cdot \frac{fI_D}{fV_G} \]  

and the transconductance from the source,

\[ g_{ms} \cdot \frac{fI_D}{fV_S} \]  

The following relationship holds between the transconductances in saturation [1],

\[ g_{ms} = n \cdot g_{mg} \]  

The slope factor \( n \) is defined as [1]

\[ n \equiv \left[ \frac{\partial V_p}{\partial V_G} \right]^{-1} = 1 + \frac{\gamma}{2 \cdot 2 \psi_0 + V_p} \]  

where the parameter \( \psi_0 \cdot 2 \phi_F + V_{ch} \) is the approximation of the surface potential in strong inversion; \( \phi_F \) is the Fermi potential of the majority carriers, \( U_T = k \cdot T / q \) is the thermodynamic potential with \( k \) being the Boltzmann constant and \( q \) the unit charge. The parameter \( \gamma \) is the body or substrate effect factor

\[ \gamma = \frac{\sqrt{2 \cdot q \cdot \epsilon_{si} \cdot N_{sub}}}{C_{ox}} \]  

where \( N_{sub} \) is the substrate doping concentration in the channel; the gate oxide capacitance per unit area \( C_{ox} \) depends on the oxide thickness \( t_{ox} \) as
The pinch-off voltage $V_p$ represents the channel voltage at a given gate voltage $V_G$, for which the inversion charge density $Q_{inv}$ of the mobile charge forming the channel becomes negligible with respect to the depletion charge density $Q_B$ [1]

$$V_p = V_G - V_{TO} - \gamma \left[ \sqrt{V_G - V_{TO}} + \left( \sqrt{\psi_0 + \frac{\gamma}{2}} \right)^2 - \left( \sqrt{\psi_0 + \frac{\gamma}{2}} \right) \right]. \quad (6)$$

The pinch-off voltage accounts for threshold voltage and substrate effects through the use of the parameters $V_{TO}$ and $\gamma$, respectively. The parameters $V_{TO}$ and $\psi_0$ are temperature dependent. An approximate expression for $V_p$ which is useful for hand calculation [1],

$$V_p \approx \frac{V_G - V_{TO}}{n(V_G)} \quad (7)$$

displays that the substrate effect is accounted for through a function of the gate voltage $V_G$, unlike conventional MOST models where it is commonly expressed as a function of $V_S$. The pinch-off voltage (6) and the slope factor (4) are shown in Fig.1 versus the gate voltage $V_G$; measured characteristics are from a long n-channel device of a 0.25\(\mu\)m CMOS technology.
Chapter 3

\[ g_{mg} = \frac{I_D}{n \cdot U_T} \quad g_{ms} = \frac{I_D}{U_T} \tag{8} \]

and in strong inversion

\[ g_{mg} = \sqrt{\frac{2 \cdot \beta \cdot I_D}{n}} \quad g_{ms} = \sqrt{2 \cdot n \cdot \beta \cdot I_D} \tag{9} \]

The normalized transconductance to current ratio in saturation can then be expressed through a single function valid in the entire region of inversion, the asymptotes of which are given by (8) and (9),

\[ \frac{g_{mg} \cdot n \cdot U_T}{I_D} = \frac{g_{ms} \cdot U_T}{I_D} = G(i_f) = \begin{cases} 1 \quad &\text{(weak inversion)} \\ \frac{1}{\sqrt{i_f}} \quad &\text{(strong inversion)} \end{cases} \tag{10} \]

where \( i_f \) is the forward normalized drain current [1]. The transconductance to current ratio is plotted in Fig.2 versus the normalized current \( i_f = I_F / I_S \) in saturation where \( I_F \equiv I_D \). The normalization factor \( I_S \) is called specific current. It is defined as the drain current corresponding to the intersection of the two asymptotes given by (10). \( I_S \) is related to the transistor aspect ratio [1]

\[ I_S = 2 \cdot n \cdot \beta \cdot U_T^2 \tag{11} \]

where \( \beta = \mu_e \cdot C_{ox} \cdot W_{eff} / L_{eff} \) is the gain factor depending on the carrier mobility \( \mu_e \). The specific current is a quantity allowing to delimit the regions of weak and strong inversion: \( I_D << I_S \) (or \( i_f << 1 \)) corresponds to weak inversion and \( I_D >> I_S \) (or \( i_f >> 1 \)) corresponds to strong inversion.

The normalized \( g_{ms} / I_D \) characteristic can be computed by numerically solving the Poisson and Gauss equations for the surface potential under long-channel and uniform doping assumptions. The result is plotted in Fig.2 versus the normalized current \( i_f \) for \( \gamma = 0.7 \sqrt{V} \). The sensitivity of the \( g_{ms} / I_D \) characteristic to \( \gamma \) is found to be small. A suitable function is required to describe this characteristic; in the past, various analytical expressions of varying degree of accuracy have been used [5][1]. A physics based expression has been proposed in [6].
\[ G(i_f) = \frac{1}{\sqrt{\frac{1}{4} + i_f + \frac{1}{2}}} \]  \hfill (12)

derived from the assumption of a linear relationship between the surface potential \( \psi_s \) and inversion charge density \( Q_{\text{inv}} \) [7]. This expression has the advantage of analytical simplicity.

As shown in Fig.2, (12) is in excellent agreement with the results obtained from the numerical computation. Therefore, the present model is in excellent agreement with the theory.

Many trade-offs must be made in realistic technology scaling. As channel lengths decrease, the vertical dimensions, oxide thickness and junction depths, are decreased, while substrate doping is increased to overcome the detrimental impact of short-channel effects. At the same time, higher channel doping concentrations also lead to reduced low-field mobility of the carriers. Low-voltage operation requires as low threshold voltages as

\[ \frac{g_{\text{ms}}} {U_T / I_D} \]

Figure 2. The normalized transconductance to current ratio \( g_{\text{ms}} U_T / I_D \) versus normalized drain current \( I_D / I_S \) in saturation, for a long-channel device, computed by numerically solving the Poisson equation (markers) and analytical expression (12) (line). The intersection of the weak and strong inversion asymptotes defines the specific current \( I_S \).
possible, while high weak inversion slope as well as high threshold voltage are required to achieve low leakage currents. Minimum practical values for threshold voltage are in the range from 0.3V to 0.4V, while supply voltages are reduced, to values of 1.5V and even 0.9V for future technologies, therefore reducing the available range of strong inversion operation. In any case, technology trends make operation in weak and moderate inversion become more and more important, and MOST models must give a precise description of these regions of operation. The present modeling approach based on transconductance to current ratio is suitable to address this situation.

Practical values for the important technological parameter $\gamma$ lie in the range of $0.3\sqrt{V}$ to $0.9\sqrt{V}$. Noting that the substrate effect depends on oxide thickness and channel doping as $\gamma \sim t_{ox} \cdot \sqrt{N_{sub}}$, its dependence on technology scaling can be assessed, depending on the scaling rules adopted. Assuming constant field scaling (e.g. [8]) with $t_{ox} \sim \kappa^{-1}$ and $N_{sub} \sim \kappa$ when the scaling factor $\kappa$ is increased, $\gamma$ would decrease as $\gamma \sim \kappa^{-1/2}$ [8]. In realistic scaling, observed over several technology generations, assuming $\gamma \sim \kappa^{-0.77}$ and $N_{sub} \sim \kappa^{1.6}$ [9], the variation of $\gamma$ is even smaller and it can
be considered almost as constant. As long as the same physical effects are dominating, the normalized $g_{ms}/I_D$ characteristic can be considered as independent of technology and scaling. In Fig.3, measurements of the normalized $g_{ms}/I_D$ characteristic from three different CMOS technology generations and from three different foundries are shown. The measurements are made on long-channel devices, with minimum feature sizes ranging from 1 µm to 0.5 µm. The normalized $g_{ms}/I_D$ characteristic has been measured on many other technologies, confirming the excellent qualitative agreement with the analytical expression (12).

In the next subsections, the normalized $g_{ms}/I_D$ characteristic is used to elaborate a complete model that is valid in all regions of operation. The large-signal static model, the charge-based dynamic model as well as the thermal noise model can all be derived from the same function (12), as will be shown in the following. Note that these following derivations have in their essence already been used to establish the former ‘EKV’ model formulations. In contrast to these, the present formulation is based on the expression (12), which has an improved physical basis and is analytically more tractable.

2.2 The static model for the drain current

The drain current $I_D$ can be expressed as the difference between a forward current $I_F$ and a reverse current $I_R$, depending only on $V_P - V_S$ and $V_P - V_D$, respectively [1]:

$$I_D = I_F - I_R = I_S \cdot (i - i_r) \quad (13)$$

In saturation (i.e. for $V_D >> V_P$), the reverse current becomes negligible compared to the forward current, and the drain current simply reduces to the forward current.

