Chapter 1.2

CMOS Low-Power Analog Circuit Design

Christian C. Enz and Eric A. Vittoz

Abstract

This chapter covers device and circuit aspects of low-power analog CMOS circuit design. The fundamental limits constraining the design of low-power circuits are first recalled with an emphasis on the implications of supply voltage reduction. Biasing MOS transistors at very low current provides new features but requires dedicated models valid in all regions of operation including weak, moderate and strong inversion. Low-current biasing also has a strong influence on noise and matching properties. All these issues are discussed, together with the particular aspects related to passive devices and parasitic effects. The design process has to be supported by efficient and accurate circuit simulation. To this end, the EKV compact MOST model for circuit simulation is shortly presented. The use of the basic concepts such as pinch-off voltage, inversion factor and specific current are highlighted thanks to some very simple but fundamental circuits and to an effective use of the model. New design techniques that are appropriate for low-power and/or low-voltage circuits are presented with an emphasis on the analog floating point technique, the instantaneous companding principle, and their application to filters.

1.2.1 Introduction

The current trend towards low-power design is mainly driven by two forces [1]: the growing demand for long-life autonomous portable equipment, and the technological limitations of high-performance VLSI systems. For the first category of products, low-power is the major goal for which speed and/or dynamic range might have to be sacrificed. High speed and high integration density are the objectives for the second application category, which has experienced a dramatic increase of heat dissipation that is now reaching a fundamental limit. These two forces are now merging as portable equipment grows to encompass high-throughput computationally intensive products such as portable computers and cellular phones.

The most efficient way to reduce the power consumption of digital circuits is definitely to reduce the supply voltage, since the average power consumption of CMOS digital circuits is proportional to the square of the supply voltage. On the other hand, the reduction of the supply voltage is also required to maintain the electric field at an acceptable level. The resulting performance loss can be overcomed for standard CMOS

technologies by introducing more parallelism [2][3][4] and/or to modify the process and optimize it for low supply voltage operation [4][5][6].

The rules for analog circuits are quite different than those applied to digital circuits. In order to clarify these differences, the fundamental limits to the reduction of the power consumption are recalled in Section 1.2.2. It is shown that decreasing the supply voltage does unfortunately not reduce the power consumption of analog circuits. This is mainly due to the fact that the power consumption of analog circuits at a given temperature is basically set by the required signal-to-noise ratio (SNR) and the frequency of operation (or the required bandwidth). A first-order analysis also shows that the absolute minimum power consumption required to process analog signals is almost independent of the supply voltage reduction. In addition to these fundamental limits, some practical limits and additional obstacles to the power reduction are also discussed.

This trend towards low-power has emphasized some aspects of MOS modelling for low-voltage and low-current analog circuit design and simulation. Particularly, the necessity to have a clear understanding of the MOS transistor operating at very low-current and to correctly model the operation of the device in the weak and moderate inversion regions, which have been ignored by most designers for years, has become a must. The availability of a good MOS transistor model has thus become a real issue for the efficient design and simulation of high performance analog and digital integrated circuits. Section 1.2.3 presents the operation and modeling of the long-channel MOS transistor with a strong emphasis on low-current. Based on this analytical model, a compact MOST model (named the EKV model) has been developed for circuit simulation and is shortly presented in Section 1.2.4.

The advantages of this model are brought out in Section 1.2.5 thanks to some simple but fundamental circuit examples. The basic concepts such as the pinch-off voltage, the specific current, the inversion factor are illustrated and exploited to better understand existing circuits or to develop new ones. They lead to the development of an attractive ratio-based design technique which is portable from one process to another, to MOS only current dividers and to low-voltage cascode bias circuits.

Some additional system considerations are given in Section 1.2.6.2. Some analog signal processing systems like hearing aids for example require a SNR much smaller than the dynamic range. A significant power reduction can be obtained by distinguishing the SNR and the dynamic range and letting the noise follow the signal level to maintain just the necessary value of SNR. This can be achieved thanks to the analog floating point approach which uses a variable gain (or range switcher) at the input and the output of the analog signal processing system to realize the compression and respectively the expansion. Distortion-free operation is obtained thanks to a proper update of all the system's state variables.

Section 1.2.6.3 presents an approach to low-voltage analog signal processing called "log-domain" filtering, where currents are compressed logarithmically when transformed into voltages and expanded exponentially when converted back to currents. The input signal has to be predistorted in order to avoid any distortion due to the non-linear operation. The log and exponential functions are implemented thanks to the exponential current-to-voltage characteristics inherent to the bipolar transistor or to the MOS transistor biased in weak inversion. Both of these techniques seem to be very attractive for low-power and low-voltage circuit design, but still a lot of effort is needed to better understand them and demonstrate their benefits for low-power.

Finally, a summary and some conclusions are drawn in Section 1.2.7.

1.2.2 Limits to low-power for analog circuit design

1.2.2.1 Fundamental limits

Power is consumed in analog signal processing circuits to maintain the signal energy above the fundamental thermal noise in order to achieve the required signal-to-noise ratio (SNR). A representative figure of merit of different signal processing systems is the power consumed to realize a single pole. The minimum power necessary to realize a single pole can be derived by considering the basic integrator presented in Fig. 2.1 assuming an ideal 100% current efficient transconductor, meaning that all the current pulled from the supply voltage is used to charge the integrating capacitor.

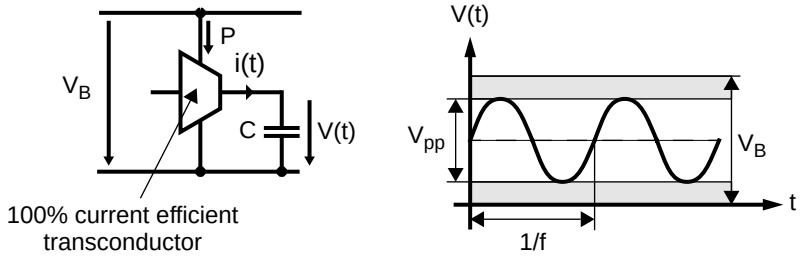


Figure 2.1 Basic integrator used to evaluate the power necessary to realize a single pole

The power consumed from the supply voltage source V_B which is necessary to create a sinusoidal voltage *V(t)* across capacitor C having a peak-to-peak amplitude V_{pp} and a frequency *f* can be expressed as:

$$P = V_B \cdot fCV_{pp} = fCV_{pp}^2 \cdot \frac{V_B}{V_{pp}} \tag{1}$$

whereas the signal-to-noise ratio is given by:

$$SNR = \frac{V_{pp}^2/8}{kT/C} \tag{2}$$

Combining (1) and (2) yields:

$$P = 8kT \cdot f \cdot SNR \cdot \frac{V_B}{V_{pp}} \tag{3}$$

According to (3), the minimum power consumption of analog circuits at a given temperature is basically set by the required SNR and the frequency of operation (or the required bandwidth). Since this minimum power consumption is also proportional to the ratio between the supply voltage and the signal peak-to-peak amplitude, power efficient analog circuits should be designed to maximize the voltage swing. The minimum power for circuits that can handle rail-to-rail signal voltages ($V_{pp} = V_B$) reduces to [7][8][9]:

$$P_{min} = 8kT \cdot f \cdot SNR \tag{4}$$

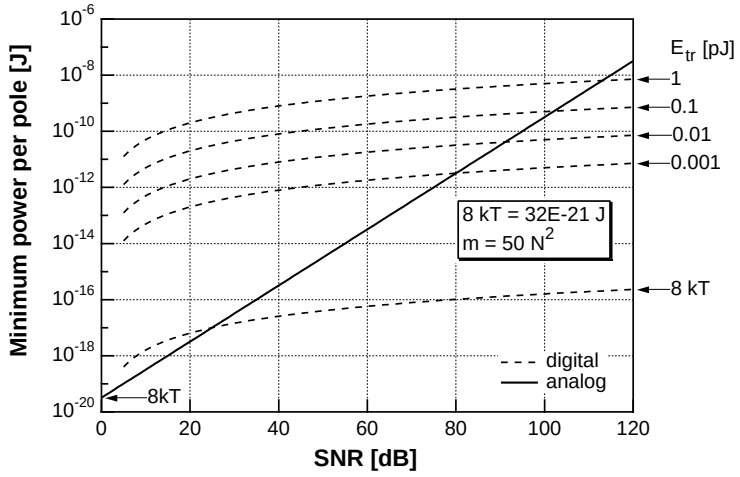


Figure 2.2 Minimum power for analog and digital circuits

This absolute limit is very steep, since it requires a factor 10 of power increase for every 10 dB of signal-to-noise ratio improvement. It applies to each pole of any linear analog filter (continuous or sampled-data as switched capacitors [8]) and is reached in the case of a simple passive RC filter, whereas the best existing active filters are still two orders of magnitude above. High-Q poles in the passband reduce the maximum amplitude at other frequencies and therefore increase the required power, according to (3).

Approximately the same result is found for relaxation oscillators, whereas the minimum power required for a voltage amplifier of gain A_v can be derived considering a single stage common-source (or common-emitter) small-signal amplifier. The signal-to-noise ratio is obtained by comparing the input rms voltage to the input-referred noise voltage:

$$SNR = \frac{V_{in}^2}{4kTR_N \cdot \Delta f} \tag{5}$$

where R_N is the input-referred thermal noise resistance, which is given by:

$$R_N = \frac{\gamma}{g_m} = \frac{\gamma}{I} \cdot \frac{I}{g_m} = \gamma \cdot \frac{V_B}{P} \cdot \frac{I}{g_m} \tag{6}$$

where γ is the noise factor defined by (48) as the product of the input-referred thermal noise resistance and the effective transconductance of the device. It is equal to $n/2$ in the case of a MOS transistor biased in weak inversion where n is the slope factor defined by (19). The power can be expressed as a function of the SNR by combining (5) and (6):

$$P = 4\gamma kT \cdot \Delta f \cdot \frac{V_B}{V_{in}^2} \cdot \frac{I}{g_m} \cdot SNR \tag{7}$$

As will be shown in Section 1.2.3, for a transistor biased in weak inversion the current-to-transconductance ratio I/g_m is equal to nU_T, where $U_T \equiv kT/q$ is the thermodynamic voltage. In this case, the power given by (7) reduces to:

$$P = 2n \cdot kT \cdot \Delta f \cdot \frac{V_B}{V_{in}} \cdot \frac{nU_T}{V_{in}} \cdot SNR \tag{8}$$

Since the peak-to-peak amplitude of the output voltage $A_v \cdot 2\sqrt{2}V_{in}$ is bounded by the supply voltage V_B, the absolute minimum power is given by:

$$P > P_{min} = 8n \cdot kT \cdot \Delta f \cdot A_v \cdot SNR \tag{9}$$

where it has been assumed that the input voltage is maximum and equal to $nU_T/\sqrt{2}$. According to (9), the minimum power for an amplifier is nA_v-times larger than the limit given by (4).

The minimum power for an analog system can be compared to that of a digital system, in which each elementary operation requires a certain number *m* of binary gate transition cycles, each of which dissipates an amount of energy E_{tr}. The minimum power is then simply given by:

$$P_{min\text{-}digital} = m \cdot f \cdot E_{tr} \tag{10}$$

where *f* is the signal bandwidth. The number *m* of transitions is only proportional to some power *a* of the number of bits *N*, and therefore power consumption is only weakly dependent on SNR (essentially logarithmically):

$$m \cong N^a \sim [\log (SNR)]^a \tag{11}$$

Comparison with analog is obtained by estimating the number of gate transitions that are required to compute each period of the signal, which for a single pole digital filter can be estimated to be approximately:

$$m \cong 50 \cdot N^2 \tag{12}$$

Immunity to thermal noise imposes an absolute minimum energy per transition E_{trmin} estimated to 8kT, which provides the absolute minimum power limit. However, in practice $E_{tr} = C \cdot V_B^2$ is forced to a much higher value (10^{-15} to 10^{-12} Joules) by the need to recharge the equivalent capacitance *C* of each gate to the supply voltage V_B. As shown in Fig. 2.2, the minimum power for digital is therefore much higher than the absolute limit $8kT \cong 32\times10^{-21} J$ at room temperature. The minimum gate capacitance is strongly dependent on the process feature size and the supply voltage is imposed by the need to achieve the required delay time and by established standards. Furthermore, if the activation rate of the circuit is very low (very small percentage of the available gates in transition on average), then the standby current of each of the gates may contribute to a non negligible additional static power consumption.

Comparison of these fundamental limits are plotted in Fig. 2.2. They clearly show that analog systems may consume much less power than their digital counterpart, provided a small signal-to-noise ratio is acceptable. But for systems requiring large signal-to-noise ratios, analog becomes very power inefficient. It is worth mentioning that a comparison of chip area basically leads to the same qualitative conclusion.

1.2.2.2 Practical limits

The limits discussed so far are fundamental since they do not depend on the technology nor on the choice of power supply voltage. However, a number of obstacles or technological limitations are on the way to approach these limits in practical circuits:

a) Capacitors increase the power necessary to achieve a given bandwidth. They are only acceptable if their presence reduces the noise power by the same amount (by reducing the noise bandwidth). Therefore, ill-placed parasitic capacitors very often increase power consumption.
b) The power spent in bias circuitry is wasted and should in principle be minimized. However, inadequate bias schemes may increase the noise and therefore require a proportional increase in power. For example, a bias current is more noisy if it is obtained by multiplying a smaller current.
c) According to (3), power is increased if the signal at any node corresponding to a functional pole (pole within the bandwidth, or state variable) has a peak-to-peak voltage amplitude smaller than the supply voltage V_B . Thus, care must be taken to amplify the signal as early as possible to its maximum possible voltage value, and to maintain this level all along the processing path. Using current-mode circuits with limited voltage swings is therefore not a good approach to reduce power, as long as the energy is supplied by a voltage source. It only becomes attractive if voltage companding techniques can be used (see Section 1.2.6.3).
d) The presence of additional sources of noise implies an increase in power consumption. These include 1/f noise in the devices, and noise coming from the power supply or generated on chip by other blocks of the circuit.
e) When capacitive loads are imposed (for example by parasitic capacitors), the current I necessary to obtain a given bandwidth is inversely proportional to the transconductance-to-current ratio g_m/I of the active device. The small value of g_m/I inherent to MOS transistors operated in strong inversion may therefore cause an increase in power consumption.
f) The need for precision usually leads to the use of larger dimensions for active and passive components, with a resulting increase in parasitic capacitors and power.
g) All switched capacitors must be clocked at a frequency higher than twice the signal frequency. The power consumed by the clock itself may be dominant in some applications.

Ways to reduce the effect of these various limitations can be found at all levels of analog design ranging from device to system.

1.2.2.3 Other obstacles to low-power

In addition to the fundamental and practical limitations discussed previously, there are also historical or even psychological barriers to the efficient design of LP analog circuits. The most important can be listed as:

a) Analog blocks must often be taken from existing libraries with bias currents at the milliampere level and with architectures that are not compatible with low-voltage or low-current.
b) The use of very low bias currents is often discarded due to a lack of adequate transistor models and correct characterization of transistors parameters as well as worst case leakage currents. Another obstacle is the fear of breaking the psychological

microampere barrier.

c) The requirements on PSRR are often exaggerated and mistaken for insensitivity to noise generated on chip.

1.2.2.4 Implications of supply voltage reduction

Unlike digital circuits, where the dynamic power decreases with the square of the supply voltage, according to (3), reducing the supply voltage of analog circuits while preserving the same bandwidth and SNR, has no fundamental effect on their minimum power consumption. However, this absolute limit was obtained by neglecting the possible limitation of bandwidth B due to the limited transconductance g_m of the active device. The maximum value of B is proportional to g_m/C. Replacing the capacitor value C by g_m/B in (2) and expressing the product of the SNR times the bandwidth yields:

$$SNR \cdot B = \frac{V_{pp}^2 \cdot g_m}{8kT} \tag{13}$$

In most cases, scaling the supply voltage V_B by a factor K requires a proportional reduction of the signal swing V_{pp}. Maintaining the bandwidth and the SNR is therefore only possible if the transconductance g_m is increased by a factor K^2. If the active device is a bipolar transistor (or a MOS transistor biased in weak inversion), its transconductance can only be increased by increasing the bias current I by the same factor K^2; power $V_B \cdot I$ is therefore increased by K. The situation is different if the active device is a MOS transistor biased in strong inversion. Its transconductance can be shown to be proportional to I/V_P, where V_P is the pinch-off or saturation voltage of the device. Since this saturation voltage has to be reduced proportionally with V_B, increasing g_m by K^2 only requires an increase of current I by a factor K and hence the power remains unchanged.

However, the maximum frequency of operation may be affected by the value of the supply voltage. For a MOS transistor in strong inversion, the frequency f_{max} for which the current gain falls to unity is approximately given by:

$$f_{max} \cong \frac{\mu \cdot V_P}{L^2} \tag{14}$$

Therefore, if the process is fixed (channel length L constant) a reduction of V_B and V_P by a factor K causes a proportional reduction of f_{max}. However, there is no fundamental reason to reduce the supply voltage of an analog circuit in a given process. On the other hand, a reduction of V_B is unavoidable to maintain the electric fields constant when scaling down a process. Both V_P and L are then scaled by the same factor K and the maximum frequency f_{max} is increased by K. For a bipolar transistor, V_P in (14) is replaced by $U_T = kT/q$ and f_{max} does not, in first approximation, depend on the supply voltage V_B.

Low-voltage limitations are not restricted to power or frequency problems. Reducing V_P increases the transconductance-to-current ratio of MOS transistors which in turn increases the noise content of current sources and drastically degrades their precision. Conductance in analog switches is difficult to ensure when the supply voltage falls below approximately the sum of the p- and n-channel transistor threshold voltages. For a given value of time constant, charge injection in a switch does not depend on V_B in absolute

value, but it increases in relative value if V_B and V_{pp} are decreased. The same is true for any constant voltage overhead such as the base-emitter voltage in bipolar transistors or the threshold voltage in MOS transistors.

As already illustrated in the previous discussion on the fundamental limits to LP and LV, many of the problems and solutions encountered in the design of LP-LV analog circuits are directly related to the properties and limitations of the MOS transistor itself, which must therefore be properly understood and correctly modelled down to very low currents. For this reason, the basic operation of the long-channel MOS transistor will be presented in Section 1.2.3, with a strong emphasis on low current and weak inversion operation. A complete but simple analytical model that can be used for the analysis and design of simple analog circuits will be elaborated. Section 1.2.4 will present the use of the MOST long-channel model and its extension to the EKV MOST model, a compact MOST model which includes all the second-order effects (so important in real-world design) and which is dedicated to LP and LV circuits simulation.

1.2.3 MOST basic long-channel static model at low current

1.2.3.1 Drain current, pinch-off voltage and modes of operation

Fig. 2.3 shows the cross-section and the corresponding symbol of an idealized n-channel MOS transistor. The intrinsic geometric and operating symmetry of the device with respect to the source and the drain can be preserved in the model by referring the source voltage V_S, the gate voltage V_G and the drain voltage V_D to the local substrate. This is clearly not the convention adopted for SPICE models for which all potentials are referred to the source electrode. The drain current is defined positive entering the drain electrode which is the most positive electrode between source and drain.

