Analog Circuits in Ultra-Deep-Submicron CMOS

Anne-Johan Annema, *Member, IEEE*, Bram Nauta, *Senior Member, IEEE*, Ronald van Langevelde, *Member, IEEE*, and Hans Tuinhout

***Abstract*—Modern and future ultra-deep-submicron (UDSM) technologies introduce several new problems in analog design. Nonlinear output conductance in combination with reduced voltage gain pose limits in linearity of (feedback) circuits. Gate-leakage mismatch exceeds conventional matching tolerances. Increasing area does not improve matching anymore, except if higher power consumption is accepted or if active cancellation techniques are used. Another issue is the drop in supply voltages. Operating critical parts at higher supply voltages by exploiting combinations of thin- and thick-oxide transistors can solve this problem. Composite transistors are presented to solve this problem in a practical way. Practical rules of thumb based on measurements are derived for the above phenomena.**

***Index Terms*—Analog design, breakdown, CMOS, distortion, evolution, future performance, gate leakage, low power, low voltage, mismatch, scaling, technology, UDSM.**

I. Introduction

THE evolution in CMOS technology is motivated by decreasing price-per-performance for digital circuitry; its pace is determined by Moore's Law. To ensure sufficient lifetime for digital circuitry and to keep power consumption at an acceptable level, the dimension-shrink is accompanied by lowering of nominal supply voltages. While this evolution in CMOS technology is by definition very beneficial for digital, this is not so for analog circuits [1]–[3].

Contemporary ICs are mixed-signal systems consisting of a large digital core including amongst others a CPU or DSP and memory, often surrounded by several analog interface blocks such as I/O, D/A, and A/D converters, RF front ends, and more. From an integration point of view all these functions would ideally be integrated on a single die. In this case the analog electronics must be realized on the same die as the digital core and consequently must cope with the CMOS evolution dictated by the digital circuit. This paper discusses a number of issues for analog designs in modern and future ultra deep submicron (UDSM) CMOS processes and possible ways to maintain performance [3].

CMOS evolution has come to a point where for analog circuits new phenomena need to be taken into account. A major issue is the decreasing supply voltage. Although the supply voltage has dropped from 5 V in the early nineties down to 1.2 V today, most analog circuits can still be designed. However, a further drop in supply voltages is expected to cause serious roadblocks for analog circuits, because the signal headroom becomes too small to design circuits with sufficient signal integrity at reasonable power consumption levels. Although the analog transistor properties do not really get worse when comparing them at identical bias conditions, lower supply voltages require biasing at lower operating voltages which results in worse transistor properties, and hence yield circuits with lower performance.

A second issue is gate leakage. Gate leakage will increase drastically when migrating to newer technologies. When the gate oxide thickness is reduced with the equivalent of one atomic layer, the gate current increases by approximately one order of magnitude. Despite technological remedies, gate leakage will become part of analog design-especially for long transistors; e.g., in 65 nm technology the current gain of a MOSFET will be as small as unity for a channel length of 30 μm. In this paper, we introduce a bias insensitive frequency $f_{\mathrm{gate}}$for quick estimation of the effect of gate leakage. Another issue is gate leakage current mismatch. For large area (long L) transistors mismatch will be dominated by gate leakage mismatch. This effect puts a new upper limit on achievable matching performance. This problem can be coped with by accepting increased power consumption or by using active cancellation techniques.

This paper is organized as follows. Section II reviews the implications of going to lower supply voltages. Section III discusses the trend in a number of bare transistor properties, also illustrating that lower supply voltages degrade circuit performance. In Section IV f_{gate} is introduced while its use and applications are discussed in Section V. Section VI discusses a solution for the reduced nominal supply voltage aspects of newer CMOS generations: operating analog circuits at relatively high voltages, using thick oxide transistors and composite high-voltage transistors.

II. Fundamental Implications of Lower Supply Voltages

From a circuit point of view, plain physics dictates that the power consumption of analog circuits is proportional to the level of signal integrity (e.g., the signal-to-noise ratio, SNR, or the signal-to-noise and distortion ratio, SINAD) and to the signal frequency [4]–[6]. In other words: for analog circuits more performance comes at the cost of higher power consumption. There is a factor between the actual and the fundamental minimum power consumption that takes into account implementation overhead, margins in operating conditions and device spread. A common observation in all power-performance relations is that power consumption rises with decreasing supply voltages. Appendix A presents a short review of a number of

Manuscript received April 11, 2004; revised May 28, 2004.

A. J. Annema and B. Nauta are with the IC Design Group, University of Twente, 7500AE Enschede, The Netherlands (e-mail: a.j.annema@utwente.nl).

R. van Langevelde and H. Tuinhout are with Philips Research Laboratories, 5656AA Eindhoven, The Netherlands.

Digital Object Identifier 10.1109/JSSC.2004.837247

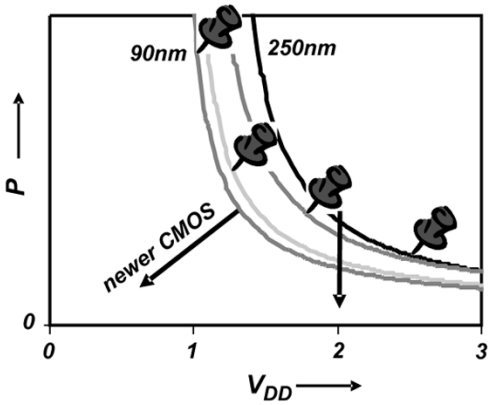


Fig. 1. Minimum power consumption for an (arbitrary) analog circuit (see Appendix A) with fixed topology and performance as a function of the supply voltage, for four technologies. Pushpins correspond to the power consumption in a technology at the nominal supply voltage for each CMOS process.

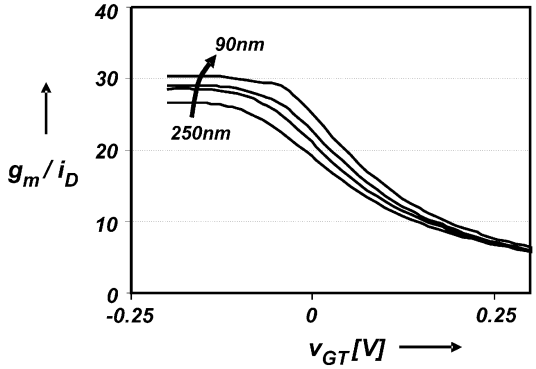


Fig. 2. The transconductance normalized with respect to the gate-overdrive voltage $V_{\rm GT}$, for four technologies.

power-performance relations and discusses exceptions to the "rule."

For a given power budget the performance drops when migrating to newer technologies, simply because of their lower supply voltages. This is probably the single most important effect that (fundamentally and practically) complicates analog designs at low supply voltages. For Fig. 1, a simple unity gain voltage buffer with fixed topology, fixed performance and fixed technology was optimized for minimum power consumption; see also Appendix A. In the optimization process, signal swing, all bias conditions of transistors and device dimensions were optimized [6]. It follows that the minimum power consumption increases with decreasing supply voltages. However, *at constant supply voltage*, porting the circuit to a newer technology *lowers* the required power consumption.

III. Scaling of Conventional Transistor Properties

Apart from issues at circuit level, basic transistor properties also change with CMOS technology evolution. This section reviews a few important properties.

A. DC Properties at Constant Voltage Headroom

This section presents the trend in dc properties of MOS transistors at constant voltage headroom ($V_{\rm DS}$) under typical analog operating conditions (low gate-overdrive voltage $V_{\rm GT}$). Note that these conditions are usually not satisfied when porting designs to newer technologies because of decreasing nominal supply voltages. For reasons of clarity, it is however a fair condition for the comparison of bare transistor properties.

Low-distortion at quasi-dc frequencies is relevant for many analog circuits. Typically, quasi-dc distortion may be due to nonlinearities in the transistors' transconductances and in their output conductances. However, nonlinearities of transconductances are usually not very relevant as they are more or less constant over technologies (under comparable biasing) and because the signal swing $v_{\rm gs}$ is usually low in a feedback or amplifier configuration. Fig. 2 shows the transconductance normalized with respect to the drain current [7] as a function of the gate-overdrive voltage $V_{\rm GT} = V_{\rm GS} - V_T$, derived from measurement on transistors in four different technologies; clearly the normalized g_m hardly changes over technology. For output conductance we find the opposite: it is heavily dependent on biasing, size, technology and typically sees large voltage swings. For this reason the remainder of this section reviews trends in output conductances. As a simple estimate of linearity, let us assume an MOS transistor with a nonlinear output conductance, and an output voltage consisting of a sine superimposed on a dc voltage. The drain current can be approximated by

$$i_D \cong I_D + g_{\rm ds}\cdot\hat{V}\cdot\sin(\omega t) - g_{\rm ds2}\cdot\frac{\hat{V}^2}{4}\cdot\sin(2\omega t) - g_{\rm ds3}\cdot\frac{\hat{V}^3}{24}\cdot\sin(3\omega t) + \cdots \quad (1)$$

The harmonic current components typically add to total harmonic distortion, but can be suppressed with voltage loop gain. Voltage loop gain at quasi-dc frequencies in transistor circuits is the combined effect of a number of $g_m/g_{\rm ds}$ratios and as such also depends on transistor output conductance. Fig. 3 shows graphs of transistor voltage gain and distortion, the latter expressed in output $\mathrm{IP}_3 = \sqrt{24\cdot g_{\rm ds3}/g_{\rm ds}}$, as a function of the effective gate-overdrive voltage $V_{\rm GT}$. The curves are derived from nonlinearly interpolated measurements on devices from four technologies; the length of all transistors was 1 μm for comparison reasons. Note that the curves are independent of the drain-current level.

