The Cryogenic Temperature Behavior of Bipolar, MOS, and DTMOS Transistors in Standard CMOS

HARALD HOMULLE1, LIN SONG2, EDOARDO CHARBON1, AND FABIO SEBASTIANO1
1 QuTech, Delft University of Technology, 2628 CD Delft, The Netherlands
2 Department of Electrical Engineering, Tsinghua University, Beijing 100084, China
CORRESPONDING AUTHOR: H. HOMULLE (e-mail: h.a.r.homulle@tudelft.nl)
This work was supported by Intel Corporations.

ABSTRACT Both CMOS bandgap voltage references and temperature sensors rely on the temperature behavior of either CMOS substrate BJTs or MOS transistors in weak inversion. Bipolar transistors are generally preferred over MOS transistors because of their lower spread. However, at deep-cryogenic temperatures, the performance of BJTs deteriorates due to a significant reduction in current gain and a substantial increase in the base resistance. On the contrary, MOS devices show more stable performance even down to 4 K, but accurate device characterization for the design of such a circuit is currently missing. We present the characterization and analysis over the temperature range from 4 K to 300 K of both substrate bipolar PNP transistors and MOS transistors in standard and dynamic threshold MOS (DTMOS) configurations implemented in a standard 0.16-μm CMOS technology. These results demonstrate that employing MOS or DTMOS enables the operation of bandgap references and temperature sensors in standard CMOS technologies even at deep-cryogenic temperatures.

INDEX TERMS Characterization, cryogenics, substrate bipolar transistors, CMOS, dynamic-threshold MOS, bandgap references, temperature sensors.

I. INTRODUCTION
Cryogenic electronics is used in several applications, either for its improved performance, e.g., for the reduced thermal noise in the read-out for high-energy and nuclear physics experiments [1], [2], or for its operability in harsh environments, e.g., in (deep) space and quantum computing applications [3], [4]. While different technologies can be operated at cryogenic temperatures, only CMOS offers the integration of billions of transistors on a single chip together with a low-power consumption and sub-kelvin functionality [5], thanks to the progress achieved following Moore’s law. Therefore, CMOS offers the best choice for the future integration of complex and scalable cryogenic electronic systems.

Any CMOS subsystem, such as voltage regulators, analog-to-digital and digital-to-analog converters, requires a reference voltage, usually generated by a bandgap-reference circuit that can be implemented with either bipolar or MOS transistors in weak inversion as core devices. Additionally, the same bandgap principle is used in CMOS bandgap temperature sensors to accurately monitor die or environment temperature [6]. Bipolar transistors, especially substrate PNP transistors (Fig. 1(a)), readily available in any CMOS process, are generally preferred for their lower spread compared to MOS transistors (Fig. 1(b)) [6]–[9]. Prior research on CMOS devices was limited to characterizing substrate PNP transistors over the industrial temperature range (−55°C–125°C) [10]. BJTs, fabricated using standard bipolar processes, have been characterized only down to 77 K [11], since at lower temperature BJTs are not well behaved due to a significant decrease in their current gain $\beta$ and an increase in their base resistance $R_B$ [12]. For the same reasons, existing circuits operating at cryogenic temperatures comprise only bandgap references relying on alternative processes, such as SiGe BiCMOS [13], and a non-conclusive attempt to implement a CMOS bandgap-based temperature sensor operating at 93 K [14].

As an alternative, standard MOS transistors or MOS in dynamic-threshold MOS (DTMOS) configuration, i.e., a MOS with short-circuited gate and body terminals, can be
used as core devices in bandgap references [8] and temperature sensors [15]. MOSFETs can potentially offer operation at lower temperature compared to bipolar transistors. In particular, DTMOS are usually preferred over MOS for their lower spread, better matching, steeper subthreshold slope and operation at lower supply voltage [8], [15], [16]. Prior work demonstrated that MOS transistors in several technologies, such as bulk, fully depleted silicon-on-insulator and FinFET, can operate down to 4 K [17]–[22] and even sub-kelvin [5], [19], [23], [24]. However, DTMOS have been characterized only down to 77 K [25]–[27] and no MOS-based bandgap core has been evaluated at lower cryogenic temperatures to the best of the authors’ knowledge.

