Design Techniques for Ultra-Low Voltage Analog Circuits Using CMOS Characteristic Curves: a practical tutorial

Alessandro G. Girardi, Lucas Compassi-Severo and Paulo César C. de Aguirre

Federal University of Pampa - UNIPAMPA
Av. Tiaraju, 810 - CEP 97546-550 - Alegrete, RS, Brazil
E-mail: alessandro.girardi@unipampa.edu.br

Abstract— The use of ultra-low-voltage (ULV) analog circuits for IoT applications, in which reduced power consumption is a mandatory specification, is becoming more and more an important design approach. Also, in many IoT applications, power is supplied with energy harvested from environmental sources. It is more efficient for the circuit to operate at a voltage level close to the provided by the energy harvester (between 0.3 and 0.6 V). To deal with this when using low-cost technology process nodes - 180-nm, for example, with $|V_T| \approx 0.5$ V - it is necessary to apply specific design techniques that take advantage of reverse short channel effect, forward bulk biasing (FBB) or bulk-driven circuits. The use of low-$V_T$ transistors is also a good alternative when they are available in the target process node. This paper presents a comprehensive scenario about modern CMOS ULV design techniques from the designer’s point of view, including design trade-offs and comments about design decisions. Four step-by-step design examples of ULV circuits are presented: a cross-coupled negative transconductor, a CMOS inverter as an analog amplifier, a pseudo-differential inverter-based amplifier, and a bulk-driven differential amplifier with active load. All designs require the biasing of transistors in moderate and weak inversion regions. The goal is to demonstrate that it is possible to design ULV analog circuits using standard-$V_T$ transistors with a supply voltage much lower than the nominal $V_{DD}$.

Index Terms— ULP; ULV; fully-differential analog circuits; inverter-based amplifier; negative transducer.

I. INTRODUCTION

The fast widespread of Internet of Things (IoT) technology is changing the world. IoT applications and services cover almost any sector. Embedded devices can replace human in performing tasks and continuously monitor environment and equipment conditions. Examples are home automation, precision agriculture, smart cities, industrial automation, healthcare systems, wearable devices, traffic management, environmental monitoring, and more [1, 2]. There are estimated 12.3 billion active IoT endpoints in 2021. By 2025, there will likely be more than 27 billion IoT connections - a growing rate of 22% - according to [3]. Most of these endpoints are composed of autonomous sensor nodes supplied with battery or non-conventional energy sources.

Powering billions of IoT devices is a major challenge, which becomes crucial with the rapid development of new technologies [4]. The integrated circuits that implement digital and analog functions in an IoT device must spend minimum energy in order to optimize battery life.

CMOS technology is still dominant in the market due to low cost, high performance and reliability. Design techniques must be adopted in order to provide circuits operating in ultra-low power mode. An important strategy is to reduce the supply voltage to levels below 1 V.

The reduction of supply voltage is a natural path in the evolution of CMOS technologies. As transistors are getting smaller, reliability issues of gate dielectrics and power dissipation limits at the maximum switching frequency make the maximum allowed node voltage drop to safe values. Voltage reductions guarantee the reliability of devices as the lower electrical fields inside oxide layers of a MOSFET produce less risk to the thinner oxides, which result from device scaling [5]. However, this is not the only reason to reduce supply voltage. There is an increasing interest in energy harvesting devices, mainly for IoT applications. These devices are capable of provide energy from the environment, but generating very low supply voltages [6].

In many IoT applications, devices need to be powered in a self-sufficient and sustainable fashion. Most of the devices will be battery operated due to cost, convenience, size, and the fact that they are implemented in hard-to-reach areas. The duration of the battery must be maximized in order to avoid battery replacement. In some cases the battery is completely removed from the system and power is supplied for the circuit with energy harvested from environmental sources (photovoltaic, vibration or piezoelectric) [1, 2, 7, 8]. In this case, circuits must operate with power consumption levels compatible to instantaneous energy harvesting, which is in the order of some $\mu W$. Also, the generated voltage level is low. A single photovoltaic solar cell can generate around 0.5 V, for example, in its maximum power point. Using lower supply voltage can improve system efficiency in energy harvesters. If DC-DC converters should be used to boost the output voltage of energy harvesters, the conversion efficiency is usually limited to 40%-70% [9]. So, it is more efficient the circuit to operate at a voltage level close to the provided by the energy harvester. It requires the development of ultra-low power (ULP) and ultra-low voltage (ULV) circuits. Transistors are biased to operate in moderate and weak inversion region and supply voltage is reduced to values typically between 0.3 and 0.6 V, which are usually close to the threshold voltage of regular MOSFETs in cheap process nodes [6].

Since ULV analog circuits should have the same (or better) performance than circuits designed for larger power supplies, it brings some additional design challenges. Due to the limited voltage range, designers can not use conventional cascode structures, or other conventional design methodologies. As transistors must operate in weak or moderate inv-
sion level, design equations are different from the traditional strong inversion. This implies additional experience for analog designers that usually limit the design space to strong inversion region. Moreover, limited supply voltage causes degradation in circuit speed, higher input-referred noise, and larger offset. Specific design techniques must be adopted to mitigate these effects.

ULV circuits are applied in several fields. An example is in healthcare where ULP bio-signal sensing devices have played an important role in wearable biomedical applications. Pursuing long lifetime and low cost, modern bi-signal sensors are demanded to work with ultra-low power consumption and usually be highly integrated with energy-harvesting and management devices [9].

