Annals of the Academy of Romanian Scientists
Series on Science and Technology of Information
ISSN 2066 - 8562 Volume **2**, Number **1**/**2009**

AN ACCURATE ANALYTICAL MODEL WITH A REDUCED PARAMETER SET FOR SHORT-CHANNEL MOS TRANSISTORS

Andrei SEVCENCO[1], Gheorghe BREZEANU[2]

Rezumat. *Este prezentat un model analitic compact, bazat pe numai câțiva parametri, pentru caracteristicile electrice ale tranzistorului MOS submicronic pentru a fi utilizat în proiectarea analogă. Efecte secundare, ca viteza de saturare și modularea lungimii canalului sunt incluse atât pentru dispozitivele nMOS cât și pentru cele pMOS. Sunt propuși noi parametri pentru descrierea vitezei efectului de saturație. Este demonstrată o aproximare liniară a dependenței vitezei purtătorilor de sarcină de câmpul electric. Este realizată o bună concordanță între model și datele experimentale pentru caracteristicile de transfer și de ieșire și pentru transconductanță, măsurate pe tranzistoare nMOS și pMOS standard (0,18 μm și 0,35 μm).*

Abstract. *A compact analytical model, based only on a few parameters, for the electrical characteristics of submicron MOS transistors to be used in analogue design is presented. Secondary effects like velocity saturation and channel length modulation are included in the model for both nMOS and pMOS devices. New parameters are proposed for the description of the velocity saturation effect. A linear approach of the dependence of the carrier velocity on the electric field is proved. A very good agreement between model and experimental data is achieved for transfer, output characteristics and transconductance measured on standard 0.18 and 0.35μm nMOS and pMOS transistors.*

Key words: MOS transistor, short channel, analytical model

1. Introduction

Technology advances have brought the microelectronic industry to never before seen levels of efficiency in production of ICs. This is a feat supported mainly by the extreme scaling of the most basic cell of electronic devices, the MOS transistor. This has not always been done in the best way needed. Since customer ease of use has always been a primordial condition for good business, in the beginning of the microelectronic industry as we know it today MOS devices were scaled through the constant voltage technique. This was due to TTL compatibility needs of the, then current, IC market. This in turn created very high fields in the channel of the MOS transistor, as channel lengths were ever smaller while voltages applied to these channels were always constant.

[1]University ”Politehnica” of Bucharest, Romania, andrei.sevcenco@catsemi.ro.
[2]University ”Politehnica” of Bucharest, Romania, gheorghe.brezeanu@dce.pub.ro.

After reaching a point where secondary effects caused by these very high electrical fields could no longer be tolerated and because it was clear that bipolar technologies would never win against MOS transistors, the constant field scaling technique took hold. This brought about a real benefit for the IC business since now the scaling, and, consequently, the efficiency of production could be taken to ever increasing heights.

After years of carefree business making and profiting from the seemingly "unending" capability of silicon to absorb our needs, it has become clearer year by year that this has brought us to a turning point. Constant field scaling has taken supply voltages so low that we are nearing the limits of our science.

While this has been true and ominous for quite some time in the digital world, analogue applications have gotten a stay of execution since what we need from MOS transistors used in most analogue devices is amplification. And since many analogue applications do not call for extremely complicated designs, design for wafer space efficiency has not been such a problem that scaling had to be taken to extremes. If we add the fact that the dynamic parameters of MOS transistors cannot be precisely guaranteed when considering using MOSFETs in highly advanced technologies, i.e. 65nm or so, we find that the analogue domain is still using channel sizes which digital devices have forgotten a decade ago.

Still, we have gotten to a point were even in the analogue domain the feature sizes used cause unwanted secondary effects. Short channel sizes bring about short channel effects like velocity saturation or channel-length modulation, which are exacerbated to a very high degree. This work focuses on an analytical MOSFET model, including these two short channel effects.

While striving for as accurate a description of MOS operation as possible, the comprehensive model presented in this paper takes into account the carrier saturation velocity for both carrier types, holes and electrons, and includes the Early effect.

A linear approximation of the quadratic expression of the electron velocity versus the electrical field in the channel is, for the first time, introduced in order to obtain an analytical solution for the *n*MOS electrical characterisitics.