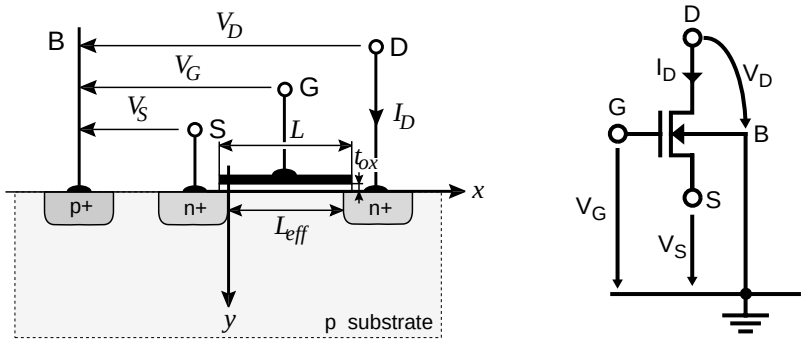


Figure 2.3 Cross-section of an idealized n-channel MOS transistor and the corresponding symbol. All voltages are referred to the local p-type substrate

A general expression for the drain current that includes both the drift and the diffusion transport mechanisms is given by [11][12][13][14][15]:

$$I_D = W \cdot \mu_n \cdot (-Q'_{inv}) \cdot \frac{dV_{ch}}{dx} \tag{15}$$

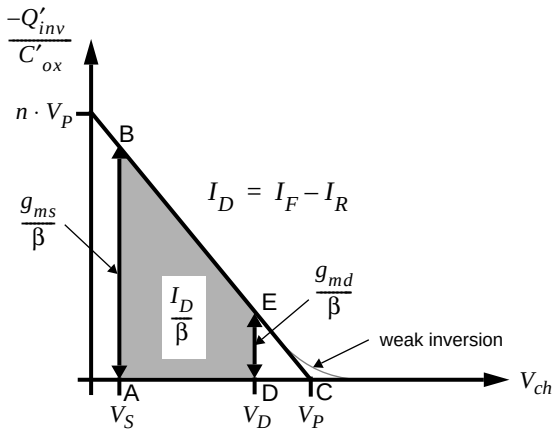


Figure 2.4 Inversion charge versus the channel potential. The drain current is proportional to the shaded surface

where it has been assumed that the mobility μ_n is constant along the *y* axis and that the carrier velocity is not saturated. According to (15), the drain current is proportional to the gradient of the channel voltage (or the longitudinal electrical field E_x) and to the *mobile inversion charge density* Q'_{inv}. The *channel voltage* V_{ch}, which depends on the position along the channel, is defined as the difference between the quasi-Fermi potential of the mobile carriers forming the channel ϕ_n and the Fermi potential Φ_F. This channel potential represents the disequilibrium in electron distribution produced by the source and the drain voltages.

As shown in Fig. 2.4, for a given gate voltage larger than the threshold voltage V_{TO}, the inversion charge Q'_{inv} is almost a linear function of the channel voltage [14][19]. The channel voltage for which Q'_{inv} becomes zero is defined as the *pinch-off voltage* V_P and corresponds in Fig. 2.4 to the intersection of Q'_{inv} with the *x*-axis (point C). V_P represents the voltage that should be applied to the equipotential channel (source and drain connected together) to cancel the effect of the gate voltage. It can be interpreted as the effect of the gate voltage referred to the channel and is thus directly related to the gate voltage by [14][19]:

$$V_P = V_G - V_{T0} - \gamma \cdot \left[\sqrt{V_G - V_{T0} + \left(\sqrt{\Psi_0} + \frac{\gamma}{2} \right)^2} - \left(\sqrt{\Psi_0} + \frac{\gamma}{2} \right) \right] \tag{16}$$

or inversely:
$$V_G = V_{T0} + V_P + \gamma \cdot \left[\sqrt{\Psi_0 + V_P} - \sqrt{\Psi_0} \right] \tag{17}$$

where V_{T0} is the *threshold voltage* which is also referred to the substrate and Ψ_0 is the approximation of the surface potential in strong inversion at equilibrium ($V_{ch} = 0$). γ is the *body effect factor* defined as:

$$\gamma = \frac{\sqrt{2 \cdot q \cdot \varepsilon_s \cdot N_{sub}}}{C'_{ox}} \tag{18}$$

The *slope factor n* corresponds to the slope of the V_G versus V_P characteristic. It is a function of the pinch-off voltage and can thus be approximated by:

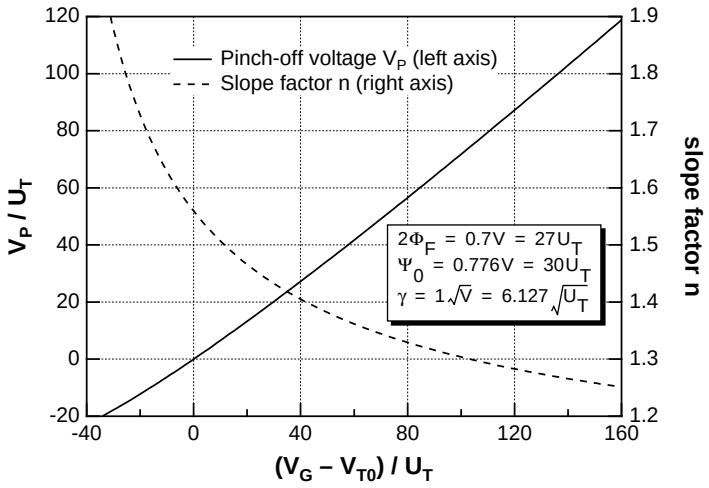


Figure 2.5 Pinch-off voltage and slope factor versus gate voltage

$$n \equiv \frac{dV_G}{dV_P} = 1 + \frac{\gamma}{2 \cdot \sqrt{\Psi_0 + V_P}} \tag{19}$$

n is also related to the slope $S \equiv (nU_T)^{-1}$ of the I-V characteristics in weak inversion plotted in a log-lin scale. Since V_P is a function of the gate voltage, n can also be expressed directly as a function of the gate voltage:

$$\frac{1}{n} = \frac{dV_P}{dV_G} = 1 - \frac{\gamma}{2 \cdot \sqrt{V_G - V_{T0} + \left(\frac{\gamma}{2} + \sqrt{\Psi_0}\right)^2}} \tag{20}$$

As shown in Fig. 2.5, for the values of γ and Ψ_0 used in practice, the pinch-off voltage can be approximated by a linear function of the gate voltage:

$$V_P \cong \frac{V_G - V_{T0}}{n(V_G)} \tag{21}$$

The drain current is obtained simply by integrating (15) from the source, where $V_{ch} = V_S$, to the drain, where $V_{ch} = V_D$. Assuming that the mobility is constant along the channel, yields:

$$I_D = \beta \cdot \int_{V_S}^{V_D} \left[\frac{-Q'_{inv}(V_{ch})}{C'_{ox}}\right] \cdot dV_{ch} \tag{22}$$

where:

$$\beta = \mu_n \cdot C'_{ox} \cdot \frac{W_{eff}}{L_{eff}} \tag{23}$$

As shown in Fig. 2.4, the drain current is proportional to the shaded surface comprised between the source and the drain voltage. It can be decomposed into a *forward current* I_F and a *reverse current* I_R defined by [13][14][19]:

$$I_D = \underbrace{\beta \cdot \int_{V_S}^{\infty} \left[\frac{-Q'_{inv}(V_{ch})}{C'_{ox}}\right] \cdot dV_{ch}}_{\text{forward current } I_F} - \underbrace{\beta \cdot \int_{V_D}^{\infty} \left[\frac{-Q'_{inv}(V_{ch})}{C'_{ox}}\right] \cdot dV_{ch}}_{\text{reverse current } I_R} = I_F - I_R \tag{24}$$

The *forward current* I_F corresponds to the triangle ABC delimited by V_S and V_P and hence depends only on the difference $V_P - V_S$ while the *reverse current* I_R corresponds to the small triangle DEC delimited by V_D and V_P and therefore depends only on $V_P - V_D$. Looking to Fig. 2.4 it is clear that when the drain voltage is increased, the area of triangle DEC is reduced and hence the reverse current is decreased. When the drain voltage reaches the pinch-off voltage, the channel is pinched-off at the drain and the reverse current becomes zero. This situation corresponds to the onset of forward saturation. Since $I_R = 0$ in forward saturation, the drain current becomes equal to the forward component I_F .

It should be noted that in reality the pinch-off point is never reached. The device progressively leaves strong inversion and enters in a region of moderate inversion followed by a region of weak inversion, where the inversion charge decreases down to zero exponentially with respect to the channel potential. For this reason the upper limits of the integrals in (24) have been set to infinity instead of V_P , extending the validity of (24) to all the regions of operation including weak, moderate and strong inversion. Although V_P has been defined as an extrapolation of the strong inversion operation, it is also used in weak inversion, which tends to show that it is a good delimiter between the strong and the weak inversion regions. The pinch-off voltage can thus be used to define the different modes of operation of the MOS transistor with respect to the source and drain voltages as illustrated in Fig. 2.6.

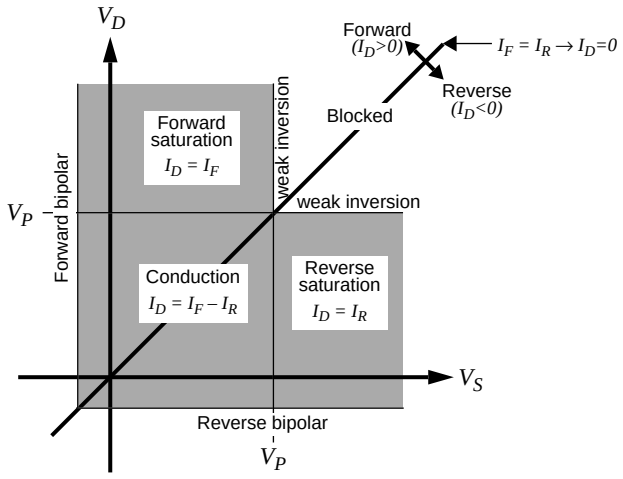


Figure 2.6 Modes of operation of the transistor

Symmetrical forward and reverse modes are possible, depending on the sign of $V_D - V_S$. For V_S and V_D both smaller than V_P, the channel is in strong inversion from the source to the drain and the transistor is in the *conduction* mode. If V_D is increased beyond V_P, the drain end of the channel is pinched-off and the device is in *forward saturation* mode. If V_S and V_D are both larger than V_P, the whole channel is pinched-off. The device operates in *weak inversion* as long as one of the source or drain voltages is still close to V_P, but becomes *blocked* if both of them are sufficiently larger than V_P.

If the drain or (and) the source junction(s) is (are) forward biased beyond a junction voltage V_J, a bipolar mode is superimposed on the MOS mode [24][25].

There are of course no abrupt limits between these various modes, but rather smooth transitions. In particular weak and strong inversion are separated by a region of moderate inversion [11].

Since in strong inversion the inversion charge Q'_{inv} can be approximated by a linear function of $V_P - V_{ch}$, the forward (reverse) current I_F (I_R) is a quadratic function of the difference $V_P - V_S$ ($V_P - V_D$) [14][16][19]:

$$I_{F(R)} = \begin{cases} \frac{n \cdot \beta}{2} \cdot (V_P - V_{S(D)})^2 & for: V_{S(D)} < V_P \\ 0 & for: V_{S(D)} \geq V_P \end{cases} \tag{25}$$

The current can be explicitly related to the gate voltage either by using (16) or the linear approximation given by (21). The complete expressions for the drain current in strong inversion are given in Table 2.1 [14][16][28]. It is interesting to note that the current-to-voltage relation in the conduction mode can be made linear by maintaining the sum of the drain and source voltage constant. This technique is used in the design of linear transconductors [29].

Table 2.1 Drain current in strong inversion

MODE	DRAIN CURRENT EXPRESSION	CONDITION
CONDUCTION	$n \cdot \beta \cdot \left[V_P - \frac{V_S + V_D}{2} \right] \cdot (V_D - V_S)$ $\cong \beta \cdot \left[V_G - V_{T0} - \frac{n}{2} \cdot (V_S + V_D) \right] \cdot (V_D - V_S)$	$V_S \leq V_P$ $V_D \leq V_P$
FORWARD SATURATION	$\frac{n \cdot \beta}{2} \cdot (V_P - V_S)^2 \cong \frac{\beta}{2n} \cdot (V_G - V_{T0} - n \cdot V_S)^2$	$V_S \leq V_P$ $V_D > V_P$
REVERSE SATURATION	$\frac{-n \cdot \beta}{2} \cdot (V_P - V_D)^2 \cong \frac{-\beta}{2n} \cdot (V_G - V_{T0} - n \cdot V_D)^2$	$V_S > V_P$ $V_D \leq V_P$
BLOCKED	$I_F = I_R \Rightarrow I_D = 0$	$V_D = V_S$

In weak inversion, it can be shown that the inversion charge Q'_{inv} is an exponential function of $V_P - V_{ch}$, which results in an exponential forward (reverse) current I_F (I_R) [11][13][14][16][28][32]:

$$I_{F(R)} = I_S \cdot \exp\left[\frac{V_P - V_{S(D)}}{U_T}\right] \tag{26}$$

where I_S is the *specific current* defined as [14]:

$$I_S \equiv 2 \cdot n \cdot \beta \cdot U_T^2 \tag{27}$$

which at a given temperature, depends essentially on the aspect ratio W/L of the device and on the mobility μ_n. As shown in Fig. 2.7, I_S corresponds in fact to the intersection of the weak and strong inversion asymptotes of the normalized transconductance versus drain current plot. I_S can therefore also be used as a delimiter between weak and strong inversion. Setting the drain current in forward saturation larger than I_S forces the device to operate in strong inversion while keeping it smaller than I_S biases the transistor in weak inversion. The specific current is thus a convenient design parameter that helps sizing the transistor according to an imposed bias current and a given mode of operation.
It is convenient to explicit the current in terms of the gate voltage by using the approximation given by (21):

$$I_D = I_F - I_R = I_{D0} \cdot \exp\left[\frac{V_G}{n \cdot U_T}\right] \cdot \left\{\exp\left[\frac{-V_S}{U_T}\right] - \exp\left[\frac{-V_D}{U_T}\right]\right\} \tag{28}$$

where:

$$I_{D0} \equiv I_S \cdot \exp\left[\frac{-V_{T0}}{n \cdot U_T}\right] \tag{29}$$

is the "leakage" current appearing for $V_G = 0$ The parameter I_{D0} is not well controlled since it depends exponentially on V_{T0}. For this reason, a transistor operating in weak inversion should always be biased by imposing a fixed current instead of a fixed gate voltage in order to avoid the high sensitivity to U_T and V_{T0} [26].
An equivalent expression of the drain current that emphasize the saturation process in weak inversion can be derived from (28):

$$I_D = I_F \cdot \left\{1 - \frac{I_R}{I_F}\right\} = I_S \cdot \exp\left[\frac{V_P - V_S}{U_T}\right] \cdot \left\{1 - \exp\left[-\frac{V_D - V_S}{U_T}\right]\right\} \tag{30}$$

For $V_D - V_S > 4 \text{ to } 5U_T$, the reverse component I_R becomes negligible with respect to I_F and the transistor enters in forward saturation. The saturation voltage in weak inversion can thus be considered to be 5 U_T, or approximately 130 mV at room temperature, which would correspond to a contribution of the reverse current to the total current of less than 1%. This obviously makes weak inversion very attractive for low-voltage applications. The expressions for the drain current in weak inversion are summarized in Table 2.2.

In the region between weak and strong inversion defined as the moderate inversion region, the current is due to both diffusion and drift mechanisms. Although the general expression for the current given by (15) holds also for this moderate inversion region, the

expressions derived for the drain current are not valid in moderate inversion because diffusion and drift currents have been considered separately for respectively weak and strong inversion. Considering these two regions more as asymptotic modes of operation, the drain current can be approximated over a wide range of current with an acceptable precision by using the following simple interpolation function [32][33]:

$$I_{F\,(R)} = I_S \cdot \left[ln\left(1 + \exp\left[\frac{V_P - V_{S\,(D)}}{2U_T} \right] \right) \right]^2 \tag{31}$$

It is quite obvious to verify that for $V_P < V_{S\,(D)}$ (or $V_G < V_{T0} + n \cdot V_{S\,(D)}$), the forward (reverse) current reduces to (26), while for $V_P > V_{S\,(D)}$ (or $V_G > V_{T0} + n \cdot V_{S\,(D)}$), I_F reduces to (25).

The currents defined by (31) can conveniently be inverted to express the voltages in terms of the forward or reverse currents as it is generally required in analog circuit design:

$$V_P \cong \frac{V_G - V_{T0}}{n} = V_{S\,(D)} + 2 \cdot U_T \cdot ln\left(\exp\left[\sqrt{i_{f\,(r)}} \right] - 1 \right) \tag{32}$$

where $i_f \equiv I_F / I_S$ and $i_r \equiv I_R / I_S$ are the normalized forward and reverse currents.

Table 2.2 Drain current in weak inversion

MODE	DRAIN CURRENT EXPRESSION	CONDITIONS
CONDUCTION	$I_S \cdot e^{\frac{V_P}{U_T}} \cdot \left[e^{\frac{-V_S}{U_T}} - e^{\frac{-V_D}{U_T}} \right]$ $= I_S \cdot e^{\frac{V_P - V_S}{U_T}} \cdot \left[1 - e^{-\frac{V_D - V_S}{U_T}} \right]$ $\cong I_{D0} \cdot e^{\frac{V_G}{n \cdot U_T}} \cdot \left[e^{\frac{-V_S}{U_T}} - e^{\frac{-V_D}{U_T}} \right]$ $= I_{D0} \cdot e^{\frac{V_G - n \cdot V_S}{n \cdot U_T}} \cdot \left[1 - e^{-\frac{V_D - V_S}{U_T}} \right]$	$V_S > V_P$ $V_D > V_P$ $I_{D0} \equiv I_S \cdot \exp\left[\frac{-V_{T0}}{n \cdot U_T} \right]$
FORWARD SATURATION	$I_F = I_S \cdot e^{\frac{V_P - V_S}{U_T}} \cong I_{D0} \cdot e^{\frac{V_G - n \cdot V_S}{n \cdot U_T}}$	$V_S > V_P$ $V_D > V_P$ $V_D - V_S >> U_T$
REVERSE SATURATION	$-I_R = -I_S \cdot e^{\frac{V_P - V_D}{U_T}} \cong -I_{D0} \cdot e^{\frac{V_G - n \cdot V_D}{n \cdot U_T}}$	$V_S > V_P$ $V_D > V_P$ $V_D - V_S >> U_T$
BLOCKED	$I_F = I_R \Rightarrow I_D = 0$	$V_S >> V_P$ or $V_S = V_D$ $V_D >> V_P$

The normalized forward current i_f is also called the *inversion factor* (or inversion coefficient), since it defines the inversion level of the transistor. It is therefore much larger than 1 for a transistor biased in strong inversion, while i_f much smaller than 1 corresponds to a transistor biased in weak inversion. For i_f around 1, the transistor operates in the region of moderate inversion.

1.2.3.2 Small-signal model

The small-signal transconductances of a MOST biased in conduction are defined using the voltages referred to the substrate:

$$g_{mg} \equiv \left.\frac{\partial I_D}{\partial V_G}\right|_{V_S,\, V_D} \qquad g_{ms} \equiv \left.\frac{\partial I_D}{\partial V_S}\right|_{V_G,\, V_D} \qquad g_{md} \equiv \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G,\, V_S} \tag{33}$$

where g_{mg}, g_{ms} and g_{md} are respectively the *gate*, *source* and *drain transconductances* of the model with voltages referred to the substrate. It is important to note that since the variation of the pinch-off voltage is n times smaller than the corresponding gate voltage variation and the forward current only depends on the voltage difference $V_P - V_S$, the gate transconductance in saturation (i.e. for $I_R = 0$) is n times smaller than the source transconductance [14][16][19]:

$$g_{mg} = g_{ms}/n \tag{34}$$

This relation is valid in saturation from weak to strong inversion.