The use of ULV for designing different CMOS analog blocks, such as A/D converters [10, 11], filters [9] and amplifiers [12, 13, 14, 15], is described in the literature. The design of \( \Sigma \Delta \) modulators in ULV has been reported with supply voltages ranging from 0.2 to 0.5 V. It has been demonstrated that it is possible to achieve a bandwidth of up to 50 kHz with a SNDR from 60 to 80 dB and dissipating some tens of \( \mu \)W [11, 16].

A number of ULV OTAs also have been proposed in the literature in recent years. It is worth noting that more and more of these circuits can operate even from deep sub-0.5 V supply [14, 13, 17, 18]. Pseudo-differential approach can be used to provide strong rejection of common-mode interferences, larger output range, and improved linearity [6]. The drawback is that it requires the implementation of an additional common-mode stabilization loop.

The introduction of digital-like blocks for emulating analog functions is another approach for designing ULV OTAs, but it requires dynamic calibration techniques to improve yield due to process variations and mismatch [19].

The main technique to deal with low supply voltages is to scale transistor threshold voltage (\( V_T \)). It can be done either by using low-\( V_T \) devices - when they are available in the target process technology - or by design strategies - when standard-\( V_T \) transistors are used. In the last case, \( V_T \) can be adjusted with design techniques that take advantage of reverse short channel effect, forward bulk biasing (FBB) or bulk-driven circuits.

The goal of this paper is to present a comprehensive scenery about modern CMOS ULV design techniques from the designer point of view, including design trade-offs and comments about design decisions.

This paper is organized as follows: Section II describes the main ULV design techniques, including the exploration of bulk-driven circuits, short-channel and reverse short-channel effects and forward bulk biasing; Section III presents the step-by-step design of four different ULV circuits: a cross-coupled negative transconductor, a CMOS inverter as an analog amplifier, a pseudo-differential inverter-based amplifier, and a bulk-driven differential amplifier with active load; and Section IV finalizes with the concluding remarks.

II. LOW VOLTAGE DESIGN TECHNIQUES

The electrical characteristics of CMOS transistors are dependent on the channel sizes, biasing voltages, and fabrication process parameters. The designer should explore the transistor channel length (\( L \)), width (\( W \)), and biasing voltages (\( V_{GS} \), \( V_{DS} \) and \( V_{BS} \)) to find the required behavior in such application. Fig. 1 shows the source-referred NMOS transistor with the indication of all biasing voltages (\( V_{GS}, V_{DS} \) and \( V_{BS} \)) and transistor aspect ratio (\( W/L \)) that should be explored to obtain the required drain current (\( I_D \)). The \( I_D \) current is one of the main design specifications of a CMOS transistor since it is directly related to the small-signal characteristics, frequency bandwidth, and speed.

The following subsections present some of the main aspects that should be considered during the design of low-voltage analog circuits.

A. Channel Inversion Level

The transistor \( W/L \) ratio required to obtain a target drain current level is related to the channel inversion level. If the \( W/L \) ratio is fixed, the lower the inversion level, the lower is the drain current level. For small values of \( V_{GS} \) the drain current is dominated by diffusion charge carriers and the transistor is operating in weak inversion level (WI). For higher values of \( V_{GS} \) the transistor is in strong inversion level (SI) and the conduction is performed by drift charge carriers. The transition between weak and strong inversion levels is called moderate inversion (MI), in which both diffusion and drift charge carriers are present on the drain current.

The inversion level can be evaluated by means of the inversion coefficient (\( I_C \)), which is a parameter dependent on the process parameters and voltage biasing [20]. For values of \( I_C \) lower than 0.1 the transistor is in WI; \( I_C \) between 0.1 and 10 is considered MI; and \( I_C \) higher than 10 can be characterized as SI. A practical way to evaluate the channel inversion level is based on the difference between the \( V_{GS} \) voltage and the threshold voltage (\( V_T \)), called over-drive voltage (\( V_{OV} \)). Fig. 2 (adapted from [21]) shows the relation of \( I_C \) and \( V_{OV} \) to the inversion levels. Values of \( V_{OV} \) from -64 mV to +201 mV are equivalent to values of \( I_C \) from 0.1 to 10.

Additionally, the transistor efficiency given by the gate-transconductance to drain current ratio (\( g_m/I_D \)) can be used

\[
\text{Channel Inversion Level}
\]

<table>
<thead>
<tr>
<th>( I_C )</th>
<th>Weak</th>
<th>Moderated</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-64</td>
<td>201</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2: Channel inversion level classification according to the inversion coefficient (\( I_C \)) and the over-drive voltage (\( V_{OV} = V_{GS} - V_T \)) [21].

![Fig. 1: Source-referred NMOS transistor biasing and channel dimensions.](image)
to quantify the inversion level. Fig. 3 shows the electrical simulation of $g_m/I_D$ as a function of $V_{OV}$ for CMOS transistors in a 180 nm process. The MI level can be defined for values of $g_m/I_D$ between 8 and 22 $V^{-1}$.

