Ultra-Low-Power CMOS Voltage Reference Topologies Regarding Technology Node

Fabián Olivera[1,2], Lucas Souza da Silva[1] and Antonio Petraglia[2]
[1]Federal Center for Technological Education Celso Suckow da Fonseca (CEFET/RJ), Nova Friburgo, RJ, Brazil
[2]Federal University of Rio de Janeiro (UFRJ), EPOLI/PEE/COPPE, Rio de Janeiro, RJ Brazil
folivera@pads.ufrj.br; lucas.souza@aluno.cefet-rj.br; antonio@pads.ufrj.br

2022 IEEE 13th Latin America Symposium on Circuits and System (LASCAS) | 978-1-6654-2008-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/LASCAS53948.2022.9789042

***Abstract*—Ultra-low-power (ULP) designs have received great attention due to the emergence of a wide variety of devices of internet-of-things (IoT) applications. The scaling down of CMOS technology introduces additional challenges for designers, since circuit topologies are usually modified in order to maintain performance under the effects of the next technology. In this paper, a comparative study of ULP voltage references (VRs) is advanced considering four CMOS technologies: 180 nm (bulk), 90 nm (bulk), 65 nm (bulk) and 28 nm (FD-SOI). In addition, major nanometer effects such as drain induced barrier lowering (DIBL), gate induced drain leakage (GIDL), gate leakage due to direct tunneling, junction leakage, among others, are carefully considered in order to suggest an appropriated VR topology regarding each technology node.**

***Index Terms*—CMOS voltage reference, deep-nanometer effects, femto-watt consumption, internet-of-things (IoT), ultra-low-power (ULP) design.**

I. Introduction

The demand for ultra-low-power (ULP) devices has gained great attention in recent years, in view of the continuous growing of internet-of-things (IoT) applications. Self-powered systems based on energy harvesting are potential candidates for improving the supply powering of systems-on-chips (SoC) [1], [2]. Low-power and low-supply voltage references (VRs) are required by low-energy relaxation oscillators of self-powered systems [3], [4], in order to operate at supply headroom of around 100∼150 mV in thermoelectric and photovoltaic harvesters. Moreover, low-dropout regulators of these energy management systems also need an ULP VR to provide an output regulation.

On the other hand, sub-threshold ULP designs are substantially deteriorated by leakage currents, which makes it difficult to identify the source of these leakages to propose the proper circuit topology at each CMOS process [5]. A well-known leakage source, that mainly affects bulk CMOS processes, is the reverse-biased junction leakage current to the substrate, which exponentially increases with temperature. Moreover, nanometer CMOS nodes (< 90 nm) produces major challenges for designers, since both drain-induced barrier lowering (DIBL) and gate-induced drain leakage (GIDL) effects cause threshold voltage shifting and substrate leakage currents, respectively. The use of silicon-on-insulator (SOI) CMOS processes may drastically reduce the substrate leakages caused by reverse-biased junctions and GIDL effects [6]. However, in addition to these issues, deep nanometer CMOS nodes (< 65 nm) are deteriorated by direct tunneling effects, which introduce undesirable currents from the gate to both drain and source device terminals [7].

VRs operating at femto-watt power consumption have been reported in [8], [3] and [4], over 65 nm, 130 nm and 180 nm bulk CMOS technologies, respectively. Substantial degradation of the supply regulation (4.7 mV/V) was obtained in [8] regarding designs [3] and [4], which shows the influence of nanometer effects on VR performance. With the purpose of indicating an appropriate VR topology to achieve femto-watt power consumption, dealing with nanometer and deep-nanometer effects, this paper presents the analysis, design and performance comparisons of VR topologies concerning four CMOS technology nodes: 180 nm (bulk), 90 nm (bulk), 65 nm (bulk) and 28 nm (FD-SOI). In addition, extensive Monte Carlo simulations are advanced in order to verify the proper operation of each design. The rest of the paper is organized as follows. Section II presents the modeling, analysis and comparisons of 2-transistor (2T), 3T and 4T ULP VRs. Section III shows the performance results of several topology designs obtained by extensive Monte Carlo simulations at each CMOS process. Concluding remarks are made in Section IV.

