A Basic Property of MOS Transistors and its Circuit Implications

E. Vittoz[1,2], C. Enz[1,2] and F. Krummenacher[2]

[1] CSEM, Jaquet-Droz 1, CH-2007 Neuchâtel, Switzerland, eric.vittoz@csem.ch
[2] EPF-Lausanne, CH-1015 Ecublens, Switzerland

ABSTRACT

The MOS transistor drain current is the (linear) **superposition** of **independent** and **symmetrical** effects of source and drain voltages. This basic property is not affected by the geometry or symmetry of the transistor, by the level of gate voltage or by narrow channel effects. However, it progressively deteriorates when the channel is shortened. Except in weak inversion, it is also degraded by structural non-homogeneities along the channel. This property can be exploited by means of the concept of pseudo-resistors to implement transistor-only linear circuits in the current domain.

Keywords: MOS transistor, superposition, symmetry, current mode.

1 INTRODUCTION

Modelling a MOS transistor is usually aimed at providing means to simulate its behavior by quantitative calculation on a computer. However, another important purpose that is often neglected is that of highlighting basic properties of the device, in order to facilitate the understanding and the synthesis of robust circuits.
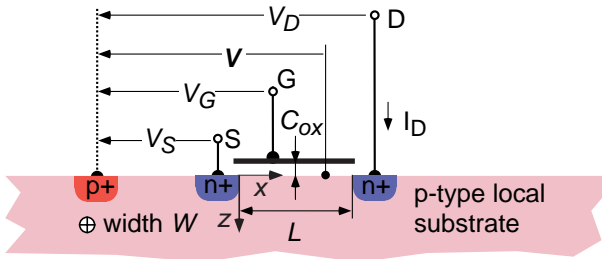

Figure 1: Definitions for n-channel MOS transistor.

The best models are those that can combine both purposes, by means of a hierarchical approach, as is available in particular in the EKV model [1]. Among other characteristics, this model uses substrate-referred definitions of source V_S, drain V_D and gate V_G voltages, as illustrated in Fig. 1 for a n-channel transistor. It also introduces a local "channel potential" V, corresponding to the splitting of electron and hole quasi-Fermi levels in the channel due to V_S and/or V_D. Since the holes remain at equilibrium, V is (within a constant) the quasi-Fermi level of electrons; it takes the value V_S at the

source end ($x = 0$), and V_D at the drain end ($x = L$) of the channel.

Using this approach we will show that a basic property of the MOS transistor can be clearly identified qualitatively.

2 BASIC PROPERTY

Assuming a long and wide channel, and with the above definition of channel potential V, the drain current I_D can be expressed as [2]

$$I_D = \mu W (-Q_i) \frac{dV}{dx} \qquad (1)$$

where Q_i is the local mobile charge per unit area and μ the local mobility.

Let us now assume that equ. (1) can be written as

$$I_D = \frac{F(V, V_G)}{G(x, V_G)} \frac{dV}{dx} \qquad (2)$$

that is separable in channel potential V and channel position x, then:

$$I_D \int_0^L G(x, V_G) dx = \int_{V_S}^{V_D} F(V, V_G) dV. \qquad (3)$$

Now, since $F(V, V_G)$ tends to zero for V large, (3) may be written as

$$I_D = \frac{1}{\int_0^L G(x, V_G) dx} \left[\int_{V_S}^{\infty} F(V, V_G) dV - \int_{V_D}^{\infty} F(V, V_G) dV \right] \qquad (4)$$

If channel length L is independent of V_S, V_D and I_D, then (4) may be rewritten in the very simple form

$$I_D = I(V_S, V_G) - I(V_D, V_G). \qquad (5)$$

This result shows that if equation (2) is fulfilled and the effective channel length L constant, then the transistor exhibits the following basic property:

The drain current is the **superposition** of **independent** and **symmetrical** effects (same function I) of source and drain voltages. One may further define a **forward** component I_F of drain current, independent of V_D

$$I_F = I(V_S, V_G) \qquad (6)$$

and a **reverse** component I_R , independent of V_S

$$I_R = I(V_D, V_G). \tag{7}$$

It is interesting to point out that the above property is similar to that of bipolar transistors as expressed by the Ebers-Moll model.

3   DOMAIN OF VALIDITY

Let us examine the necessary and sufficient conditions by which equation (1) has the form of (2).

If channel width W is not too small, its effective electrical value does not depend on V. It may thus depend on position x along the channel and be included in $G(x, V_G)$ without affecting the basic property. Now, the 1-dimensional equation (1) is no more applicable if the variation of W becomes comparable to channel length L. It must then be replaced by a 2-dimensional equation, but this does not affect relation (5).