Analytical equations are derived for the drain current dependence on gate and drain voltage and for the transconductance in strong inversion.

An intuitive physical model of MOS electrical behaviour which uses a small number of parameters is obtained.

The model is confirmed by comparison to the experimental data, measured on both *n*MOS and *p*MOS transistors obtained in standard (0.18, 0.35 and 1 μm) CMOS processes.

2. Velocity saturation

In analogue applications the most important short-channel effect in MOS transistors is, undoubtedly, the carrier velocity saturation.

This effect stems from the increase of the longitudinal electrical field present in the channel.

Scaling is now generally done at constant voltages and the electrical field in the channel has reached values much greater than the critical electrical field, ξ_c. Due to this fact, for submicron channels, the model with constant mobility is no longer valid.

Mobility is a measure of the velocity acquired by the carrier as a result of the electrical field. At low longitudinal electrical fields, ξ (when the drain voltage, V_{DS}, is small and/or L is large), the carrier velocity, v, increases linearly with the increase of the field and the slope is the surface mobility, μ.

At high electrical fields, this slope decreases with the increase of the field, and the velocity reaches a maximum or saturation value, called the scattering-limited velocity, v_{scl}.

A first order analytical approximation of the dependence of velocity as a function of electrical field is given by[1]-[10]

$$v = \frac{\mu\xi}{\left(1+\left(\xi/\xi_c\right)^{\alpha}\right)^{1/\alpha}} \quad (1)$$

where ξ_c is the critical field, defined as $\xi_c = v_{scl}/\mu$. ξ_c has a value of 1V/μm (1kV/cm) for electrons and 3V/μm (3kV/cm) for holes in silicon [5][7]-[10].

Hence, in an *n*MOS device with $L_{min} = 0.5\mu m$ a voltage drop of only 0.5V along the channel is needed to produce an average electrical field equal to the critical field.

For holes, at 300K, $\alpha=1$ as verified in [1] and $\alpha=2$ for *n*MOS devices [3]. This difference is visible in Fig.1, where the carrier velocity versus longitudinal electrical field is plotted for both MOS device types.

Eq.(1) is used in order to obtain the drain current expression of MOS transistors. In obtaining the solution for the drain current, the α parameter has always been considered with a value of 1 for all carrier types, even though it was experimentally proved that $\alpha=2$ for electrons. In this paper we will, for the first time, take into account the differences determined by the values of α between electrons and holes.

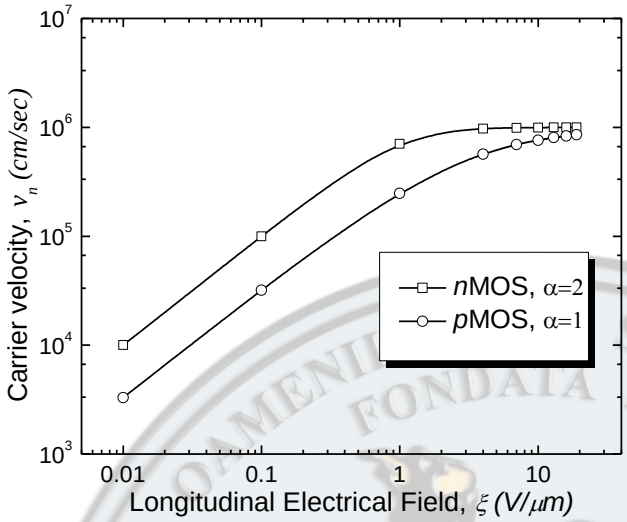


Fig. 1. Carrier velocity dependence on longitudinal electrical field.

3. The model

The velocity saturation effect was included up until now in our other papers [7]-[10]. Models for *p*MOS ($\alpha=1$) operation in weak and strong inversion were achieved.

A set of formulas for the drain current, I_D, and transconductance, g_m, based on a very reduced set of material parameters were obtained [7]-[10]. This made the model extremely easy to handle on paper, what one might call a "back-of-the-envelope" quality. This simple model is extended for *n*MOS transistors in the present paper.