The operation of analog circuits in ultra-low voltage (ULV) is limited to weak and moderated inversion levels since it is difficult to exceed the $V_{OV}$ of 201 mW. Thus the drain current density $I_D/W$ is an important design parameter since it is related to the transistor area required to reach a target value of $I_D$. Fig. 4 shows the simulated values of $I_D/W$ as a function of $g_m/I_D$. Between the limits of MI there is a difference of three orders of magnitude in $I_D/W$ (from $\mu A/\mu m$ to $nA/\mu m$). The lower the inversion coefficient, the higher the transistor area needed to obtain a certain value of $I_D$. Moreover, the higher the area the higher the parasitic capacitances. Because of that, the operation in WI is suitable only to low frequency high efficient circuits.

Small-signal parameters are also affected by the inversion level and transistor sizes. The gate and bulk conductances ($g_m$ and $g_{nb}$) and output conductance ($g_{ds}$) are proportional to the current level, but their ratios are changed according to the inversion level. The intrinsic voltage gain of a transistor is proportional to $g_m/g_{ds}$ and $g_{nb}/g_{ds}$ for gate and bulk input amplifiers, respectively. Fig. 6 shows the simulation results of the $g_m/g_{ds}$ for different values of $g_m/I_D$. As expected, the intrinsic gain is proportional to $g_m/I_D$, presenting higher values in WI.

Equation 1 could be used as a cost function in a sizing process, and its behavior for a NMOS transistor is depicted in Fig. 5. Its maximum value occurs for overdrive voltages around 100 mV, i.e., with the transistor operating in moderate inversion. It means that for such optimization the $V_{DD}$ should be equal to $V_{TN} + |V_{TP}| + 2 \cdot 100$ mV.

$$F = \left( \frac{g_m}{I_D} \right) \cdot \left( \frac{g_m}{g_{ds}} \right) \cdot \left( \frac{I_D}{W} \right)$$  

B. Bulk-driven circuits

A traditional design technique for ULV OTAs is the adoption of bulk-driven differential input pairs. This technique exploration backs from the late 80’s but wide exploration occurs since 2000, when a bulk-driven Miller-compensated opamp working at 1 V has been presented [22]. Bulk-driven OTAs are able to operate at very low-voltages such as 0.5 V [23, 12] down to 0.3 V [14] and 0.25 V [13]. The main advantage of this approach is to set the rail’s voltage at the gate of the input transistors and then defining their operating point region (weak, moderate or strong inversion) and the minimum $V_{DS}$ saturation voltage, thus providing rail-to-rail
input capability. This is a key characteristic for low-voltage analog applications such as buffers and gm-C filters. Also, bulk-driven input OTAs are known for providing an almost constant transconductance for a wide common-mode input voltage range.

However, the \( g_{mb}/g_m \) ratio is quite low - roughly 0.3 in a typical 180-nm process. Thus, bulk-driven OTAs present some disadvantages such as lower voltage gain, GBW and slew-rate. The input referred noise of bulk-input OTAs is higher than a gate input one, as demonstrated in [14].

C. Short-Channel and Reverse Short Channel Effects

The reduction of channel length in CMOS transistors impacts on \( V_T \). Short channel effects generate a decrease of \( V_T \) with the decrease of \( L \) [24]. It occurs since the depletion regions between source/channel and channel/drain tend to be symmetric in charge. Thus, the electric charges belonging to the drain side depletion region fill part of the area below the gate, thus reducing the amount of immobile charge that is imaged by the gate. As a result, a lower \( V_{GS} \) should be used to induce the required inversion region below the gate. This phenomenon is modeled as an increase of the threshold voltage with the increase of \( L \), since the ratio between the drain-side depletion region area below the gate and the total gate area decreases.

Another short channel effect is the drain-induced barrier lowering (DIBL). In essence, the DIBL effect generates a reduction of \( V_T \) with the increase of \( V_{DS} \). The increase of drain voltage expands the drain/substrate depletion region under the gate. Thus, the same gate voltage value is able to attract more carriers into the channel, in a similar manner that a \( V_T \) reduction caused by short \( L \).

To cope with DIBL, nanometer CMOS processes adopt the halo implant [24]. With this technique, the substrate near the source and drain regions is highly doped in order to reduce the depletion region under the gate and improve the device characteristics. For small channel lengths, the halo doping invades the channel region, and, in some cases, the halo doping near source and drain regions can overlap. This phenomenon increases the average substrate doping under the gate with the channel length reduction. It is modeled as a \( V_T \) increase with the decrease of \( L \) [25], which is called reverse short channel effect. This effect is present in multi-\( V_T \) tech nodes. Fig. 7 depicts the \( V_T \) behavior for different channel lengths in standard and low-\( V_T \) transistors in a 180-nm CMOS process.

D. Forward Bulk Biasing - FBB

Another possible solution to control \( V_T \) is to use a forward bulk biasing (FBB) approach. This strategy aims to change the transistors threshold voltage by controlling the source-bulk voltage \( (V_{SB}) \), as can be observed in eq. 2 [26].

\[
V_T = V_{T0} + \gamma (\sqrt{V_{SB} + |2\phi_F|} - \sqrt{|2\phi_F|})
\]  

(2)

Here, \( V_{T0} \) is the threshold voltage for \( V_{SB} = 0 \) \( V \), \( 2\phi_F \) is the surface potential and \( \gamma \) is the parameter that defines the MOS transistor body effect. The change of \( V_T \) could be performed by forward bulk-bias \( (V_{SB} < 0 \ V) \) to decrease \( V_T \) or by reverse bulk-bias \( (V_{SB} > 0 \ V) \) to increase \( V_T \). In ULV designs the FBB technique has been adopted both to reduce the \( V_T \) and to control the common-mode output voltage of pseudo-differential amplifiers [15]. The FBB technique requires access to the transistors bulk terminal. It can be easily employed in PMOS transistors in most fabrication processes. However, it can be used in NMOS transistors only if buried deep nwell or triple-well are available. Fig. 8 shows the \( V_T \) versus \( V_{SB} \) behavior for standard and low-\( V_T \) PMOS and deep nwell NMOS transistors in a 180-nm CMOS process.