The total charge per unit area in silicon Q_{Si} is obtained by applying Gauss' law to a short unit area vertical element of channel :

$$Q_{Si} = -C_{ox}(V_G - V_{FB} - \Psi_s) \tag{8}$$

whereas the bulk depletion charge Q_b is given by

$$Q_b = -\sqrt{2qN_b\varepsilon_{si}\Psi_s} \tag{9}$$

where V_{FB} is the flat-band voltage, Ψ_s the local surface potential, N_b the substrate doping concentration and ε_{si} the dielectric constant of silicon. Thus, the local mobile charge can be expressed as

$$Q_i = Q_{Si} - Q_b = -C_{ox}(V_G - V_{FB} - \Psi_s) + \sqrt{2qN_b\varepsilon_{si}\Psi_s}. \tag{10}$$

If (and only if) $V_G - V_{FB}$, C_{ox} and N_b are all independent of x (homogeneous channel), then Q_i is only a function of Ψ_s that in turn can only be a function of V for V_G constant. This result is still valid if any other term in (10) also depends on V (or Ψ_s but not on x). This includes the effect of polydepletion, which can be modelled by an equivalent C_{ox} that is function of Ψ_s (Ψ_s-dependent dielectric thickness).

The local mobility depends on the local vertical field, which for an homogenous channel only depends on Ψ_s and thus on V. This variation can thus be included in $F(V, V_G)$ and does not affect the basic property. This is only true if the velocity of carriers along the channel remains a negligible fraction of its saturation value.

As another necessary condition, the effective value of channel length L along which $G(x, V_G)$ is integrated must depend neither on drain current I_D nor on drain or source voltage V_D or V_S.

In summary, the basic property of MOS transistors expressed by equation (5) is available when the channel is long and homogeneous, independently of its shape. It remains valid for large gate voltages in spite of the resulting mobility reduction due to the vertical field. It will be shown in subsection 4.3 that it is also maintained for narrow channel transistors.

4   CAUSES OF DEGRADATION

4.1   Short channel effects

When the channel is not very long, several independent mechanisms degrade the basic property. This is the most important reason why this property is not fully available in practice.

As the drain (or source) voltage increases beyond forward (reverse) saturation, the depleted region around the drain (source) extends inside the channel, thus shortening the effective channel length and increasing I_F or I_R. As a result, at least one of these two components becomes dependent on both V_D and V_S. This absolute reduction of channel length takes an increasingly relative importance when the channel is shortened. For very short channel, additional effects (barrier lowering, 2-dimensional effects) further degrade or destroy the property.

If the drain current per unit width is increased by reducing L while maintaining the density Q_i of carriers, their velocity must increase; but as velocity saturation is approached, the mobility is reduced, which makes it a function of I_D, thus destroying the validity of (2).

4.2   Non-homogeneous channel

Referring to equation (10), if any term of the right-hand side depends on position x in the channel, it makes Q_i a (non-separable) function of both x and Ψ_s (thus of V); relation (2) is no more valid and the basic property is lost. This is even true if the non-homogenous channel is symmetrical with respect to source and drain: indeed, the effects of source and drain voltages on I_D remain symmetrical, but they are no more independent nor linearly superimposed.

Since ε_{si} and V_G can reasonably be assumed to be constant along the channel, three terms remain to be examined in (10).

Variations of substrate doping N_b can be due to an intentional source to drain doping asymmetry, as in lightly doped drain (LDD) structures, or to some artifact of the process, like the piling-up of impurities at both ends of the channel. Since such a variation of N_b always occurs at the very ends of the channel, this is yet another reason why the basic property is lost in very short channel devices.

Since the flatband voltage V_{FB} depends on the Fermi level of silicon in the channel, it is variable as soon as the doping itself is variable. Further variation could be induced by variations of the fixed interface charge, as a consequence of non source-drain symmetrical channel engineering.

There is no reason to intentionally change the value of oxide capacitance C_{ox} along the channel, but a variation at both ends is unavoidable, which further contributes to the degradation of the basic property for short channel devices.

Weak inversion represents a special case. It is characterized by Q_i negligible with respect to Q_b. Equation (10) is thus no more applicable to calculate Q_i, which must be

computed directly from the electron density n given by

$$n = \frac{n_i^2}{N_b} e^{\frac{\Psi - V}{U_T}} \qquad (11)$$

where n_i is the intrinsic carrier density and Ψ the local electrostatic potential. Now, by definition, Q_i is obtained by integrating n vertically from $z=0$ (surface of silicon) into deep in the substrate:

$$-Q_i = q \int_0^\infty n \, dz. \qquad (12)$$

By introducing the vertical field $E_z = -d\Psi/dz$, integration in space may be replaced by integration in potential:

$$
\begin{aligned}
-Q_i &= q \int_0^{\Psi_s} \frac{n}{E_z} d\Psi = \frac{q n_i^2}{N_b} \int_0^{\Psi_s} \frac{e^{\Psi/U_T}}{E_z} d\Psi \cdot e^{-V/U_T} \\
&= G_q e^{-V/U_T}.
\end{aligned}
\qquad (13)
$$