To accommodate velocity saturation begin with the general equation for the drain current in strong inversion [1]

$$I_D = Wv|Q| \tag{2}$$

where Q is the induced charge in the channel and W is the channel width. At a distance y along the channel, the voltage with respect to the source is V_{cS} and the gate-to-channel voltage in this point is V_{GS}-V_{cS}. We assume that this voltage exceeds the threshold voltage V_T. Thus, the induced electron charge per unit area in the channel is [1][2]

$$|Q|(y) = C_{ox}(V_{GS} - V_T - V_{cS}(y)) \tag{3}$$

where C_{ox} is the specific oxide capacitance. The electrical field in the channel is

$$\xi = \frac{dV_{cS}}{dy}, \tag{4}$$

where dV_{cS} is the incremental voltage drop along the length of channel dy at a distance y from the source. Substituting eqs.(1), (3) and (4) in eq.(2) gives

$$I_D = W[C_{ox}(V_{GS} - V_T - V_{cS}(y))]\frac{\mu\frac{dV_{cS}}{dy}}{\left[1+\left(\frac{dV_{cS}}{dy}\Big/\xi_c\right)^{\alpha}\right]^{1/\alpha}}. \quad (5)$$

Separating the variables and integrating along the whole length of the channel gives

$$\int_0^L I_D dy\left[1+\left(\frac{dV_{cS}}{dy}\Big/\xi_c\right)^{\alpha}\right]^{1/\alpha} = \int_0^{V_{DS}} W\mu\, C_{OX}\left[V_{OV} - V_{cS}(y)\right]dV_{cS} \quad (6)$$

where $V_{OV} = V_{GS} - V_T$ is overdrive voltage.

The integral from the first term has an analytical solution for $\alpha = 1$ (*p*MOS only).

For electrons (i.e. $\alpha=2$), we propose to linearize the carrier velocity expression, since for the *p*MOS case the drain current integral proved solvable.

Therefore, we equate

$$v = \frac{\mu\xi}{\left(1+\left(\xi/\xi_c\right)^2\right)^{1/2}} = \frac{\mu\xi}{1+\left(\xi/(a\cdot\xi_c)\right)} \quad (7)$$

where a is a linearization parameter. Solving with respect to this parameter we obtain

$$a = \frac{L\xi_C}{V_{DS.m}} + \sqrt{1+\frac{(L\xi_C)^2}{V_{DS,m}{}^2}} \quad (8)$$

The longitudinal electric field along the channel has been approximated by its average value

$$\xi = V_{DS,m} / L \quad (9)$$

$V_{DS,m}$ is the extracted value of V_{DS} for which the error in using the linear expression to calculate the velocity is minimum.

This observation is confirmed by plotting the velocity versus electric field in Fig.2. A good fitting is obtained between the curves based on the quadratic expression and the linearization equation, respectively, for a 0.13μm *n*MOS device.

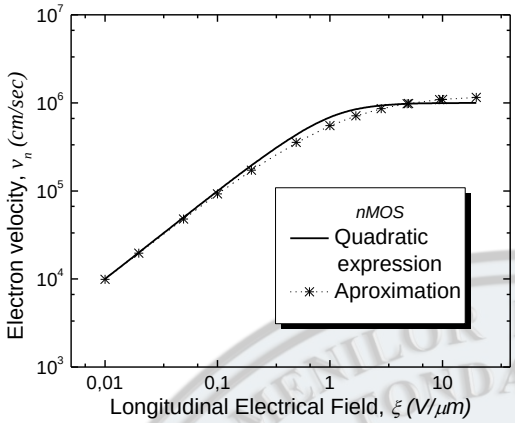


Fig. 2. Comparison of the real dependence on electrical field to linearized version of electron velocity.

In practice the accuracy of the linearization is important only in the electric field range which corresponds to the longitudinal field values from actual submicron MOS devices. While focusing on practical biasing, we observed an important increase in linearization performance for shorter channel lengths. This was obvious in the drop in linearization error from a maximum of 9% in the *0.35μm* process to a negligible 3% for the *0.1μm* technology. This increase in accuracy can be observed in Fig.3.