Table 2.3 Transconductances in strong and in weak inversion

	STRONG INVERSION		**WEAK INVERSION**
	conduction	**forward saturation**	
g_{mg}	$\beta \cdot (V_D - V_S)$	$\beta \cdot (V_P - V_S) = \sqrt{\frac{2 \cdot \beta \cdot I_D}{n}}$ $= \frac{2I_D}{n \cdot (V_P - V_S)}$ $= \frac{2I_D}{V_G - V_{T0} - n \cdot V_S}$	$\frac{I_D}{n \cdot U_T}$
g_{ms}	$n \cdot \beta \cdot (V_P - V_S) = \sqrt{2 \cdot n \cdot \beta \cdot I_F}$ $= \frac{2I_F}{V_P - V_S} = \frac{2 \cdot n \cdot I_F}{V_G - V_{T0} - n \cdot V_S}$		$\frac{I_F}{U_T}$
g_{md}	$n \cdot \beta \cdot (V_P - V_D) = \sqrt{2 \cdot n \cdot \beta \cdot I_R}$ $= \frac{2I_R}{V_P - V_D}$ $= \frac{2 \cdot n \cdot I_R}{V_G - V_{T0} - n \cdot V_D}$	0	$\frac{I_R}{U_T}$

The value of the transconductances in strong and in weak inversion can be calculated respectively from (25) and (26). They are summarized in Table 2.3. Source and gate transconductances in saturation are proportional to the drain current when the transistor is biased in weak inversion and proportional to the square root of the drain current in strong inversion.

In forward saturation, the drain transconductance becomes negligible compared to the output conductance g_{ds} due to the channel length modulation. It is important to note that although the gate transconductance g_m and output conductance g_{ds} of the model having the voltages referred to the source are respectively equal to g_{mg} and g_{ds} of the model with voltages referred to the substrate, the transconductance from the bulk is related to g_{mg} , g_{ms} and g_{ds} according to [16]:

$$g_{mb} \equiv \left.\frac{\partial I_D}{\partial V_{BS}}\right|_{V_{GS},\, V_{DS}} = g_{ms} - g_{mg} - g_{ds} \cong g_{ms} - g_{mg} = (n-1)\, g_{mg} = \frac{n-1}{n} g_{ms} \quad (35)$$

Fig. 2.7 shows the gate and source transconductances in saturation normalized to their value in weak inversion versus the inversion coefficient. As mentioned earlier, the strong inversion asymptote crosses the weak inversion asymptote (=1) at a saturation current equal to the specific current. At this particular point the error made on the transconductance estimation when using the weak inversion expression of Table 2.3 reaches its maximum which is about 40 %. This is clearly unacceptable for doing analog design. Furthermore, transistors are very often biased in this moderate inversion region since it is a good compromise between current efficiency and gate area. A better estimation of the transconductance in the moderate inversion region is obtained by differentiating the large-signal drain current interpolation given by (31):

$$G(i_f) \equiv \frac{g_{mg} \cdot nU_T}{I_F} = \frac{g_{ms} \cdot U_T}{I_F} \cong \frac{1 - e^{-\sqrt{i_f}}}{\sqrt{i_f}} \quad (36)$$

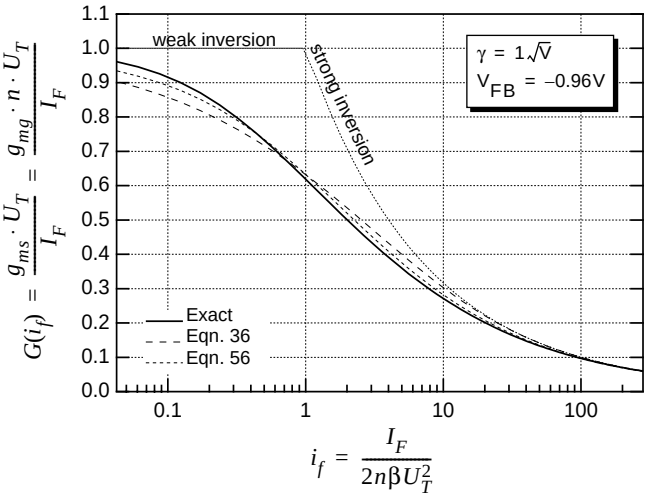


Figure 2.7 Transconductance interpolation functions

This interpolation function and the exact result obtained from a numerical computation are plotted in Fig. 2.7. The simple approximation given by (36) tends to overestimate the exact result in strong inversion, while it is slightly pessimistic in weak inversion. The maximum approximation error is now reduced from 40 % to typically 5 %, which is sufficient for hand calculations. Of course for computer simulation more accurate small-signal interpolation function are required. Although simple small-signal interpolation functions can be found, they have to be integrated to obtain the corresponding large-signal interpolation function used for the current. This generally leads to relatively complicated expressions which cannot always be inverted in order to express the current in terms of the voltage as required by the simulator. This has little consequence in the case of the computer simulation model, since the small- and the large-signal interpolation functions are independent of the process parameters and of the bias and can therefore be computed once and tabulated. Furthermore, in some circuit simulators the derivatives are calculated numerically instead of using an analytic expression. In such a case, the use of a large-signal interpolation function obtained by integration of the small-signal function ensures not only an accurate estimation of the operating point but also of the corresponding transconductances. This is the approach which has been taken for developing the EKV model (see Section 1.2.4).

1.2.3.3 Noise model

As shown in Fig. 2.8 (a), the noise of the MOS transistor at low-frequency can be modeled by two independent sources of random noise: a noisy current source connected between drain and source representing the thermal noise due to the noisy channel, and a noisy voltage source in series with the gate and modeling the flicker (or 1/f) noise related to the Si-SiO_2 interface.

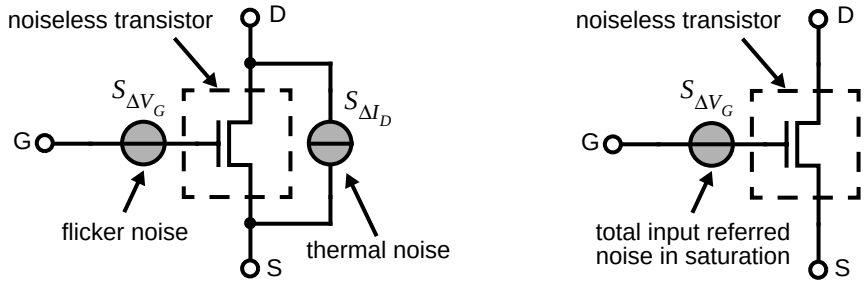


Figure 2.8 Model of the noisy transistor at low frequency

A general expression for the thermal noise can be derived by first considering the noise produced by a single elementary piece of the conductive channel and assuming the rest of the channel is noiseless (c.f. Fig. 2.9).

Like any resistor, this elementary piece of resistive channel produces a local fluctuation of the channel voltage dV_{ch} having a Power Spectral Density (PSD):

$$dS_{\Delta V_{ch}} = 4kT \cdot dR \tag{37}$$

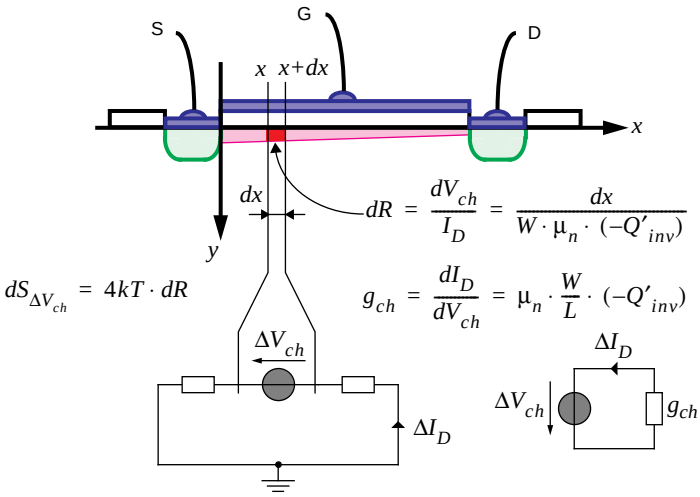


Figure 2.9 Two transistor model of the channel thermal noise

where dR is resistance of the elementary channel section which can be derived from (15):

$$dR = \frac{dV_{ch}}{I_D} = \frac{dx}{W \cdot \mu_n \cdot (-Q'_{inv})} \tag{38}$$

The channel voltage fluctuation results in a variation of the drain current having a PSD given by:

$$dS_{\Delta I_D} = g_{ch}^2 \cdot dS_{\Delta V_{ch}} = 4kT \cdot \frac{\mu_n}{L^2} \cdot W \cdot (-Q'_{inv}) \cdot dx \tag{39}$$

where g_{ch} is the channel conductance at point x in the channel:

$$g_{ch} = \frac{dI_D}{dV_{ch}} = \mu_n \cdot \frac{W}{L} \cdot (-Q'_{inv}) \tag{40}$$

Integrating (39) from source to drain gives the total PSD of the drain current fluctuations, which can be expressed in terms of a thermal noise conductance G_{Nth}:

$$S_{\Delta I_D} = 4kT \cdot G_{Nth} \tag{41}$$

$$G_{Nth} = \frac{\mu_n}{L^2} \cdot W \cdot \int_0^L -Q'_{inv} \cdot dx = \frac{\mu_n}{L^2} \cdot |Q_{inv}| \tag{42}$$

According to (42), the current fluctuation PSD is proportional to the total charge "stored" in the channel. Since no assumption has been made on the mode of operation of the transistor, (42) is valid from weak to strong inversion including in moderate inversion. The total charge in the channel is directly related to the source and drain

transconductances. The thermal noise conductance can thus be written in terms of g_{ms} and g_{md} according to:

$$G_{Nth} = \begin{cases} \frac{1}{2} \cdot (g_{ms} + g_{md}) = \frac{g_{ms}}{2} \cdot (1 + i_r/i_f) & \textit{weak inv.} \\ \frac{2}{3} \cdot \frac{g_{ms}^2 + g_{ms}g_{md} + g_{md}^2}{g_{ms} + g_{md}} = \frac{2}{3} \cdot g_{ms} \cdot \frac{1 + \sqrt{i_r/i_f} + i_r/i_f}{1 + i_r/i_f} & \textit{strong inv.} \end{cases} \quad (43)$$

which can be related to the bias current or voltages using the expressions given in Table 2.3. For a transistor operating as a conductance (i.e. for $V_D = V_S$ or equivalently for $i_f = i_r$), the source and drain transconductances are both equal to the channel "on" conductance G_{on} and therefore (43) reduces to:

$$G_{Nth} = g_{ms} = G_{on} \qquad \textit{for:} \qquad i_f = i_r \qquad \textit{or:} \qquad V_D = V_S \quad (44)$$

in weak and in strong inversion. It should be noted that even though the drain current is zero for $V_D = V_S$, g_{ms} is not equal to zero since it depends on the forward current which for $V_D = V_S$ is equal to the reverse current.

It is also interesting to note that in weak inversion the source and drain transconductances can be replaced by their respective expression given in Table 2.3 resulting in the following current noise PSD:

$$S_{\Delta I_D} = 4kT \cdot \frac{1}{2} \cdot (g_{ms} + g_{md}) = 2kT \cdot \left(\frac{I_F}{U_T} + \frac{I_R}{U_T} \right) = 2 \cdot q \cdot (I_F + I_R) \quad (45)$$

which corresponds to full shot noise of *both the forward and the reverse components* [34][35].

Since in saturation $g_{md} = 0$ (or $i_r = 0$), the total charge in the channel only depends on the source transconductance. The thermal noise conductance is then equal to:

$$G_{Nth} = \begin{cases} \frac{1}{2} \cdot g_{ms} & \textit{weak inversion} \\ & \textit{and saturation} \\ \frac{2}{3} \cdot g_{ms} & \textit{strong inversion} \end{cases} \quad (46)$$

In saturation $g_{mg} \neq 0$ and the thermal noise conductance can be conveniently referred to the gate as a thermal noise resistance by dividing by the square of the gate transconductance:

$$R_{Nth} = \frac{G_{Nth}}{g_{mg}^2} = \begin{cases} \frac{n}{2 \cdot g_{mg}} & \textit{weak inversion} \\ & \textit{and saturation} \\ \frac{2}{3} \cdot \frac{n}{g_{mg}} & \textit{strong inversion} \end{cases} \quad (47)$$

Since the noise current PSD of a transconductor is always proportional to the transconductance, the noise performance of this transconductor is conveniently expressed in terms of the noise factor γ defined as the product of the input referred noise resistance by the transconductance:

$$\gamma \equiv g_{mg} \cdot R_{Nth} = \frac{G_{Nth}}{g_{mg}} = \begin{cases} \frac{2 \cdot n}{3} \cong 1 & \textit{weak inversion} \\ \frac{n}{2} & \textit{strong inversion} \end{cases} \quad \textit{and saturation} \qquad (48)$$

In addition to the thermal noise, the MOS transistor is also strongly affected by low-frequency 1/f noise due to the fluctuation of the carrier density caused by trapping of carriers in traps located in the oxide and close to the Si-SiO2 interface via tunneling effect and fast surface states. Several theories exist that give rise to different expressions of the PSD and different relations to the bias. Nevertheless, all these theories agree on the fact that the PSD referred to the gate varies nearly in inverse proportion to frequency and to the gate area:

$$S_{V_{flicker}} = 4kT \cdot \frac{\rho}{W \cdot L \cdot f} \qquad (49)$$

where:

$$\rho \approx \textit{constant} \quad \left(\frac{A \cdot s}{m}\right)^2 \qquad (50)$$

Factor ρ is strongly process dependent and the p-channel very often has a smaller value than the n-channel (the ratio can be as high as 100). In general the ρ factor depends on the bias condition and may also depend on temperature. Measurements have shown that the bias dependence is weak [19]. The ρ factor can thus be considered as constant for a given temperature.

In saturation it is often convenient to refer the total noise to the gate and characterize it with a frequency dependent noise resistor $R_N(f)$ defined as:

$$S_{\Delta V_G} = 4kT \cdot R_N(f) \qquad \textit{with: } R_N(f) = \frac{\rho}{WL \cdot f} + \frac{\gamma}{g_{mg}} \ (\textit{in saturation}) \qquad (51)$$

where γ / g_{mg} is given by (47).

According to (51), the value of the total input referred noise at a given frequency reduces down to the 1/f noise when increasing the transconductance and to the thermal noise when enlarging the gate area.

1.2.3.4 Mismatch

Like noise, the mismatch between two identical transistors M_1-M_2 must be characterized by two statistical parameters: the threshold voltage mismatch $\delta V_{T0} \equiv V_{T01} - V_{T02}$, having a standard deviation σ_{VT}, and $\delta\beta/\beta \equiv (\beta_1 - \beta_2)/\beta$ having a standard deviation σ_β. In a well designed layout, each of these mismatch components have a mean value very close to zero and are usually very weakly correlated. They are typically ranging from 2 to 20 mV for δV_{T0} and 0.2 to 20 % for $\delta\beta/\beta$. Assuming σ_{VT} and σ_β are uncorrelated, the standard deviation σ_{ID} of the drain current mismatch $\delta I_D / I_D$ of two transistors having the same gate and source voltages (like for example a simple current mirror) is given by:

$$\sigma_{ID} = \sqrt{\sigma_\beta^2 + \left(\frac{g_{mg}}{I_D} \cdot \sigma_{VT}\right)^2} \qquad (52)$$

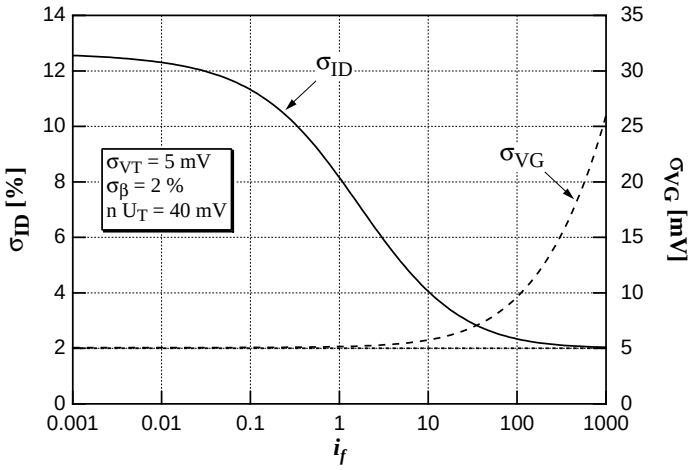


Figure 2.10 Matching of drain current (left axis) and gate voltage (right axis) as functions of the inversion coefficients of the transistors.

whereas the standard deviation σ_{VG} of the gate voltage mismatch $\delta V_G \equiv V_{G1} - V_{G2}$ of two transistors biased at the same drain current and having the same source voltage (like for example a differential pair) is given by:

$$\sigma_{VG} = \sqrt{\sigma_{VT}^2 + \left(\frac{I_D}{g_{mg}} \cdot \sigma_\beta\right)^2} \tag{53}$$

These results are plotted in Fig. 2.10 for $\sigma_{VT} = 5\ mV$ and $\sigma_\beta = 2\ \%$, by using the continuous transconductance shown in Fig. 2.7 with $nU_T = 40\ mV$.

It can be seen from Fig. 2.10 and from equations (52) and (53) that biasing the transistor in weak inversion (i.e. imposing $i_f << 1$ and thus a large g_{mg}/I_D ratio), results on one hand in a very poor drain current matching:

$$\sigma_{ID} \cong \frac{g_{mg}}{I_D} \cdot \sigma_{VT} \cong \frac{\sigma_{VT}}{nU_T} \qquad \textit{in weak inversion} \tag{54}$$

and on the other hand in a minimum mismatch between the gate voltages, equal to the threshold voltage mismatch σ_{VT}. A very large inversion factor is required to reduce σ_{ID} to the minimum possible value set by σ_β.

Weak inversion is thus not appropriate when current matching is required like for example in current mirrors, but it is beneficial for voltage matching like for example in a differential pair, the offset of which can be reduced to its minimum value set by the threshold voltage mismatch.

1.2.4 The EKV compact MOST model for circuit simulation

1.2.4.1 From hand calculation to circuit simulation

The development of the EKV compact MOST model for circuit simulation started at the end of the 80s on the basis of a strong experience in low-power CMOS analog circuit design acquired over the years at CSEM [32][33] and at EPFL [19]. It was primarily motivated by the lack of simulation models able to cover correctly the full range of operation modes the MOST can provide and particularly the moderate and weak inversion regions. This gap artificially limited the circuit design creativity that could be expected by exploiting all these modes. Although at this time many analog circuits did not require any simulation, their complexity increased rapidly, up to a point where it was essential to be able to correctly simulate them.