II. Ultra-Low-Power CMOS Voltage References

A. Modeling at Sub-threshold Region

Nanoscale bulk-CMOS transistors operating at sub-threshold region and ultra-low current levels, have four currents that should be considered for proper modeling [5], [6] (see Fig. 1(a)): the sub-threshold current from drain to source (I_{DS}), the gate-to-drain/source (or vice versa) leakage current ($I_{G(D/S)L}$) due to the direct tunneling effect, the reverse-biased junction leakage current from drain/source-to-substrate ($I_{(D/S)L}$), and the gate induced drain leakage current (I_{GIDL}) caused by the high field effect in the drain junction. Substrate leakages can be considerably reduced in FD-SOI CMOS processes in view of the fact that both channel and drain/source areas are isolated from the substrate by the ultra-thin buried oxide (see Fig. 1(b)). An appropriate expression for the drain-to-source sub-threshold current that includes the main nanometer effects, such as, drain induced barrier lowering (DIBL) and body biasing, is given by ([4], [9])

$$I_{DS} = I_{So}\frac{W}{L}\exp\left(\frac{V_{GS}-V_{T,\text{eff}}}{nU_T}\right)\left(1-\exp\left(\frac{-V_{DS}}{U_T}\right)\right) \quad , \quad (1)$$

where U_T is the thermal voltage n, $V_{T,\text{eff}}$, I_{So} and W/L are slope factor, effective threshold voltage, specific current and aspect ratio of the MOS transistor, respectively, and V_{GS} and V_{DS} denote the gate-to-source and drain-to-source voltages, respectively. A first-order approximation for the effective threshold voltage $V_{T,\text{eff}}$ in (1) can be modeled by using

$$V_{T,\text{eff}} = V_{To} - \lambda V_{DS} - \eta V_{BS} \quad , \tag{2}$$

where V_{To} is the zero-bias threshold voltage extrapolated at $V_{DS} = V_{BS} = 0$ V, λ and η are parameters for modeling DIBL and body bias effects, respectively, and V_{BS} is the substrate-to-source voltage.

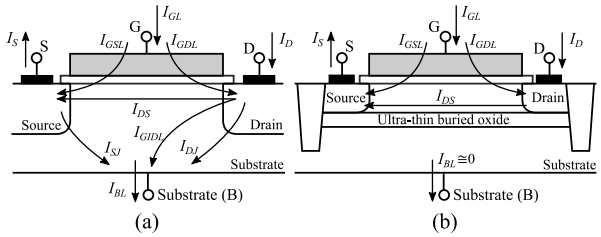

Fig. 1. Main sources of leakage in a nanoscale NMOS transistor: (a) bulk CMOS and (b) FD-SOI CMOS.

Fig. 2 shows the sub-threshold leakage over the total leakage current ratio, where $I_{\text{Leak}} = |I_{BL}| + |I_{GL}|$ for 28 nm FD-SOI, 65 nm bulk, 90 nm bulk and 180 nm bulk CMOS processes, using some available devices at each process design kits: high threshold voltage (HVT), standard threshold voltage (SVT) and low threshold voltage (LVT). As expected, in view of their low current levels in off state ($V_{GB} = 0$ V), HVT devices are the most affected ones by leakage currents, worsening at low temperature operation, since sub-threshold currents become of similar order or even less than leakage. It should be observed that narrow channel dimensions present the best ratio in bulk CMOS processes, due to the fact that the narrower the junction widths and gate-drain borders, the lower leakages caused by reverse junctions and tunneling effects are, respectively. In 28 nm FD-SOI CMOS process, narrow channel devices do not improve leakage ratio for two reasons: (1) the gate leakage increases at almost the same rate as does the sub-threshold one regarding the device width, and (2) the use of ultra-thin buried oxide considerably reduces the substrate leakage.

B. 2T Voltage References

The self-biased concept, which takes advantage of the sub-threshold leakage for circuit biasing, have been widely used to implement ULP VRs [3], [4], [10], [11]. A 2-transistor (2T) ULP VR can be performed either by using the gate-to-ground (GG) self-bias (SB) [10] (Fig. 3(a)) structure or gate-to-source (GS) SB [11] (Fig. 3(b)) in series with a diode-connected MOS. The diode-connected MOS can be implemented either with NMOS (Fig. 3(c)) or PMOS (Fig. 3(d)) devices.

