Delft University of Technology

Cryo-CMOS Voltage References for the Ultrawide Temperature Range From 300 K Down to 4.2 K

van Staveren, Job; Padalia, Pinakin M.; Charbon, Edoardo; Almudever, Carmen G.; Scappucci, Giordano; Babaie, Masoud; Sebastiano, Fabio

DOI
10.1109/JSSC.2024.3378768

Publication date
2024

Document Version
Final published version

Published in
IEEE Journal of Solid-State Circuits

Citation (APA)
van Staveren, J., Padalia, P. M., Charbon, E., Almudever, C. G., Scappucci, G., Babaie, M., & Sebastiano, F. (2024). Cryo-CMOS Voltage References for the Ultrawide Temperature Range From 300 K Down to 4.2 K. *IEEE Journal of Solid-State Circuits*, *59*(9), 2884-2894. https://doi.org/10.1109/JSSC.2024.3378768

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Cryo-CMOS Voltage References for the Ultrawide Temperature Range From 300 K Down to 4.2 K

Job van Staveren, Pinakin M. Padalia, Edoardo Charbon, *Fellow, IEEE*,
Carmen G. Almudever, *Member, IEEE*, Giordano Scappucci,
Masoud Babaie, *Senior Member, IEEE*,
and Fabio Sebastiano, *Senior Member, IEEE*

***Abstract*— This article presents a family of sub-1-V, fully-CMOS voltage references adopting MOS devices in weak inversion to achieve continuous operation from room temperature (RT) down to cryogenic temperatures. Their accuracy limitations due to curvature, body effect, and mismatch are investigated and experimentally validated. Implemented in 40-nm CMOS, the references show a line regulation better than 2.7%/V from a supply as low as 0.99 V. By applying dynamic element matching (DEM) techniques, a spread of 1.2% (3σ) from 4.2 to 300 K can be achieved, resulting in a temperature coefficient (TC) of 111 ppm/K. As the first significant statistical characterization extending down to cryogenic temperatures, the results demonstrate the ability of the proposed architectures to work under cryogenic harsh environments, such as space- and quantum-computing applications.**

***Index Terms*— Body effect, cryogenic CMOS (cryo-CMOS), DTMOS, extreme environment, MOS-based, quantum computing, voltage references.**

I. Introduction

VOLTAGE references are a key component in many electronic systems, such as sensor readouts [1], data converters [2], [3], and supply regulators [4]. Although electronic circuits must typically only ensure operation over the military temperature range from −55 °C to +125 °C, several applications, such as space exploration [5], require electronics capable of operating over a significantly extended temperature range, for example, lunar temperatures are ranging from −230 °C to +120 °C [6]. Control electronics for particle detectors [7] or quantum computing applications even require operating temperatures as low as 100 mK and below, and up to a few tens of kelvins due to self-heating [8]. Given its very-large-scale-of-integration (VLSI) capabilities, high-frequency operation, and wide operating temperature range, nanometer-CMOS technology is an ideal candidate to implement such cryogenic electronics. Cryogenic CMOS (cryo-CMOS) voltage references are, therefore, extremely relevant for the development of such wide-temperature-range applications.

For the standard temperature range, state-of-the-art voltage references typically use Si bipolar transistors (BJTs) [9], [10], [11], where a proportional-to-absolute-temperature voltage (PTAT) and a complementary-to-absolute-temperature (CTAT) voltage are summed to generate a first-order temperature-independent reference voltage, fundamentally equal to the bandgap voltage of silicon. However, bandgap references suffer from poor performance at cryogenic temperatures due to freeze-out effects in the base region [12], [13], rendering Si BJTs not useful for cryogenic electronics. Moreover, BJTs are fundamentally incompatible (at cryogenic temperatures) with the low supply voltages used in nanometer CMOS technologies, because the base–emitter voltage V_{be} is higher than 1.1 V at cryogenic temperatures, even for nA collector currents. The SiGe heterojunction bipolar transistor (HBT) can overcome such limitations, as it is functional down to mK temperatures, and has already been used in references [13], [14]. However, HBTs are not available in standard CMOS processes and are not suitable for cryogenic sub-1-V designs, since they also require a V_{be} above 1 V at cryogenic temperatures.

Alternatively, MOS devices in weak inversion have been employed at room temperature (RT) [15], [16] and remain well-behaved down to mK temperatures [17], [18], [19]. However, all prior works employing MOS devices instead of BJTs in cryo-CMOS voltage references lack the statistical characterization and require high supply voltages [12], [20] (3 and 5.5 V), thus being unsuitable for sub-1-V applications. Next to combining voltages with complementary temperature dependence, MOS-based references can exploit the

Manuscript received 7 August 2023; revised 22 November 2023 and 3 February 2024; accepted 7 March 2024. Date of publication 3 April 2024; date of current version 27 August 2024. This article was approved by Associate Editor Taekwang Jang. This work was supported in part by Intel and in part by the Research Program OTP through The Netherlands Organization for Scientific Research (NWO) under Project 16278. *(Corresponding author: Job van Staveren.)*

Job van Staveren and Fabio Sebastiano are with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CJ Delft, The Netherlands, and also with QuTech, 2628 CJ Delft, The Netherlands (e-mail: j.vanstaveren@tudelft.nl).

Pinakin M. Padalia was with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CJ Delft, The Netherlands, and also with QuTech, 2628 CJ Delft, The Netherlands. He is now with Renesas Electronics India Pvt., Ltd., Bengaluru 560093, India.

Edoardo Charbon is with the Kavli Institute of Nanoscience Delft, 2628 CJ Delft, The Netherlands, and also with the École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

Carmen G. Almudever was with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CJ Delft, The Netherlands, and also with QuTech, 2628 CJ Delft, The Netherlands. She is now with the Departamento de Informática de Sistemas y Computadores, Universitat Politècnica de València, 46022 Valencia, Spain.

Giordano Scappucci is with QuTech and the Kavli Institute of Nanoscience Delft, Delft University of Technology, 2628 CJ Delft, The Netherlands.

Masoud Babaie is with the Department of Microelectronics, Delft University of Technology, 2628 CJ Delft, The Netherlands, and also with QuTech, 2628 CJ Delft, The Netherlands.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2024.3378768.

Digital Object Identifier 10.1109/JSSC.2024.3378768

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

zero-temperature-coefficient (ZTC) point, which is a specific gate–source voltage V_{gs} corresponding to the drain current I_d being constant over temperature [21], [22]. However, extending this principle to cryogenic temperatures would require reliable CAD-compatible cryogenic device models, which are only scarcely available and have significant limitations, such as coverage for only a limited set of geometries [17], [19], [23]. Although a cryogenic ZTC-based reference has been demonstrated [13], the lack of statistical characterization still leaves uncertainty on the robustness with respect to process variations.

As an alternative, this article presents a series of MOS-based voltage references employing NMOS, PMOS, or DTMOS as core elements and capable of operating from a sub-1-V supply from 300 down to 4.2 K. Extending on [24], we present extensive characterization over process, supply voltage and temperature, together with the assessment of the performance improvement when using dynamic element matching (DEM) and trimming. By providing a systematic study of several main error sources, this work lays the basis for the design of the accurate low-voltage wide-temperature range cryo-CMOS voltage references presented in this article.

The article's organization is as follows. Section II presents a brief study of the changes in CMOS device behavior at cryogenic temperatures, after which Section III describes the implementation of the proposed voltage reference architectures. Finally, Section IV shows the measurements of the fabricated chip, and Section V provides a conclusion.

II. Cryo-CMOS Design Challenges

One of the major design challenges for cryo-CMOS circuits is the lack of CAD-compatible cryogenic device models, making it difficult to quantitatively predict circuit performance. Due to the cryogenic shift in device performance and the numerical instability in the foundry device models when extrapolated beyond their range of validity, also standard foundry models cannot be used at cryogenic temperatures. Still, by comparing characterization data [25], [26] at 300 and 4.2 K, boundaries for the main relevant changes in device and circuit behavior can be derived to ensure robust circuit design, although unfortunately no circuit simulations can be performed.

First, the threshold voltage V_{th} increases by 100–150 mV, which effectively reduces the available headroom by the same amount, implying that cryo-CMOS low-voltage circuit design is even more challenging than at 300 K. For example, pass-gates can stop conducting in a dead-zone around mid-supply due to the increased threshold voltage of both the PMOS and NMOS transistor [27]. In this work, this challenge is overcome by maximizing overdrive on the switches, or using pass-gates only when higher ON-resistance is tolerated.

Second, the subthreshold slope (SS) is steeper at cryogenic temperatures, causing transistors to exhibit behavior closer to an ideal switch. As a consequence, V_{gs} cannot be significantly reduced, even in weak inversion, thus exacerbating the headroom limitations.

Third, mismatch increases at cryogenic temperatures [25], [26]. Due to the steep SS, the impact of V_{th} mismatch on the drain current is more significant. DEM techniques will be employed to mitigate and investigate these effects.

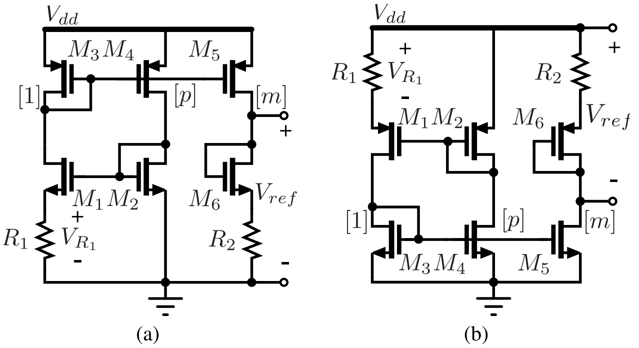


Fig. 1. Simplified schematic of the proposed CMOS voltage references with core devices $M_{1,2,6}$: (a) NMOS as core devices and (b) PMOS as core devices when the bulk of $M_{1,2,6}$ is connected to their source, and, alternatively, DTMOS as core devices when the bulk of $M_{1,2,6}$ is connected to their gate.