Fig. 3(a) shows that at constant gate-overdrive voltage $V_{\rm GT}$ and fixed drain source voltage (here 0.3 V) the voltage gain of transistors decreases somewhat with newer technologies: a factor 2 in four generations. Fig. 3(b) shows that with the same scaling the output IP3 improves. The combined effect of these two trends is that at fixed transistor length and at constant voltage headroom, the quasi-dc circuit performance hardly changes over technology. Note that constant voltage headroom implies that no supply voltage downscaling is done.

B. DC Properties at Decreasing Voltage Headroom

With migration to more advanced CMOS processes usually the supply voltage of a circuit realized in that technology is decreased, implying that the voltage headroom and signal swing of individual transistors are also decreased. Fig. 4(a) shows the transistor voltage gain, following from nonlinearly interpolated measurements, now under the assumption that the quiescent

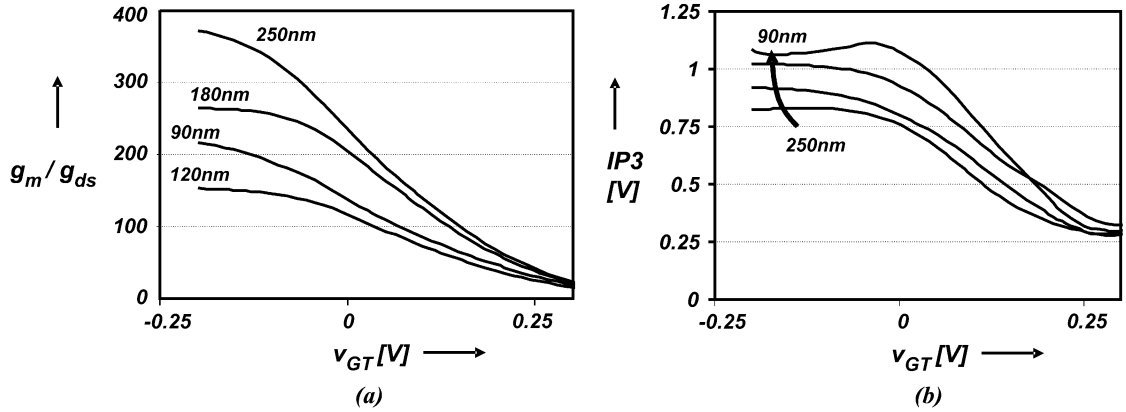


Fig. 3. Various DC-properties of transistors as a function of the gate-overdrive voltage at fixed $V_{\rm DS} = 0.3$ V and $L = 1\ \mu$m for four technologies: (a) the gain and (b) the output IP3.

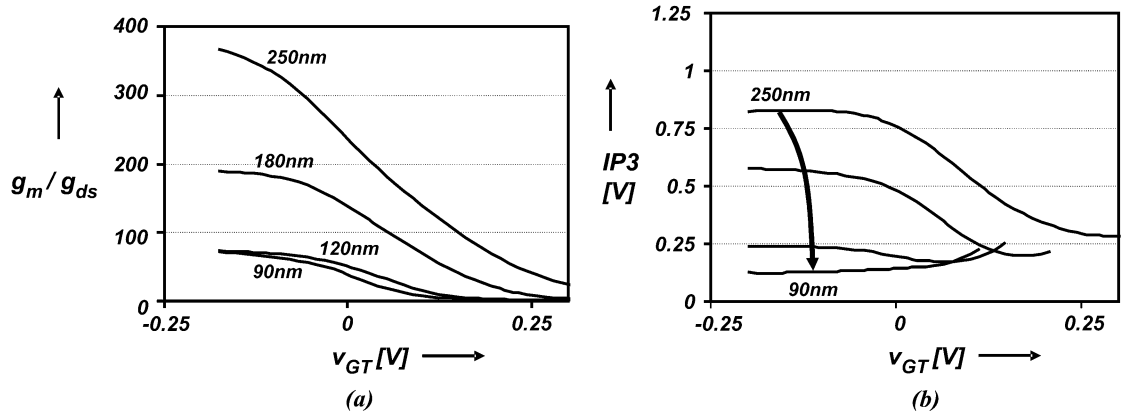


Fig. 4. Various dc properties of transistors as a function of the gate-overdrive voltage with $V_{\rm DS}$ proportional to the nominal supply voltage (0.3 V for 250 nm) and $L = 1\ \mu$m for four technologies: (a) the gain, and (b) the output IP3.

$V_{\rm DS}$ and the signal swing are decreased proportional to the nominal supply voltage, but still at constant transistor length; Fig. 4(b) shows the corresponding output IP3. Note that in order to compare only transistors in saturation, the $V_{\rm GT}$ranges are different for the four technologies.

Under the more realistice assumptions leading to Fig. 4, we see that porting typically results in significantly lowered transistor voltage gain and output IP3. With this scaling scenario, higher harmonic components may increase in amplitude despite the smaller signal; the THD increases significantly. At circuit level the degraded quasi-dc performance can be compensated by techniques that boost gain, such as (regulated) cascodes. These are, however, harder to fit within decreasing supply voltages. Other solutions include a more aggressive reduction of signal magnitude which requires a higher power consumption to maintain SNR levels.

AC Properties: The ac performance of transistors improves with newer technologies: it is one of the main technology drivers. As a first order estimate, two classes of capacitance are important for speed: the intrinsic capacitances of transistors and the junction capacitances. In this context intrinsic capacitances are all capacitances related to actual MOS operation, including overlap capacitances.

The impact of intrinsic capacitances hardly changes over technology for transistors under analog operating conditions with a fixed length. This can be illustrated using simple scaling theories [8], [9] combined with square-law relations that are satisfactory for illustrating trends. It follows that to first order the unity gain frequency of transistors depends only on effective gate-overdrive voltage and on channel length.

$$f_{UG} \approx \frac{g_m}{2\pi \cdot C_{\rm ox} \cdot WL} \propto \frac{V_{\rm GT}}{L^2} \tag{2}$$

where $C_{\rm ox}$ is the oxide capacitance per unit area. With a fixed transistor length, e.g., for voltage gain or accuracy reasons, the transistor's intrinsic speed hardly changes over technology. Combined with the findings in the previous parts of this section, it follows that there is a clear trade-off between gain and speed via transistor length (see also [10]). Circuits in newer CMOS technologies can hence achieve higher bandwidths but at the cost of degraded quasi-dc performance.

Another aspect of ac performance is the junction capacitance. With technology-scaling both the LDD structures and the actual junctions become shallower, roughly proportional to the technology feature size. Also, the junction area roughly scales in proportion to the minimum gate-length, while the dope level increase does not significantly increase the capacitance per area. Altogether this leads to a significantly reduced junction capacitance per g_m with newer technologies. This allows for better HF and RF performance with technology evolution.

IV. Gate Leakage and F_{GATE}

Besides its impact on conventional properties of circuits and devices, CMOS evolution introduces several new problems in analog design. One of the new phenomena is gate leakage [11]: gate current due to direct tunneling through the thin gate oxide. This leakage depends mainly on gate-source voltage bias and gate area.

One obvious implication of gate leakage is that the gate input impedance includes the conventional input capacitance $C_{\rm in}$ in parallel to a tunnel conductance $g_{\rm tunnel}$. These two have identical area dependence, resulting in an $f_{\rm gate}$ that is area *in*dependent and fairly independent of the drain-source voltage $v_{\rm DS}$. However, the tunnel current density for electrons and holes is different mainly due to the differences in oxide barrier height, resulting in (see Appendix B)

$$\begin{aligned} f_{\rm gate} &= \frac{g_{\rm tunnel}}{2\pi C_{\rm in}} \\ &\cong 1.5\cdot 10^{16}\cdot v_{\rm GS}^2\cdot e^{t_{\rm ox}(v_{\rm GS}-13.6)} \quad \text{(NMOST)} \\ &\cong 0.5\cdot 10^{16}\cdot v_{\rm GS}^2\cdot e^{t_{\rm ox}(v_{\rm GS}-13.6)} \quad \text{(PMOST)} \end{aligned} \tag{3}$$

where $t_{\rm ox}$ is in [nm] and $v_{\rm GS}$ is in [V].