The adoption of FBB impacts on \( g_m \) and \( g_{mb} \) of CMOS transistors. Figure 9 shows the \( g_{mb}/g_m \) ratio for a NMOS transistor with different \( V_{BS} \) voltages. There is an increase of \( g_{mb}/g_m \) due to the increase of \( V_{BS} \). Also, considering the same \( V_{BS} \), one can verify an increase of \( g_{mb}/g_m \) as the overdrive voltage decreases. This behavior is explained by eq. 3 and 4, which model \( g_{mb} \) in strong and weak inversion regions, respectively.

\[
g_{mb(SI)} = \mu C_{ox} \frac{W}{L} (V_{GS} - V_T) \frac{\gamma}{2\sqrt{V_{SB} + |2\phi_F|}}
\]  

(3)

!![Fig. 7: \( |V_T| \) versus \( L \) for standard and low-\( V_T \) transistors \((V_{BS}=0)\) in 180-nm process node. Reverse short channel effect causes a slightly increase in \( |V_T| \) with the decrease of \( L \).]

!![Fig. 8: \( |V_T| \) versus source-bulk voltage \((eq. \ 2)\) for standard and low-\( V_T \) transistors in 180-nm process node \((L = 1.0 \ \mu m)\).]
\[ g_{mb(WT)} \approx \left( \frac{n-1}{n} \right) \frac{I_D}{V_T} \]  

(4)

Here, \( \mu \) is the carrier mobility, \( C_{ox} \) is the gate oxide capacitance per unit area and \( n \) is the slope factor (derivative of the gate voltage with respect to the pinch-off voltage [27]).

The disadvantage of FBB is the existence of bulk leakage current and the risk of latch-up due to the bulk parasitic bipolar transistors. As the \( V_{BS} \) of NMOS transistor is positive (and negative in PMOS), the bulk to source and bulk to drain junctions are forward biased. So, the bulk leakage current level follows the diode equation, being proportional to the exponential variation on \( V_{BS} \). However, \( V_{BS} \) lower than 0.6 V is safe for latch-up and generally, the leakage current is much lower than the drain current [23].

E. Low-\( V_T \) Transistors

CMOS process with multi-\( V_T \) devices are becoming popular in the last years, despite the cost for extra masks and fabrication process steps. Also, to provide multi-\( V_T \) transistors in 7-nm and beyond process is a challenging task [28]. In planar CMOS process, low-\( V_T \) transistors are built in a slightly p-doped silicon. Thus, the low concentration of holes below the gate (in comparison of the the high concentration of holes in a \( P^+ \) substrate of standard transistors) results in a lower threshold voltage. Low-\( V_T \) transistors enable the supply voltage reduction while keeping reasonable overdrive voltage for analog IC design.

As shown in Fig. 8, in a 180 nm CMOS process the \( |V_T| \) of the standard devices is around 0.5 V with \( V_{SB} = 0 \) V, while low-\( V_T \) NMOS and PMOS present \( |V_T| \) equal to 0.31 V and 0.23 V, respectively. The value of \( V_T \) can be further reduced employing the FBB technique. This approach is limited to low-\( V_T \) PMOS devices since the NMOS insulated bulk option is normally available only for standard-\( V_T \) transistors.

III. DESIGN EXAMPLES

This section presents design examples of four ULV analog circuits: a cross-coupled negative transductor, an inverter-based amplifier, a pseudo-differential inverter-based amplifier, and a bulk-input amplifier. These circuits are designed with standard-\( V_T \) and low-\( V_T \) transistors in a 180-nm CMOS process. Schematic-level simulations are performed with Spectre simulator using BSIM4 MOSFET model. First and foremost, it is important to have a fully understanding about the technology characteristic and main transistors parameters behavior. With the advance of CMOS technologies and transistors shrinking, analog CMOS designers should use first order transistor’s equations with caution, since different physical effects such as short channel and reverse short channel effects, which have huge impact on \( V_T \), are not adequately modeled. More accurate (and complex) MOSFET models, such as BSIM4, should be used for modeling these effects, with the aid of a SPICE simulator. The quality of process parameters provided by the foundry for the target design process is fundamental for a good estimation of circuit performance during the design stage. A good alternative is to use pre-computed lookup tables to explore the design trade-offs, as proposed by Jespers and Murmann in [29]. ULV circuits are expected to work with transistors in weak and moderate inversion, thus the \( g_m/I_D \) design methodology is a good approach to help analog designers in the sizing task [30]. The results obtained in this paper were generated with the aid of a tool called “Gmm/I_D Starter Kit”, introduced in [29]. It is an easy and straightforward tool for exploring all transistor inversion regions.

For sake of clarity, all design examples analyzed in this sections consider a 0.5-V power supply.