In this paper, we extend our prior results on the characterization of substrate PNPs presented in [12] by also including the characterization of MOSFETs in standard and dynamic threshold configuration fabricated in a standard 0.16 μm CMOS technology over the temperature range from 4 K to 300 K. The measurement setup is described in Section II; experimental results and their discussion are shown in Section III for the bipolar transistor, whereas Section IV covers MOS characterizations. Finally, the impact on cryogenic circuit design is discussed in Section V and conclusions are drawn in Section VI.

II. MEASUREMENT SETUP
In bandgap references and temperature sensors, the voltage drop on a diode-connected device (BJT or MOS in weak inversion) is compared or combined with the voltage difference on two diode-connected devices biased at different current densities, which is expected to be proportional to absolute temperature (PTAT) [6]. Since the accuracy of those voltages directly influences the performance of bandgap references and temperature sensors, they are thoroughly analyzed hereafter.

The PNPs have been characterized in a Janis ST-500 cryogenic probe station with reference temperature sensor on the sample holder, while the MOS transistors have been mounted on a PCB immersed in liquid helium or in helium vapours with the reference temperature sensor mounted on the PCB. Electrical characterization has been performed using Keithley 2636B SourceMeter Units (SMUs).

As depicted in Fig. 3, the drop on one of the diode-connected devices (\(V_{BE}\) for PNP, \(V_{GS}\) for MOS and DTMOS) has been characterized while applying a bias current (\(I_1\)) and measuring the base current (\(I_B\)) for the PNP, whereas the drain current (\(I_D\)) for MOS transistors is measured, and a small bias current is applied for the DTMOS. The measurement setup consists of a dual channel system with the ability to control and measure both the collector and the base current of the PNPs, as well as the drain current of the MOS transistors and the DTMOS. The setup allows for the measurement of the differential voltage \(\Delta V_{BE/DS}\) and the drain current \(I_D\) for MOS and DTMOS.

FIGURE 1. Simplified cross section of a substrate PNP (a) and a PMOS (b) in a standard CMOS process.

FIGURE 2. Micrographs of the PNP (a) and MOSFET (b) test structures, indicating the relevant contacts mentioned in Fig. 1.

FIGURE 3. (a) Schematic diagram of the measurement set-up with details of the device under test (DUT) for (b) PNP, (c) MOS and (d) DTMOS characterization. Current is injected in the emitter or source, the collector or drain and gate are always grounded, whereas the base is grounded in the case of \(\Delta V_{BE}\) measurements, but connected to a current meter for \(V_{BE}\) measurements. For MOS measurements, the transistor’s body is connected to the standard supply, i.e., 1.8 V in this technology.
difference in the voltage drop over two diode-connected devices ($\Delta V_{BE} = V_{BE1} - V_{BE2}$ for the PNP, $\Delta V_{GS} = V_{GS1} - V_{GS2}$) has been characterized while biasing the pair with a current ratio $p (I_1 = pI_2)$. For the standard MOS, the drain and gate were shorted to ground, whereas the body was biased with the standard supply voltage (1.8 V). The current was injected into the source. For the DTMOS, the drain, gate and body were shorted to ground, and again current was injected into the source. To improve measurement accuracy in the $\Delta V_{BE}$ measurements, the base has been shorted to ground. Mismatch of the PNP and MOS pairs has been averaged using a dynamic element-matching (DEM) technique by averaging the results for $I_1 = pI_2$ and $I_2 = pI_1$, as shown in [10].

### III. PNP Experimental Results

The collector current $I_C = I_E - I_B$ was computed from the measured base and emitter currents, and plotted in Fig. 4 as a function of $V_{BE}$, showing that the exponential relation between $I_C$ and $V_{BE}$ holds over a wide range of currents and down to approximately 70 K. From the same data, the current gain $\beta = I_C/I_B$ is computed and found to be relatively current independent over more than three decades of $I_E$ for temperatures above 70 K (Fig. 5). As common for vertical PNPs in modern CMOS technologies [28], [29], the PNPs show a low current gain ($\beta \approx 6$ at room temperature), which drops below unity for temperatures below 130 K. $\beta$ degrades quasi-exponentially with cooling, which is expected due to the apparent bandgap narrowing associated with the emitter [11].