Prior to simulation, the designer has to correctly size each transistor in order to obtain the desired performance. Very often the sizing procedure is the result of a compromise between several design parameters such as current, overdrive or saturation voltage, transconductance (or noise), output conductance and gate area. The asymptotic weak and strong inversion models presented previously are clearly not sufficient to find an adequate solution to this multivariable problem. It requires a simple analytical model valid in all regions of operation. Among the design parameters mentioned above, probably the most important are the bias current, the transconductance and the inversion factor. As already mentioned, the use of the asymptotic models could lead to an unacceptable error when estimating the transconductances in the moderate inversion region. This has been corrected by the small-signal interpolation function $G(i_f)$ defined for example by (36). The effective transconductance in saturation can easily be estimating from the bias current $I_F \cong I_D$ and the inversion coefficient i_f:

$$g_{mg} = \frac{I_F}{nU_T} \cdot G(i_f) \qquad g_{ms} = \frac{I_F}{U_T} \cdot G(i_f) \tag{55}$$

Although (36) gives a sufficient accuracy, it is not convenient since it cannot be solved to find the inversion factor as a function of the required transconductance and the fixed bias. A simpler and more accurate interpolation function is given by [14][16][19]:

$$G(i_f) = \frac{g_{mg} \cdot nU_T}{I_F} = \frac{g_{ms} \cdot U_T}{I_F} = \frac{1}{\sqrt{i_f + \frac{1}{2} \cdot \sqrt{i_f} + 1}} \tag{56}$$

which is also plotted in Fig. 2.7. This equation can be conveniently solved to express the inversion factor in terms of the fixed bias current and the targeted transconductance:

$$i_f = \left(\sqrt{\frac{1}{16} + \left(\frac{1}{G^2} - 1 \right)} - \frac{1}{4} \right)^2 \tag{57}$$

A dedicated calculator shown in Fig. 2.11 has been developed to help the designer to properly size a single transistor according to given specifications and to explore the multivariable space of solution for his particular problem in order to find a satisfactory compromise.

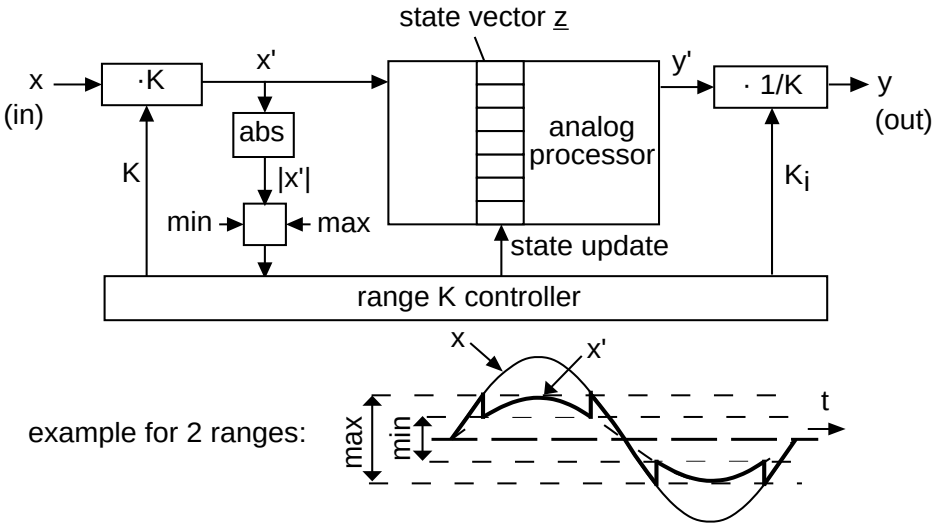


Figure 2.11 MOS transistor sizing calculator

The design variables have been divided in five groups named specifiers: the voltage specifier (including i_f, G, $V_{GS}-V_T$, V_P-V_S), the current specifier, the transconductance specifier (including $g_{ms}$and g_{mg}) and the W/L specifier (including W_{eff}/L_{eff} and β). The first axis (in log scale) of each group is the master axis from which the other axes are calculated resulting in curious scales. The cursors can be moved preserving the coherence between the variables of each specifier group. In order to solve a given sizing problem, two among the five specifiers have to be fixed while the other can be moved to explore the solution space until a good compromise is found. Additional informations such as gate area, parasitic capacitances and thermal noise resistance are monitored to help the designer adjusting his choice.

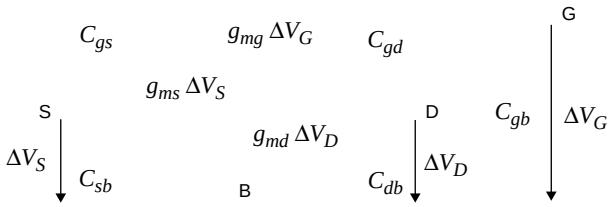
The next step of the design is the simulation of the circuit and its related problems. Probably the best compilation of MOS modelling problems to date has been presented in the work of Y. Tsividis [20]-[22] and completed by G. Machado in a discussion of some non-technical but fundamental issues in [23]. The basic requirements a MOS model should fulfil are well described in [22] and will not be reproduced here. The EKV model meets most of these criteria and fulfils most of the benchmarks defined in [22]. It is important to note that in addition to these requirements, the EKV model is also hierarchically structured, presenting several coherent levels, from simple analytical expressions to support creative synthesis, to detailed expressions for precise computer simulation. This allows the designers to understand the operation of the MOS transistor and then to correctly exploit and master its numerous characteristics in order to develop new high performance analog circuits. The latter aspect is also fundamental from an educational point of view [23].

Table 2.4 EKV MOST intrinsic model parameters [17]

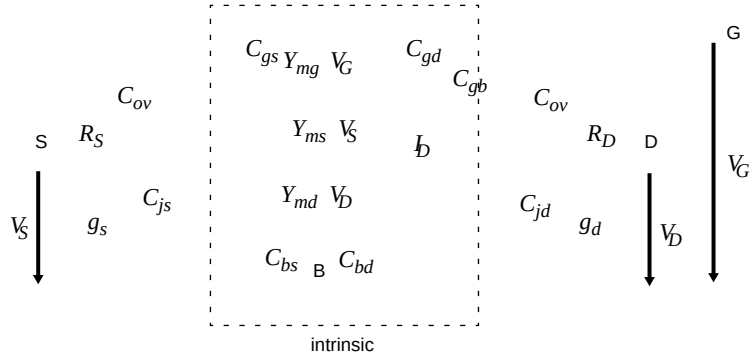
NAME	SYMBOL	DESCRIPTION
COX	C'_{ox}	gate oxide capacitance per unit area
VTO	V_{T0}	nominal threshold voltage
GAMMA	γ	body effect parameter
PHI	ψ_0	bulk Fermi potential (*2)
KP	$\mu \cdot C'_{ox}$	transconductance parameter
THETA	θ	mobility reduction coefficient due to vertical field
UCRIT	E_c	longitudinal critical field (velocity saturation)
DW	ΔW	channel narrowing
DL	ΔL	channel shortening
LAMBDA	-	depletion length coefficient
XJ	x_j	junction depth
WETA	-	narrow channel effect coefficient
LETA	-	short channel effect coefficient

The EKV compact MOST model for circuit simulation is based on the static long-channel theory presented in Section 1.2.3, but it also includes all the main second-order effects such as: channel length modulation, mobility reduction due to vertical field, velocity saturation, impact ionization and short- and narrow-channel effects [16][17][18][39]. One of the strong advantage of the EKV model in comparison to other models is its few number of intrinsic parameters which are limited to only 13 compared with typically more than 50 for other models. Most of the EKV model intrinsic parameters listed in Table 2.4 have their traditional SPICE meaning. In addition to these parameters, the model also includes 3 temperature parameters (TCV, BEX, UCEX) and 2 noise parameters (KF, AF) used to specify the 1/f noise. Like in any other models, the EKV model has also its own extrinsic model which is fully described by a set of approximately 20 parameters.

The dynamic model includes both a quasi-static and a first-order non-quasi-static model. The intrinsic part of the quasi-static small-signal model is shown in Fig. 2.12 (a). Since it only includes five intrinsic capacitances, this model does not take into account the transcapacitances [11]. A more elaborated high-frequency model that includes both the intrinsic and extrinsic part of the device is presented in Fig. 2.12 (b). The intrinsic part corresponds in fact to the first-order non-quasi-static model, where the transconductances of the circuit shown in Fig. 2.12 (a) have been replaced by first-order transadmittances. Unlike the model of Fig. 2.12 (a), it can be shown that the complete model of Fig. 2.12 (b) takes into account the effect of the transcapacitances which are included into the transadmittances. Since it is a non-quasi-static model, it gives a correct prediction of the transadmittance magnitude at high frequency unlike other models that include quasi-static transcapacitances [37].



a) Quasi-static intrinsic model



b) Complete high-frequency small-signal model including non-quasi-static transadmittances

Figure 2.12 EKV small-signal dynamic models

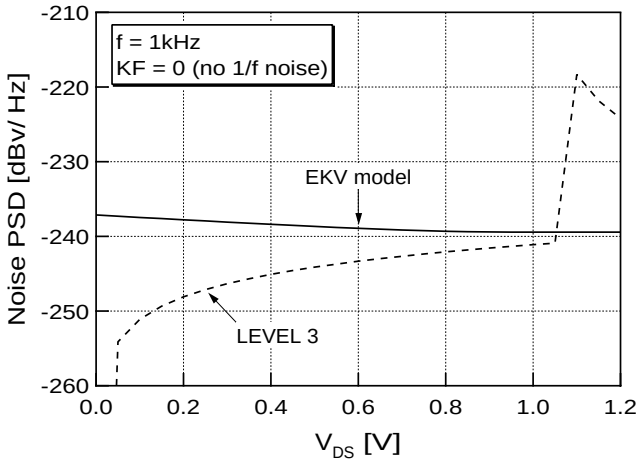


Figure 2.13 Thermal noise PSD versus V_{DS} voltage for the LEVEL 3 and the EKV model

It is worth mentioning that the thermal noise in most SPICE simulators is modelled as a noisy current source having a PSD given by:

$$S_{\Delta I_D} = 4kT \cdot \frac{2}{3} \cdot g_m \tag{58}$$

Since the gate transconductance is proportional to V_{DS} in the linear region, according to (58) the thermal noise of the device becomes zero for $V_D = V_S$, which is obviously unphysical! This is clearly illustrated in Fig. 2.13 where the thermal noise PSD at 1 kHz has been plotted versus the V_{DS} voltage for the UCB LEVEL 3 and the EKV model [16]. Fig. 2.13 additionally shows a discontinuity in the thermal noise PSD, the origin of which lies in a discontinuity of the gate transconductance at the limit between conduction and saturation. This problem has been solved in the EKV model by properly interpolating the thermal noise PSD from weak to strong inversion and from conduction to saturation, insuring a correct value in the linear region and smooth transitions between the different modes of operation as shown in Fig. 2.13 [14][16][19].

1.2.5 Examples of robust circuits illustrating the use of the EKV model

This section aims at illustrating the effective use of the model for the analysis and synthesis of some very simple but fundamental circuits. It should help the reader to get familiarized with the abstract concepts of pinch-off voltage, inversion factor, specific current through some circuit examples. The section does not cover all the well-known basic building blocks which are already described in details in some other references [26][27][67].

1.2.5.1 Pinch-off voltage extraction

The pinch-off voltage is not only a convenient concept established for the development of the EKV MOST model, but it can be measured or extracted and used to properly bias some circuits. Since the pinch-off voltage can be interpreted as the effect of the gate voltage referred to the channel, it should be possible to measure it for example at the source end of the device. Considering the expression of the drain current in saturation ($I_D = I_F$) and using the interpolation function given by (26) while setting $V_S = V_P$ leads to [16][38]-[40]:

$$I_D = I_S \cdot \left[\ln\left(1 + \exp\left[\frac{V_P - V_S}{2U_T} \right] \right) \right]^2 \overset{V_P = V_S}{=} I_S \cdot [\ln(2)]^2 \cong \frac{I_S}{2} = n \cdot \beta \cdot U_T^2 \tag{59}$$

The particular value of the drain current for which the source voltage is equal to the pinch-off voltage is thus equal to half the specific current. Hence, the pinch-off voltage can be measured by setting the drain current to $I_S/2$ in a transistor biased in saturation as shown in Fig. 2.14 (a). Since this circuit should provide the pinch-off voltage for any voltage applied to its gate, it can be used to measure the V_P versus V_G characteristic, corresponding to (16), by simply sweeping the gate voltage and measuring the source voltage. It is indeed used to extract the key parameters `VTO`, `GAMMA` and `PHI` used in the EKV model [16][38]-[40].

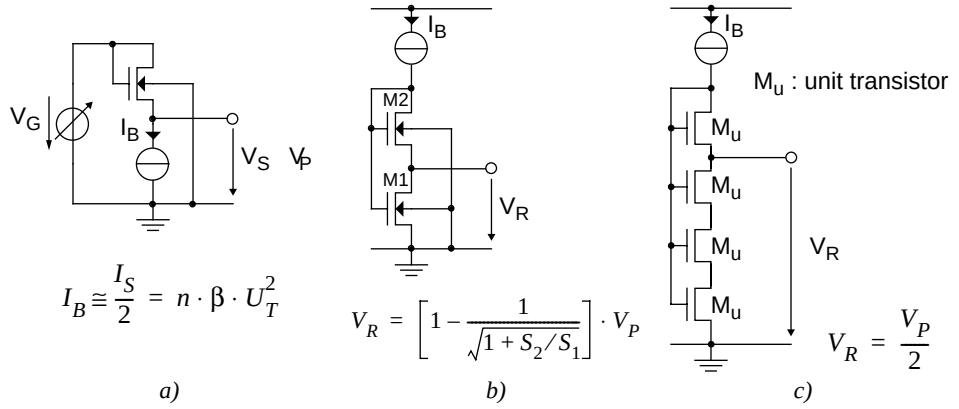


Figure 2.14 Pinch-off voltage extractors.

A fraction of the pinch-off voltage can be extracted using the circuit presented in Fig. 2.14 (b). In this case the biasing current I_B is not equal to $I_S/2$ but is such that both transistors operate in strong inversion. Since the gate of M_2 is connected to its drain terminal, M_2 operates in saturation, while M_1 is assumed to operate in conduction. Since both transistors M_1 and M_2 have the same gate voltage, they share the same pinch-off voltage V_P and the same slope factor n . Equating their drain currents yields:

$$I_{D1} = I_{D2} \Rightarrow \frac{n\beta_1}{2}(V_P - V_R)^2 = \frac{n\beta_2}{2}\left[V_P^2 - (V_P - V_R)^2\right] \tag{60}$$

The value of the reference voltage V_R between the drain of M_1 and ground is then given by solving (60):

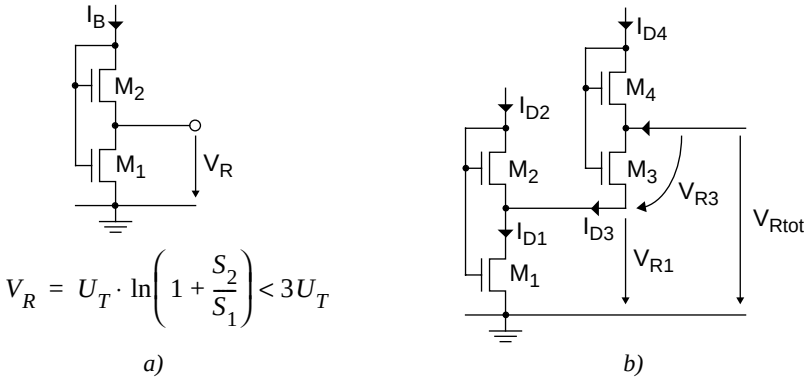
$$V_R = \left[1 - \frac{1}{\sqrt{1 + S_2/S_1}}\right] \cdot V_P \tag{61}$$

where $S_i \equiv W_{eff\text{-}i}/L_{eff\text{-}i}$ for $i = 1, 2$ are the aspect ratio of M_1 and M_2 respectively. As will be shown latter, it is convenient to implement the ratio S_2/S_1 by using series association of several identical "unit" transistors [42]. For example, M_1 and M_2 can be implemented by four stacked "unit" transistors in the same well as shown in Fig. 2.14 (c). M_1 and M_2 will therefore have the same width (the width of one unit transistor) and if V_R is tapped at the source of the upper transistor, then $S_2 = 3S_1$ and thus $V_R = V_P/2$.

The circuit shown in Fig. 2.14 (b) can also be used in weak inversion to realize a simple PTAT voltage reference [30][31]. M_1 is again assumed to be in conduction, whereas M_2 is in saturation. Equating the drain currents using the current expression given in Table 2.2 and solving for V_R results in:

$$V_R = U_T \cdot \ln\left(1 + \frac{S_2}{S_1}\right) \tag{62}$$

This circuit is only useful for realizing PTAT voltages smaller than typically $3U_T$ due to the logarithmic dependence of V_R on S_2/S_1. Larger PTAT voltages can be achieved by simply stacking the same two transistor scheme and trading β ratios against



$$V_R = U_T \cdot \ln\left(1 + \frac{S_2}{S_1}\right) < 3U_T$$

a) *b)*

Figure 2.15 PTAT voltage references in weak inversion [30]

current ratios as shown in Fig. 2.15 (b). The original S_2/S_1 ratio is then multiplied by a factor equal to the ratio of the currents flowing in the lower and respectively in the upper transistor:

$$V_{R1} = U_T \cdot \ln\left(1 + \frac{S_2}{S_1} \cdot \frac{I_{D1}}{I_{D2}}\right) \qquad V_{R3} = U_T \cdot \ln\left(1 + \frac{S_4}{S_3} \cdot \frac{I_{D3}}{I_{D4}}\right) \tag{63}$$

If $I_{D2} = I_{D4} = I$ and $S_2/S_1 = S_4/S_3 = \alpha$, the stacked voltage V_{Rtot} is then equal to:

$$V_{Rtot} = U_T \cdot \ln\left[(1+\alpha)\,(1+2\alpha)\right] \tag{64}$$

which can be extended to$N$ stacked identical PTAT voltage sources driven by equal currents:

$$\frac{V_{Rtot}}{U_T} = N \cdot \ln(\alpha) + \ln\left[\Gamma(N + 1 + 1/\alpha)\right] - \ln\left[\Gamma(1 + 1/\alpha)\right] \tag{65}$$

where $\Gamma(x)$ is the Gamma function. Eqn. 65 is plotted in Fig. 2.16 which shows that a 1 V PTAT voltage source at room temperature requires 8 current branches with $\alpha = S_2/S_1 = 30$.

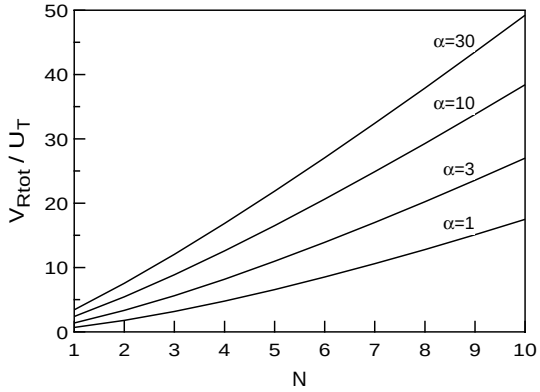


Figure 2.16 Voltage of N stacked PTAT sources

1.2.5.2 Specific current extractor and ratio-based design technique

In analog circuit design it is crucial to correctly set the operating point of a MOS transistor in order to precisely control its small-signal characteristics like for example its transconductances. The specific current represents a well defined operating point on the DC transfer characteristic of the MOS transistor which is just between weak and strong inversion in the middle of the so-called moderate inversion region. If a bias current equal to the specific current $I_{S\text{-}ref}$ of a reference transistor is available (say for example a square transistor having a unity aspect ratio $S_{ref} = 1$), any transistor M_x of aspect ratio S_x can then be operated at a given inversion factor i_{fx} by means of a weighted copy of this current $I_{S\text{-}ref}$ according to:

$$K = \frac{S_x \cdot i_{fx}}{S_{ref} \cdot i_{f\text{-}ref}} = S_x \cdot i_{fx} \tag{66}$$

since by assumption $S_{ref} = 1$ and by definition the inversion factor $i_{f\text{-}ref}$ of the reference transistor is equal to 1 if it is biased by a current equal to its specific current $I_{S\text{-}ref}$. The latter factor K can be implemented for example by a weighted current mirror. It is therefore convenient to generate a reference current equal to the specific current of a given reference transistor in order to precisely set the inversion factor of any transistor by using a proper series/parallel combination of this reference transistor [42]. This ratio-based design technique is very attractive since it is independent to the first-order of the temperature and of the technology [47].

