A 0.12–0.4 V, Versatile 3-Transistor CMOS Voltage Reference for Ultra-Low Power Systems

Arthur Campos de Oliveira, Student Member, IEEE, David Cordova, Student Member, IEEE, Hamilton Klimach, Member, IEEE, and Sergio Bampi, Senior Member, IEEE

Abstract—In this paper, we propose an ultra-low power compact 3-transistor voltage reference capable of operating at ultra-low supply voltages. The proposed circuit is based on the self-cascode MOSFET (SCM), which provides a reference voltage proportional to the threshold voltage ($V_T$) difference of the two NMOS transistors that compose it. Reverse short-channel and narrow-width effects are explored to obtain such $V_T$ difference while using the same type of transistor. Ultra-low power operation and low line sensitivity is achieved by biasing the SCM with a zero-$V_T$ (native) transistor, also leading to an area efficient design. To show its versatility, three versions of the proposed circuit were fabricated in a standard 0.13-µm CMOS process. Measurement performed over five samples showed an average temperature coefficient of 150–1500 ppm/°C. Minimum supply voltages of 0.12–0.4 V was observed while providing reference voltages around tens of mV. The proposed circuits consume 0.33–50 pW at room temperature and minimum supply voltage. The occupied area for any version is less than 0.0012 mm$^2$.

Index Terms—Subthreshold, voltage reference, ultra low power, ultra low voltage, 3 transistor.

I. INTRODUCTION

ULTRA-low power (ULP) sensing and biomedical systems are emerging as the primary targets within the wide range of Internet of Things (IoT) applications. To make these systems feasible, the analog and mixed-signal modules that compose them, such as wake-up timers [1], [2] and temperature sensors [3], must comply with an extreme constrained area and power budget. The voltage reference is an essential building block that composes these modules and frequently has to operate not only constrained by power but also by minimum voltage limitations.

The subthreshold operation has been widely explored to design ULP voltage references [4]–[14]. Many of these can consume units to tens of nanowatts [4]–[9], while the requirement of today’s IoT building blocks can reach the sub-nW power consumption range [15]. Sub-nW voltage references can be obtained through the difference of the threshold voltages ($V_T$) of different devices [10], [13], [14], with a PMOS-only configuration [12], or through the $V_T$ dependence on the transistors geometry [11]. However, many of these designs provide a minimum output voltage of around 0.2 V [10], [12]–[14], which is a limiting factor. Fig. 1 shows the conventional approach for a voltage-to-current (V-I) converter, and for a relaxation oscillator. For a V-I converter shown in Fig. 1(a), achieving low power consumption can be impractical in area-constrained systems due to the large resistor required for such reduction. By generating a sub-100 mV voltage source ($V_S$), nW power consumption can be achieved without significant area penalty [3]. Ultra-low energy relaxation oscillators [1], [2] can also be achieved through sub-100 mV voltage generators, as depicted in Fig. 1(b). For a given target frequency ($f_{OUT}$), considering $V_{REF}/I_{REF} \gg$ comparator and buffer delay, and fixed capacitor value $C$, down-scaling its voltage reference ($V_{REF}$) a smaller current reference ($I_{REF}$) is needed. Consequently, the oscillator energy per cycle ($E_{cycle} = P_{avg}/f_{OUT}$), which is a hard constraint for heavily duty-cycled systems, is reduced.

Fig. 1. Conventional (a) voltage-to-current converter and (b) relaxation oscillator.
II. THE 3-TRANSISTOR VOLTAGE REFERENCE

A. Circuit Description

Fig. 2(a) shows the voltage reference circuit, where M₁ and M₂ compose the self-cascode MOSFET (SCM) which is biased by the current source ILOAD. The current source ensures that both transistors of the SCM operate in weak inversion (WI, or subthreshold). The difference between the gate-to-source voltages of transistors M₁ and M₂ defines the output voltage reference (VREF).

To begin with the circuit analysis, we need a MOSFET model for the weak inversion operation, which is the most appropriate condition for ultra-low-power operation. According to the Unified Current-Control Model (UICM) [16], the drain current of a long channel NMOS transistor operating in WI is given by

\[ I_D = 2eI_S \exp \left( \frac{V_G - V_T}{n\phi_I} \right) \left[ \exp \left( -\frac{V_S}{\phi_I} \right) - \exp \left( -\frac{V_D}{\phi_I} \right) \right] \]

where \( e \) is the Euler’s number, \( I_S = I_{S0}S \), \( S \) is the transistor aspect ratio (W/L), and \( W \) and \( L \) are the channel width and length, respectively. \( I_{S0} = \mu C'_{ox} n\phi_I^2/2 \) is the sheet normalization current, which is process related, \( \mu \) represents the carrier mobility, \( n \) the subthreshold slope factor, \( C'_{ox} \) is the gate capacitance per unit area, and \( \phi_I = kT/q \) is the thermal voltage. \( V_G \), \( V_S \), and \( V_D \) are the gate, source and drain voltages, respectively, all referred to the bulk, and \( V_T \) is the geometry-dependent threshold voltage.

