Design methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model

A. Mangla a,⁎, M.-A. Chalkiadaki a, F. Fadhuile b, T. Taris b, Y. Deval b, C.C. Enz a,c

a École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
b IMS-Bordeaux, France
c Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland

1. Introduction

Analog circuit design is a complex exercise involving multiple tradeoffs, e.g. between power consumption and speed, and several degrees of freedom like the drain current and transistor dimensions. Design with advanced deep-submicron technology nodes is further complicated because of the short-channel effects [1]. The importance of accurate device models which stems from the fact that analog design has special modeling needs has long been recognized [2], and all the more emphasized with the aggressive scaling of the CMOS technology and the short-channel effects becoming dominant.

The requirement that a model be accurate as well as computationally fast makes the development of advanced compact models quite complex. The modern circuit simulators coupled with sophisticated compact models are powerful tools for the design and analysis of circuits with advanced technology nodes, yet most experienced analog designers would still rely on their design intuition and do ‘hand calculations’ before performing any simulation. For this pre-simulation phase the designer should adopt a charge based method, it is more physics oriented; it is not only more accurate than its predecessors but can also be faster than the surface potential based models [3,4]. Even though the full compact model includes most of the real-device effects that are present in the advanced bulk CMOS technologies, the core model is based on simple analytical equations which are continuous in all regions of operation. This exempts the designer to use asymptotic approximations that may be valid in weak and strong inversion regions but are inaccurate in moderate inversion (the preferred operating region for low-power analog and RF circuits [5]) and do not account for short-channel effects.

The charge based BSIM6 model would extend elegantly to the existing design methodologies too. The inversion coefficient (IC) based design methodology [6] that has been extensively used in conjunction with the charge based EKV model [7] for the design of low power analog and RF circuits [8,5] can now also be used with the new BSIM6 model. It will be shown in Section 2 that the charge based methodology and its application to the next generation BSIM6 compact MOSFET model.

2. Charge based analysis

2.1. Drain current and transconductance

The basic charge equation in a MOSFET is given by [7]

\[2q_i \ln q_i = \nu_p - \nu. \] (1)

This work is partially funded by the Swiss National Science Foundation Grant no. 200021-127241 and ENIAC project MIRANDELA.

⁎ Corresponding author. Tel.: +41 21 693 4 607.
E-mail address: anurag.mangla@epfl.ch (A. Mangla).

Please cite this article as: A. Mangla, et al., Design methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.02.022
where \( q_i \) is the normalized inversion charge density
\( (q_i = Q_i/Q_{\text{spec}}) \), normalized using specific charge \( Q_{\text{spec}} \).
\[ Q_{\text{spec}} = -2nU_tC_{\text{ox}}. \]  
(2)

\( V_p \) and \( V_d \) are the pinch-off and channel voltages respectively, normalized to the thermal voltage \( U_T(\approx kT/q) \). The normalized drain current, normalized using the specific current \( I_{\text{spec}} \) [7]
\[ I_{\text{spec}} = I_{\text{spec}} C_{\text{ox}} W / L \]  
(3a)

\[ I_{\text{spec}} = 2n\mu_0 C_{\text{ox}} U_T^2 \]  
(3b)

where \( \mu_0 \) is the constant low-field mobility, is given by [9]
\[ \mu_0 = \frac{ue}{C_{\text{ox}} L} \]  
(4)

In (4) \( u \) is the effective mobility normalized to the low-field surface mobility \( \mu_{\text{spec}} \), \( e \) is the longitudinal electric field normalized to the critical electric field \( E_c \) and \( \lambda_c \) is defined as
\[ \lambda_c = \frac{2U_T}{E_c} = \frac{2\mu_0 U_T}{\lambda_{\text{sat}} L} \]  
(5)

where \( \lambda_{\text{sat}} \) is the saturation velocity.