The source transconductance in saturation can be expressed as

$$g_{ms} \equiv -\frac{\partial I_D}{\partial V_S}_{V_G, V_D} = \frac{\partial I_F}{\partial (V_P - V_S)} = \frac{I_S}{U_T} \cdot \frac{\partial i}{\partial V} \quad (14)$$

where $v \cdot (V_P - V_{S(D)})/U_T$ is the normalized voltage for the forward (respectively reverse) current. From (12), the source transconductance in saturation is also given by...
Equating (14) and (15) and using (12) results in

\[
\frac{\partial i_f}{\partial v} = i_f \cdot G(i_f) = \frac{1}{4} + i_f - \frac{1}{2}
\]

which can be integrated to find a large-signal interpolation function for the normalized voltage as a function of the normalized current

\[
v = 2 \cdot \left( \sqrt{\frac{1}{4} + i - \frac{1}{2}} \right) + \ln \left( \sqrt{\frac{1}{4} + i - \frac{1}{2}} \right)
\]

(17)

Eq. 17 needs to be inverted for the purpose of using the model for computer simulation, to obtain the normalized currents in terms of the normalized voltages. Unfortunately, (17) cannot be inverted analytically. A simple Newton-Raphson algorithm is used to obtain (18) [10][11]. It requires three iterations only for a maximum residual error below $10^{-6}$ when choosing adequate initial conditions. Little computational penalty is therefore incurred in using this simple numerical scheme. The inverted function will have the following asymptotes:

\[
i = F(v) = \begin{cases} \left( \frac{v}{2} \right)^2 & (v >> 0) \\ e^v & (v << 0) \end{cases}
\]

The normalized currents are now expressed as a function of the normalized voltages, for the whole current range from weak through moderate to strong inversion, as well as from conduction to saturation, as is required for the formulation of the compact MOST model.

### 2.3 Hand calculation model and circuit design

For analog circuit design, simplified model expressions can be obtained for the basic long-channel model. The simplified expressions for drain current and transconductances are obtained for the asymptotic regions of device operation, weak and strong inversion, conduction and saturation, and are summarized in Table I to Table III [1]. Apart from use of the hand calculation expressions in design, they often are also useful for the development of the parameter extraction method.

The relationship between the pinch-off voltage and the weak inversion slope should be noted, since it is important for model development and
design practice. The long-channel expression for the drain current in weak inversion, using the approximate expression (7) for the pinch-off voltage and assuming \( i_f \gg i_r \) (requiring \( V_D - V_S > \sim 5U_T \), in weak inversion), is given by

\[
I_D \cong I_S \cdot \exp \left( \frac{V_p - V_S}{U_T} \right) \cong I_S \cdot \exp \left( \frac{V_G - V_{TO} - n \cdot V_S}{n \cdot U_T} \right)
\]  

The substrate effect thus determines the weak inversion slope through the substrate effect factor (or weak inversion slope factor) \( n \). The subthreshold swing \( S_G \), describing the change in gate voltage needed to change \( I_D \) by one decade, is also often used in this context, and is related to \( n \) simply as:

\[
S_G \equiv \left[ \frac{\partial (\log I_D)}{\partial V_G} \right]^{-1} = \ln(10) \cdot n \cdot U_T \cong 2.3 \cdot n \cdot U_T
\]  

An ideal slope factor of \( n = 1 \) therefore corresponds to the ideal subthreshold swing of 60mV/dec. Similarly, the subthreshold swing for a modulation from the source can be defined as

\[
S_S \equiv \left[ \frac{\partial (\log I_D)}{\partial V_S} \right]^{-1} = \ln(10) \cdot U_T \cong 2.3 \cdot U_T
\]  

Note that no parameter specific to weak inversion has been introduced in the model so far. In general, the weak inversion slope is correctly predicted for long-channel transistors for modulation from both the gate and the source. This holds for technologies for which surface states are negligible, which is the case in most modern CMOS technologies.

Numerous applications of the model principles and its use in analog and mixed analog-digital circuit design can be found in [3]. In particular, a ratio-based design technique is described, which allows design of circuits that are insensitive to process variations and temperature to a first order. Using the specific current of a reference transistor, generated from a dedicated circuit, allows to set the inversion coefficient (or normalized current) of a device by simple scaling of this reference current by series and/or parallel combination of reference transistors. This ratio-based design technique is therefore attractive for designing circuits that can be ported from one technology to another, avoiding major redesign while preserving the main performance. The EKV model can therefore be considered as a tool enabling design-reuse and portability. Further aspects related to model formulation can also be found in [12].
The approach described so far offers the advantage of having a model with continuous drain current from weak to strong inversion and continuous $n^\text{th}$-order derivatives with respect to any terminal voltage. This is essential for the design and simulation of analog ICs and particularly for analog circuits that have to operate at a low supply voltage imposing that the operating points of many transistors be in the middle of the moderate inversion region. The continuity and accuracy of the derivatives is also important for the correct computation of distortion and intermodulation products which are fundamental limitations of RF circuits.

No mobility reduction effects, nor short-channel effects, have been taken into account so far. The long-channel model needs to be complemented to account for vertical field dependent mobility, for which an adequate expression will be developed in the next subsection. Another important aspect of precision MOS modeling is accounting for the effects of the non-uniformity of channel doping profiles with the depth from the Si-SiO$_2$ interface, resulting from ion implantation to reduce short-channel effects and to prevent punch-through. This effect is accounted for in the $V_p(V_G)$ relation as described in [13], requiring two additional parameters.

### Table 1: Drain current in strong inversion.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Expression for drain current</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduction</td>
<td>$n \cdot \beta \cdot \left[ V_p - \frac{V_s + V_d}{2} \right] \cdot (V_d - V_s)$</td>
<td>$V_s \leq V_p$, $V_d \leq V_p$</td>
</tr>
<tr>
<td></td>
<td>$\cong \beta \cdot \left[ V_G - V_{TO} - \frac{n}{2} \cdot (V_s + V_d) \right] \cdot (V_d - V_s)$</td>
<td></td>
</tr>
<tr>
<td>Forward saturation</td>
<td>$\frac{n \cdot \beta}{2} \cdot (V_p - V_s)^2 \cong \frac{\beta}{2 \cdot n} \cdot (V_G - V_{TO} - n \cdot V_s)^2$</td>
<td>$V_s \leq V_p$, $V_d &gt; V_p$</td>
</tr>
<tr>
<td>Reverse saturation</td>
<td>$\frac{-n \cdot \beta}{2} \cdot (V_p - V_d)^2 \cong \frac{-\beta}{2 \cdot n} \cdot (V_G - V_{TO} - n \cdot V_d)^2$</td>
<td>$V_s &gt; V_p$, $V_d \leq V_p$</td>
</tr>
<tr>
<td>Blocked</td>
<td>$I_F = I_r \Rightarrow I_D = 0$</td>
<td>$V_d = V_s$</td>
</tr>
</tbody>
</table>

### 2.4 Vertical field dependent mobility

While oxide thickness is reduced when scaling CMOS technology, the power supply voltage is usually not reduced quite accordingly—as would be required by the constant field scaling—to maintain maximum speed for digital circuit applications. Thus higher fields result across the gate oxides in more advanced technologies [14]. At high vertical fields, surface roughness
scattering is considered as the main mechanism limiting mobility, while other scattering mechanisms dominate at lower fields. Mobility dependence on effective field has exposed a ‘universal’ behavior—regardless of doping levels—at high and intermediate vertical field strengths, when the effective mobility $\mu_{eff}$ is plotted with respect to the effective field $E_{eff}$. The latter is expressed as a function of the depletion charge density $Q_B$ and the inversion charge density $Q_{inv}$ [15]:

$$E_{eff} = \frac{Q_B' + \eta \cdot Q_{inv}'}{\varepsilon_s}$$

(22)

<table>
<thead>
<tr>
<th>Mode</th>
<th>Expression for $I_D$</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduction</td>
<td>$I_S \cdot e^{V_S - V_T} \cdot \left[ e^{\frac{V_D}{V_T}} - e^{\frac{V_B}{V_T}} \right]$</td>
<td>$V_S &gt; V_P$</td>
</tr>
<tr>
<td></td>
<td>$= I_S \cdot e^{V_S - V_T} \cdot \left[ 1 - e^{\frac{V_D}{V_T} - \frac{V_B}{V_T}} \right]$</td>
<td>$V_D &gt; V_P$</td>
</tr>
<tr>
<td></td>
<td>$\equiv I_{D0} \cdot e^{V_G - V_T} \cdot \left[ e^{\frac{V_D}{V_T}} - 1 \right]$</td>
<td>$I_{D0} = I_S \cdot e^{-\frac{V_D}{nU_T}}$</td>
</tr>
<tr>
<td></td>
<td>$\equiv I_{D0} \cdot e^{V_G - V_T} \cdot \left[ 1 - e^{-\frac{V_D}{V_T}} \right]$</td>
<td></td>
</tr>
</tbody>
</table>

| Forward saturation | $I_F = I_S \cdot e^{\frac{V_S - V_T}{U_T}} \equiv I_{D0} \cdot e^{\frac{V_G - V_T}{nU_T}}$ | $V_S > V_P$                 |
|                   | $V_D > V_P$ | $V_D - V_S >> U_T$ |

| Reverse saturation | $-I_R = I_S \cdot e^{\frac{V_S - V_T}{U_T}} \equiv I_{D0} \cdot e^{\frac{V_G - V_T}{nU_T}}$ | $V_S > V_P$                 |
|                   | $V_D > V_P$ | $V_S - V_D >> U_T$ |

| Blocked           | $I_F = I_R \Rightarrow I_D = 0$ | $V_D = V_S$ or $V_S >> V_D, V_D >> V_P$ |

Depending on the temperature range, the field strength and the type of carriers, mobility degradation can be expressed using different exponents of
$E_{\text{eff}}$. At high field strengths, a $\mu_{\text{eff}} \sim E_{\text{eff}}$ dependence is observed for holes (p-channel), while for electrons (n-channel) the dominant dependence is $\mu_{\text{eff}} \sim V_{\text{eff}}$.

