Low Power Analog Design in Scaled Technologies

A. Baschirotto\textsuperscript{1,2}, V. Chironi\textsuperscript{2}, G. Cocciolo\textsuperscript{2}, S. D’Amico\textsuperscript{2}, M. De Matteis\textsuperscript{2}, P. Delizia\textsuperscript{2}

\textsuperscript{1}Dept. of Physics “G. Occhialini”
University of Milano-Bicocca
Milan – Italy

\textsuperscript{2}Dept. of Innovation Engineering
University of Salento
Lecce – Italy

Abstract

In this paper an overview on the main issues in analog IC design in scaled CMOS technology is presented. Decreasing the length of MOS channel and the gate oxide has led to undoubted advantages in terms of chip area, speed and power consumption (mainly exploited in the digital parts). Besides, some drawbacks are introduced in term of power leakage and reliability. Moreover, the scaled technology lower supply voltage requirement has led analog designers to find new circuital solution to guarantee the required performance.

I. INTRODUCTION

The evolution of the analog performance of MOS devices through technology scaling can be seen in Table I (\cite{6}) for the most important parameters. The influence of these and other effects will be discussed in the next sections.

Table I – MOS DEVICE PARAMETER TRENDS

\begin{tabular}{|c|c|c|c|c|c|}
\hline
Node & \text{Nm} & 250 & 180 & 130 & 90 & 65 \\
\hline
L_{\text{GATE}} & \text{Nm} & 180 & 130 & 92 & 63 & 43 \\
\hline
\text{t}_{\text{OX (inv.)}} & \text{Nm} & 6.2 & 4.45 & 3.12 & 2.2 & 1.8 \\
\hline
\text{Peak} g_{\text{m}} & \mu S/\mu m & 335 & 500 & 720 & 1060 & 1400 \\
\hline
g_{\text{ds}} & \mu S/\mu m & 22 & 40 & 65 & 100 & 230 \\
\hline
\text{g}_n g_{\text{ds}} & - & 15.2 & 12.5 & 11.1 & 10.6 & 6.1 \\
\hline
V_{\text{DD}} & V & 2.5 & 1.8 & 1.5 & 1.2 & 1 \\
\hline
V_{\text{TH}} & V & 0.44 & 0.43 & 0.34 & 0.36 & 0.24 \\
\hline
f_T & GHz & 35 & 53 & 94 & 140 & 210 \\
\hline
\end{tabular}

II. CMOS TECHNOLOGY MAIN TRENDS

A. Power Reduction

In digital CMOS circuits, the power consumption is mainly due to three current components: (i) the leakage current due to the reverse biased diodes formed between the substrate, the well, and the source and drain diffusion regions of the transistors, (ii) the short circuit current due to the presence of
current carrying path from the supply voltage to ground when certain PMOS and NMOS transistors are simultaneously ON for a short period due to the signal transitions at the input to the logic gates, and (iii) switching current due to charging and discharging of the load capacitance. Among the three sources of power dissipation, the last component is by far the most dominant. Ignoring the internal capacitances of logic gates, the average power consumption for a logic gate due to charging and discharging of load capacitance $C$ is given by:

$$ P_{dig} = f C V_{DD}^2 $$

where $V_{DD}$ is the supply voltage, and $f$ is the operation frequency. From Eq. 1, the power consumption in digital circuits is reduced in scaled technology.

In analog circuits, the performances are often limited by the thermal noise (this is the case, for instance, of an acquisition channel), which is inversely proportional to the bias current, i.e.:

$$ \frac{kT}{C} = \frac{\alpha}{I}, \quad I = \frac{P_{an}}{(\beta \cdot V_{DD})} $$

where $\alpha$ and $\beta$ are two constants properly sized.

To achieve a target Dynamic-Range (DR) with a maximum output swing (i.e. signal amplitude) of $SW = V_{DD} - 2V_{sat}$, this can be written as:

$$ DR = \frac{(V_{DD} - 2V_{sat})^2}{(\alpha / I)} $$

the power consumption in the analog circuits depends on DR and $V_{DD}$ as follows:

$$ P_{an} \propto \frac{DR}{V_{DD}} $$

As a consequence, for a given DR, $V_{DD}$ reduction, as required by technology scaling, brings an increase of analog power consumption. This result is the opposite than that for digital circuits. Thus, the technology scaling results detrimental for analog circuits design.

### B. $(V_{DD}$–$V_{TH})$ Reduction

Technology scaling forces a reduction of both $V_{DD}$ (as seen before) and $V_{TH}$. However $V_{TH}$ scales faster that $V_{DD}$, and this reduces node by node the distance $(V_{DD}$–$V_{TH})$. From an intuitive point of view, the distance $(V_{DD}$–$V_{TH})$ represents the “free” voltage space for analog design. The reduction of this space in scaled technologies makes critical the analog block design.