For signal frequencies higher than this $f_{\rm gate}$ the input impedance is mainly capacitive and the MOSFET behaves as a conventional MOSFET. Otherwise, below $f_{\rm gate}$ it is mainly resistive and the gate leakage is dominant. Fig. 5 shows tens of $f_{\rm gate}$ curves based on measurements on transistors with different sizes and bias conditions in a 180-nm technology, and shows the prediction using (3). It follows that in this technology the gate appears to be capacitive for signal frequencies higher than roughly 0.1 Hz, while it appears resistive only at very low signal frequencies.

Analog applications typically apply transistors biased at low and moderate gate-overdrive voltages. The corresponding $f_{\rm gate}$ of a technology is then inside a relatively small frequency band. Fig. 6 shows such $f_{\rm gate}$ bands for four technologies as derived from measurements. This figure clearly illustrates that signal frequencies for which the input impedance appears to be resistive change from roughly 0.1 Hz in 180-nm technologies to about 1 MHz in 65-nm CMOS.

$f_{\rm gate}$ will prove to be useful in the estimation of the impact of gate leakage on other relevant properties of MOS transistors. A number of estimations are given in the next section of this paper.

V. Impact of Gate Leakage

A. Limited Current Gain

Input bias current due to gate leakage is very similar to base current in bipolar technologies and hence known solutions for bipolar circuits can usually be applied in analog CMOS circuits with leaky gates. There are two major differences between the bipolar base current and the CMOS gate current. First, in CMOS the width and length can be selected—and are optimized in analog designs—while in bipolar designs only the emitter area (equivalent to MOSFET width) can be set while the base width (equivalent to MOSFET length) is fixed. The result is that in ultra-deep-submicron processes long CMOS transistors (that are frequently required in conventional analog circuit designs,

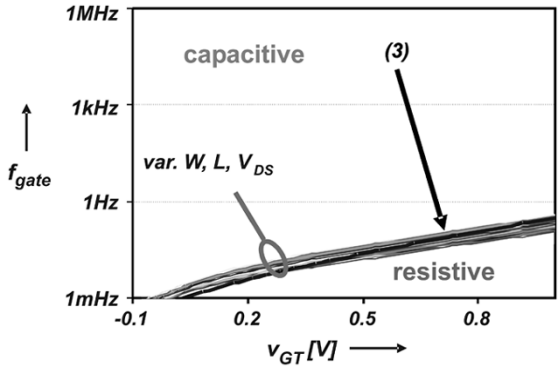


Fig. 5. $f_{\rm gate}$ as a function of the effective gate-overdrive voltage for different NMOS-transistors in 180-nm CMOS, based on measurements and fitted using (3).

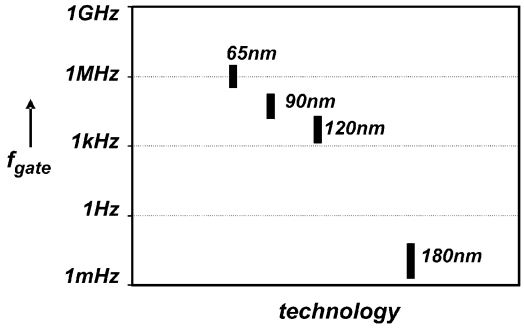


Fig. 6. $f_{\rm gate}$ ranges for typical analog applications, for NMOS transistors in different CMOS technologies. For PMOS transistors, $f_{\rm gate}$ is roughly a factor 3 lower.

for high output resistance or for low mismatch and flicker noise reasons) may have a lower-than-unity current gain. Secondly the bipolar base current is near zero under reverse bias conditions, whereas the gate current is not which should be taken into account, e.g., for switching applications.

The dc current gain can be estimated using $f_{\rm gate}$. In strong inversion and saturation at low gate-overdrive voltages, the drain current is by rough approximation given by the square-law relation. Using the expression for $g_{\rm tunnel}/i_{\rm GS}$ for strong inversion (see Appendix C)

$$\frac{g_{\rm tunnel}}{i_{\rm GS}} \approx \frac{1}{v_{\rm GS}-V_T}+\frac{1}{v_{\rm GS}}+\alpha_{gtun}\cdot t_{\rm ox} \qquad t_{\rm ox}\,\text{in}\,[\text{nm}],\quad \alpha_{gtun}\approx 1/\text{nm}\cdot\text{V}$$

with (3) the gate current at frequencies lower than $f_{\rm gate}$ can be rewritten into

$$\begin{aligned} i_{\rm GS} &= \frac{2\pi C_{\rm ox}\cdot W\cdot L\cdot f_{\rm gate}}{g_{\rm tunnel}/i_{\rm GS}} \\ &= \frac{2\pi C_{\rm ox}\cdot W\cdot L\cdot f_{\rm gate}}{\frac{1}{v_{\rm GS}-V_T}+\frac{1}{v_{\rm GS}}+\alpha_{gtun}\cdot t_{\rm ox}} \quad t_{\rm ox}\,\text{in}\,[\text{nm}] \end{aligned} \tag{4}$$

Substituting the gate-oxide thickness and with typical voltages for analog applications yields the rule-of-thumb estimation of $i_{\rm GS}$ for frequencies below $f_{\rm gate}$

$$i_{\rm GS}\approx \beta_{i\,\rm gs}\cdot C_{\rm ox}\cdot W\cdot L\cdot f_{\rm gate} \quad \text{with}\ \beta_{i\,\rm gs}\approx 1\,\text{V}. \tag{5}$$

Now, using the well-known quadratic approximation for the drain current

$$i_D = \frac{W}{2L} \cdot \mu C_{\rm ox} \cdot (v_{\rm GS} - V_T)^2$$

it follows that at frequencies lower than $f_{\rm gate}$, the current gain of an MOS transistor in strong inversion and saturation is given by (6). A similar relation can be derived for the weak-inversion region:

$$\frac{i_D}{i_{\rm GS}} \approx \frac{1}{L^2} \cdot \frac{\mu \cdot (v_{\rm GT} - V_T)^2}{2 \cdot \beta_{i\,\rm gs} \cdot f_{\rm gate}}. \tag{6}$$

In (6) both the carrier mobility and $f_{\rm gate}$ are about a factor 3 different for NMOS and PMOS transistors; the mobility ratio is specific to silicon, the $f_{\rm gate}$ ratio is due to oxide barrier differences and therefore also material-related. This leads to (7) for both types of transistors:

$$\frac{i_D}{i_{\rm GS}} \cong \vartheta \cdot \frac{v_{\rm GT}^2}{v_{\rm GS}^2 \cdot e^{t_{\rm ox} v_{\rm GS}} \cdot L^2}$$
$$\vartheta = 7 \cdot 10^{-6} \cdot e^{13.6 \cdot t_{\rm ox}} \quad L \text{ in } [\mu\text{m}]. \tag{7}$$

Fig. 7 shows the results of (7) and measurement results for a few transistors in two leaky technologies (gate currents are estimated for the 65-nm generation) at low effective gate-overdrive voltages. Clearly visible from this figure are the strong length dependence and the low current-gain for long transistors in ultra-deep-submicron CMOS technologies. It follows that long transistors cannot be usefully applied anymore in UDSM technologies. The figure also illustrates that the square-law estimation used in the derivation is adequate.

B. Self-Discharge Effects and Droop Rates

A large number of circuit designs apply MOS capacitances for storing charge. Examples include switched-current circuits, PLL loop filters, hold circuits and some switched capacitor circuits. Gate leakage causes a nonzero droop rate of the voltage across MOS capacitances (see Fig. 8) and thereby puts a bound on the maximum usable hold time and the minimum operating frequency. The droop rate of the leaky gate capacitance is

$$\frac{dv_C}{dt} = -\frac{i_G}{C_{\rm in}} = -\frac{i_G}{g_{\rm tunnel}} \cdot \frac{g_{\rm tunnel}}{C_{\rm in}}.$$

With the relation derived in Appendix C it follows that the droop rate in strong inversion is to a good approximation given by (8). A similar relation follows in weak inversion.

$$\frac{dv_C}{dt} \approx -\frac{2\pi \cdot f_{\rm gate}}{\frac{1}{v_{\rm GS}-V_T} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox}}$$
$$t_{\rm ox} \text{ in } [\text{nm}], \quad \alpha_{gtun} \approx 1/\text{nm} \cdot \text{V}. \tag{8}$$

For typical UDSM gate-oxide thicknesses and typical analog operating conditions a rule-of-thumb estimation for the droop rate of MOS capacitances is given in (9):

$$\frac{dv_C}{dt} \approx -\gamma_{dv\,dt} \cdot f_{\rm gate} \left[\frac{V}{s}\right] \quad \text{with } \gamma_{dv\,dt} \approx 1\,\text{V}. \tag{9}$$

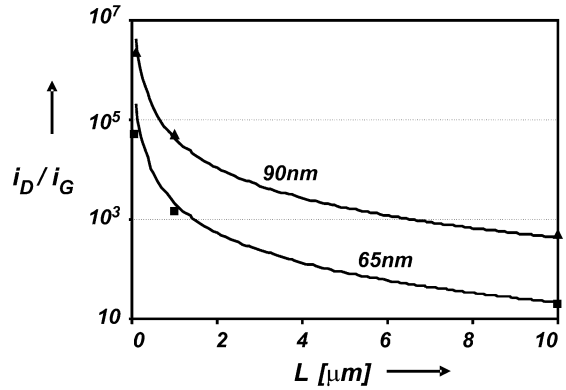


Fig. 7. Low-frequency current gain of MOS transistors in advanced CMOS technologies as a function of gate length, at $V_{\rm GS} = 0.5$ V. The curves follow from (7), 90-nm markers are based on measurements.