The normalized source transconductance
\[ g_{ms} = \frac{G_{ms}}{I_{\text{spec}}} = \frac{g_m n_{Gm}}{I_{\text{spec sat}} \text{saturation}} \]  
(6)

and \( G_m \) and \( n \) being the total gate transconductance and the slope factor, respectively, can then be calculated as
\[ g_{ms} = \frac{\lambda_c \lambda_e q_i}{\lambda_{\text{sat}} E_c}, \]  
(8)

\( q_i \) being the value of \( q_i \) at the source.

2.2. Inversion coefficient

The inversion coefficient \( IC \) is the measure of the level of channel inversion and is equal to the drain current in forward saturation:
\[ IC = \frac{I_d}{I_{\text{spec}}} = \left. I_d \right|_{\text{saturation}} = I_{\text{dsat}}. \]  
(9)

The different regions of operation in saturation can then be defined as
\( IC \leq 0.1 \) weak inversion
\( 0.1 < IC \leq 10 \) moderate inversion
\( 10 < IC \) strong inversion.

2.3. Short channel effects

As discussed in Section 1, the short-channel effects become prominent at sub-micron channel lengths. Two of the especially important effects are drain induced barrier lowering (DIBL) and velocity saturation (VS). While DIBL pertains to weak inversion, velocity saturation is dominant at higher values of \( IC \). In addition, there is mobility reduction due to the vertical field (MRVF). To maintain sufficient accuracy these effects should be accounted for even in the simplistic model for hand calculations. Both velocity saturation and mobility reduction due to vertical field are related to the mobility of the carriers and affect not only the magnitude but also the slopes of \( I_d \) vs. \( V_C \) and \( G_m/I_d \) vs. \( I_C \) curves [7].

To include the velocity saturation effect we choose the simple piecewise linear velocity-field model for its simplicity [7], which gives the normalized effective mobility as
\[ u_{\text{eff}} = \frac{\mu_{\text{eff}}}{\mu} = \begin{cases} 1 & \text{for } e < 1 \\ 1/e & \text{for } e \geq 1. \end{cases} \]  
(10)

where \( \mu_{\text{eff}} \) includes the mobility reduction due to the vertical electric field and is related to the low field mobility \( \mu_0 \) by [7]
\[ \frac{\mu_{\text{eff}}}{\mu_0} = \frac{1}{1 + \theta(q_b + q_i/2)} = \frac{1}{k_1 q_i + k_2} \]  
for \( e < 1, \)  
(11)

where
\[ \theta = \frac{Q_{\text{spec}}}{C_{\text{ox}} U_T^2} \]  
(12a)

\[ k_1 = \theta \left( 2 + \frac{1}{2} \frac{\psi_p}{\lambda_b} \right), \]  
(12b)

\[ k_2 = \frac{1}{1 + \frac{\theta \psi_p}{\lambda_b} \psi_p} \]  
(12c)

\( E_b \) is the electric field intensity at which the mobility starts to decrease significantly, \( \lambda_b \) is the normalized body factor and \( \psi_p \) is the normalized pinch-off surface potential (corresponding to \( q_i = 0 \)). In (4), \( u \) is the effective mobility normalized to low-field surface mobility:
\[ u = \frac{u_{\text{eff}}}{\mu_0} = \frac{\mu_{\text{eff}}}{\mu_0} \frac{\mu}{\mu_0} \]  
(13)

Therefore, from (10) and (11),
\[ u = \begin{cases} \frac{1}{k_1 q_i + k_2} & \text{for } e < 1 \\ 1/e & \text{for } e \geq 1. \end{cases} \]  
(14)

DIBL is modeled as a shift in the threshold voltage [7] which can simply be accounted for in (1) as a shift in \( V_p \).

2.4. Effect of short channel on drain current and transconductance

Following the derivation in [9] for velocity saturation and in [7] for MRFV, we obtain the drain current in saturation
\[ I_{\text{dsat}} = \frac{4(q_i + q_i^2)}{2 + \lambda_{ck} \sqrt{4(1 + \lambda_{ck}) + \lambda_{ck}^2 (1 + 2q_i)^2}}, \]  
(15)

where
\[ \lambda_{ck} = \frac{\lambda_c}{k_2} \]  
(16)

accounts for mobility reduction due to vertical field and is slightly bias dependent because of the gate bias dependent parameter \( k_2 \) given by (12c).