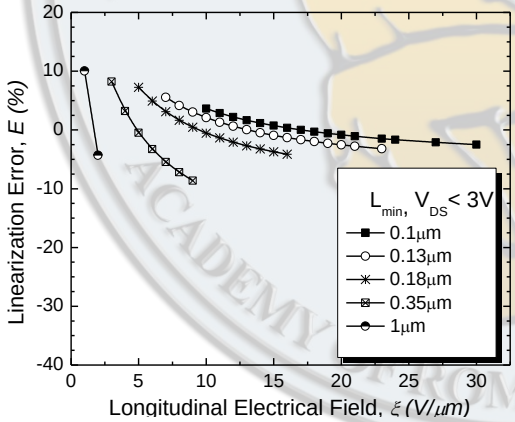


Fig.3. Resulting error in the linearization process, for various MOS technologies.

By using the general linearized expression of carrier velocity, we obtained an integrable form of eq.(6):

$$\int_0^L I_D dy = \int_0^{v_{DS}} Wk' \left[V_{OV} - V_{cS}\right] dV_{cS} - \int_0^{v_{DS}} i_D \frac{dV_{cS}}{a \cdot \xi_c} \,. \quad (10)$$

Carrying out these integrations we obtain

$$I_D = \frac{1}{1+\frac{V_{DS}}{aL\xi_c}} \frac{W}{L} k' \left[V_{OV} V_{DS} - \frac{V_{DS}^2}{2} \right]. \quad (11)$$

4. Electrical characteristics

Eq.(11) is valid in strong inversion for the triode operating region. In order to determine an expression for the drain current in the active (or saturation) operating region, let $V_{DS,act}$ be the value of V_{DS} at the boundary between triode and active regions of the transistor that sets $\frac{dI_D}{dV_{DS}} = 0$. From (11) results

$$V_{DS,act} = \frac{V_{OV}}{\frac{1}{2}\left(\sqrt{1+\frac{2(V_{GS}-V_T)}{V_{DS,c}}} + 1 \right)} \quad (12)$$

where $V_{DS,c} = aL\xi_c$ corresponds to the critical drain voltage. By substituting (12) for V_{DS} in (11) we obtain the active region expression for I_D:

$$I_D = \frac{1}{2} \frac{W}{L} k_n' V_{DS,act}^2 \quad (13)$$

Thus, we observe that the resulting general forms of the drain current in strong inversion we obtained here are almost identical to the expressions presented in the *p*MOS case in our previous papers [7]-[10].

The difference is contained in the multiplication of $V_{DS,c}$ by the linearization factor, *a*.

Eq.(13) predicts that the drain current is independent of V_{DS} in the active region, because channel- length modulation was not included. But the Early effect becomes important in submicron transistors.

Consider the effect of changes in V_{DS} on the channel length. Since drain voltage increases over $V_{DS,act}$, the pinch-off region of the channel is extended and this causes a decrease in the effective channel length, L_{eff}. A Taylor series can be used for this dependence:

$$L_{eff} = L(V_{DS,act}) + \frac{dL}{dV_{DS}}(V_{DS} - V_{DS,act}) + \ldots \cong L\left(1 - \frac{(V_{DS} - V_{DS,act})}{V_E}\right) \quad (14)$$

where

$$V_E \cong \left(-\frac{dL}{dV_{DS}} \right)^{-1} \cdot L \tag{15}$$

is called the Early voltage of the transistor, which is a constant (i.e. independent of I_D).

To find the drain current in the active region affected by the Early effect substitute L_{eff} given by (14) for L in eq.(13). The result is:

$$I_D = \frac{1}{2}\frac{W}{L_{eff}} k' V_{DS,act}^2 \cong \frac{1}{2}\frac{W}{L} k' V_{DS,act}^2 \left(1 + \frac{V_{DS} - V_{DS,act}}{V_E} \right) \tag{16}$$

The above expression evinces for the first time that the channel modulation effect is due to the effective voltage (V_{DS} - $V_{DS,act}$) which biases the pinch-off region. In classical SPICE modelling, the Early effect has always been attributed to the whole drain voltage, V_{DS} [1]-[3]. Even though this was done in order to assure absolute convergence of simulated equations, it cannot be ignored that it can be a supplemental source of errors.

5. Transconductance

The transconductance, g_m, is the measure of the transistor's amplifier function, because it reflects the transfer efficiency from input to output. Transconductance in short channel devices is modified significantly by velocity saturation. Therefore an accurate analytical description of MOS transconductance must include the velocity saturation effect.