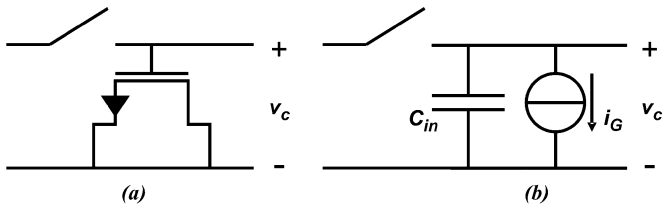


Fig. 8. (a) An MOS capacitance used in an (idealized) track-and-hold circuit, and (b) the equivalent circuit in "hold."

In words, the droop rate of a "stored" voltage on a MOS capacitor, in [V/s], is approximately equal to $f_{\rm gate}$ (in [Hz]). This relation implies that for track-and-hold circuits the maximum hold time for a droop amounting to ΔV is

$$\Delta t \approx \frac{\Delta V}{\gamma_{dv\,dt} \cdot f_{\rm gate}} \quad [\text{s}] \tag{10}$$

Allowing, e.g., 1-mV drop on a sampled-and-held value, the maximum usable hold time is in the millisecond range in 180-nm technologies, which is usually sufficient. However the maximum hold time decreases rapidly with newer technologies, down to a typical value in the low nano second range for 65-nm technologies. Note that this low maximum hold time makes it impossible to apply MOS capacitors in low and medium sample-rate A/D converters. Capacitors must then be realized either using thick-oxide devices, or inter-metal capacitances. If one can only use standard thin-oxide transistors as capacitances, PMOS transistors are half an order better than the NMOS transistors. Similar conclusions hold for PLL loop filters and switched-current circuits.

C. Gate-Leakage Matching and Its Implications

Gate leakage is caused by quantum-mechanical tunneling and depends on the layer thickness and the field strength. As such, it also exhibits spread. Relative spread, or matching, usually limits the achievable level of performance of analog circuits: it sets a lower bound on figures such as offsets in amplifiers and the accuracy in A/D converters.

Because spread and mismatch are dc effects, they do not (from a fundamental point of view) require any additional

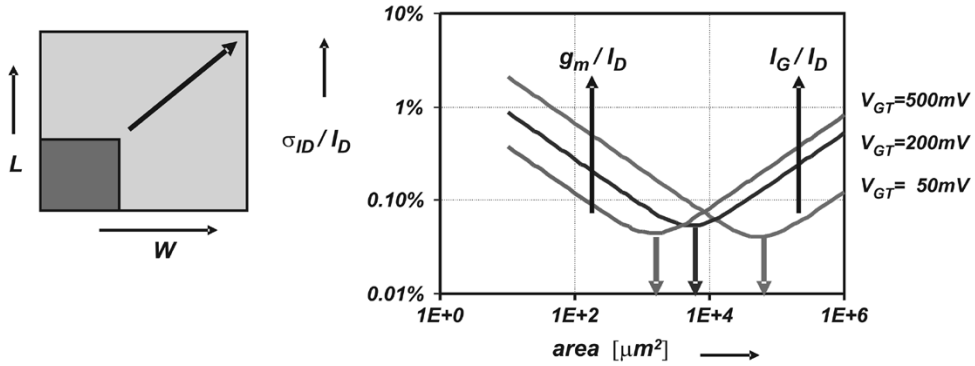


Fig. 9. The spread of an MOS transistor in 65-nm CMOS with linear scaling of W and L as a function of area: there are clear optima in the attainable matching.

power. However, in practice they prove to be a major implementation problem [12]. The usual way to get a sufficient level of matching between MOS transistors is to simply spend area [14], thereby increasing power consumption at a given speed, because larger capacitances have to be charged [15].

Compared to the conventional mismatch sources, gate-leakage mismatch now comes as an extra mismatch source with a *different* area dependency. We found that, excluding defect-like outliers, mismatch of gate leakage current is proportional to the gate current level with a proportionality constant of roughly $0.03/\sqrt{Area}$, where $Area$ is the transistor's gate area in square-microns. Assuming that conventional mismatch and gate current mismatch are uncorrelated, the total relative mismatch of a transistor's drain current is roughly

$$\frac{\sigma_{id}^2}{i_D^2} = \frac{\sigma_{id,\text{conventional}}^2}{i_D^2} + \frac{\sigma_{id,\text{gate leakage}}^2}{i_D^2} = \left(\frac{A_{\text{VT}}}{\sqrt{WL}} \cdot \frac{g_m}{i_D}\right)^2 + \left(\frac{X_{\text{IGS}}}{\sqrt{WL}} \cdot \frac{i_G}{i_D}\right)^2 \quad (11)$$

where $X_{\text{IGS}} \approx 0.03$.

The first term is the conventional mismatch due to mismatch in threshold voltage, with A_{VT} the matching coefficient [12]–[14]. This A_{VT} is a technology-related factor that is roughly proportional to the gate-oxide thickness, saturating in UDSM technologies around 2–3 mVμm [13]. The second term is the mismatch in gate current as introduced here. Note that the mismatch in the current factor β of the transistors is neglected, which is allowed for practical values of V_{GT} [14].

1) Improving Matching: The Classical Way: The classical way—without incorporating gate current mismatch [14]—to decrease mismatch is to spend area and to set an optimum g_m/I_D ratio. For a number of applications the input referred mismatch of transistors $\sigma_{vgs} = \sigma_{id}/g_m$ is relevant; minimization of input-referred mismatch typically comes down to maximizing g_m/I_D or increasing the gate-area in some way. Decreasing output referred mismatch σ_{id} can be achieved by lowering g_m/I_D or by increasing the gate-area of the transistor. By linearly scaling W *and* L of a transistor, its bias settings are unchanged while the matching improves proportionally to the scale factor. When scaling only device widths, and keeping L constant the current level increases proportionally and matching improves as the square root of the width scale factor [16].

2) Improving Matching: Including Gate-Leakage Effects: If gate leakage mismatch is accounted for, like in (11), the conventional mismatch decreasing rules cannot be applied any more. The impact of gate-leakage mismatch on the overall mismatch of MOS transistors is again best illustrated using the square-law relation; this relation is sufficient for rough estimation purposes. With expression (7) for the current gain of a MOS transistor it follows that

$$\frac{\sigma_{ID}^2}{i_D^2} = \left(\frac{\varsigma}{\sqrt{WL}}\right)^2 + \left(\frac{\xi \cdot L^2}{\sqrt{WL}}\right)^2 \quad (12)$$

where ς is related to conventional mismatch and ξ is related to gate current mismatch. In the classical way the matching could be improved by spending more area in any way. In UDSM technologies, (12) shows that by spending more area the conventional mismatch contribution always decreases but that at the same time the gate-mismatch contribution may increase. This latter contribution increases if the transistor length increases, as is the case with linear scaling of W and L of the transistors. Hence, with linear scaling by increasing W and L, the total relative mismatch in drain current may increase for large gate areas, which effectively limits the maximum usable transistor area. For 180-nm and 120-nm CMOS technologies this maximum usable area is very large and hence the attainable levels of matching are very good. However, for the 90-nm (measured matching) and 65-nm (estimated matching) generation this yields maximum usable areas in the order of respectively 10^4 μm^2 and 10^3 μm^2: then gate-leakage mismatch is a significant effect that limits the attainable matching. This is illustrated in Fig. 9. In this case, active mismatch cancellation techniques or matching-insensitive designs are required.

On the other hand, scaling only the transistor width and scaling the current levels in proportion, both the gate-leakage contribution and the conventional matching term improve with the square root of the scale factor. This is illustrated in Fig. 10. In this case essentially any level of matching can be obtained at the expense of area and power consumption. Active matching cancellation techniques are not required here, but can be used to break the area–power-matching relation.