An important characteristic of GG-SB and GS-SB alternatives is the current behavior regarding voltage V_x (see Fig. 3), since this current is responsible to determine the reference

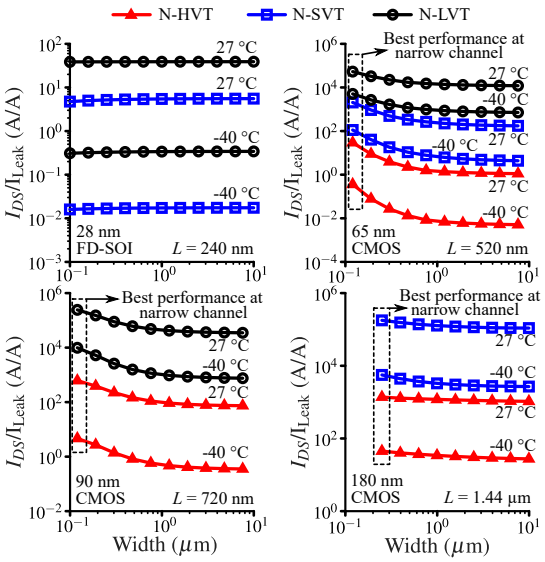

Fig. 2. Sub-threshold leakage over total leakage ratio as a function of transistor width. Simulation conditions: $V_{DB} = 0.3$ V, $V_{GB} = V_{SB} = 0$ V and $L = 8L_{\text{min}}$.

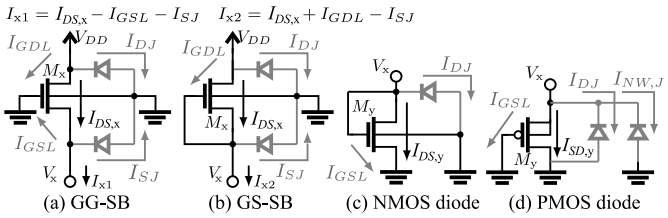

Fig. 3. 2T VR components along with the main leakage sources (in grey): (a) gate-to-ground (GG) SB; (b) gate-to-source (GS) SB; (c) diode-connected NMOS; (d) diode-connected PMOS.

voltage over the diode-connected MOS. Fig. 4 shows the current behavior of GG-SB ($I_{x,1}$), GS-SB ($I_{x,2}$), and NMOS diode (I_y) as a function of V_x in 65 nm bulk-CMOS. It should be observed that the current of GG-SB structure decreases at higher rate than that of GS-SB, thereby causing a reference voltage degradation (ΔV_x) at -30 °C in view of the fact that leakage currents dominate $I_{x,1}$. At the cost of additional power consumption, GS-SB structure can operate at lower temperatures at the same dimensions.

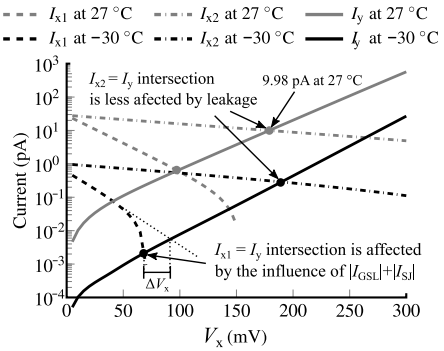

Fig. 4. Currents I_{x1}, I_{x2} and I_y simulations by using long ($L_x = L_y = 3$ µm) and wide ($W_x = 11$ µm and $W_y = 4$ µm) channel transistor in 65 nm bulk.

Another characteristic is the power supply sensitivity. This behavior can be obtained by differentiating (1) for both GG-SB and GS-SB connections, whose results are presented in Table I, which also include the sensitivity of the relative current variation regarding V_x. As can be appreciated, these

TABLE I
SENSITIVITY BEHAVIOR OF GG AND GS IMPLEMENTATIONS.

Topology	$\Delta V_x/\Delta V_{DD}$ [V/V]	$(\Delta I_{DS}/I_{DS})/\Delta V_x$ [1/V]
GG	$\lambda/(\lambda+\eta+1)$	$-(\lambda+\eta+1)/(nU_T)$
GS	$\lambda/(\lambda+\eta)$	$-(\lambda+\eta)/(nU_T)$

sensitivities can be expressed as a function of λ and η. Although the GS-SC structure has lower leakage influence at low temperatures, it presents higher sensitivity to supply variations, which can be a limitation on the design depending on the CMOS technology. Fig. 5 shows the values of λ and η parameters extracted from simulations for several device dimensions. As can be observed, 65 and 90 nm processes have higher λ values than do the 28 and 180 nm ones, above 20 mV/V in some cases. It deteriorates the supply sensitivity and hence more complex 3T and 4T voltage references have been proposed in the literature to improve the supply rejection.