Finally, the resistors that are required for most references also suffer from a temperature dependence. To minimize these effects, n-type unsilicided poly resistors will be used, which vary less than 5% over temperature [28]. Furthermore, the reference voltage will mostly be set by a ratio of resistors, hence making it less vulnerable to changes in absolute resistance.

III. Circuit Design

A. Working Principle

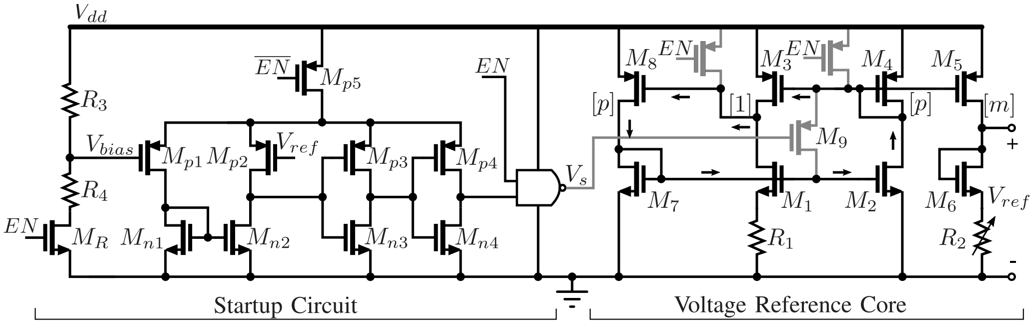
A MOS transistor operating in weak inversion can emulate the exponential I–V characteristic of a BJT that is required for classical bandgap references. The drain current I_d of a MOS transistor is then given as

$$I_d = \mu C_{\text{ox}} \frac{W}{L}(n-1)V_T^2 \exp\left(\frac{V_{\text{gs}} - V_{\text{th}}}{nV_T}\right) \tag{1}$$

where μ is the mobility, C_{ox} is the oxide capacitance per unit area, W and L are the width and length, respectively, n is the nonideality factor, and V_T is the thermal voltage. Looking at Fig. 1(a), and assuming M_1 and M_2 are in weak inversion and have nominally equal size, the voltage V_{R_1} across R_1 can be computed as

$$V_{R_1} = V_{\text{gs2}} - V_{\text{gs1}} = nV_T \ln(p) = n\frac{kT}{q}\ln p \tag{2}$$

where $V_{\text{gs}1,2}$ is the gate–source voltage of $M_{1,2}$, and p is the ratio of current densities between M_2 and M_1 set by the 1:p gain of the current mirror M_3–M_4. Note that V_{R_1} is a PTAT voltage. Due to the source of M_1 and M_2 being freely available (unlike the collector in parasitic pnp BJTs), V_{R_1} can be generated without using the typically adopted operational amplifier (e.g., in [9]), resulting in lower power consumption, higher accuracy, and improved reliability under unexpected environmental conditions. Resistor R_1 converts V_{R_1} into a current (as in [29]), which is mirrored into the series connection of M_6 and R_2 using M_3 and M_5, hence the voltage across R_2 is a scaled version of V_{R_1}. A corresponding CTAT voltage is generated from the gate–source voltage of



Core Device	NMOS	PMOS	DTMOS
R_1	165 kΩ	194 kΩ	164 kΩ
R_2	92 kΩ	109 kΩ	94 kΩ
R_3	308 kΩ	92 kΩ	92 kΩ
R_4	92 kΩ	308 kΩ	308 kΩ
$M_{1,2,6,7}$	{8,8,8,8}×15 µm/2 µm		
$M_{3,4,5,8}$	$\{1,p,m,p\}\times$ 1 µm/7.5 µm	$\{1,p,m,p\}\times$ 1 µm/20 µm	

Fig. 2. Schematic and sizing of the proposed architecture based on NMOS core transistors, where $p = 10$ and $m = 5$. All transistors are low-Vt (LVT) devices, except for M_R and $M_{p1,2}$ (standard Vt, SVT). No stacked devices were needed to obtain the desired transistor channel length. The bulk of $M_{1,2,6,7}$ is connected to the ground. The arrows indicate the main feedback loop. Resistors are implemented as unsalicided n-poly resistors. The startup and enable transistors are depicted in gray. A dual architecture was also implemented with PMOS and DTMOS as core devices.

M_6, provided that M_6 is also in weak inversion. The reference voltage V_{ref} is then given as

$$\begin{aligned} V_{\text{ref}} &= m\frac{R_2}{R_1}\cdot V_{R_1} + V_{\text{gs}6} \\ &= n\cdot m\frac{R_2}{R_1}\frac{kT}{q}\ln p + V_{\text{th}} + n\frac{kT}{q}\ln\left[\frac{I_{d6}}{\mu C_{\text{ox}}\frac{W}{L}(n-1)V_T^2}\right] \end{aligned} \tag{3}$$

where m is the gain of the current mirror $M_{3,5}$, and I_{d6} is the drain current of M_6. By appropriately choosing $mR_2/R_1\cdot\ln(p)$, the temperature coefficient (TC) of the PTAT component can be scaled to obtain a first-order temperature-independent reference voltage V_{ref}, approximately equal to the threshold voltage V_{th}. Since V_{gs} of a MOS transistor is typically lower than V_{be} of a BJT, MOS-based architectures do not necessarily require low-voltage techniques to implement sub-1-V references, unlike traditional bandgap references. Fig. 1(b) shows the dual-circuit implemented with PMOS as core devices and NMOS as current sources, resulting in V_{R_1} and V_{ref} now being referred to V_{dd}. By placing the core PMOS transistors in separate n-wells, their bulk can either be connected to their source to avoid the body effect or to their gate to create a reference based on DTMOS transistors [15]. Compared to PMOS transistors, (P-)DTMOS transistors require a lower V_{gs}, have a nonideality factor n closer to unity, and, at least at RT, exhibit lower process variations [15], [18], reducing the minimum V_{dd} and improving linearity and variation of V_{ref}.

B. Proposed Architecture

A drawback of the circuit in Fig. 1 is the limited supply rejection due to the noncascoded current sources. Via the finite output impedance, the difference in the drain–source voltage between M_3 and M_4, $\Delta V_{\text{ds}} = V_{\text{ds3}} - V_{\text{ds4}}$, translates into an error in the 1:p current ratio. Furthermore, ΔV_{ds} depends linearly on V_{dd}, thereby limiting the supply rejection. However, inserting cascodes in this architecture is nontrivial due to the required biasing and the limited headroom. A 5× change in absolute current is expected (due to the current being set by V_{R_1}/R_1), which is likely to bring the cascodes from strong into weak inversion. Due to the lack of accurate cryogenic device models, reliably designing bias networks dealing with such widely shifting operating points is challenging. Moreover, using an operational amplifier (opamp) to keep V_{ds3} and V_{ds4} equal is challenging since the required input common-mode of such an opamp (equal to V_{gs2}) would not leave sufficient headroom to reliably implement the opamp, especially in the absence of accurate device models. As current-mode voltage references typically need an opamp with similar requirements [30], current-mode references are not suitable for the target wide-temperature-range low-voltage applications.

As a solution, the proposed architecture in Fig. 2 employs an additional feedback branch to keep the drain voltage $V_{\text{d}3,4}$ of $M_{3,4}$ at the same potential, inspired by [31], but now further reducing the required headroom. The transistor M_7 (M_8) is a copy of M_2 (M_4). Since $V_{\text{gs7}} = V_{\text{gs2}}$, M_7 and M_2 carry equal currents, resulting in $V_{\text{gs8}} = V_{\text{gs4}}$ and thus $V_{\text{ds3}} = V_{\text{gs8}} = V_{\text{gs4}} = V_{\text{ds4}}$, which is independent of V_{dd} and hence reduces the supply sensitivity. The proposed architecture (Fig. 2) ensures a much better matching of V_{ds3} and V_{ds4} than the simplified architecture (Fig. 1), showing a simulated sensitivity to supply variations of the difference $V_{\text{ds3}} - V_{\text{ds4}}$ of only −64 mV/V (Fig. 2) versus −960 mV/V (Fig. 1). Simulations then show that the supply sensitivity is now limited by the limited impedance in the output branch. Similar to the simplified architecture in Fig. 1, also PMOS and DTMOS flavors of the proposed architecture have been implemented, where all voltages are referred to V_{dd}. Adding the feedback branch also affects the loop-gain in this architecture, thus potentially impacting stability. The simplified architecture [Fig. 1] has a loop-gain equal to $A_{\text{simp}} \approx (\text{gm}_4/\text{gm}_2)\cdot(\text{Gm}_1/\text{gm}_3)$, set by the gain of the two gm/gm amplifiers formed by M_4 and M_2, and M_1 and M_3, where $\text{Gm}_1 = \text{gm}_1/(1+\text{gm}_1 R_1)$ is the equivalent transconductance of the source-degenerated M_1, and gm_i the transconductance of M_i. Since $\text{Gm}_1 < \text{gm}_1$, $A_{\text{simp}} < \text{gm}_1\cdot\text{gm}_4/(\text{gm}_2\cdot\text{gm}_3) = 1/p\cdot p = 1$, the loop gain is positive and below unity ($A_{\text{simp}} = 0.4$ for the simplified NMOS architecture), and hence the circuit is stable. For the proposed architecture, the gain from the feedback loop equals $A_{\text{fb}} = -\text{gm}_8/\text{gm}_7$, noting that M_8 and M_7 form a gm/gm amplifier. Effectively, this can be modeled by increasing Gm_1 to $A_{\text{fb}}\cdot\text{Gm}_1$. The gain of the loop can now be expressed as $A_{\text{prop}} = (\text{gm}_3/(A_{\text{fb}}\cdot\text{Gm}_1))\cdot\text{gm}_2/\text{gm}_4$, which can be rewritten as $A_{\text{prop}} = A_{\text{simp}}^{-1}\cdot\text{gm}_7/\text{gm}_8 \approx -9$. Note that the direction of the loop is now opposite to the direction as in Fig. 1. Since M_7 and M_8 carry the same current, with M_7 in weak inversion,

$\mathrm{gm}_7 > \mathrm{gm}_8$, and therefore $A_{\mathrm{prop}} < -A_{\mathrm{simp}}^{-1} < -1$, and the circuit is stable.