The normalized source transconductance can then be calculated as
\[ g_{ms} = \frac{2q_i}{\sqrt{4(1 + \lambda_{ck}) + \lambda_{ck}^2 (1 + 2q_i)^2}}, \]  
(17)

equivalently expressed in terms of \( i_d \) (\( = I_{\text{dsat}} = IC \)) as
\[ g_{ms} = \frac{\sqrt{I_d^2 + 4I_d \lambda_{ck} + 4I_d + 1}}{I_d \lambda_{ck} + \lambda_{ck} + 2} \]  
(18)
which gives
\[ g_{m2} = \frac{g_{m1}}{\lambda_d} = \sqrt{\frac{2g_{ck}^2 + 4g_{ck} + 1}{g_{ck}^2 + 2g_{ck} + 2d}}. \]

Velocity saturation and mobility reduction due to vertical field slightly impact the drain current \( I_D \) and gate transconductance \( G_m \) even in weak inversion [9]. Nevertheless, the transconductance remains proportional to the drain current in weak inversion and hence \( G_m/I_D \) remains unaffected in this region.

### 2.5. Comparison of the simple analytical model with BSIM6 and measurements

Fig. 1 shows the transfer characteristics (variation of the normalized drain current with normalized gate over-drive voltage) for three different channel lengths, viz. 40, 60 and 80 nm. The analytical curves are nothing but a plot of (15) in which \( \lambda_d \) is eliminated using (1). Each channel length corresponds to a different value of the parameter \( \lambda_d \). Measurements on NMOS transistors fabricated in a standard bulk 40 nm CMOS technology as well as the BSIM6 fitting results at the same channel lengths (shown on the same figure) compare very well with the simple model.

Eq. (19) for \( g_{m2}/I_D \) is plotted in Fig. 2(a) as a function of \( IC \) with \( \lambda_c = 0.45 \). The normalized \( G_m/I_D \) \( g_{m2}/I_D \), is equal to 1 in weak inversion and decreases on moving towards strong inversion. In strong inversion it varies as \( 1/\sqrt{I_C} \) if no velocity saturation is present, otherwise it tends asymptotically to \( 1/(\lambda_c I_C) \). The parameter \( \lambda_c \), called the velocity saturation factor and defined in (5), is a measure of velocity saturation effect and hence is an important parameter for short-channel devices that have strong velocity saturation effects. For a given technology it depends only on the channel length; the shorter the channel, the higher the value of \( \lambda_c \) and the higher the velocity saturation effect. In Fig. 2(a) the asymptote \( 1/(\lambda_c I_C) \) intersects the weak inversion asymptote \( g_{m2}/I_D = 1 \) at \( IC = 1/\lambda_c \), which can be interpreted as the demarcation between weak inversion and velocity saturated strong inversion region.

Fig. 2(b) shows a comparison of the analytical plots of the normalized \( G_m/I_D \) with measurements and BSIM6. It can be seen that the simple analytical equation (15) is valid from weak to strong inversion regions and is reasonably close to the measurements and the BSIM6 model.

It should be re-emphasized that the purpose of this simple analytical model is to provide initial design guidelines and not to replace a numerically accurate full compact model like BSIM6.\(^1\) Of course, the initial values should be reasonably accurate to justify their use. As an example, the relative error in \( G_m/I_D \) between the simple analytical model and the measurements is limited to 10\%, as can be seen in the inset in Fig. 2(b).

It can be observed in Fig. 2(b) that for the 40 nm device, \( g_{m2}/I_D \) starts degrading as \( 1/(\lambda_c I_C) \) already around \( IC = 3 \) and that there is practically no region with \( 1/\sqrt{I_C} \) dependence. For shorter devices this point would only move towards \( IC = 1 \) or lower, implying a strong degradation of transconductance efficiency (due to velocity saturation) already at low biases. Evidently, to avoid this degradation the bias points for short-channel devices would need to be pushed from strong inversion towards moderate and eventually weak inversion.