For proper operation, the SCM transistor M₁ can be either in triode or saturation, while M₂ is always saturated. By applying (1) for both transistors (M₁ and M₂) we have

\[ I_{D1} = 2eI_{S1} \exp \left( \frac{V_X - V_T}{n_1\phi_I} \right) \left[ 1 - \exp \left( -\frac{V_{REF}}{\phi_I} \right) \right] \]
\[ I_{D2} = 2eI_{S2} \exp \left( \frac{V_X - V_T - n_2V_{REF}}{n_2\phi_I} \right) \left[ 1 - \exp \left( -\frac{V_{REF}}{\phi_I} \right) \right] \]

From Fig. 2(a), \( I_{D1} = I_{D2} \), and since we are using the same type of transistors, \( n_1 = n_2 \) can be assumed. Thus, a general expression for \( V_{REF} \) can be obtained as

\[ V_{REF} = \phi_I \ln \left[ 1 + \frac{I_{S2}}{I_{S1}} \exp \left( \frac{V_{T1} - V_{T2}}{n\phi_I} \right) \right] \]

Considering that \( V_{T1} > V_{T2} \), which is determined in the design phase, and \( \ln(1 + x) \to \ln(x) \) for \( x \gg 1 \), a simplified expression for \( V_{REF} \) is obtained as

\[ V_{REF} = \frac{V_{T1} - V_{T2}}{n} + \phi_I \ln \left( \frac{I_{S2}}{I_{S1}} \right) \]

which considers that M₁ is saturated, i.e., \( V_{REF} > 3 \sim 4\phi_I \).

The threshold voltage presents a near-linear negative dependence on temperature [17]

\[ V_T(T) = V_T(T_0) - \alpha V_T(T - T_0) \]

where \( T_0 \) is the reference temperature, and \( \partial V_T/\partial T = \alpha V_T \) is the threshold voltage temperature slope.

From (5) and (6), and setting \( \partial V_{REF}/\partial T = 0 \), a value of \( S_2/S_1 \) for optimal temperature compensation can be obtained

\[ \left( \frac{S_2}{S_1} \right)_{OPT} = \frac{I_{S0}^{Q1}}{I_{S0}^{Q2}} \exp \left[ \frac{q}{k} \left( \frac{\alpha V_{T1} - \alpha V_{T2}}{n} \right) \right] \]
Replacing (7) into (5), and neglecting the $\alpha V_T$ terms at room temperature, the temperature compensated $V_{REF}$ is now given by

$$V_{REF} = \frac{V_{T1}(T_0) - V_{T2}(T_0)}{n}$$

(8)

Therefore, the SCM results in a reference voltage equal to the $V_T$s difference from the $M_1$ and $M_2$ transistors divided by the subthreshold slope factor $n$. To achieve $V_{T1} > V_{T2}$, reverse short-channel effect (RSCE), and narrow-width effect (NWE) are explored, as detailed in Section III.

B. $I_{LOAD}$ for Power Supply Rejection Improvement

Considering the small-signal equivalent of the SCM depicted in Fig. 2(b), the Power Supply Rejection (PSR) of the $V_{REF}$ can be simplified as

$$PSR = \frac{v_{ref}/v_{dd}}{v_{x}/v_{dd}} = \frac{v_{ref}/v_{x}}{v_{x}/v_{dd}}$$

(9)

where it represents the voltage transfer function between the supply voltage ($v_{dd}$) and the reference voltage ($v_{ref}$), and it is conveniently split into two transfer functions: $v_{x}/v_{dd}$ and $v_{ref}/v_{x}$, which are the PSR contributions coming from the supply voltage to the current load ($I_{LOAD}$), and that of the SCM to the reference voltage, respectively.

The voltage reference becomes less sensitive to supply variations when $v_{ref}/v_{dd}$, or in turn $v_{x}/v_{dd}$ and $v_{ref}/v_{x}$, are minimized. The value $v_{ref}/v_{x}$ is obtained using the equivalent circuit of Fig. 2(b), for $M_1$ in triode, yielding

$$\frac{v_{ref}}{v_{x}} = \frac{g_{m2} - g_{m1} + g_{ds2}}{g_{ms2} + g_{ds2} + g_{m1}} \approx \frac{g_{ds2}}{g_{ms2}}$$

(10)

where $g_m$ and $g_{ms}$ are the gate and source transconductances and $g_{ds}$ is the drain-source conductance, respectively. The expression shows that $v_{ref}/v_{x}$ can be reduced by either increasing $I_{LOAD}$ or reducing $M_2$ output conductance. As previously discussed, $M_1$ and $M_2$ are sized for the temperature compensation of $V_{REF}$. Therefore, the minimization of $v_{ref}/v_{dd}$ relies only on the implementation of $I_{LOAD}$.