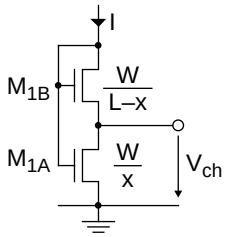


Figure 2.17 Tapped transistor

The specific current generator (or specific current extractor SCE [47]) is based on the stacked transistor circuit of Fig. 2.14 (b) which is investigated hereafter from an other point of view. Consider a single transistor of width W and length L where a tap has been inserted at a distance x from the source terminal in order to measure the channel voltage at this particular position. As shown in Fig. 2.17, it can be considered as two stacked transistors M_{1A} and M_{1B} having the same gate and substrate and the same width, but different length: x for the bottom transistor and $L - x$ for the top. The top transistor operates in saturation since it has its gate terminal connected to its drain, while the bottom transistor operates in conduction. The channel voltage at position x from the source can be deduced by using (32):

$$V_P = 2U_T \cdot \ln\left(\exp\left[\sqrt{i_{f1A}}\right] - 1 \right) \tag{67}$$

$$V_P - V_{ch} = 2U_T \cdot \ln\left(\exp\left[\sqrt{i_{f1B}}\right] - 1 \right) \tag{68}$$

where i_{f1A} and i_{f1B} are the inversion coefficients of transistor M_{1A} and M_{1B} respectively. Combining (67) and (68) leads to:

$$V_{ch} = 2U_T \cdot \ln\left(\frac{\exp\left[\sqrt{i_{f1A}}\right]-1}{\exp\left[\sqrt{i_{f1B}}\right]-1}\right) \tag{69}$$

It should be noted that:

$$i_{f1B} = i_{r1A} = \left[ln\left(1+\exp\left[\frac{V_P - V_{ch}}{2U_T}\right]\right)\right]^2 \tag{70}$$

since the source voltage of M_{1B} is equal to the drain voltage of M_{1A} . On the other hand, the inversion factor of the bottom transistor i_{f1A} is equal to the inversion factor of the overall transistor M_1 which will be defined as i_{f1} . Equating the currents flowing in M_{1A} and M_{1B} leads to:

$$I_{S1B} \cdot i_{f1B} = I_{S1A} \cdot (i_{f1A} - i_{r1A}) \qquad or: \qquad \frac{W}{L-x} \cdot i_{f1B} = \frac{W}{x} \cdot (i_{f1A} - i_{r1A}) \tag{71}$$

Introducing (70) into (71) allows to express the inversion coefficient of M_2 as a function of the inversion coefficient of M_1 and distance x:

$$i_{f1B} = \left(1 - \frac{x}{L}\right) \cdot i_{f1A} \tag{72}$$

which can be introduced into (69), resulting in:

$$V_{ch} = 2U_T \cdot \ln\left(\frac{\exp\left[\sqrt{i_{f1}}\right]-1}{\exp\left[\sqrt{(1-x/L)\cdot i_{f1}}\right]-1}\right) \tag{73}$$

where i_{f1A} has been replaced by i_{f1} since they are equal. Equation (73) gives the value of the channel voltage at a relative distance x/L from the source as a function of the inversion factor. The channel voltage normalized to U_T has been plotted versus x/L for different inversion coefficients in Fig. 2.18 (a) and versus i_{f1} for different relative distances in Fig. 2.18 (b).

Equation (73) is rather complicated, but it can be strongly simplified by considering only the weak and strong inversion asymptotes:

$$V_{ch} = \begin{cases} -U_T \cdot \ln\left(1 - \dfrac{x}{L}\right) & i_{f1} << 1 \quad (weak\ inv.) \\ 2U_T \cdot \sqrt{i_{f1}} \cdot \left(1 - \sqrt{1 - \dfrac{x}{L}}\right) = & \\ = V_P \cdot \left(1 - \sqrt{1 - \dfrac{x}{L}}\right) & i_{f1} >> 1 \quad (strong\ inv.) \end{cases} \tag{74}$$

According to (74), the channel voltage in weak inversion becomes independent of the inversion factor, which is consistent with the curves shown in Fig. 2.18 (b) for low i_{f1} , whereas it is proportional to the square-root of i_{f1} (or the pinch-off voltage) in strong inversion.

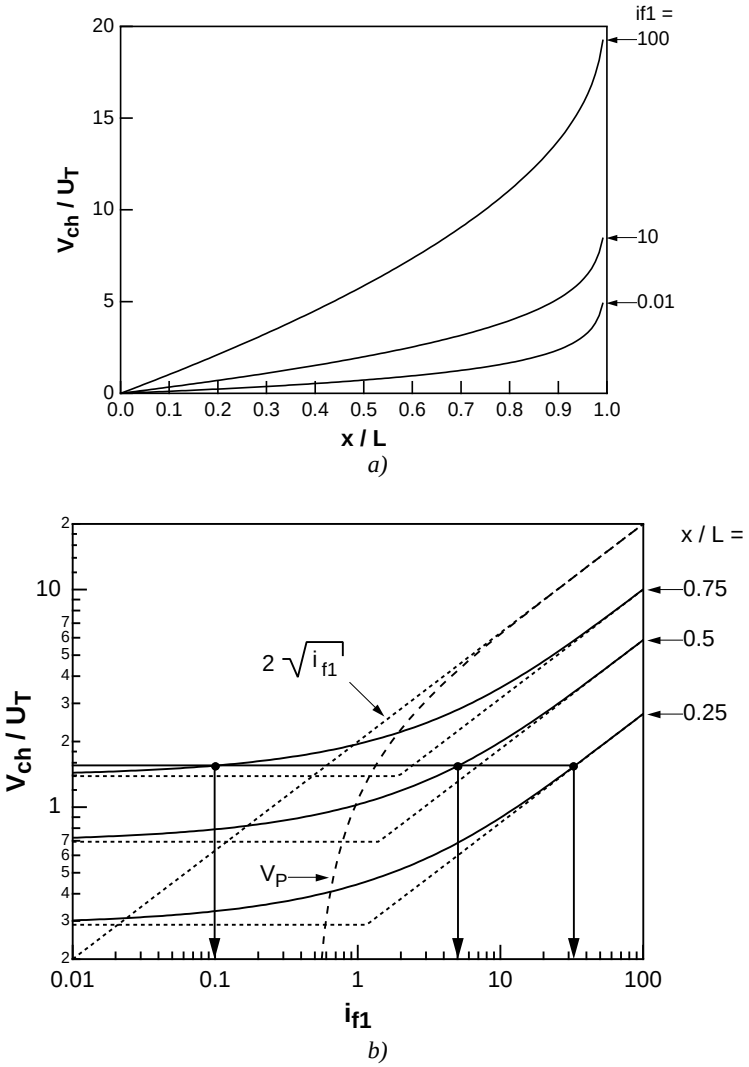


Figure 2.18 Channel voltage versus the distance of the tap from the source terminal of M_1 (a) and versus the inversion coefficient of M_1 (b).

Note that the two asymptotes all cross at a value of i_{f1} comprised between 1 and 2. The precise position of the tap can be calculated from (74) in order to extract a defined fraction of the pinch-off voltage. For example by setting $x/L = 3/4$ the tap voltage is exactly equal to half the pinch-off voltage.

Fig. 2.18 (b) also shows that two transistors tapped at different relative values can still have equal channel voltages provided they are biased at the two adequate but different inversion coefficients. This property can be used to generate a specific current by comparing the channel voltages of two differently tapped transistors. This principle is implemented by the circuit shown in Fig. 2.19, which is inspired from the well known current reference proposed in [31].

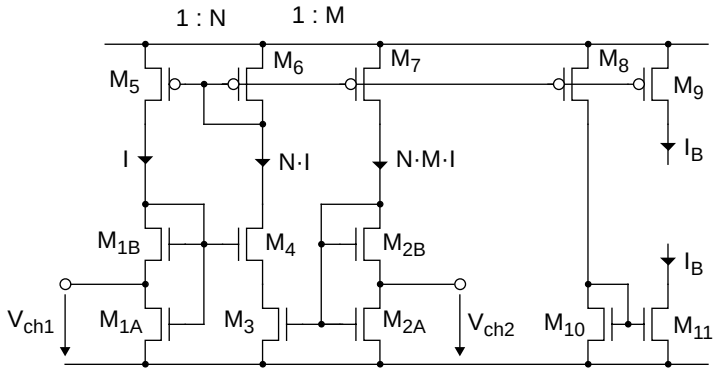


Figure 2.19 Specific current extractor (SCE) [47]

In the circuit of Fig. 2.19, the tapped transistor M_1 is biased at low current so that it operates in weak inversion, whereas the other tapped transistor M_2 is biased at a current $N \cdot M$ times larger so that it operates in strong inversion. The comparison between the two channel voltages V_{ch1} and V_{ch2} is made in the median branch at the drain of M_3. Since M_4 is made N times wider than M_{1B} its source voltage is equal to V_{ch1}, whereas the drain voltage of M_3, which is M times longer than M_{2A}, is equal to V_{ch2}. Connecting the source of M_4 to the drain of M_3 forces the two channel voltages to be equal:

$$-U_T \cdot \ln\left(1 - \frac{x_1}{L_1}\right) \cong 2U_T \cdot \sqrt{i_{f2}} \cdot \left(1 - \sqrt{1 - \frac{x_2}{L_2}}\right) \tag{75}$$

The inversion factor of M_2 can be calculated from (75) as a function of the relative tap positions of M_1 and M_2 and is plotted in Fig. 2.20.

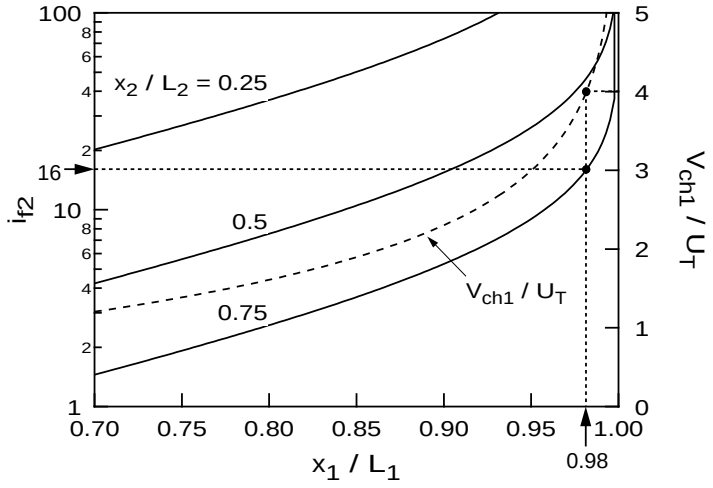


Figure 2.20 Inversion factor of M_2 versus the relative tap position of M_1 for different relative tap position of M_2

Fig. 2.20 also shows the value of the channel voltage V_{ch1} (and V_{ch2} since they are forced to be equal) as a function of x_1/L_1. This channel voltage should be made as large as possible to reduce the effect of threshold mismatch. This requires x_1/L_1 to be very close to 1. As an example, x_1/L_1 should be made equal to $53/54 \cong 0.98$ in order to obtain $V_{ch1} \cong 4U_T$. Choosing $x_2/L_2 = 0.75$ would lead to $i_{f2} = 16$. The inversion factor of M_1 is linked to that of M_2 by $I_{D2} = N \cdot M \cdot I_{D1}$ or:

$$i_{f1} = \frac{S_2}{S_1} \cdot \frac{i_{f2}}{N \cdot M} \tag{76}$$

Any n-channel transistor M_x of aspect ratio S_x can then be operated at a given inversion factor i_{fx} by means of a weighted copy of current I_{D2}. The weight can be set for example by the ratio of the aspect ratio of transistors M_8 and M_7 according to:

$$K = \frac{S_8}{S_7} = \frac{S_x \cdot i_{fx}}{S_2 \cdot i_{f2}} \tag{77}$$

The circuit of Fig. 2.19 can be improved on several aspects by cascoding and using the technique presented in Fig. 2.15 (b) to help implementing large channel voltage while preserving reasonable area [47].

1.2.5.3 The concept of pseudo-conductance and its application to MOS networks

Resistors are generally used to perform a voltage-to-current (or inversely a current-to-voltage) conversion. Since the sheet resistivities of the available layers that can be used to implement passive resistors are relatively low (typically 100 Ω/□ for the polysilicon layer), high-value resistors may be very area-consuming and are therefore prohibited. Moreover, passive resistors cannot be electrically tuned as it may be required for some applications. Passive resistors can thus be advantageously be replaced by MOS transistors operating in the conduction regime and having a sheet resistance given by:

$$R_{\square} = \frac{1}{n \cdot \mu_n C'_{ox} \cdot (V_P - V_S)} \cong \frac{1}{\mu_n C'_{ox} \cdot (V_G - V_{T0} - n \cdot V_S)} \tag{78}$$

which may reach several 10 kΩ/□ for a 1 volt $V_P - V_S$ voltage range. Since the value of the resistor is voltage dependent, dedicated techniques have to be used in order to improve their linearity [29]. An example of such a technique can be illustrated by looking to the expression of the current in strong inversion and in conduction given in Table 2.1 and note that if the source and drain voltages are balanced around a constant common mode voltage $\frac{1}{2}(V_D + V_S)$, the current becomes a linear function of the V_{DS} voltage. The other techniques exploit very similar properties and can be used for example to build MOSFET-C continuous-time filters [29].

Resistors are also required in resistive networks to split currents linearly among the network branches. They are typically used in analog neural networks [41] to perform a spatial averaging or in D/A converters to implement reference currents like for example in a R2R ladder. Such a function can also be realized using solely MOS transistors [15][44][43]. As mentioned in Section 1.2.3.1, the drain current is the difference of a forward current I_F depending only on $V_P - V_S$ through a function F and a reverse current

I_R depending only on $V_P - V_D$ but through the same function F:

$$I_D = I_S \cdot [F(V_P - V_S) - F(V_P - V_D)] \tag{79}$$

where $F(V_P - V)$ is the interpolation function linking weak to strong inversion and which can be approximated by the log function defined in (31). If function $F(V)$ would be a linear function $F(V_P - V) = -(V_P - V)/V_0$ where V_0 is an arbitrary scaling voltage, (79) would represent the simple Ohm's law with a conductance given by I_S/V_0. Unfortunately $F(V_P - V)$ is not linear with respect to V and therefore does not correspond to the original Ohm's law. Nevertheless, it can be interpreted as if it would be by defining transformed voltages or pseudo-voltages by $V^* = -V_0 \cdot F(V_P - V)$ [44]:

$$I_D = G^* \cdot \left(V_D^* - V_S^* \right) \tag{80}$$

which corresponds to a pseudo-Ohm's law with a corresponding pseudo-conductance defined as $G^* = I_S/V_0$ and which can be sized by adjusting the transistor's W/L ratio [44]. Linear current splitting can therefore be realized with networks of transistors interconnected by their source and drain terminals and sharing a common gate voltage. Since $F(V)$ is a positive function of V the pseudo voltage V^* cannot change sign but its zero-reference level (pseudo-ground) is reached as soon as the corresponding real voltage V is large enough to saturate the transistor [44]. Any current flowing to the pseudo-ground can therefore be easily mirrored by means of a complementary transistor operating in saturation [44].

Making all the transistors of the network operate in weak inversion has the advantage that the term depending on V_G can be factorized. The pseudo-voltages and pseudo-conductances can then be redefined by including the term depending on V_G [44]:

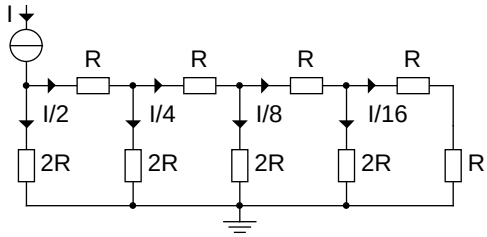
$$V^* = -V_0 \cdot \exp\left[\frac{-V}{U_T}\right] \tag{81}$$

and:

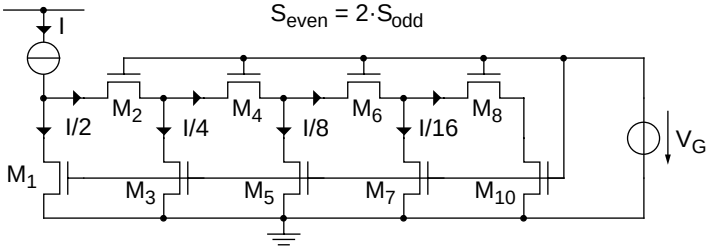
$$G^* = \frac{I_S}{V_0} \cdot \exp\left[\frac{V_P}{U_T}\right] \cong \frac{I_S}{V_0} \cdot \exp\left[\frac{V_G - V_{T0}}{nU_T}\right] \tag{82}$$

In addition to the individual sizing provided by the aspect ratio of each transistor (hidden in the specific current I_S), the pseudo-conductances of each transistor of a network can then be adjusted by controlling its gate voltage. Any network of linear resistors can thus be implemented by means of MOS transistors, with the additional possibility of electrically controlling the value of each equivalent resistor if the transistor is operating in weak inversion.

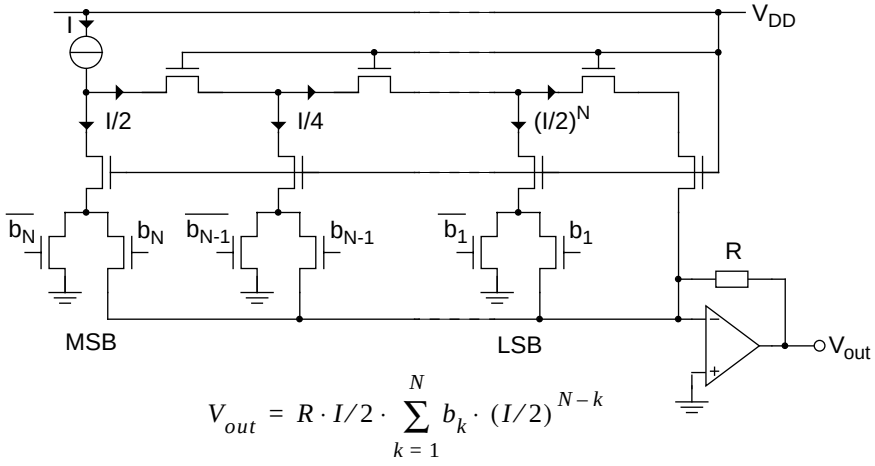
In particular, the previous concept can be applied to the R2R ladder shown in Fig. 2.21 (a) resulting in the MOST-only ladder depicted in Fig. 2.21 (b). In order to split the current equally at each node, the aspect ratios of the even transistors need to be twice that of the odd transistors, which can be simply realized either by connecting two unit transistors in parallel for the series branch and one for the shunt branch (or inversely one for the series branch and two in series for the shunt branch). It is interesting to mention that the current division is independent of the current and therefore of the mode of inversion of the MOSTs. The ladder has been used in a volume control circuit [15], but it can obviously also be used in a simple and compact D/A converter as depicted in Fig. 2.21 (c).



a) Original R2R ladder

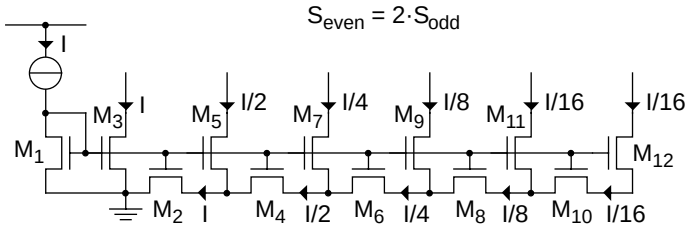


b) Transformed transistor only ladder (M2M ladder)



$$V_{out} = R \cdot I/2 \cdot \sum_{k=1}^{N} b_k \cdot (I/2)^{N-k}$$

c) Compact M2M ladder D/A converter



d) Binary weighted current sources

Figure 2.21 Linear current splitting using the pseudo-conductance concept

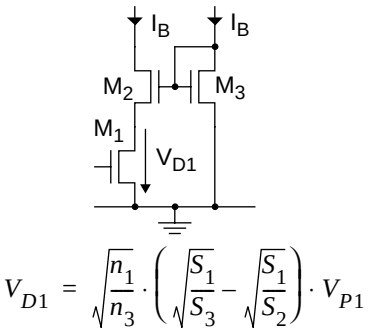
In this circuit, the lower transistors are doubled and simultaneously used as switches controlled by the input digital word [43]. The accuracy of such D/A converter is typically limited to 6 to 8 bits, due to mismatch and to the variation of the virtual ground imposed by the opamp. The digitally-controlled M2M ladder can also be used as a programmable V-to-I converter which can be used in MOSFET-C continuous-time filters [43].