Table 3: Transconductances in strong and weak inversion.

<table>
<thead>
<tr>
<th></th>
<th>Strong inversion</th>
<th>Weak inversion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conduction</td>
<td>Forward saturation</td>
</tr>
<tr>
<td>$g_{mn}$</td>
<td>$\beta \cdot (V_D - V_S)$</td>
<td>$\beta \cdot (V_P - V_S) = \sqrt{\frac{2 \cdot \beta \cdot I_D}{n}}$</td>
</tr>
<tr>
<td></td>
<td>$= \frac{2 \cdot I_D}{n \cdot (V_P - V_S)}$</td>
<td>$\equiv \frac{2 \cdot I_D}{V_G - V_{TO} - n \cdot V_S}$</td>
</tr>
<tr>
<td>$g_{ms}$</td>
<td>$n \cdot \beta \cdot (V_P - V_S)$</td>
<td>$\sqrt{2 \cdot n \cdot \beta \cdot I_F}$</td>
</tr>
<tr>
<td></td>
<td>$\frac{2 \cdot I_F}{V_P - V_S}$</td>
<td>$\equiv \frac{2 \cdot n \cdot I_F}{V_G - V_{TO} - n \cdot V_S}$</td>
</tr>
<tr>
<td>$g_{md}$</td>
<td>$n \cdot \beta \cdot (V_P - V_D)$</td>
<td>$\sqrt{2 \cdot n \cdot \beta \cdot I_R}$</td>
</tr>
<tr>
<td></td>
<td>$= \frac{2 \cdot I_R}{V_P - V_D}$</td>
<td>$\equiv \frac{2 \cdot n \cdot I_R}{V_G - V_{TO} - n \cdot V_D}$</td>
</tr>
</tbody>
</table>

Accounting for these dependencies is particularly important for correct prediction of intermodulation. Also, a different factor for the dependency on inversion charge is observed, $\eta \equiv 0.5$ for n-channel and $\eta \equiv 0.3$ for p-channel [16] at room temperature. While most other MOST models use approximate and simplified expressions accounting for the global mobility dependence, the approach taken here is to consider the local effective field depending on the position, which is then integrated along the channel. The expression for the localized mobility dependence on vertical field,
accounting for the two terms of mobility dependence on effective field as discussed above, is formulated as [11]

\[ \mu_{\text{eff}}(x) = \mu_0 \left( \frac{E_{\text{eff}}(x)}{E_1} + \frac{E_{\text{eff}}(x)}{E_2} \right)^{-1} \]  \hspace{1cm} (23)

where \( E_1 \) and \( E_2 \) are the model parameters related to the first- and second-order effective fields, respectively. The above expression (23) is comparable to the model e.g. as discussed in [17], which however does not consider local field dependent mobility. To account for mobility reduction globally throughout the whole channel, (23) is integrated along the channel [11], using the expressions for the charge densities in terms of the normalized current instead of position, \( Q_{\text{inv}}(i) \) and \( Q_{\text{B}}(i) \). The integration can be carried out, following the procedure outlined in Section 4 for the charges integration. The resulting mobility expression, in terms of the normalized currents \( i_f \) and \( i_r \) (which will be omitted here for brevity), will also depend on the longitudinal field, i.e. \( V_{DS} \): mobility reduction due to vertical field is attenuated in saturation, since a part of the channel near the drain is less strongly inverted than at the source. Unlike other MOST models, the 2nd-order term of mobility reduction used here represents the fully integrated form of the local effective field dependence. This mobility model has shown very accurate results for technologies with thin gate oxides in strong inversion, both for n-channel transistors where the 2nd-order term is privileged, and for p-channel transistors adequately modeled by using the 1st-order term. A slight temperature dependence of the related parameters can be observed. This dependence can be neglected in a first-order approximation for usual temperature ranges from 0°C to 100°C, using only the temperature dependence of the low-field mobility.

Another point should be noted for this formulation of the mobility model. No specific parameters are introduced for the dependence on substrate bias, as it is often necessary in other models. The choice of a fixed parameter \( \eta \) has proven to simplify parameter extraction. Good results are obtained for the temperature range considered above--for both n- and p-channel--for different substrate biases, providing the substrate effect related parameters have been correctly extracted. Also, the asymptotic behavior of this model at very high fields is correct. Unphysical predictions, such as increasing drain current with increased substrate bias, as can sometimes be observed in other models, are avoided.
3. THE STATIC MODEL FOR SHORT AND NARROW GEOMETRIES

The previous derivation has been done assuming a long-channel device. In short-channel devices, the two-dimensional nature of the fields would actually also require a two-dimensional analysis. For the purposes of a circuit simulation model which should remain reasonably simple, approximations can be used, yielding a model with sufficient accuracy and efficiency.

For short-channel devices in strong inversion, the effects of velocity saturation and channel length modulation (CLM) are introduced. Another important effect affecting short-channel devices is the resistivity of the source and drain diffusion regions, which is typically bias dependent in technologies using lightly doped drain (LDD) structures.

In weak inversion, the short- and narrow-channel effects are introduced by modifying the expression of the pinch-off voltage $V_p$ using a charge-sharing approach. The pinch-off voltage $V_p$ for a long-channel device only depends on the gate voltage, whereas for short- and narrow-channel devices it becomes also a function of the source and drain voltages to account for the charge-sharing effect. The charge-sharing modeling approach predicts the classical threshold-voltage roll-off for short channels, as well as a decreased substrate effect. However, the so-called reverse short-channel effect (RSCE) introduces a threshold voltage roll-up for shorter channels. This effect can be noted since submicron technologies are used, and can become even more important in certain deep submicron technologies. Furthermore, a stronger dependence of the threshold voltage on drain-to-source voltage than predicted by the charge-sharing approach is sometimes required, which is introduced by a model for the drain-induced barrier lowering (DIBL). These different effects will mainly affect the pinch-off voltage, through the effective threshold voltage and effective substrate effect factor, which become a function of geometry and bias. Another effect noticeable on short-channel transistors is the weak inversion slope which can be strongly degraded for shortest channels, due to onset of punchthrough, requiring further flexibility in the model. The effect of the weak inversion slope degradation will be accounted for by changing the argument of the current interpolation function.

As in common practice, the offsets between drawn and effective device geometries are first introduced, using the two parameters $D_L$ and $D_W$.

\[
L_{eff} = L + D_L \quad W_{eff} = W + D_W
\]  

(24)
The channel length correction $D_L$ will be treated as a constant for a given technology as usual in many other models. Some models consider also a bias-dependency of this parameter. However, considering both dependencies of series resistance and channel length with bias would lead to a non-solvable system of equations. Therefore, the bias dependency of series resistance, which will be retained, also accounts for eventual bias dependent effective channel length.

In the following, the models for the different effects are presented. Their fundamental form often corresponds to classical formulations also found in the literature. A number of similarities can for instance be noted with the PCIM model [18]. Mathematical conditioning is required to ensure continuity and correct behavior of the equations in all regions of operation; while this is necessary also elsewhere, this is particularly the case for the formulation of the short-channel effects. Therefore, no exhaustive presentation is made here; rather the basic forms are shown. Yet another effect affecting mainly short-channel devices is substrate current, generated through impact ionization of high energy carriers. Substrate current is added to the drain current, causing a degradation of the output conductance. This effect is also included in the model but will not be described here. While these models are reasonably simple, their combination has shown good accuracy in most situations, and also allows reasonably simple parameter extraction.

### 3.1 Velocity saturation and channel length modulation (CLM)

An increase in the field along the channel due to increasing drain-to-source voltage causes the velocity of carriers to saturate, resulting in a considerable reduction of drain current of short-channel transistors. A suitable velocity-field relationship has to be included in the model. Furthermore, as the device operates in saturation, the point in the channel where velocity saturation occurs moves towards the source, resulting in a degraded output conductance, which is referred to as channel-length modulation. Velocity saturation is accounted for in the effective mobility term as

\[
\mu_{eff}' = \mu_{eff} \left(1 + \frac{V_{DS}}{E_C \cdot L_{eff}}\right)
\]  

(25)
The auxiliary function $V_{DS}'$ ensures adequate behavior among conduction and saturation regions,

$$
V_{DS}' = \begin{cases} 
V_D - V_S & \text{for : } V_D - V_S << V_{DSS} \\
V_{DSS} & \text{for : } V_D - V_S >> V_{DSS}
\end{cases}
$$

(26)

where the drain-to-source saturation voltage is expressed in strong inversion as

$$
V_{DSS} \approx E_C \cdot L_{\text{eff}} \cdot \left[ \frac{1 + \frac{2 \cdot (V_p - V_S)}{E_C \cdot L_{\text{eff}}}}{1 + \frac{V_{DS}''}{E_C \cdot L_{\text{eff}}}} - 1 \right]
$$

(27)

and where the longitudinal critical field $E_C$ is a temperature dependent model parameter.

