An MOS Transistor Model for RF IC Design Valid in All Regions of Operation

Christian Enz, *Member, IEEE*

Invited Paper

***Abstract*—This paper presents an overview of MOS transistor modeling for RF integrated circuit design. It starts with the description of a physical equivalent circuit that can easily be implemented as a SPICE subcircuit. The MOS transistor is divided into an intrinsic part, representing mainly the active part of the device, and an extrinsic part responsible for most of the parasitic elements. A complete charge-based model of the intrinsic part is presented. The main advantage of this new charge-based model is to provide a simple and coherent description of the dc, ac, nonquasi-static (NQS), and noise behavior of the intrinsic MOS that is valid in all regions of operation. It is based on the forward and reverse charges q_f and q_r defined as the mobile charge densities, evaluated at the source and at the drain. This intrinsic model also includes a new simplified NQS model that uses a bias and frequency normalization allowing one to describe the high-order frequency behavior with only two simple functions. The extrinsic model includes all the terminal access series resistances, and particularly the gate resistance, the overlap, and junction capacitances as well as a substrate network. The latter is required to account for the signal coupling occurring at RF from the drain to the source and the bulk, through the junction capacitances. The bias dependence of these components is discussed, mainly focusing on the overlap capacitances. The noise model is then presented, including the effect of the substrate resistances on the RF noise parameters. All the aspects of the model are validated for a 0.25-μm CMOS process. The simulation results of the Y-parameters, the transit frequency, the noise parameters, and the RF large-signal characteristics show an excellent agreement with the measured data.**

***Index Terms*—Modeling, MOS devices, MOSFETs, RF CMOS, semiconductor device modeling.**

I. Introduction

THE aggressive scaling of deep-submicrometer CMOS technologies that has been going on for more than 25 years has allowed an increase in the number of transistors per chips and, hence, has extended the functionality while at the same time dramatically pushing the speed performance [1]. Although these tremendous speed improvements have been mainly driven by the requirements of very large scale integration (VLSI) digital chips, they can also be exploited for analog RF circuits [2], [3]. Today, ultra deep-submicrometer (UDSM) technologies have cached-up or even surpassed the transit frequencies achieved by bipolar transistors [4]. The transit frequency of a 0.13-μm gate length MOS transistor can reach the 80-GHz region compared to the 50–70-GHz region obtained with a bipolar transistor from a SiGe process [4]. This clearly opens the door to full CMOS highly integrated solutions for wireless applications. Of course, high f_t and $f_{\max}$ are not enough. Good noise performance and low-power consumption are required as well. Since the noise figure is also decreasing when increasing the transit frequency [5]–[8], it also takes advantage of the down-scaling of the transistor length [8]. The power consumption is still larger than what you would get from bipolar implementation, but this is changing thanks to scaling. The operating points are moving from strong inversion toward weak inversion, improving the current efficiency and making it closer to what is obtained with a bipolar transistor.

After several years of intensive research that has demonstrated the feasibility of using CMOS technologies for RF applications, real products using CMOS also for the RF portion of a chip are now emerging. Several examples of single-chip systems for Bluetooth, including the radio transceiver together with the baseband digital processor, and fully integrated in CMOS have been presented recently [9]–[11]. Other examples of RF CMOS chips can be found in the literature [12].

Nevertheless, the design of RF integrated circuits (ICs) for real products remains a challenge due to the strong constraints on power consumption and noise that leave little margin for the RF IC designer. It is, therefore, crucial to be able to accurately predict the performance of CMOS RF circuits in order to improve design efficiency and reduce the time-to-market. This requires MOS transistor compact models that are accurate over a wide range of bias, from dc to RF and for a large set of geometries.

This paper presents an overview of MOS transistor modeling for RF IC design. Section II describes the equivalent circuit of an RF MOS transistor starting from its structure and layout. It also discusses some practical issues related to the implementation of the complete RF MOST model in a simulator such as SPICE. Section III briefly presents a very compact charge-based model of the MOST valid in all modes of operation. This is required since the down-scaling of the transistor geometry is accompanied by a reduction of the supply voltage and of the overdrive voltage that pushes the operating points more and more from the usual strong inversion regime toward moderate or even weak inversion. A complete small-signal model is described in Section IV. It includes a simple charge-based model of the intrinsic

Manuscript received May 3, 2001.

The author is with the Swiss Center for Electronics and Microtechnology, CH-2007 Neuchatel, Switzerland (e-mail: christian.enz@csem.ch).

Publisher Item Identifier S 0018-9480(02)00846-3.

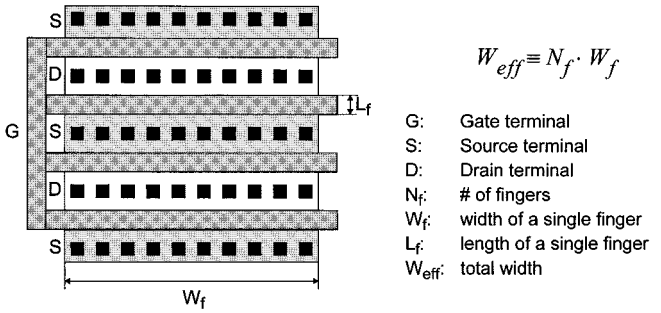

Fig. 1. Layout of a typical RF MOS transistor.

part of the MOST that accounts for nonquasi-static (NQS) operation. Section V discusses the noise properties of the MOS at HF and gives simple expressions for the four noise parameters. Finally, the large-signal operation at RF is briefly presented in Section VI.

II. MOS Transistor Equivalent Circuit at RF

A. RF MOS Transistor Structure and Layout

RF MOS transistors are usually designed as large devices in order to achieve the desired transconductance required to make the transistor operate up to RF. As shown in Fig. 1, they are usually laid out as multifinger devices, because in deep-submicrometer CMOS processes the maximum finger length (corresponding to the unit transistor width W_f) is limited. This is due to the so-called "narrow-line effect" increasing the silicided polysilicon sheet resistance as the finger width (corresponding to the transistor gate length L_f) decreases due to grain boundary problems [13]. Typical devices have up to 10 or more fingers. The total transistor effective width W_{eff} is then simply $N_f \cdot W_f$.

A cross section of a single-finger MOS transistor is presented in Fig. 2(a) together with the corresponding simplified equivalent circuit, which is discussed in the following section.

B. Equivalent Circuit at RF

Although it is always possible to have a detailed equivalent circuit that accounts for all the physical elements that are part of the RF MOS transistor, it is often too complex to be implemented as a compact model or a subcircuit for circuit simulation purposes. Moreover, many of the component values would be difficult or even impossible to extract and the subcircuit would contain too many internal nodes, which would significantly reduce the simulation efficiency. As is often the case in modeling, a tradeoff has to be found between accuracy and efficiency. A good compromise is obtained when simplifying the complete detailed equivalent circuit to the one presented in Fig. 2(b), which from experience has shown to be sufficient for most RF circuit simulations. This equivalent circuit is made of the intrinsic part of the MOST, corresponding to the active part of the device and represented in Fig. 2(b) by the MOST symbol M_i. All the other elements are only parasitic components corresponding to the extrinsic part of the device. They are made essentially of capacitances and resistances that play an increasingly important role as the operating frequency rises. Both the

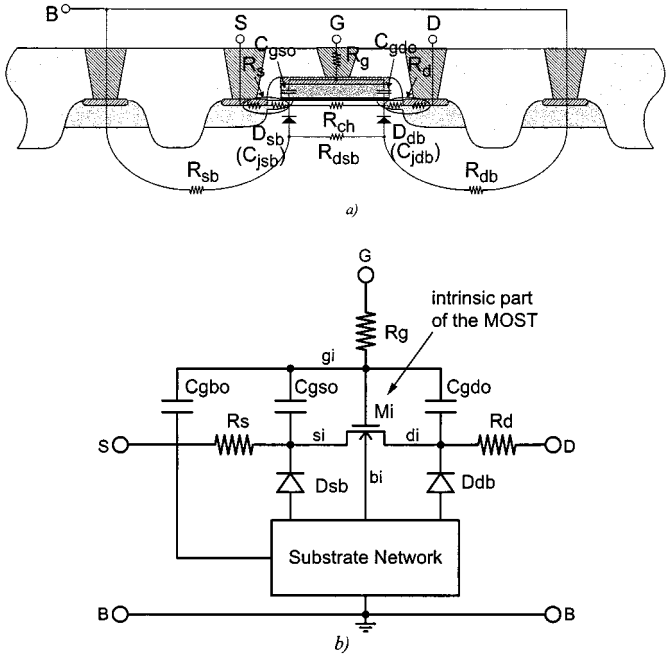

Fig. 2. MOS transistor equivalent circuit. (a) MOS transistor cross section. (b) Equivalent subcircuit.

intrinsic and extrinsic models of the MOST will be described in details in the following sections.

Note that the equivalent circuit of Fig. 2(b) does not include the parasitic components related to the test structure, such as the pad capacitances and the lead series resistances and inductances. The latter will have to be carefully deembedded from the measurements to bring the reference planes close to the useful device. For example, all the measurements presented afterwards have been cautiously deembedded using a two-step procedure [14]–[16].

C. Practical Implementations Issues

The MOS compact models available in circuit simulator such as SPICE have four terminals but do usually not include the gate resistance nor the substrate network. In order to have access to the internal nodes of the RF MOS transistor and implement the equivalent circuit of Fig. 2(b) in a SPICE simulator, most of the time a subcircuit approach is used. Note that not all the extrinsic components that are already available in the compact model (i.e., source and drain resistances, overlap capacitances, and junction diodes) can be used. For example, the source and drain series resistors in most compact models are only "soft" resistances embedded in the expression used to calculate the drain current. They account for the dc voltage drop across the source and drain resistances and its effect on the static drain current, but they do not add any poles and are, therefore, invisible for ac simulation. Therefore, they have to be added outside of the compact model as "real" resistors. Also, the source-to-bulk and drain-to-bulk diodes of the compact model have their anodes connected to the same node. Depending on the substrate network, their anodes have to be connected to two separate nodes [as shown in Fig. 2(b)]. In this case, the diodes internal to the compact model have to be turned off and two external diodes D_{sb} and D_{db} have to be added. The overlap capac-

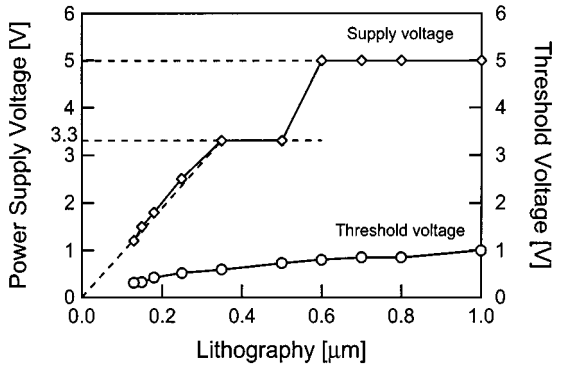

Fig. 3. CMOS process supply voltage scaling [17].

itances C_{gso} and C_{gdo} are usually also part of compact models, but not all provide good bias-dependent models. This bias dependence must be accounted for in order to obtain a RF MOST model that is valid over a large bias range. All these extrinsic components will be discussed in detail in Section IV.

Before looking at the RF operation, it is important to have a good dc model, since all the small-signal parameters are derived from it. A very compact charge-based dc model is summarized in the following section.

III. DC Model

As illustrated in Fig. 3 [17], due to technology scaling, the supply voltage had to be decreased progressively from the traditional 5 V used for older CMOS technologies (typically larger than 0.5 μm), down to 1.2 V for more recent 0.13-μm UDSM technologies, in order to reduce high electrical fields within the device and, hence, avoid the related effects. The threshold voltage can unfortunately not be scaled in the same proportion without strongly increasing the drain leakage current, which affects the static power consumption of digital chips. This results in a decrease of the effective overdrive voltage $V_G - V_{TO}$, which, in turn, moves the operating point of analog transistors more and more from strong inversion to moderate inversion and eventually even into weak inversion. In this perspective, it is important to have models that accurately predict the behavior of the MOST in all regions of operation, from strong to weak inversion, through moderate inversion. This motivates the following dc model description, which is valid in all modes of operations.

The dc and ac operations as well as the noise model of the MOST can be described in terms of the different charges stored in the device and more specifically of the inversion mobile charge density Q'_i, evaluated at the source and drain ends of the channel and defined as Q'_{iS} and Q'_{iD}, respectively [18]–[22]. This charge-based modeling approach has the following strong advantages:

1) it is grounded on the device physics;
2) it provides simple analytical formulations (at least for long-channel devices) that are valid in all modes of operations, from weak to strong inversion;
3) it insures the coherence between the dc, ac and noise models;
4) it covers a large frequency range.

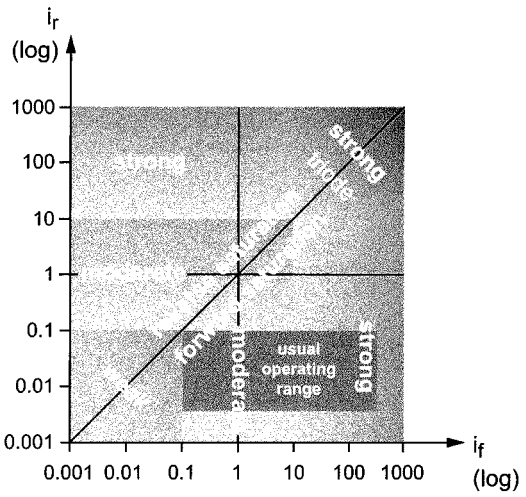

Fig. 4. Different modes of operation of the MOST.