Considering, for instance, the analog switch realized with a pass-gate, as shown in Fig. 3, it can process a rail-to-rail input signal only with a minimum $V_{DD, min}$ given by:

$$ V_{DD, min} > 2V_{TH} + 2V_{OW} $$

It gives that $V_{DD, min}$ is technology dependent. Fig. 4 shows $V_{DD}$ margin tapering with scaling technology till 22nm node where analog switch won’t be possible anymore.

### C. $V_{TH}$ variations

$(V_{DD}$–$V_{TH})$ reductions are critical for analog design. This occurs for the technology scaling and, for a given technology node, also for the analog design choices. In fact $V_{TH}$ depends on several effects. Among them the most important ones are:

- technology variations: process, supply voltage ($V_{DD}$ does not depend strongly on $V_{DD}$) and temperature (PVT) variation;
- analog design choices: device mismatch, short and narrow channel effects;
- layout choice: STI effects;

#### 1) PVT Variation

For a 65nm technology, $V_{TH}$ variations due to PVT variations can be very large.

#### Table II – $V_{TH}$ variation at corner simulations

<table>
<thead>
<tr>
<th>Nominal</th>
<th>Fast</th>
<th>Slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{TH}$ [mV]</td>
<td>344</td>
<td>341</td>
</tr>
<tr>
<td>$g_{ds}$ [μA/V]</td>
<td>548</td>
<td>486</td>
</tr>
<tr>
<td>$g_{ps}$ [μA/V]</td>
<td>15.9</td>
<td>14.3</td>
</tr>
<tr>
<td>$g_{ds}/g_{ps}$</td>
<td>17.6</td>
<td>17.6</td>
</tr>
</tbody>
</table>

For instance Table II gives the operating parameters for a NMOS device (W=650nm, L=65nm, $V_{DD}=730$mV, $V_{DD}=1.2$). The nominal value of $V_{TH}$ (nominal case & 27°C) is 547mV. This value would change with PVT from 425mV to 646mV, i.e. $\pm$110mV, a large amount. Considering, for instance, the
design of the cascode current mirror of Fig. 6, the minimum supply voltage required for this block operation would be larger, for all the worst-cases than the maximum \( V_{DD} \) allowed by the 65nm technology (1.2V).

\[
Eq. 6 \quad V_{DDmin_{Cascode}} > V_{sat} + 2 \cdot V_{GS} > 2 \cdot V_{TH} > 1.2V
\]

This simple example shows how one of the most popular building blocks has to be reconsidered in scaled technologies.

2) Analog design choices

\( V_{TH} \) value is affected also by statistical variation around its actual value. Mismatch observations based on transistor pairs can be described as well with a normal distribution with mean \( \mu \) and standard deviation \( \sigma \).

\[
Eq. 7 \quad \sigma_{\Delta V_T} = \frac{\gamma}{W \cdot L}
\]

where \( \gamma \) is a technology conversion constant (in mV·\( \mu \)m). The usual rule-of-thumb for \( \gamma \) vs. technology node is

\[
Eq. 8 \quad \gamma \approx \frac{1}{T_{ox}}
\]

i.e., “1mV·\( \mu \)m x nm \( T_{ox} \)”, where \( T_{ox} \) is the MOS oxide thickness. This means that for the same device area (W·L) scaled technology features a better matching. Thus, all these circuits whose power consumption is limited by the device matching (for instance flash ADC, or multipath/multichannel analog systems) can exploit the improved scaled technologies where the analog designer achieves the same matching performance with lower device size.

The \( V_{TH} \) value is also affected by the device size, due to the edge phenomena (for short and narrow channel cases) that are typically negligible in larger device size ([7]).

Narrow-channel effect becomes significant when the channel width is of the same order of magnitude as the thickness of the depletion region under the gate oxide. For MOSFET’s with non-recessed oxide-isolation structures, a decrease of the channel width (W) leads to a \( V_{TH} \) increase. In fact for W large, the additional inversion layers charge at the edge of the channel (\( Q_{CHL} \)) is negligible, while for narrow W, \( Q_{CHL} \) becomes important and results in increasing \( V_{TH} \) (see Fig. 6).