A. Negative Transconductor

The cross-coupled negative transconductor is a very important basic building circuit that can be used in several applications [31]. In ULV amplifiers and filters it is employed to improve the bandwidth and voltage gain. Fig. 10 (a) shows the schematic of the cross-coupled negative transconductor. Transistors M1a and M1b implement the negative transconductance while transistors M2a, M2b, and M2c work as current sources for biasing.

The small-signal model of the transconductor is presented in Fig. 10 (b). Based on this, the low-frequency equivalent negative transconductance is obtained as:

\[ g_{mneg} = -g_{m1} + g_{ds1} + g_{ds2} \]  

(5)

The required equivalent negative transconductance \( g_{mneg} \) can be found by using some iterative process in order to size NMOS and PMOS transistors. However, it can be performed in a few steps using the transistor curves presented in Section II. First of all, the transistor channel length and the inversion level should be selected. The inversion level is limited by the supply voltage \( (V_{DD}) \), since the input DC voltages should be at \( V_{DD}/2 \) to improve voltage swing. Due to this, the absolute value of \( V_{GS} \) of both PMOS and NMOS transistors is equal to \( V_{DD}/2 = 250 \) mV.

To find the exact value of \( g_{mneg} \), the \( g_{ds1} \) and \( g_{ds2} \) of eq. 5 should be rewritten as a functions of \( g_{m1} \). The key-strategy is to assume both transistors with the same drain current \( (I_{D_{M2}} = I_{D_{M1}} = I_{ref}) \) and to use the \( g_m/I_D \) and \( g_m/g_{ds} \) parameters, as shown in eqs. 6, 7 and 8.

\[ g_{ds1} = \left( \frac{1}{g_{ds1}} \right) g_{m1} \]  

\[ g_{ds2} = \left( \frac{1}{g_{ds2}} \right) g_{m1} \]  

(6)
In this design, all devices are low-$V_T$ transistors, $V_{DD} = 0.5$ V and $L = 1\ \mu$m. The value of $L$ was defined to reduce the reverse short channel effect in the low-$V_T$ transistors. Thus, $V_{GS1}$ and $|V_{GS2}|$ are equal to 0.25 V, $V_{TN} = 0.31$ V, $V_{TP} = 0.23$ V and both transistors are operating in moderated inversion. With the value of $L$, curves of $g_m/I_D$, $g_m/g_{ds}$, $I_D/W$ as a function of $V_{GS}$ are generated by electrical simulation and the values for $V_{GS} = 0.25$ V are extracted. By using eqs. 10 to 12 the circuit is sized. The obtained design parameters for a negative transconduance of -10 $\mu$S are presented in Table I. The simulated frequency response of the negative transconduance is shown in Fig. 11. It can be seen that the value of -10 $\mu$S is kept constant from DC to 25 MHz without considering any extra capacitive load.

![Diagram of cross-coupled negative transconductor](image)

**Table I:** Design parameters obtained for the 0.5 V cross-coupled negative transconductor of Fig. 10.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_1$</td>
<td>2.17 $\mu$m</td>
</tr>
<tr>
<td>$L_1$</td>
<td>1.00 $\mu$m</td>
</tr>
<tr>
<td>$W_2$</td>
<td>5.09 $\mu$m</td>
</tr>
<tr>
<td>$L_2$</td>
<td>1.00 $\mu$m</td>
</tr>
<tr>
<td>$I_{ref}$</td>
<td>580.5 nA</td>
</tr>
</tbody>
</table>

**B. Inverter-based amplifier**

Digital-like inverter-based amplifiers are known due to their high-transconductance efficiency and are good candidates for low-voltage applications since they require only two stacked transistors as verified in the Nauta transconductor [32]. Different design strategies and circuit topologies with CMOS inverters as the main amplifier have been published over the last decade to improve the power efficiency of low-voltage continuous-time (CT) filters [31] and both discrete-time (DT) [33] and CT sigma-delta ADCs [11, 16]. The drawback is that inverter-based amplifiers are highly affected by process, voltage and temperature (PVT) variations if no compensation strategy is adopted, thus reducing the DC voltage gain $A_{v0}$ and gain-bandwidth product (GBW).

This design example considers the inverter-based amplifier illustrated in Fig. 12 and its respective small-signal model. The amplifier low-frequency voltage gain is expressed by eq. 13.

$$A_{v0} = -\frac{g_{m+n} + g_{m+p}}{g_{ds+n} + g_{ds+p}}$$

The maximum DC voltage gain occurs for an input common-mode voltage equal to the CMOS inverter trip point. This value can be estimated using the drain current quadratic model and considering that both transistors are in saturation [32], as follows:

$$V_{trip} = \frac{V_{DD} - V_{TN} + V_{TP}}{1 + \sqrt{\frac{2}{\beta p}}} + V_{TN}$$

![Frequency response of the designed cross-coupled negative transconductor](image)
where $\beta_{n,p} = \mu C_{ox}(W_{n,p}/L_{n,p})$. It means that $V_{trip}$ is controlled by transistors sizing and also highly dependent on process variations.

Sizing a CMOS inverter for digital applications is usually done by setting the transistors minimal length and adjusting the NMOS and PMOS transistors width to provide symmetrical rise and fall times. For analog applications, the sizing is done focusing on setting $V_{trip}$ equal to the input common-mode voltage, which is usually $V_{DD}/2$. To accomplish this, the transistors sizing can be done as stated hereafter.