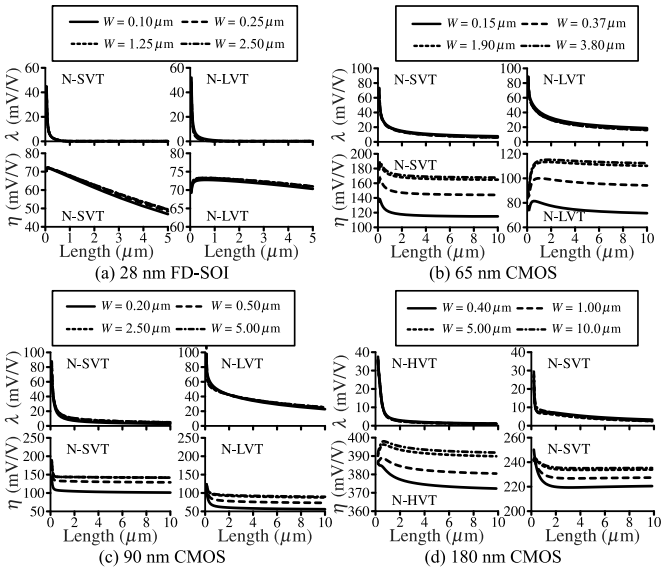

Fig. 5. Extraction results from simulations for λ and η as a function of transistor dimensions at 27 °C.

C. 3T and 4T Voltage References

Based on the 2T VR concept, 3T and 4T VRs were reported by including additional devices in order to improve their performances. The 3T versions of GG and GS structures are depicted in Figs. 6(a) [3] and 6(b) [12], respectively, in which the NMOS-diode was substituted by the self-cascode structure, therefore reducing power consumption and improving supply regulation regarding 2T versions. Further improvements in terms of line regulation can be obtained by using the 4T double-gate-to-ground (DGG) [3] of Fig. 6(d). The 3T-GG and 3T-GS topologies need 3 transistors operating in saturation region, having a minimum supply voltage limitation. To reduce this minimum supply by keeping low-power consumption, a triode-regulated (TR) VR was proposed in [4] (see Fig. 6(c)), which only requires 2 transistors operating in saturation, thereby realizing minimum supply of approximately 120 mV. On the other hand, with the purpose of achieving a stable reference voltage in the presence of high gate leakage current, a 4T GG-based topology with an auxiliary leakage compensation (LC) device (M_4) was reported in [7] (see Fig. 6(e)).

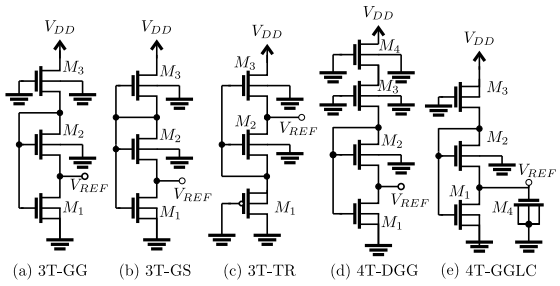

Fig. 6. 3T and 4T ULP CMOS voltage reference topologies.

III. SIMULATION PERFORMANCE OF ULP VRS

In order to demonstrate the proper femto-watt operation of 3T and 4T VRs discussed in Section II, and compare their performances, the design of appropriate topologies at each CMOS technology node were carried out by following the methodology reported in [4]. Table II shows the performance summary from simulation of each design, and the corresponding device dimensions are presented in Table III. The reference voltage (V_{REF}) and temperature coefficient (TC) were obtained by 1000 runs of Monte Carlo simulations, including mismatch and process variations. A traditional metric to asses the performance of VRs is the well-known figure-of-merit (FoM) [3]

$$\text{FoM} = (T_{\max} - T_{\min})^2/(\text{TC}_{\text{avg}} \cdot \text{Power} \cdot \text{Area} \cdot 1\text{e}21) \quad , \quad (3)$$

where $T_{\max}$ and $T_{\min}$ are the maximum and minimum measured temperatures, respectively. Fig. 7(a) shows this metric as a function of the occupied area for all designs. In view of its extreme low-power consumption, the 3T-GG topology presents higher FoM than other topologies from 180 nm down to 65 nm. In 28 nm technology, over high gate leakage conditions, only the 4T-GGLC topology achieves stable operation. Although 3T-GG and 4T-GGLC are appropriate candidates for ultra-low-power applications in all processes, a lower supply voltage can be obtained by using 3T-TR and 4T-DGG as can be appreciated in Fig. 7(b), since they can be designed with only two transistors operating in saturation region. An advantage of 3T-TR topology is that it is completely implemented only by using standard CMOS. It should also be observed from Fig. 7(b) that, due to the high DIBL effect, 65 nm and 90 nm process VRs suffer with a deteriorated LR, and the 4T-DGG topology can improve this regulation.