In this charge-based model, the drain current I_D is split into the difference between a forward current I_F and a reverse current I_R [23]

$$I_D = I_F - I_R = I_{\text{spec}} \cdot (i_f - i_r) \tag{1}$$

where $i_f = I_F/I_{\text{spec}}$ and $i_r = I_R/I_{\text{spec}}$ are the forward and reverse current normalized to the specific current $I_{\text{spec}} \equiv 2 \cdot n \cdot \beta \cdot U_T^2$ [23] delimiting the regions of weak and strong inversion with $U_T = kT/q$ being the thermodynamic voltage. Factor n is the weak inversion slope, which depends on the gate-to-bulk voltage and typically ranges from 1.6 in weak inversion to 1.3 in strong inversion for an n-channel transistor (1.4–1.2 for p-channel) [23]. Note that I_{spec} depends on the transistor geometry according to $\beta = \overline{\mu}_{eff} C'_{\text{ox}} W_{\text{eff}}/L_{\text{eff}}$. The normalized currents i_f and i_r characterize the state of inversion of the channel at the source and at the drain, respectively. They can, therefore, be used to define different modes of operation as illustrated in Fig. 4. The lines $i_f = 1$ and $i_r = 1$ delimit four different regions: weak inversion for i_f and i_r much smaller than one (typically smaller than 0.1), strong inversion for i_f and/or i_r much larger than one (typically larger than ten), and moderate inversion for i_f and/or i_r close to unity (typically comprised between 0.1–10). The 45° line corresponds to $I_D = 0$. Reverse saturation corresponds to the region above this line where $i_f < i_r$ and, therefore, the drain current is negative (it circulates from the source to the drain). RF MOST are usually biased in forward saturation corresponding to the region below the 45° line where $i_f > i_r$ and $I_D > 0$ and shown by the darker zone in Fig. 4. In saturation, i_f is sometimes also called the inversion factor. The upper right region corresponds to the triode region. The forward and reverse normalized currents can be expressed in terms of the normalized forward and reverse charges q_f and q_r [18], [20]

$$i_f \equiv \frac{I_F}{I_{\text{spec}}} = q_f^2 + q_f \quad i_r \equiv \frac{I_R}{I_{\text{spec}}} = q_r^2 + q_r \tag{2}$$

with q_f and q_r defined as

$$q_f \equiv \frac{-Q'_{iS}}{2nU_T C'_{\text{ox}}} \quad q_r \equiv \frac{-Q'_{iD}}{2nU_T C'_{\text{ox}}}. \tag{3}$$

Note that the first quadratic terms in (2) correspond to the drift current circulating in strong inversion, whereas the linear terms are the diffusion current dominating in weak inversion. In moderate inversion, both drift and diffusion currents are present and, therefore, both terms have to be accounted for. The normalized forward and reverse charges are also related to the dc small-signal parameters, namely the source and drain transconductances g_{ms} and g_{md} according to

$$q_f = g_{ms}/Y_0 \quad q_r = g_{md}/Y_0 \tag{4}$$

where $Y_0 \equiv I_{\rm spec}/U_T$ and g_{ms} and g_{md} are defined as

$$g_{ms} \equiv -\left.\frac{\partial I_D}{\partial V_S}\right|_{V_G, V_D} \qquad g_{md} \equiv \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G, V_S}. \tag{5}$$

The gate transconductance is then simply calculated from (5) as

$$g_m \equiv \left.\frac{\partial I_D}{\partial V_G}\right|_{V_S, V_D} = \frac{g_{ms} - g_{md}}{n}. \tag{6}$$

From (2) and (4), it is easy to find a relation between the transconductances and the current [20], [22]

$$\begin{aligned} g_{ms} &= Y_0 \cdot q_f = Y_0 \cdot \frac{2i_f}{\sqrt{4i_f+1}+1} = \frac{I_F}{U_T} \cdot \frac{2}{\sqrt{4i_f+1}+1} \\ g_{md} &= Y_0 \cdot q_r = Y_0 \cdot \frac{2i_r}{\sqrt{4i_r+1}+1} = \frac{I_R}{U_T} \cdot \frac{2}{\sqrt{4i_r+1}+1}. \end{aligned} \tag{7}$$

In saturation, $i_f \gg i_r$ and, therefore, $g_{ms} \gg g_{md}$ resulting in the gate transconductance becoming simply proportional to the source transconductance $g_m \cong g_{ms}/n$. In the linear region, the gate transconductance becomes zero since $i_f = i_r$ and $g_{ms} = g_{md}$. Equation (7) can also be written in terms of the transconductance-to-current ratio, which is a good factor of merit to evaluate the current efficiency of a device. The normalized g_m/I_D relation in saturation is then given by [20], [22]

$$\frac{g_{ms} \cdot U_T}{I_D} = \frac{g_m \cdot n \cdot U_T}{I_D} \cong \frac{2}{\sqrt{4i_f+1}+1}. \tag{8}$$

The higher the g_m/I_D, the more efficient the device is. The maximum of (8) equals one and is reached in weak inversion (for $i_f \ll 1$), whereas it decreases like $1/\sqrt{i_f}$ in strong inversion (for $i_f \gg 1$). The g_m/I_D curve is very useful for the designers since it helps to choose the appropriate operating point of the device by selecting the right inversion factor i_f.

The current and transconductances are related to the terminal voltages through the forward and reverse charges q_f and q_r according to [21], [22]

$$\begin{aligned} \frac{V_P - V_S}{U_T} &\cong 2q_f + \ln(q_f) \\ &= \begin{cases} 2q_f & q_f \gg 1 \quad \text{(strong inversion)} \\ \ln(q_f) & q_f \ll 1 \quad \text{(weak inversion)} \end{cases} \end{aligned} \tag{9}$$

and

$$\begin{aligned} \frac{V_P - V_D}{U_T} &\cong 2q_r + \ln(q_r) \\ &= \begin{cases} 2q_r & q_r \gg 1 \quad \text{(strong inversion)} \\ \ln(q_r) & q_r \ll 1 \quad \text{(weak inversion)} \end{cases} \end{aligned} \tag{10}$$

where $V_P \cong (V_G - V_{TO})/n$ is the pinchoff voltage corresponding to the value of the channel voltage for which the inversion charge density becomes zero. For a long-channel transistor, $V_P - V_S$ is also the drain-to-source saturation voltage above which the transistor enters in forward saturation. This value is then reduced by the effect of velocity saturation as described below.

Strictly speaking, the partitioning of the drain current in a forward and reverse mode is only valid as long as the mobility can be considered as constant, which is the case at low electrical fields. Mobility does not affect the charges Q'_{iS} and Q'_{iD}, which depend on the terminal voltages. However, mobility mainly affects the relations between the charges and the current and between the charges and the transconductances. The effects of mobility reduction due to both the vertical electrical field and the velocity saturation of the carrier due to a large lateral field, can easily be incorporated into the specific current using the effective mobility defined as [21]

$$\overline{\mu}_{\rm eff} = \frac{\mu_0}{1 + \dfrac{Q_B + \eta \cdot Q_i}{\varepsilon_{si} \cdot E_0 \cdot W_{\rm eff} \cdot L_f} + \dfrac{2 \cdot U_T}{L_f \cdot v_{\rm sat}} \cdot (q_f - q_r)} \tag{11}$$

where μ_0 is the low-field mobility, Q_B and Q_i are, respectively, the total depletion and inversion charges, η is a parameter typically equal to one-half, E_0 is a parameter defining the field above which the effect of mobility reduction starts, and $v_{\rm sat}$ is the saturated velocity of the carriers.

Since RF MOST are usually minimum length devices, another important effect that has to be accounted for is the channel length modulation (CLM). At the onset of forward saturation, $V_D = V_P$ and the channel is pinched off on the drain side. As V_D increases above V_P, the pinchoff point moves toward the source causing the channel length to be reduced by an amount ΔL. The saturation effect is a very complex two-dimensional (2-D) or even three-dimensional (3-D) phenomenon. The channel reduction ΔL can be estimated from a quasi-2-D analysis of the velocity saturation region near the drain [24]

$$\Delta L \cong l \cdot \ln\left(1 + \frac{V_D - V_{Dsat}}{V_{pp}}\right) \tag{12}$$

where l depends on the depth of the diffusion of the source and drain extension (SDE) regions x_j according to [24]

$$l \equiv \sqrt{\frac{\varepsilon_s \cdot x_j}{C'_{\rm ox}}}. \tag{13}$$

V_{pp} is a parameter proportional to $l \cdot E_c$ and ε_s is the silicon dielectric constant. The saturation voltage $V_{\rm Dsat}$ corresponding

to the drain voltage at which the output conductance becomes zero is given by [44]

$$\frac{V_{\mathrm{Dsat}}-V_S}{E_c\cdot L_{\mathrm{eff}}}=\sqrt{2\cdot\frac{V_P-V_S}{E_c\cdot L_{\mathrm{eff}}}+1}-1 \tag{14}$$

where E_c is the critical electrical field above which the carrier velocity starts to saturate. As can be seen from (14), the saturation voltage $V_{\mathrm{Dsat}}-V_S$ is equal to V_P-V_S when velocity saturation can be neglected (i.e., for $(2\cdot(V_P-V_S)/(E_c\cdot L_{\mathrm{eff}}))\ll 1$) and becomes smaller than the pinchoff voltage V_P-V_S when velocity saturation appears. The CLM effect strongly impacts the output conductance in saturation.

Additional effects such as drain induced barrier lowering (DIBL), reverse short-channel effect (RSCE) can be included in the expression of the threshold voltage. The description of these effects are beyond the scope of this paper and will not be discussed here. The interested reader can refer to [25].

As mentioned earlier, the above equations are valid in all modes of operation from weak to strong inversion. They will be used below to describe the small-signal circuit.

IV. Small-Signal Modeling at RF

A. Intrinsic Part

A very general small-signal equivalent circuit of the intrinsic part of the MOST valid in all regions of operation, including NQS mode, is presented in Fig. 5(a) [26], [28]. The currents of the voltage-controlled current sources (VCCSs) are defined by

$$\begin{aligned} I_m &= Y_m\cdot(V(gi)-V(bi))\\ I_{ms} &= Y_{ms}\cdot(V(si)-V(bi))\\ I_{md} &= Y_{md}\cdot(V(di)-V(bi)) \end{aligned} \tag{15}$$

where the gate transadmittance Y_m depends on the source and drain transadmittances Y_{ms} and Y_{md} according to

$$Y_m=\frac{Y_{ms}-Y_{md}}{n}. \tag{16}$$

The source and drain transadmittances can be expressed as a product of a bias-dependent term corresponding to the low-frequency transconductances g_{ms} and g_{md} and a function ξ_m that accounts for NQS effects and, hence, for the frequency dependence [28]

$$Y_{ms}\cong g_{ms}\cdot\xi_m \quad Y_{md}\cong g_{md}\cdot\xi_m \tag{17}$$

with [28]

$$\xi_m=\frac{\lambda}{\sinh(\lambda)}\cong\frac{1}{1+j\omega\cdot\tau_{\mathrm{qs}}},\qquad \text{for: } \omega\cdot\tau_{\mathrm{qs}}\ll 1 \tag{18}$$

where $\lambda\equiv(1+j)\cdot\sqrt{3\omega\tau_{\mathrm{qs}}}$ is a function of the frequency, which is normalized to $\omega_{\mathrm{qs}}\equiv 1/\tau_{\mathrm{qs}}$. As indicated in (18), for $\omega\cdot\tau_{\mathrm{qs}}\ll 1$, ξ_m reduces to a first-order function having a pole equal to ω_{qs}. The cutoff frequency ω_{qs} corresponds, therefore, to the frequency limit between quasi-static (QS) and NQS operation: for frequencies below ω_{qs}, the intrinsic transistor operates in QS mode, whereas for $\omega>\omega_{\mathrm{qs}}$ NQS effects appear due to the distributed nature of the oxide capacitance and the resistive channel below. ξ_m is plotted versus $\omega\cdot\tau_{\mathrm{qs}}$ in Fig. 6(a) together with the first- and second-order approximations. The time constant τ_{qs} is bias-dependent according to

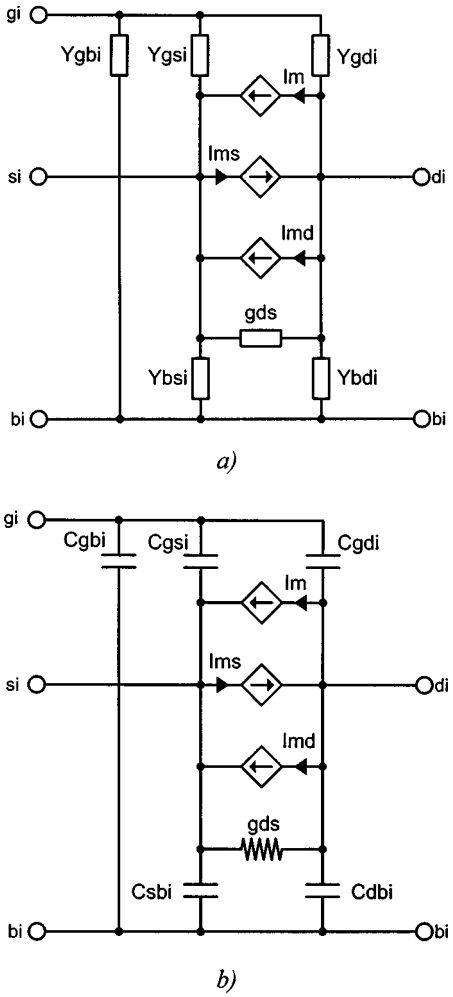

Fig. 5. Small-signal equivalent circuit of the intrinsic part of the MOST in NQS and QS operation.

$$\frac{\tau_{\mathrm{qs}}}{\tau_0}=\frac{1}{30}\cdot\frac{4q_f^2+10q_f+5+12q_fq_r+10q_r+4q_r^2}{(q_f+1+q_r)^3} \tag{19}$$

with $\tau_0\equiv L_{eff}^2/(\overline{\mu}_{\mathrm{eff}}U_T)$. In saturation ($q_f\gg q_r$), (19) reduces to

$$\frac{\tau_{\mathrm{qs}}}{\tau_0}=\frac{1}{30}\cdot\frac{4q_f^2+10q_f+5}{(q_f+1)^3}\cong\begin{cases}\dfrac{2}{15}\cdot\dfrac{1}{\sqrt{i_f}}\cong\dfrac{4}{15}\cdot\dfrac{nU_T}{V_G-V_{TO}-nV_S} & q_f\gg 1\quad\text{(strong inversion)}\\[2ex] \dfrac{1}{6} & q_f\ll 1\quad\text{(weak inversion).}\end{cases} \tag{20}$$