Device parameters can be extracted by observing that

\[
I_C = I_s \left( e^{\frac{qV_{BE,i}}{kT}} - 1 \right)
\]

where $V_{BE,i}$ is the intrinsic base-emitter voltage (i.e., $V_{BE}$ for $R_B = 0$), $n_{BJS}$ is the effective emission coefficient, $I_s$ is the saturation current, $k$ is the Boltzmann constant, $T$ is the absolute temperature and $q$ is the electron charge. Consequently,

\[
V_{BE} = V_{BE,i} + R_B I_B = n_{BJS} \frac{kT}{q} \ln \left( \frac{I_C}{I_s} + 1 \right) + R_B \frac{I_E}{\beta + 1}
\]

Applying (2) to a PNP pair biased at current ratio $p = I_1/I_2$, we get

\[
\Delta V_{BE} \equiv n_{BJS} \frac{kT}{q} \ln (p) + R_B \frac{p - 1}{\beta + 1} I_{E2}
\]

under the following assumptions: $I_C \gg I_s$, i.e., biasing far from the low-injection region; $R_B$ is current independent; $I_E$ is current independent, which is valid for a large range of currents as shown in Fig. 5; accurate PNP matching, which is achieved by proper device layout and dynamic element matching. Fig. 6(a) shows the measured $\Delta V_{BE}$ as a function of the injected emitter current, indicating a stable $\Delta V_{BE}$ over a current range from approximately $10^{-10} - 10^{-5}$ A and

---

**FIGURE 4.** PNP Collector current versus emitter-base voltage for temperatures ranging from 7 K to 294 K.

**FIGURE 5.** PNP current gain $\beta$ versus emitter current.

**FIGURE 6.** $\Delta V_{BE} = V_{BE1} - V_{BE2}$ versus (a) $I_{E1}$ and (b) temperature for $I_{E1} = pI_{E2}$ with $p = 3$. Where $I_E$ is the emitter current, $R_B$ the base resistance and $\beta$ the current gain.
for temperatures above 70 K. $\Delta V_{BE}$ is approximately PTAT above 70 K (Fig. 6(b)) as expected from (3), especially for emitter currents in the medium-low current regime.

Using the measured current gain to fit (3) to the measured $\Delta V_{BE}$ (Fig. 6(b)), the base resistance $R_B$ and the emission coefficient $n_{BJT}$ have been extracted. As shown in Fig. 7, the emission coefficient is close to unity at room temperature, as expected in CMOS substrate PNP's [10], and is approximately constant over a wide range of emitter currents at least down to 130 K. At lower temperatures, $n_{BJT}$ shows a steep increase and is no longer current independent. The base resistance $R_B$ increases significantly at lower temperatures, as discussed in more details in [12].

Finally, in order to study the temperature behavior of $V_{BE}$, the following standard model is fitted to the measured data [6], [10]:

$$V_{BE}(T) = V_{g0} \left( 1 - \frac{T}{T_r} \right) + \frac{T}{T_r} V_{BE}(T_r) - \eta \frac{kT}{q} \ln \left( \frac{T}{T_r} \right) + \frac{kT}{q} \ln \left( \frac{I_C(T)}{I_C(T_r)} \right)$$

(4)

where $V_{g0}$ and $\eta$ are constant fitting parameters and $T_r$ is a reference temperature. Fitting (4) to the measurement data over both the temperature range from 111 K to 294 K with $T_r = 294$ K and the current range from 10 pA to 100 $\mu$A results in an average fitting error below 2 mV over the fitting range (Fig. 8) and in fitting parameters very close to prior work [10] ($\eta = 4.20 \pm 0.4$, $V_{g0} = 1.13$ V $\pm$ 5 mV, with 95% confidence intervals).