Now, since Q_i is negligible with respect to Q_b, it has no effect on Ψ, Ψ_s and E_z. Thus G_q in (13) is independent of Q_i (which is not the case in strong or moderate inversion). Furthermore, since V only affects the negligible Q_i, G_q is also independent of V; G_q can therefore be any function of position x along the channel without affecting the validity of (2): indeed, $F(V, V_G) = e^{-V/U_T}$ and G_q can be included in $G(x, V_G)$.

In weak inversion, the basic property expressed by (5) is thus valid **even for a non-homogeneous channel** (N_b, C_{ox}, $V_G - V_{FB}$ variable with position along the channel). This property, which is also found in bipolar transistors operated in moderate injection, can be traced back to the fact that the current is a linear function of the mobile charge density, as long as this mobile charge does not affect the electrostatic potentials (dominated by Q_b in weak inversion and by majority carriers in moderate injection).

4.3 Narrow channel effects

The distribution of the gate to channel field lines is modified by the side structure of the channel. This is equivalent to connecting in parallel many transistors having different characteristics. However, if each transistor i fulfils equation (5), with

$$I_{Di} = I_i(V_S, V_G) - I_i(V_D, V_G) \qquad (14)$$

then the sum of I_{Di} fulfils it as well.

5 PSEUDO-RESISTOR CONCEPT

By defining a pseudo-voltage V^* [3], [4] given by

$$V^* = -K_0 \int_V^\infty F(V, V_G) dV \qquad (15)$$

and a pseudo-resistance R^* defined by

$$R^* = K_0 \int_0^L G(x, V_G) dx \qquad (16)$$

equation (4) can be rewritten in the form of the linear pseudo-Ohm's law:

$$I_D = (V_D^* - V_S^*)/R^* \qquad (17)$$

This demonstrates that, as long as the basic property is valid, any network obtained by interconnecting transistors characterized by the same function $F(V, V_G)$ and the same gate voltage V_G is **linear with respect to currents** (linear current splitting [5]). Thus, any prototype network made of real linear resistors may be converted to a pseudo-resistor network made of transistors only, provided only currents are considered. A ground in the resistor network ($V = 0$) corresponds to a **pseudo-ground** in the transistor network ($V^* = 0$) obtained by saturating the corresponding side of the transistor.

The value of constant K_0 introduced in (15) and (16) is irrelevant, since it disappears in (17), but its dimension can be chosen so as to obtain the dimension of V^* in volts and that of R^* in ohms.

It should be pointed out that, although a narrow channel does not affect the basic property of the transistor, it has an effect on the function $F(V, V_G)$, which may differ from that of a wider channel, thereby degrading the linearity of current splitting.

For the special case of weak inversion, equation (13) shows that $F(V, V_G)$, thus V^*, is independent of V_G, whereas R^* depends on V_G. The linear pseudo-Ohm's law is thus valid even for transistors having different gate voltages, which can be used to modulate the values of R^*. Furthermore, linear current splitting is maintained even with narrow channel transistors.

6 CIRCUIT EXAMPLES

The most immediate application of (pseudo-)resistive circuits is in attenuators [5]; a special case is that of the $R - 2R$ network generating binary weighted currents for DA conversion [6]. Further applications are found in processing of spatial information [4]: n^{th}-order moment computation, isotropic and non-isotropic diffusion networks, 2-D emulation of physical media, path finding. In these examples, normal resistors can simply be replaced by more IC-compatible transistors with a common and fixed gate voltage.

Pseudo-resistive circuits that have no usable resistive prototype exploit the possibility to make the common gate voltage dependent on the signals. Fig. 2 shows a general current mirror structure together with its resistive prototype.

As long as the two upper transistors are saturated (pseudo-grounds), the corresponding resistors are grounded. It is then easy to write the linear equation relating the various currents from the resistive network:

$$I_1(R_1 + R_2) + I_a R_2 = I_3(R_3 + R_4) + I_b R_4 \qquad (18)$$

thus, in the transistor circuit with $A_i = W_i/L_i \propto 1/R_i$:

$$I_3 = I_1 \frac{A_1 + A_2}{A_3 + A_4} \frac{A_3 A_4}{A_1 A_2} + I_a \frac{A_3 A_4}{A_2(A_3 + A_4)} - I_b \frac{A_3}{A_3 + A_4}. \qquad (19)$$
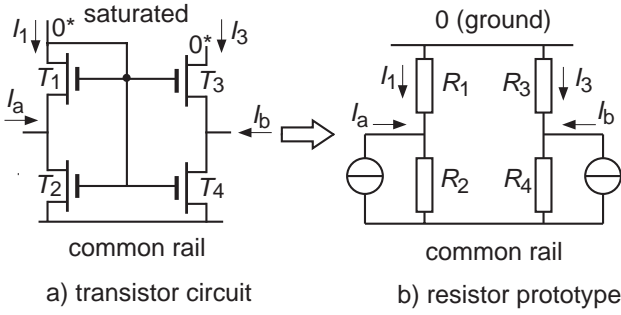
Figure 2: General current mirror.