In this example, $V_{DD}$ is defined in 0.5 V and the design has as goal to achieve a GBW equal to 1 MHz for a 10-pF capacitive load using standard-$V_T$ transistors. Thus, it is possible to estimate the transistors’ overdrive voltages and their respective $g_{m}/I_D$ ratio. At this time, we are considering both NMOS and PMOS with $V_{BS} = 0$ V. The $|V_{GS}|$ of both NMOS and PMOS transistors in this case are equal to $V_{DD}/2 = 250$ mV, and thus for $|V_T| = 0.5$ V the overdrive voltage is $-250$ mV and the transistors operate in weak inversion. As shown in Fig. 3, it leads to a $g_{m}/I_D$ around 28.6 $V^{-1}$ and 27.7 $V^{-1}$ for NMOS and PMOS transistors, respectively.

One of the most important steps is to estimate the transistors channel length. As a rule of thumb, the minimum $L$ should be equal to at least five times the minimum length ($L = 5L_{min}$) to minimize the impact of process variations and minimize the increase of $V_T$ due to reverse short channel effects. Also, the trade-off regarding transistor output conductance and silicon area should be kept in mind. A higher $L$ impacts in a higher intrinsic gain ($g_{m}/g_{ds}$). However, more silicon area will be required to achieve the same transconductance for a given overdrive voltage. Analog designers are advised to plot the $g_{m}/g_{ds}$ for different channel lengths before choosing the $L$. As a designer’s choice, we will set $L = 1$ $\mu$m since in this technology the intrinsic gain is higher than 100 for negative overdrive voltages. As shown in Fig. 6, the intrinsic gain is higher than 150 for both NMOS and PMOS transistors.

The GBW of a CMOS inverter is defined according to:

$$GBW = \frac{g_{m_{tot}}}{2\pi C_L} = \frac{g_{m_n} + g_{m_p}}{2\pi C_L} = \frac{g_{m_n}}{2\pi C_L} \cdot I_D + \frac{g_{m_p}}{2\pi C_L} \cdot I_D$$

(15)

Since the GBW specification is 1 MHz and the capacitive load is 10 pF, the amplifier total transconductance $g_{m_{tot}}$ must be 6.2832 $\mu$S. Both PMOS and NMOS transistors present similar $g_{m}/I_D$ ratio, thus the bias current could be estimated by the $g_{m}/I_D$ to be equal to 1.11 $\mu$A. Then, the current density ($I_D/W$) helps defining the transistors width. From Fig. 4, which relates $I_D/W$ with $g_{m_n}/I_D$, it is possible to extract an $I_D/W$ equal to 2.05 nA/$\mu$m for a $g_{m_n}/I_D$ around 28.6 $V^{-1}$. It leads to an NMOS transistor width of 20 - 26.81 $\mu$m. The electrical simulation of the drain current for this NMOS transistor using BSIM4 model results in 897.11 nA (19.2 % difference with respect to previously estimate $I_D$). So, an interactive size tuning is required, leading to a final width of 20 - 33 $\mu$m. In this case, the size tuning is done by increasing the number of multiples in all transistors.

The next step is to define the PMOS width. The same approach can be performed to estimate this parameter. For fine tuning the input and output of the CMOS inverter should be shorted since the $V_{trip}$ voltage appears at this shorted node. The final PMOS width is 20 - 107 $\mu$m.

Fig. 13 (red) shows the frequency response of the designed amplifier. The achieved GBW is 718.48 kHz, i.e. 28.15% lower than the desired value of 1 MHz. Thus, the designer must keep in mind that estimating the transistor’s intrinsic capacitive load is mandatory for the first initial calculations. For weak inversion designs, the intrinsic capacitive load could exceed a few pF. It occurs due to the degradation of $I_D/W$ with the increase of $g_{m}/I_D$, as shown in Fig. 4 for $g_{m}/I_D$ higher than 25 $V^{-1}$. Increasing the multiplicity from 20 to 33 for both NMOS and PMOS provided a GBW equal to 1 MHz.

A 500-run Monte Carlo simulation was performed and provided a $V_{trip}$ average value of 250.08 mV with a standard deviation of 7.29 mV. The CMOS inverter AC response is highly dependent on this voltage value, and with such variation, the inverter does not work properly under PVT variations. There are some compensation strategies available on the literature to cope with such variation for single inverters, as the on-chip body bias [34] and the use of a dedicated low-dropout regulators to power the inverters [35]. Single CMOS inverters can be used to design switched capacitor
circuits using the correlated double sample technique [36] to reduce the inverter offset and a passive common-mode feedback circuit to guarantee a common-mode voltage gain lower than the differential gain [34] [35].

C. Pseudo-differential inverter-based amplifier

In this section, the design of a pseudo-differential amplifier powered by a 0.5-V power supply is explored aiming to achieve a GBW equal to 1 MHz for a 10-pF capacitive load. The aim is to achieve additional stability with respect to PVT variation by means of a pseudo-differential approach.