IV. MOS EXPERIMENTAL RESULTS

Thanks to the nature of MOS devices, the gate current $I_G$ can be neglected, thus simplifying the MOS current and voltage relationship in the subthreshold region to:

$$I_{DS} = I_e e^{\frac{V_{GS}-V_t}{n_{MOS}kT}} \left( 1 - e^{-\frac{qV_{DS}}{kT}} \right)$$

(5)

in which $n_{MOS}$ is the non-ideality factor. Fig. 10 shows the measured $I_{DS}$ vs $V_{GS}$ curves for the four MOS combinations. The drain current exhibits an exponential behavior over a wide range of currents, as expected from (5). Although this range is smaller compared to that of the PNP, the exponential relationship holds even at temperatures as low as 4 K. As shown in Fig. 9, the subthreshold slope $SS$ decreases at lower temperatures, with the $SS$ of DTMOS being lower with respect to MOS and closer to the $SS$ of bipolar transistors as expected from theory [8], [16]. Interestingly, the large DTMOS shows a steeper slope compared to the other devices above 100 K, whereas the small DTMOS improves beyond that down to 4 K.

$\Delta V_{GS}$ was measured with the same current ratio $p = 3$ as the bipolar transistor to enable a fair comparison (Fig. 11). $\Delta V_{GS}$ is slightly less current dependent for larger devices, which can be attributed to the absence of short-channel effects. Mismatch was evident for the small devices when...
operated without dynamic element matching (not shown), since the $\Delta V_{GS}$ could even become negative at the lowest temperatures. For the large DTMOS, $\Delta V_{GS}$ is stable, i.e., current independent, over a current range from roughly $10^{-11} - 10^{-8}$ A, which is two orders of magnitude less than the BJT.

The large DTMOS pair (LL) shows the lowest current dependency of $\Delta V_{GS}$, as highlighted in Fig. 12. In a similar way as done for the emission coefficient of the PNP (Fig. 7), its non-ideality factor $n_{MOS}$ is shown in Fig. 13. $n_{MOS}$ is close to its ideal value of 1 for low currents and for temperatures down to roughly 130 K. At lower temperatures, $n_{MOS}$ increases similarly as in the bipolar device and deviates significantly from ideality.

Similar to Fig. 8, $V_{GS}$ is plotted versus temperature in Fig. 14 for the large DTMOS. To clearly assess the effect of replacing the BJTs with (DT)MOS devices in standard bandgap references and temperature sensors, the same model used for $V_{BE}$ (4) is used to fit the measured $V_{GS}$. However, fitting over both the temperature range and the bias current
range gives significant deviation, as shown in Fig. 14(c,d), already below 200 K (13-mV average fitting error for the large DTMOS, 26 mV for the large MOS) over the complete temperature range when using $V_{g0} = 0.47$ V and $\eta = 1.37$. On the contrary, fitting the same model against the individual curves results in a lower fitting error (3.4 mV for the large DTMOS, 4.6 mV for the large MOS), but a considerable spread in both $V_{g0}$ and $\eta$ is observed. $\Delta V_{GS}$ extrapolated at 0 K, i.e., $V_{g0}$ in (4), appears to be roughly exponentially dependent on the biasing current $I_D$. Over the current range from 10 pA to 100 nA, $V_{g0}$ of the large DTMOS shifts by approximately 100 mV from 0.43 V to 0.53 V. For the large MOS, this shift is even larger (from 0.79 to 0.94 V). $\eta$ spreads from 3.6 to $-0.5$ in the DTMOS.

Finally, the non-ideality factor (for the MOS and DTMOS) and the emission coefficient (for the BJT) of all devices is shown in Fig. 15. The $n_{MOS}$ in DTMOS devices is closer to the BJTs emission coefficient $n_{BJT}$ compared to their MOS counterparts. In all cases, however, $n$ increases rapidly at temperatures below 100 K.

V. DISCUSSION

$V_{BE}$ shows a temperature dependence in good agreement with the room-temperature model above 50 K (Fig. 8) and with high accuracy above 110 K. Although $V_{GS}$ of the large DTMOS follows fairly well the same model down to 4 K, an apparent current dependence of the extrapolated $V_{GS}$ at 0 K is observed and may lead to a lower performance of DTMOS-based references, especially when employing a single point trim and assuming a fixed $V_{g0}$. This is also confirmed by the larger spread of DTMOS-based circuits when compared to their BJT counterparts. For example, Souri et al. [9], [15] reports an inaccuracy of $\pm 0.48$°C versus $\pm 0.15$°C for, respectively, a DTMOS- and BJT-based temperature sensor.