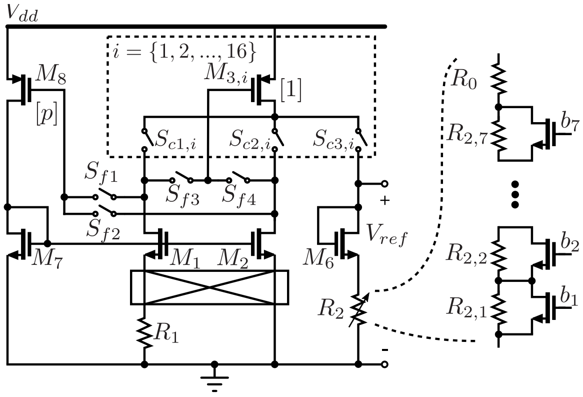


Fig. 3. Schematic showing the proposed reference implemented with core-transistor DEM (on $M_{1,2}$), current-source DEM (on $M_{3,i}$), and a resistive trimming network (on R_2). The 16 unit current sources from $M_{3,4,5}$ in Fig. 2 have been combined into the transistor indicated as $M_{3,1-16}$. Each of the units in $M_{3,1-16}$ can be uniquely configured to be connected to either the drain of M_1, M_2, or M_6. The bulk of $M_{1,2,6,7}$ is connected to the ground. The chopper, cascode, and trimming switches are SVT devices, and S_{f1-4} are LVT devices.

The sizing of the proposed architecture is shown in the table in Fig. 2. The sizing process starts by finding the current density range in which the core transistors are in weak inversion. This range, divided by the expected change in (PTAT-)current (due to the temperature change) determines the maximum current density ratio p. Having a larger p reduces the required scaling of V_{R_1} and therefore reduces error propagation from V_{R_1} to V_{ref}. The available cryogenic device characterization data shows that when devices are in weak inversion at 300 K, the devices can be assumed to be in weak inversion also at cryogenic temperatures [12]. Moreover, the PTAT nature of the bias current ensures that the current at cryogenic temperatures is fundamentally lower than at 300 K. As a next step, the absolute currents can be set based on leakage considerations, to ensure that the leakage currents, such as the gate leakage, are negligible with respect to the bias currents. This current can be defined using R_1 according to (2). In this design, the current at 300 K equals 425 nA to limit the effect of leakage. Given that the core transistors ($M_{1,2,6,7}$) need to be in weak inversion, the minimum current will set their aspect ratio. To avoid the current sources ($M_{3,4,5,8}$) entering weak inversion when current decreases at cryogenic temperatures, thereby compromising their matching [25], they must be biased far into strong inversion, hence the long channel length. The remaining m and R_2 can be set based on the scaling factor required for V_{R_1} [see (2)], where there is a tradeoff between power (higher m) and area (higher R_2). Due to the scaling factor being dependent on a ratio of resistors, the scaling factor $m \cdot R_2/R_1$ will be independent of the resistor TC.

The left part of Fig. 2 shows the implementation of the startup network. When the reference is in the OFF-state and no current is flowing, the gate–source voltage $V_{\mathrm{gs}} = 0$ for all transistors, and $V_{\mathrm{ref}} = 0$. A comparator ($M_{p1,2}$ and $M_{n1,2}$) senses whether the circuit is on ($V_{\mathrm{ref}} \approx 500$ mV) or off ($V_{\mathrm{ref}} = 0$) by comparing V_{ref} to $V_{\mathrm{bias}} = 0.23 \cdot V_{\mathrm{dd}} \approx 250$ mV ($\approx V_{\mathrm{ref}}/2$). Two cascaded inverters ensure that the comparator output is reconstructed to full logic levels. Although a basic digital inverter may be employed to efficiently detect the reference being in the OFF-state, the target reference voltage is close to the midsupply and hence to the threshold of the digital logic, thus affecting the PVT robustness of digital-based detectors. Using the comparator avoids such an issue and improves the startup's robustness. In the OFF-state, the startup transistor M_9 is enabled, forcing a current to flow in the reference. After startup is detected by the comparator, M_9 is disabled again. For characterization purposes, an enable signal EN was added to allow turning off the reference, startup circuit, and resistive divider. Measurements (see Section IV) showed that the startup network in Fig. 2 is not effective below 60 K. The low V_{dd} will limit $V_{\mathrm{gs9}} + V_{\mathrm{gs4}}$ to 1.1 V, causing those transistors to be either off, or too far in subthreshold due to the high threshold voltage at low temperatures (about 600 mV for PMOS [26]). In the second batch, the startup transistor was modified into NMOS with the drain connected to the gate of M_3 and M_4, and the source connected to ground. This startup is also not yet fully reliable, as it does not guarantee the startup of the feedback branch. For future designs, it is recommended to connect the source of the (NMOS) startup transistor to the ground, and the drain to the gate of M_8. This will ensure the startup of the feedback loop, which in turn starts up the rest of the circuit, as proven in a different test chip (not shown in this work).

C. Trimming

By making R_2 tunable, the PTAT term in V_{ref} in (3) can be scaled. Consequently, all errors resulting in a PTAT error in V_{ref}, such as a mismatch in the ratio R_2/R_1, can be compensated for by trimming R_2. To allow for this, R_2 has been implemented as a fixed resistor R_0, in series with a 7-bit, binary weighted resistor ladder, as depicted in Fig. 3. A series structure is chosen to optimize the required area. To circumvent the switch limitations mentioned in Section II, R_2 is not placed at the drain of M_6, but at its source. The transistors switching $R_{2,i}$ thus have a source voltage ranging from ground to <70 mV, allowing for sufficient overdrive. In case R_2 and M_6 were interchanged, a voltage of roughly 450 mV would be on the source of the switches at cryogenic temperatures. The smaller resistors are then arranged to be closer to the ground to minimize the switches' source voltage. The switches were sized to optimize their ON/OFF-resistance by taking into account their different source voltages. The simulated worst case error due to the nonzero ON-resistance is limited to below 700 μV (or 5 ppm/K in terms of TC).

D. Dynamic Element Matching

Any mismatch in the current mirrors will affect the 1:p:m mirror ratio and therefore the accuracy of V_{ref}. By applying DEM on the current sources, this error can be removed. Given that $p = 10$ and $m = 5$, it is a natural choice to implement 16 unit current sources. As confirmed by the simulations, any mismatch in the feedback branch translates into a mismatch between V_{ds3} and V_{ds4}, which is negligible with respect to the residual error after applying DEM. Fig. 3

shows the implementation of the 16 unit current sources, each having three switches ($S_{c1,i}$–$S_{c3,i}$) that can be individually and statically controlled by an on-chip SPI module, allowing the current to be directed to any of the branches. The switches are implemented as PMOS transistors, which can be opened by applying V_{dd} to their gate. To close a switch, 150 mV is applied (via an external bias source) to the gate of the switch. By using 150 mV instead of ground, the supply rejection of the circuit could be optimized by using the switch as a cascode. As their source is at 1.1-V_{ds}, sufficient overdrive can be guaranteed at cryogenic temperatures. In the first phase, $M_{3,1}$ is connected to the drain of M_1, $M_{3,2-11}$ to the drain of M_2, and $M_{3,12-16}$ to the drain of M_6. In the next phase, this will be $M_{3,2}$, $M_{3,3-12}$, and $M_{3,13-16}$ and $M_{3,1}$, respectively. After a total of 16 phases, M_1 and M_2 are interchanged with the chopping switches, and the procedure is repeated, yielding 32 phases. As the branch with only one unit current source is the dominant source of variation, each of the 16 unit current sources will now be connected to this branch once every 16 phases. Behavioral simulations with Spectre and MATLAB show that the statistical error in p and m is around 2.8% at −40 °C before DEM and is expected to reduce about two orders of magnitude to below 0.025%.