Pseudo-differential inverter-based amplifiers have been used in state-of-the-art continuous-time sigma-delta ADCs [11, 16]. Fig. 14 presents the schematics of this amplifier described in [11] with a full bulk-based common-mode feedback (CMFB) described in [15]. The CMOS inverters composed of $M_{1,2}$ and $M_{11,12}$ are the main inverters and are sized to set the desired transconductance. The CMOS inverters composed of $M_{3,4}$ and $M_{9,10}$ set the main inverters output voltage. Meanwhile, the cross-coupled inverters composed of $M_{5,6}$ and $M_{7,8}$ implement a negative resistance at the output that can be used to boost the DC gain in face of resistive loads. The amplifier low-frequency voltage gain is given by eq. 16.

$$A_{v0} = \frac{g_{m1} + g_{m2}}{\sum_{i=1}^{6} g_{ds_i} + (g_{m3} + g_{m4}) - (g_{m5} + g_{m6})}$$

(16)

The output common-mode voltage is kept constant at $V_{DD}/2$ by controlling the bulk voltage of PMOS and NMOS transistors. The VCMFB control signal is set to $V_{DD}/2$ at the typical corner. This signal controls the $V_{trip}$ of the CMOS inverters (see eq. 14) by applying the effect of forward and reverse bulk bias. The CMFB is performed by a single differential difference amplifier (DDA) to avoid loading the main amplifier.

In our design approach, we suggest to use equal sized, but scaled CMOS inverters through the use of different multiples of $W$. In this sizing we considered standard-$V_T$ PMOS and NMOS transistors with equal channel length. Initially, the NMOS transistor $M_2$ is sized to achieve the desired transconductance with $V_{GS} = V_{DS} = V_{BS} = V_{DD}/2$. It is suggested to use multiples of 10-20 in order to scale the auxiliary inverters. Then, the PMOS transistor $M_1$ is sized with the same number of multiples to provide the voltage $V_{trip}$ equal to $V_{DD}/2$ and ensures the maximum output swing. The auxiliary inverters composed of $M_{3,4}$ and $M_{9,10}$ are sized to drain 20% of the bias current of the main CMOS inverters. However, this biasing could be even reduced since these inverters are used to stabilize the DC output voltage of the main inverters. The CMOS inverters composed of $M_{3,4}$ and $M_{9,10}$ are sized to provide a negative output conductance to compensate for part of the total OTA output conductance. In this design with a pure capacitive load the sum of $(g_{m5} + g_{m6})$ is set to null the effect of $(g_{m3} + g_{m4})$. It is important to be aware that if a resistive load is connected at the OTA output, the negative conductance $-(g_{m5} + g_{m6})$ should be carefully designed accounting for possible resistive load variations.

The DDA is designed with standard-$V_T$ transistors biased with $V_{BS}$=0, thus operating in weak inversion. It is possible to achieve a DC gain of almost 40 dB with this topology since its load is a pure capacitive load (bulk terminal). The voltage gain of the DDA is given by eq. 17. Thus, transistors $M_{A-D}$ are optimized for high transconductance and transistors $M_{E-D}$ are optimized to present a low output conductance. The sizing is done using multiples to enable an easy tuning of the CMFB frequency response.

$$A_{v0(DDA)} = \frac{2 \cdot g_{m,A}}{g_{ds,F} + g_{ds,B} + g_{ds,C}}$$

(17)

Table II summarizes the obtained transistors sizes for this pseudo-differential inverter-based amplifier. Fig. 15 presents the OTA frequency response. The designed amplifier achieves a DC gain of 41.65 dB and provides a GBW of 1.051 MHz with a phase margin (PM) of 89.7° at typical corner. The gain for common-mode input signals is -33.26 dB at DC and -20 dB at 9.73 kHz.

The performance of the circuit under process and mismatch variations is evaluated with a 200-run MC simulation at temperatures of $27^\circ C$ and $85^\circ C$. Fig. 16 shows the behavior of the GBW and DC gain under such variations. The GBW of this OTA is highly dependent on the operating temperature, especially due to the lack of a current reference for

![Fig. 14: Pseudo-differential inverter-based amplifier. a) Main amplifier with bulk-based CMFB; b) CMFB amplifier.](image)
biasing (as such a constant-gm current source). The mean GBW is 1.07 MHz (σ = 0.22 MHz) and 3.16 MHz (σ = 0.53 MHz) for temperatures of 27°C and 85°C, respectively. On the other hand, the DC gain remains almost constant: the mean DC gain is 41.39 dB (σ = 2.04 dB) and 40.61 dB (σ = 1.44 dB) for temperatures of 27°C and 85°C, respectively.

Finally, it is essential to verify the CMFB loop stability. In this design, the CMFB loop frequency response has a unity gain frequency of 174 kHz with a PM of 45°. This OTA topology could be used as the input stage of multi-stage low-voltage CMOS amplifiers. This design example considers as design specification: VDD equal to 0.5 V, GBW = 100 kHz and a 10-pF capacitive load. The technology node is 180 nm and all transistors are standard-VT. This OTA has three stacked transistors (one of them in a diode-connected configuration), thus reducing the output swing when compared with the previously presented inverter-based OTA. The OTA low-frequency gain is given by eq. 18 while GBW is given by eq. 19.

\[
A_v0 = \frac{g_{m2}}{(g_{ds2} + g_{ds3})} \tag{18}
\]

\[
GBW = \frac{g_{m2}}{C_L} \tag{19}
\]

Some assumptions are made to start sizing the circuit to achieve the designed specification. The headroom voltages are specified as: \(V_{GS1} = V_{DS3} = 250 \text{ mV, } V_{DS1} = 100 \text{ mV and } V_{DS5} = 150 \text{ mV.}\) Then, the bulk transconductance should be defined to achieve the required GBW. In this, we considered for sizing an extra inherent circuit capacitive load of 3 pF (30% of \(C_L\)). The required \(g_{m2}\) to achieve a 100-kHz GBW is 8.16 \(\mu\)S. The bulk of \(M_2\) is at 250 mV and its source terminal estimated voltage is 350 mV. It leads to \(|V_{BS}| = 100 \text{ mV.}\) Thus, the \(g_{m2}/g_m\) ratio should be estimated for the PMOS transistor in a similar way that is shown in Fig. 9 for NMOS transistors. In the adopted CMOS process, the PMOS \(g_{m2}/g_m\) ratio for such conditions is 0.34. Thus, \(M_2\) should be biased to provide \(g_m = 24 \mu\)S.