Output current I_3 is thus the **weighted sum** of input currents I_1, I_a and $-I_b$. Furthermore, since T_2 and T_4 are not saturated, I_a and I_b are injected at a voltage level below V_{DSsat}, providing a **low input voltage** current mirror (I_b may be ignored with T_3 and T_4 merged into a single transistor if only positive weighting is needed).

If T_4 is short-circuited and all other transistors are identical with $I_a = 0$, then $I_3 = 2I_1$. A current ratio N is obtained if N transistors are connected in series instead of two. Further connecting M transistors in parallel at the output provides a large ratio $N.M$ with only $N + M$ identical transistors (much better matching than by using different channel lengths). A very small ratio can be obtained by exchanging input and output.

If operation is limited to weak inversion, a rich variety of current-linear circuits is possible, since each transistor becomes equivalent to a linear voltage- or current-controlled resistor [4].

A transistor is (forward) saturated when $I_R \ll I_F$. This can be exploited to implement the bias circuit of a low-voltage cascode, as illustrated in Fig. 3 [7], [8].
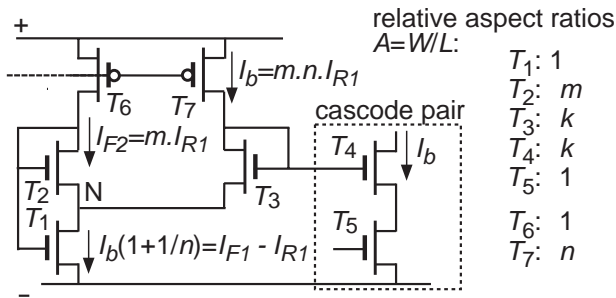

Figure 3: Low-voltage cascode [7], [8].

Saturated transistors T_3 and T_4 have the same current density (same $I_F L/W$), thus their sources are at the same voltage. Therefore, T_1 and T_5 have the same drain voltage. If $n \gg 1$, they also have the same current density, thus I_F/I_R of T_5 is equal to that of T_1.

Since T_2 is saturated and m-times larger than T_1, its drain current I_{F2} is m-times the reverse current I_{R1} of T_1. Applica-

tion of Kirchhoff's law at node N results in

$$I_{F1}/I_{R1} = 1 + m(n + 1) \qquad (20)$$

which is independent of bias current I_b and can be large with reasonable values of ratios m and n. The drain voltage of T_5 may be further increased by reducing the size of T_3 (for example to that of T_1, as proposed in [8]).

7 CONCLUSION

For a long channel MOS transistor that is homogeneous along the channel, the drain current is the linear superposition of independent and symmetrical effects of source and drain voltages (forward and reverse components). This basic property is valid for any shape of transistor, even if the channel is very narrow. It is not affected by the reduction of mobility due to large gate voltage, nor by polydepletion effects.

Due to a combination of effects, this property is progressively degraded when the channel is shortened. It is further degraded if the channel is not homogeneous from source to drain, except if the device is operated in weak inversion (mobile charge negligible with respect to depletion charge).

This basic property may be exploited to obtain a linear splitting of currents [5] among similar transistors. This principle can be formalized by means of the concept of pseudo-resistor [3], which finds numerous applications in current mode circuits [4].

REFERENCES

[1] C. Enz, F. Krummenacher and E. Vittoz, Analog Integrated Circuits and Signal Processing, 8, 83-114, 1995.

[2] Y. Tsividis, Operation and Modeling of The MOS Transistor, p.152, McGraw-Hill,1999.

[3] E. Vittoz and X. Arreguit, El. Letters, 29, 297-299, 1993.

[4] E. Vittoz, Proc. MicroNeuro'97, Dresden,163-173, 1997.

[5] K. Bult and G. Geelen, ISSCC Digest Tech. Papers, 198-199, 1992.

[6] C. Enz and E. Vittoz, Emerging Technologies: Designing Low-Power Digital Circuits, IEEE Piscataway, 79-133, 1996.

[7] E. Vittoz, Design of VLSI Circuits for Telecommunications and Signal Processing, Chapter 5, Prentice-Hall, 1994.

[8] B. Minch, Proc. ISCAS'02, 3, 619-622, 2002.