Mismatch in the core transistors M_1 and M_2 affects the reference voltage, as any mismatch-induced difference between V_{gs1} and V_{gs2} directly appears in V_{R_1}, which is then amplified to V_{ref} by $m \cdot R_2/R_1$. Note that since the TC of the V_{gs} of an MOS (below $-0.9\ \mathrm{mV\,K^{-1}}$ in our case) is smaller than for a BJT (typically $-2\ \mathrm{mV\,K^{-1}}$), a lower value for $m \cdot R_2/R_1$ can be used compared to BJT-based references (for the same p), which reduces the amplification of error sources associated with $M_{1,2,3,4}$ and R_1 to the output. This is a beneficial property of MOS-based references, especially for uncompensated error sources. In case there is both a threshold voltage- and beta-mismatch between the two core transistors, V_{R_1} can be computed as

$$\begin{aligned} V_{R_1} &= V_{\mathrm{gs2}} - V_{\mathrm{gs1}} \\ &= \underbrace{n\frac{kT}{q}\ln p}_{\text{PTAT Term}} + \underbrace{\left[(V_{\mathrm{th2}} - V_{\mathrm{th1}}) + n\frac{kT}{q}\ln\left(\frac{\beta_1}{\beta_2}\right)\right]}_{\text{Mismatch Term}} \end{aligned} \tag{4}$$

where $\beta_{1,2}$ and $V_{\mathrm{th1,2}}$ are the beta-factor and threshold voltage of $M_{1,2}$, respectively. By exchanging M_1 and M_2 and averaging V_{ref}, the mismatch is removed. Ignoring the body effect (see following subsection), the residual V_{th} mismatch is below 0.2 mV. The implementation of the required switches is shown in Fig. 3. A chopper using NMOS switches at the source of the core transistors can be made sufficiently low impedance, as V_{R_1} is below 100 mV. The chopper at the drain of the core transistors can be conveniently combined with the already present cascode switches $S_{c1,i}$–$S_{c3,i}$. Pass-gate S_{f1-4} ensures proper feedback is maintained when interchanging M_1 and M_2. Since these pass-gates are in series with a gate (with a gate leakage below 5 nA at 300 K), this would only require $V_{\mathrm{gs}} > V_{\mathrm{th}}$. Since V_{th} is larger for PMOS than for NMOS in this process, this requirement is always met. At 4.2 K, the ON-resistance is estimated to be below 12.5 kΩ.

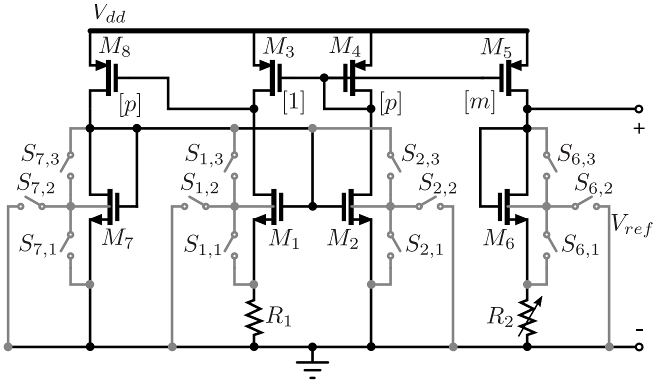


Fig. 4. Proposed architecture implemented with NMOS (in separate deep n-wells) as core devices, where the bulk of the core devices can be connected to either source (ϕ_1), ground (ϕ_2), or gate (ϕ_3). Switches $S_{\{1,2,6,7\},3}$ are LVT devices, all other switches are SVT devices.

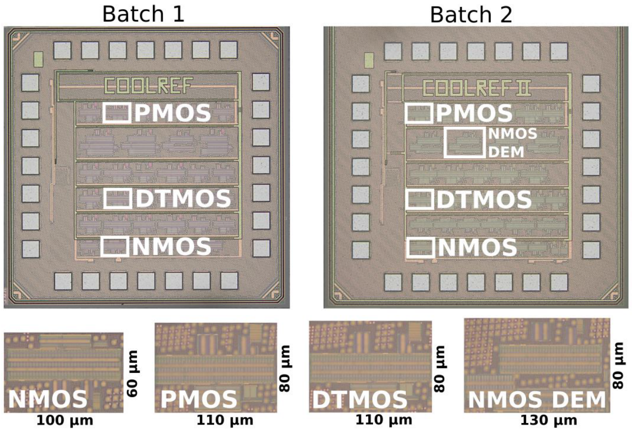


Fig. 5. Die micrographs for both batches. Insets show instances of the proposed architecture in Fig. 2 with NMOS, PMOS, and DTMOS as core device, as well as the architecture in Fig. 3 (NMOS DEM).

E. Configurable Bulk

Whereas DEM can be used to remove statistical mismatch between M_1 and M_2, it cannot remove systematic mismatch due to the body effect, since M_1 and M_2 have a different source potential. Next to M_1, also M_6 suffers from the body effect due to the drop on R_2. Interchanging M_6 and R_2 would solve this problem, but it also makes it challenging to implement a tunable R_2 (see Section II). Using the available deep n-well, the architecture in Fig. 4 has been implemented, where the NMOS core transistors are all placed in isolated p-wells. Using the switches, the potential of the p-wells can be connected to either the source (ϕ_1) or ground (ϕ_2), allowing to assess the effects of the body effect on the PTAT, CTAT, and the reference voltage. For PMOS references, source and bulk are always shorted. Finally, the bulk can also be connected to the gate (ϕ_3), essentially creating an N-DTMOS configuration. As the gate voltage in N-DTMOS configuration is expected to be below 450 mV, leaving 650 mV headroom, these switches can also be implemented with NMOS transistors without the risk of insufficient headroom.

IV. Measurement Results

Two batches have been fabricated in a commercial 40-nm bulk CMOS process (Fig. 5), similar to the nanometer

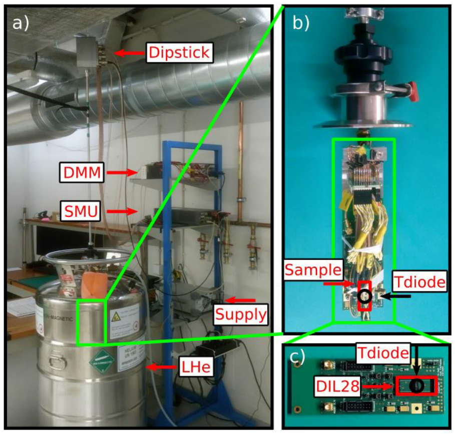
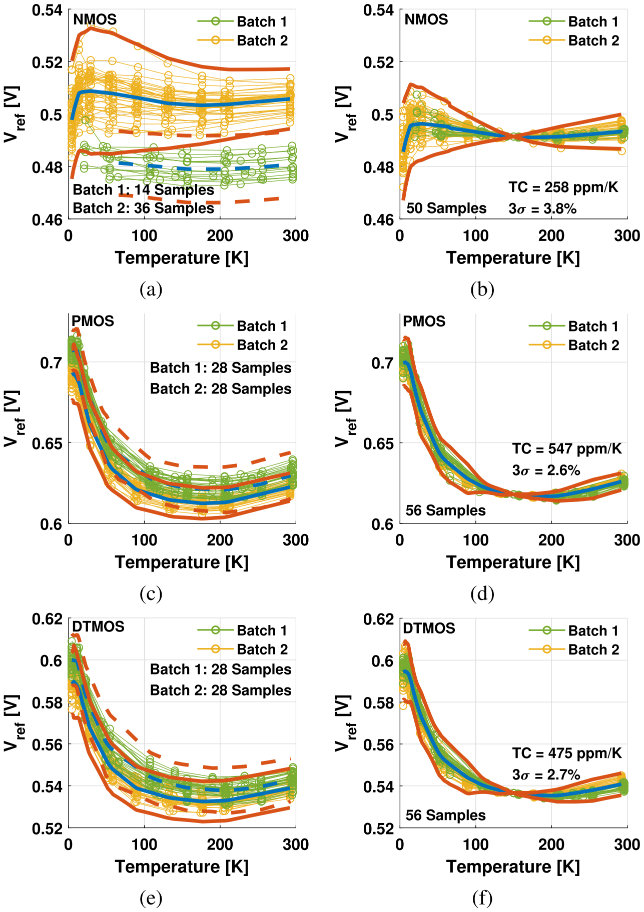


Fig. 6. Cryogenic measurement setup, showing (a) the dipstick in LHe, and measurement equipment, (b) PCB mounted in dipstick, and (c) PCB itself.



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 7. (a), (c), and (e) Measured V_{ref} from the proposed references implemented with either NMOS, PMOS, or DTMOS as core device, without trim, and (b), (d), and (f) after applying a single-point scaling trim in MATLAB at 150 K. The mean and $\pm 3\sigma$ are indicated using the red and blue lines, respectively, where the dashed lines are for batch 1 and the solid lines for batch 2.

processes commonly used in cryo-CMOS quantum computing applications [3], [32], and packaged in ceramic DIP packages. Characterization was performed using a dipstick in a Dewar with liquid helium (LHe) (Fig. 6). Due to the high input impedance (>100 GΩ) from the multimeter (Keithley 2002), no buffering for the references was needed. Seven chips from the first batch (two NMOS-, four PMOS-, and four DTMOS instances per chip) were measured, and four chips from the second batch (nine NMOS-, seven PMOS-, and seven DTMOS instances per chip). All architectures are exactly the same in both batches, except for the slight modification in the startup network in the NMOS-based architecture. The NMOS architectures with DEM and configurable bulk connection are only present in the second batch. Data for all presented plots can be found in [33].