When short channel effects (SCE) occur the depletion region under the gates includes all the charge from source to drain (Fig. 7). At source and drain, a part of the charge (\( Q_{CHL} \)) is due to the depletion region and then it has not to be generated by the gate voltage. This results in a \( V_{TH} \) reduction. This situation is increased by the drain voltage movement, which can further reduce \( V_{TH} \) (this is the Drain-Induced Barrier Lowering effect – DIBL). This \( V_{TH} \) reduction could reach very low \( V_{TH} \) values. To avoid this situation, some additional technological steps (typically a modified doping profile at the channel edges, like HALO) are introduced to maintain a certain \( V_{TH} \) value. In this situation the channel length reduction results in a larger \( V_{TH} \) value.

3) Layout design

The device size shrinking in scaled technologies allows a strong reduction of the overall die size. In this situation other dimensions limit the die size reduction. One of the most critical limitations appeared to be the LOCOS size, which is the technology step used to separate two different active areas. The cross section of the LOCOS is shown in Fig. 8, where the "bird's beak" is an evident limitation of its size reduction. For this reason in order to reduce the separation space between two active areas, a different technology step has been adopted. This is the shallow trench isolation (STI), whose cross section is shown in Fig. 9 ([8], [9]). This process step, which consists of an oxide deposition into a trench, achieves a completely abrupt transition between the active area and the isolation. In a simplified description, this abrupt transition applies a mechanical stress to the active area edge that increases \( V_{TH} \) (in some simulation tools the STI effects is taken into account as a mobility variation).

\[
\text{Fig. 10 shows the STI effects for different layout design. Case (a) refers to the single device layout, where the mechanical pressure is applied to both device edges. This means that } V_{TH0} \text{ is the maximum value for the threshold voltage. In case (b), in both devices an edge is immune from STI pressure and then } V_{TH1} \text{ is lower than } V_{TH0}. \text{Finally in case (c), the external devices feature a threshold voltage given by } V_{TH2}, \text{ while the internal devices appear immune from STI and the threshold voltage } V_{TH3} \text{ is lower than } V_{TH2} \text{ and } V_{TH0}. \text{Notice that the STI effects is often dominant with respect to the narrow channel effect previously described and then for narrower gate size } V_{TH} \text{ tends to decrease.}
\]
**D. DC-Gain Reduction**

Analog signal processing is often based on circuits embedding opamp. An opamp key parameter is the dc-gain, which depends on the MOS device intrinsic gain (i.e. the $g_m/g_{ds}$). Technology scaling introduces a $g_m$ increase. However this is worsened by a stronger $g_{ds}$ increase, which results in a lower intrinsic gain (see Table I). The $g_{ds}$ increase can be seen in Fig. 11 that shows the output characteristics of MOS devices of different $L$ in a 65nm technology. The slope of these curves corresponds to the $1/g_{ds}$. This strong reduction of the intrinsic gain forces the development of improved opamp structures to achieve a sufficiently large dc-gain.

**E. Velocity saturation**

With scaling technology, the electric field across the channel increases and the carriers in the channel have an increased velocity. However, at high fields there is no longer a linear relation between the electric field and the velocity as the velocity gradually saturates reaching the saturation velocity ($v_{sat}$), which increases the transit time of carriers through the channel. At low electric field ($\varepsilon$), the velocity ($v$) increases proportionally to $\varepsilon$:

$$\mu_0 = \frac{v}{\varepsilon} \quad \text{Eq. 9}$$

For high electric field (i.e. small $L$) the velocity saturates to $v_{sat} (=10^5 \text{ m/s})$. The main consequence is that the current depends linearly with $(V_{GS} - V_{TH})$ and, then, transconductance saturate to $g_{msat}$:

$$I_D = W \cdot C_{ox} \cdot (V_{GS} - V_{TH}) \cdot v_{sat} \quad \text{Eq. 10}$$

$$g_{msat} = \frac{\delta I_D}{\delta (V_{GS} - V_{TH})} = W \cdot C_{ox} \cdot v_{sat} \quad \text{Eq. 11}$$
III. SCALTECH ANALOG DESIGN

A. ScalTech at transistor level

The reduced “free” space (VDD–VTH) allowed in scaled technology forces to consider different MOS operation conditions where the VTH “cost” has not to be fully “paid”. This is the case of operating MOS devices in sub-threshold region (VGS< VTH). In this condition MOS device presents the advantage of the minimum overdrive, small gate capacitance, large g_m/Id and large voltage gain (the gain is typically 25%-30% higher than the gain in saturation region). On the counterpart it suffers of larger drain current mismatch (input offset), large output noise current for a given Id, low speed. In fact the mismatch AVT parameter for the device in sub-threshold is typically three times higher than the value in saturation region [5]. This means that when the offset is critical, sub-threshold devices need some offset compensations scheme, while when offset can be tolerated they can fully exploited (like in band-pass sigma-delta modulators). Finally, nonetheless the sub-threshold devices exhibit a lower speed, this is compensated by the higher speed of the scaled technology and then they can be used in typical analog baseband applications.