The gate of \(M_2\) is connected to 0 V and its source terminal estimated voltage is 350 mV, leading to \(|V_{GS}| = 350 \text{ mV.}\) The channel length of \(M_2\) is chosen to be \(L = 1 \mu\mu\m). According to Fig. 8, a \(|V_{BS}| = 100 \text{ mV provides a } V_T = 460 \text{ mV, resulting in } V_{OV} = -110 \text{ mV.}\) Such overdrive voltage results in a \(g_m/I_D\) ratio of 24 V\(^{-1}\), leading to \(I_{D2} = 1.0 \mu\)A. As a designer’s choice we will set the tail current exactly to 2.0 \(\mu\)A.

The ideal current source \(I_{ref}\) is set to 500 nA. The channel length of \(M_6\) is set to \(L = 3 \mu\mu\m\) to reduce the tail current dependence on \(V_{DS5}.\) Also, the \(V_T\) of \(M_6\) is reduced to 470 mV, and its overdrive voltage is -90 mV. Transistor \(M_2\) is sized using the \(I_D/W\) relation already stated along the paper. Transistors \(M_{3-4}\) are sized considering a channel

Fig. 15: AC response for the designed pseudo-differential inverter-based OTA of Fig. 14: voltage gain (solid) and phase (dotted).

Fig. 16: 200-run Monte Carlo simulations for the pseudo-differential inverter-based amplifier of Fig. 14. a) GBW; b) Low-frequency voltage gain.

Fig. 17: PMOS bulk-input differential amplifier with active load.
length of $L = 1.5 \mu m$ as a trade-off regarding its output conductance and intrinsic capacitance loading since they operate with an overdrive voltage of $-225$ mV.

Table III summarizes the transistors sizes of the bulk-driven OTA. Figure 18 presents the AC response of designed OTA. It achieved a DC gain of $23.30$ dB and a GBW of $118.56$ kHz ($PM = 89.58^\circ$) and the total power consumption, including the biasing current, is $1.24 \mu W$. The simulated GBW is higher than the specified and is explained due to the reduced intrinsic load capacitance, considered to be $3$ pF at the initial design phase. It means that for reduced GBW (lower than $500$ kHz) there is a low inherent circuit loading. Also, a DC simulation shows that the achieved $g_{mb}$ is very close of the specified: $g_{mb} = 8.45 \mu S$ (difference of only $3.6\%$).

This example focused on sizing only to achieve a specific GBW, thus specific circuit optimizations such as the input common-mode range and output swing are not explored along with the paper. An important comment is that the DC gain could be increased by increasing the $L$ of the input pair of transistors.

### IV. CONCLUSION

This paper described the main design techniques for ULV analog circuits. Trade-offs and design decisions were highlighted, showing the sizing techniques for transistors operating in weak and moderate inversion regions. The $g_{m}/I_D$ technique allows the exploration of all design regions using a single characteristic curve.

It is possible to notice that ULV analog design is not an exact procedure. Although simplified design equations provide a good estimate of circuit bias point, they do not model short channel effects. An interactive size tuning must be done with the aid of an electrical simulator to achieve the required specification.

Four design examples have shown that it is possible to achieve good performance in terms of AC response and power dissipation using a supply voltage much lower than the nominal value. The main challenge is to deal with transistors operating with gate-to-source voltages below $V_T$. It is important to highlight that the design steps described here are not exhaustive. Other design specifications must be taken into account in the sizing process, according to the application requirements. The general procedure, however, was described and exemplified.

The use of ULV analog circuits in IoT applications is increasing and the search for optimum sizing that results in ultra-low power consumption while keeping good performance is mandatory. This leads to the exploration of all device inversion regions and the use of specific design techniques such as bulk-driven, forward bulk biasing, negative transconductance and bulk-based CMFB.

### REFERENCES


---

**Table III:** Summary of the bulk-driven amplifier (Fig. 17) transistors dimensions.

<table>
<thead>
<tr>
<th>Transistors</th>
<th>multiples $\times$ W/L ($\mu\text{m} \times \mu\text{m}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1, M_2$</td>
<td>$8 \times (8.0/1.0)$</td>
</tr>
<tr>
<td>$M_3, M_4$</td>
<td>$8 \times (83.0/1.5)$</td>
</tr>
<tr>
<td>$M_5$</td>
<td>$8 \times (30.0/3.0)$</td>
</tr>
<tr>
<td>$M_6$</td>
<td>$2 \times (30.0/3.0)$</td>
</tr>
</tbody>
</table>

---

Fig. 18: AC response for the designed bulk-driven OTA of Fig. 17: voltage gain (solid) and phase (dotted).


