Subthreshold Mismatch in Nanometer CMOS at Cryogenic Temperatures

P. A. 't Hart¹, M. Babaie¹, E. Charbon⁴, A. Vladimirescu¹,²,³ and F. Sebastiano¹
¹TU Delft, Delft, The Netherlands; ²ISEP, Paris, France; ³UC Berkeley, Berkeley, CA, USA; ⁴EPFL, Lausanne, Switzerland; ⁵Intel, Hillsboro, OR, USA

Abstract—Cryogenic device models are essential for the reliable design of the cryo-CMOS interface that enables large-scale quantum computers. In this paper, mismatch characterization and modeling of a 40-nm bulk CMOS process over the 4.2–300 K temperature range is studied, towards an all-operating-region mismatch model. An overall increase of variability is shown, in particular in the subthreshold region at cryogenic temperatures due to a dramatic increase of the subthreshold slope mismatch. Mismatch in strong inversion is modeled by the Croon model while the weak-inversion region is modeled by taking subthreshold slope variability into account. This results in the first model capable of predicting mismatch over the whole range of operating regions and temperatures.

I. INTRODUCTION

Quantum computers promise an exponential speed-up in solving several problems that are beyond the capability of today’s computers. Their operation is based on processing the information stored in quantum bits (qubits), which must be typically cooled to deep cryogenic temperatures below 1 K for proper operation [1]. Since current implementations comprise only a few tens of qubits, they can be directly wired to conventional room-temperature control electronics. This approach, however, strictly limits the scalability of the system in terms of the number of qubits due to the reliability of cables and their physical size. To overcome these constraints, cryogenic CMOS control electronics can be placed in close proximity to the qubits, and hence, also operate at cryogenic temperatures with a preferred operating temperature of 4.2 K, at which typical dilution refrigerators can provide significant cooling power [2]. Nanometer CMOS is the technology of choice by virtue of its maturity, high speed/bandwidth capability and high level of integration, which together enable the complex circuits required to control a practical quantum computer.

The availability of accurate device models aimed for and validated at cryogenic temperatures is indispensable for the design of such a CMOS controller. In addition to the DC/RF characterization and modeling recently presented [3], [4], complete models for cryogenic nanometer CMOS must include device mismatch, since mismatch can pose severe limits on the performance of sensitive analog and digital circuits. CMOS device mismatch at room temperature has been extensively studied [5], [6], and even its temperature dependence has been characterized over a limited range (273–400 K) [7]. Prior work on cryogenic mismatch covering various CMOS nodes (0.5-µm Silicon-on-Sapphire CMOS [8], 0.35-µm bulk CMOS [9] and 40-nm bulk CMOS [10]) indicates that CMOS transistor matching strongly deteriorates at cryogenic temperatures. However, no data has ever been presented on cryogenic mismatch in the subthreshold region, although such region of operation is crucial for cryogenic circuits, as the limited cooling power pushes for ultra-low-voltage and ultra-low-power circuits exploiting the subthreshold behavior.

This paper focuses on filling this gap by presenting, for the first time, subthreshold-region device mismatch characterization and modeling of 40-nm CMOS transistors over the temperature range from 4.2 K to 300 K. It is shown that while taking into account the mismatch in threshold voltage ($V_{TH}$) and transconductance factor ($β$) is sufficient to accurately model mismatch in moderate- and strong-inversion regimes, it is necessary to include the effect of mismatch in subthreshold swing ($SS$) to extend the mismatch modeling to all operating regions, from subthreshold to strong inversion and over the full temperature range (4.2–300 K).

II. TEST STRUCTURES AND MEASUREMENT SETUP

Thin oxide, standard-threshold-voltage, N- and PMOS devices were fabricated in an 1.1-V, 40-nm bulk CMOS process. Per die, three different geometries were available for mismatch characterization, layed out in a matched pair configuration: $W/L = 120$ nm/40 nm (minimum size), 360 nm/120 nm and 1.2 µm/400 nm. Care was taken to reduce the impact of systematic mismatch, edge effects and mechanical stress by placing...
the devices at minimum distance with symmetrical connections, surrounding them by identical dummy devices and by keeping metallization away from the pair as indicated in Fig. 1.