TABLE II
PERFORMANCE SUMMARY OF CMOS VRS DESIGNED FOR ULP OPERATION

Ref#	Topology	Technology	$V_{DD,\min}$ (mV)	LR $\left(\frac{\text{mV}}{\text{V}}\right)$	PSRR (dB)[a]	V_{REF} (mV)	Power (pW)	TC_{avg} / TC_{std} $\left(\frac{\text{ppm}}{^\circ C}\right)$[b]	Area (μm^2)[c]	FoM $\left(\frac{^\circ C^3 mm^2}{W}\right)$
D01	3T-GG	180 nm bulk	160	0.919	−69.2	54.32	0.146	67.96 / 22.32	2.72	949,522
D02	3T-GS	180 nm bulk	230	0.683	−59.7	121.3	0.289	36.64 / 28.29	19.2	126,110
D03	4T-DGG	180 nm bulk	150	0.733	−84.6	59.70	0.069	44.35 / 24.93	12.0	697,470
D04	3T-TR	180 nm bulk	120	2.200	−61.0	65.70	0.252	90.31 / 59.32	5.04	223,189
D05	3T-GG	90 nm bulk	180	2.923	−62.0	85.22	0.142	98.26 / 9.00	1.50	1,228,364
D06	3T-GS	90 nm bulk	220	4.874	−55.4	155.4	1.237	91.33 / 19.85	7.05	32,186
D07	4T-DGG	90 nm bulk	110	2.000	−65.8	57.37	0.192	75.57 / 36.12	4.46	396,456
D08	3T-TR	90 nm bulk	150	3.972	−61.6	67.27	1.178	124.5 / 21.36	3.97	44,033
D09	3T-GG	65 nm bulk	190	4.796	−51.9	125.1	0.181	104.6 / 38.22	5.83	232,099
D10	3T-GS	65 nm bulk	240	4.857	−55.4	157.3	2.066	114.1 / 16.44	3.19	34,065
D11	4T-DGG	65 nm bulk	220	1.953	−65.7	111.2	0.429	173.2 / 36.01	5.39	63,990
D12	4T-GGLC	28 nm FD-SOI	170	0.869	−71.6	77.74	3.290	87.57 / 21.39	0.55	162,357

[a] Power supply rejection ratio (PSRR) at 100 Hz. [b] Average (avg) and standard deviation (std) values of the temperature coefficient (TC) were computed from −40 to 120 °C. [c] The occupied area was estimated as 3 times the total gate area.

TABLE III
DEVICE DIMENSIONS OF CMOS VRS DESIGNED FOR ULP OPERATION

Ref#	W_4/L_4 $\left(\frac{\mu m}{\mu m}\right)$	W_3/L_3 $\left(\frac{\mu m}{\mu m}\right)$	W_2/L_2 $\left(\frac{\mu m}{\mu m}\right)$	W_1/L_1 $\left(\frac{\mu m}{\mu m}\right)$
D01	-	0.35/0.71 (NS)	0.37/0.60 (NS)	0.35/1.26 (NH)
D02	-	0.35/5.82 (NS)	0.36/2.38 (NS)	0.35/10.00 (NH)
D03	0.39/2.89 (NS)	0.56/1.80 (NS)	0.67/0.85 (NS)	0.36/3.77 (NH)
D04	-	0.40/2.50 (NS)	0.60/0.28 (NS)	0.40/1.28 (PS)
D05	-	0.20/1.08 (NS)	0.68/0.29 (NS)	0.55/0.17 (NH)
D06	-	1.33/0.35 (NS)	0.54/2.36 (NS)	0.25/2.55 (NH)
D07	0.21/3.07 (NL)	0.27/0.16 (NL)	2.54/0.20 (NS)	0.45/0.74 (NH)
D08	-	0.21/2.86 (NS)	0.38/0.59 (NS)	0.92/0.58 (PS)
D09	-	0.16/6.25 (NL)	0.35/1.34 (NS)	0.17/3.13 (NH)
D10	-	2.16/0.24 (NS)	0.64/0.60 (NS)	0.23/0.75 (NH)
D11	1.90/0.27 (NL)	0.17/4.18 (NL)	0.80/0.53 (NS)	0.16/1.12 (NH)
D12	0.17/0.45 (NL)	0.27/0.04 (NL)	0.11/0.50 (NL)	0.09/0.59 (NS)