Fig. 3 shows two possible implementations for $I_{LOAD}$. It can be implemented using a PMOS or NMOS transistor acting as a current source. In this case, we are considering both transistors saturated. The expression for $v_{x}/v_{dd}$ is obtained from their small-signal analysis and yields to

$$\left(\frac{v_{x}}{v_{dd}}\right)_{PMOS} \approx \frac{g_{ms}}{g_{ds}} \lor \left(\frac{v_{x}}{v_{dd}}\right)_{NMOS} \approx \frac{g_{ds}}{g_{ms}}$$

(11)

Since, $g_{ms} > g_{ds}$, $v_{x}/v_{dd}$ is higher than unit for a PMOS, worsening the PSR, leading us to discard this option for the current load. On the other hand, this parameter is lower than a unit for an NMOS, which is desired to improve the PSR.

Another reason to avoid the usage of a PMOS for $I_{LOAD}$ implementation is that its n-well behaves as an RC network of the parasitic devices within the well and the interface with the substrate, which increases the $v_{x}/v_{dd}$ as frequency increase [18].

Taking into account that the voltage reference circuit is working in WI and the gate voltage of the NMOS of Fig. 3 is $V_B < V_{DD}$, we have two options for $V_B$ to achieve ultra-low power consumption while improving the PSR:

- Case I: $V_B = V_X$, $v_{x}/v_{dd}$ is increased from $g_{ds}/g_{ms}$ to $g_{ds}/g_{mb}$, Fig. 4(a).
- Case II: $V_B = 0$ V, $v_{x}/v_{dd}$ remains the same as (11), Fig. 4(b).

From the two previous cases, $V_B = 0$ V is chosen for our circuit design versions since it represents the lowest $v_{x}/v_{dd}$, thus resulting in the higher PSR. Since the reference voltage does not depend directly on the $I_{LOAD}$ value when the SCM is kept in WI, the current source design is constrained to obtain the lowest possible area while operating at ultra-low power consumption. Therefore, the SCM is biased by the leakage of only one transistor, which guarantee a minimum voltage drop across $I_{LOAD}$ of $3\sim4\phi$. Thus, the area occupied by the NMOS that implements $I_{LOAD}$ depends only on its channel current density ($I_D/W$). Fig. 5 shows the current densities of the different transistors available in the target process for $V_{GS} \leq 0$ V. The lower $I_D/W$ of the standard-$V_T$ or low-power-$V_T$ transistors would yield larger areas compared to the zero-$V_T$ or low-$V_T$ transistors. Due to this fact, the zero- and low-$V_T$ transistors are considered the best candidates to implement $I_{LOAD}$. The best among them can be obtained through Fig. 6, which shows that the zero-$V_T$ transistor offers the lowest $v_{x}/v_{dd}$ and thus improves the PSR.

The circuit we herein propose has the merit of being a simple 3-Transistor (3T) voltage reference, as shown in Fig. 7(a).
Another reason for the zero-\(V_T\) transistor to provide the best PSR is that the \(V_X\) voltage is mainly defined by the difference of the \(V_T\) of transistors \(M_1\) and \(M_3\) [10], thus providing a regulated voltage above \(V_{REF}\).

Through Fig. 4(b), one can note that there is no explicit solution for the simulated transistors. \(M_1\) is considered saturated, i.e., \(V_{REF} \approx 3 \sim 4\phi_t\). If near or sub-\(\phi_t\) output is desired, (5), and consequently (7), are no longer valid. Therefore, an adequate optimal temperature compensation condition for Sub-\(\phi_t\) \(V_{REF}\) must be defined.

Through (4), one can note that there is no explicit solution to obtain such ratio. Thus, rewriting it, we have

\[
\ln \left( \exp \left( \frac{V_{REF}}{\phi_t} \right) - 1 \right) = \ln \left( \frac{I_{S2}}{I_{S1}} \right) + \frac{V_{T1} - V_{T2}}{n\phi_t} \tag{13}
\]

By making a linear approximation of the log function as

\[
\ln \left( \exp \left( \frac{V_{REF}}{\phi_t} \right) - 1 \right) \approx A + B \frac{V_{REF}}{\phi_t} \tag{14}
\]

where \(A\) and \(B\) are the fitting parameter which depends on the defined interval of \(V_{REF}\). Since we want \(V_{REF} < \phi_t\), the chosen range for \(V_{REF}\) is from \(\phi_t/2\) to \(\phi_t\) [11]. With these values the obtained fitting parameters are \(A = -1.4\) and \(B = 1.94\).