The matched pairs (24 per geometry per die) share their source and drain terminals, while the gate is either connected to a common bus when the device is selected or pulled to the appropriate rail when deselected, as shown in Fig. 2. A digital interface controls the selection of devices and enables automated characterization. Kelvin connections are employed to mitigate the effects of parasitic resistances. The potential of the nwell/bulk are fixed by contact rings biased at 1.1 V and 0 V, respectively. It should be noted that this is a redesigned test structure, fabricated in a different 40-nm bulk CMOS process than the one presented in [10]. This was necessary to minimize leakage enabling accurate subthreshold-current measurement.

For the electrical characterization, two Keithley 2636B Source Measurement Units (SMUs) were connected to the samples by low-leakage triaxial guarded connections required for the measurement of the subthreshold currents. Since the time required to characterize a single die exceeds 12 hours due to the low currents and associated instrument settling times, an automated setup with samples directly dipped into liquid helium is preferred over a manned setup with a cryogenic probe station. A total of three dies (72 matched pairs per geometry) have been characterized both in triode (\(|V_D| = 50 \text{ mV}\)) and saturation region (\(|V_D| = 1.1 \text{ V}\)) at Room Temperature (RT, \(T = 300 \text{ K}\)) and Liquid Helium Temperature (LHT, \(T = 4.2 \text{ K}\)). After recording the drain current (\(I_D\)) as a function of gate voltage (\(V_{GS}\)) for each device, the leakage of the array is measured by deselecting all devices and later subtracted from \(I_D\) to extend the measurement range on the low side. Individual devices directly accessible via reserved bond pads have been used to validate the array data.

\(V_T\) and \(\beta\) were extracted from the drain current by means of the Extrapolation in Linear Region (ELR) method [11]. To handle the increased mismatch at cryogenic temperatures, drain-current mismatch is described by \(\Delta \log(I_D)\).

### III. Experimental Results

Fig. 3 shows \(I_D = V_{GS}\) characteristics for 48 NMOS devices on a single die at both RT and LHT. The temperature impact on the three relevant parameters can clearly be identified at LHT: \(V_{TH}\) increases (\(\Delta V_{TH} \approx 100 \text{ mV}\)), mobility increases (\(\beta_{4.2K}/\beta_{300K} \approx 2\times\)) and SS decreases (\(SS_{300K} \approx 90 \text{ mV/dec} \rightarrow SS_{4.2K} \approx 20 \text{ mV/dec}\)). At LHT, NMOS devices start to show anomalous behavior when biased around and below the threshold voltage, while PMOS suffers from an SS varying with gate bias. NMOS devices also show this phenomenon to a lesser extent. Both effects are not present at RT and have not been observed in another previously characterized 40-nm process [10]. Similar effects are observed in pad-accessible devices, pointing to the effects being inherent to the devices. The differences between NMOS and PMOS and between technologies can most likely be explained by different fabrication steps and process parameters, with related defects.

To investigate drain-current mismatch, \(\Delta \log(I_D) = \log(I_{D1}) - \log(I_{D2})\) data was computed from 72 device pairs and plotted in Fig. 4. The increased \(V_{TH}\) and decreased SS at LHT causes \(I_D\) to reach the instrument current floor at much higher \(V_{GS}\) compared to RT. The effect of the anomalous behavior around \(V_{TH}\) can be clearly seen, e.g., at \(V_G = 0.6 \text{ V}\) for the largest device. This, and the increased SS mismatch exacerbated by the slope changes, results in a subthreshold mismatch orders of magnitude higher at LHT.
Drain-current mismatch is analyzed and modeled by $\sigma_{\Delta \log I_D}$ derived from the data in Fig. 4. The Croon model [6] is used to predict RT/LHT drain-current mismatch in moderate- to strong-inversion, as such model has already been proven to apply at cryogenic temperatures [10]:

$$\sigma_{\Delta \log I_D}^2 = \frac{1}{(\ln 10)^2} \left[ \sigma_{\Delta g_m}^2 \left(\frac{\bar{g}_m}{I_D}\right)^2 + \sigma_{\Delta V_{TH}}^2 \right],$$  

(1)

where $I_D$ is the drain current and $g_m$ is the transconductance, the overline indicating the average value over all matched pairs with a given geometry. The In 10 originates from the use of the logarithm with base 10.