A. Reference Voltage-NMOS

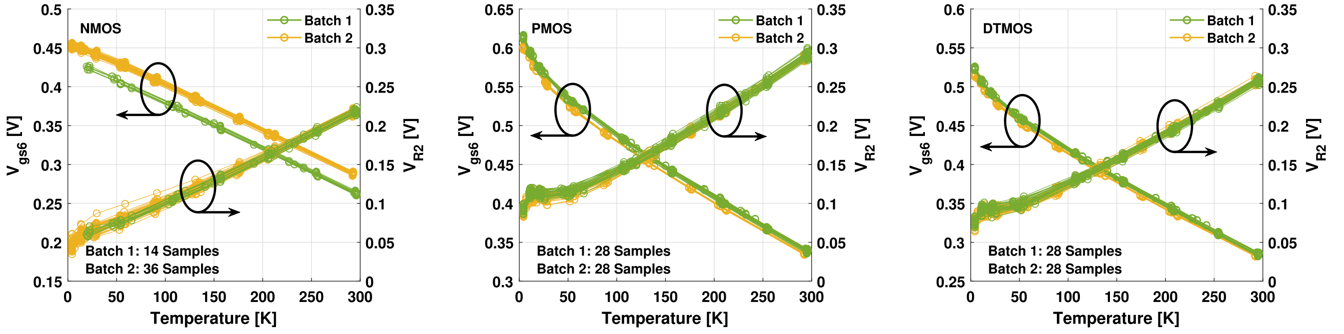
Fig. 7(a) shows V_{ref} versus temperature of the NMOS-based architecture for both batches. The value of R_2 is set to optimize the TC determined from the box method over the temperature range from 4 to 300 K. The same value for R_2 is used for all instances in both batches. Applying a single-point scaling trim in MATLAB at 150 K to both batches, where a temperature-independent scaling factor is applied postmeasurement to the reference voltage, such that at 150 K all references coincide, yields the curves in Fig. 7(b). A TC of 258 ppm/K and spread of 3.8% (3σ) is achieved, where the TC is computed using the box method, in which the box fits all curves from both batches. It is clearly visible that the box size, and therefore the TC is dominated by the variation at cryogenic temperatures, attributed to the more severe effects of mismatch at cryogenic temperatures [25], [26], and the systematic nonlinearity below 20 K. Before trimming, a TC and 3σ spread of 141 ppm/K and 2.7% for batch 1, and 348 ppm/K and 4.8% for batch 2 are achieved. Due to the startup issue, batch 1 has a temperature range limited to above 60 K (see Section III), hence explaining the performance difference between batches 1 and 2.

Next to V_{ref}, the PTAT voltage was characterized by measuring the voltage V_{R_2} across the output resistor R_2, in turn allowing also the CTAT voltage V_{gs6} to be computed using (3). As can be seen in Fig. 8(a), the CTAT voltage V_{gs6} shows an offset between the two batches, but the PTAT voltage V_{R_2} overlaps. This low susceptibility to process corners is attributed to the spread of V_{R_2} in (3) mainly depending on the mismatch rather than spread (between M_1/M_2 and R_1/R_2 in Fig. 2), in addition to any spread in the nonideality factor n.

The CTAT voltage V_{gs6} in (3) is directly affected by spread in V_{th}, R_1 (via I_{d6}), μ, and n. Given that V_{th} is outside the logarithm, batch-to-batch spread in V_{th} will thus be the main source of offset in V_{ref} in Fig. 7(a) and the CTAT voltage in Fig. 8(a). This is also confirmed by corner simulations (about 60 mV change in V_{ref} and 80 mV in V_{th} between extreme corners). The saturation in V_{gs6} at low temperatures is caused by saturation in V_{th}, induced by the saturation in bulk Fermi potential [34], which has been previously observed [25], [35].

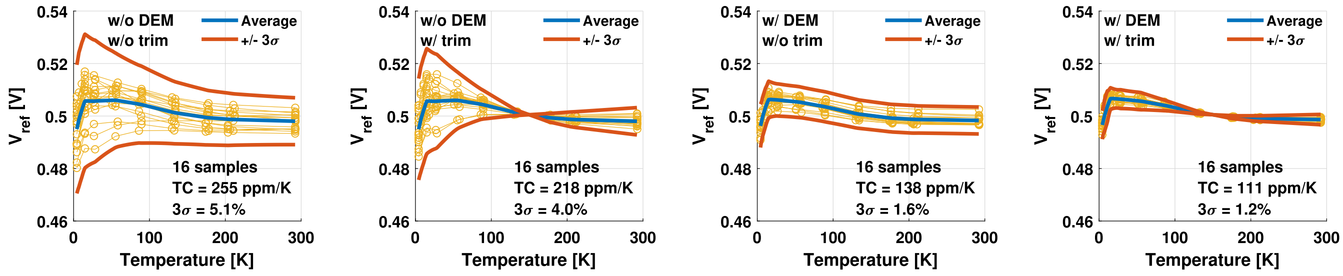
B. Reference Voltage-P/DTMOS

As can be seen from the measured reference voltage generated by the PMOS-(c) and DTMOS-based (e) references in Fig. 7, the reference voltage for the DTMOS-based reference is roughly 100 mV lower than for the PMOS-based references.



(a) (b) (c)

Fig. 8. PTAT-(V_{R_2}) and CTAT (V_{gs6}) voltage corresponding to the measured V_{ref} in Fig. 7(a), (c), and (e) for the proposed architecture, implemented with either (a) NMOS, (b) PMOS, or (c) DTMOS as core device. The setting of R_2 is the same for all curves of the same device flavor.



(a) (b) (c) (d)

Fig. 9. Measurements of the architecture in Fig. 3 showing V_{ref} (a) without any compensation, (b) a single-point scaling trim, (c) core transistor- and current source DEM, and (d) core transistor- and current source DEM, together with a single-point scaling trim. The setting of R_2 is the same in all plots.

This is caused by the lower threshold voltage of the DTMOS devices, resulting from the bulk of the DTMOS being at a potential lower than V_{dd}. Similar to Fig. 7(a), a small offset is present between the two batches, which is again mainly attributed to the spread in V_{th} between both batches and is well within the corner simulations (50 mV change in V_{ref} and 60 mV in V_{th} between extreme corners). In Fig. 7(d) and (f), the reference voltage is depicted after a single-point scaling trim at 150 K. The TC and 3σ spread are computed on all samples from both batches together. Contrary to the NMOS, it can be observed that both for PMOS and DTMOS, the TC is limited by the systematic nonlinearity below 100 K rather than statistical errors. In fact, the variation for PMOS and DTMOS is lower than for NMOS (2.6% and 2.7% versus 3.8%). Again, the 3σ spread is larger below 50 K, pointing to the mismatch at cryogenic temperatures as the dominating factor for the variation.

Similar to the NMOS, for the PMOS and DTMOS, V_{R_2} in Fig. 8(b) and (c) from both batches overlaps. Furthermore, an offset is present when comparing V_{gs6} from both batches. As observed in Fig. 7(c) and (e), the PMOS and DTMOS V_{ref} suffers from a large systematic nonlinearity. Based on Fig. 8(b) and (c), this can be traced back to both the PTAT and CTAT voltage. First, a saturation in V_{R_2} can be observed, which is fundamentally caused by a saturation in the SS [18], [36]. Second, V_{gs6} starts increasing below 50 K, which is attributed to the increase in PMOS V_{th} also observed in literature [34], although also a saturation in PMOS V_{th} has been reported [25]. Given that both the increase in V_{gs6} and saturation in V_{R_2} have the same sign, a significant systematic nonlinearity appears in V_{ref} below 100 K, which turns out to be the dominant error that sets the TC. Mostly for the P/DTMOS-based references, but also for the NMOS-based references, a strong nonlinearity in the V_{ref} below 20 K appears (Fig. 7), which can be traced back to the PTAT voltage V_{R_2}. A similar nonlinearity was observed in [18], where the data suggested the nonlinearity may depend on the operating region of the transistor. Using the model and data in [23], it was verified that the core transistors in the proposed architecture are in weak inversion for all temperatures, hence making it unlikely that the nonlinearity is caused by the core transistors being out of weak inversion below 20 K. Whereas the model in [23] can be used to investigate whether the devices are in weak inversion, numerical issues cause the model to be inconclusive about the physical origin of the nonlinearity.

C. Dynamic Element Matching

When DEM is not enabled [Fig. 9(a)], that is, for V_{ref} in the first DEM phase out of 32 phases, the circuit in Fig. 3 exhibits comparable TC (255 versus 348 ppm/K) and 3σ spread (5.1% versus 4.8%) as the second-batch NMOS V_{ref} in Fig. 7(a). The same holds when considering the single-point scaling trim as in Figs. 7(b) and 9(b) (218 versus 258 ppm/K, and 4.0% versus 3.8%). Enabling DEM on the current sources and the core transistors reduces the spread by up to 3× [Fig. 9(c)]. V_{ref} is now computed by taking the average of all 32 DEM phases. By only applying DEM (w/o trim) on the current sources, the variation reduces to 3.4%, and to 4.1% (w/o trim) if only applied on the core transistors.

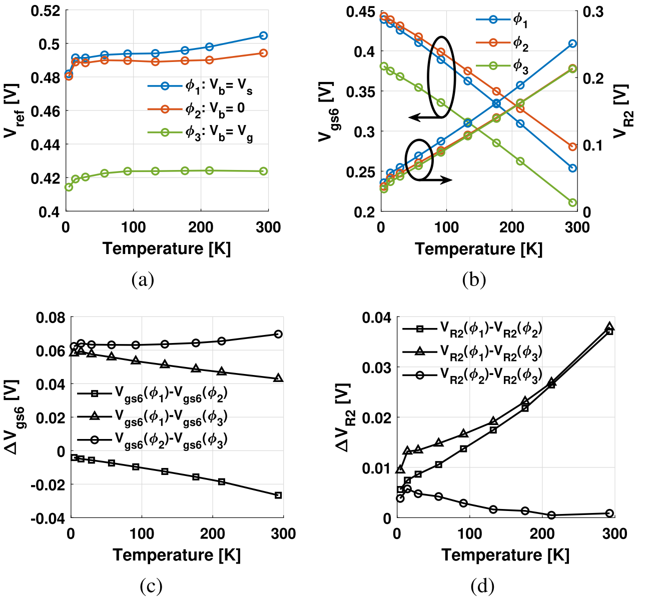


Fig. 10. Output voltage from a typical sample (a) V_{ref}, (b) V_{gs6} and V_{R_2}, and the differences between (c) V_{gs6} and (d) V_{R_2} in the three configurations for the circuit in Fig. 4. The setting of R_2 is the same for all curves.