Both BJT and DTMOS exhibit an increased curvature of $V_{BE}$ or $V_{GS}$ at deep-cryogenic temperatures, but much more pronounced in the BJT. Although it can potentially be corrected by a compensation circuit, the performance of BJT-based bandgap cores may be limited below 70 K by this curvature and the increased sensitivity of $V_{BE}$ with respect to $I_E$ at low temperatures. DTMOS transistors show lower $V_{GS}$ at 4 K compared to MOS and bipolar transistors, thus allowing for sub-1-V cryogenic circuits.

$\Delta V_{BE}$ is sensitive with respect to current gain and base resistance, whereas $\Delta V_{GS}$ does not show such limitations thanks to the negligible gate current in MOS. For the large DTMOS, $\Delta V_{GS}$ is approximately PTAT over the complete temperature range and standard circuit topologies could be adapted to take the residual non-linearity into account. In contrast, the BJT exhibits a $\Delta V_{BE}$ behavior that significantly deviates from being PTAT below 70 K, possibly leading to additional curvature spread that cannot be tackled by standard circuit techniques [6]. Moreover, both the emission coefficient $n_{BJT}$ in the BJT and the non-ideality factor $n_{MOS}$ in the MOS and DTMOS show non-negligible deviation from unity already at 130 K. While this could be cancelled by the ratiometric nature of temperature sensors [6], its spread could strongly affect the accuracy of bandgap references.

VI. CONCLUSION

Substrate PNP transistors have been shown to operate reliably down to 70 K, however their usability at lower temperatures is heavily limited by the decreased current gain, increased base resistance and non-unity emission coefficient. MOS devices, and particularly DTMOS, offer proper functionality extending even at temperatures as low as 4 K, although their subthreshold slope and non-ideality factor significantly deteriorates approaching 4 K. Based on the
extensive characterization of MOS and DTMOS devices, we believe that, despite the above-mentioned limitations, MOS or DTMOS can replace BJTs in voltage references and temperature sensors operating over the entire range from 4 K to 300 K and potentially at even lower temperatures, as required in several applications, such as the electronics for satellites and spacecrafts, cryogenic temperature sensors, and the control interface of quantum processors.

ACKNOWLEDGMENT

The authors would like to thank NXP for chip fabrication.

REFERENCES


[9] M. A. P. Pertijis and J. Huijsing, Precision Temperature Sensors in


HARALD HOMULLE received the bachelor’s and master’s degrees from the Delft University of Technology in 2012 and 2014, respectively, both in electrical engineering. During the master project, he did an internship with EPFL, Lausanne, for the development of a multichannel TDC on FPGA.

In 2014, he joined the research group of E. Charbon as a Ph.D. student on the topic of cryogenic electronics, with the main focus on the demonstration of the first cryogenic qubit read-out system.

LIN SONG received the bachelor’s degree in electrical engineering from Tsinghua University, Beijing, China, in 2014, and the master’s degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 2016. He is currently an Analog Designer with Analog Devices Inc., Beijing Design Office, China.
EDOARDO CHARBON received the Diploma degree from ETH Zurich in 1988, the master’s degree from the University of California at San Diego in 1991, and the Ph.D. degree from the University of California at Berkeley in 1995, all in electrical engineering and EECS.

From 1995 to 2000, he was with Cadence Design Systems. In 2000, he joined Canesta Inc. as a Chief Architect. Since 2002, he has been a member of the Faculty of EPFL. In 2008, he joined the Faculty of Delft University of Technology as a Full Professor in VLSI design. Recently, he became a PI with the QuTech research initiative, focusing on the development of cryogenic electronics for quantum computing purposes. His current research interests include (3-D, bio-, and medical) imaging, quantum integrated circuits, and cryogenic electronics.

FABIO SEBASTIANO received the bachelor’s and master’s degrees in electrical engineering from the University of Pisa, Italy, in 2003 and 2005, respectively, the master’s degree from the Sant’Anna School of Advanced Studies, Pisa, Italy, in 2006, and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011.

From 2006 to 2013, he was with NXP Semiconductors Research, Eindhoven, The Netherlands, where he was involved in conducting research on fully integrated CMOS frequency references, deep-submicron temperature sensors, and area-efficient interfaces for magnetic sensors. In 2013, he joined the Delft University of Technology, where he is currently an Assistant Professor. His current research interests include sensor read-outs, fully integrated frequency references, and cryogenic electronics for quantum applications.