To account for channel length modulation, the effective channel length in the specific current (11) is replaced by an equivalent channel length $L_{\text{eff}}''$.

$$
L_{\text{eff}}'' = L_{\text{eff}}' \cdot (1 - \frac{\Delta L}{L_{\text{eff}}'})
$$

(28)

The term $\Delta L$ has a logarithmic dependence on $V_D - V_S$, according to a simplified pseudo two-dimensional analysis [19]

$$
\Delta L = \lambda \cdot L_C \cdot \ln(1 + \frac{V_{DS}''}{E_C \cdot L_C})
$$

(29)

where $\lambda$ is a model parameter. The characteristic length $L_C$ is defined as a function of the oxide capacitance and junction depth $X_J$ [19]

$$
L_C = \sqrt{\frac{X_J \cdot E_p}{C_{ox}}}
$$

(30)

The auxiliary function $V_{DS}''$ ensures that channel length modulation only occurs in saturation,

$$
V_{DS}'' = \begin{cases} 
0 & \text{for : } V_D - V_S << V_{DSS} \\
V_D - V_S - V_{DSS} & \text{for : } V_D - V_S >> V_{DSS}
\end{cases}
$$

(31)

Further details of the exact model formulation can be found in [10].
3.2 Charge-sharing and reverse short-channel effect (RSCE)

Normally, the threshold voltage decreases monotonically with decreasing the channel length [8], which is modeled using the conventional charge-sharing approach. However, in present submicron CMOS technologies, the threshold voltage initially increases when decreasing the channel length, reaches a maximum value, and then rolls off when the usual charge-sharing effect becomes dominant. The RSCE effect is often observed with n-channel devices, but may not be present in p-channel devices.

The charge-sharing principle [20] is based on geometrical considerations, for short-channel devices where the source and drain depletion regions overlap with the channel region. A reduction in the effective depletion charge controlled by the gate is predicted, depending on the extensions of the two depletion regions controlled by the source-bulk and drain-bulk junctions. The depletion depths $W_{S(D)}$ around the junctions depend on $V_{S(D)}$ as

$$W_{S(D)} = \zeta \sqrt{\psi_0 + V_{S(D)}} \quad \text{with} \quad \zeta = \sqrt{\frac{2 \cdot \mathcal{E}_d}{q \cdot N_{sub}}}$$

(32)

In narrow channel devices, the depletion region beneath the channel inversion region is not strictly confined laterally but fringes out. Therefore, an increased gate voltage is required to control the total depletion charge, resulting in an increased threshold voltage and substrate effect.

The charge-sharing effect for both short and narrow channels can be accounted for by introducing an effective substrate effect factor, which is formulated as follows [8][1][10]:

$$\gamma_{eff} = \gamma - g_L \left( \sqrt{\psi_0 + V_D} + \sqrt{\psi_0 + V_S} \right) + g_W \sqrt{\psi_0 + V_p}$$

(33)

where

$$g_L = \frac{\eta_L \cdot \varepsilon_{ox} \cdot L_{eff}^{-lex}}{C_{ox}} \quad g_W = \frac{3 \cdot \eta_W \cdot \varepsilon_{ox} \cdot W_{eff}^{-wex}}{C_{ox}}$$

(34)

The above relationship shows that the substrate effect factor is reduced for shorter channel length, while it is increased for narrow channel widths. Two empirical parameters $\eta_L$ and $\eta_W$ are used, related to the extension of the respective depletion regions. The exponent of the effective channel length usually takes a value of $lex = 1$, and similarly $wex = 1$ for effective channel width; the latter are handled as empirical model parameters. The
Chapter 3

The charge-sharing effect is in general only weakly temperature dependent. $\gamma_{\text{eff}}$ is then used in the equation of the pinch-off voltage (6) [1]:

$$V_{P\text{eff}} = V_G' - \psi_0 - \gamma_{\text{eff}} \cdot \left[ \sqrt{V_G'} + \left( \frac{\gamma_{\text{eff}}}{2} \right)^2 - \frac{\gamma_{\text{eff}}}{2} \right]$$

(35)

where $V_G = V_G - V_{TO} + \psi_0 + \gamma \cdot \psi_0 \cdot \psi_0$. The resulting effective slope factor is obtained as

$$n_{\text{eff}} = 1 + \frac{\gamma_{\text{eff}}}{2 \cdot \sqrt{\psi_0 + V_{PS}}}$$

(36)

Note that $\gamma_{\text{eff}}$, besides changing the substrate effect, also introduces a drop of the effective threshold voltage (which is referred to bulk) for short channels, while it is increased for narrow channels:

$$\Delta V_{T(\text{CS})} = -\psi_0 \cdot \left[ g_L \cdot \left( \sqrt{\psi_0 + V_D} + \sqrt{\psi_0 + V_S} \right) - g_W \cdot \sqrt{\psi_0} \right]$$

(37)

The reverse short-channel effect (RSCE) has been found to depend on the energy and dose of punchthrough implant and reoxidation time [21]. It is explained on the basis of non-uniform channel dopant distribution along the channel region due to transient enhanced diffusion of dopants. RSCE not only considerably affects device behavior, but also complicates MOST modeling due to its strong variability with channel length. Peaks of the threshold voltage of as much as $100mV$ or even more above the long-channel value $V_{TO}$ can be observed. The location of the peak is variable as well, and it can even be situated at the shortest channel length, such that no threshold voltage roll-off is observed at all. Various models for the RSCE have been proposed, either as a correction of substrate doping [22], or as a correction on the threshold voltage [23]. They have the disadvantage to use CPU-expensive exponential terms, which also may introduce serious convergence problems during the related parameter extraction, due to the high sensitivity of exponential expressions.

A simple but reasonably accurate model of the change in threshold voltage, without use of exponentials, is derived from the original expression used by Arora [23]. It is given by [10][11]
Figure 4. Variation of the threshold voltage due to RSCE as a function of the drawn channel length for n-channel devices of a 0.5\( \mu \)m technology.

\[
\Delta V_T^{(RSCE)} = \frac{2 \cdot Q_0}{C_{ox}} \left[ 1 + \frac{1}{2} \left( \frac{\xi - \varphi}{\sqrt{\xi^2 + 4 \cdot \varphi^2}} \right) \right]^2
\]

(38)

where \( \xi = C_1 \cdot \left( 10 \cdot L_{eff} / L_K - 1 \right) \), \( \varphi \) and \( C_1 \) are two constants. (38) contains two parameters, related to the peak doping density at the source/drain ends, \( Q_0 \), and the characteristic length \( L_K \) over which the doping distribution is spread.

The threshold voltage variation due to RSCE given by (38) is combined with the threshold voltage reduction resulting from the charge-sharing model. The overall threshold voltage variation with respect to its long-channel value \( V_{TO} \) is plotted in Fig. 4 as a function of the drawn gate length in the case of a 0.5\( \mu \)m technology. Fig. 4 shows that the RSCE is well modeled without the need for exponential functions, making the extraction of the related parameters \( Q_0 \) and \( L_K \) more robust.

Note that narrow-channel effects also may show a more complicated behavior than introduced so far in the model. Generally, a strong dependency on the isolation scheme is observed [17]. In particular, inverse narrow-width effects (INWE) have been observed, resulting in an analogous but opposite behavior as with the RSCE: threshold voltage may decrease for narrower channel widths [24], before it starts to increase when the usual narrow-channel effects start to dominate. Since these effects however can even be considerably more complicated, they are not introduced here.
3.3 Drain induced barrier lowering (DIBL)

In short-channel devices, and in particular for deep submicron technologies in use at present, an additional effect can be noted to have an important impact on the device characteristics: drain induced barrier lowering (DIBL) is commonly observed as a decrease of the threshold voltage for an increased drain bias (see e.g. [17]). DIBL models commonly introduce a bias- and channel-length-dependent threshold voltage variation. Various dependencies on drain-to-source voltage and channel length have been proposed. Many models use a linear relationship between threshold voltage shift and \( V_D - V_S \), which is adequate at large \( V_D - V_S \). Theoretical analysis and measurements show that this behavior may change considerably at lower \( V_D - V_S \), where a stronger than linear relationship is observed. However, experimental results suggest that the linear relationship in combination with the charge-sharing model provides reasonable accuracy [11]:

\[
\Delta V_{T(DIBL)} = -\sigma \cdot (V_D - V_S) = -\frac{\sigma_0 \cdot \varepsilon_n}{C_{ox}} \cdot L_{eff}^{-slex} \cdot (V_D - V_S)
\]

(39)

where \( \sigma_0 \) is the DIBL related model parameter. The formulation used here is comparable to the one used in [18]. The exponent for effective length dependence usually ranges from 2 to 3. To simplify computation and parameter extraction, it is linked to the exponent of the charge-sharing model by fixing its value to \( slex = lex + 1 \). In several modeling approaches, the effect of the threshold voltage shift is maintained only in weak inversion, while it is eliminated in strong inversion. Experimental results however suggest that maintaining the shift even in strong inversion does not adversely affect the simulated device behavior. The charge-sharing already predicts a lowering of the threshold voltage in the same direction as DIBL. The global threshold voltage model is now a result of the combined effects of the charge-sharing, RSCE and DIBL effects [17] with geometry and bias:

\[
V_{T_{eff}} = V_{T_{0}} + \Delta V_{T(CS)} + \Delta V_{T(RSCE)} + \Delta V_{T(DIBL)}
\]

(40)

In practice, the related parameters are strongly process-dependent.