D. Noise

Just as any current across a junction, gate leakage exhibits shot noise with current density $S_{IG} = 2q \cdot I_G$. As such, it is

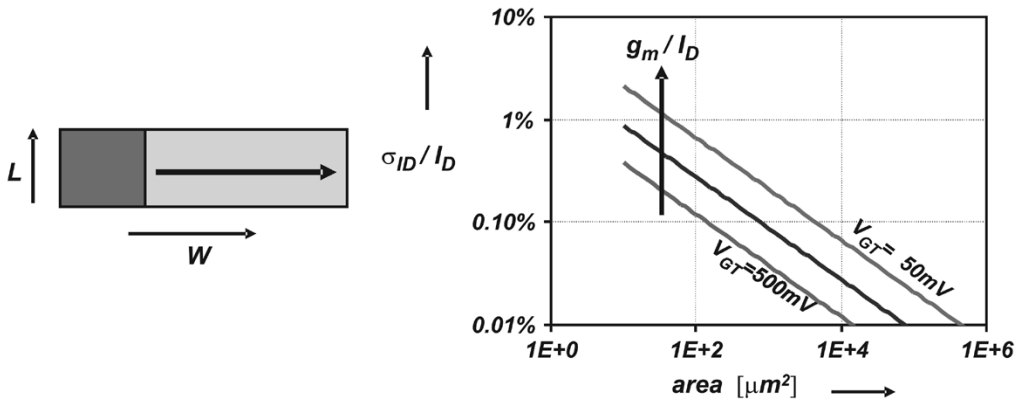


Fig. 10. The spread of an MOS transistor in 65-nm CMOS with width scaling, and constant L as a function of area: no minimum in the attainable matching at the cost of increased power consumption.

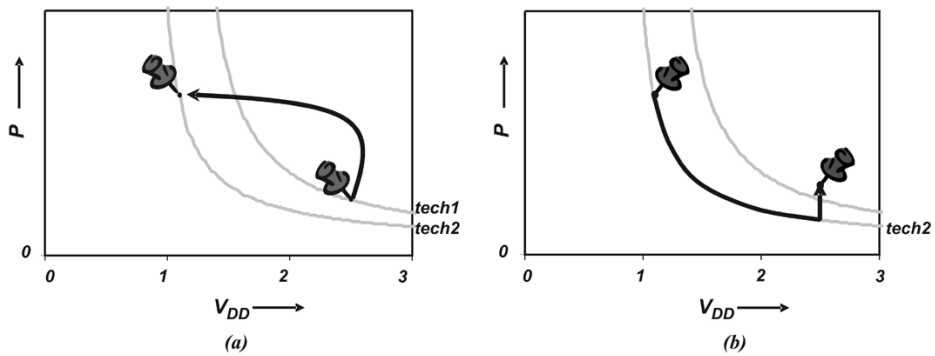


Fig. 11. Porting older designs to newer CMOS technology: (a) with supply voltage down scaling with a significant power consumption increase, and (b) a high-voltage version of the original circuit can operate at higher-than-nominal supply voltages with its associated lower power consumption. Implementation overhead to enable high-voltage operation increases power consumption.

equivalent to base currents in bipolar transistors. Note that at RF frequencies effects like the induced gate noise also contribute to the total gate noise [17], [18]. Noise in the gate current will therefore limit noise performance in analog circuits in UDSM CMOS.

VI. A Solution: Living Outside Rails

The two major problems associated with analog circuit design in UDSM technologies are the low supply voltage and the gate-leakage related effects. One strategy to deal with the low-supply drawback is to operate critical parts of analog circuits at a supply voltage significantly higher than the nominal supply voltage for the CMOS process used. This typically reduces the power consumption significantly at a given level of performance [19]–[21], but requires a focus on lifetime issues such as oxide breakdown [22], hot carriers [23], [24], NBTI [25] and junction breakdown [26]. Generally junction breakdown is not a major issue while hot carrier degradation does not play a significant role at supply voltages lower than 1 V. The other two effects need to be limited by a suitable limitation of terminal-pair voltages. Techniques known from high-voltage I/O circuits can be readily used for this. A brief review is presented at the end of this section.

Analog Circuits at High Supply Voltages: The effect of operation of analog circuits in UDSM CMOS at a high supply voltage, up to a few times as high as the nominal supply voltage $V_{\mathrm{DD,nom}}$ for the process, is illustrated in Fig. 11. Both curves in Fig. 11(a) and (b) show the minimum power consumption as a function of the supply voltage, for a circuit in a CMOS technology at constant performance, with optimized bias settings and device dimensions for each technology and supply voltage [6]. In these figures, the upper curves (tech1) correspond to an older technology while the lower curve (tech2) is a newer technology. For analog, the migration to a newer CMOS technology with its associated lower nominal supply voltage results in increased power consumption at fixed performance, indicated by the a) arrow in Fig. 11(a). Note that the jump to another curve corresponds to going to another technology with the same circuit topology. However, implementing the circuit in such a way that it runs at a high supply voltage can significantly decrease power consumption; this is indicated by arrow b) in Fig. 11(b). For reliability reasons typically circuit overhead is required, resulting in the upward part of the b) arrow.

An obvious advantage of migration to more advanced CMOS technologies is that it enables selective application of low-voltage transistors with their specific advantages and disadvantages. Especially interesting is the digital computational power that can solve many deterministic analog inaccuracies (e.g., mismatch and distortion).

Running Analog Circuits at High Supply Voltages: For circuits operating at high supply voltages, a number of robust high-voltage-tolerant transistors can be used to replace the standard transistors that can only reliably operate up to nominal supply

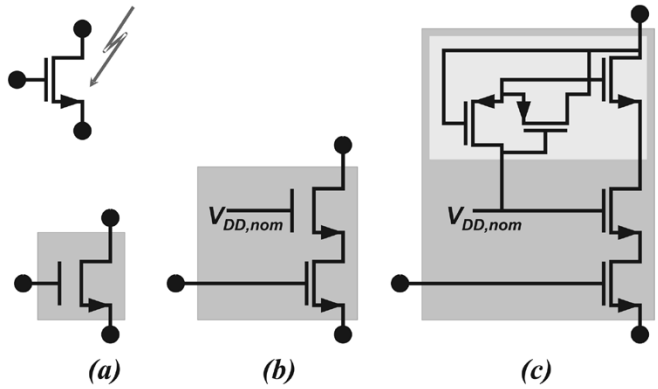


Fig. 12. Ways to implement high-voltage tolerant transistors in standard CMOS: (a) thick-oxide transistor; (b) (thick-oxide) cascode; (c) retractable cascode composite transistor.

voltages. Fig. 12 presents three examples known from high-voltage (HV) I/O circuits. A number of circuits and blocks with analog behavior implementing the retractable cascode HV transistors shown in Fig. 12(c) were discussed in [3] and [27]. These HV transistors enable direct reuse of most older circuit architectures, running at supply voltages corresponding to the original design: high supply voltages when related to the nominal supply voltage of the CMOS process used. Extended-drain transistors that can be realized in standard CMOS could also be used; such structures are described in [28].

The easiest way is to use the (commonly available) thick-oxide transistor [Fig. 12(a)] that is comparable to a two-generations-old standard transistor. However, in order to benefit from technology scaling compound structures using thin-oxide transistors, as in Fig. 12(b) or (c),[1] typically outperform the thick-oxide transistor in Fig. 12(a) in the fields of matching, $1/f$ noise and output impedance. These compound structures have some disadvantages: they are asymmetric, do not solve gate-leakage issues, and require suitable cascode voltages at power-up.

Careful selection of the analog sections to run at high supply voltages, and careful selection of the best type of transistor (thin oxide, thick oxide, or compound) will to a great extent circumvent one of the main roadblocks in UDSM CMOS technolgies: the low nominal supply voltage. Note that thick-oxide transistors also solve gate-leakage issues as their gate leakage is usually negligible.

VII. Conclusion

Modern and future UDSM CMOS introduce several new problems for analog circuit design. From a fundamental point of view, lowering the analog supply voltage leads to an increase in power dissipation at constant performance. This increase in power dissipation becomes drastic as the supply voltage approaches the threshold voltage plus a few hundred millivolts, as illustrated in Fig. 1.

[1]The retractable cascode structure typically uses one cascode with a fixed gate voltage, and one or more cascodes with variable gate voltages. In Fig. 12(c), the variable gate voltage is soft-switched by two PMOS transistors between the drain voltage of the switched transistor and the nominal supply voltage, whichever is highest.

When migrating to modern technologies the quasi-dc analog transistor properties hardly change, as long as constant transistor lengths and terminal voltages are used. However, if the supply voltage is reduced according to the technology roadmap, the analog performance is lowered because of the lower bias voltages. Nonlinear output conductance in combination with reduced voltage gain pose limits with respect to linearity of (feedback) circuits.

Gate leakage becomes a serious problem in upcoming technologies, especially if long transistors are used. A parameter $f_{\rm gate}$ is introduced that enables quick estimations of gate-leakage related effects. Besides gate leakage itself, mismatch in gate leakage introduces new limitations. Mismatch cannot be tackled anymore by simply spending more area for transistors: when the transistor length L is increased an upper limit to matching is encountered. Here, increasing area does not automatically improve overall matching anymore, except if higher power consumption is accepted or active cancellation techniques are applied.