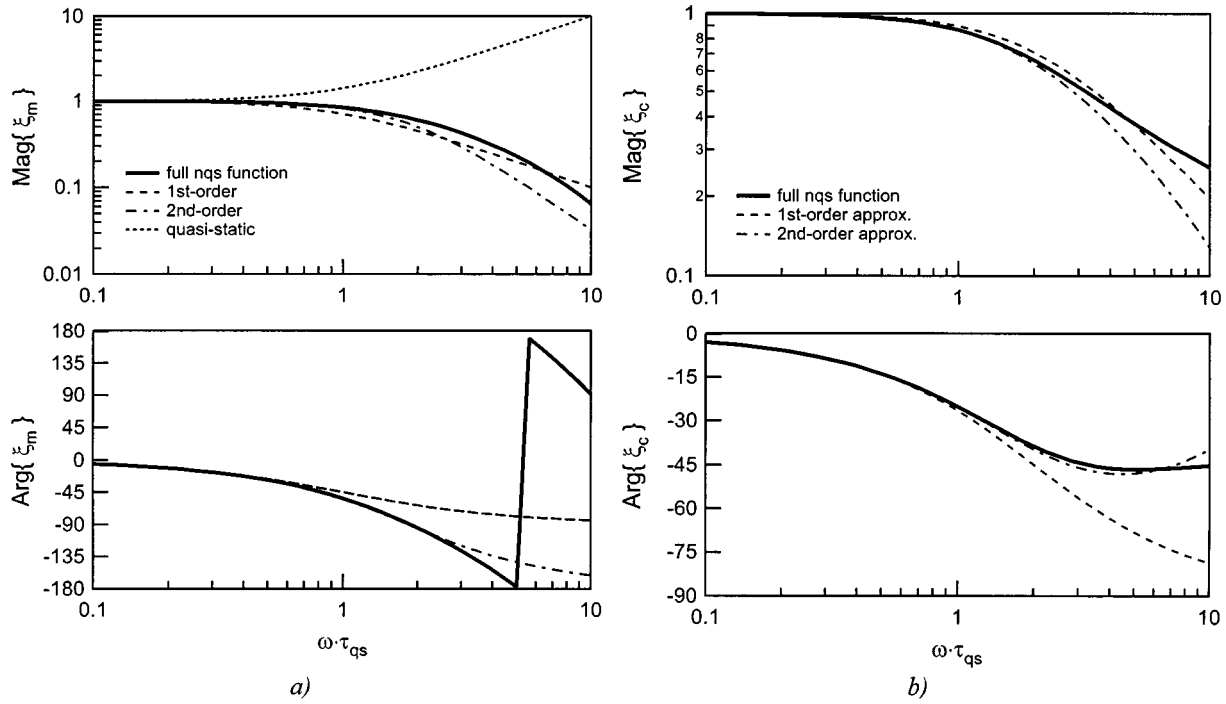

Fig. 6. Transadmittance ξ_m and admittance ξ_c NQS functions. (a) Transadmittance ξ_m NQS function. (b) Admittance ξ_c NQS function.

In strong inversion, $\tau_{\rm qs}$ decreases as the inverse of the square root of the drain current (or equivalently inversely to the gate overdrive voltage), whereas in weak inversion, it becomes bias-independent. Note that it is identical to what is obtained for the diffusion time constant in the bipolar transistor.

The transadmittance NQS function ξ_m has been verified experimentally on a long-channel NMOS transistor for different operating points in saturation and is plotted in Fig. 7. The fact that all the measured curves fall close to each other and are in excellent agreement with the theory clearly validates the transadmittance normalization $Y_{ms}/g_{ms} = \xi_m$ as well as the frequency normalization with $\tau_{\rm qs}$. Similar results have been obtained on shorter channel devices at different biases [28].

The admittance Y_{gsi} can also be decomposed into a product

$$Y_{gsi} = j\omega C_{ox} \cdot c_c \cdot \xi_c \tag{21}$$

where $C_{\rm ox} \equiv W_{\rm eff} \cdot L_f \cdot C'_{\rm ox}$ is the total gate oxide capacitance, c_c is a normalized capacitance that depends only on the bias condition according to

$$c_c = \frac{1}{3} \cdot \frac{q_f \cdot (2q_f + 4q_r + 3)}{(q_f + q_r + 1)^2} \tag{22}$$

and ξ_c is a transfer function that accounts for NQS effects. A general expression valid in all regions of operation is given in [28]. In saturation, it simplifies to

$$\xi_c \cong 2 \cdot \frac{\cosh(\lambda) - 1}{\lambda \cdot \sinh(\lambda)} \cong \frac{1}{1 + j\omega \cdot \tau_{\rm qs}/2}, \qquad \text{for } \omega \cdot \tau_{\rm qs} \ll 1. \tag{23}$$

As shown in (23), for $\omega \cdot \tau_{\rm qs} \ll 1$, ξ_c reduces to a first-order function having a cutoff frequency equal to twice that of the transadmittance function ξ_m given by (18) [25]. NQS function ξ_c is plotted versus $\omega \cdot \tau_{\rm qs}$ in Fig. 6(b) together with the first- and second-order approximations. It clearly shows the $-45°$ asymptote of the phase reached for $\omega \cdot \tau_{\rm qs} \gg 1$, which is characteristic of NQS effects. In order to verify the NQS function ξ_c experimentally, a very long channel PMOS transistor was chosen in order to bring the NQS frequency down to sufficiently low frequency and keep the effect of parasitic extrinsic components as small as possible in order to make the measurement easier and

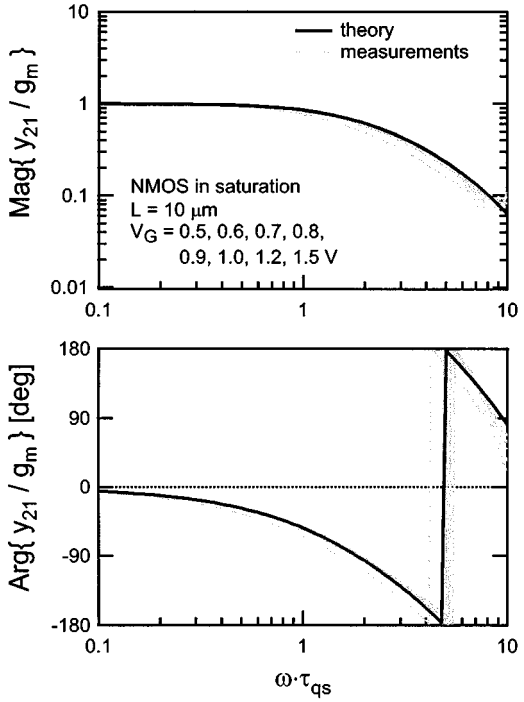

Fig. 7. Forward transadmittance y_{21} illustrating the analytical model accounting for NQS effects.

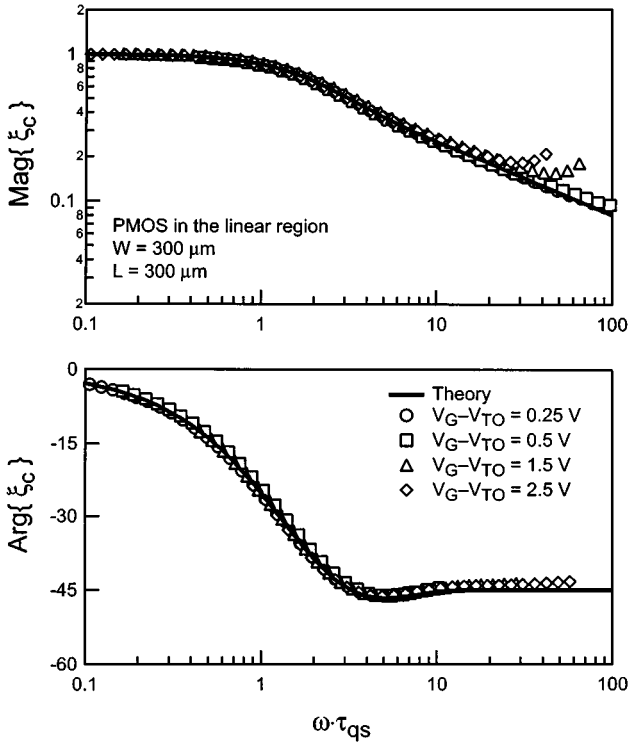

Fig. 8. NQS function ξ_c measurements.

avoid deembedding. The magnitude and phase of ξ_c are plotted in Fig. 8, where the symbols correspond to the data measured in the linear region ($V_{DS} = 0$) for several gate voltages. It shows an excellent agreement with theory for a broad range of gate voltages and over three decades of frequency. Other measurements made on shorter channel devices gave similar results [28].

Note that in strong inversion and in saturation (i.e., for $q_f \gg 1$ and $q_r \ll 1$), the normalized capacitance (22) reduces to the well-know 2/3 value, whereas in the linear region (i.e., for $q_f = q_r \gg 1$) it reduces to 1/2. In weak inversion (i.e., for both $q_f \ll 1$ and $q_r \ll 1$), the capacitance function c_c reduces to q_f, which corresponds to a diffusion capacitance since q_f is proportional to the current in weak inversion.

Admittance Y_{gdi} is obtained by symmetry by permuting q_f and q_r in (22), whereas Y_{bsi} is related to Y_{gsi} according to [23], [26]

$$Y_{bsi} = (n-1) \cdot Y_{gsi}. \tag{24}$$

Symmetrically Y_{bdi} is linked to Y_{gdi} by [23], [26]

$$Y_{bdi} = (n-1) \cdot Y_{gdi}. \tag{25}$$

Admittance Y_{gbi} can be computed from Y_{gsi} and Y_{gdi} according to [28]

$$Y_{gbi} = \frac{n-1}{n} \cdot (j\omega C_{\rm ox} - Y_{gsi} - Y_{gdi}). \tag{26}$$

The above formulation is very compact although it covers a wide range of bias and frequency range. It requires to know the forward and reverse normalized currents i_f and i_r extracted from the operating point bias information. The transconductances g_{ms} and g_{md} and the forward and reverse charges q_f and q_r can then be computed from i_f and i_r using (7). Knowing q_f and q_r allows for the calculation of $\tau_{\rm qs}$ and λ, ξ_m, ξ_c and c_c and all the transadmittances and admittances of the small-signal circuit of Fig. 5(a).

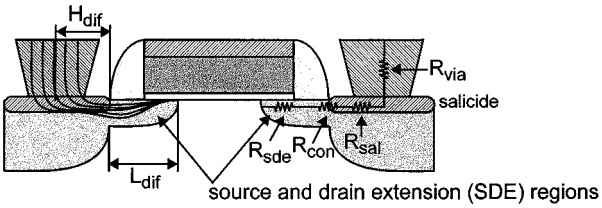

Fig. 9. Source and drain resistances.

For frequencies much below $\omega_{\rm qs}$, the NQS function can be further simplified to its QS approximation $\xi_m \cong 1 - j\omega \cdot \tau_{\rm qs}$ resulting in [26]

$$\begin{aligned} Y_{ms} &\cong g_{ms} \cdot (1 - j\omega \cdot \tau_{\rm qs}) = g_{ms} - j\omega \cdot C_{ms} \\ Y_{md} &\cong g_{md} \cdot (1 - j\omega \cdot \tau_{\rm qs}) = g_{md} - j\omega \cdot C_{md} \end{aligned} \tag{27}$$

and

$$Y_m \cong g_m \cdot (1 - j\omega \cdot \tau_{\rm qs}) = g_m - j\omega \cdot C_m \tag{28}$$

in saturation. $C_{ms} = g_{ms} \cdot \tau_{\rm qs}$, $C_{md} = g_{md} \cdot \tau_{\rm qs}$, and $C_m = g_m \cdot \tau_{\rm qs}$ are the source, drain and gate transcapacitances accounting for the nonreciprocity of the capacitances [25]. The transadmittances given by (27) and (28) correspond to the result obtained from a simple QS analysis by partitioning the channel charge linearly between the source and drain [30]. It is worth mentioning that any charge-based compact model, such as BSIM3v3, MOS Model 9 or EKV, are inherently QS models and, therefore, have transcapacitances given by (27) and (28). As illustrated in Fig. 6(a), although the phase of the transadmittances given by (27) are equal to that of the first-order model, their magnitude increases with frequency above $\omega_{\rm qs}$ instead of decreasing as the first-order and higher order models predict. This difference has to be accounted for when implementing an NQS model on top of a charge-based model to avoid wrong magnitude and phase characteristics.

In QS operation, the admittance NQS function simply reduces to unity and, hence, the admittances Y_{gsi}, Y_{gdi}, Y_{gbi}, Y_{bsi}, and Y_{bdi} become simple capacitors C_{gsi}, C_{gdi}, C_{gbi}, C_{bsi}, and C_{bdi}. The general NQS circuit of Fig. 5(a), therefore, reduces to the simple QS small-signal circuit shown in Fig. 5(b).