3.4 Weak inversion slope degradation

The pinch-off voltage concept allows to accurately predict the weak inversion slope for long-channel devices. However, as noted earlier, short-
channel devices can be affected by a reduced weak inversion slope, indicating that punchthrough is starting to occur. The behavior with channel length can however be strongly process-dependent. If no threshold voltage adjust implant is used, the weak inversion slope can be gradually degraded with shorter channel length, while a threshold voltage adjust implant can even improve the weak inversion slope for shorter channels, before it starts degrading for very short channels [25]. The improvement in weak inversion slope is explained on the basis of charge-sharing between gate and drain. Note that the charge-sharing model introduced before actually predicts an improved weak inversion slope for shorter channels due to the use of the effective slope factor $n_{eff}$ (36). The weak inversion slope degradation calls for additional flexibility in the model, which is introduced through a modification of the argument of the drain current interpolation function $F(v)$ in the following manner:

$$F_{\eta}(v) = \eta_{v}^2 \cdot F\left(\frac{v}{\eta_{v}}\right)$$

Note that this change has a similar effect as if the temperature were changed when calculating $U_{T}$ used in the normalized voltage $V \sim 1/U_{T}$ and in the specific current $I_{S} \sim U_{T}$. The resulting current in weak inversion is

$$I_{D} \equiv I_{S} \cdot \eta_{v}^2 \cdot \exp\left(\frac{V_{poff} - V_{S}}{\eta_{v} \cdot U_{T}}\right) \equiv I_{S} \cdot \exp\left(\frac{V_{G} - V_{L} - n_{eff} \cdot V_{S}}{\eta_{v} \cdot n_{eff} \cdot U_{T}}\right)$$

The factor $\eta_{v}^{-1}$ ensures that the asymptotic behavior remains unchanged in strong inversion, while its influence in weak inversion is negligible compared to the effect of $\eta_{v}$ in the argument of the exponential. Also note the similarity of this formulation with the one used in [18]. Here however the slope factor correction $\eta_{v}$ is itself linked to $L_{eff}$ through the charge-sharing and DIBL models:

$$\eta_{v} = 1 + V_{L} \cdot f\left(\gamma_{eff}, \sigma\right)$$

where the parameter $V_{L}$ allows the weak inversion slope to be adapted; the function $f\left(\gamma_{eff}, \sigma\right)$ accounts for the change in the slope factor with respect to the long-channel value due to the charge-sharing and DIBL models. This function also guarantees that the normalized transconductance-to-current ratio, for modulation from either the gate or the source, cannot exceed the
physical limit of 1. The effective slope factor for short channel devices is therefore equal to \( \eta \cdot n_{\text{eff}} \), while the subthreshold swing \( S_G \) becomes

\[
S_G \approx 2.3 \cdot \eta \cdot n_{\text{eff}} \cdot U_T
\]  

(44)

For long channels, \( \eta \) tends to 1, making the influence on weak inversion slope vanish.

### 3.4 Gate voltage dependent series resistance

Series resistance is a critical parameter for process engineering, device modeling and circuit design of present submicron CMOS technologies using most often lightly doped drain (LDD) structures. In this context, the series resistance cannot be considered as constant and independent of the gate voltage. The resistivity of the gate overlapped accumulated LDD region will be affected by the gate voltage. Typically, a decreasing series resistance with increasing gate voltage is observed. Not accounting for this effect may lead to additional errors in modeling and of course in simulation. The variation of the series resistance with the gate voltage may become less important for advanced technologies such as 0.15 \( \mu m \) and 0.1 \( \mu m \), to nearly constant values when the effective gate bias \( V_{\text{Geff}} = V_G - V_{\text{Teff}} \) is increased [26]. Nevertheless, even small variations of the resistance with the effective gate bias have a non-negligible effect on the extraction of both channel length reduction and series resistance with \( V_{\text{Geff}} \) [26], when the popular so-called resistance based method is used.

Various models of the bias dependent series resistance have been proposed [26][27]. Unfortunately, they are rather complicated and use computationally expensive exponential functions. The simple model for the bias dependent series resistance proposed here avoids the use of exponentials without noticeable loss of accuracy [11]

\[
R_{\text{s,deff}} = \frac{R_s}{W_{\text{eff}}} \left[ 1 - \left( \frac{1}{2} \cdot r + \sqrt{r^2 + 4 \cdot \varepsilon_2^2} \right) \right]
\]  

(45)

where \( r = \frac{S_{\text{VK}} \cdot (V_{\text{Geff}} / V_k - 1)}{2} \) and \( \varepsilon_2 \) is a constant. The parameters in (45) are \( V_k \), \( S_{\text{VK}} \) and \( R_s \), the diffusion resistivity at \( V_{\text{Geff}} = 0 \). Equation (45) is plotted versus the effective gate voltage in Fig. 5 and compared to measured data for a 0.35 \( \mu m \) technology [22]. The approximation error is well below 1\% for the whole range of \( V_{\text{Geff}} \).
Including the effect of series resistance explicitly in the expression for drain current has several advantages: it allows the gate voltage dependence to be introduced, whereas if external resistances are used, such a dependence is difficult to establish; also, the number of iterations when solving the circuits is reduced. Parameter extraction itself is simplified as well since solving for the extra nodes is avoided. The following relations are used as discussed in [28]:

$$\frac{I_D}{I_{D0}} \cong \frac{g_{mg}}{g_{mg0}} \cong \frac{g_{ms}}{g_{ms0}} \cong \frac{g_{md}}{g_{md0}} \cong \frac{1}{1 + g_{ms0} \cdot R_{Seff} + g_{md0} \cdot R_{Deff}}$$  \hspace{1cm} (46)$$

where the subscript 0 denotes current or conductances calculated without series resistance. This relation shows that the impact of series resistance is greatest in conduction, where $g_{md}$ is highest. In saturation, the series resistor on the drain side will have a negligible effect.

Figure 5. Gate bias dependence of the series resistance for an n-channel device of a 0.35\textmu m CMOS technology [22].

This approach is more accurate than the usual scheme of accounting for series resistances which neglects the substrate bias conductance. Note however also that this approach, while it yields accurate results for DC, will have a negative impact on AC and transient analysis, especially for high frequency applications. The series resistance is therefore accounted for
through constant external resistors independent of bias, and through an internally accounted bias dependent part using (46).

4. THE CHARGE AND THERMAL NOISE MODELS

4.1 Charges integration

The derivation of the quasi-static charge model is based on the fundamental relation existing between the inversion charge density \( Q_{\text{inv}} \) and the transconductance \( g_{ms} \) at point \( x = 0 \) in the channel where the channel voltage is equal to the source voltage \( V_S \):

\[
-Q_{\text{inv}}'(V_S) = \frac{C_{ox}}{B} \cdot g_{ms}(V_S)
\]  \hspace{1cm} (47)

Replacing \( g_{ms} \) by (15) leads to

\[
-Q_{\text{inv}}'(x = 0) = \frac{C_{ox}}{\beta} \cdot \frac{I_T}{U_T} \cdot G(i_f) = 2 \cdot n \cdot U_T \cdot C_{ox} \cdot i_f \cdot G(i_f)
\]  \hspace{1cm} (48)

The same relation exists at any point along the channel where the channel voltage is equal to

\[
-Q_{\text{inv}}'(x) = \frac{C_{ox}}{\beta} \cdot \frac{I_x}{U_T} \cdot G(i_x) = 2 \cdot n \cdot U_T \cdot C_{ox} \cdot i_x \cdot G(i_x)
\]  \hspace{1cm} (49)

where \( I_x = I_S \cdot i_x \) is the current variable, evaluated for \( V_S = V_x \). Using the relationship \( dx = -L(i_f - i_r) \cdot di \) between the position and the normalized current [1], the total charge stored in the channel is then obtained by integrating (49) from the source, where \( i \) is equal to the forward normalized current \( i_f \), to the drain where \( i \) is equal to the reverse normalized current \( i_r \) [1].
Figure 6. Node charges versus gate voltage for $V_S = 0V$ and for $V_D = 0V$ and $V_D = 2V$.

\[ Q_{\text{inv}} \equiv W \int_0^L Q_m'(x) \, dx = -\frac{W \cdot L}{i_f - i_r} \int_i^i' Q_m'(i') \, di \]

\[ = -\frac{2 \cdot n \cdot U_T \cdot C_{ox}}{i_f - i_r} \int_i^i' i \cdot G(i') \cdot di = -\frac{2 \cdot n \cdot U_T \cdot C_{ox}}{i_f - i_r} \int_i^i' \left( \frac{1}{\sqrt{4 + i}} - \frac{1}{2} \right) \cdot di \]

\[ = -n \cdot U_T \cdot C_{ox} \left( \frac{4}{3} \frac{\chi_f^2 + \chi_r \chi_f + \chi_r^2}{\chi_f + \chi_r} \right) \]

where $C_{ox} = W \cdot L \cdot C_{ox}'$ and $\chi_{f(r)} = \frac{1}{2} \sqrt{4 + i_{f(r)}}$. The charge associated to the drain node is computed using a linear charge partitioning while performing the integration.
The charge associated to the source terminal is calculated as
\[ Q_S = Q_{inv} - Q_D \]
and the gate charge is given by
\[ Q_G = Q_{inv} + Q_B + Q_{ox}, \]
where \( Q_B \) is the depletion charge given by
\[ Q_B = -C_{ox} \cdot \gamma \cdot \psi_0 + V_p \left( 1 - \frac{1}{n} \right) \cdot Q_{inv} \]
and \( Q_{ox} \) is an eventual fixed oxide charge.