Operating critical parts at higher supply voltages, by exploiting combinations of thin- and thick-oxide transistors can solve the low voltage as well as the gate leakage problems. Composite transistors are presented in Fig. 12 to solve this problem in a practical way. In summary: unlike digital designs, analog circuits benefit from technology scaling if the supply voltages are *not* scaled down.

Appendix A

In this Appendix, known performance–power relations for active circuits are briefly reviewed and their impact is discussed.

A. Performance in SNR: Total Integrated Noise

A number of papers on the relation between analog performance and power consumption specify the performance in only its signal-to-noise-ratio (SNR) and the signal bandwidth [4], [5]. Typically, only the total integrated thermal noise is taken into account. In these papers, a system as depicted in Fig. 13(a) is used: an unspecified analog circuit represented by a resistance or conductance.

Including only thermal noise integrated over the total noise bandwidth of the circuit, the integrated noise voltage and the required (class-A) bias current are

$$\bar{v}_n^2 = \frac{kT}{C} \quad \text{and} \quad I_{\rm bias} = 2\pi f_{\rm sig} C \hat{V}.$$

These equations lead to a minimum power consumption of an analog circuit given by

$$P = \frac{8\pi kT \cdot \mathrm{SNR} \cdot f_{\rm sig}}{\eta_{\rm vol} \cdot \eta_{\rm cur}}$$

where $\eta_{\rm vol}$ is the ratio between the peak-peak signal swing and the supply voltage [15], $\eta_{\rm cur}$ is the efficiency of using supply current [15], kT are Boltzmann's constant and the temperature, respectively, SNR is the circuit's signal-to-noise ratio as a power ratio, $f_{\rm sig}$ is the signal frequency, $\hat{V}$ is the signal amplitude.

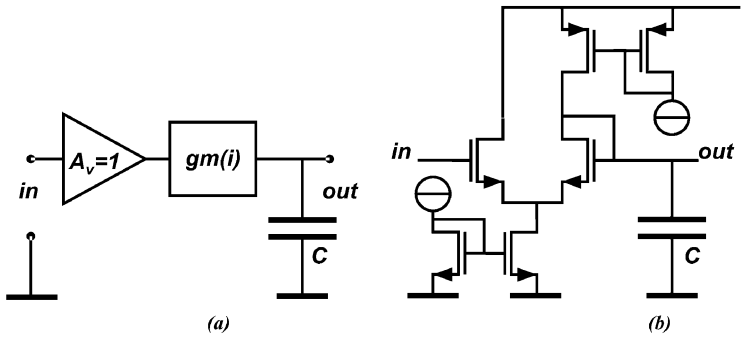


Fig. 13. (a) Circuit representation assumed for the power-performance relations: a circuit with an output resistance and an output load, and (b) actual circuit topology used in [6].

Taking into account only thermal noise, the power consumption required for some SNR is technology independent if both the parameters $\eta_{\rm vol}$ and $\eta_{\rm cur}$ are invariant over technology. In general, $\eta_{\rm vol}$ and $\eta_{\rm cur}$ increase with newer technologies because of voltage overhead (e.g., to accommodate gate-source overdrive voltage and saturation [15]) and due to the use of folded structures.

B. Performance in SNR: Bandwidth Limited Noise

For many circuits the total integrated thermal noise at the output is irrelevant: only the noise within some frequency band is of interest. In this case, neither the thermal noise nor the required bias current is related to the output capacitance: this capacitance is now purely parasitic. For this type of circuit, the integrated thermal noise voltage is[2]

$$\bar{v}_n^2 = \frac{4kT}{gm} BW$$

where BW is the relevant frequency band.

For ordinary active electronic components, the transconductance g_m is a function of the bias current and of some voltage. For bipolar transistors and MOS transistors

$$g_m = \frac{qI_{\rm BIAS}}{kT} \quad \text{and} \quad g_m \cong \frac{2I_{\rm BIAS}}{v_{GT}}$$

where a lower bound to the effective gate-source overdrive voltage v_{GT} for MOS transistors is weak-inversion (bipolar-like) operation. These two expressions can be captured in one: $g_m = I_{\rm BIAS}/v_{\rm OD}$ where $v_{\rm OD}$ is something like equivalent effective overdrive voltage. The minimum power consumption to reach a certain SNR is then

$$P = \frac{16kT \cdot {\rm SNR} \cdot BW}{\eta_{\rm vol} \cdot \eta_{\rm cur}} \cdot \frac{v_{\rm OD}}{\widehat{V}}.$$

It also follows that the power consumption is proportional to the targeted SNR. However, lowering the supply voltage (and signal swing) without decreasing the $v_{\rm OD}$ increases the required power consumption. Note that this situation always occurs with bipolar transistors and MOS transistors in moderate or weak inversion.

[2]Actually the noise should be corrected with the circuit noise excess factor. For simplicity reasons this is neglected here.

C. Performance in SINAD: Total Integrated Noise

In many analog circuits, both noise and distortion are relevant to the performance. For those circuits the signal-to-noise-and-distortion ratio (SINAD or SNDR) may be the right way to express the performance. Examples of circuits for which this type of performance is relevant include switched-capacitor circuits and track-and-hold circuits. The circuit corresponds to that in Fig. 13(b). In general, it is an unspecified analog circuit driving an explicit load capacitance as shown in Fig. 13(a). The actual analog circuit can be anything ranging from a nonlinear resistance to an analog amplifier with any amount of global or local feedback. In [6], it was shown that, taking the total integrated thermal noise into account, for weakly nonlinear analog circuits with a dominant load capacitor the absolute minimum power consumption is

$$P_{\rm min} = 2 \cdot \hat{V} \cdot gmi\left(2kT \cdot \pi \cdot f_{\rm sig} \sqrt[n]{\frac{2\alpha_n}{\hat{V}}} \cdot {\rm SINAD}^{\frac{2n+1}{2n}} \cdot \frac{(2n+1)^{\frac{2n+1}{2n}}}{n\hat{V}}\right)$$

where n is the ordinal number of the dominant harmonic, α_n links higher harmonics to the first harmonic[3] as $h_n = \alpha_n (h_1)^n$, and $gmi(g)$ is the inverse of the function between the conductance and bias current of a circuit.

In the derivation of this expression, distortion, and noise are traded against each other in such a way that a maximum performance–power consumption ratio is reached. Note that this general expression is much more complex than its SNR-based counterpart, presented in Section A. For ordinary weakly nonlinear analog circuits, substituting the $gmi(x)$ function and $\alpha_n$results in a circuit-specific SINAD-P relation with many similarities to the SNR-P relation mentioned above; see [6] for two examples. The biggest difference between the SNR expression and the SINAD expression is that the latter contains a multiplicative term that increases with decreasing signal swing, and hence with the lower supply voltage that comes with newer CMOS generations.

[3]These variables describe the nonlinearity of the analog circuit and follow from a Taylor series expansion of the dc-transfer curve of the circuit.

With the assumptions leading to the SNR limit (marginally preventing both clipping and slewing) straightforward mathematics shows that the complex SINAD-P expression collapses to the SNR-P one[4] in Section A:

$$\lim_{n\to\infty} P_{\min} = 8\pi kT f_{\rm sig}\,\mathrm{SINAD}|_{D\to 0} \equiv 8\pi kT f_{\rm sig}\,\mathrm{SNR}.$$

D. Performance in SINAD: Bandwidth Limited Noise

For circuits dealing with a SINAD performance specification, either the total integrated noise or bandwidth limited noise may be relevant. In the case of bandwidth-limited noise, there is no need for an explicit load capacitance. As a direct consequence of the absence of any required bandwidth limitation, from a mathematical point of view harmonic distortion is zero. The relation between power consumption and bandwidth-limited SINAD therefore equals the relation for power and bandwidth-limited noise under B.

E. Summary

From a fundamental point of view it can be concluded that lowering the supply voltage increases the power-performance ratio, with exception of the simplest case described under A. Moreover, any voltage overhead ΔV worsens the power-performance ratio with decreasing supply voltage $V_{\rm DD}$, typically introducing a multiplicative term $V_{\rm DD}/V_{\rm DD}-\Delta V$ in the power relation.

F. Exceptions to the Rule

Most analog circuits comply with the power-performance relations discussed here. Obvious exceptions to this rule are circuits that are overly robust in some aspect, for example most flash A/D converters that minimize mismatch issues by spending area, or that cannot satisfy some scaling issues, e.g., low-noise amplifiers.

Flash A/D Converters: Practical findings indicate that the power consumption in flash A/D converters is not determined by thermal noise issues, see, e.g., [15]. Typically, low-resolution flash converters aim to reach a certain level of matching, by spending area, at some operating frequency. Under these conditions matching and speed requirements determine the power consumption, and consequently the power consumption of flash A/D converters decreases with newer CMOS generations because of better matching properties. However, when using active matching techniques [15], [29], [30], the need to spend area for matching is absent and the power–performance relation is determined by SNR issues again [15]. Note that it is a fundamental property of dc-type disturbances that no power is needed for their minimization. The apparent independence of SNR and power consumption in flash A/D converters is therefore due to the practical way mismatch is dealt with.