B. Extrinsic Part

At RF, the extrinsic part of the MOST plays an increasingly important role. It is mainly made of the series resistances R_s, R_d, and R_g accessing, respectively, the source, drain, and gate intrinsic nodes. As shown in Fig. 9, the source and drain resistances are made of different parts including the via resistance $R_{\rm via}$, the salicide resistance $R_{\rm sal}$, the salicide-to-silicon contact resistance $R_{\rm con}$ and the SDE region resistance $R_{\rm sde}$. The total

resistance is usually dominated by the contact and SDE resistances [13], [26], [27]

$$R_s = R_{\rm sde} + R_{\rm con} + R_{\rm sal} + R_{\rm via} \cong R_{\rm sde} + R_{\rm con} \tag{29}$$

with

$$R_{\rm sde} = \frac{L_{\rm dif}}{N_f W_f} \cdot R_{\text{sde-}sh} \quad R_{\rm sal} = \frac{H_{\rm dif}}{N_f W_f} \cdot R_{\text{sal-}sh} \tag{30}$$

where $L_{\rm dif}$ is the length of the SDE region and $H_{\rm dif}$ is the half width of the diffusion regions as shown in Fig. 9. The typical values of the sheet resistances are 0.9–1.1 kΩ/square for $R_{\text{sde-}sh}$ and 2–4 Ω/square for $R_{\text{sal-}sh}$. The total via resistance $R_{\rm via}$ depends on the number of via per finger with typically 3–6 Ω resistance per via. As can be seen from the above numbers, the total resistance is usually dominated by the contact and the SDE resistances. Note that the contact resistance $R_{\rm con}$ does not scale with the length of the salicide above a certain minimum value (typically $H_{\rm dif}$). This is due to the fact that most of the current flows within the salicide instead of going to the diffusion because of the latter higher resistivity. Therefore, increasing the salicide width above the minimum $H_{\rm dif}$ does not reduce the contact resistance even though the bottom contact area between salicide and silicon is increased. Since the SDE region length $L_{\rm dif}$ is almost constant, the total source and drain resistances R_s and R_d only scale with the finger width and the number of fingers according to

$$R_s = R_d \cong R_{\rm sde} + R_{\rm con} \cong \frac{0.5 R_{\rm dsw}}{N_f W_f} \tag{31}$$

where $R_{\rm dsw}$ is the drain-to-source resistance, which is typically ranging from 0.8–1 k$\Omega\mu$m for a 0.25-μm process and may become much less for more advanced UDSM processes.

The gate resistance is made of the salicide and polysilicon resistances and is given by [6], [27]

$$R_g = \kappa \cdot \frac{1}{3} \cdot \frac{W_f}{N_f \cdot L_f} \cdot R_{g\text{-}sh} \tag{32}$$

where $R_{g\text{-}sh}$ is the gate salicide sheet resistance. RF MOS transistor are usually very wide and the factor 1/3 accounts for the distributed nature of the *RC* line across the channel (in the width direction) [31], [27]. It is well known that, if the gate is connected by metal lines on both the source and drain sides, the gate resistance is four times smaller than what is obtained for a gate connected only on one side [25]. Factor κ accounts for this and is equal to unity for a gate contacted on one side and to 1/4 for a two-side contacted gate.

Note that the real part of the input impedance when looking into the gate of a common-source transistor in saturation is not constant over bias as (32) would predict. Measurements have shown a strong bias dependence [32], which cannot come from the gate nor from the source series resistances, but is due to the NQS behavior of Y_{gsi} and Y_{gbi}. Indeed, in saturation and in strong inversion, it can be shown that the real part of the impedance when looking into the *intrinsic* gate terminal of Fig. 5(a) is given by

$$\begin{aligned} R_{in} &= \mathrm{Re}\{Z_{in}\} \cong \mathrm{Re}\left\{\frac{1}{Y_{gsi} + Y_{gdi}}\right\} \\ &\cong \frac{4n}{5(3n-1)^2 \cdot g_m} \\ &\cong \frac{1}{8 \cdot g_m} \end{aligned} \tag{33}$$

where $n = 1.3$ has been used. $R_{\rm in}$ decreases inversely proportional to the overdrive voltage V_G–V_{TO}, as observed in [32]. This bias dependence can strongly affect the maximum available power gain and cause large deviations resulting in an overestimation of $f_{\max}$ [32]. It is, therefore, interesting to note that, even though NQS effects may not show up in the y_{21} parameter, they have to be accounted for in y_{11} to correctly capture the bias dependence of the input impedance. This is currently not the case of most compact models, which only offer very poor NQS models.

All the parts of the gate electrode that are not directly on top of an active part of the device contribute to the overlap capacitances. As shown in Fig. 9, the source and drain diffusion extensions under the gate create the gate-to-source and gate-to-drain overlap capacitances C_{gso} and C_{gdo}, and the extensions of the gate over the substrate corresponds to the gate-to-bulk overlap capacitance C_{gbo}. Capacitances C_{gso} and C_{gdo} are strongly bias-dependent due to the SDE regions [33]. The overlap portion of the SDE regions behave similarly to a MOS capacitor of the opposite type than the channel. Therefore, when the transistor is biased in inversion, the overlap SDE regions may be in accumulation [33]. The parasitic capacitances due to fringing electrical fields also significantly contribute to the overlap capacitances (typically more than 20%). This is particularly true for UDSM processes where the vertical dimensions become larger than the lateral dimensions and, hence, favor the fringing capacitances. The bias dependence is, therefore, very difficult to model since it involves 3-D effects. Good results have been obtained in saturation by using the following simple empirical relations [34], [35], [27]:

$$\begin{aligned} C_{gso} &= C_{gso\text{-}\min} - \frac{\Delta C_{gso}}{2} \\ &\quad \cdot \left(1 + \tanh\left(\frac{V_G - V_S - \Delta V_{gso}}{V_{gso\text{-}norm}}\right)\right) \\ C_{gdo} &= C_{gdo\text{-}\min} + \frac{\Delta C_{gdo}}{2} \\ &\quad \cdot \left(1 + \tanh\left(\frac{V_G - V_D - \Delta V_{gdo}}{V_{gdo\text{-}norm}}\right)\right) \end{aligned} \tag{34}$$

where $C_{gso\text{-}\min}$, ΔC_{gso}, ΔV_{gso}, $V_{gso\text{-}norm} C_{gdo\text{-}\min}$, ΔC_{gdo}, ΔV_{gdo}, and $V_{gdo\text{-}norm}$ are all parameters. Fig. 10 shows the gate-to-source and gate-to-drain capacitances extracted from an S-parameter measurement made at 1 GHz (symbols of the upper plots). The overlap capacitances are then obtained by subtracting the intrinsic capacitance obtained from simulations using a compact model (EKV v2.6 in this case) (symbols of

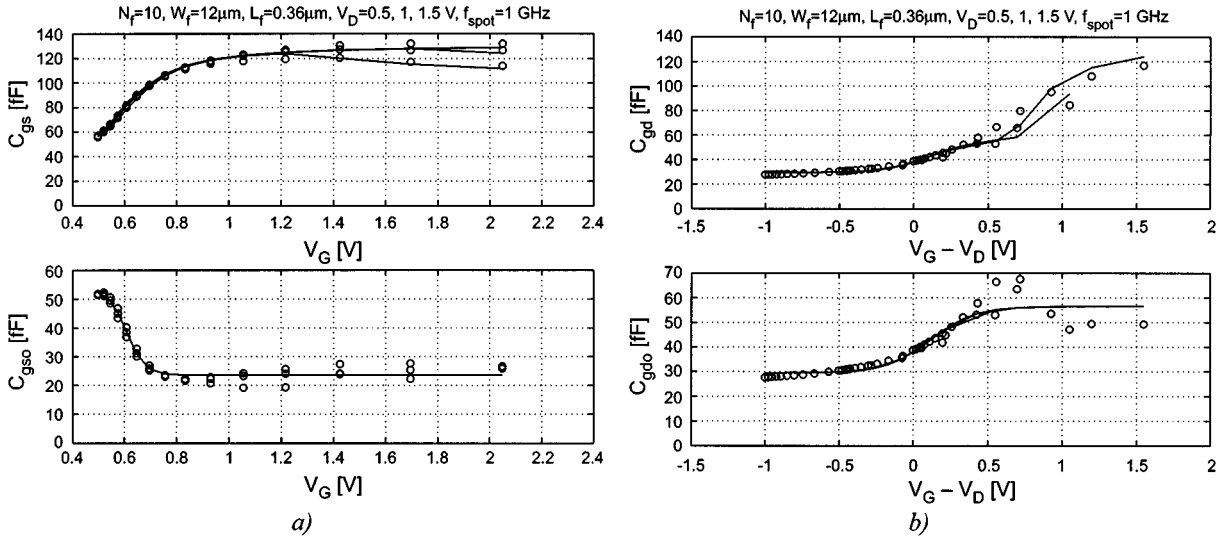

Fig. 10. Overlap capacitances bias dependence. (a) Total gate-to-source capacitance (top) and extracted overlap capacitance (bottom). (b) Total gate-to-drain capacitance (top) and extracted overlap capacitance (bottom).

lower plots). The result is then fitted with (34). The fitted capacitances are shown in Fig. 10 by the lines. In the scalable model, parameters $C_{gso\text{-}\min}$, ΔC_{gso}, $C_{gdo\text{-}\min}$, and ΔC_{gdo} are scaled with geometry.

The substrate network shown in Fig. 2(b) will be discussed in the following section.

C. Intra-Device Substrate Coupling

At high frequency, the impedances of the junction capacitances become small enough so that the RF signal at the drains couples to the nearby source diffusions and to the bulk contact through the junction capacitances and the substrate. The doping levels of UDSM CMOS processes are sufficiently high so that the substrate can be considered as purely resistive and, hence, this coupling can be modeled by a resistive network. Depending on the technology and on the frequency range to be covered, this network can reduce to a simple resistance or may need to be more complex. A good compromise is to use the Π resistive circuit made of resistances R_{dsb}, R_{sb} and R_{db}, as shown in Fig. 11 [26], [27]. Resistance R_{dsb} represents all the coupling occurring from drains to sources, whereas R_{sb} and R_{db} correspond to the coupling from source and drain to bulk. Since all the source (drain) diffusions are connected together via metal layers (assumed to have negligible resistances compared to the substrate resistances), the junction capacitances C_{jsb} and C_{jdb} as well as the substrate resistances R_{dsb}, R_{sb}, and R_{db} can reasonably be approximated as the parallel connection of all individual source and drain junction capacitances, source-to-drain, source-to-bulk and, respectively, drain-to-bulk resistances. Resistances R_{sb} and R_{db} are basically dominated by the source (drain) diffusions, which are the closest to the substrate contact. Their scaling strongly depends on the geometry of the bulk contact. For example, if there are only bulk contacts at each end of the device as shown in Fig. 12(a), R_{sb} and R_{db} are determined mainly by the source and drain diffusions that are closest to the substrate contact, resulting in a scaling with the finger width. The scaling law becomes much more complex in the more realistic case where the substrate contact partly surrounds the diffusions ["horse shoe" substrate contact of Fig. 12(b)]. In this case, part of $1/R_{sb}$ scales with the finger width and part depends on the length of the lateral substrate contact, which is proportional to the number of fingers. The substrate resistances are in principle also bias-dependent due to changes of the depletion width around the diffusions, which affect the length of the resistive path. It is shown in [40] that, for frequencies typically below 10 GHz, the Π resistive circuit can be replaced by a single substrate resistance $R_{\rm sub}$ corresponding to the parallel connection of R_{sb} and R_{db} in the circuit of Fig. 11 with $R_{dsb} = 0$. This substrate resistance $R_{\rm sub}$ shows only a weak bias dependence [40]. Other substrate networks have been published in the literature [36]–[41].

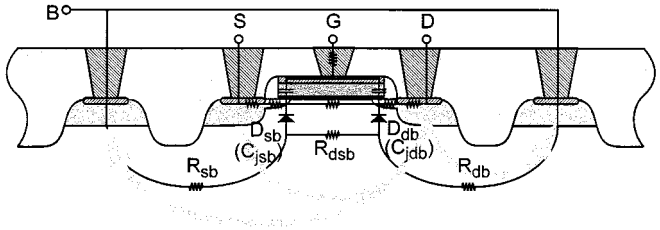

Fig. 11. Intra-device substrate coupling.

D. Y-Parameters Analysis and Measurements

Since the capacitances are all approximately proportional to the total gate width $N_f \cdot W_f$ and since the terminal resistances are inversely proportional to $N_f \cdot W_f$, the time constant due to the terminal resistances depend only on the gate length L_f, the overlap length $L_{\rm ov}$ or the diffusion width $H_{\rm dif}$. The latter dimensions are usually taken as minimum to achieve the highest cutoff frequency. Therefore, the poles due to the terminal resistances are typically at a much higher frequency than the transit frequency, so that they basically can be neglected when calculating the Y-parameters and the related quantities. Neglecting also the substrate resistances in the small-signal circuit of Fig. 2(b) (i.e., assuming that they are zero) allows one to derive the following

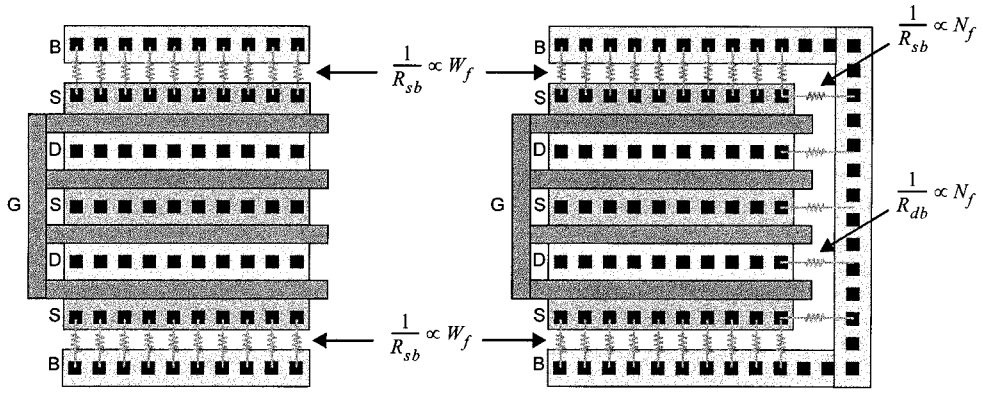

Fig. 12. Substrate contact and approximative scaling of the substrate resistances.