The transistors forming the ladder are not necessarily required to be all biased in the linear region. Fig. 2.21 (d) shows an example of a string of binary weighted current sources using the same principle but where the transistors with even numbers are all in saturation.

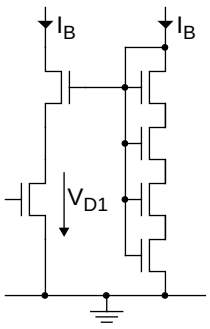
The performance of the latter circuits are of course limited by a number of second-order effects. Mismatch of device geometry and oxide thickness will only affect the accuracy of the division. Threshold voltage mismatch affects the linearity of current-division in strong inversion, but only its accuracy in weak inversion. The effects of channel-length modulation, velocity saturation and drain-induced barrier lowering are mainly affecting the devices when they are operating in saturation and should therefore not influence the ladder of Fig. 2.21 (b) if all transistors are operating in conduction.

1.2.5.4 Low-voltage cascode stage bias circuits

Low-voltage operation prevents the use of any stack of transistors. In particular, cascode stages must be implemented in such a way that the common-source device is biased just at the onset of saturation in order to preserve maximum voltage swing. A well-known low-voltage cascode stage bias circuit is shown in Fig. 2.22 (a) [27][45].



$$V_{D1} = \sqrt{\frac{n_1}{n_3}} \cdot \left(\sqrt{\frac{S_1}{S_3}} - \sqrt{\frac{S_1}{S_2}} \right) \cdot V_{P1}$$



all transistors are identical

$$V_{D1} \cong V_{P1}$$

a) *b)*

Figure 2.22 Low-voltage cascode stage bias circuit in strong inversion

All transistors of Fig. 2.22 (a) are assumed to be biased in strong inversion and in saturation. Equating the current through M_2 and M_3 and setting $V_{P3} = V_{P2}$ and $n_3 = n_2$ since M_2 and M_3 have the same gate voltage, results in an expression of the

drain voltage of M_1 :

$$V_{D1} = \left(1 - \sqrt{S_3/S_2}\right) \cdot V_{P2} \tag{83}$$

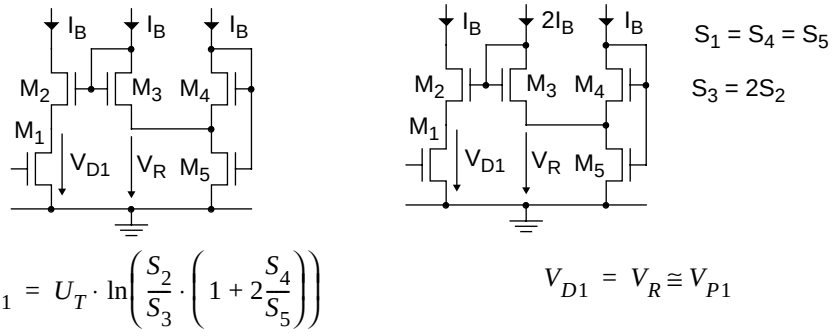
A relation between V_{P1} and V_{P2} can be found by equating the current through M_1 and M_3 :

$$\frac{n_3\beta_3}{2} \cdot V_{P3}^2 = \frac{n_1\beta_1}{2} \cdot V_{P1}^2 \Rightarrow V_{P2} = V_{P3} = \sqrt{\frac{n_1 S_1}{n_3 S_3}} \cdot V_{P1} \tag{84}$$

Substituting (84) into (83) leads to:

$$V_{D1} = \sqrt{\frac{n_1}{n_3}} \cdot \left(\sqrt{\frac{S_1}{S_3}} - \sqrt{\frac{S_1}{S_2}}\right) \cdot V_{P1} \cong \left(\sqrt{\frac{S_1}{S_3}} - \sqrt{\frac{S_1}{S_2}}\right) \cdot V_{P1} \tag{85}$$

which can be made equal to the minimum value necessary to maintain M_1 in saturation independently of the bias current I_B by setting $V_{D1} = V_{P1}$. This can be done for example by choosing M_2 identical to M_1 ($S_1 = S_2$) and implementing M_3 as a series connection of four transistors identical to M_1 (resulting in $S_3 = S_1/4$) [45]. Another possibility is to make S_2 much larger than S_1 so that the second term of (85) can be neglected and then to choose $S_1 = S_3$.



$$V_{D1} = U_T \cdot \ln\left(\frac{S_2}{S_3} \cdot \left(1 + 2\frac{S_4}{S_5}\right)\right) \qquad V_{D1} = V_R \cong V_{P1}$$

a) In weak inversion *b) In strong inversion*

Figure 2.23 Low-voltage cascode stage bias circuit

The latter result remains valid even if M_2 is in moderate or weak inversion. On the other hand, the simple circuit of Fig. 2.22 (a) can unfortunately not be used when both M_1 and M_2 are biased in weak inversion, since it would require a prohibitive ratio S_1/S_3 to obtain the 4 to $6U_T$ required to bring M_1 in saturation. This voltage must be provided by an external PTAT reference as the one described in Section 1.2.5.1. The complete circuit used to bias a cascode stage in weak inversion is drawn in Fig. 2.22 (a). The value of the reference voltage is deduced from (63) by setting $I_{D1}/I_{D2} = 2$:

$$V_R = U_T \cdot \ln\left(1 + 2\frac{S_4}{S_5}\right) \tag{86}$$

which is limited to a value typically smaller than $4U_T$. A larger value of V_R and hence of V_{D1} can be achieved by using the technique presented in Fig. 2.15 (b) or by choosing $S_2 >> S_3$ resulting in:

$$V_{D1} = U_T \cdot \ln\left(\frac{S_2}{S_3} \cdot \left(1 + 2\frac{S_4}{S_5}\right)\right) \tag{87}$$

For example, choosing $S_2/S_3 = S_4/S_5 = 8$ results in $V_R = V_{D1} \cong 5U_T \cong 130mV$, which is sufficient to maintain M_1 in saturation. The minimum value of the output voltage is then approximately $10U_T \cong 260mV$.

It is worth mentioning that the circuit of Fig. 2.22 (a) can also be used in strong inversion. The value of the reference voltage V_R is similar to (61) except for the additional factor 3 due to the current flowing through M_3 and absorbed by M_5 :

$$V_R = \left[1 - \frac{1}{\sqrt{1 + 3 \cdot S_4/S_5}}\right] \cdot V_{P5} \tag{88}$$

Choosing M_4 identical to M_5 leads simply to $V_R = V_{P4}/2 = V_{P5}/2$. On the other hand, M_1 and M_4 have the same drain current:

$$\frac{n_1\beta_1}{2}V_{P1}^2 = \frac{n_4\beta_4}{2}\left(V_{P4} - \frac{V_{P4}}{2}\right)^2 \Rightarrow V_{P5} = V_{P4} = 2 \cdot \sqrt{\frac{n_1 S_1}{n_4 S_4}} \cdot V_{P1} \tag{89}$$

and hence:

$$V_R = \frac{V_{P4}}{2} = \frac{V_{P5}}{2} = \sqrt{\frac{n_1 S_1}{n_4 S_4}} \cdot V_{P1} \tag{90}$$

The drain voltage of M_1 is imposed equal to V_{P1} by simply choosing M_1 identical to M_4 which corresponds to set $S_4 = S_1$ in (90). In the final circuit, M_1 , M_4 and M_5 are all identical, while M_3 is implemented by connecting two transistors identical to M_2 in parallel.

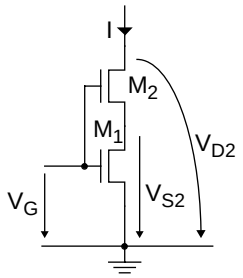


Figure 2.24 Low-voltage self-cascode stage [46]

The additional current branch used to bias the cascode stage can be saved by using the very compact and power efficient self-cascode scheme presented in Fig. 2.24 [46]. It can be shown that the output conductance of this combination of two transistors M_1 and M_2 having each an Early voltage V_{E1} and V_{E2} respectively, is given by:

$$g_o = g_{ds2} \cdot \frac{1 + g_{ds1}/g_{md1}}{1 + g_{ms2}/g_{md1}} = g_{ds2} \cdot \frac{1 + \xi \cdot g_{ds1}/g_{ms2}}{1 + \xi} \tag{91}$$

where $g_{dsi} = I/V_{Ei}$ for $i = 1, 2$ are the output conductances of M_1 and M_2 , g_{md1} is the drain transconductance of M_1 (in conduction) and g_{ms2} is the source

transconductance of M_2 (in saturation). Factor ξ is defined as the ratio of the forward current of M_2 to the reverse current of transistor M_1:

$$\xi \equiv \frac{g_{ms2}}{g_{md1}} = \frac{I_{F2}}{I_{R1}} = \frac{I}{I_{R1}} = \frac{I_{S2}}{I_{S1}} = \begin{cases} \frac{S_2}{S_1} \cdot \exp\left(\frac{V_{T01} - V_{T02}}{nU_T}\right) & \textit{in weak inv.} \\ \frac{S_2}{S_1} \cdot \left(\frac{V_G - V_{T02} - nV_{S2}}{V_G - V_{T01} - nV_{S2}}\right)^2 & \textit{in strong inv.} \end{cases} \quad (92)$$

For transistors having the same threshold voltages $V_{T01} = V_{T02}$ (neither short- nor narrow-channel effects), factor ξ reduces to S_2/S_1 in both weak and strong inversion. If S_2/S_1 is made much larger than 1, factor ξ also becomes much larger than 1 and the output conductance given by (91) tends to that of a traditional cascode stage:

$$g_o \cong g_{ds2} \cdot \frac{g_{ds1}}{g_{ms2}} \qquad \textit{for: } \xi >> 1 \quad (93)$$

This can be realized by choosing an equal channel length for both M_1 and M_2 and by setting the channel width W_2 much larger than W_1. Setting W_1 to the minimum width results in an increase of V_{T01} with respect to V_{T02} due to the narrow-channel effect on M_1 and consequently helps to further increase factor ξ significantly, while maintaining a reasonable S_2/S_1 ratio. Although this is true in both weak and strong inversion, it is really effective mainly in weak inversion thanks to the exponential term in (93).

There are of course other cascode bias circuits, but probably the most compact and current efficient is the one shown in Fig. 2.24. P. Heim proposed a cascode biasing circuit which has the interesting property that it can operate at any current level with a minimal output saturation voltage and independently of the technology since the design is based on ratios [48]. Unfortunately, this circuit requires a relatively large number of transistors and is therefore not suited for dynamic applications but rather to constant current which is in contradiction with its ability to work at any current level.

1.2.6 Some additional system considerations

1.2.6.1 General considerations

Power minimization must already be addressed at the system level. A first aspect is the management of the power delivered to the various blocks of a chip, which will be very important for digital sub-circuits. For analog blocks, voltage is not critically related to power, and power management can be limited to shutting off some functions when they are not needed, and to possibly multiplying (on- or off-chip) the supply voltage if it comes from a very low voltage source.

High-frequency operation tends to require a power much above the limits presented in Section 1.2.2, essentially because of the presence of parasitic capacitors. The architecture of low-power RF receivers should thus be selected to minimize the number of active devices operating at the carrier frequency. An extreme solution would be to directly sample the RF signal at a subharmonic of the usual local oscillator frequency. All the image bands produced by this undersampling process would have to be eliminated by passive SAW filters just after the antenna. Of course, the signal power available after mixing would be reduced by the undersampling factor. The sampling mixer should thus be

preceded by RF amplification stages.

Certain applications such as paging or the Global Positioning System (GPS) do not require full-time operation of the receiver. Indeed, depending on the protocol, the duty cycle may be drastically reduced, with a proportional reduction in power consumption. Commercially available watch pagers already operate continuously for 30 days on a small battery.

1.2.6.2 The analog floating point technique

For continuous operation, and even when high frequency is not an obstacle, it will never be possible to pass the limit given by (4), and at least ten times more power will probably be needed for practical reasons. This means that it will never be possible to realize a 16-bit audio A/D converter (SNR = 98 dB) that consumes less than about 50-100 W, and even more power will be needed for amplifying the analog signal before conversion. On the other hand, a few microwatts per pole will be sufficient to implement the subsequent digital filtering with an advanced process and low-voltage operation. Thus, the necessary analog interfaces will (and do in fact already) consume most of the power in a signal processing chain when a dynamic range larger than 40 to 60 dB is required.

This is true only if the dynamic range must be assimilated with the maximum signal-to-noise ratio SNR that can be achieved in the whole bandwidth. However, in many applications, the SNR necessary for a certain level of signal is much smaller than the full dynamic of the signal. For example, speech transmission only requires a SNR of 40 dB, but the range of signals to be processed can be as large as 100 dB. By letting the noise follow the signal level to maintain just the necessary value of SNR, the 60 dB difference in this example provides the possibility to reduce power consumption by a factor of 10^6!

A well-known solution along this line is to use automatic gain control (AGC), in which the gain is slowly adapted to maintain constant the RMS or peak level of the signal, without affecting its instantaneous wave form. This solution necessarily causes some distortions, which must be minimized by carefully selecting the time constant(s) of the control loop.

Another known approach is the automatic range selection used in instrumentation. This approach can be extended to general analog processing systems by using the architecture shown in Fig. 2.25.

This approach, called analog floating point (AFP), amounts to multiply the instantaneous signal $x(t)$ by a factor K which is adapted to maintain the signal $x'(t)$ entering the processor within a min-max range. The scaled signal x' is then processed to produce a signal y' which is divided by the same factor K. Distortions by the processor are avoided if its state vector z (set of state variables) is multiplied by K^+/K each time the factor changes from K to a new value K^+. This updating of the state vector must be carried out in a time shorter than half the period of the highest frequency to be processed. It can be done between two sampling instants if the system is operating in discrete time.

1.2.6.3 Instantaneous companding and “log-domain” filtering

A possible way to circumvent the dynamic range problems introduced by the reduction of the supply voltage is to use companding. Companding systems traditionally used syllabic compression where the gains of the compressor and the expander are adjusted according to slowly varying characteristics of the signal such as envelope or

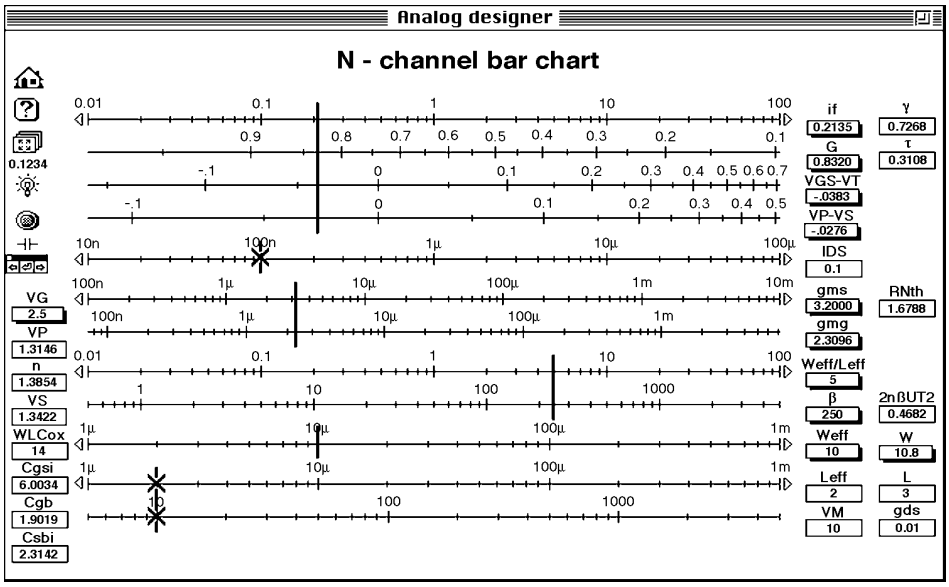


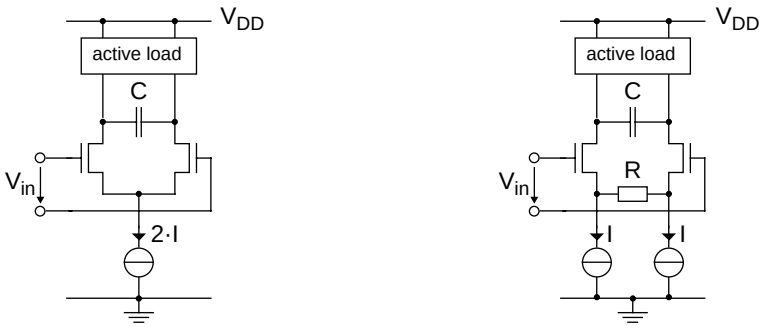
Figure 2.25 The analog floating point technique [49]

power [50]. More recent techniques called voltage companding current-mode filtering or "log-domain" filtering use instantaneously varying gains instead of syllabic detection or in other words non-linear characteristics [53]-[65]. The currents, having inherently a large dynamic range, are compressed when transformed into voltages (for example prior to the integration on a capacitor) and expanded afterwards when transformed back to currents. This technique makes the voltage swings across the integrator's capacitors almost independent of the supply voltage, which can then be reduced to the minimum required for a proper operation of the circuit. The "log-domain" technique additionally exploits the properties of the exponential function that can easily be implemented using either the well known relation between the collector current and the base-emitter voltage of a bipolar transistor or the drain current of a MOS transistor operating in weak inversion.

In addition to a low-voltage operation, voltage-companding also provides the most efficient use of current for implementing a given transconductance. A comparison can be elaborated by considering the g_m/I ratio as a factor of merit. The latter is maximum and equal to unity for a bipolar transconductor operating in small-signal. The maximum g_m/I for a MOST transconductor as the one used in the integrator shown in Fig. 2.26 (a), is obtained in weak inversion but is unfortunately *n* times smaller than for a BJT. The linear range of the circuit shown in Fig. 2.26 (a) is strongly limited to typically U_T.