Replacing (14) in (13)

\[
V_{REF} = \frac{V_{T1} - V_{T2}}{nB} + \frac{\phi_t}{B} \ln \left( \frac{I_{S2}}{I_{S1}} \right) - \frac{A}{B} \phi_t \tag{15}
\]
Including the temperature dependence of $V_T$ through (6) in (15), and setting $\partial V_{REF}/\partial T = 0$, the optimal value for the ratio $S_2/S_1$ can be obtained as

$$\left( \frac{S_2}{S_1} \right)_{OPT} = \frac{I_{S01}}{I_{S02}} \exp \left[ \frac{a V_{T1} - a V_{T2}}{n} \right] + A$$

(16)

Through the optimal $S_2/S_1$ ratio, the sub-$\phi_t$ temperature compensated voltage reference can be expressed as

$$V_{REF} = \frac{V_{T1}(T_0) - V_{T2}(T_0)}{n B}$$

(17)

where, by considering $M_1$ to be in triode, the output voltage will also depend on the fitting parameter $B$.

III. CIRCUIT DESIGN

A. Design of the SCM

After defining the best way to implement the bias of the SCM for LS and PSR improvement, the first step to design the proposed circuits is to choose the type of the transistors to be used in the SCM. Since our goal is low power operation, first we choose from the target process the two available transistors with higher $V_T$s. The presented voltage references were designed in a standard 0.13-$\mu$m CMOS process that offers four types of NMOS transistors, being the so-called standard and low power the ones with higher $V_T$s. The dependence of $V_T$ with the transistor dimensions is defined by the reverse short-channel and narrow-width effects, RSCE and NWE, respectively. As well known, the main contributor to such variations is the RSCE. Thus, to achieve low TC, the temperature dependence of $V_T (aV_T)$ with respect to the channel length (L) must be considered, as shown in Fig. 8.

As a conclusion, for this step, the low power type was chosen for the SCM implementation.

Fig. 9(a) shows the variation of $V_T$ with the geometry of the low power transistor, in which $V_T$ is extracted through the $g_m/I_D$ method [16], against channel length (L) for different transistors ratios W/L. As shown, in this CMOS process as L increases there is a decrease in $V_T$ (RSCE), and the same happens when W is increased (NWE). It is important to consider that these effects do not scale with the use of multipliers. With that said, unit transistors to implement $M_1$ and $M_2$ were used to have a well controlled $V_T$ through their $W/L$.

To guarantee that $V_{T1} > V_{T2}$, $M_1$ is sized with a small L and W/L between 1 to 2. A larger value of $V_{T1}$ could be achieved by using the minimum allowed L, but this was avoided to reduce the mismatch variation. Transistor $M_2$ is sized taking into consideration the following:

- $L_2 > L_1$ to guarantee $V_{T1} > V_{T2}$.
- To prevent charging damage in the polysilicon, the maximum thin oxide area for any single gate in this process is 230 $\mu$m$^2$.
- To guarantee an efficient layout area and compliant with $(W \times L)_2 \leq 230$ $\mu$m$^2$, a range of [1 - 2] for $(W/L)_2$ and $L_2 = [10 - 15]$ $\mu$m is defined, as illustrated in Fig. 9(b).

From the previous considerations, the maximum value for $V_{REF}$ is limited by the maximum area of $M_2$. As shown in Fig. 10(a), for a fixed size of $M_1$, the maximum value of the voltage reference is 70 mV, meaning that any value within Sub-$\phi_t < V_{REF} < 3\phi_t$ can be achieved.

B. Design of $I_{LOAD}$

Using the range defined for $(W/L)_2$ with $L_2$ as a starting point, the voltage references are designed to achieve the lowest
possible temperature coefficient (TC) for the minimum $V_X$, as illustrated in Fig. 10(b). As previously stated, the area of $M_2$ must satisfy the constraint of $(W \times L)_2 \leq 230 \mu m^2$. Besides complying with the minimum $M_2$ area, $I_{LOAD}$ must also be chosen to attain a minimum $V_X$ value across temperature. $V_X$ is mainly defined by $M_1$ and $M_3$, but since $M_1$ is set to have a fixed size, only the size of $M_3$ will define both $I_{LOAD}$ and $V_X$. From Fig. 10(b), the resulting $V_X$ that satisfies these conditions is defined by $S_3/S_1$ ratio between 0.32~0.41 [10]. The ratio $S_2/S_1$ for optimal temperature compensation predicted by (7) is around 0.86, while the actual ratio chosen after simulations is $S_2/S_1 = 0.826$. The difference between the predicted and simulated values comes from the fact that $V_{REF} < 3\phi_t$, i.e., $M_1$ is not saturated. However, the proximity of the predicted and simulated values for temperature compensation validates the theoretical approach.