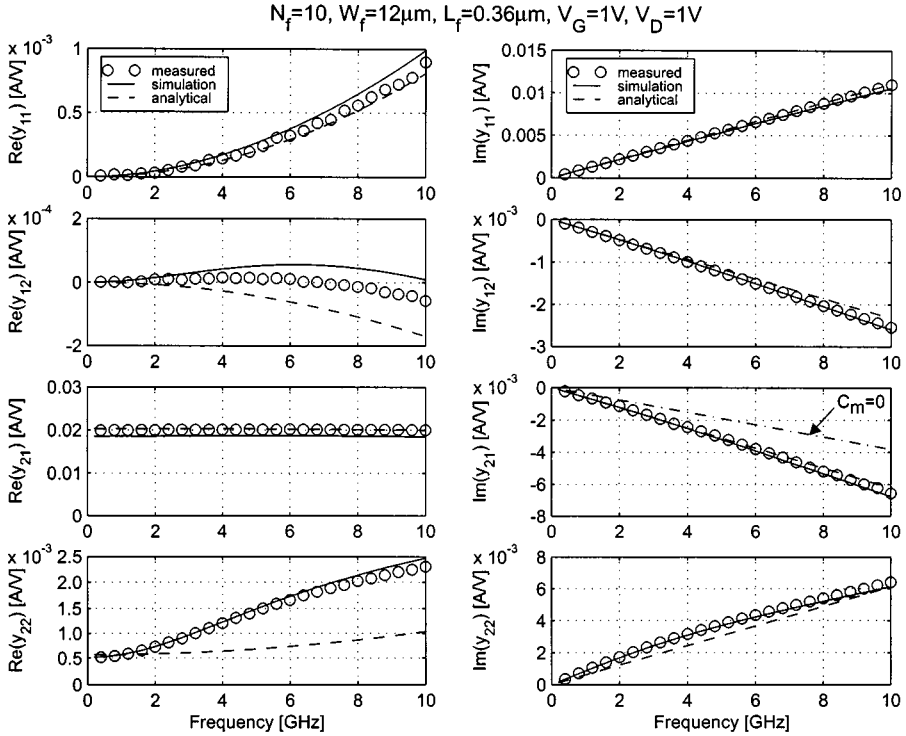

Fig. 13. Comparison between the measured, simulated, and analytical Y-parameters (n-channel, $N_f = 10$, $W_f = 12$ μm, $L_f = 0.36$ μm, and $V_G = V_D = 1$ V).

analytical expressions for the Y-parameters [26]:

$$\begin{aligned}
y_{11} &\cong \frac{j\omega C_{gg}}{1+j\omega R_g C_{gg}} \\
&\cong \omega^2 R_g C_{gg}^2 + j\omega C_{gg} \\
y_{12} &\cong \frac{-j\omega C_{gd}}{1+j\omega R_g C_{gg}} \\
&\cong -\omega^2 R_g C_{gg} C_{gd} - j\omega C_{gd} \\
y_{21} &\cong \frac{g_m - j\omega(C_m + C_{gd})}{1+j\omega R_g C_{gg}} \\
&\cong g_m - \omega^2 R_g C_{gg}(C_m + C_{gd}) \\
&\quad - j\omega(C_m + C_{gd} + g_m R_g C_{gg}) \\
y_{22} &\cong \frac{g_{ds} + j\omega(C_{bd} + C_{gd})}{1+j\omega R_g C_{gg}} \\
&\cong g_{ds} + \omega^2 R_g C_{gg}(C_{bd} + C_{gd}) \\
&\quad + j\omega(C_{bd} + C_{gd} - g_{ds} R_g C_{gg})
\end{aligned} \tag{35}$$

where $C_{gg} = C_{gs} + C_{gd} + C_{gb}$ is the total gate capacitance and all capacitances in (35) include both the intrinsic and extrinsic parts

$$\begin{aligned}
C_{gd} &= C_{gsi} + C_{gso} \\
C_{gd} &= C_{gdi} + C_{gdo} \\
C_{gb} &= C_{gbi} + C_{gbo} \\
C_{sb} &= C_{sbi} + C_{jsb} \\
C_{db} &= C_{dbi} + C_{jdb}
\end{aligned} \tag{36}$$

with C_{jsb}, C_{jdb} being the source-to-bulk and drain-to-bulk junction capacitances, respectively. The right-hand side approximations of (35) are obtained by assuming that $\omega R_g C_{gg} \ll 1$. One of the advantage of having the simple analytical expressions for the Y-parameters given by (35) is that they can be used for a direct extraction of the RF model parameters from measurements as presented in [42] and [43]. For example, C_{gd} can be extracted

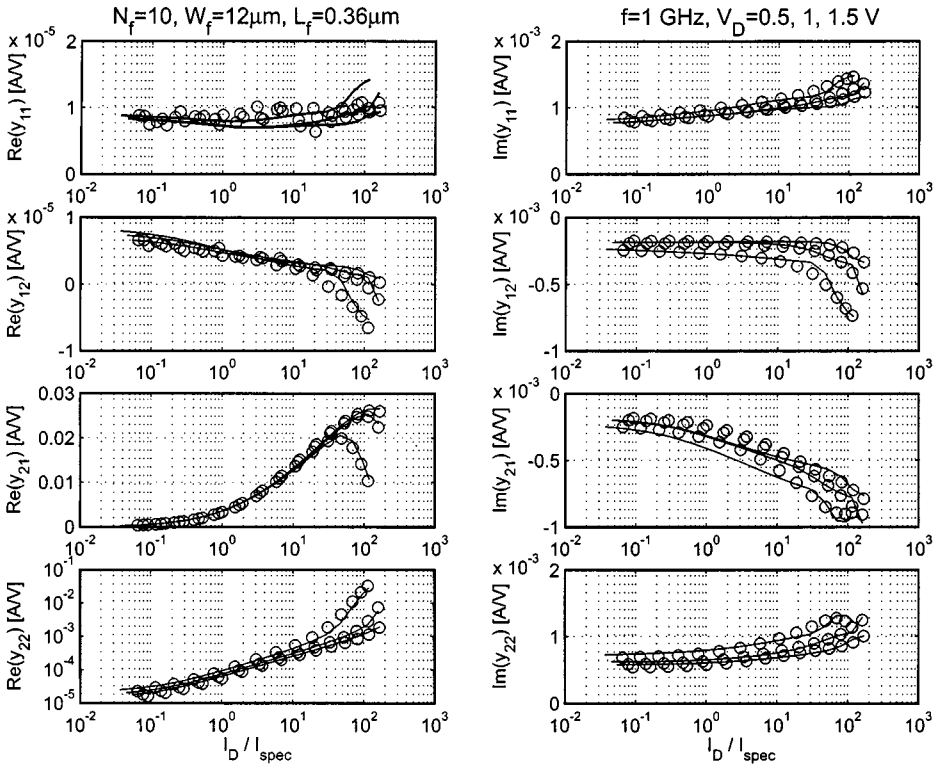

Fig. 14. Comparison between the measured and simulated Y-parameters versus inversion factor $i_f = I_D/I_{\rm spec}$ (n-channel, $N_f = 10$, $W_f = 12\ \mu$m, $L_f = 0.36\ \mu$m).

from $|\mathrm{Im}\{y_{12}\}/\omega|$ and C_{gg} from $|\mathrm{Im}\{y_{11}\}/\omega|$. In the linear region, $C_{gs} \cong C_{gd}$ and $C_{gb} \cong C_{gg} - C_{gs} - C_{gd}$, whereas in saturation $C_{gb} \ll C_{gs}$, C_{gd} and, therefore, $C_{gs} \cong C_{gg} - C_{gd}$. The gate resistance can be extracted as $\mathrm{Re}\{y_{11}\}/(\mathrm{Im}\{y_{11}\})^2$. The extraction of the substrate resistances requires a more complicated procedure described in [42] and [43].

The equations comprising (35) have been verified experimentally in Fig. 13, which shows a comparison between the measured (after a two-step deembedding process), the simulated, and the analytical Y-parameters for an n-channel transistor with $N_f = 10$, $W_f = 12\ \mu$m, and $L_f = 0.36\ \mu$m. The simulations have been performed with the complete QS model of Fig. 2(b) using the EKV v2.6 compact model. As can be seen from Fig. 13, the analytical expressions (dashed lines in Fig. 13) match the measurements very well even up to 10 GHz, except for $\mathrm{Re}\{y_{22}\}$. This discrepancy is due to the intra-device substrate coupling effect, which was not accounted for in the above analysis. Note that Fig. 13 also shows that there may be a big discrepancy in $\mathrm{Im}\{y_{21}\}$ if the transcapacitance C_m in (35) is neglected. The simulations with the complete QS model (straight lines in Fig. 13) show a very good match with measurements including the output admittance y_{22}. Note that the discrepancies in $\mathrm{Re}\{y_{12}\}$ are not critical since y_{12} is dominated by its imaginary part corresponding approximately to ωC_{gd}. Similar results have been obtained for other operating points and other device geometries using the same scalable model [26], [27]. Good results have also been achieved when using the compact model BSIM3v3 instead of EKV v2.6 [26], [27].

A good fit of the Y-parameters over frequency at a particular operating point is not to difficult to achieve. On the other hand, having the simulated Y-parameters fit the measurements over a wide range of bias at a given frequency is much more difficult to achieve. It not only requires an accurate intrinsic compact model but also accurate bias-dependent models for the extrinsic components as discussed above. The bias dependence of the described model has been checked at 1 GHz over a wide bias range by sweeping the gate voltage. The Y-parameters measured and simulated for the same device are plotted versus the inversion factor i_f in Fig. 14. An excellent fit is obtained over more than three decades of currents, which validates the bias-dependent model presented above.

Another way to evaluate the bias dependence of a model is by looking at the transit frequency f_t. It is defined as the frequency for which the extrapolated small-signal current gain $h_{21} = y_{21}/y_{11}$ drops to unity, which results in $f_t \cong g_m/(2\pi \cdot C_{gg})$. Since g_m is proportional to $W_{\rm eff}/L_f$ and C_{gg} is proportional to $W_{\rm eff} \cdot L_f$, f_t is inversely proportional to L_f^2. Note that this is only true at low lateral electrical field where there is no velocity saturation. If the carriers' velocity is saturated, the g_m is proportional to $W_{\rm eff}$, but becomes independent of $L_{\rm eff}$ and, therefore, f_t then roughly scales as $1/L_f$ instead of $1/L_f^2$ as would be expected.

The f_t measured for two n-channel transistors having different gate lengths are plotted in Fig. 15 versus the inversion factor i_f. The simulation results agree well with the measured data over a wide bias range. Note that the peaking of f_t is due to the transistor leaving saturation and entering the triode region. At this point, the transconductance starts to saturate or even decrease, while the gate-to-drain intrinsic capacitance starts to increase, making the f_t reach a maximum and then decrease.

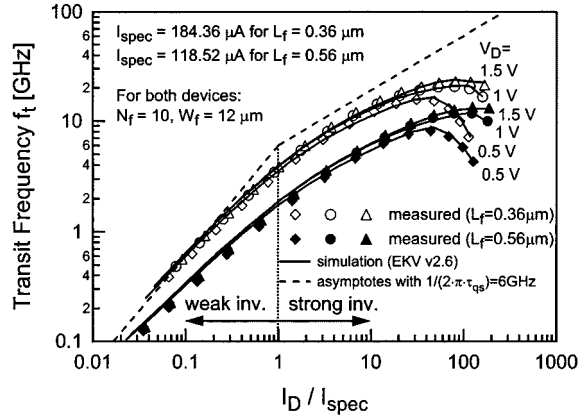

Fig. 15. Transit frequency f_t versus inversion factor $i_f = I_D/I_{\rm spec}$ for two different transistors.

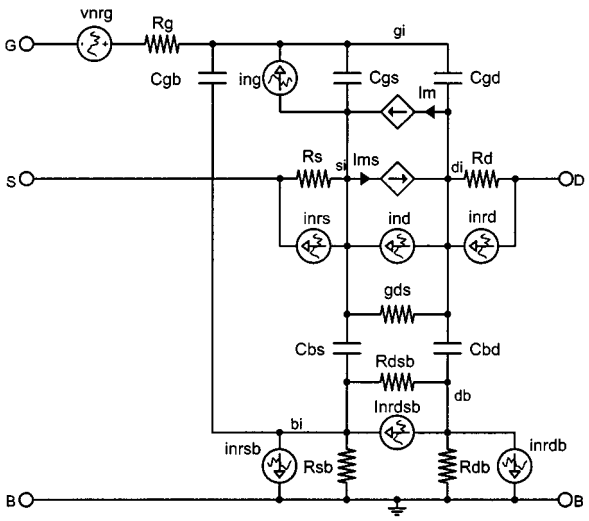

Fig. 16. Noise sources in the MOS transistor in saturation.

V. Noise Modeling at RF

A. Noise Sources in the MOS Transistor

The different noise sources in the MOS transistor are shown in Fig. 16. They include: the noise at the drain having a power spectral density (PSD) $S_{\rm ind}$, constituted by the channel thermal noise $S_{nch} = 4kT \cdot G_{nch}$ and the flicker noise $S_{\rm flicker} = g_m^2 \cdot K_f/(C'_{\rm ox}W_{\rm eff}L_{\rm eff}f^{\alpha})$, the terminal resistances thermal noise $S_{vnrg} = 4kT \cdot R_g$, $S_{inrs} = 4kT/R_s$, $S_{inrd} = 4kT/R_d$, and the substrate resistances thermal noise $S_{inrsb} = 4kT/R_{sb}$, $S_{inrdb} = 4kT/R_{db}$, and $S_{inrdsb} = 4kT/R_{dsb}$. The flicker noise mainly affects the low-frequency performance of the device and can be ignored at high frequency.[1] In addition to the channel thermal noise at the drain, at high frequency the local noise sources within the channel are capacitively coupled to the gate and generate an induced gate noise $S_{\rm ing}$.