B. ScalTech at circuit level

The use of scaled technology in analog design needs some new developments. This has to be introduced for any functional block. In the following the case of the basic analog switch, of the opamp, and of analog filters are discussed.

1) Analog Switch

A critical problem in designing analog sampled-data systems (like SC circuits, ADC, etc…) operating at low-voltage supply is the implementation of a MOS switch. Using a NMOS switch as a sampling switch in the T/H circuit has main issue of input-dependent finite ON-resistance given by:

\[ R_{ON} \propto \frac{1}{(W/L)(V_{GS}-V_{TH})} \]

Since VGS=(VDD−Vin), for W/L given, RON is signal dependent and results to be more resistive (performing lower bandwidth) at low supply voltage. This problem is more critical when VDD decreases as in scaled technologies. A popular solution is the use of a “bootstrapped” switch, whose functional and circuit scheme is shown in Fig. 13. Fig. 13-(b) shows that during the on-state the gate-to-channel voltage is kept constant, guaranteeing constant switch conductance. This is done by connecting a capacitance (precharged at VDD during the off-state) between the gate and source terminals of the main switch ([10]). As a results the switch during the on-state operates with a constant VGS, i.e. with a constant on-resistance, as shown in Fig. 14. Several circuit implementation of the conceptual scheme of Fig. 13-(b) are present in literature. One of most popular of them is shown in Fig. 13-(c) whose complexity indicates the increased cost of this solution (in terms of area, power consumption, additional load for the previous stage, etc….).

Fig. 13 - Bootstrapped Switch: (a-b) conceptual scheme, (c) circuit implementation.

Fig. 14 - Gate voltage VDD boosted.

2) Operation Amplifier

The design of an opamp in scaled technologies has to face several problems. Among them the most critical ones regards the bias point and the frequency response.

Regarding the bias point, the differential pair of Fig. 15 has to be considered, since it is the opamp input stage.
At low voltage it is mandatory to maximize the dynamic range, so a rail-to-rail output signal has to be processed with large linearity. To maximize the voltage swing the input and output common mode voltage of the cell has then to be fixed at $V_{DD}/2$.

**Eq. 13**

$$V_{i\_DC} = V_{o\_DC} = \frac{V_{DD\_min}}{2}$$

The opamp input node operating point requirements are:

**Eq. 14**

$$V_{i\_DC} = \frac{V_{DD\_min}}{2} - V_{GS} - V_{DS\_Sat} = V_{TH} - 2 \cdot V_{ov}$$

As a consequence, $V_{DD\_min}$ is given by

**Eq. 15**

$$V_{DD\_min} = 2 \cdot V_{TH} + 4 \cdot V_{ov}$$

This value can be quite large and in some cases disable the use of standard opamp topologies.

Regarding the frequency response, the dc-gain of a CMOS opamp is lowering with technology scaling, due to the reduced intrinsic gain. In addition, due to the lower supply voltage, high-gain stacked-device structures like cascode cannot be used. Thus opamps in scaled technologies uses multistage structure, where each stage introduces a pole in the overall frequency response. This means that an opamp typically presents several gain stages and then several poles. Then the frequency response compensation becomes fundamental. Several compensation schemes can be exploited which are based on capacitive feedback and/or transconductance feedforward ([11], [12]). These multistage opamp compensation topologies have to be compared in terms of ac-performances (gain, bandwidth, phase margin), load driving capability, power consumption, complexity and occupied area (since compensation capacitor is not scaling with technology).

An example of the combination of this technique is given by the three-stage opamp shown in Fig. 15 ([13]). The compensation scheme uses a Single-Miller capacitor Feed-Forward Compensation (SMFFC). It uses a transconductance feed-forward path to provide a left-half-plane (LHP) zero to compensate the second pole (first non-dominant pole).

![Fig. 16 - Structure of the three-stage SMFFC amplifier](image)

The compensation scheme is also shown in Fig. 17. In this scheme, together with the differential mode architecture the Common-mode feedback Circuit (CMFB) is shown. In fact, the feedforward paths, used for the differential mode compensation, are not effective for the CMFB compensation. A critical point in low-voltage multistage opamp is then also the frequency compensation of the CMFB loop. In the scheme of Fig. 17, a feedforward path in the feedback loop is introduces by the “D” stage which is effective only for common-mode signals. Table III summarizes the achieved performance with this opamp.