In strong inversion, transconductance is obtained from eqs.(16) and (12) by differentiating:

$$g_m = \frac{dI_D}{dV_{GS}} = \frac{W}{L} k' \frac{V_{DS,act}}{\sqrt{1 + 2V_{OV}/V_{DS,c}}} \left(1 - \frac{V_{DS,act}}{2V_E} \right) \cong \frac{W}{L} k' \frac{V_{DS,act}}{\sqrt{1 + 2V_{OV}/V_{DS,c}}} \tag{17}$$

Rearranging with respect to the drain current, the above expression becomes

$$g_m = \frac{g_{m,c}}{1 + \sqrt{I_{D,K}/I_D}} \tag{18}$$

where

$$g_{m,c} = Wk' a \xi_c \tag{19}$$

is the saturated transconductance, caused by the velocity saturation effect. $I_{D,K}$, the knee drain current, is a new parameter of the model. The expression of this current results from eqs.(17) and (18).

$$I_{D,K} = \frac{WL}{2} k' a^2 \xi_c^2 = \frac{W}{2} a k' \xi_c \cdot V_{DS,c} \qquad (20)$$

Figs.4 plot the transconductance versus I_D and V_{OV}, respectively. At low bias, g_m is inversely proportional to the square root of the current or overdrive voltage, respectively, and the velocity saturation is not significant. When the velocity is saturated, the transconductance reaches its maximum value given by eq.(19). $g_{m,c}$ is a constant independent of I_D, V_{OV} and L. This corresponds to high longitudinal electrical fields (i.e. $\xi > \xi_c$).

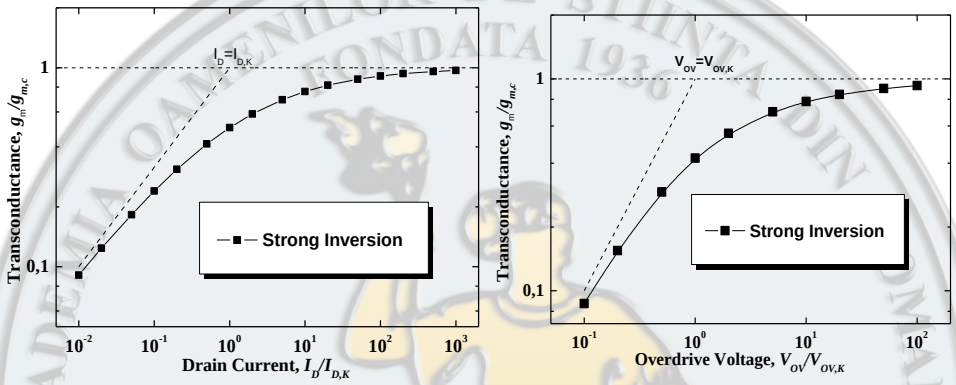


Fig. 4. MOS transconductance relative to maximum transconductance versus: (a) drain current; (b) overdrive voltage

Each of the curves from Figs.4 has two asymptotes, which correspond to the limit cases where the velocity saturation effect is insignificant and dominant, respectively. Extrapolation of these straight lines gives a flexing (knee) point where $I_D = I_{D,K}$ and $V_{OV,K} = \frac{3}{2} aL\xi_c$. Note that the V_{DS} value which corresponds to the knee current is still the critical drain voltage ($V_{DS,k} = V_{DS,c}$).

Hence, at high biases, over the knee point, the velocity saturation effect becomes predominant.

The transconductance is almost unchanged by the channel-length modulation due to $V_{DS,act} \square V_E$. (see eq.(17)).

One more dynamic parameter which must be added to the small signal model is the output conductance. It is derived from eq. (16)