Consequently, 3.8% can be attributed to the current sources, and 3.0% to the core transistors. DEM is particularly effective at cryogenic temperatures, as it prevents mismatch from the current sources and core transistors to be the dominant source of variation, which is expected in view of the increased mismatch in both weak- and strong inversion. By applying DEM and a scaling trim, the residual TC of 111 ppm/K in Fig. 9(d) is not anymore limited by the spread but by the systematic nonlinearity below 20 K. As such a nonlinearity does not benefit from mismatch-compensation techniques in the circuit core, its cause cannot be attributed to random or systematic mismatch effects. Because the current magnitude in the circuit significantly reduces at cryogenic temperatures (by approximately 5×), gate leakage could potentially induce nonlinearity. Simulations from −40 °C to 27 °C indicate a maximum gate leakage of about 5 nA, which would lead to an error in V_{R_2} at 4.2 K up to 4%. However, the lack of suitable cryogenic device models, and even the absence of cryogenic gate-leakage characterization data, prevents us from drawing a definitive conclusion.

D. Impact of Body Effect

The impact of body effect can be analyzed by observing the reference voltage V_{ref} [Fig. 10(a)] and the corresponding CTAT (V_{gs6}) and PTAT (V_{R_2}) components [Fig. 10(b)] when switching the core-transistor bulk in the circuit in Fig. 4 to their source, to ground, and their gates, respectively. Looking at Fig. 4 and neglecting the statistical mismatch between M_1 and M_2, V_{R_2} can be written as

$$V_{R_2} = m\frac{R_2}{R_1}\left(V_{\mathrm{gs2}} - V_{\mathrm{gs1}}\right) = m\frac{R_2}{R_1}\left(n\frac{kT}{q}\ln p + \Delta V_{\mathrm{th}}\right) \quad (5)$$

where $\Delta V_{\mathrm{th}} = V_{\mathrm{th2}} - V_{\mathrm{th1}}$ is due to the body effect. When the bulk of each of the NMOS is connected to the source, V_{th} of all core devices is nominally equal to $V_{\mathrm{th0}} = V_{\mathrm{th}}|_{V_{\mathrm{bs}}=0}$, and $\Delta V_{\mathrm{th}} = 0$. When the bulk is connected to ground, M_7 and M_2 have the same $V_{\mathrm{th}} = V_{\mathrm{th0}}$, but since $V_{\mathrm{bs1}}, V_{\mathrm{bs6}} < 0$, $V_{\mathrm{th1}}, V_{\mathrm{th6}} > V_{\mathrm{th0}}$. As $\Delta V_{\mathrm{th}} < 0$ in this case, V_{R_2} is lower than for $V_{\mathrm{bs}} = 0$, as shown in [Fig. 10(b)]. Due to the lower PTAT voltage, the bias current reduces (since R_1 is fixed), and also V_{gs6} is expected to reduce. Given that $V_{b6} = 0$, the V_{th6} increases, which has a stronger effect on V_{gs6} than the reduced bias current, hence explaining why V_{gs6} is higher than for $V_{\mathrm{bs6}} = 0$. As the source voltage of M_1 and M_6 is a PTAT voltage, the circuit with $V_b = 0$ (ϕ_2) converges to the configuration with $V_{\mathrm{bs}} = 0$ (ϕ_1) when the temperature approaches absolute zero. As a result, both V_{gs6} and V_{R_2} converge at low temperatures in this case, which is indeed observed in Fig. 10(b) as well. By computing $\Delta V_{\mathrm{th}} = \Delta V_{R_2}/(mR_2/R_1)$, ΔV_{th} can now also be computed to be −13 and –2.0 mV at 300 and 4.2 K, respectively, corresponding to a body-effect coefficient of 0.17 V/V and 0.15 V/V. Moreover, the behavior of $V_{R_2}(\phi_1) - V_{R_2}(\phi_2)$ shows that also ΔV_{th} is essentially a PTAT voltage, implying that the body effect can be mitigated by applying a PTAT trim. By trimming of R_2 and V_{ref} for $V_b = V_s$ and $V_b = 0$ can be made equal up to 0.6 mV, thereby making it not the limiting factor for the TC. It is, therefore, not required to use a deep n-well process to achieve a lower TC.

When the gate is connected to the bulk (ϕ_3, $V_b = V_g$), we form an N-DTMOS device. As the bulk–source voltage $V_{\mathrm{bs1}} < V_{\mathrm{bs2}}$, also $V_{\mathrm{th1}} > V_{\mathrm{th2}}$ and thus $\Delta V_{\mathrm{th}} < 0$, implying that V_{R_2} in ϕ_3 is lower than in ϕ_1, where $V_{\mathrm{bs}} = 0$. Due to both the reduced bias current (since V_{R_1} is smaller and R_1 is fixed) and the reduced V_{th} of M_6, V_{gs6} for ϕ_3 is therefore smaller than for $V_{\mathrm{bs}} = 0$ (ϕ_1). This reduction in V_{gs6} is mostly induced by the N-DTMOS configuration, which essentially lowers V_{th}. In terms of headroom, using the deep n-well to form an N-DTMOS structure is thus beneficial for cryogenic low-voltage designs where headroom is a limiting factor. Note that because the nonlinearity in Fig. 10(a) is consistent over the bulk arrangements, it can be excluded that the systematic nonlinearity below 20 K in V_{ref} is caused by the body effect.

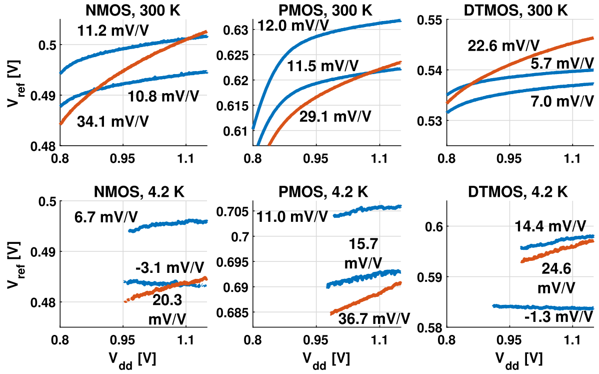


Fig. 11. Measured supply dependence of V_{ref} for two instances (blue curves) of the proposed architecture and one instance (red curves) of the simplified architecture, measured at 300 and 4.2 K.

E. Line Regulation and Power Consumption

To assess the effectiveness of the additional feedback loop in the proposed architecture in Fig. 2, the line regulation was

TABLE I

Performance Comparison

		This work			EDL 2009 [14]	SSCL 2018 [12]	SSCL 2020 [37]	JSSC 2021 [11]
Technology		40-nm CMOS			SiGe BiCMOS	40-nm CMOS	28-nm FDSOI	0.18-μm CMOS
Core Device		NMOS	PMOS	DTMOS	SiGe HBT	Thick-ox. DTMOS	Thick-ox. MOS	Si BJT
Temperature Range [K]		4.2-300			0.7-293	4.2-300	4.2-300	233-398
Supply Voltage [V]		0.96-1.1	0.99-1.1	0.98-1.1	3.3	1.8-3.3	1.2-1.8	1.8±10%
Power [μW]	300 K	13.7[a]	14.9[a]	15.1[a]	N.A.	368	15.8	31
	T_{min}	5.1[a]	8.2[a]	7.8[a]	130	132	13.9	N.A.
Active Area [mm^2]		0.006	0.009	0.009	N.A.	0.0004	0.041	0.38
Line Regulation [%/V]	300 K	2.2	2.0	1.3	N.A.	6.9	0.4	0.01
	T_{min}	1.3	2.6	2.7	N.A.	8.3	0.6	N.A.
Integrated noise at 300 K [μV$_{\text{rms}}$]		19[b]	9.9[b]	10.3[b]	N.A.	N.A.	N.A.	44[c]
TC (1-point trim) [ppm/K]		258(111[d])	547	475	160	833	1214	3.2-5.5
Spread (3σ) (1-point trim) [%]		3.8(1.2[d])	2.6	2.7	N.A.	5.3	N.A.	-0.02, 0.12
Samples [#]		50[e](16[d])	56	56	1	5	1	18

[a] In static operation, excluding external bias; [b] In 1-10 Hz; [c] In 0.01 Hz-2.5 Hz (estimated from [11]);
[d] For the architecture in Fig. 3 after averaging in MATLAB; [e] The 14 samples from batch 1 are working down to roughly 67 K.