All the charges are plotted versus the gate voltage for two different drain voltages (\( V_D = 0V \) and \( V_D = 2V \)) in Fig.6, which clearly demonstrates the continuity with respect to \( V_G \) and the symmetry of \( Q_S \) and \( Q_D \) with respect to source and drain sides.

Note that this charges model and the related transcapacitances model described in the next section, which are both used in [10], have been independently reported in [7] where they are obtained in a very similar manner, using the pinch-off voltage and normalized current concepts.

### 4.2 Transcapacitances model

The above formulation of the charge model allows charge conservation in transient simulation. The simulation algorithms require the partial derivatives of the charges, or transcapacitances, with respect to the terminal voltages to be formulated:
\[ C_{xy} = \pm \frac{fQ_x}{fV_y} \text{ for } x, y = G, S, D, B \]  

where the + sign is used in case \( X = Y \) and the - sign otherwise. The corresponding expressions are not shown here for brevity. Some of the intrinsic capacitances are compared in Fig. 7 with measurements from a 0.25 \( \mu \)m technology. Entirely symmetrical transcapacitances with respect to source and drain voltages are obtained, since the node charges are formulated in symmetric terms of \( i_f \) and \( i_r \), as illustrated for the case \( V_D = V_S \) in Fig. 7. The agreement of the model with measurements is excellent.
Figure 7. Measured and simulated transcapacitances versus gate voltage $V_G$, at $V_S = V_D = 0\text{V}$ from depletion to strong inversion, for a long n-channel device from a 0.25$\mu$m technology.

Note also that the pinch-off voltage $V_p$, used to calculate the basic variables used in the charges model, includes the short-channel effects used for the static model in the pinch-off voltage, namely charge-sharing, DIBL, and RSCE. Therefore, the most important short-channel effects are included in the dynamic charges/transcapacitances model. Additional short-channel effects, such as bias-dependent overlap capacitances, will be addressed elsewhere.

4.3 Noise model

The noise power spectral density can be expressed for each point of the channel as [1]

$$dS_{M_p} = 4 \cdot k \cdot T \cdot \frac{\mu(x)}{L^2} \cdot W \cdot Q_{av}'(x) \cdot dx$$  \hspace{1cm} (54)

The total power spectral density of the drain current fluctuations is then obtained through integration from source to drain, assuming constant mobility along the channel.
\[ S_{M_d} \equiv 4 \cdot k \cdot T \cdot \frac{\mu_{\text{eff}}}{L_{\text{eff}}^2} \cdot W \cdot \int_0^l Q_{\text{inv}}(x) \cdot dx \]

\[ = 4 \cdot k \cdot T \cdot \frac{\mu_{\text{eff}}}{L_{\text{eff}}^2} \cdot |Q_{\text{inv}}| \]  

(55)

however using the effective value of mobility resulting from (23) and effective channel length. Similarly as for the charges model, part of the short-channel effects are included through the pinch-off voltage. Since the expression (54) is general, the thermal noise model is valid from weak to strong inversion. Note that thermal noise does not vanish at \( V_{DS} = 0 \), as is incorrectly predicted by the models used in early SPICE versions, where the thermal noise expression is proportional to \( g_{mg} \) which can be seen to be equal to zero in Table III at \( V_{DS} = 0 \).

The thermal noise model is completed by a flicker or \( 1/f \) noise model, corresponding to expressions widely used with other MOST models:

\[ S_{f_{1/f}} \sim 4 \cdot k \cdot T \cdot \frac{K_f}{W_{\text{eff}} \cdot L_{\text{eff}} \cdot f^{A_f}} \]  

(56)

where the two parameters \( A_f \approx 1 \) and \( K_f \) are used. The parameter \( K_f \) can significantly vary with the processes; often, p-channel devices, in case they use buried channel, exhibit lower flicker noise than n-channel devices. Measurements show some bias dependency of the parameter \( K_f \). For many design applications, this dependence can however be neglected.

5. MODEL APPLICATION AND EXPERIMENTAL RESULTS

5.1 The computer simulation model

The scalable model described previously has been implemented as a compact model in the circuit simulator ELDO. Coding of the entire set of model equations requires much more than just assembling the different pieces of the model. One of the important requirements for MOST models for circuit simulation is continuity of drain current and higher-order derivatives, among all operating regions, as well as outside of normal ranges.
of operation. Such conditions may occur while the simulator establishes
operating points; therefore, mathematical conditioning is required that avoids
functions to over- or underflow and ensures robustness. The drain current as
well as the other model quantities are in principle formulated as single
equations, conditioned such that they are valid in all regions of operation.
Convenient smoothing functions are intensely used (see e.g. the set
documented in [29]). However, much care is needed in choosing them so
that fundamental physical behavior is not adversely affected.

The complete intrinsic model including series resistance requires a total
of 23 process and DC parameters, which are described in Table IV. Two
more parameters are required if vertical non-uniform doping is accounted
for; three parameters are required to account for substrate current, for a total
of 28 parameters. This number compares favorably with the more than 65
process and DC parameters required by the BSIM3v3 model [27][30]
without counting its effective length and width sensitivity parameters.
Temperature effects are formulated to affect the parameters $V_T$, $\psi_0$,
$K_p(=\mu_0 \cdot C_{ox})$, $E_C$, series resistance and substrate current, allowing good
modeling of temperature behavior over the usual temperature range from 0-
100°C. A non-quasistatic small-signal model is also provided, which uses
first-order transadmittances [1], requiring no additional parameters.

The public-domain MOST model called ‘EPFL-EKV v2.6’ [10] (see also
http://legwww.epfl.ch/ekv/), is available in most circuit simulators, among
which ADS, ANTRIM-AMS, APLAC, ELDO, HSPICE, PSPICE, SABER,
SMARTSPICE, SMASH, SPECTRE. This model is built using the same
fundamental approach for the transconductance-to-current ratio as described
here. The mobility model and short-channel effects are slightly simpler,
using a reduced parameter set. Good adaptation for CMOS technologies in
the submicron to quarter-micron range has been found and numerous analog
ICs have been successfully designed.

This model also includes all the additional features as described in this
section, making it suitable in particular for low-voltage, low-current
applications in analog and mixed analog-digital circuit simulation.

5.2 Hierarchical model structure

The EKV model is hierarchically structured so that a designer can
arbitrarily include or exclude particular physical effects, helping him to gain
insight into the underlying physics of the MOS transistor and their impact on the circuit.

Table 4: Intrinsic model parameters, symbols and units. Parameters for vertical non-uniform doping, substrate current, temperature effects and noise are not included. All parameters are available in the public domain version of the model, EPFL-EKV v2.6, except those marked (*).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>COX</td>
<td>$C_{ox}'$</td>
<td>Gate oxide capacitance</td>
<td>F/m²</td>
</tr>
<tr>
<td>XJ</td>
<td>$X_J$</td>
<td>Source &amp; Drain Junction depth</td>
<td>M</td>
</tr>
<tr>
<td>VTO</td>
<td>$V_{TO}$</td>
<td>Nominal threshold voltage</td>
<td>V</td>
</tr>
<tr>
<td>GAMMA</td>
<td>$\gamma$</td>
<td>Body effect factor</td>
<td>V/1/2</td>
</tr>
<tr>
<td>PHI</td>
<td>$\psi_0$</td>
<td>Bulk Fermi potential (2x)</td>
<td>V</td>
</tr>
<tr>
<td>KP</td>
<td>$K_p$</td>
<td>Transconductance parameter</td>
<td>A/V²</td>
</tr>
<tr>
<td>E1 (E0)</td>
<td>$E_1$</td>
<td>1st order mobility reduction coefficient</td>
<td>1/V</td>
</tr>
<tr>
<td>E2 (*)</td>
<td>$E_2$</td>
<td>2nd order mobility reduction coefficient</td>
<td>1/V</td>
</tr>
<tr>
<td>UCRIT</td>
<td>$E_C$</td>
<td>Longitudinal critical field</td>
<td>V/m</td>
</tr>
<tr>
<td>LAMBDA</td>
<td>$\lambda$</td>
<td>Depletion length coefficient</td>
<td>-</td>
</tr>
<tr>
<td>DL</td>
<td>$D_L$</td>
<td>Channel length correction</td>
<td>M</td>
</tr>
<tr>
<td>LETA</td>
<td>$\eta_L$</td>
<td>Charge sharing coefficient</td>
<td>-</td>
</tr>
<tr>
<td>LEX (*)</td>
<td>$\eta_{ex}$</td>
<td>Length exponent for charge-sharing</td>
<td>-</td>
</tr>
<tr>
<td>SIGMA0 (*)</td>
<td>$\sigma_0$</td>
<td>DIBL coefficient</td>
<td>AsV/m²</td>
</tr>
<tr>
<td>NUL (*)</td>
<td>$\eta_L$</td>
<td>Short-channel slope factor parameter</td>
<td>AsV/m²</td>
</tr>
<tr>
<td>Q0</td>
<td>$Q_0$</td>
<td>RSCE peak charge density</td>
<td>AsV/m²</td>
</tr>
<tr>
<td>LK</td>
<td>$L_K$</td>
<td>RSCE characteristic length</td>
<td>M</td>
</tr>
<tr>
<td>DW</td>
<td>$D_W$</td>
<td>Channel width correction</td>
<td>M</td>
</tr>
<tr>
<td>META</td>
<td>$\eta_N$</td>
<td>Narrow width effect coefficient</td>
<td>-</td>
</tr>
<tr>
<td>WEX (*)</td>
<td>$w_{ex}$</td>
<td>Width exponent for charge-sharing</td>
<td>-</td>
</tr>
<tr>
<td>RL (*)</td>
<td>$R_L$</td>
<td>Diffusion resistivity</td>
<td>M</td>
</tr>
<tr>
<td>VK (*)</td>
<td>$V_K$</td>
<td>Characteristic gate voltage for resistivity</td>
<td>V</td>
</tr>
<tr>
<td>SVK (*)</td>
<td>$S_{VK}$</td>
<td>Resistivity coefficient</td>
<td>-</td>
</tr>
</tbody>
</table>
The minimum parameter set required would typically be $C_{OX}$, $K_P$, $V_{TO}$, $\Gamma$, and $\phi$, while other parameters are disabled. This results in the fundamental “ideal” long-channel behavior, corresponding to the simple expressions used for hand-calculation in the asymptotic operating regions, but also including the transition regions of moderate inversion. Therefore a direct link between the hand calculation and the computer simulation models can be established, a unique feature of this MOST model. Other effects, such as mobility degradation and short-channel effects, can then be gradually added and their impact on device characteristics and circuits studied. Such a procedure imposes a number of constraints on the model formulation, rarely met in other models: the different effects modeled should remain sufficiently independent of each other and the fundamental asymptotic behavior should remain correct.