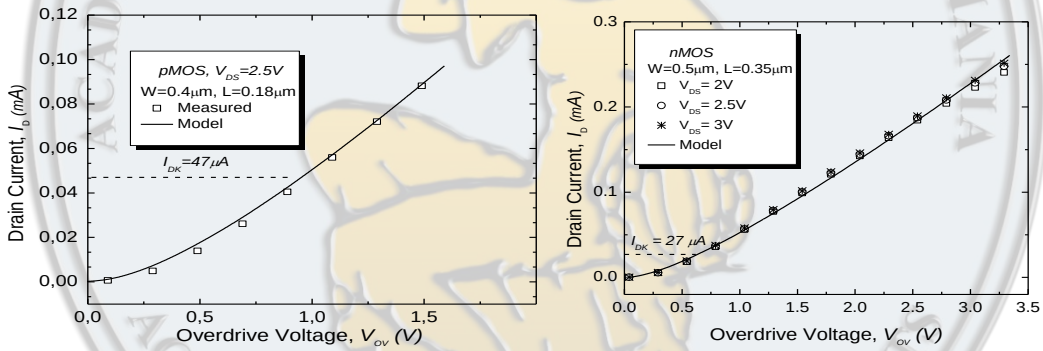
$$g_{ds} = \frac{dI_D}{dV_{DS}} \cong \frac{I_D}{V_E}. \qquad (21)$$

g_{ds} is a result of the channel-shortening effect.

6. Theory-experiment comparison

The validation of the model presented in the previous section is accomplished by comparison to experimental data measured on *p*MOS and *n*MOS transistors produced in 0.18, 0.35 and 1 μm commercial CMOS processes, which represent the classical technologies for analogue circuits. While these minimum channel sizes cannot be considered thoroughly deep submicron, it is still true that phenomena that exist in devices with feature sizes smaller than these values are governed by quantum mechanics, rather than by the classical theory. In this case, MOS device capabilities for analogue applications are diminished. However, even in submicron CMOS processes the MOS transistors in a given analogue circuit are deliberately designed to have larger channels for high performances.

Figs.5 show a comparison between the model and experimental transfer characteristics measured on a *p*MOS device realised in the 0.18μm technology and a *n*MOS device fabricated in the 0.35μm technology. The knee drain currents calculated using eq.(20) are also inserted in Figs.5. A very good agreement of the theoretical curves, described by eqs.(12) and (13), and the measured data, has been achieved.



(a) (b)

Fig.5. Drain current as a function of overdrive for: (a) pMOS transistor;(c) nMOS transistor

We can discern on these plots the two limit situations of the drain current variation. When the saturation velocity effect is negligible, from (12) and (13) results

$$I_D = \frac{1}{2} k' \frac{W}{L} V_{OV}^2 . \qquad (22)$$

This square law dependence of I_D on V_{OV} is evinced in Figs 5, for $I_D << I_{D,K}$.

To examine the limit case when the velocity is completely saturated, let $I_D >> I_{D,K}$.

From eqs.(12) and (13) we obtain

$$I_D = \frac{1}{2} k' \frac{W}{L} V_{DS,c} V_{OV} \quad (23)$$

A linear increase of drain current versus overdrive voltage can be observed at high currents when the carrier velocity is saturated (Figs.5), in agreement with eq.(23).

The measured output characteristics are compared with the theoretical curves, based on eq.(16), in Figs.6.

Only the active region is represented, where the output characteristics are linear. By extrapolating this part of the characteristics, back to the V_{DS} axis, the Early voltage can be deduced.

A good theory-experiment fit was achieved for the plotted characteristics from Figs.6.

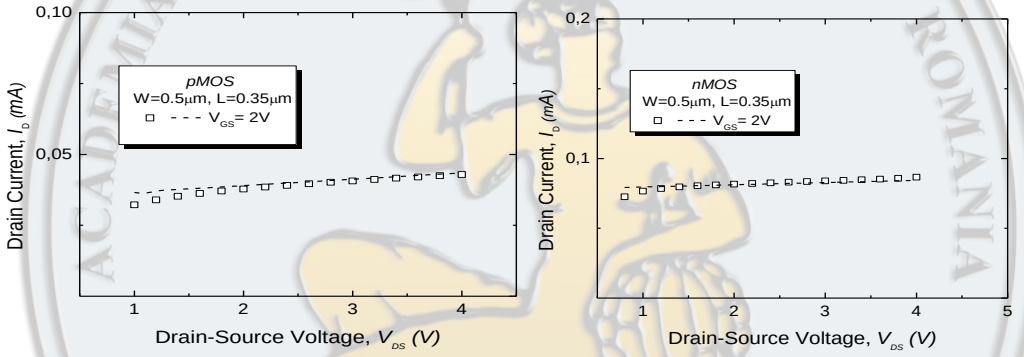


Fig.6. Drain current as a function of V_{DS} for: (a) pMOS transistor;(c) nMOS transistor

Table 1 contains values for extracted parameters from the fit between the model and the measured data, for both transfer and output characteristics.