![Fig. 17 - Structure of three-stage SMFFC amplifier with the CM-control](image)

**Table III – MULTISTAGE OPAMP PERFORMANCE SUMMARY**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology CMOS</td>
<td>65nm</td>
</tr>
<tr>
<td>Differential Gain/UGB</td>
<td>84dB / 200MHz</td>
</tr>
<tr>
<td>Common Mode Gain/UGB</td>
<td>85dB / 136MHz</td>
</tr>
<tr>
<td>PSRR@1MHz</td>
<td>60dB</td>
</tr>
<tr>
<td>CMRR@1MHz</td>
<td>38dB</td>
</tr>
<tr>
<td>HD3@5MHz</td>
<td>-92dB</td>
</tr>
<tr>
<td>Output Noise@1MHz</td>
<td>27nV/√Hz</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>10mW</td>
</tr>
</tbody>
</table>

**C. Analog Filters**

Continuous-time analog filters are typically implemented using Gm-C, Active-RC or Active-Gm-RC topologies. The Active-RC and the Active-Gm-RC architectures exhibit a feedback structure and then they could presents a frequency response limitation (limited by the opamp GBW). However they can perform large linear range [16]. On the other hand, open-loop filters (like Gm-C) appear attractive in terms of noise and power consumption minimization, but large overdrive voltage is needed in order to perform large linear range [17].

At low supply voltage, while Active-RC and ActiveGm-RC can perform rail-to-rail signal processing capability, this is not the case of Gm-C filters, which results extremely inefficient in scaled technologies. As a consequence, closed-loop circuits (like Active-RC and Active- Gm-RC) have then to be considered. Among them, thanks to the single opamp topology, the multi-path Active-RC cell of Fig. 18 allows reducing the power consumption if compared to the typical...
two-opamp biquadratic cell [18]. However, the frequency response of this cell is affected by the opamp GBW which could be reduced when high-gain multi-stage opamp structures (with compensation schemes reducing the GBW) are used. In a robust design the opamp GBW has to be 50-to-100 higher than the filter pole frequency. This problem can be solved by the correspondent Active-Gm-RC structure of Fig. 19, where the opamp frequency response is taken into account in the overall filter frequency response. In this way the opamp GBW can be only 2-to-3 times higher than the filter pole frequency. This is much less demanding than the multipath structure.

Another key problem of both Active-RC and Active-Gm-RC (and any virtual ground based structure) is the bias voltages to be applied at the filter and opamp input and output nodes. The typical approach is to bias input and output nodes at the same voltage level. This however occurs in the bias problem as well as the multipath structure.

\[ V_{oa\_DC} = \frac{V_{DD\_min}}{2} - I_1 \cdot \frac{R_1 \cdot R_2}{R_1 + R_2} \]

Using this structure has been possible to design a 0.55V analog filter in a 65nm technology, performing rail-to-rail input and output swing.

Table IV – 4th-ORDER 65NM FILTER PERFORMANCE SUMMARY

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_{DD_min} ) [V]</td>
<td>0.55</td>
</tr>
<tr>
<td>CMOS [( \mu m )]</td>
<td>0.13</td>
</tr>
<tr>
<td>( V_{TH} ) [V]</td>
<td>0.3</td>
</tr>
<tr>
<td>Current Cons. [mA]</td>
<td>5.8</td>
</tr>
<tr>
<td>Power Cons. [mW]</td>
<td>3.5</td>
</tr>
<tr>
<td>Filter Order</td>
<td>40</td>
</tr>
<tr>
<td>( G ) [dB]</td>
<td>0</td>
</tr>
<tr>
<td>( f_{\text{lim}} ) [MHz]</td>
<td>11.3</td>
</tr>
<tr>
<td>In-band IIP3 [dBm]</td>
<td>10</td>
</tr>
<tr>
<td>Out-Band IIP3 [dBm]</td>
<td>13</td>
</tr>
<tr>
<td>( 1\text{dBcP} ) [dBm]</td>
<td>0.5</td>
</tr>
<tr>
<td>Noise [in(V/m)]</td>
<td>110</td>
</tr>
<tr>
<td>DR [dB] - THD@40dBc</td>
<td>60</td>
</tr>
<tr>
<td>Area [mm(^2)]</td>
<td>0.43</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

In this paper an overview of the challenges imposed by the use of scaled technologies in the analog circuit design is presented. In particular, intrinsic gain decreasing, \( V_{DD}\_V_{TH} \) reduction and lower supply voltage pushed analog designers to develop new circuit solutions for the analog functional blocks. The case of analog switch, opamp and Active-RC filters is here studied to demonstrate that it is possible to develop new circuit solutions in order to guarantee the same analog performance also in scaled technologies.

REFERENCES