NS: Standard VT N-type; NH: High VT N-type ; NL: Low VT N-type
PS: Standard VT P-type

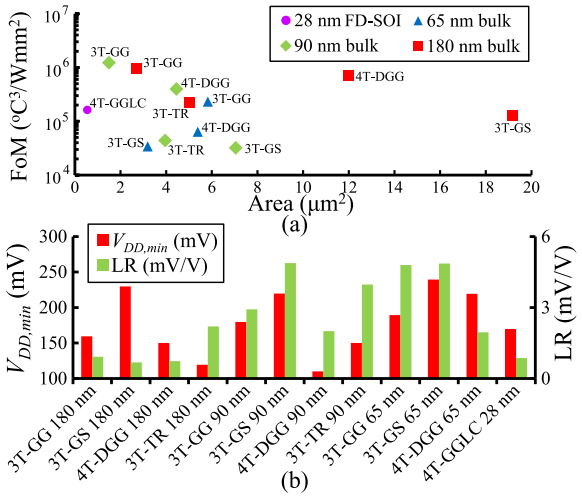

Fig. 7. Performance summary of the designed ULP VRs: (a) FoM as a function of occupied area; (b) minimum supply and line regulation (LR) performances as a function of both circuit topologies and CMOS technology.

IV. CONCLUSIONS

This paper advanced performance comparisons of ULP VR topologies over four CMOS technologies, in order to evaluate the influence of nanometer (<90 nm) and deep-nanometer (<65 nm) effects. Simulation results showed that 3T-GG (65 nm, 90 nm and 180 nm) and 4T-GGLC (28 nm) designs have excellent figure-of-merit performance in comparison with other ones reported in literature. However, in 65 nm and 90 nm processes at high DIBL effect conditions, 4T-DGG can reduce LR. On the other hand, the 3T-TR (90 nm and 180 nm) topology achieves minimum supply of around 120 mV, being a simple VR that can be implemented only by standard devices, hence avoiding the use of special masks.

REFERENCES

[1] J. Zhao, R. Ghannam *et al.*, "Self-powered implantable medical devices: Photovoltaic energy harvesting review," *Adv. Healthcare Mater.*, vol. 9, no. 17, p. 2000779, 2020.

[2] C. Xu, Y. Song *et al.*, "Portable and wearable self-powered systems based on emerging energy harvesting technology," *Microsyst. Nanoeng.*, vol. 7, no. 25, pp. 2055–7434, 2021.

[3] A. C. de Oliveira, D. Cordova *et al.*, "A 0.12–0.4 v, versatile 3-transistor cmos voltage reference for ultra-low power systems," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 65, no. 11, pp. 3790–3799, 2018.

[4] F. Olivera, L. S. da Silva, and A. Petraglia, "A 120 mv supply, triode-regulated femto-watt cmos voltage reference design," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 68, no. 2, pp. 587–591, 2021.

[5] S. G. Narendra and A. P. Chandrakasan, *Leakage in Nanometer CMOS Technologies*. New York: Springer, 2006.

[6] M. Alioto, "Understanding dc behavior of subthreshold cmos logic through closed-form analysis," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 57, no. 7, pp. 1597–1607, 2010.

[7] F. Olivera and A. Petraglia, "Gate leakage compensation technique for self-cascode based voltage references," *Electronics Letters*, vol. 56, no. 22, pp. 1174–1176, 2020.

[8] H. Wang and P. P. Mercier, "A 420 fw self-regulated 3t voltage reference generator achieving 0.47%/v line regulation from 0.4-to-1.2 v," in *IEEE European Solid State Circ. Conf.*, 2017, pp. 15–18.

[9] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd ed. Oxford University Press, 2011.

[10] D. Albano, F. Crupi *et al.*, "A sub- kt/q voltage reference operating at 150 mv," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 23, no. 8, pp. 1547–1551, 2015.

[11] H. Wang and P. P. Mercier, "A 14.5 pw, 31 ppm/°c resistor-less 5 pa current reference employing a self-regulated push-pull voltage reference generator," in *IEEE International Symposium on Circ. and Syst.*, 2016, pp. 1290–1293.

[12] A. C. de Oliveira, J. G. Caicedo *et al.*, "0.3 v supply, 17 ppm/°c 3-transistor picowatt voltage reference," in *IEEE 7th Latin American Symposium on Circ. and Syst*, 2016, pp. 263–266.