There are two versions of the proposed voltage reference concerning the implementation of the current reference ($I_{LOAD}$) for the SCM. In one version the current reference was designed with one transistor (3T), and another with two series transistors (4T). As stated before, in WI the transistor is said to be saturated when $V_{DS} > 3 \sim 4\phi_t$. Thus, the minimum supply voltage for the conventional 3T version will be given as $V_{DDMIN} = V_{DS3} + V_{DS2} + V_{REF}$, resulting in a minimum supply voltage of approximately 300 mV for $V_X \approx 200$ mV. For the PSR enhanced 4T version, the minimum supply voltage is increased by 100 mV to guarantee saturation of the additional series transistor.

C. Design for Sub-$\phi_t$ $V_{REF}$

From the previous section, the value of $V_{REF}$ is defined by the $V_T$ difference between $M_1$ and $M_2$, where the upper bound is defined by $V_{T1}$ while $V_{T2}$ defines the lower bound of this difference, meaning that any reference voltage value smaller than $\Delta V_{MAX}$ (Fig. 10(a)) can be obtained. Even though optimally temperature compensated, $V_{REF}$ still presents a temperature variation around 150-200 $\mu$V due to the non-linear behavior of the compensated $\Delta a_{V_T}$. Meaning that the same non-linear variation is expected for any reference voltage value. Therefore, one can conclude that as the average reference voltage is reduced, its TC increases.

By following the same procedure used for the conventional 3T approach, $S_1 = \frac{22\mu m}{\sqrt{W}}$ defines the $V_T$ upper bound, being $S_2$ sized considering the trade-off between TC and the value of $V_{REF}$ in order to obtain a $V_{REF} < \phi_t$ with the lowest TC. The simulated dependence of $V_{REF}$ and TC with respect to $S_2$ for a fixed $L_2 = 15 \mu m$, which defines the design space, is shown in Fig. 11. As for the conventional approach, the usage of $L_2$ as a longer channel, while $L_1$ is smaller, implies in a wider voltage reference. Thus, in the case of the sub-$\phi_t$ $V_{REF}$ the dependence of $V_T$ with respect to the W (NWE) will play a major role.

For this design, a voltage reference of less than 20 mV was set as the objective. Considering the TC and output voltage trade-off, from Fig. 11 we can achieve a 15.9 mV output voltage with a TC of 240 ppm/°C. The optimal temperature compensation ratio predicted by (16) was around 0.24, while the one obtained from simulation was 0.267.

As previously defined, the minimum supply voltage of the 3T structure is $V_{DDMIN} = V_{DS3} + V_{X}$, where $V_X = V_{DS2} + V_{REF}$. This implies that for the sub-$\phi_t$ approach $V_{DDMIN}$ would be approximately the same as before since $V_X$ is defined by $M_1$-$M_3$ and has not significantly changed. In the case when PSR and LS specifications are relaxed, $V_{DS3}$ could be reduced, or even not considered, and the minimum supply voltage would be given approximately by the voltage drop across the SCM, resulting in $V_{DDMIN} = V_{DS2} + V_{REF} \approx 120$ mV. This condition would make $M_3$ to operate in triode region, thus reducing the power consumption substantially, at the expense of a much higher LS. This consideration is only possible due to the higher-drive capability of zero-$V_T$ transistors even at extremely low $V_{DS}$ values [19].

As important to note that $\mu_1 \approx \mu_2$ is considered for expressions (7) and (16) since the same type of transistor is used for $M_1$ and $M_2$. Then, the temperature dependent terms $\left(\mu \frac{d^2 \phi}{dT}\right)$ of the specific currents ($I_{SQ}$) cancel each other. For this reason, $I_{SQ}$ is considered to be constant with temperature to provide
Fig. 13. Simulated temperature and supply voltage dependence of the (a) 3T, 4T and (b) Sub-$\phi_t$ versions.