The threshold voltage was extracted through the classical technique, using the square root of the drain current.

The current factor, *k'*, and the linearization parameter, *a*, were obtained by comparing the transfer characteristics of the measured MOSFETs in strong inversion with the theoretical curve, given by eqs.(12) and (13) (Figs.5). We can see (Table 1) that the current factor increases at low channel lengths with the same rate for both *p*MOS and *n*MOS transistors.

The value of *a* is multiplied in the table by the critical electrical field. While *p*MOS devices have a constant critical field for any CMOS process, in the *n*MOSFET's case, ξ_c is different from the classical value [1] and it decreases with shorter channel length.

The Early voltage was extracted from the output characteristics (Figs.6).

TABLE 1. Parameters extracted from the measured data for all studied devices

Parameter	**CMOS Process**					
	nMOS		***pMOS***			
L(μm)	1	0.35	1	0.35	0.18	0.18
W(μm)	25	0.5	25	0.5	2	0.4
$V_T(V)$	0.623	0.659	-0.7	-0.623	-1.42	-0.34
$k'(\mu A/V^2)$	160	210	45	60	116	100
$V_E(V)$	-353	-50.7	55.6	16.3	–	–
$a^*\xi_c(V/\mu m)$	1.74	1.22	3	3	3	3

A complete strong inversion plot for the output characteristic of a $1\mu m$ channel *n*MOS device is shown in Fig.7.

As we have seen earlier, the active region measurements agree completely with the model. Still, when considering the triode region, a noticeable overestimation of the real current in the channel is obtained with the model.

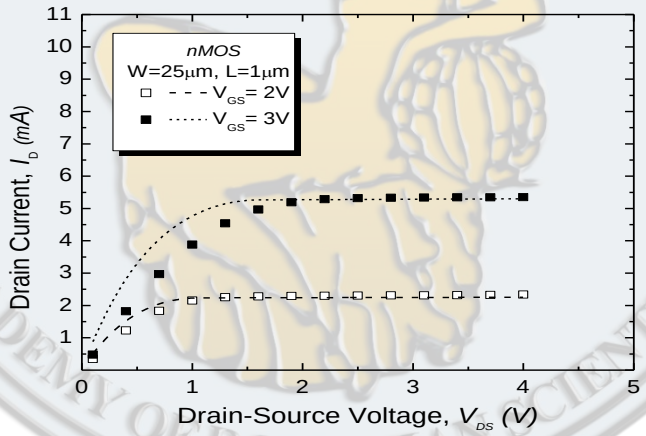
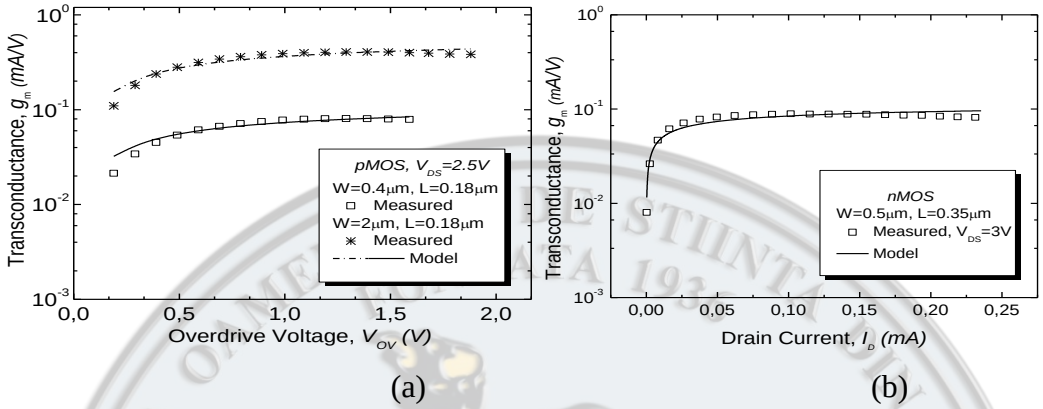


Fig.7. Complete strong inversion plot for drain current as a function of V_{DS} for an nMOS transistor

This difference, which increases at higher gate voltage (Fig.7), arises from the modulation of the carrier mobility in the channel.