Low-Noise Amplifiers: Other circuits that do not comply with the discussed power-performance relations include RF low-noise amplifiers (LNAs). In the derivation of these relations, the signal swing is optimized for a given supply voltage. However, in LNA-type circuits the signal swing is fixed and lower supply voltages result in somewhat degraded bias circuitry and in a lower implementation overhead [15]. The overall result is that for circuits with a fixed very low signal swing, the power–performance ratio can improve with newer CMOS generations [31], [32].

[4]This is true for at least most of the $g_m(I)$ functions that can be realized in standard electronics.

Appendix B

Using a simplified relation for gate current based on the gate current model in MOS Model 11 [33], we can write the following relation for a MOSFET in saturation. In this relation, the effects of overlap regions are neglected:

$$i_{\rm GS} = A\cdot v_{\rm INV}\cdot v_{\rm GS}\cdot\exp(B\cdot v_{\rm GS})$$

where $v_{\rm INV}$ is the effective gate bias, given by

$$v_{\rm INV} = m\cdot\varphi_T\cdot\ln\left(1+\exp\left[\frac{v_{\rm GS}-V_T}{m\cdot\varphi_T}\right]\right)$$

and A and B are constants given by

$$A = \frac{I_{\rm GINV}}{2}\cdot\exp\left[-\frac{3}{2}\cdot\frac{B_{\rm INV}}{\chi_B}\right] \quad\text{and}\quad B = \frac{3}{8}\cdot\frac{B_{\rm INV}}{\chi_B^2}.$$

In the above, m determines the subthreshold slope ($\mathrm{m}=1.3$), χ_B is the oxide potential barrier ($\chi_B = 3.1$ V for electrons, $\chi_B = 4.5$ V for holes), and $I_{\rm GINV}$ and $B_{\rm INV}$ are physical parameters dependent on oxide thickness $t_{\rm ox}$, channel length L and channel width W. For electrons, we can write

$$I_{\rm GINV} = 1.6\cdot 10^{-4}\cdot\frac{WL}{t_{\rm ox}^2} \quad t_{\rm ox}\text{ in [m]}$$
$$B_{\rm INV} = 2.9\cdot 10^{10}\cdot t_{\rm ox} \quad t_{\rm ox}\text{ in [m]}.$$

As a simple approximation, we get the following expressions for the factors A and B, now with the oxide thickness in [nm] and assuming NMOS transistors. Note that gate current is proportional to the total gate area.

$$A = WL\cdot\frac{1.6\cdot 10^{14}}{t_{\rm ox}^2}\cdot\exp(-14\cdot t_{\rm ox}) \quad t_{\rm ox}\text{ in [nm]}$$
$$B = 4.5\cdot t_{\rm ox}.$$

For the input capacitance $C_{\rm GG}$, we find (also neglecting the overlap regions) the following relation, where $C_{\rm OX}$ is the total oxide capacitance. Note that this term is also proportional to the gate area.

$$C_{GG} = \frac{2}{3}\cdot C_{\rm OX}\cdot\frac{\partial v_{\rm INV}}{\partial v_{\rm GS}} = \frac{2}{3}\cdot\frac{WL\cdot\varepsilon_{\rm ox}}{t_{\rm ox}}\cdot\frac{\partial v_{\rm INV}}{\partial v_{\rm GS}}$$

with

$$\frac{\partial v_{\rm INV}}{\partial v_{\rm GS}} = \left(1+\exp\left[-\frac{v_{\rm GS}-V_T}{m\cdot\varphi_T}\right]\right)^{-1}.$$

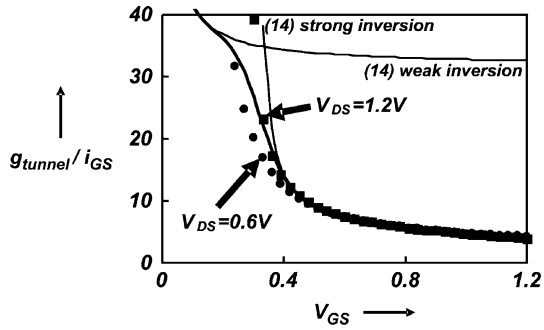


Fig. 14. Ratio $g_{\rm tunnel}/i_{\rm GS}$ as a function of gate bias $v_{\rm GS}$. Markers are measurement results and curves are predicted results using (13) and both expressions of (14).

The frequency where the imaginary part and the real part of the input impedance are equal is

$$f_{\rm gate} = \frac{1}{2 \cdot \pi \cdot C_{GG}} \cdot \frac{\partial i_{\rm GS}}{\partial v_{\rm GS}}$$

with

$$\frac{\partial i_{\rm GS}}{\partial v_{\rm GS}} = A \cdot \exp(B \cdot v_{\rm GS}) \cdot \left[v_{\rm INV} \cdot (1 + B \cdot v_{\rm GS}) + v_{\rm GS} \cdot \frac{\partial v_{\rm INV}}{\partial v_{\rm GS}} \right].$$

it follows that in strong inversion and saturation, where $v_{\rm INV} \approx v_{\rm GS} - V_T$ and $\partial v_{\rm INV}/\partial v_{\rm GS} = 1$ and $C_{GG} = (2/3) \cdot C_{\rm OX}$, $f_{\rm gate}$ for NMOS transistors can be approximated by

$$f_{\rm gate} \cong \frac{\vartheta_{fgate} \cdot v_{\rm GS}^2}{\exp(\upsilon_{fgate} \cdot t_{\rm ox})} \exp(\zeta_{fgate} \cdot t_{\rm ox} v_{\rm GS}) \quad t_{\rm ox} \text{ in } [\rm nm]$$
$$\vartheta_{fgate} \approx 1.5 \cdot 10^{16}\ \rm Hz/V^2$$
$$\upsilon_{fgate} \approx 13.6/\rm nm$$
$$\zeta_{fgate} \approx 1/\rm nm \cdot V.$$

Similar relations can be derived for the moderate and weak inversion regions, and the linear region. For PMOS transistors $f_{\rm gate}$ is roughly a factor 3 lower due to the higher oxide potential barrier $\chi_{\rm B}$.

APPENDIX C

The relation between gate conductance and gate current can be used in a number of expressions that link some gate-leakage related property to the size-independent $f_{\rm gate}$. In this paper the properties discussed are the dc-current gain and the self-discharge droop-rate of MOS-capacitances. Using Appendix B, the tunnel conductance $g_{\rm tunnel} = \partial i_{\rm GS}/\partial V_{\rm GS}$ can be readily calculated. The normalized gate conductance is then

$$\frac{g_{\rm tunnel}}{i_{\rm GS}} = \frac{\partial v_{\rm INV}/\partial v_{\rm GS}}{v_{\rm INV}} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox} \quad t_{\rm ox} \text{ in } [\rm nm], \quad \alpha_{gtun} \approx 1.13/\rm nm. \quad (13)$$

In strong inversion and weak inversion, respectively, this relation can be approximated by the following ones. Note that the dimensions are correct because of implicit multiplication by appropriate scale factors.

$$\frac{g_{\rm tunnel}}{i_{\rm GS}} \approx \frac{1}{v_{\rm GS} - V_T} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox} \quad t_{\rm ox} \text{ in } [\rm nm], \quad \alpha_{gtun} \approx 1.13/\rm nm$$
$$\frac{g_{\rm tunnel}}{i_{\rm GS}} \approx \frac{1}{m \cdot \varphi_T} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox} \quad t_{\rm ox} \text{ in } [\rm nm], \quad \alpha_{gtun} \approx 1.13/\rm nm. \quad (14)$$

As an example, Fig. 14 shows measured data for a 10 μm $*$ 10 μm transistor in a 120-nm technology, with the more exact expression and the two approximations for the weak inversion and strong inversion region. Clearly the simple relations comply very well to measurements.