TABLE II
MONTE CARLO SIMULATED PERFORMANCE OF THE PROPOSED VOLTAGE REFERENCES FOR 1000 RUNS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Voltage Reference</th>
<th>3T</th>
<th>4T</th>
<th>Sub-$\phi_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process (μm)</td>
<td></td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp Range (°C)</td>
<td></td>
<td>-25</td>
<td>-125</td>
<td>0 - 125</td>
</tr>
<tr>
<td>$V_{DD\text{MIN}}$ (V)</td>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>0.12</td>
</tr>
<tr>
<td>$V_{REF}$ (mV)</td>
<td></td>
<td>62.2</td>
<td>62.3</td>
<td>15.9</td>
</tr>
<tr>
<td>$\sigma/\mu$ (%)</td>
<td></td>
<td>3.4</td>
<td>3.3</td>
<td>3.8</td>
</tr>
<tr>
<td>LS (%/V)</td>
<td></td>
<td>0.07</td>
<td>0.006</td>
<td>0.216</td>
</tr>
<tr>
<td>TC (ppm/°C)</td>
<td></td>
<td>30.7</td>
<td>33</td>
<td>240.4</td>
</tr>
<tr>
<td>PSR @ 100 Hz (dB)</td>
<td></td>
<td>-67.3</td>
<td>-98</td>
<td>-38.4</td>
</tr>
<tr>
<td>Power (μW)</td>
<td></td>
<td>21</td>
<td>28</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Fig. 14. Measured supply dependence of the (a) 3T, 4T and (b) Sub-$\phi_t$ versions.

IV. SIMULATION AND MEASUREMENT RESULTS

The proposed voltage references were fabricated in a standard 0.13-μm CMOS process. The micrograph of the fabricated chip and the circuit's layout are shown in Fig. 12. The total area is 0.0006, 0.0012 and 0.0009 mm$^2$ for the 3T, 4T and Sub-$\phi_t$ versions, respectively.

Simulated temperature and supply dependence for the proposed circuits is shown in Fig. 13. The 3T reference presents an LS of 0.07 %/V and TC of 30.7 ppm/°C, while the 4T reduces the LS by 10× and keeps the TC the same for all the $V_{DD}$ range. The TC of the Sub-$\phi_t$ is around 750 ppm/°C at 120 mV supply voltage, and this value reduces to 240 ppm/°C for $V_{DD}$ greater than 300 mV, when $M_3$ saturates. In the case of the Sub-$\phi_t$ reference, $M_1$ operates in triode, and the temperature dependency of $V_{REF}$ loses linearity, increasing TC as mentioned before. The low-temperature range was restricted for this reference as a way to impose a limit to TC. For comparison purposes, the simulated performance results for 1000 Monte Carlo runs considering both process and mismatch variations are presented in Table II, where each parameter represents the average of all simulations.

Measurements were performed through triaxial cables using a Keysight 4156C Semiconductor Parameter Analyzer for DC sweep, and a Tenney TPS thermal chamber for temperature control. A total of five chips were packaged in ceramic and measured.

The average measured $V_{DD}$ dependence of the reference voltages for the 3T and 4T are shown in Fig. 14(a), resulting in an average LS of 0.188 and 0.08 %/V, respectively. From the detail shown, one can see the LS improvement that results from the fourth series transistor in the 4T version at the cost of 100 mV increase in the minimum supply voltage. The same dependence was measured for the Sub-$\phi_t$ version as presented in Fig. 14(b), showing that it achieves an LS of 1.62 %/V for $V_{DD}$ of 120 mV and 300 mV, respectively.

The temperature behavior for the five measured samples is presented in Fig. 15, showing TCs from 120 to 208 ppm/°C for the 3T and 4T, and an average of 1537 ppm/°C for the Sub-$\phi_t$. The variability spread can be perceived from this figure, but no significant statistical parameter could be calculated since our measurements were derived from five samples only.
Fig. 16 presents the reference voltage variation of the three versions with respect to its value at 25 °C. From this figure one can see a similar TC slope for all samples of the same version, meaning the TC was not canceled as a result from fabrication process variation impact. The TC tendency could be adjusted by calibration in batches, where only a few samples are measured in different temperatures and the resulting average adjust is applied to all dies in batch, thus reducing calibration cost. This could be implemented by adjusting the sizes of $M_1$ and $M_2$.

The average measured reference voltages were 26, 27.2 and 8.4 mV for the 3T, 4T and Sub-$\phi_t$ versions, respectively, which represents around half of the predicted through circuit simulation. The measured $V_T$ difference for the 3T voltage reference NMOS transistors (Table I) was about 40–50 mV, being 80–90 mV the simulated difference, indicating that even though their $V_T$ difference is lower, the output voltage still follows (5). One possible cause for this discrepancy between simulated and measured $V_{\text{REF}}$ results could be that reverse short-channel effect (RSCE) and narrow-width effect (NWE) is not precisely modeled in the target technology, especially in the subthreshold region, being overestimated for design safety purposes. No ESD protection was used for the output pins.