Due to the high electrical fields created by the gate biasing the carrier mobility is additionaly reduced determining an increase of the importance velocity saturation effect. This fact shows the importance of the inclusion of velocity saturation in any analytical model for the MOSFET.

The transconductance curves calculated using (16) are compared with the experimental data in Figs.8. The same good theory – experiment agreement as for the transfer characteristics (Figs. 5) is obtained, for any channel width (Figs.8).



(a) (b)

Fig.8. Transconductance dependence on drain current for: (a) pMOS; (b) and (c) nMOS

6. Conclusions

An accurate analytical model with a reduced parameter set for submicron MOS transistors was presented.

The model takes into account the velocity saturation in the channel and the channel-length modulation.

Simple equations with just a few parameters were obtained for the drain current, drain voltage and transconductance for both *p*MOS and *n*MOS transistors.

For *n*MOS transistors, a linearization of the carrier velocity dependence on the longitudinal electrical field was necessary.

Several new parameters were introduced in the model in order to evince the velocity saturation importance.

The knee point corresponds to the bias which creates the critical eletrical field.

For biasing over the knee point the velocity saturation effect becomes predominant.

A new expression of the drain current, which takes into account the channel length modulation, was achieved.

A new expression of the drain current in the active region, which takes into account the channel-length modulation, was obtained.

This equation evinces that the drain current increases due to the voltage drop on the effective pinch-off region.

The accuracy of the proposed model was proven by comparing it with the experimental data measured on *p*MOS and *n*MOS transistors fabricated in the standard 0.18, 0.35 and 1 μm CMOS processes.

A very good theory-experiment agreement was obtained for the transfer characteristics, output characteristics and the transconductance.

It was demonstrated that the model correctly predicts the significant effect of the carrier velocity saturation on several analogue parameters of the MOS transistor.

The model, being based on a reduced set of parameters, can be easily implemented in an analogue ICs simulator.

REFERENCES

[1] P.R.Gray, P.J. Hurst, S.H.Lewis, R.G.Meyer, "Analysis and Design of Analog IC's," Fourth Edition, J.Whiley&Sons, 2001.

[2] T.H.Lee, "The Design of CMOS Radio Frequency Integrated Circuits," Cambridge Univ. Press, 1999.

[3] C.Enz, E. Vittoz, J. Steengaard, "Low Voltage, Low Power Analog IC Design," *Summer Course*, Lausanne, 2005.

[4] C. Leifso, and J.W. Haslett, "An Analytic Model for Estimating the Length of the Velocity Saturated Region in GaAs MESFET's," *IEEE Trans. Electron Devices*, Vol. 47, pp. 905-909, May 2000.

[5] F. Assaderaghi et al., "Saturation velocity and Velocity Overshoot of Inversion Layer Electrons and Holes", *Proc. IEDM*, pp. 479-482, 1994.

[6] D. Sinitsky, F. Assaderaghi, C. Hu, J. Bokor, "High field hole velocity and velocity overshoot in silicon inversion Layers," *IEEE Electron Devices Let.*, vol. 18, no. 2, pp.54-56, February 1997.

[7] G. Brezeanu, A. Sevcenco, "A Compact Model of Submicron Channel MOS Transistor Transconductance for Analog Circuit Simulation," *Proc. ISSCS*, Iasi, Romania, pp. 149- 152, 2007.

[8] A. Sevcenco, G. Brezeanu "Experimental Validation of a Simple Analytical Model for the Electrical Behaviour of Nanoscale MOSFETS," *Proc. CAS*, Sinaia, Romania, pp. 551- 554, 2007.

[9] G. Brezeanu et al. "A Complete Analytical Submicron MOS Transistor Model for Analogue Applications" *Proc. CAS*, Sinaia, Romania, pp. 309- 312, 2008.

[10] A. Sevcenco, Gh. Brezeanu "A Compact Analytical Modelling of the Electrical Characteristics of Submicron Channel MOSFETs," Romanian Journal of Information Science and Technology, 2008