[1]Flicker noise is nevertheless important for some RF circuits such as mixers or oscillators that up-convert the low-frequency noise around the carrier and deteriorate the phase noise or the signal-to-noise ratio.

B. Channel Thermal Noise

Although all the noise sources contribute to the total noise at high- frequency, the dominant contribution still comes from the channel thermal noise having a PSD given by

$$S_{nch} = 4kT \cdot G_{nch} \tag{37}$$

with

$$G_{nch} = \gamma \cdot g_{ms} \tag{38}$$

where $k = 1.38 \times 10^{-23} J/K$ is the Boltzman constant and T the absolute temperature. G_{nch} is the channel thermal noise conductance and γ is a bias-dependent noise excess factor [44], defined as

$$\gamma \equiv \frac{G_{nch}}{g_{ms}} \simeq \frac{|q_i|}{2 \cdot q_f} \tag{39}$$

where q_i is the total channel charge $-Q_i$ normalized to $nU_TC_{\rm ox}$

$$q_i \equiv \frac{-Q_i}{nU_TC_{\rm ox}} = \frac{1}{3} \cdot \frac{4q_f^2+3q_f+4q_fq_r+3q_r+4q_r^2}{q_f+1+q_r}. \tag{40}$$

For long-channel devices biased in the linear region (i.e., for $i_f = i_r$ and $q_f = q_r$), $q_i = 2q_f$ and, therefore, the noise excess factor is equal to unity, whereas in saturation q_i simplifies to

$$q_i = q_f \cdot \frac{4/3 \cdot q_f + 1}{q_f + 1} \tag{41}$$

which reduces to q_f in weak inversion and $4/3 \cdot q_f$ in strong inversion. The value of the noise excess factor in saturation, defined as $\gamma_{\rm sat}$, varies from 1/2 in weak inversion to 2/3 in strong inversion. $\gamma_{\rm sat}$ is related to another factor defined as [44]

$$\alpha_{\rm sat} \equiv \frac{G_{nch}}{g_m} = \frac{g_{ms}}{g_m} \cdot \gamma_{\rm sat} = n \cdot \gamma_{\rm sat}. \tag{42}$$

$\alpha_{\rm sat}$ is a good figure of merit to compare the thermal noise performance of different transconductor implementations. $\alpha_{\rm sat}$ is proportional to $\gamma_{\rm sat}$ and is typically equal to unity in strong inversion.

For short-channel devices biased in strong inversion and in saturation, the lateral electrical field might become larger than the critical field E_c resulting in the carrier velocity to saturate close to the drain and eventually even all along the channel. Since the carrier velocity is limited, an additional charge builds up in the saturation region close to the drain resulting in additional thermal noise and, therefore, in an increase of the noise excess factor $\gamma_{\rm sat}$ $(\alpha_{\rm sat})$ compared to the long-channel value 2/3 $(n \cdot 2/3)$. For electrical fields that are above some threshold field E_t, often taken equal to E_c, the cinetic energy of the carriers increases and the collisions with the lattice become more frequent. The thermal equilibrium existing between the mobile carriers and the lattice is then perturbed, giving rise to hot carriers. This effect can be modeled by a carrier effective temperature larger than the lattice temperature [45] resulting in a higher thermal noise and a larger noise excess factor than at low electrical field.

It is still not clear whether the increase of $\gamma_{\rm sat}$ is mainly due to velocity saturation or hot carriers or both. Measurements made on an old process have shown a noise excess factor as high as five, attributing these high values mainly to hot carriers [46].

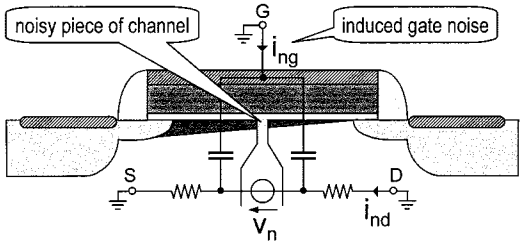

Fig. 17. Channel thermal noise capacitive coupling to the gate resulting in induced gate noise.

More recent measurements have been made on several CMOS technologies and have shown that the noise excess factor indeed increases for deep-submicrometer devices but remains below two [47]. It is also shown in [47] that usually the contribution of the hot carriers can be neglected and that the increase of γ is mainly due to velocity saturation. A more recent work [48], following the analysis initially made in [49], presents an analytical model for noise that include both the effect of velocity saturation and hot carriers. Measurements show that the contribution of hot carriers in the saturated part of the channel on the drain becomes equal to the contribution due to thermal noise in the part of the channel that is not saturated and where the carriers are at thermal equilibrium [48].

A recent work [51] claims that the traditional approach using a voltage noise source in the channel to model the contribution of an elementary noisy piece of channel is incorrect due to spatial correlation of the different voltage sources that have to be summed to get the total noise due to the channel. Numerical noise simulations using local current noise sources [50] have shown that the physical origin of the excess noise is caused by a higher local ac resistance near the source junction [51]. Noise simulations resulted in a noise excess factor increase for short-channel devices, reaching a typical value of two for devices having a 0.25-μm channel length [50], [51].

C. Induced Gate Noise

As illustrated in Fig. 17, at high frequency, the local channel voltage fluctuations due to thermal noise couple to the gate through the oxide capacitance and cause an induced gate noise current to flow [5], [25], [44], [54]. In saturation, most of the channel charge is located on the source side and, hence, this noise current can be modeled by a single noisy current source i_{ng} connected in parallel with C_{gsi}, as shown in Fig. 16, with a PSD given by [5]

$$S_{\rm ing} = 4kT \cdot G_{ng}(\omega) \tag{43}$$

where [5]

$$G_{ng}(\omega) = \delta \cdot \frac{(\omega C_{gs})^2}{5g_{ms}} = \beta_{sat} \cdot \frac{(\omega C_{gs})^2}{g_m} \tag{44}$$

where δ is a bias-dependent factor that is equal to 4/3 for a long-channel device in saturation and $\beta_{\rm sat} \equiv \delta/(5n)$ [5], [54]. Since the physical origin of the induced gate noise is the same as for the channel thermal noise at the drain, the two noise sources i_{nd} and i_{ng} are partially correlated with a correlation factor [5], [44], [54]

$$c \equiv \frac{S_{\rm ing,ind}}{\sqrt{S_{\rm ind} S_{\rm ing}}} \cong jc_g \tag{45}$$

where $S_{\rm ing,ind}$ is the cross-power spectral density [25]. For a long-channel transistor in strong inversion, $c_g \cong 0.4$ in saturation [5], [44], [54]. The frequency dependence and scaling of the correlation coefficient has recently been investigated using device noise simulations. Contradictory results have been reported. In [52], it is shown that both the imaginary and real parts of c are frequency-dependent. The real part of c is negative and tends to zero at low frequency [52]. At a given positive frequency, its value tends to decrease when reducing the channel length [52]. On the other hand, the imaginary part of c is positive and tends to increase when reducing the channel length [52]. The results published in [53] show that the correlation factor c remains mainly imaginary (the real part is about ten times smaller than the imaginary part) and that its value is slightly smaller than the long-channel value of 0.4 for short-channel devices (it typically ranges from 0.35 to 0.3 for $0.1 \leq f/f_t \leq 0.5$) [53]. A more recent study has compared the results obtained from drift-diffusion (DD) and hydro-dynamic (HD) simulations [50]. It is shown that only the HD formulation give rise to an increase of both γ and δ when reducing the gate length [50]. Values as high as two and four were obtained for γ and δ, respectively, for a 0.25-μm length n-channel MOST. Note that the ratio δ/γ remains approximately independent of the channel length ($\delta/\gamma \approx 2$) [50]. On the other hand, both DD and HD models yield an increase of the correlation factor c_g up to a value comprised between 0.7 and 0.8 for the same device [50]. Today, it is hard to draw any conclusions from these simulation results. The only observation that can be made from these different results is that for frequencies far below the cutoff frequency (typically ten times lower), the correlation coefficient remains mainly imaginary.

The induced gate noise is not implemented in all compact models yet (except for MOS Model 9 that includes the induced gate noise but without the correlation and BSIM4 [55]).

By symmetry, there is also a substrate induced noise due to the coupling of the channel thermal noise to the substrate [56], but its effect seems to be negligible due to the smaller substrate transconductance and the contribution of the substrate resistance to thermal noise.

D. HF Noise Parameters

It is well known that a noisy two-port can be represented by the same noiseless two-port and an input noise voltage source v_n having a PSD $S_v = 4kTR_v$ together with an input noise current source i_n having a PSD $S_i = 4kTG_i$ [5]. S_v and S_i depend on the internal physical noise sources and are, therefore, generally correlated. This correlation is accounted for by the correlation admittance Y_c defined as [5]

$$Y_c \equiv \frac{\overline{i_n v_n^*}}{\overline{v_n^2}} = G_c + jB_c. \tag{46}$$

A complete description of the noise always requires the following four parameters: R_v, G_i, G_c and B_c. The noise param-

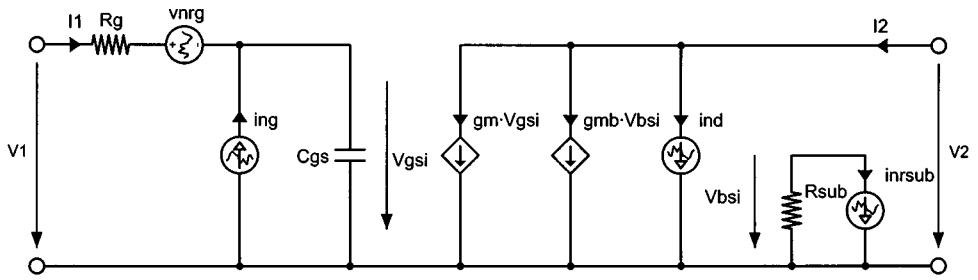

Fig. 18. Simplified small-signal schematic used for noise calculation.

eters can be evaluated from the simplified small-signal circuit shown in Fig. 18,[2] where it has been assumed that $R_{dsb} \ll R_{sb}$ and $R_{dsb} \ll R_{db}$ leading to a single substrate resistance $R_{\rm sub} = R_{sb}//R_{db}$. The capacitances C_{bs} and C_{bd} have also been neglected. After simplification, the following noise parameters are obtained [26], [27]:

$$\begin{aligned} R_v &\cong \frac{\alpha_{\rm sat}}{g_m} \cdot D_c \\ G_i &\cong \frac{\alpha_{\rm sat}}{g_m} \cdot (\omega C_{gs})^2 \cdot \psi \\ G_c &\cong \omega^2 R_g C_{gs}^2 \cdot \psi / D_c \\ B_c &\cong \omega C_{gs} \cdot \chi / D_c \end{aligned} \tag{47}$$

where $D_c \cong 1 + \alpha_g + \alpha_{sub}$ accounts for the noise of the gate and substrate resistances. α_g is the noise PSD of the gate resistance normalized to the input referred channel noise and $\alpha_{\rm sub}$ is the ratio of the output referred substrate resistance noise PSD normalized to the output referred channel noise [26], [27]

$$\alpha_g \equiv \frac{g_m R_g}{\alpha_{sat}} \qquad \alpha_{\rm sub} \equiv \frac{g_{mb}^2 R_{\rm sub}}{\alpha_{\rm sat} g_m}. \tag{48}$$

Parameters ψ and χ in (47) account for the induced gate noise and its correlation to the drain noise [26], [27]

$$\begin{aligned} \psi &\cong 1 + \alpha_{\rm sub} + \frac{\beta_{\rm sat}}{\alpha_{\rm sat}} + 2c_g\sqrt{\frac{\beta_{\rm sat}}{\alpha_{\rm sat}}} \\ \chi &\cong 1 + \alpha_{\rm sub} + c_g\sqrt{\frac{\beta_{\rm sat}}{\alpha_{\rm sat}}}. \end{aligned} \tag{49}$$

Note that both variables ψ and χ reduce to unity when the substrate noise and the induced gate noise are both ignored. Equation (49) can be simplified by using the long-channel values of $\alpha_{\rm sat}$ and $\beta_{\rm sat}$

$$\frac{\beta_{\rm sat}}{\alpha_{\rm sat}} = \frac{\delta}{\gamma} \cdot \frac{1}{5n^2} = \frac{2}{5n^2} \cong 0.24 \tag{50}$$

where $n \cong 1.3$ has been assumed. As already mentioned above, to the first-order, the simplification $\delta/\gamma \approx 2$ may even hold for short-channel devices since both $\alpha_{\rm sat}$ and $\beta_{\rm sat}$ are affected the same way by the short-channel effects. Using the above simplification, ψ and χ reduce to $\psi \cong 1.62 + \alpha_{\rm sub}$ and $\chi \cong 1.19 + \alpha_{\rm sub}$. Equation (47) shows that the induced gate noise does not affect R_v, but contributes to G_i, G_c, and B_c through the factors ψ and χ. On the other hand, substrate noise may typically contribute to 20% of R_v whereas R_g typically contributes to about 5% [26], [27].