### 3. Simple example of \( IC \) based design methodology

#### 3.1. Current optimization in a common-source stage

Following [11], the goal is to optimize a capacitively loaded common-source stage, as shown in Fig. 3, to minimize its current consumption for a desired gain \( A_v \) at a given frequency of operation \( \omega \), which is assumed to be higher than the cut-off frequency \( \omega_c \), i.e., \( \omega > \omega_c \).

The small-signal voltage gain of the circuit (loaded by an external load capacitance \( C_L \)) can then be approximated as

\[ A_v \approx \frac{\Delta V_{out}}{\Delta V_{in}} = \frac{G_m}{\omega C_L}. \]

---

\(^{1}\) The interested reader is referred to [10] for an overview of the benchmarking of the scaling properties of BSIM6 against 40 nm CMOS technology.
where
\[ C_{\text{out}} = C_w W + C_L. \]  
Here \( C_w \) is the self-loading capacitance per unit width of the gain transistor and \( W \) its width. At a given length \( L \) (usually taken as the minimum provided by the technology), the bias current through the transistor is
\[ I_b \propto I_C \cdot W. \]  
Solving for \( W \) and normalizing, we obtain (see Appendix A for the detailed derivation)
\[ w = \frac{\Omega}{g_{\text{ms}} - \Omega \ell} \cdot \ell; \]  
\[ I_b = \frac{\Omega}{g_{\text{ms}} - \Omega \ell} I_C \]  
where,
\[ \Omega = \frac{\omega}{\omega_0} \ell \]  
\[ \ell = \frac{L}{I_0} \]  
\[ \omega_0 = \frac{I_{\text{spec}}}{n U T C_{\text{SS}} L_0} \]  
and \( L_0 \) is the minimum channel length provided by the chosen technology. In Fig. 4, \( I_b \) is plotted as a function of \( I_C \) for three different values of \( \Omega \) at \( L = 40 \) nm \( (\ell = 1) \), using (18) for \( g_{\text{ms}} \). The plots of \( I_b \) without including velocity saturation \( (\ell_c = 0 \text{ in (18))} \) are also shown for comparison. We can observe that, to obtain a given \( \Omega \), higher bias current is needed in the device with velocity saturation than for the one without. This is especially prominent for high \( \Omega \) and in strong inversion.

From Fig. 4, it is evident that the bias current \( I_b \) of the common-source stage can be minimized for a given gain and operating frequency, i.e. at a given \( \Omega \). For a target \( \Omega \) at a given channel length (fixing \( \ell_c \) and \( \ell \)), the optimum inversion coefficient for achieving minimum bias current \( (I_{\text{opt}}) \) can be obtained by minimizing (24) numerically and plotting the curve as in Fig. 4. Interestingly, this optimum lies in the moderate or the weak inversion region. Having obtained \( I_{\text{opt}} \) we can calculate the required width \( W_{\text{opt}} \) by plugging-in \( I_{\text{opt}} \) for \( I_b \) in (18) and using in (23). Finally, the required optimum bias current can be calculated as
\[ I_{b_{\text{opt}}} = I_{\text{spec}} \frac{W_{\text{opt}}}{L} I_{\text{opt}}. \]