ACKNOWLEDGMENT

The authors would like to thank J. Schmitz for fruitful discussions on lifetime issues, A. Kumar and A. Heringa from Philips Research for providing the 65-nm CMOS parameters used in this paper, and the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] Y. Taur, "CMOS design near the limit of scaling," *IBM J. Res. & Dev.*, vol. 46, no. 2/3, pp. 213–222, 2002.
[2] D. D. Buss, "Technology in the internet age," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2002, pp. 18–21.
[3] A. J. Annema, B. Nauta, R. van Langevelde, and H. Tuinhout, "Designing outside rails constraints," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2004, pp. 134–135.
[4] J. O. Voorman and D. Blom, "Noise in gyrator capacitor filters," *Philips' Res. Rep.*, vol. 26, no. 4, pp. 114–133, 1971.
[5] E. A. Vittoz, "Low-power design: Ways to approach the limits," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 1994, pp. 14–18.
[6] A. J. Annema, "Analog circuit performance and process scaling," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 6, pp. 711–725, 1999.
[7] D. Foty, D. Binkley, and M. Bucher, "Starting over: $\rm g_m/I_d$-based MOSFET modeling as a basis for modernized analog design methodologies," in *Proc. Nanotech*, vol. 1, 2002, pp. 682–685.
[8] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. 9, pp. 256–268, 1974.
[9] H. Iwai, "CMOS technology—Year 2010 and beyond," *IEEE J. Solid-State Circuits*, vol. 34, no. 3, pp. 357–366, Mar. 1999.
[10] M. Garg, S. S. Suryagandh, and J. S. Woo, "Scaling impact on analog performance of sub-100 nm MOSFET's for mixed mode applications," in *Proc. ESSDERC*, 2003, pp. 371–374.
[11] R. van Langevelde, A. J. Scholten, R. Duffy, F. N. Cubaynes, M. J. Knitel, and D. B. M. Klaassen, "Gate current: Modeling, ΔL extraction and impact on RF performance," in *IEDM Tech. Dig.*, 2001, pp. 289–292.
[12] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications," in *IEDM Tech. Dig.*, 1998, pp. 915–918.
[13] J. Dubois, J. Knol, M. Bolt, H. Tuinhout, J. Schmitz, and P. Stolk, "Impact of source/drain implants on threshold voltage matching in deep-submicron CMOS technologies," in *Proc. ESSDERC*, 2002, pp. 115–118.
[14] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 10, pp. 1433–1440, Oct. 1989.
[15] K. Bult, "Analog design in deep-submicron CMOS," in *Proc. ESSCIRC*, 2000, pp. 11–17.
[16] E. Klumperink and B. Nauta, "Systematic comparison of HF CMOS transconductors," *IEEE Trans. Circuits Syst. II*, vol. 50, no. 10, pp. 728–741, Oct. 2003.

[17] A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven, and V. C. Venezia, "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618–632, Mar. 2003.

[18] R. van Langevelde, J. C. J. Paasschens, A. J. Scholten, R. J. Havens, L. F. Tiemeijer, and D. B. M. Klaassen, "New compact model for induced gate current noise," in *IEDM Tech. Dig.*, 2003, pp. 867–870.

[19] T. A. F. Duisters and E. C. Dijkmans, "A -90-dB THD rail-to-rail input opamp using a new local charge pump in CMOS," *IEEE. J. Solid-State Circuits*, vol. 33, no. 7, pp. 947–955, Jul. 1998.

[20] E. van Tuijl, J. van den Homberg, D. Reefman, C. Bastiaansen, and L. van de Dussen, "A 128 fs multi-bit $\Sigma\Delta$ CMOS audio DAC with real-time DEM and 115 dB SFDR," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2004, pp. 368–369.

[21] M. Keskin, U. K. Moon, and G. C. Temes, "A 1-V 10-MHz clock-rate 13-bit CMOS $\Sigma\Delta$ modulator using unity-gain-reset opamps," *IEEE J. Solid-State Circuits*, vol. 37, no. 7, pp. 817–824, Jul. 2002.

[22] "Special issue on scaling limits of gate oxides," *Semicond. Sci. Technol.*, vol. 15, no. 5, 2000.

[23] P. Heremans, R. Bellens, G. Groeseneken, and H. Maes, "Consistent model for the hot-carrier degradation in n-channel and p-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 35, no. 12, pp. 2194–2209, Dec. 1988.

[24] R. Woltjer, G. Paulzen, H. Pomp, H. Lifka, and P. Woerlee, "Three hot-carrier degradation mechanisms in deep-submicron PMOSFET's," *IEEE Trans. Electron Devices*, vol. 42, no. 1, pp. 109–115, Jan. 1995.

[25] T. Yamamoto, K. Uwasawa, and T. Mogami, "Bias temperature instability in scaled p+ polysilicon gate p-MOSFET's," *IEEE Trans. Electron Devices*, vol. 46, no. 5, pp. 921–926, May 1999.

[26] G. A. M. Hurkx, H. C. de Graaff, W. J. Kloosterman, and M. P. G. Knuvers, "A new analytical diode model including tunneling and avalanche breakdown," *IEEE Trans. Electron Devices*, vol. 39, pp. 2090–2098, 1992.

[27] A. J. Annema, G. G. Geelen, and P. C. de Jong, "5.5 V tolerant I/O in a 2.5 V 0.25 μm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 528–538, Mar. 2001.

[28] H. Ballan and M. Declerq, *High Voltage Devices and Structures in Standard CMOS Technologies*. Boston, MA: Kluwer, 1999.

[29] K. C. Dyer, D. Fu, S. H. Lewis, and P. J. Hurst, "An analog background calibration technique for time-interleaved analog-to-digital converters," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 1912–1919, Dec. 1998.

[30] H. V. D. Ploeg, G. Hoogzaad, H. A. H. Termeer, M. Vertregt, and R. L. J. Roovers, "A 2.5-V 12-b 54-Msample/s 0.25-μm CMOS ADC in 1-mm^2 with mixed-signal chopping and calibration," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 1859–1867, Dec. 2001.

[31] Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The impact of scaling down to deep submicron on CMOS RF circuits," *IEEE J. Solid-State Circuits*, vol. 33, no. 7, pp. 1023–1035, Jul. 1998.

[32] P. H. Woerlee, M. J. Knittel, R. van Langevelde, D. B. M. Klaassen, L. F. Tiemeijer, A. J. Scholten, and A. T. A. Zegers-van Duijnhoven, "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, Aug. 2001.

[33] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical background of MOS model 11, level 1101," Philips Electronics N.V., 2003. NL-TN 2003/00239, http://www.semiconductors.philips.com/Philips_Models/mos_models.

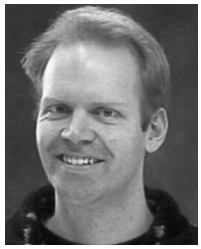
Anne-Johan Annema (M'00) received the M.Sc. degree in electrical engineering and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1990 and 1994, respectively.

In 1995, he joined Philips Research, Eindhoven, The Netherlands, where he worked on a number of electronics-physics-related projects ranging from low-power low-voltage circuits, fundamental limits on analog circuits in conjunction with process technologies, high-voltage in baseline CMOS to feasibility research of future CMOS processes for analog circuits. Since 2000, he has been with the IC Design Group, Department of Electrical Engineering, University of Twente, Enschede, The Netherlands. He is also a part-time consultant in industry. In 2001, he co-founded ChipDesignWorks. His current research interest is in physics, analog and mixed-signal electronics, and deep-submicrometer technologies and their joint feasibility aspects.

Bram Nauta (S'89–M'91) was born in Hengelo, The Netherlands, in 1964. In 1987, he received the M.Sc. degree (*cum laude*) in electrical engineering from the University of Twente, Enschede, The Netherlands. In 1991, he received the Ph.D. degree from the same university on the subject of analog CMOS filters for very high frequencies.

In 1991, he joined the Mixed-Signal Circuits and Systems Department, Philips Research, Eindhoven, the Netherlands, where he worked on high-speed A/D converters. From 1994, he led a research group in the same department, working on analog key modules. In 1998, he returned to the University of Twente, as full Professor heading the IC Design Group in the MESA+ Research Institute and Department of Electrical Engineering. His current research interest is analog CMOS circuits for transceivers. He is also a part-time consultant in industry. In 2001, he co-founded Chip Design Works. His Ph.D. thesis was published as a book: *Analog CMOS Filters for Very High Frequencies* (Boston, MA: Kluwer, 1993). He holds 11 patents in circuit design.

Dr. Nauta received the Shell Study Tour Award for his Ph.D. work. From 1997 to 1999, he served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: ANALOG AND DIGITAL SIGNAL PROCESSING, and in 1998 he served as Guest Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS. In 2001, he became an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS and he is also a Member of the Technical Program committee of ESSCIRC and ISSCC. He is a corecipient of the IEEE ISSCC 2002 Van Vessem Outstanding Paper Award.



Ronald van Langevelde (S'94–M'95) received the M.Sc. degree in electrical engineering in 1994 and the Ph.D. degree in 1998, both from the Eindhoven University of Technology, The Netherlands.

He is currently with Philips Research Laboratories, Eindhoven, The Netherlands, where he has developed the surface-potential-based MOS Model 11. His research interests include MOSFET device physics, circuit-level MOSFET modeling, and distortion analysis in circuit design.

Hans P. Tuinhout received the M.Sc. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 1980.

Since then, he has worked at Philips Research, Eindhoven, The Netherlands, on process and device characterization for silicon CMOS and BiCMOS integrated circuit technologies. His current research activities are focused on high precision dc I-V measurements for characterization of device mismatch.