The HF noise is often characterized by a set of four other parameters, namely, the minimum noise factor $F_{\rm min}$ [or minimum noise figure $NF_{\rm min} \equiv 10\log(F_{\rm min})$], the input referred noise resistance R_v, and the optimum source admittance $Y_{\rm opt} \equiv G_{\rm opt} + jB_{\rm opt}$ for which the minimum noise figure is obtained [5]. Parameters $F_{\rm min}$, $G_{\rm opt}$, and $B_{\rm opt}$ can be expressed in terms of the two-port noise parameters R_v, G_i, G_c, and B_c according to [26], [27]

$$\begin{aligned} F_{\rm min} &= 1 + 2R_v \cdot (G_{\rm opt} + G_c) \\ G_{\rm opt} &= \sqrt{G_i/R_v - B_c^2} \\ B_{\rm opt} &= -B_c. \end{aligned} \tag{51}$$

$G_{\rm opt}$ and $B_{\rm opt}$ can written in terms of device parameters as

$$\begin{aligned} G_{\rm opt} &\cong \omega C_{gs} \cdot \frac{\sqrt{D_c\psi - \chi^2}}{D_c} \cong \omega C_{gs} \cdot 0.47 \\ B_{\rm opt} &\cong -\omega C_{gs} \cdot \frac{\chi}{D_c} \cong -\omega C_{gs} \cdot 1.1 \end{aligned} \tag{52}$$

where the right-hand-side values have been obtained assuming that $n \cong 1.3$, $\alpha_g \cong 0.06$, $\alpha_{\rm sub} \cong 0.21$, and $D_c \cong 1.27$. From (52), it is seen that the source susceptance required for noise matching is about 1.1 times larger than that required for power matching. For $\omega R_g C_{gs} \ll 1$, $F_{\rm min}$ simplifies to

$$\begin{aligned} F_{\rm min} &\cong 1 + 2R_v \cdot G_{\rm opt} \\ &\cong 1 + \frac{\omega C_{gs}}{g_m} \cdot 2\alpha_{\rm sat}\sqrt{D_c\psi - \chi^2} \\ &\cong 1 + \frac{\omega}{\omega_t} \end{aligned} \tag{53}$$

where the transit frequency ω_t has been approximated by g_m/C_{gs}.

The noise parameters of an n-channel device have been measured and carefully deembedded using the methodology presented in [57], [58]. They are plotted in Fig. 19 and compared to the results obtained from simulation using the complete subcircuit of Fig. 2 with the additional induced gate noise source added to the subcircuit (but not accounting for the correlation between induced gate noise and drain thermal noise). Also plotted in Fig. 19 are the results obtained from (47) and (51). The discrepancies between the analytical and the measured results mainly come from a wrong frequency behavior due to the simple equivalent circuit used for the analytical derivation. Nevertheless, the simple approximation given by (53) already gives a reasonable estimation of the minimum noise figure. It can also be shown

[2]Notice the signs of i_{ng} and i_{nd}, which are important for getting the correct sign for the correlation factor. The sign convention is the same as in [5].

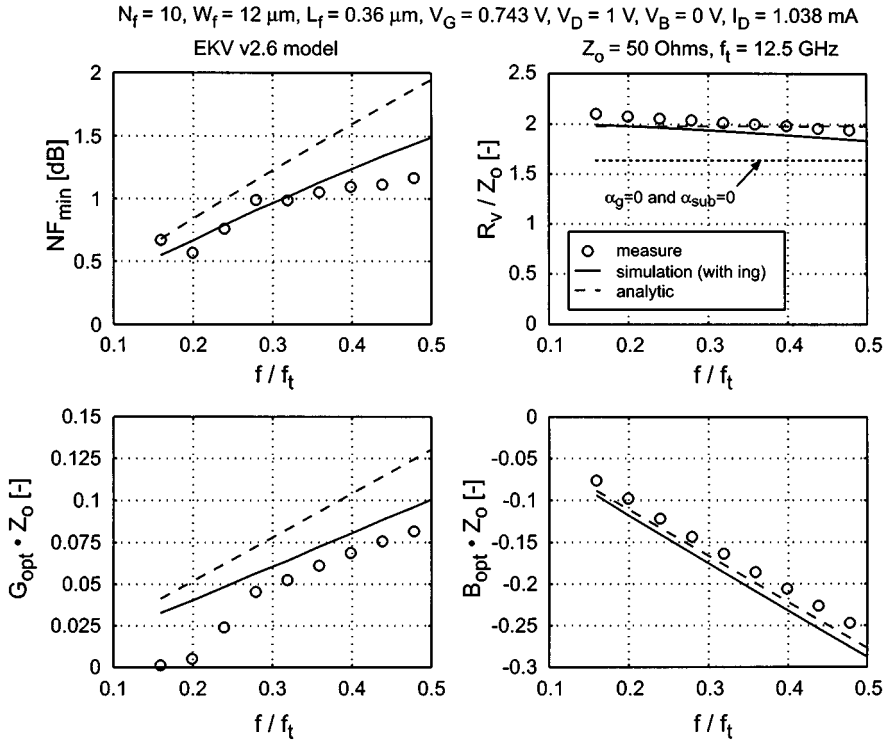

Fig. 19. Comparison between measured and simulated noise parameters.

that the effect of the correlation factor c_g on $F_{\min}$ is negligible for the long-channel values of the parameters used above.

VI. Large-Signal Modeling

The previous sections mainly described the small-signal behavior, but a good model should also account for the large-signal operation not only at low frequencies but also at RF. The low-frequency large-signal behavior is mainly captured by the static nonlinear I–V characteristics, whereas at RF the nonlinearities of the capacitances may also contribute. The subcircuit model given in Fig. 2(b) has been evaluated and compared to measurements performed at 900 MHz. The simulations have been performed with the BSIM3v3 compact model for the intrinsic MOS using a harmonic balanced simulator. Note that the subcircuit parameters have been extracted from dc and y-parameter measurements and no additional fitting was required. Fig. 20(a) shows the fundamental, second, and third harmonics versus the input power for a given bias, whereas Fig. 20(b) presents the same frequency components as a function of the bias drain current for a given input power. The match between measurement and simulation is good. The model even captures the different nulls appearing in the second and third harmonics of Fig. 20(b). The discrepancy observed in Fig. 20(a) for the third harmonic at low input power is due to a measurement limitation. The match slightly degrades at low bias (moderate and weak inversion). Additional results for the MOS Model 9 compact model can be found in [59].

VII. Moderate and Weak Inversion for RF Circuits

The high transit frequency of UDSM CMOS processes can be traded with power consumption to implement RF circuits operating in the gigahertz frequency range. This can be done by moving the operating point from strong inversion to moderate or even weak inversion, in order to spend just the required power to achieve the desired performance. There are several advantages to bias the transistor in moderate or weak inversion. The first advantage is the increase of the current efficiency (measured by the g_m/I_D ratio), which results in a further reduction of the power consumption. Secondly, the decrease of the bias voltages results in lower electrical fields within the device. This avoids velocity saturation and hot electron effects. Having no velocity saturation results in f_t scaling as $1/L_f^2$ compared to only $1/L_f$ when velocity saturation is present. This means that scaling is more effective for devices biased in the weak and moderate inversion region than in strong inversion. Thirdly, having no hot electron effects avoids the increase of the noise excess factor. Finally, the reduction of the bias voltages better accommodates the use of low supply voltages that are imposed by the scaling of UDSM technologies.

On the other hand, moving toward weak inversion changes the I_D–V_G characteristic from a quasi-quadratic to an exponential function, which clearly degrades the device linearity. Moderate inversion, therefore, represents a good tradeoff between power consumption, noise and linearity.

Part of the power is just used to fight against the extrinsic components such as overlap and junction capacitances. There might be a concern that the time constants in moderate and weak inversion might be completely dominated by these extrinsic components and, therefore, counterbalance the advantage of the current efficiency increase. A way to investigate this issue is by looking at the total transit time τ_t defined as $\tau_t \equiv 1/(2\pi f_t)$, which can be decomposed into $\tau_t = \tau_i + \tau_e$, where τ_i corresponds to the transit time of the intrinsic part $\tau_i =$

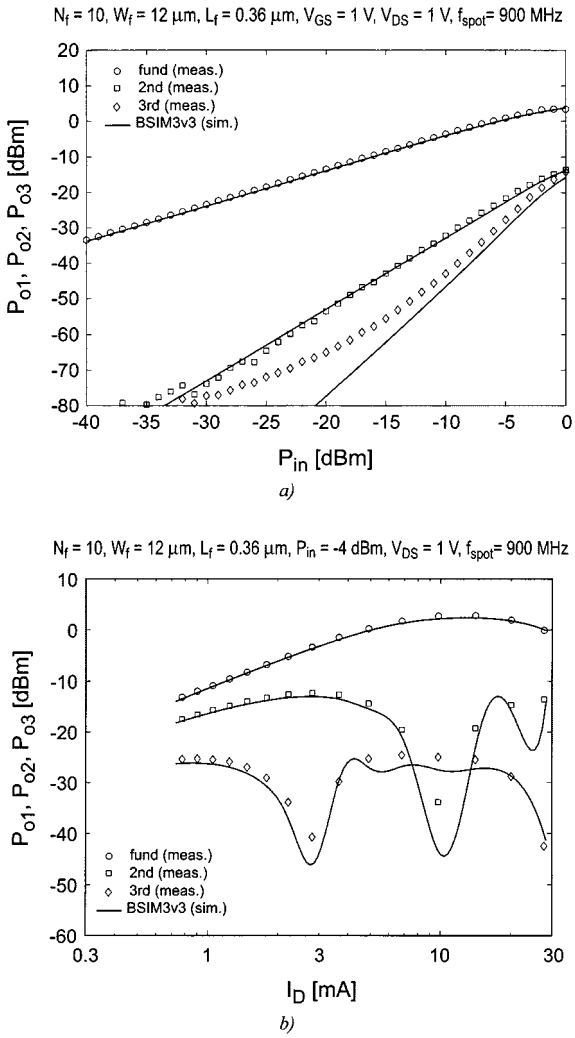

Fig. 20. Large-signal behavior. (a) Fundamental, second, and third harmonics versus the input power. (b) Fundamental, second, and third harmonics versus the bias current.

C_{ggi}/g_m with C_{ggi} being the total gate intrinsic capacitance $C_{ggi} = C_{gsi}+C_{gdi}+C_{gbi}$. The time constant τ_i represents ultimately the lower time constant the device can achieve for a given operating point. The time constant τ_e corresponds to the additional delay introduced by the extrinsic part of the device due mainly to the overlap capacitances $\tau_e \cong C_{ggo}/g_m$, where C_{ggo} is the total gate overlap capacitance $C_{ggo} = C_{gso}+C_{gdo}+C_{gbo}$. It is shown in [7] that the extrinsic parasitics typically account for about 40% of the total transit time in strong inversion and about 50% in moderate inversion. This means that the increase of parasitic to intrinsic time constant ratio does not degrade dramatically when moving the operating point from strong to moderate inversion. This is another good reason for moderate inversion to be considered for RF operation with deep-submicrometer devices in order to meet the low-voltage and low-power requirements.

VIII. Conclusion

The scaling of deep-submicrometer CMOS technologies has pushed the transit frequency to frequencies typically beyond 80 GHz. This makes CMOS a real challenger for realizing systems that include also the RF front-ends of integrated transceivers. The design of such RF CMOS integrated circuits requires MOS transistor models that are accurate up to and even beyond the GHz frequency range. In addition, to model all the physical effects that are affecting their dc characteristics, the models also have to account for a number of effects that become predominant at higher frequency. The latter include a NQS description of the intrinsic part of the MOS transistor as well as all the extrinsic components, such as the gate, source, and drain resistances, the substrate coupling resistances, the overlap and junction capacitances, that all play an increasingly important role at high frequency. The downscaling of the technology is also combined with a decrease of the supply voltage that results in a reduction of the overdrive voltage V_G–V_{TO} and consequently in a shift of the operating points from the traditional strong inversion region, into moderate or even weak inversion. In this perspective, it is important that the model accurately predicts the behavior of the MOST in all regions of operation, from strong to weak inversion, through moderate inversion.

This paper presents a complete charge-based model of the intrinsic part of the MOST that include a new simplified NQS model. The main advantage of this new charge-based model is to provide a simple and coherent description of the dc, ac, and noise behavior of the intrinsic part of the MOS transistor. It basically uses the forward and reverse charges q_f and q_r defined as the mobile charge densities, evaluated at the source and at the drain, to describe the current, the small-signal components, and the noise in all regions of operation. The new NQS model uses a bias and frequency normalization that allows one to describe the high-order frequency behavior of all the intrinsic transadmittances and admittances by only two functions ξ_m and ξ_c. At lower frequency, this model simplifies to the well-known QS model.

All the extrinsic components are described in detail, particularly the important resistive substrate network. The latter accounts for the ac coupling of the high-frequency signal from the drain diffusions, through the junction capacitances, to the nearby source diffusions and to the substrate contacts. This effect mainly affects the output admittance y_{22} and can be modeled by a simple Π resistive network. Although all these extrinsic components show a bias dependence, it is generally sufficient to account for the bias dependence of the junction and overlap capacitances, leaving the resistances bias-independent. A simple model for the bias dependence of the overlap capacitance valid in saturation is also proposed.

The different noise sources in the transistor are described, including the induced gate noise due to the capacitive coupling of the local noise source within the channel to the gate terminal. The increase of the thermal noise excess factor γ in short-channel devices is mainly due to velocity saturation and eventually to hot carriers. The noise at high frequency is characterized by four parameters, namely the input referred noise resistance R_v, the minimum noise factor $F_{\min}$, and the real and imaginary parts $G_{\rm opt}$ and $B_{\rm opt}$ of the optimum source admittance for which the noise factor is minimum. These parameters can be transformed into the power spectral densities of the voltage and current noise sources and the correlation admittance of the cor-

responding noisy two-port $ABCD$-parameters representation. The noise parameters of a simplified MOS transistor model including the induced gate noise, the correlation between the induced gate noise and the drain thermal noise, the substrate resistance noise, and the gate resistance noise are derived. The analytical results as well as the simulation results obtained from the complete subcircuit model are favorably compared to measurements. This analysis also shows that the contribution of the substrate resistance may be significant and should, therefore, not be neglected, as is often the case. An example shows that the substrate and gate resistances can typically contribute to about 20% and 5%, respectively, to the total input referred noise resistance.

The complete RF MOS model has been implemented in a SPICE simulator as a scalable subcircuit. It has been successfully validated over frequency up to 10 GHz, over geometry and over a wide bias range on a 0.25-μm CMOS process for both n- and p-channel devices.