### 3.2. Comparison with BSIM6

To verify the design methodology outlined above, we attempt the design of a common-source stage loaded with another common-source transistor of width 100 \( \mu \)m (implying a load capacitance \( C_L = 18.5 \) fF). We target a gain \( A_v = 15 \) dB and a working frequency \( f = 24 \) GHz, giving \( \Omega = 0.83 \). Choosing the minimum channel length, 40 nm \( (\ell = 1) \), provided by the 40 nm standard CMOS technology, we obtain the optimum value of \( I_C, I_{\text{opt}} = 6.3 \), that is in moderate inversion. The normalized optimum width \( W_{\text{opt}} = 1.41 \) is calculated using (23) and the normalized bias current \( I_{b_{\text{opt}}} = 8.78 \) is calculated using (24). Knowing the values of technology parameters we can calculate the optimum bias current \( I_{b_{\text{opt}}} \) and width \( W_{\text{opt}} \). Eventually, plugging the values of \( W_{\text{opt}} \) and \( I_{b_{\text{opt}}} \) for simulating with the BSIM6 model, we obtain a gain of 14 dB at 24 GHz, which is close to the target gain of 15 dB.

### 4. Low-power RF design

A more practical requirement for low-power RF design would be not only to minimize the consumption to obtain a given gain but also to minimize the noise. Maximizing the gain, minimizing the noise and minimizing the consumption involves a mutual tradeoff. To quantify this tradeoff we can introduce a figure-of-merit and then optimize this FoM (rather than dealing with the gain, noise and current individually), as we show in the following.

Consider a common-source stage loaded with an identical transistor such that it presents a load \( C_{\text{SS}} \) to the first stage (see Fig. 5), we can write
\[ A_v = \frac{g_m}{\omega C_{\text{SS}}} = \frac{\omega_1}{\omega} \]  
where
\[ \omega_1 \approx \frac{g_m}{C_{\text{SS}}} \]  
is approximately equal to the transit frequency. The noise factor is given by [7]
\[ F \approx 1 + \frac{\gamma_{\text{nd}}}{g_m R_S} \]  
where \( \gamma_{\text{nd}} \) is the excess noise factor of the gain transistor and \( R_S \) is the source resistance. In order to maximize the gain \( A_v \) and minimize the noise factor \( F \) at a given bias current \( I_0 \), we can define a FoM as
\[ \text{FoM}_{\text{GainNoise}} \triangleq \frac{A_v}{(f - 1) I_0} \frac{R_S}{\gamma_{\text{nd}}} \frac{g_m \omega_1}{I_0} \]  
\[ \text{FoM}_{\text{GainNoise}} \] is proportional to an important FoM for low-power RF

---

Please cite this article as: A. Mangla, et al., Design methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.02.022
This work presents a higher current consumption. Significant gain at RF, better noise and linearity, albeit at the cost of a higher current consumption. Devices are biased in strong inversion to obtain the small-signal current gain of the transistor in common-source configuration defined as the frequency at which the extrapolated small-signal calculations. Because $C_m/I_D$ is maximum in weak inversion, it is advantageous to bias the device towards this region from the point of view of dc gain and power consumption. However, the linearity and noise of the device degrade in this region. The transistor frequency $f_{T}$ is defined as the frequency at which the extrapolated small-signal current gain of the transistor in common-source configuration falls to unity. Devices are biased in strong inversion to obtain significant gain at RF, better noise and linearity, albeit at the cost of a higher current consumption. $F_{oM_{RF}}$, which combines these tradeoffs into a single metric, was introduced in [12] and discussed further in [13]. Since it combines the two quantities which have their maxima in opposite directions of the $IC$ axis, it is particularly suited for low-power RF design, serving as a tool to locate the optimum operating point ($IC$) of the device. This optimum can be rapidly located by plotting the normalized $F_{oM}$ as a function of $IC$ using (18) and (28), as shown in Fig. 6. A comparison with measurements and BSIM6 simulations on 40, 60 and 80 nm, shown on the same figure, shows a reasonable match. Interestingly, for short channel devices, this FoM has a maximum in moderate inversion, indicating that the optimum operating point lies in this region. This corroborates the advantageous use of moderate inversion for low-power RF circuits as already mentioned in Section 1. The use of this FoM for RF design in moderate inversion region has been demonstrated in [13].