5.3 Statistical circuit simulation including matching

Device mismatch, resulting from various random processes during device fabrication, can be observed between the parameters of equally designed devices [31]. In many analog or mixed analog-digital circuits, device mismatch is the main factor limiting their performance. The standard deviation of parameters for matched transistors is often observed to follow a dependence on geometry as [31]:

$$\sigma_P \approx \frac{A_P}{\sqrt{W_{eff} \cdot L_{eff}}}$$

(57)

where $A_P$ is the area proportionality constant for the parameter $P$. The statistical variation between matched devices will therefore be degraded as the area of the devices is decreased. Such dependencies can be observed over a considerable range of geometries, as long as identically laid out transistors are considered. The device mismatch is also degraded with increasing distance; in practice, this can often be neglected if distances between matched devices are kept minimal.

Statistical parameters, used in Monte Carlo simulation, need to be made available in the model equations to account for the geometry dependence of mismatch. With the exception of a few dedicated tools [32], MOST models in most circuit simulators almost never address this crucial need. Introducing statistical parameters at the model level allows a significant reduction of effort for the designer when optimizing analog layouts. Therefore, the
statistical parameters $A_{V_{T0}}$, $A_{K_P}$ and $A_\gamma$ [31] are introduced, affecting the model parameters $V_{T0}$, $K_P$ and $\gamma$ in the following way [10]:

$$V_{T0a} = V_{T0} + A_{V_{T0}} / \sqrt{W_{\text{eff}}} \cdot L_{\text{eff}}$$ (58)

$$K_{Pa} = K_P \cdot (1 + A_{K_P} / \sqrt{W_{\text{eff}}} \cdot L_{\text{eff}})$$ (59)

$$\gamma_a = \gamma + A_\gamma / \sqrt{W_{\text{eff}}} \cdot L_{\text{eff}}$$ (60)

For each matched MOST in the circuit, an individual parameter set $V_{T0a}$, $K_{Pa}$ and $\gamma_a$ is generated according to the standard deviation of the parameter and depending on the device geometry. The matching parameters are sometimes supplied by foundries, since the knowledge of these parameters is a key figure of merit of a technology. An area efficient method to obtain such parameters from high density integrated matrices of MOS transistors has recently been presented [33].

5.4 The pinch-off voltage measurement and parameter extraction method

The parameter extraction method proposed here follows the principles outlined in [34]. Of particular importance is parameter extraction from the pinch-off versus gate voltage characteristic measured at a constant current, from which the threshold voltage and doping related parameters are obtained [13][34], as well as charge-sharing and RSCE parameters. This method allows direct extraction of some of the parameters involved, making the extraction sequence considerably simpler, while other parameters are obtained by local optimization.

The pinch-off voltage versus gate voltage characteristic can be directly measured at the source of the transistor, by biasing it at a constant current [34]. The bias current is chosen such that the transistor is operating close to the middle of moderate inversion, equal to approximately half the specific current $I_S$ [34]. This level of current is determined as follows: imposing the normalized voltage $V = 0$ in (17) , such that $V_p \equiv V_s$ , and solving for the normalized current, yields $i_e \equiv I_p / I_s \equiv 0.618$ . Using the circuit in Fig. 8 a) , the pinch-off voltage is then measured at the source while sweeping the gate voltage over the bias range of interest. The characteristics depend on device sizes as seen in the measurements from a $0.7 \mu m$ technology in Fig. 9 , showing typical behavior. For short-channel, a decreased substrate effect (increased slope of the $V_p$ vs. $V_G$ characteristic) is commonly observed.
compared to long-channel, while it is increased for narrow-channel. Similarly, the threshold voltage decreases for short channel and increases for narrow channel, unless RSCE or INWE are present.

Figure 8. Circuit for pinch-off voltage $V_P$ vs. $V_G$ measurement with constant current bias in moderate inversion; a) simple measurement setup using gate-drain connected device, b) enhanced measurement setup using an OPAMP for measurement at constant $V_{DS}$.

In the simple measurement scheme of Fig. 8 a) using the gate-drain connection, the drain-to-source voltage varies only slowly, since the source voltage increases when sweeping the gate voltage. While this method has the advantage of simplicity, the measurement scheme of Fig. 8 b) can also be used, where an operational amplifier allows to control $V_{DS}$ such that it remains strictly constant during the measurement. Instead, if available, an automatic feedback unit (AFU) can be used in instruments such as the HP4142 DC parameter analyzer, so that no dedicated measurement circuitry
is required to perform such a measurement. A small drain-to-source voltage of $V_{DS} \approx 5 \cdot U_T$ is sufficient to ensure that the transistor is saturated, so that the measured source voltage $V_S \approx V_P$ can be interpreted as the pinch-off voltage.

![Figure 9. Measurement and parameter extraction from the $V_P$ vs. $V_G$ characteristics for long, short and narrow n-channel devices of a 0.7 µm CMOS technology.](image)

Long-channel parameters are determined from the measurement performed on a long and large transistor. The threshold voltage $V_{TO}$ is determined as the particular value of $V_G$ corresponding to the $V_P \approx V_S = 0V$ crossing point [34]. $\text{GAMMA}$ and $\text{PHI}$ are extracted by fitting the analytical expression (6) of the long-channel pinch-off voltage to the measured characteristic. For the case where vertical non-uniform doping is considered, the related parameters can also be determined from the same measured characteristic [13].

The pinch-off voltage measured for transistors with varying channel length can then be used to extract the charge-sharing and RSCE related model parameters. Note that the current bias needs to be adapted for each geometry, so that the same inversion level is maintained for all devices. Similarly as for long-channel, the analytical expression of the pinch-off
voltage $V_{Peff}$, including now the short-channel effects, is adapted to the measured characteristics. The charge-sharing related parameters $LETA$ and $LEX$ are determined first, by adjusting the slope of the $V_p$ vs. $V_G$ characteristics for various channel lengths. Typically three or more different channel lengths are used. The effective threshold voltage for each device geometry corresponds to the intersection points $V_{Peff}$; plotting $\Delta V_T = V_{Peff} - V_{TO}$ vs. $L_{eff}$ results in the RSCE characteristic shown in Fig. 4. The charge-sharing parameters already imply a roll-off of the threshold voltage. The parameters $Q0$ and $LK$ can be determined by matching the analytical expression $V_{reff}$ to the measured effective threshold voltage. The DIBL related parameter $SIGMA0$ can be obtained if the pinch-off voltage measurement is performed at various $V_{DS}$, using if necessary the modified measurement circuit. A simple transformation of the pinch-off voltage characteristic $V_{reff} = V_G - V_p$ allows to obtain the $V_{TB}$ vs. $V_p$ characteristic in a very simple manner [13]. An increased sensitivity is obtained when using the transform $V_{TB}(V_p)$ instead of $V_p(V_G)$ when extracting the related parameters.

Note that this method of characterization of the substrate effect, using a constant current bias in moderate inversion to measure the $V_p$ vs. $V_G$ characteristic, is not restricted to be used only with the EKV MOST model, but can also be applied with other MOST models. An important feature of this technique is that it allows to obtain threshold voltage and substrate effect from a single measured characteristic, as opposed to methods using extrapolated threshold voltage, which needs to be repeated for each back-bias. Another advantage is that this measurement technique is quite insensitive to variation of vertical field mobility and series resistance due to the low level of current that is applied.

An estimate of the specific current needs to be found for each transistor size, since the specific current is geometry- (and bias-) dependent. A reasonably simple scheme is the following: for a given gate voltage, i.e. a fixed pinch-off voltage, $I_S$ can be determined, as indicated in Fig. 10, from the strong inversion slope of the $\sqrt{I_D}$ vs. $V_S$ characteristic in saturation, which is an almost linear function except for the influence of series resistance which varies with current level. To minimize influence of mobility reduction due to vertical field and series resistance for short-channel transistors, $V_G$ and $V_D$ are chosen as low as possible, while maintaining the strong inversion and saturation condition. At any rate, a weak sensitivity of
the measured pinch-off voltage characteristic and the extracted parameters with respect to the bias current chosen is observed [34].