For some RF applications in the gigahertz range using existing or future UDSM processes, it may be advantageous to move the operating point of the RF devices from strong inversion to moderate inversion to take advantage of the higher g_m/I_D ratio and, therefore, of the higher current efficiency and benefit from the lower electrical fields within the transistor. This avoids velocity saturation and hot carrier effects, resulting in a smaller excess noise factor $\alpha_{\rm sat}$. It also allows one to take full advantage of the $1/L_{\rm eff}^2$ scaling of the transit frequency compared to the $1/L_{\rm eff}$ scaling when velocity saturation is present. Finally, moderate inversion also better accommodates the increasingly stringent low-voltage constraint.

REFERENCES

[1] H. Iwai, "CMOS technology—Year 2010 and beyond," *IEEE J. Solid-State Circuits*, vol. 34, pp. 357–366, Mar. 1999.

[2] H. Iwai and H. S. Momose, "Technology toward low-power/low-voltage and scaling of MOSFETs," *Microelectron. Eng.*, vol. 39, no. 1–4, pp. 7–30, Dec. 1997.

[3] H. S. Momose, E. Morifuji, T. Yoshitomi, T. Ohguro, M. Saito, T. Morimoto, Y. Katsumata, and H. Iwai, "High-frequency AC characteristics of 1.5 nm gate oxide MOSFETs," in *Proc. Int. Electron Devices Meeting*, Dec. 1996, pp. 105–108.

[4] G. Weinberger, "The new millennium: Wireless technologies for a truly mobile society," in *Int. Solid-State Circ. Conf. Tech. Dig.*, Feb. 2000, pp. 20–25.

[5] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[6] T. Manku, "Microwave CMOS—Device physics and design," *IEEE J. Solid-State Circuits*, vol. 34, pp. 277–285, Mar. 1999.

[7] C. Enz, "MOS transistor modeling for RF integrated circuit design," in *Proc. IEEE Custom Integrated Circuits Conf.*, May 2000, pp. 189–196.

[8] H. S. Momose, R. Fujimoto, S. Otaka, E. Morifuji, T. Ohguro, T. Yoshitomi, H. Kimijima, S. I. Nakamura, T. Morimoto, Y. Katsumata, H. Tanimoto, and H. Iwai, "RF noise in 1.5 nm gate oxide MOSFET's and evaluation of the NMOS LNA circuit integrated on a chip," in *Proc. VLSI Technol. Symp.*, June 1998, pp. 96–97.

[9] F. Op't Eynde *et al.*, "A fully-integrated single-chip SOC for Bluetooth," in *Int. Solid-State Circ. Conf. Tech. Dig.*, Feb. 2001, pp. 196–197.

[10] A. Ajjikuttira *et al.*, "A fully-integrated CMOS RFIC for bluetooth applications," in *Int. Solid-State Circ. Conf. Tech. Dig.*, Feb. 2001, pp. 198–197.

[11] H. Darabi *et al.*, "A 2.4 GHz CMOS transceiver for bluetooth," in *Int. Solid-State Circ. Conf. Tech. Dig.*, Feb. 2001, pp. 200–201.

[12] A. A. Abidi, P. R. Gray, and R. G. Meyer, *Integrated Circuits for Wireless Communications*. Piscataway, NJ: IEEE Press, 1999.

[13] I. C. Chen and W. Liu, *High-Speed or Low-Voltage, Low-Power Operations in ULSI Devices*, C. Y. Change and S. M. Sze, Eds. New York: Wiley, 2000.

[14] M. C. A. M. Koolen, J. A. M. Geelen, and M. P. J. G. Versleijen, "An improved de-embedding technique for on-wafer high-frequency characterization," in *IEEE Proc. Bipolar Circuit Technol. Meeting*, 1991, pp. 188–191.

[15] M. C. A. M. Koolen, "On-wafer high-frequency device characterization," in *Proc. Eur. Solid-State Devices Res. Conf.*, Sept. 1992, pp. 679–686.

[16] H. Cho and D. E. Burk, "A three-step method for the de-embedding of high-frequency S-parameter measurements," *IEEE Trans. Electron Devices*, vol. 38, pp. 1371–1375, June 1991.

[17] K. Bult, "Analog design in deep sub-micron CMOS," in *Proc. Eur. Solid-State Circuits Conf.*, Sept. 2000, pp. 11–17.

[18] M. A. Maher and C. A. Mead, "A physical charge-controlled model for the MOS transistors," in *Advanced Research in VLSI, 1987 Stanford Conference*, P. Losleben, Ed. Cambridge, MA: MIT Press, 1987.

[19] A. I. A. Cunha, M. C. Schneider, and C. Galup-Montoro, "An explicit physical model for long-channel transistor including small-signal parameters," *Solid-State Electron.*, vol. 38, no. 11, pp. 1945–1952, 1995.

[20] ——, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1510–1519, Oct. 1998.

[21] M. Bucher, J.-M. Sallese, C. Lallement, W. Grabinski, C. C. Enz, and F. Krummenacher, "Extended charges modeling for deep submicron CMOS," in *Proc. Int. Semiconductor Device Res. Symp.*, Charlottesville, NC, Dec. 1999, pp. 397–400.

[22] M. Bucher, C. Lallement, C. Enz, F. Theodoloz, and F. Krummenacher, "Scalable G_m/I based MOSFET model," in *Proc. Int. Semiconduct. Device Res. Symp.*, Charlottesville, NC, Dec. 1997, pp. 615–618.

[23] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits Signal Processing J.*, vol. 8, pp. 83–114, July 1995.

[24] P. K. Ko, *Approaches to Scaling in Advanced MOS Device Physics*, N. G. Einspruch and G. S. Gildenblat, Eds. New York: Academic, 1989, pp. 1–37.

[25] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. New York: McGraw-Hill, 1999.

[26] C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 186–201, Feb. 2000.

[27] ——, "MOS transistor modeling issues for RF circuit design," in *Advances in Analog Circuit Design (AACD'99)*, W. Sansen, J. Huijsing, and R. van de Plassche, Eds. Norwell, MA: Kluwer, 1999, pp. 191–225.

[28] A.-S. Porret, J.-M. Sallese, and C. C. Enz, "A compact nonquasi-static extension of a charge-based MOS model," *IEEE Trans. Electron Devices*, vol. 48, pp. 1647–1654, Aug. 2001.

[29] J.-M. Sallese and A.-S. Porret, "A novel approach to charge based non quasi static model of the MOS transistor valid in all modes of operation," *Solid-State Electron.*, vol. 44, no. 6, pp. 887–894, June 2000.

[30] D. E. Ward and R. W. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE J. Solid-State Circuits*, vol. SC-13, pp. 703–707, Oct. 1978.

[31] E. Abou-Allam and T. Manku, "A small-signal MOSFET model for radio frequency IC applications," *IEEE Trans. Computer-Aided Design*, vol. 16, pp. 437–447, May 1997.

[32] L. F. Tiemeijer, L. M. F. de Maaijer, and R. van Langevelde, "RF CMOS modeling," in *Advances in Analog Circuit Design (AACD'99)*, W. Sansen, J. Huijsing, and R. van de Plassche, Eds. Norwell, MA: Kluwer, 1999, pp. 129–149.

[33] P. Klein, "A compact-charge LDD-MOSFET model," *IEEE Trans. Electron Devices*, vol. 44, pp. 1483–1490, Sept. 1997.

[34] C. E. Biber, M. L. Schmatz, T. Morf, U. Lott, and W. Bachtold, "A nonlinear microwave MOSFET model for spice simulators," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 604–610, May 1998.

[35] C. E. Biber, "Microwave modeling and circuit design with sub-micron CMOS technologies," Ph.D. dissertation, Swiss Fed. Inst. Technol., Zurich, Switzerland, 1998.

[36] W. Liu, R. Gharpurey, M. C. Chang, U. Erdogan, R. Aggarwal, and J. P. Mattia, "R.F.MOSFET modeling accounting for distributed substrate and channel resistances with emphasis on the BSIM3v3 SPICE model," in *Proc. Int. Electron Devices Meeting*, Dec. 1997, pp. 309–312.

[37] D. R. Pehlke, M. Schroter, A. Burstein, M. Matloubian, and M. F. Chang, "High-frequency application of MOS compact models and their development for scalable RF MOS libraries," in *Proc. IEEE Custom Integrated Circuits Conf.*, May 1998, pp. 219–222.

[38] L. F. Tiemeijer and D. B. M. Klassen, "Geometry scaling of the substrate loss of RF MOSFETs," in *Proc. Eur. Solid-State Dev. Res. Conf.*, Sept. 1998, pp. 481–483.

[39] S. F. Tin and K. Mayaram, "Substrate network modeling for CMOS RF circuit simulation," in *Proc. IEEE Custom Integrated Circuits Conf.*, May 1999, pp. 583–586.
[40] Y. Cheng and M. Matloubian, "On the high-frequency characteristics of substrate resistance in RF MOSFETs," *IEEE Electron Device Lett.*, vol. 21, pp. 604–606, Dec. 2000.
[41] H. Hjelmgren and A. Litwin, "Small-signal substrate resistance effects in RF CMOS identified through device simulation," *IEEE Trans. Electron Devices*, vol. 48, pp. 397–399, Feb. 2001.
[42] S. H. Jen, C. Enz, D. R. Pehlke, M. Schroter, and B. J. Sheu, "Accurate MOS transistor modeling and parameter extraction valid up to 10-GHz," in *Proc. Eur. Solid-State Dev. Res. Conf.*, Bordeaux, France, Sept. 1998, pp. 484–487.
[43] ——, "Accurate MOS transistor modeling and parameter extraction valid up to 10-GHz," *IEEE Trans. Electron Devices*, vol. 46, pp. 2217–2227, Nov. 1999.
[44] A. van der Ziel, *Noise in Solid-State Devices and Circuits*. New York: Wiley, 1986.
[45] P. Klein, "An analytical thermal noise model of deep-submicron MOSFET's," *IEEE Electron Device Lett.*, vol. 20, pp. 399–401, Aug. 1999.
[46] A. Abidi, "High-frequency noise measurements on FET's with small dimensions," *IEEE Trans. Electron Devices*, vol. ED-33, pp. 1801–1805, Nov. 1986.
[47] A. J. Scholten, H. J. Tromp, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, and P. W. H. de Vreede, "Accurate thermal noise model for deep-submicron CMOS," in *Proc. Int. Electron Devices Meeting*, Dec. 1999, pp. 155–158.
[48] G. Knoblinger, P. Klein, and M. Tiebout, "A new model for thermal channel noise of deep-submicron MOSFET's and its application in RF-CMOS design," *IEEE J. Solid-State Circuits*, pp. 831–837, May 2001.
[49] D. P. Triantis and A. N. Birbas, "Thermal noise modeling for short-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, pp. 1950–1955, Nov. 1996.
[50] J.-S. Goo, C.-H. Choi, F. Danneville, E. Morifuji, H. S. Momose, Z. Yu, H. Iwai, T. H. Lee, and R. W. Dutton, "An accurate and efficient high frequency noise simulation technique for deep submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 47, pp. 2410–2419, Dec. 2000.
[51] J.-S. Goo, C.-H. Choi, A. Abramo, J.-G. Ahn, Z. Yu, T. H. Lee, and R. W. Dutton, "Physical origin of the excess thermal noise in short channel MOSFETs," *IEEE Electron Device Lett.*, vol. 22, pp. 101–103, Feb. 2001.
[52] S. Donati, M. A. Alam, K. S. Krisch, S. Martin, M. R. Pinto, and H. H. Vuong, "Physics-based RF noise modeling of submicron MOSFETs," in *Proc. Int. Electron Devices Meeting*, Dec. 1998, pp. 81–84.
[53] T. Manku, "RF simulations and physics of the channel noise parameters within MOS transistors," in *Proc. IEEE Custom Integrated Circuits Conf.*, May 1999, pp. 369–372.
[54] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS low-noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745–759, May 1997.
[55] W. Liu, *MOSFET Models for SPICE Simulation, Including BSIM3v3 and BSIM4*, 1st ed. New York: Wiley, 2001.
[56] L.-J. Pu and Y. Tsividis, "Small-signal parameters and thermal noise of the four-terminal MOSFET in nonquasi-static operation," *Solid-State Electron.*, vol. 33, p. 521, 1990.
[57] C. Chen and M. J. Deen, "High frequency noise of MOSFET's I: Modeling," *Solid-State Electron.*, vol. 42, pp. 2069–2081, Nov. 1998.
[58] C. Chen, M. J. Deen, Z. X. Yan, M. Schroter, and C. Enz, "High frequency noise of MOSFET's II: Experiments," *Solid-State Electron.*, vol. 42, pp. 2083–2092, Nov. 1998.
[59] R. van Langevelde *et al.*, "RF-distortion in deep-submicron CMOS technologies," in *Proc. Int. Electron Devices Meeting*, Dec. 2000, pp. 807–810.

Christian C. Enz (M'84) received the M.S. and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1984 and 1989, respectively.

From 1984 to 1989, he was a Research Assistant at the EPFL, where he was involved in the field of micropower analog CMOS integrated circuits (IC) design. In 1989, he was one of the founders of Smart Silicon Systems S.A. (S3), where he developed several low-noise and low-power ICs, mainly for high-energy physics application. From 1992 to 1997, he was an Assistant Professor at the EPFL, where he was involved in the field of low-power analog CMOS and BiCMOS IC design and device modeling. From 1997 to 1999, he was Principal Senior Engineer at Conexant (formerly Rockwell Semiconductor Systems), Newport Beach, CA, where he was responsible for the modeling and characterization of MOS transistors for the design of RF CMOS circuits. In 1999, he joined the Swiss Center for Electronics and Microtechnology (CSEM), Neuchatel, Switzerland, where he headed the RF and Analog IC design group and was promoted to Executive Vice President heading the Microelectronics Department. He is also an Adjunct Professor at the EPFL. His technical interests and expertise are in the field of very low-power analog and RF IC design and MOS transistor modeling. He has authored and co-authored over 90 scientific papers and has contributed to numerous conference presentations and advanced engineering courses.