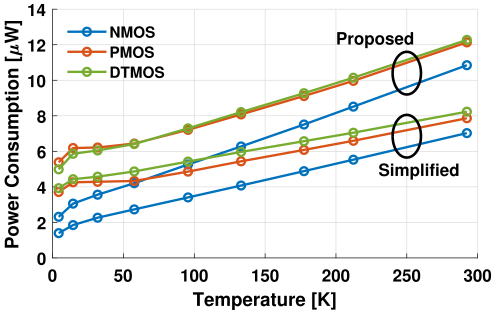


Fig. 12. Measured power consumption from a 1.1 V supply for the proposed architecture core (average of 9/7/7 samples for N/P/DTMOS) versus the simplified architecture (1 sample for N/P/DTMOS). This plot is excluding the 2.8 μA drawn by the resistive divider formed by R_3 and R_4, which varies less than 5% over temperature.

measured for both the proposed architecture and the simplified architecture. The line regulation has been computed using a first-order fit of V_{ref} for $V_{\text{dd}} \in \{1.05, 1.15\}$ V at 300 K and $V_{\text{dd}} \in \{1.0, 1.15\}$ V at 4.2 K. Datapoints for which the reference did not startup were discarded (mostly below 0.95 V). As can be seen in Fig. 11, the proposed architecture achieves better line regulation than the simplified architecture, demonstrating the effectiveness of the additional feedback branch. An important observation is that at 4.2 K, the reference is either on or off, and there is no smooth transition region as there is at 300 K. This effect is caused by the steeper SS at cryogenic temperatures, making the transistor behave closer to an ideal switch. In case there is not sufficient headroom available, the circuit will then fully turn off. Combined with the increased V_{th} at cryogenic temperatures, the references consistently need a higher minimum V_{dd} than at 300 K. Two instances even exhibit negative line regulation, which is likely caused by the vastly shifting operating point of the circuit (and thus the variation of the loop-gain) during the measurements, combined with the very low current levels, cryogenic device effects, and mismatch effects.

The measured power consumption is shown in Fig. 12, where the power consumption from the proposed architecture (Fig. 2) is about 1.5× higher than the simplified architecture (Fig. 1) due to the additional feedback branch. The microwatt power consumption is in line with the typically assumed power budget of roughly 1 m watt/qubit for quantum computing applications [3]. The absence of the typically adopted amplifier in the proposed architecture (as mentioned in Section III-A) allows for low power and low noise. However, as the DEM in this architecture is only static, $1/f$-noise is the dominant factor in terms of noise. A performance comparison with other works is presented in Table I.

V. Conclusion

Harsh-environment applications, such as quantum computing, require electronics to operate far below the standard temperature range. A family of voltage references is presented that can reliably operate from 300 down to 4.2 K from a sub-1-V supply. Prototypes fabricated in a commercial 40-nm CMOS process achieve a TC below 547 ppm/K and 3σ variation below 3.8% after a single-point trim over 56 samples from 2 batches. The adoption of a feedback-regulated architecture ensures a line regulation below 2.7%/V for sub-1-V operation. After applying DEM techniques, the TC and the spread can be reduced to 111 ppm/K and 1.2%, respectively, mainly limited by systematic nonlinearity below 20 K. When no deep n-well is employed, the body effect manifests itself mainly as a PTAT error and can, therefore, be easily removed with a PTAT trim. Furthermore, nonlinearity, core-transistor, and current-source mismatch have been experimentally analyzed. Thus, the proposed architectures reliably provide a PVT-robust reference voltage, allowing for use down to extremely low temperatures.

References

[1] K. Souri and K. A. A. Makinwa, *Energy-Efficient Smart Temperature Sensors in CMOS Technology*, 1st ed. Berlin, Germany: Springer, 2017.

[2] L. Enthoven, J. van Staveren, J. Gong, M. Babaie, and F. Sebastiano, "A 3V 15b 157 μW cryo-CMOS DAC for multiplexed spin-qubit biasing," in *Proc. IEEE Symp. VLSI Technol. Circuits*, Jun. 2022, pp. 228–229.

[3] G. Kiene et al., "A 1GS/s 6-to-8b 0.5 mW/qubit cryo-CMOS SAR ADC for quantum computing in 40 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 64, Feb. 2021, pp. 214–216.

[4] D. Andrade-Miceli et al., "Cryogenic low-drop-out regulators fully integrated with quantum dot array in 22-nm FD-SOI CMOS," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2021, pp. 635–637.

[5] W. Kuhn et al., "A microtransceiver for UHF proximity links including Mars surface-to-orbit applications," *Proc. IEEE*, vol. 95, no. 10, pp. 2019–2044, Oct. 2007.

[6] J. D. Cressler and H. A. Mantooth, *Extreme Environment Electronics*. Boca Raton, FL, USA: CRC Press, 2013.

[7] C. Enz, *Cryogenic Particle Detection* (Topics in Applied Physics). Berlin, Germany: Springer, 2005.

[8] P. A. 't Hart, M. Babaie, A. Vladimirescu, and F. Sebastiano, "Characterization and modeling of self-heating in nanometer bulk-CMOS at cryogenic temperatures," *IEEE J. Electron Devices Soc.*, vol. 9, pp. 891–901, 2021.

[9] G. Ge, C. Zhang, G. Hoogzaad, and K. A. A. Makinwa, "A single-trim CMOS bandgap reference with a 3σ inaccuracy of ±0.15% from −40 °C to 125 °C," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2693–2701, Nov. 2011.

[10] G. Maderbacher et al., "A digitally assisted single-point-calibration CMOS bandgap voltage reference with a 3σ inaccuracy of ±0.08% for fuel-gauge applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 1–3.

[11] J.-H. Boo et al., "A single-trim switched capacitor CMOS bandgap reference with a 3σ inaccuracy of ±0.02%, −0.12% for battery-monitoring applications," *IEEE J. Solid-State Circuits*, vol. 56, no. 4, pp. 1197–1206, Apr. 2021.

[12] H. Homulle, F. Sebastiano, and E. Charbon, "Deep-cryogenic voltage references in 40-nm CMOS," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 5, pp. 110–113, May 2018.

[13] Y. Yang, K. Das, A. Moini, and D. J. Reilly, "Cryo-CMOS band-gap reference circuits for quantum computing," 2019, *arXiv:1910.01217*.

[14] L. Najafizadeh et al., "Sub-1-K operation of SiGe transistors and circuits," *IEEE Electron Device Lett.*, vol. 30, no. 5, pp. 508–510, May 2009.

[15] A.-J. Annema, "Low-power bandgap references featuring DTMOSTs," *IEEE J. Solid-State Circuits*, vol. 34, no. 7, pp. 949–955, Jul. 1999.

[16] T. Ytterdal, "CMOS bandgap voltage reference circuit for supply voltages down to 0.6V," *Electronics Lett.*, vol. 39, pp. 1427–1428, Oct. 2003. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/el_20030937

[17] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenic temperatures," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 996–1006, 2018.

[18] H. Homulle, L. Song, E. Charbon, and F. Sebastiano, "The cryogenic temperature behavior of bipolar, MOS, and DTMOS transistors in standard CMOS," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 263–270, 2018.

[19] A. Beckers, F. Jazaeri, and C. Enz, "Characterization and modeling of 28-nm bulk CMOS technology down to 4.2 K," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 1007–1018, 2018.

[20] L. Varizat, G. Sou, M. Mansour, D. Alison, and A. Rhouni, "A low temperature 0.35 μm CMOS technology BSIM3.3 model for space instrumentation: Application to a voltage reference design," in *Proc. IEEE MetroAeroSpace*, Jun. 2017, pp. 74–78.

[21] I. M. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Trans. Circuits Syst. I, Fund. Theory Appl.*, vol. 48, no. 7, pp. 876–884, Jul. 2001.

[22] J. Jiang, W. Shu, and J. S. Chang, "A 5.6 ppm/°C temperature coefficient, 87-dB PSRR, sub-1-V voltage reference in 65-nm CMOS exploiting the zero-temperature-coefficient point," *IEEE J. Solid-State Circuits*, vol. 52, no. 3, pp. 623–633, Mar. 2017.

[23] P. A. 't Hart, J. van Staveren, F. Sebastiano, J. Xu, D. E. Root, and M. Babaie, "Artificial neural network modelling for cryo-CMOS devices," in *Proc. IEEE 14th Workshop Low Temp. Electron. (WOLTE)*, Apr. 2021, pp. 1–4.

[24] J. van Staveren et al., "Voltage references for the ultra-wide temperature range from 4.2K to 300K in 40-nm CMOS," in *Proc. IEEE 45th Eur. Solid State Circuits Conf. (ESSCIRC)*, 2019, pp. 37–40.

[25] P. A. 't Hart, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Subthreshold mismatch in nanometer CMOS at cryogenic temperatures," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 797–806, 2020.

[26] P. A. 't Hart, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and modeling of mismatch in cryo-CMOS," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 263–273, 2020.

[27] J. van Dijk et al., "Cryo-CMOS for analog/mixed-signal circuits and systems," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Mar. 2020, pp. 1–8.

[28] P. A. 't Hart, T. Huizinga, M. Babaie, A. Vladimirescu, and F. Sebastiano, "Integrated cryo-CMOS temperature sensors for quantum control ICs," in *Proc. IEEE 15th Workshop Low Temp. Electron. (WOLTE)*, Jun. 2022, pp. 1–4.

[29] E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operations," *IEEE J. Solid-State Circuits*, vol. SSC-12, no. 3, pp. 224–231, Jun. 1977.

[30] H. Banba et al., "A CMOS bandgap reference circuit with sub-1-V operation," *IEEE J. Solid-State Circuits*, vol. 34, no. 5, pp. 670–674, May 1999.

[31] Y.-H. Lam and W.-H. Ki, "CMOS bandgap references with self-biased symmetrically matched current-voltage mirror and extension of sub-1-V design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 6, pp. 857–865, Jun. 2010.

[32] B. Prabowo et al., "A 6-to-8 GHz 0.17mW/qubit cryo-CMOS receiver for multiple spin qubit readout in 40 nm CMOS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2021, pp. 212–214.