Figure 10: Measured $\sqrt{I_D}$ vs. $V_S$ characteristic at various gate voltages $V_G$ used to determine the specific current $I_S$.

5.5 Parameter extraction sequence

All process related model parameters can be initialized with known values from process technology, among which the gate oxide capacitance $C_{OX}$ and the junction depth $X_J$. For increased precision, the gate oxide capacitance can be obtained from AC measurements. This is of importance when thin gate oxides are used, in particular for oxide thicknesses well below $t_{ox} \approx 10\,\text{nm}$. A DC parameter extraction sequence is proposed here which includes the above described method of the pinch-off voltage measurement and parameter extraction principle. As commonly required, a set of transistors with geometries covering the range of interest of $W$ and $L$ is used.

The first step includes extraction of geometrical offset (the parameter $D_L$) and series resistance parameters, using e.g. the method in [26]. Long-
channel threshold voltage and substrate effect related parameters are then obtained from the pinch-off voltage measurement method. From the same long-channel device, the mobility parameters $K_P$, $E_1$ and $E_2$ can then be obtained from $I_D$ vs. $V_G$ at low $V_{DS}$ using local optimization. Long-channel parameters are now fixed.

In the following, short-channel parameters are obtained. Threshold voltage and charge-sharing related parameters are obtained from the pinch-off voltage measured for each device, where a best fit over geometry and bias is achieved through the adjustment of the parameters $\text{LETA}$, $\text{LEX}$, $Q_0$, $L_k$ and $\text{SIGMA}_0$. The latter can also be obtained from $I_D$ vs. $V_G$ in weak inversion at different source biases $V_S$ measured at low and high $V_{DS}$ respectively. The parameter for weak inversion slope degradation $\text{NU}$ is best obtained from the transconductance-to-current ratio $g_{md}/I_D$ in weak inversion. Short-channel parameters for velocity saturation ($U_{\text{CRIT}}$) and channel length modulation ($\lambda$) can now be obtained from the output characteristics $I_D$ and $g_{md}$ vs. $V_D$ in strong inversion at different gate voltages $V_G$. Substrate current parameters are best obtained from measured $I_S$ vs. $V_G$ characteristics at high $V_D$. In general, substrate current is only important for n-channel devices, for which $g_{md}$ is degraded at high $V_D$.

Various methods exist to obtain the channel width correction $D_W$, similarly to the channel length correction $D_L$. It can also be simply obtained by adapting the $I_D$ vs. $V_G$ characteristics in strong inversion. The narrow-channel effect related parameters ($W_{\text{ETA}}$, $W_{\text{EX}}$) are then obtained in a similar manner as short-channel parameters, using the pinch-off voltage method. Generally, reasonable results are obtained also for devices having both short and narrow channels, even though no particular modeling of the joined short- and narrow-channel effects has been introduced in the present model formulation.

Clearly, the pinch-off voltage measurement/extraction method considerably simplifies the parameter extraction. If this measurement capability is not at hand, it can be replaced by conventional techniques using I-V curve fitting more extensively, at the cost of reduced efficiency and accuracy.

The above sequence can be refined and adapted to different technologies and/or particular operating regions if needed. It is sometimes necessary to reextract a given parameter if its value depends on another parameter to be subsequently extracted. The usage of mixed direct extraction and local optimization has shown to give a reasonable compromise between accuracy
of model fit and efficiency. An automated procedure is in use when large amounts of data need to be gathered for statistical circuit simulation. The extraction method as described above has also shown to give good results with many different technologies, including deep submicron technologies.

5.6 Experimental results

In this subsection, a model validation, based on the simplified formulation of the public-domain model EPFL-EKV version 2.6 [10], is made. The scaling behavior for threshold voltage with channel length is illustrated for a 0.5µm technology in Fig. 4, where the simulated equivalent threshold voltage is shown to match the experimental data well. A small error of 6mV maximum for transistors with drawn channel lengths ranging from 0.4µm to 10µm is obtained [11].

For the same technology, the measured and simulated $I_D$ vs. $V_G$ characteristics in saturation and for different $V_S$ are compared for three different channel lengths in Fig. 11 to Fig. 13, for a long ($W = L = 10\mu m$), intermediate ($W = 10\mu m, L = 1\mu m$) and short channel device ($W = 10\mu m, L = 0.5\mu m$) respectively. The model fits well the experimental data from weak to strong inversion for all different channel lengths. Note also that the level of current corresponding to the specific current, approximately in the middle of moderate inversion is indicated for each geometry in the same figures.

Fig. 14 to Fig. 16 present the output characteristics $I_D$ and $g_{md}$ vs. $V_D$ for different $V_G$ for the same devices as above. The output conductance $g_{md}$ both in conduction and saturation is well fitted for all geometries taking into account the small number of parameters that are used. These figures demonstrate the continuity of the model among all operating regimes and the good scaling behavior, since a single set of parameters has been used for all geometries.
Figure 11. Transfer characteristics $I_D$ vs. $V_G$ of a 10µm/10µm n-channel device for different $V_S$ (0.5µm technology; symbols: measurement; lines: model). The level of the specific current $I_S$ is indicated.

Figure 12. Transfer characteristics $I_D$ vs. $V_G$ of a 10µm/1µm n-channel device for different $V_S$ (0.5µm technology; symbols: measurement; lines: model). The level of the specific current $I_S$ is indicated.
Figure 13. Transfer characteristics $I_D$ vs. $V_G$ of a 10\(\mu\)m/0.5\(\mu\)m n-channel device for different $V_S$ (0.5\(\mu\)m technology; symbols: measurement; lines: model). The level of the specific current $I_S$ is indicated.

Figure 14. Output characteristics $I_D$ and $g_{md}$ vs. $V_D$ at different $V_G$ of a 10\(\mu\)m/10\(\mu\)m n-channel device (0.5\(\mu\)m technology; symbols: measurement; lines: model).
Figure 15. Output characteristics $I_D$ and $g_{md}$ vs. $V_D$ at different $V_G$ of a 10µm/1µm n-channel device (0.5µm technology; symbols: measurement; lines: model).

Figure 16. Output characteristics $I_D$ and $g_{md}$ vs. $V_D$ at different $V_G$ of a 10µm/0.5µm n-channel device (0.5µm technology; symbols: measurement; lines: model).
6. CONCLUSIONS

An analytical compact charge-sheet model of the MOS transistor based on a physical description of the normalized transconductance-to-current ratio has been presented. The model uses the framework of the ‘EKV’ model approach, within which the pinch-off voltage, slope factor and normalized current are the principal model variables. It has been shown that the normalized transconductance-to-current ratio characteristic is independent of the technological parameters and therefore constitutes an ideal physical basis for the development of a complete compact circuit simulation model. The long-channel static model, the dynamic charge model as well as the thermal noise model, are all obtained by integration of the function describing the transconductance-to-current ratio. Therefore the present modeling approach is highly consistent and unified for all its aspects. The continuity of the ideal model equations and their $n^{th}$-order derivatives are therefore guaranteed. Hand-calculation expressions are developed, which are intensely used in analog circuit design.

The model for computer simulation includes all major physical effects present in submicron CMOS technology. Universal mobility dependence on vertical effective field is accounted for as well as effects of vertical non-uniform doping. Short-channel effects are formulated to complement the long-channel model: velocity saturation, channel-length modulation, source and drain charge-sharing, reverse short-channel effect, drain induced barrier lowering, weak inversion slope degradation and bias-dependent series resistance, making the model applicable to deep submicron CMOS technologies. A set of only 28 intrinsic model DC parameters is used, including vertical non-uniform doping, substrate current and bias dependent series resistance, comparing favorably to other models using typically more than 65 parameters. Thanks to the comparatively simple formulation, the model is also efficient and robust. To allow evaluation of geometry dependent device mismatch on analog circuits, statistical parameters are introduced for the main intrinsic model parameters for use in Monte-Carlo simulations.

The quality of circuit simulation results not only depends on the simulation model being used, but also critically depends on the quality of extracted device parameters. A parameter extraction method specially adapted to the model’s structure and formulation, is therefore introduced. It is based on coupled direct extraction and local optimization techniques, allowing the sequence to be automated for the gathering of large amounts of data. In particular, an original measurement method, using constant current bias in moderate inversion, is used to determine threshold voltage and
substrate effect related model parameters. Experimental results have been provided using measurements from CMOS technologies in the range of 0.7µm to 0.25µm. The model’s scalability is demonstrated for a 0.5µm standard CMOS technology, using a single parameter set.

The model presented in this chapter, in the version available in public domain, called EPFL-EKV v2.6, is used in a variety of contexts including deep submicron analog design. The basic model formulation is obtained from clear physical concepts and the model is well adapted to a large range of MOS technologies. Due to the model’s strong link to circuit design practice, it also facilitates the portability and reuse of analog and mixed analog-digital ICs.

7. ACKNOWLEDGMENTS

It is a pleasure to the authors to acknowledge Wladek Grabinski for experimental support and Gerson A. S. Machado for numerous suggestions regarding the manuscript.

8. REFERENCES