The increased subthreshold mismatch resulting from the increased $SS$ variability cannot adequately be modeled by the Croon model, therefore a subthreshold mismatch model was adopted that takes both $V_{TH}$ and $SS$ terms into account [12]:

$$\sigma_{\Delta \log I_D}^2 = \left(\frac{\sigma_{\Delta V_{TH}}}{SS}\right)^2 + \left(\frac{V_{GS} - V_{TH}}{\sigma_{\Delta SS}}\right)^2. \quad (2)$$

The RT correlation between $V_{TH}$ and $SS$ was found to be insignificant. At LHT, the slope changes mask any correlation that could be present. Hence, the correlation term is not considered in this analysis. It is important to note that the model parameters in Eq. (1) and (2) ($\sigma_{\Delta g_m}$, $\sigma_{\Delta V_{TH}}$, $\sigma_{\Delta \beta}$ and $\sigma_{\Delta SS}$) are not fitted, but the mismatch in each parameter ($V_{TH}$, $\beta$, $SS$) has been extracted from $I_D - V_{GS}$ data of each pair in the triode region, used to compute the relative standard deviation and then directly applied to the models. This is a strong argument for the validity of the adopted model. Drain-current mismatch in triode operation ($|V_{DS}| = 50 \text{ mV}$) at RT is plotted in Fig. 5a and 6a. The Croon model is able to predict mismatch accurately in moderate- and strong-inversion and relatively well in subthreshold. In this region, the Croon model underestimates the mismatch as it does not take the $SS$ variability into account. The subthreshold model gives a better estimation, although the two models predict similar mismatch, since the main mismatch contributor is the $V_{TH}$ variability due to the small $SS$ variability at RT. A large increase in subthreshold mismatch can be seen at LHT in Fig. 5b and 6b. The Croon model is still able to accurately predict mismatch in strong inversion as shown for RT, however, it breaks down in the subthreshold region as a result of the poorly behaved $g_m/I_D$ due to the slope changes and the much larger variability in $SS$ with respect to RT. The Croon model is therefore only used for $V_{GS} > V_{TH}$ with a small guard band that avoids the effects of the anomalous behaviour around $V_{TH}$. The subthreshold model does not suffer from this as it does not rely on $g_m/I_D$ but only on the extracted parameters $\sigma_{\Delta V_{TH}}$ and $\sigma_{\Delta SS}$. It is then able to predict subthreshold mismatch with much higher accuracy and is only used for $V_{GS} < V_{TH}$ since its validity is limited to subthreshold.

Drain-current mismatch in saturation is particularly relevant to circuit designers since this is the most commonly used operating regime in circuits. Thus, mismatch in saturation ($|V_{DS}| = 1.1 \text{ V}$) was explicitly characterized and modeled, as shown in Fig. 5c,d and 6c,d. Since the high $|V_{DS}|$ lowers
$V_{TH}$ via Drain Induced Barrier Lowering (DIBL), the $V_{TH}$ extracted in triode with ELR does not capture this effect and therefore $V_{TH}$ is extracted in saturation with the use of the Extrapolation in Saturation Region (ESR) method [11]. Apart from this, similar remarks as those for the triode region are valid for the data and models in saturation. The effectiveness of the model is also shown in Fig. 7, for which the model parameters have been obtained by fitting, instead of by direct extraction, to improve model fitting.

In order to mitigate parameter variability, the device area can be increased according to Pelgrom’s scaling law [5]:

$$\sigma_{\Delta V_{TH}} = \frac{A_{VT}}{\sqrt{WL}}, \quad \sigma_{\Delta \beta/\beta} = \frac{A_{\beta}}{\sqrt{WL}}, \quad \sigma_{ASS/SS} = \frac{A_{SS}}{\sqrt{WL}}$$

(3)

where $A_{VT}$, $A_{\beta}$ and $A_{SS}$ are the threshold voltage, transconductance factor and subthreshold swing area-scaling parameters respectively, and $WL$ is the device active area. To check the validity of Pelgrom’s scaling law, Pelgrom plots for $\sigma_{\Delta V_{TH}}, \sigma_{\Delta \beta/\beta}$ and $\sigma_{ASS/SS}$ are shown in Fig. 8. The linear fittings have been inversely weighted with the error bars and the slopes represent the $A$-factors. It can be seen that the variability of these 3 parameters can be modeled by Eq. (3) both at RT and LHT, as already shown in [10] for the threshold voltage and transconductance factor of a different 40-nm process. While RT $A_{VT}$ and $A_{\beta}$ are very similar to the ones previously extracted in [10], the LHT values reported here are significantly larger. $A_{\beta}$ increases by 2× from RT to LHT and is the main contributor to mismatch (for $V_{GS} - V_{TH} > 200$ mV), although the increase of $A_{VT}$ is significant. The SS variability shows a very large increase at LHT for both NMOS and PMOS devices, therefore split axes are used to accommodate them in one plot. $A_{SS}$ for the NMOS device increases with a factor of 10×. Due to the excessive slope-changing behavior of PMOS device at LHT, it was not possible to fit the model of Eq. (3) to the data and thus $A_{SS}$ is not reported in this case.