The transconductor can of course be linearized in order to get the required dynamic range. Unfortunately, any linearization technique, as for example the one shown in Fig. 2.26 (b), results in a degradation of the effective g_m/I ratio which is proportional to the increase of the linear voltage range with respect to U_T:

$$\frac{g_m \cdot U_T}{I} = \frac{U_T}{R \cdot I} = \frac{U_T}{V_{in\text{-}max}} = \frac{1}{g_m \cdot R} << 1 \qquad (94)$$

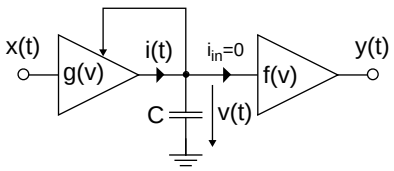


$V_{in\text{-}max} \le U_T << V_{DD}$ — *a) Small-signal integrator*

$U_T << V_{in\text{-}max} \le R \cdot I$ — *b) Linearized integrator*

Figure 2.26 Basic MOS integrators

The companding technique allows thus to extend the dynamic range while preserving the maximum g_m / I ratio.



expanding function: $y = f(v)$

companding function: $i = g(v) \cdot x$

linear capacitor: $i = C \cdot \frac{dv}{dt}$

Figure 2.27 Basic integrator illustrating the principle of instantaneous companding

The principle of instantaneous companding is illustrated by the integrator presented in Fig. 2.27, where the output signal *y(t)* is expanded by a given expanding function *f(v)* from the voltage across the integration capacitor *C* which is assumed to be linear. In order to preserve a global linear transfer function:

$$y(t) = \frac{1}{\tau} \cdot \int_{-}^{t} x(\tau) \cdot d\tau \qquad or \qquad \tau \cdot \frac{dy}{dt} = x \tag{95}$$

the current *i(t)* provided by the input current amplifier (or transconductor) has to be predistorted with a non-linear gain function *g(v)* which should satisfy the following necessary and sufficient condition [51]:

$$g(v) = \frac{C}{\tau} \cdot \left[\frac{df}{dv}\right]^{-1} \tag{96}$$

This general relation demonstrates that it is not necessary to have an exponential expanding function in order to perform companding. Fig. 2.28 shows an example of cubic

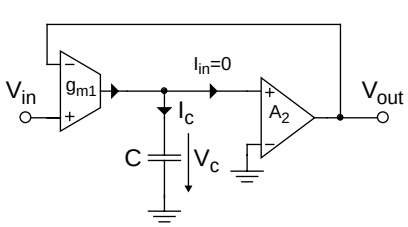
expansion:

$$f(v) = v^3 + v \tag{97}$$

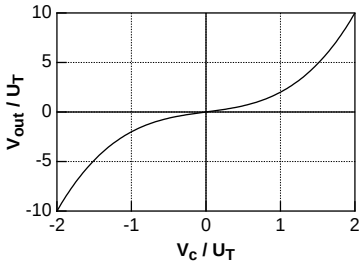
which can be applied to the first-order low-pass filter presented in Fig. 2.28 (a).

The cubic expansion is implemented using a voltage amplifier having a non-linear gain given by:

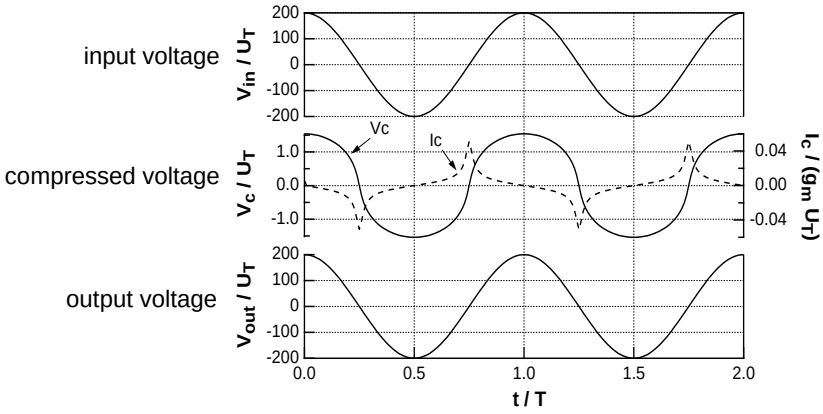
$$A_2 = A_0 \cdot \left[(V_c / U_T)^2 + 1 \right] \tag{98}$$



a) First-order low-pass filter



b) Expansion characteristic



c) Large-signal transient analysis

Figure 2.28 Example of a cubic expansion function

The input signal is predistorted by the input non-linear transconductor having a transconductance equal to:

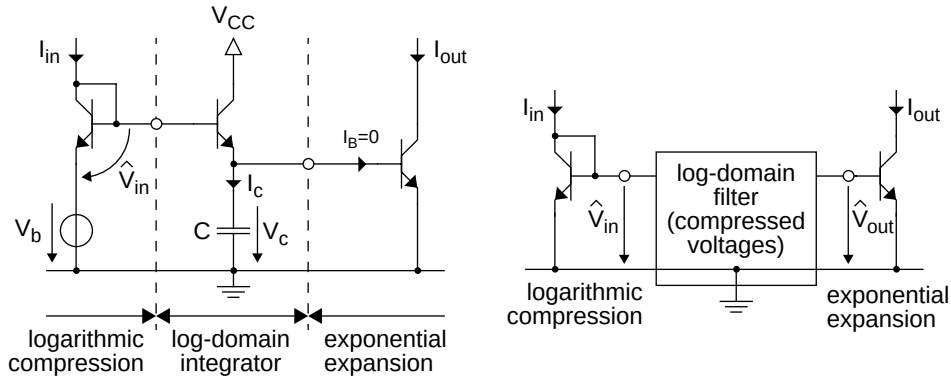
$$g_{m1} = \frac{g_m}{3\,(V_c / U_T)^2 + 1} \tag{99}$$

It is easily verified that the large-signal transfer function corresponds indeed to a first-order low-pass filter having a cut-off frequency given by:

$$f_c = A_0 \cdot \frac{g_m}{2\pi C} \tag{100}$$

The circuit of Fig. 2.28 (a) has been simulated for $A_0 = 1$ and $f_c T = 100$. The voltages are plotted in Fig. 2.28 (c) showing particularly the voltage across the capacitance which looks very distorted but which finally leads to a non-distorted output voltage thanks to the predistortion introduced by the input non-linear transconductor g_{m1} .

Although linear integrators can ideally be built with any expanding function, it requires a non-linear amplifier (or transconductor) satisfying (96) which is not always easy or even possible to realize. The exponential function is very well suited to implement companding circuits thanks to the fact that it is invariant to the differentiation operator. The compressor can therefore be built with the same function as the expander.



a) Basic integrator principle *b) Log-domain filters principle [59][63]*

Figure 2.29 Principle of log-domain filters

Moreover, the exponential function can be implemented with a reasonable accuracy using BJTs or MOSTs in weak inversion. An example of a BJT implementation (without bias circuit) is shown in Fig. 2.29 (a). The expanded output current is given by:

$$y = I_{out} = f(V_c) = I_S \cdot \exp\left[V_c / U_T\right] \tag{101}$$

where I_S is the BJT saturation current (and not the specific current). According to (96), the non-linear gain function $g(V_c)$ is given by:

$$g(V_c) = \frac{C \cdot U_T}{\tau \cdot I_S \cdot \exp\left[V_c / U_T\right]} = \frac{C \cdot U_T}{\tau \cdot I_S} \cdot \exp\left[\frac{-V_c}{U_T}\right] \tag{102}$$

The $\exp\left[-V_c / U_T\right]$ term in (102) can easily be obtained by simply connecting the capacitor to the emitter of a BJT as shown in Fig. 2.29 (a). The current flowing into the capacitor assuming the base current of the expander is negligible is then given by:

$$I_c = I_S \cdot \exp\left[\frac{\hat{V}_{in} + V_b - V_c}{U_T}\right] = I_0 \cdot \frac{\exp\left[\hat{V}_{in} / U_T\right]}{\exp\left[V_c / U_T\right]} = \tag{103}$$

$$= \frac{I_0 \cdot I_{in} / I_S}{\exp\left[V_c / U_T\right]} = g(V_c) \cdot I_{in}$$

where $I_0 = I_S \cdot \exp\left[V_b / U_T\right]$ is the bias current setting the time constant, $\hat{V}_{in} = U_T \ln\left(I_{in} / I_S\right)$ is the compressed input voltage and I_{in} is the input current.

Comparing (103) to (102) allows to determine the integration time constant:

$$\tau = \frac{C \cdot U_T}{I_0} = \frac{C}{g_m} \tag{104}$$

which is equal to the time constant of the integrator working in small-signal but valid also for large-signal operation. Fig. 2.29 (a) shows that a linear current-mode integrator can be realized with an exponential integrator surrounded by a logarithmic voltage compressor and an exponential expander.

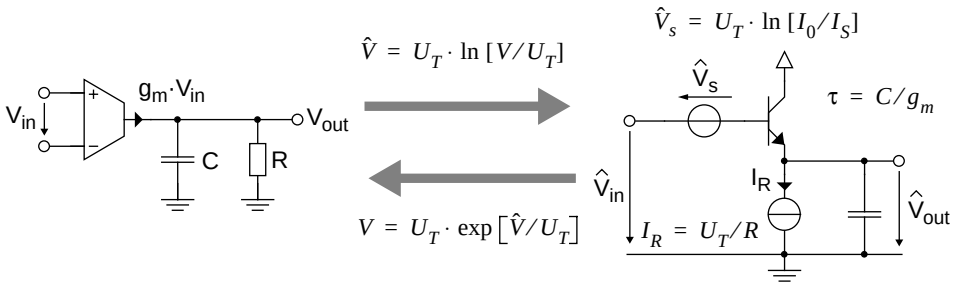


Figure 2.30 Mapping between linear- and log-domain [53][63][64]

As shown in Fig. 2.29 (b), this principle can be extended to higher order filters where the central part is called a log-domain filter since all the voltages are compressed logarithmically [59][63].

Log-domain filters can be designed by simple component substitution from a linear g_m-C filter using the mapping between the linear- and the log-domain [53][64] presented in Fig. 2.30. The mapping is defined by the linear capacitance and the expression of the time constant (or equivalently of the transconductance) being the invariants of the transformation. Note that a resistor in the linear-domain is transformed into a current source in the log-domain. This property can be used to tune the quality factor of a filter or an oscillator.

Log-domain filters can also be implemented in CMOS by simply replacing the bipolar transistors by MOS transistors biased in weak inversion. As an example, Fig. 2.31 (b) shows a MOST version of the log-domain integrator originally proposed by Seevinck [55]. The original bipolar integrator is based on a translinear multiplier which is redrawn with MOS transistors in Fig. 2.31 (a). In order for this MOST multiplier to conform to the translinear principle as described in [66], the source-to-bulk voltages have to be set to zero, which requires that M_3 and M_5 are implemented in separate wells. Writing the equation of the translinear voltage loop yields:

$$-V_{GB1} - V_{GB3} + V_{GB4} + V_{GB2} = 0 \tag{105}$$

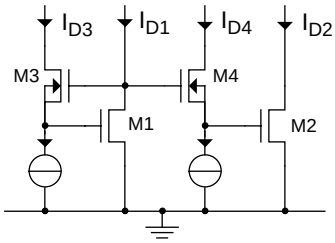
Assuming all the transistors are in saturation, the gate-to-bulk voltages are related to the drain currents I_{Di} according to:

$$V_{GBi} = n_i U_T \ln (I_{Di}/I_{Si}) + V_{T0i} \qquad i = 1 \ldots 4 \tag{106}$$

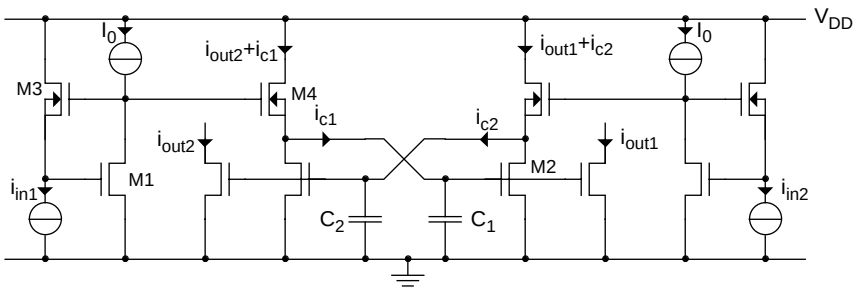
where it has been assumed that $V_{SBi} = 0 \quad i = 1, \ldots, 4$. The product of the drain currents is derived from (105) and (106) assuming a perfect matching of the leakage

current I_{D0i} and identical slope factors n_i :

$$I_{D1} \cdot I_{D3} = I_{D2} \cdot I_{D4} \tag{107}$$



a) MOS multiplier



b) CMOS voltage companding current mode integrator

Figure 2.31 CMOS multiplier and log-domain class AB integrator [55]

The current transfer function of the differential log-domain integrator shown in Fig. 2.31 (b) is obtained by using (107):

$$I_0 \cdot i_{in1} = i_{out1} \cdot (i_{out2} + i_{c1}) \tag{108}$$

$$I_0 \cdot i_{in2} = i_{out2} \cdot (i_{out1} + i_{c2}) \tag{109}$$

Subtracting (109) from (108) results in:

$$(i_{in1} - i_{in2}) \cdot I_0 = i_{out1} \cdot i_{c1} - i_{out2} \cdot i_{c2} \tag{110}$$

Since the capacitors C_1 and C_2 are assumed to be linear, the products of the output and of the capacitive currents are given by:

$$i_{outi} \cdot i_{ci} = nU_T \frac{di_{outi}}{dt} \qquad i = 1, 2 \tag{111}$$

Substituting (111) in (110) and integrating yields the equation of a linear integrator:

$$i_{out} = \frac{1}{\tau} \cdot \int i_{in} dt \tag{112}$$

where $i_{in} \equiv i_{in1} - i_{in2}$ and $i_{out} \equiv i_{out1} - i_{out2}$ are respectively the differential input

and output currents and τ is the integrator time constant given by:

$$\tau = \frac{1}{\omega_u} = \frac{C}{g_m} = \frac{nU_T \cdot C}{I_0} \tag{113}$$

which can be tuned by varying the bias current I_0.

The integrator of Fig. 2.31 (b) can easily be cascaded since the inputs and outputs are compatible. Moreover, it may be driven by several input current sources and can be extended to multiple outputs by adding several output transistors.

Thanks to the companding of the input current, the variation of the voltages across the integrating capacitors C_1 and C_2 stays small (typically smaller than $4nU_T$). Capacitors C_1 and C_2 can therefore be implemented by the non-linear parasitic C_{GB} and C_{GS} capacitances of the n-channel MOS transistors connected to these nodes without significantly degrading the distortion. This makes the integrator suited for integration in a standard digital CMOS process.

The original bipolar integrator suffers from a low DC gain and a high sensitivity to mismatch and base currents. The latter problem is of course solved by using MOSTs instead of BJTs, but the low DC gain and particularly the relative high sensitivity to mismatch becomes a strong handicap for this CMOS implementation. Moreover, the multiplier uses stacked gate-to-source voltages and is therefore not suited to low-voltage operation (it typically requires at least a 1.8 V supply). The circuit of Fig. 2.31 (b) can be improved by using a folded multiplier as suggested in [62] and [65].

It is worth mentioning that due to the class AB operation, the SNR is no longer a linear function of the input current as it would be for a class A circuit. For input currents larger than the nominal bias current, the noise current can no more be considered as constant but it increases with the square root of the signal current and so does the SNR. As shown in Fig. 2.32, the slope of the SNR versus the modulation index in a log-log scale starts at 20 dB/dec and progressively decreases down to 10 dB/dec for a signal amplitude larger than the bias current (or a modulation index *m* larger than 1). Class AB and class A current mode circuits may be compared by considering that they have initially the same bias current I_0, the same bandwidth B (or small-signal transconductance g_m) and therefore the same idle current noise. Both SNR versus input signal characteristics are thus superimposed for currents smaller than the bias current as shown in Fig. 2.32. The maximum signal or modulation index m_{max} of a class AB circuit is set by the maximum acceptable amount of distortion, whereas it is bounded by the bias current for a class A circuit. The class AB circuit can thus extend the dynamic range by a factor m_{max} and improve the SNR by a factor $\sqrt{m_{max}}$, without any increase of the standby bias current. This makes class AB companding integrators very attractive to extend the dynamic range and the SNR, while preserving the power consumption.

The integrator described previously has been used in the 4th-order low-pass (LP) Tchebycheff filter shown in Fig. 2.33 which has been synthesized from a LC ladder prototype. All integrators have the same transconductance $g_m = I_0/nUt$, which can be tuned by adjusting a common bias voltage V_{G0}. The filter requires an input signal conditioner in order to maintain all currents positive and enable the class AB operation [59]. The difference of the conditioner output currents is equal to the input current, while their product is kept constant and equal to I_0.

The simulated transfer functions are shown in Fig. 2.34 for different bias currents. They are computed from the step responses obtained from a transient analysis in order to

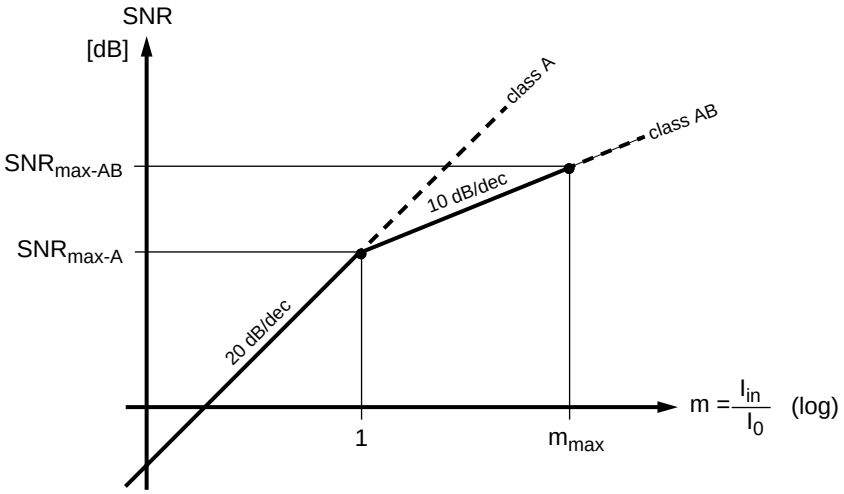


Figure 2.32 SNR versus input signal amplitude for a class A and a class AB circuit

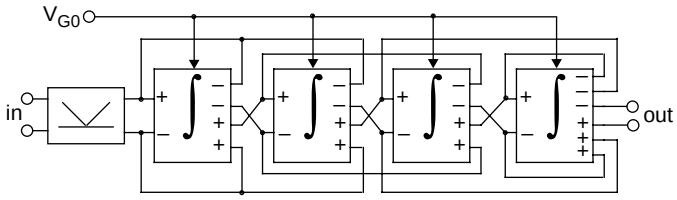


Figure 2.33 4th-order low-pass Tchebycheff filter

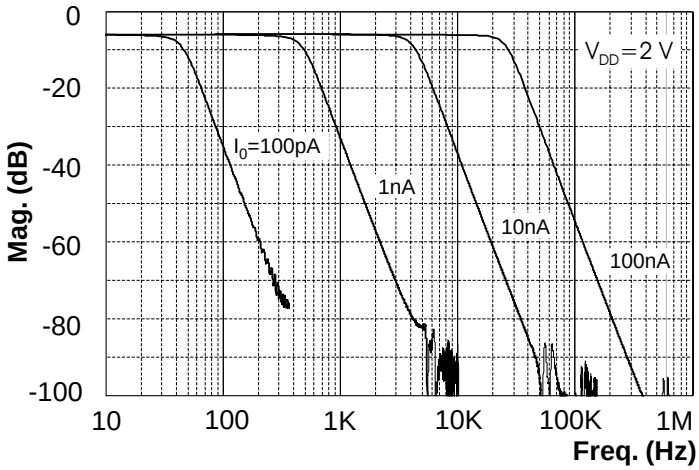


Figure 2.34 Large signal transfer function simulated for $m = 1$

demonstrate the correct operation of the filter even for large signals. The step amplitude is set equal to the bias current, but thanks to the class AB operation, the filter accepts

currents much larger than I_0 . The DC gain variation over the 3 decades tuning range stays smaller than 0.2 dB. The effect of the non-linearities of the capacitors are negligible thanks to the small variation of the voltages and to the property of the LC ladder filter which have a small sensitivity to component variations (Fettweiss-Orchard theorem).