5. Conclusion

Accurate MOSFET compact models like BSIM6, integrated with sophisticated circuit simulators, place powerful tools in the hands of analog circuit designers to design circuits using the state-of-the-art technology nodes. However, most analog design cycles begin with simple hand calculations where tradeoffs (like that between power consumption, gain, operating frequency and noise) are evaluated and first-cut approximations about the operating points and dimensions of the transistors are made. Simple, yet sufficiently accurate, models are needed at this stage of the design cycle.

This paper summarized a set of equations and figures-of-merit especially suitable for low-power analog and RF design derived from the charge based core of the BSIM6 model, that are simple enough for use with basic numerical tools, yet accurate enough to yield a good first-order approximation of the operating point, viz., inversion coefficient $IC$. It was shown that the optimum operating point for low-power RF circuits lies in the moderate inversion region emphasizing the importance of accuracy of the model in this region. A comparison with the BSIM6 compact model and measurements from NMOS devices fabricated with 40 nm standard bulk CMOS technology was also presented. The comparison points out to the advantage of the charge based approach adopted in BSIM6: it can be used both as a full featured compact model and as a pre-simulation design methodology.

Acknowledgments

The authors would like to thank Drs. Patrick Sheer, Didier Belot and André Juge of ST Microelectronics, Crolles, France, for the measurements used in this paper, and the BSIM group at UC Berkeley, USA, for providing the beta version of the BSIM6 bulk MOSFET model.

Appendix A

The small-signal voltage gain of the capacitively loaded common-source transistor shown in Fig. 3 is given by

$$A_v = \frac{G_m}{\omega C_{GS}}.$$  \hspace{1cm} (A.1)

Therefore

$$C_{tot} = \frac{C_m}{\omega A_v} = \frac{g_m G_{spec}}{\omega A_v} = \frac{g_m I_{spec}(W/L)}{nU_T \omega A_v}.$$  \hspace{1cm} (A.2)

$C_{GS}$ is the total load capacitance seen by the gain transistor and is the sum of the external load capacitance $C_L$ and the load capacitance of the gain transistor itself ($C_u$ per unit width):

$$C_{tot} = C_u W + C_L.$$  \hspace{1cm} (A.3)

Let us define a constant

$$\omega_0 = \frac{I_{spec}}{nU_T C_u L_0}$$  \hspace{1cm} (A.4)

where $L_0$ is the minimum channel length provided by the chosen technology. Thus $\omega_0$ is a parameter dependent entirely on the design cycle.

Please cite this article as: A. Mangla, et al., Design methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.02.022
technology and temperature and independent of the operating point and frequency. Solving (A.2) and (A.3) for $W$ and substituting using (A.4), we get

$$W = \frac{C_L}{C_W} \frac{\omega A_v \frac{L}{I_0}}{g_{ms} - \Omega \ell}.$$  \hspace{1cm} (A.5)

Rewriting (A.5) in terms of normalized variables, we obtain

$$w = \frac{W}{W_0} = \frac{\Omega}{g_{ms} - \Omega \ell} \ell$$  \hspace{1cm} (A.6)

where

$$W_0 = \frac{C_L}{C_W}$$  \hspace{1cm} (A.7a)

$$\Omega = \frac{\omega}{\omega_0} A_v$$  \hspace{1cm} (A.7b)

$$\ell = \frac{L}{I_0}.$$  \hspace{1cm} (A.7c)

Now the bias current is then given by

$$I_{db} = I_{spec} \frac{W}{I_0} = I_{spec} \frac{W_0}{I_0} \frac{\Omega}{g_{ms} - \Omega \ell} \ell \frac{I_0}{I_{spec}}.$$  \hspace{1cm} (A.8)

In the normalized form

$$i_{db} = \frac{I_{db}}{I_0} = \frac{\Omega}{g_{ms} - \Omega \ell} \frac{I_0}{I_{spec}}.$$  \hspace{1cm} (A.9)

where,

$$I_{C_0} = I_{C_{spec}} \frac{W_0}{I_0}.$$  \hspace{1cm} (A.10)

It should be remembered that $g_{ms}$ in (A.6) and (A.8) is a function of $I_C$ through (18).

References