The power consumption of the 3T voltage reference version and its dependency on temperature and $V_{DD}$ are shown in Fig. 17. As expected, since the SCM is biased by the $M_3$ leakage or subthreshold current of a zero-$V_T$ transistor, its power consumption increases exponentially with respect to temperature. As shown, the circuit consumes less than 50 pW at minimum supply and room temperature, while it increases to less than 8 nW at 125 °C and 1.2 V supply. The same power consumption behavior was measured in the 4T and the Sub-$\phi_t$ circuits, with the Sub-$\phi_t$ version presenting sub-pW consumption at room temperature due to the lower supply voltage.

In order to compare the performance of the proposed circuits with the state-of-the-art, a Figure of Merit (FoM) that considers the main performance parameters of a voltage reference, such as temperature range, TC, power, and silicon area, is used. It can be expressed as [14]

$$\text{FoM} = \frac{(T_{\text{MAX}} - T_{\text{MIN}})^2}{\text{TC} \times \text{Power} \times \text{Area}} \times \frac{1}{10^{17}} \quad (18)$$

First-order compensated voltage references usually present a parabola-like curve across temperature, meaning that as the temperature goes far from the reference point, typically room temperature, the output reference voltage temperature sensitivity increases. Therefore, the temperature range $(T_{\text{MAX}} - T_{\text{MIN}})$ is squared in the presented FoM expression,
TABLE III
COMPARISON WITH STATE-OF-THE-ART ULTRA-LOW POWER CMOS VOLTAGE REFERENCES

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology (μm)</td>
<td>0.35</td>
<td>0.35</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.13</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Vref (V)</td>
<td>0.9-5</td>
<td>1.14</td>
<td>0.45-2</td>
<td>0.45-1.8</td>
<td>0.9-3</td>
<td>0.9-1</td>
<td>0.5-3</td>
<td>0.15-1.8</td>
<td>1.2-2.2</td>
<td>1.4-3.6</td>
<td>0.45-3.3</td>
</tr>
<tr>
<td>Vpp (nV)</td>
<td>670</td>
<td>96.6</td>
<td>263.5</td>
<td>118.41</td>
<td>713</td>
<td>460</td>
<td>176</td>
<td>17.69</td>
<td>986.2</td>
<td>1250</td>
<td>225.3</td>
</tr>
<tr>
<td>Temp. Range (°C)</td>
<td>0-80</td>
<td>-20-80</td>
<td>0-120</td>
<td>-40-85</td>
<td>-20-40</td>
<td>-45-120</td>
<td>-20-80</td>
<td>0-120</td>
<td>-40-85</td>
<td>0-100</td>
<td>0-120</td>
</tr>
<tr>
<td>LS (%/V)</td>
<td>0.27</td>
<td>0.09</td>
<td>0.44</td>
<td>0.033</td>
<td>0.3</td>
<td>0.059</td>
<td>0.036</td>
<td>2.03</td>
<td>0.38</td>
<td>0.31</td>
<td>0.15</td>
</tr>
<tr>
<td>PSR @ 100Hz (dB)</td>
<td>-47</td>
<td>-60</td>
<td>-45</td>
<td>-50.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Power (pW)</td>
<td>36,000</td>
<td>22,000</td>
<td>2,600</td>
<td>15,600</td>
<td>2,970</td>
<td>83,000</td>
<td>29.5</td>
<td>26.1</td>
<td>114</td>
<td>35</td>
<td>54.8</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>0.045</td>
<td>0.0189</td>
<td>0.045</td>
<td>0.0132</td>
<td>0.054</td>
<td>0.061</td>
<td>0.0093</td>
<td>0.0012</td>
<td>0.0048</td>
<td>0.0025</td>
<td>0.002</td>
</tr>
<tr>
<td>FoM (°C/W⋅mm²)</td>
<td>0.0004</td>
<td>0.002</td>
<td>0.009</td>
<td>0.0013</td>
<td>0.0025</td>
<td>0.0002</td>
<td>0.125</td>
<td>0.315</td>
<td>0.23</td>
<td>0.36</td>
<td>1.26</td>
</tr>
</tbody>
</table>

*After trimming, *Simulated;

![Fig. 18. Ultra-low power state-of-the-art voltage references FoM performance with respect to power consumption.](http://www.itrans24.com/landing1.html)

a benefit for designs where a wider temperature range is covered. The product between TC, power at room temperature and silicon area, must be as low as possible, meaning that resistorless solutions are preferred since integrated resistors tend to occupy larger areas when designed for low current operation and good matching when compared to MOS-only configurations.