The total harmonic distortion (THD) has been evaluated for each cut-off frequency and is plotted in Fig. 2.35 versus the ratio of the input to bias current defined as the modulation index *m*. The THD for a cut-off frequency fixed at about 500 Hz is less than 1 % even for a modulation index as high as 30. The dynamic range is estimated to about 65 dB and the power consumption is 150 nW for a cut-off frequency of 500 Hz.

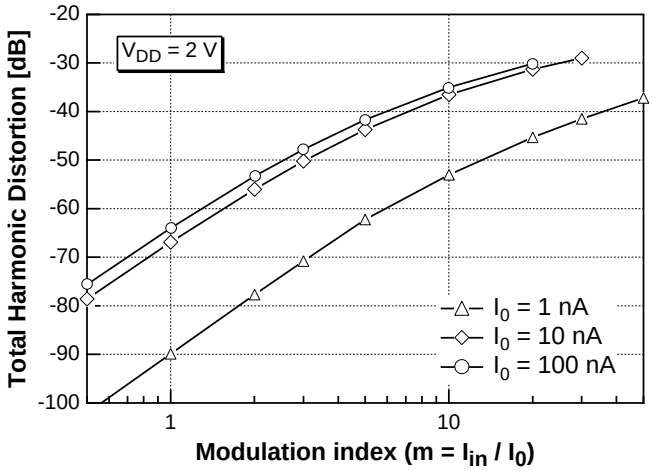


Figure 2.35 Simulation of the THD versus the input current normalized to the bias current (defined as the modulation index *m*)

Although log-domain filters seem to be very attractive for low-power and low-voltage, a considerable effort is still needed to better understand them and demonstrate their benefits for low-power and low-voltage analog signal processing applications.

1.2.7 Summary and conclusion

Unlike digital circuits where the dynamic power decreases with the square of the supply voltage, a first-order analysis shows that the minimum power consumption required to process analog signals is almost independent of the supply voltage reduction. This is mainly due to the fact that the power consumption of analog circuits at a given temperature is basically set by the required SNR and the frequency of operation (or the required bandwidth). A more detailed analysis indicates that this minimum power consumption of analog circuits is proportional to the ratio between the supply voltage and the signal peak-to-peak amplitude. Power efficient analog circuits should thus be designed to maximize the voltage swing and therefore should handle rail-to-rail signal voltages.

Many of the problems and solutions encountered in CMOS low-power analog circuit design are directly related to the properties of the MOS transistor itself, which must

therefore be properly understood and modelled down to very low currents. The essential features of a transistor can be captured in a symmetrical model where the drain current I_D is the superposition of a forward component I_F and reverse components I_R, which are proportional to the same function F of $V_P - V_S$ and $V_P - V_D$ respectively, where V_P is the pinch-off voltage which is directly related to the gate voltage. This function F is quadratic in strong inversion, while it is exponential in weak inversion. The proportionality factor is the specific current which depends on the aspect ratio of the device and defines a limit between weak and strong inversion. The ratio of the actual drain current to the specific current defines the inversion level or inversion factor of a MOS transistor in saturation. If this inversion factor is lower than one, the device operates in weak inversion. An inversion factor larger than one corresponds to a device operating in strong inversion and a transistor having an inversion factor close to one operates in the moderate inversion region. The transconductance-to-current ratio increases when the inversion factor is decreased and reaches a maximum in weak inversion. An unacceptable large error is made when estimating the effective transconductance of a transistor biased in the middle of the moderate inversion from the asymptotic expressions valid either in weak or in strong inversion. Continuous expression of the drain current as well as the transconductances are thus required for doing a correct sizing of each transistor operating in moderate inversion. Simple analytical expressions can be used and checked by simulation using for example the complete EKV MOST model. The scaling of advanced process is an additional motivation for having such a continuous analytical model, valid from weak to strong inversion. As a matter of fact, the scaling gradually shifts the specific current of a minimum size transistor to higher values of currents which for a given bias current progressively moves the corresponding inversion coefficient to lower values. In other words, for given performance, MOS transistors are more and more biased in the moderate or even in the weak inversion regions of operation.

The pinch-off voltage, the specific current and the inversion coefficient are not only abstract concepts which are only useful for the elaboration and formalism of the transistor model. These definitions show to be also very useful for designing efficient circuits. The pinch-off voltage can indeed be measured and is used for extracting the key parameters of the EKV model. It can also be generated from simple circuits and used for example to bias cascode stages. The specific current of a reference transistor can be generated by means of a dedicated circuit. It can then be used to precisely set the inversion coefficient of a device by simple scaling of this reference current using a series and/or parallel combination of reference transistors. This ratio-based design technique has the advantage to be insensitive to the first-order to temperature and process variations. It is therefore attractive for designing circuits that can be ported from one process to another without any major redesign while preserving the main performance.

The formulation of the drain current as the combination of a forward and a reverse component can be advantageously used to show that currents can be divided in given proportion among the branches of a network composed only of MOS transistors behaving as non-linear resistors. This property is valid independently of the current level and can be used for replacing any resistors of a resistive network by transistors. The same principle can also be used to realize a string of binary weighted current sources.

Voltage gain should be achieved by single-stage operational transconductance amplifier in order to avoid any compensation capacitor other than the load itself and therefore prevent any additional useless power consumption. The gain can be further increased without adding any extra current branch, by using a cascode stage which has to

be properly biased at the minimum voltage in order to preserve sufficient voltage swing.

In many analog signal processing applications, such as hearing aids for example, the input signal dynamic range is much larger than the required SNR. Hence, power can be saved if the SNR can be reduced to the minimum required while maintaining the necessary dynamic range. An ideal analog signal processing systems should even be able to maintain the SNR constant independently of the signal: when the signal is weak, the noise floor has to be low, when the signal is large the noise may be large as long as the required SNR is still achieved. This can be achieved by means of the analog floating point technique, which basically consists in scaling the input signal by an adequate factor adjusted in such a way that the signal fits within a given range. Distortions are avoided by correctly updating the state variables of the analog signal processor. This technique is obviously well adapted to analog sampled-data systems such as switched-capacitor circuits, since the updating can take place between two sampling instants leaving sufficient time for the undesirable but unavoidable transients to vanish. The analog floating point technique can also be applied to continuous-time analog signal processing systems as long as the state-variables are also updated continuously.

A possible way to maintain a sufficient dynamic range when reducing the supply voltage without degrading the power consumption of analog signal processing circuits is to use the instantaneous companding technique. In this approach, the currents are compressed when transformed into voltages and expanded when transformed back to currents. The input current has to be predistorted in order to preserve a linear operation. The expanding function is theoretically not restricted to the exponential function but since the predistortion of the signal requires the derivative of this expanding function, it is much easier to realize it using the exponential function since it is invariant to the differentiation operator and can be implemented either by the current-to-voltage characteristic of a bipolar transistor or a MOS transistor biased in weak inversion. The instantaneous voltage-companding technique additionally allows to extend the dynamic range while preserving the maximum g_m/I available in small-signal operation.

1.2.8 References

[1] M. Declercq and M. Degrauwe, “Low-Power / Low-Voltage IC Design: An Overview,” *Advanced Engineering Course on Low-Power / Low-Voltage IC Design*, Lausanne (EPFL), Switzerland, June 1994.

[2] A. P. Chandrakasan, S. Sheng and R. W. Brodersen, “Low-Power CMOS Digital Design,” *IEEE Journal of Solid-State Circuits*, Vol. 27, pp. 473-484, Apr. 1992.

[3] T. G. Noll and E. de Man, “Pushing the Performance Limits due to Power Dissipation of Future ULSI Chips,” *Int. Solid-State Circ. Conf. Dig. of Tech. Papers*, pp. 1652-1655, 1992.

[4] D. Liu and C. Svensson, “Trading Speed for Low Power by Choice of Supply and Threshold Voltages,” *IEEE Journal of Solid-State Circuits*, Vol. 28, pp. 10-17, 1993.

[5] G. Schrom, D. Liu, C. Pichler, C. Svensson and S. Selberherr, “Analysis of Ultra-Low-Power CMOS with Process and Device Simulation,” *Proc. of ESSDERC'94*, Sept. 1994.

[6] J. Burr and J. Shott, "A 200 mV Self-testing Encoder/Decoder Using Standford Ultra-Low Power CMOS," *Int. Solid-State Circ. Conf. Dig. of Tech. Papers*, pp. 84-85, 1994.

[7] E. A. Vittoz, "Low-Power Design: Ways to Approach the Limits," *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 14-18, May 1994.

[8] R. Castello and P. R. Gray, "Performance limitation in switched-capacitor filters," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 865-876, Sept. 1985.

[9] G. Groenewold, "Optimal dynamic range integrators," *IEEE Trans. Circuits Syst. I*, vol. 39, pp. 614-627, Aug. 1992.

[10] E. A. Vittoz, "Future of Analog in the VLSI Environment," *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 1372-1375, May 1990.

[11] Y. P. Tsividis, *Operation and modelling of the MOS Transistor*, Mc Graw-Hill, 1987.

[12] H. C. Pao and C. T. Sah, "Effects of diffusion current on the characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electronics*, vol. 10, pp. 927-937, 1966.

[13] J. D. Chatelain, *Dispositifs à semiconducteurs*, Traité d'électricité, vol. VII, Presses Polytechniques Romandes, 1979.

[14] 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," special issue of the *Analog Integrated Circuits and Signal Processing journal* on Low-Voltage and Low-Power Design, vol. 8, pp. 83-114, July 1995.

[15] K. Bult and J. G. M. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE Journal of Solid-State Circuits*, vol. SC-27, pp. 1730-1735, Dec. 1992.

[16] C. C. Enz, *The EKV model: a MOST model dedicated to low-current and low-voltage analog circuit design and simulation* in *Low-power HF microelectronics*, edited by G. S. Machado, IEE Circ. and Syst. Series N 8, IEE publications, 1996.

[17] M. Bucher, C. Lallement, C. C. Enz and F. Krummenacher, "Accurate MOS modelling for analog circuit simulation using the EKV model," to be published in the *Proc. IEEE Int. Symp. Circuits Syst.*, May 1996.

[18] C. Lallement, C. Enz and M. Bucher, "Simple solutions for modelling the non-uniform substrate doping", to be published in the *Proc. IEEE Int. Symp. Circuits Syst.*, May 1996.

[19] C. C. Enz, "High Precision CMOS Micropower Amplifiers," PhD Thesis N 802, EPFL, Lausanne, 1989.

[20] Y. Tsividis and G. Masetti, "Problems in precision modeling of the MOS transistor for analog applications," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, pp. 72-79, Jan. 1983.

[21] Y. Tsividis and K. Suyama, "MOSFET Modeling for analog circuit CAD: problems and prospects," *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 14.1.1-14.1.6, May 1993.

[22] Y. Tsividis and K. Suyama, "MOSFET Modeling for analog circuit CAD: problems and prospects," *IEEE Journal of Solid-State Circuits*, vol. SC-29, pp. 210-216, March 1994.

[23] G. A. S. Machado and C. Toumazou, "Systematic Design-Oriented Characterisation of MOS Devices and Circuit Building Blocks in Engineering Education," *Proceedings of the 9th conference of the Sociedade Brasileira de Microelectronica*, pp. 243-257, Aug. 1994.

[24] E. A. Vittoz, "MOS transistors operated in the lateral bipolar mode and their application in CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. SC-18, pp. 273-279, June 1983.

[25] X. Arreguit, "Compatible lateral bipolar transistors in CMOS technology: model and applications," PhD Thesis N 817, EPFL, Lausanne, 1989.

[26] E. A. Vittoz, "The Design of High-Performance Analog Circuits on Digital CMOS Chips," *IEEE Journal of Solid-State Circuits*, vol. SC-20, pp. 657-665, June 1985.

[27] E. A. Vittoz, *Micropower techniques*, in *Design of MOS VLSI circuits for telecommunications*, edited by J. Franca and Y. Tsividis, Prentice-Hall, 1994.

[28] E. A. Vittoz, "MOS transistor," *Intensive Summer Course on CMOS & BiCMOS VLSI Design, Analog&Digital*, Lausanne (EPFL), Switzerland, September 1993.

[29] Y. Tsividis and V. Gopinathan, *Continuous-time filters*, in *Design of MOS VLSI circuits for telecommunications*, edited by J. Franca and Y. Tsividis, Prentice-Hall, 1994.

[30] E. A. Vittoz and O. Neyroud, " A low-voltage CMOS bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-14, pp. 573-577, June 1979.

[31] E. Vittoz and J. Fellrath, "MOS analog integrated circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. SC-12, pp. 224-231, June 1977.

[32] H. J. Oguey and S. Cserveny, "MOS modelling at low current density," Summer Course on "Process and Device Modelling", ESAT Leuven-Heverlee, Belgium, June 1983.

[33] H. Oguey et S. Cserveny, "Modèle du transistor MOS valable dans un grand domaine de courants," *Bull. SEV/VSE*, Feb. 1982.

[34] J. Fellrath, "Shot noise behavior of subthreshold MOS transistors," *Revue de Physique Appliquée*, vol. 13, pp. 719-723, December 1978.

[35] R.Sarpeshkar, T. Delbrück and C. A. Mead, "White Noise in MOS Transistors and Resistors," *IEEE Circuits &Devices Magazine*, vol. 9, pp. 23-29, Nov. 1993.

[36] N. D. Arora, *MOSFETs models for VLSI Circuit Simulation, Theory and Practice*, Computational Microelectronics, Ed. S. Selberherr, Springer-Verlag 1993.

[37] A. Juge, private communications.

[38] G. Machado, C. C. Enz and M. Bucher, "Estimating key parameters in the EKV MOST model for analogue design and simulation", *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 1588-1591, April 29-May 3, 1995.

[39] C. Lallement, M. Bucher and C. Enz, "The EKV MOST Model and the Associated Parameter Extraction", *Proc. of the European HP IC-CAP Users Meeting*, 1995.

[40] M. Bucher, C. Lallement and C. C. Enz, “An efficient parameter extraction methodology for the EKV MOST model,” to be published in the *Proc. IEEE Int. Conf. on Test Structures*, March 1996.

[41] C. A. Mead, *Analog VLSI and neural systems*, Addison-Wesley, Reading, 1989.

[42] C. Galup-Montoro, M. C. Schneider and I. J. B. Loss, “Serie-parallel association of FET’s for high gain and high frequency applications,” *IEEE Journal of Solid-State Circuits*, vol. SC-29, pp. 1094-1100, Sept. 1994.

[43] R. T. Goncalves, S. Noceti Filho, M. C. Schneider and C. Galup-Montoro, “Digitally Programmable Switched Current Filters,” to be published in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1996.

[44] E. A. Vittoz and X. Arreguit, “ Linear networks based on transistors,” *Electron. Lett.*, vol. 29, pp. 297-298, Feb. 1993.

[45] T. C. Choi, R. T. Kaneshiro, R. W. Brodersen, P. R. Gray, W. B. Jett and M. Wicox, “High-frequency CMOS switched-capacitor filters for communications applications,” *IEEE Journal of Solid-State Circuits*, vol. SC-18, pp. 652-664, Dec. 1983.

[46] R. C. Schober et al., “Ultra low power circuit techniques for focal plane electronics applications,” JPL Task Plan 80.3295, 1993.

[47] P. Heim, S. Schultz and M. A. Jabri, “Technology-independent biasing technique for CMOS analogue micropower implementations of neural networks,” *Proceedings of the 4th IEEE International Workshop on Cellular Neural Networks and their Applications* (CNNA-95)

[48] P. Heim and M. A. Jabri, “MOS cascode-mirror biasing circuit operating at any current level with minimal output saturation voltage,” *Electron. Lett.*, vol. 31, pp. 690-691, Apr. 1995.

[49] E. M. Blumenkrantz, “The analog floating point technique,” *IEEE Symp. on Low-Power Electronics Dig. of Tech. Papers*, pp. 72-73, Oct. 1995.

[50] Y. P. Tsividis, V. Gopinathan, and L Tòth, “Companding in signal processing,” *Electron. Lett.*, vol. 26, pp. 1331-1332, Aug. 1990.

[51] Y. Tsividis, “On Linear Integrators and Differentiators Using Instantaneous Companding”, *IEEE Trans. Circuits Syst. II*, vol. CAS-42, pp. 561-564.

[52] Y. Tsividis, “General approach to signal processors employing companding,” *Electron. Lett.*, vol. 31, pp. 1549-1550, Aug. 1995.

[53] C. C. Enz, “Low-power log-domain continuous-time filters: an introduction,” *Proc. Low-Power - Low-Voltage Workshop of the European Solid-State Circuit Conf. (ESSCIRC’95)*, Sept. 1995.

[54] R. W. Adams, “Filtering in the log domain,” Preprint 1470, presented at 63rd AES Conf., New York, May 1979.

[55] E. Seevinck, “Companding current-mode integrator: A new circuit principle for continuous-time monolithic filters,” *Electron. Lett.*, vol. 26, pp. 2046-2047, Nov. 1990.

[56] D. R. Frey, “A general class of current mode filters,” *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 1435-1438, May 1993.

[57] D. R. Frey, “Log domain filtering: An approach to current mode filtering,” in IEE Proc., pt. G, Dec. 1993, vol. 140, pp. 406-416.

[58] D. R. Frey, “A 3.3 V electronically tunable active filter usable to beyond 1 GHz,” *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 493-496, May 1994.

[59] D. R. Frey, “Current mode class AB second order filter,” *Electron. Lett.*, Vol. 30, pp. 205-206, Feb. 1994.

[60] C. Toumazou, J. Ngarmnil, and T. S. Lande, “Micropower log-domain filter for electronic cochlea,” *Electron. Lett.*, vol. 30, pp. 1839-1841, Oct. 1994.

[61] D. Perry and G. W. Roberts, “Log-domain filters based on LC Ladder synthesis,” *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 311-314, May 1995.

[62] M. Punzenberger and C. C. Enz, “Low-Voltage Companding Current-Mode Integrators”, *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 2112-2115, May 1995.

[63] G. van Ruymbeke, “Filtres continus intégrés programmables,” PhD Dissertation, EPFL, N 1397, 1995.

[64] F. Yang and C. C. Enz, “Design of Low-Power &Low-Voltage Log-Domain Filters,” to be published in the *Proc. IEEE Int. Symp. Circuits Syst.*, May 1996.

[65] M. Punzenberger and C. C. Enz, “A new 1.2 V BiCMOS log-domain integrator for companding current-mode filters,” to be published in the *Proc. IEEE Int. Symp. Circuits Syst.*, May 1996.

[66] B. Gilbert, “Translinear circuits: a proposed classification,” *Electron. Lett.*, vol. 11, pp. 14-15, 1975.

[67] Vittoz E., Analog VLSI implementation of neural networks in Handbook of neural computation, Institute of Physics Publishing and Oxford University Press, 1995.