[33] J. van Staveren, "Data corresponding to: Cryo-CMOS voltage references for the ultra-wide temperature range from 300 K down to 4.2 K," 4TU ResearchData, Delft, The Netherland, 2024. [Online]. Available: https://doi.org/10.4121/61411994-3527-4055-8283-2d393becadab

[34] A. Beckers, F. Jazaeri, A. Grill, S. Narasimhamoorthy, B. Parvais, and C. Enz, "Physical model of low-temperature to cryogenic threshold voltage in MOSFETs," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 780–788, 2020.

[35] N. C. Dao, A. E. Kass, M. R. Azghadi, C. T. Jin, J. Scott, and P. H. W. Leong, "An enhanced MOSFET threshold voltage model for the 6–300 K temperature range," *Microelectron. Rel.*, vol. 69, pp. 36–39, Feb. 2017.

[36] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical limit of low temperature subthreshold swing in field-effect transistors," *IEEE Electron Device Lett.*, vol. 41, no. 2, pp. 276–279, Feb. 2020.

[37] Y. Yang, K. Das, A. Moini, and D. J. Reilly, "A cryo-CMOS voltage reference in 28-nm FDSOI," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 186–189, 2020.

Job van Staveren received the B.Sc. degree in electrical engineering and the M.Sc. degree in electrical engineering, with a specialization in microelectronics from Delft University of Technology, Delft, The Netherlands, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering.

His research interests include analog/mixed-signal design, with a particular focus on design for the cryogenic temperature range and measurement of CMOS-electronics at cryogenic temperatures.

Pinakin M. Padalia was born in India, in 1993. He received the B.E. degree (Hons.) in electrical and electronics engineering from the Birla Institute of Technology and Science (BITS), Pilani, India, in 2017, and the M.S. degree in electrical engineering with a specialization in microelectronics from Delft University of Technology, Delft, The Netherlands, in 2019.

From 2019 to 2020, he was a Research Assistant at QuTech, Delft, with the cryogenic electronics group focusing on wide temperature range voltage and current references. Since 2020, he has been back in India and joined a young company, QPiAI, Bengaluru, India, as a Quantum Circuit Designer for developing hardware for controlling qubits. He is a Senior Engineer with the Analog and Communication Group, Renesas Electronics, Bengaluru. His current research interests include analog mixed-signal IC design and systems.

Mr. Padalia was a recipient of the Junior Research Fellowship from the Indian Academia of Sciences in 2016.

Edoardo Charbon (Fellow, IEEE) received the Diploma degree from ETH Zürich, Zürich, Switzerland, in 1988, the M.S. degree from the University of California at San Diego, La Jolla, CA, USA, in 1991, and the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 1995, all in electrical engineering and EECS.

He has consulted with numerous organizations, including Bosch, X-Fab, Texas Instruments, Maxim, Sony, Agilent, and Carlyle Group. From 1995 to 2000, he was with Cadence Design Systems, San Jose, CA, USA, where he was the Architect of the company's initiative on information hiding for intellectual property protection. In 2000, he joined Canesta Inc., Sunnyvale, CA, USA, as the Chief Architect, where he led the development of wireless 3-D CMOS image sensors. From 2008 to 2016, he was a Full Professor with Delft University of Technology, Delft, The Netherlands. Since 2002, he has been a Faculty Member with the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, where he is currently a Full Professor. He has been the driving force behind the creation of deep-submicrometer CMOS SPAD technology, which has been mass-producing since 2015 and is present in telemeters, proximity sensors, and medical diagnostics tools. He is a fellow of the Kavli Institute of Nanoscience, Delft. He was the Chair of the VLSI Design. He is a Distinguished Visiting Scholar of the W. M. Keck Institute for Space, Caltech, Pasadena, CA, USA. He has authored or coauthored over 400 articles and two books, and he holds 24 patents. His interests span from 3-D vision, LiDAR, FLIM, FCS, and NIROT to super-resolution microscopy, time-resolved Raman spectroscopy, cryo-CMOS circuits, and systems for quantum computing.

Dr. Charbon is a Distinguished Lecturer of the IEEE Photonics Society.

Carmen G. Almudever (Member, IEEE) received the Ph.D. degree in electronic engineering from the Universitat Politècnica de Catalunya (UPC BarcelonaTech), Barcelona, Spain, in 2014.

From 2014 to 2021, she was an Assistant Professor with the Quantum and Computer Engineering Department and the Group Leader of the Quantum Computing Division, QuTech, Delft University of Technology, Delft, The Netherlands, where she worked on the definition and implementation of scalable quantum computer architectures. In February 2021, she joined the Computer Engineering Department, Universitat Politècnica de València, Valencia, Spain, as a Distinguished Researcher (Beatriz Galindo Program for Attracting Talented Researchers). She has authored or coauthored one book chapter and more than 60 peer-reviewed technical publications. Her research focuses on different aspects of full-stack quantum computing systems, including quantum programming languages and compilers, quantum error correction, fault-tolerant quantum computation, mapping of quantum algorithms, and benchmarking and scalability of (modular) quantum computers.

Dr. Almudever serves as an Associate Editor for the *ACM Transactions on Quantum Computing*. She served as the Technical Program Co-Chair for the Computing Frontiers Conference 2021 and the Program Track Co-Chair (Quantum Systems Software) for QCE22. She was a co-recipient of several awards including the 2017 MICRO Best Paper Award, three HiPEAC awards (2017 and 2023), and the Intel Doctoral Student Honor Program Award (2012). She received the Beatriz Galindo Grant from the Spanish Ministry of Universities in 2020. She is the coordinator of an EiC Pathfinder Open Project on scalable multi-core quantum computer architectures.

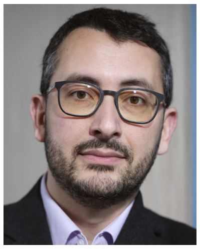
Giordano Scappucci received the M.Sc. degree from Sapienza Università di Roma, Rome, Italy, in 2000, and the Ph.D. degree from Università Roma Tre, Rome, in 2004.

Since 2005, he has been a Researcher with the Centre for Quantum Computing Technology, University of New South Wales, Sydney, NSW, Australia, and he has been the Group Leader of QuTech, Delft University of Technology, Delft, The Netherlands, since 2015. He leads the development of semiconductor materials for quantum computing at QuTech, Delft University of Technology. The silicon–germanium quantum materials developed by the Scappucci Lab enabled landmark experiments in the field of quantum information and are used by researchers globally for the development of quantum technologies. He has published more than 100 journal articles (e.g., in Nature, Science, and Nature Review Materials) and has given over 50 invited talks at international conferences, universities, research institutes, and industry.

Masoud Babaie (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2004, the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2006, and the Ph.D. degree (cum laude) in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2016.

From 2006 to 2011, he was with the Kavoshcom Research and Development Group, Tehran, where he was involved in designing wireless communication systems. From 2014 to 2015, he was a Visiting Scholar Researcher with Berkeley Wireless Research Center, Berkeley, CA, USA. In 2016, he joined Delft University of Technology, where he is currently an Associate Professor. He has authored or coauthored one book, three book chapters, 11 patents, and more than 100 peer-reviewed technical articles. His research interests include RF/millimeter-wave integrated circuits and systems for wireless communications and cryogenic electronics for quantum computation.

Dr. Babaie was a co-recipient of the 2015–2016 IEEE Solid-State Circuits Society Pre-Doctoral Achievement Award, the 2019 IEEE ISSCC Demonstration Session Certificate of Recognition, the 2020 IEEE ISSCC Jan Van Vessem Award for Outstanding European Paper, the 2022 IEEE CICC Best Paper Award, and the 2023 IEEE IMS Best Student Paper Award (second place). He received the Veni Award from The Netherlands Organization for Scientific Research (NWO) in 2019. He is the Co-Chair of the Emerging Computing Devices and Circuits Subcommittee of the IEEE European Solid-State Circuits Conference (ESSCIRC) and on the Technical Program Committee of the IEEE International Solid-State Circuits Conference (ISSCC). He is currently serving as an Associate Editor for the IEEE Solid-State Circuits Letters.



Fabio Sebastiano (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees (cum laude) in electrical engineering from the University of Pisa, Pisa, Italy, in 2003 and 2005, respectively, the M.Sc. degree (cum laude) from Sant'Anna School of Advanced Studies, Pisa, in 2006, and the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 2011.

From 2006 to 2013, he was with NXP Semiconductors Research, Eindhoven, The Netherlands, where he researched fully integrated CMOS frequency references, nanometer temperature sensors, and area-efficient interfaces for magnetic sensors. In 2013, he joined Delft University of Technology, where he is currently an Associate Professor. He has authored or coauthored one book, 11 patents, and over 100 technical publications. His research interests include cryogenic electronics, quantum computing, sensor read-outs, and fully integrated frequency references.

Dr. Sebastiano is on the Technical Program Committee of the ISSCC and the IEEE RFIC Symposium and has been on the Program Committee of IMS. He was a co-recipient of several awards, including the 2008 ISCAS Best Student Paper Award, the 2017 DATE Best IP Award, the ISSCC 2020 Jan van Vessem Award for Outstanding European Paper, and the 2022 IEEE CICC Best Paper Award. He has served as a Distinguished Lecturer of the IEEE Solid-State Circuit Society. He is currently serving as an Associate Editor for IEEE Transactions on Very Large Scale Integration and an Associate Editor for the IEEE Journal of Solid-State Circuits (JSSC) and served as the Guest Editor of JSSC.