Table III and Fig. 18 summarize the performance of the proposed circuit versions in comparison with the state-of-the-art ultra-low power voltage references. Voltage references operating at nW power consumption range [4]–[9] usually provide output voltages greater than 100 mV and present larger occupied areas due to the need of additional current sources for biasing. Also, the lowest minimum supply voltage reported for this kind of circuit is around 0.45 V [6], [7]. Compared with the state-of-the-art voltage references operating at pW consumption range [10]–[14], all proposed versions are competitive in performance for TC, LS and occupied area, while presenting the lowest minimum supply voltage and the highest FoM among them. The voltage reference proposed in [11] presents similar TC and LS performances when compared with our Sub-$\phi_t$ version, but with $\sim$80x more power consumption. To the best of our knowledge, the proposed Sub-$\phi_t$ voltage reference circuit represents one of the lowest minimum supply voltage and power consumption presented in literature, reaching 120 mV supply and sub-pW operation. Additionally, the proposed 3T structure can also be used as a sensing element for temperature sensors while offering ultra-low power consumption and low LS.

V. CONCLUSION

A compact and versatile 3T voltage reference capable of operating at very low supply voltages and ultra-low power has been implemented in a standard 0.13-μm CMOS technology. The proposed circuit is based on the SCM, which provides a voltage reference that is proportional to the $V_T$ difference of the transistors that compose it. Distinct $V_T$s using the same type of device are obtained through the geometry dependence of $V_T$, namely the reverse short-channel and narrow-width effects. Due to its versatility, the proposed circuit can operate at minimum supply voltages of 120 to 400 mV, while having a consumption in the femto to picowatt range. Measurement results show that the proposed circuit has the best FoM performance when compared to the state-of-the-art. These results show that the 3T circuit is suitable for ultra-low power systems, and could be applied to area-efficient V-I converters, ultra-low energy relaxation oscillators, and as sensing element for temperature sensors.

ACKNOWLEDGMENT

MOSIS MEP support is acknowledged for chip fabrication.

REFERENCES


Arthur Campos de Oliveira (S’13) received the B.Sc. degree in electrical engineering from the Federal University of Pampa, Brazil, in 2015, and the M.Sc. degree in microelectronics from the Federal University of Rio Grande do Sul, Brazil, in 2017. He is currently pursuing the Ph.D. degree with the Delft University of Technology, Delft, The Netherlands.

His current research interests include analog subthreshold circuit design, power management, and sensing systems.

David Cordova (S’12) received the B.Sc. degree in electronics engineering from the Pontifical Catholic University of Peru, Lima, Peru, in 2007, and the M.Sc. degree in microelectronics from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 2014. He is currently pursuing the Ph.D. degree with the Conception Group on IMS and the University of Bordeaux, Bordeaux, France, with a focus on frequency generator circuits toward 5G. He completed the IC-Brazil Program with RFIC design specialization at the IC Brazil Training Center 1, Porto Alegre, Brazil, in 2010.

His research interests include analog and RF IC design in CMOS technologies.

Hamilton Klimach (S’04–M’09) received the B.Eng. and M.Sc. degrees from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 1998 and 1994, respectively, and the Dr.Eng. degree from the Federal University of Santa Catarina, Florianopolis, Brazil, in 2008, all in electrical engineering.

Since 1990, he has been a Full Professor with the Department of Electrical Engineering, Federal University of Rio Grande do Sul, where he was the Head of the Department from 2010 to 2011, and the Dr.Eng. and M.Sc. advisor in the Microelectronics Graduate Program (PGMicro) of the same university. Since 2016, he has been the Local Coordinator of the Training Center 1 of the Brazilian Design Training Program (IC Brazil Program), where he has been serving as a Lecturer since 2009.

His research interests are in the areas of analog, mixed-signal and RF IC design, in semiconductor device modeling, and in variability modeling. He is a member of the IEEE Circuits and Systems Society.

Sergio Bampi (M’86–SM’18) received the B.Sc. degree in electronics engineering and physics from the Federal University of Rio Grande do Sul (UFRGS) in 1979, and the M.Sc. and Ph.D. degrees in electrical engineering from Stanford University, USA, in 1986.

He joined the UFRGS Microelectronics Group, Informatics Institute, in 1986, where he is currently a Full Professor. He serves as a Ph.D. Advisor and a Principal Investigator in the Microelectronics and Computer Science Programs at UFRGS since 1988.

He was a Distinguished Lecturer (2009–2010) of the IEEE Circuits and Systems Society, as the Technical Director of the Brazilian Design Training Program (PGMicro) of the same university. Since 2016, he has been the Local Coordinator of the Training Center 1 of the Brazilian Design Training Program (IC Brazil Program), where he has been serving as a Lecturer since 2009.

His research interests are in the areas of analog, mixed-signal and RF IC design, in semiconductor device modeling, and in variability modeling. He is a member of the IEEE Circuits and Systems Society.