IV. CONCLUSION

This paper reports a study of the subthreshold device mismatch of 40-nm CMOS devices at cryogenic temperatures for the first time. Mismatch increases at cryogenic temperatures in all operating regions and, in particular, subthreshold matching worsens more than several orders of magnitude. Although the Croon model accurately describes mismatch in strong inversion, it is unable to do so in subthreshold where the SS variability becomes dominant. Therefore, a different model taking the SS variability into account is used to adequately predict the subthreshold mismatch. Improving matching by increasing the device area as predicted by Pelgrom’s model however, proves to be valid at cryogenic temperatures in all operating regions including subthreshold. As a result, the proposed mismatch models can become an essential tool for the design of the cryo-electronics that will enable scalable quantum computers.

ACKNOWLEDGMENT

The authors would like to thank Dr. H.P. Tuinhout and Dr. M.J.M. Pelgrom for the helpful discussions and Intel for funding.

REFERENCES

[1] L. Vanderven, “Quantum Computing - The Next Challenge in Circuit and System Design,” in 2017 IEEE ISSCC, Feb 2017, pp. 24–29.
[2] F. Sebastianio, H. Homulle, B. Patra, R. Incandela, J. van Dijk, L. Song, M. Babaie, A. Vladimirescu, and E. Charbon, “Cryo-CMOS Electronic Control for Scalable Quantum Computing,” in 2017 54th ACM/EDAC/IEEE DAC, 2017, pp. 1–6.
[3] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastianio, “Characterization and Compact Modeling of Nanometer CMOS Transistors at Deep-Cryogenic Temperatures,” IEEE JEDS.
[4] A. Beckers, F. Jazaeri, and C. Enz, “Characterization and Modeling of 28 nm Bulk CMOS Technology down to 4.2 K,” IEEE JEDS.
[5] M. J. M. Pelgrom and A. C. J. Duijnmaker, “Matching Properties of MOS Transistors,” in Proc. ESSCIRC, Sept 1988, pp. 327–330.
[6] J. A. Croon, W. Sanoen, and H. Maes, Matching Properties of Deep Sub-Micron MOS Transistors. Springer, 2005.
[7] P. Andricicola and H. P. Tuhnout, “The Temperature Dependence of Mismatch in Deep-Submicron Bulk MOSFETs,” IEEE EDL, vol. 30, no. 6, pp. 690–692, June 2009.
[8] K. Das and T. Lehmann, “Effect of Deep Cryogenic Temperature on Silicon-On-Insulator CMOS Mismatch: A Circuit Designers Perspective,” Cryogenics, vol. 62, pp. 84–93, Jul. 2014.
[9] N. C. Dao, A. E. Kais, C. T. Jin, and P. H. W. Leong, “Impact of Series Resistance on Bulk CMOS Current Matching Over the 5-300K Temperature Range,” IEEE EDL, vol. 38, no. 7, pp. 847–850, July 2017.
[10] P. A. ‘t Hart, J. P. G. van Dijk, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastianio, “Characterization and Model Validation of Mismatch in Nanometer CMOS at Cryogenic Temperatures,” in 2018 48th ESSDERC, Sep. 2018, pp. 246–249.
[11] A. Ortiz-Conde, F. G. Sánchez, A. Cerdeira, M. Estrada, and Y. Yue, “A Review of Recent MOSFET Threshold Voltage Extraction Methods,” Microelectronics Reliability, vol. 42, no. 4, pp. 583 – 596, 2002.
[12] P. Magnone, F. Crupi, P. Andricicola, H. Tuhnout, and R. J. P. Lander, “FinFET Mismatch in Subthreshold Region: Theory and Experiments,” IEEE TED, vol. 57, no. 11, pp. 2848–2856, Nov 2010.