Design of Cryogenic Fully Differential Gain Boosting-OTA by the $g_m/I_d$ methodology used for a 14 bit Pipelined-SAR ADC

Mingjie Wen, Chao Luo, BoLun Zeng and Guoping Guo

Abstract—Quantum computing (QC) requires cryogenic electronic circuits as control and readout sub-systems of quantum chips to meet the qubit scale-up challenges. At this temperature, MOSFETs transistors exhibit many changes such as higher threshold voltage, higher mobility, and steeper subthreshold slope. We present a cryogenic fully differential gain boosting-OTA used for a 14 bit Pipelined-SAR ADC operating at 4.2K as the readout circuit for semiconductor-based quantum computing system. Using $g_m/I_d$ methodology to get pre-computed lookup tables based on the cryogenic 110nm BSIM4 model. The proposed OTA achieves very high unity-gain frequency @1.23GHz and open-loop low frequency gain @101dB. The total power consumption is 2.66mW at 4.2K, and a setting accuracy better than 0.01% with $f_{closed}$ of 37MHz in a closed-loop application.

Index Terms—quantum computing, cryogenic CMOS analog circuit, OTA, low threshold voltage-MOSFETs, $g_m/I_d$ methodology

I. INTRODUCTION

Quantum processors based on arrays of quantum bits (qubits) including semiconductor quantum dots or superconducting transmons operate at sub-Kelvin cryogenic temperatures. Bringing the quantum-classical interface closer to the qubit is the core solution for overcoming the qubits scale-up challenge. Classical electronic components will be greatly changed even invalid in such a condition and the use of MOSFETs operating at cryogenic temperatures (cryo-CMOS) has been studied extensively [1]. Cryogenic MOSFETs' modeling plays a key role in the control and readout circuits for quantum computing. MOSFETs' cryogenic characteristics have been widely reported in recent years [2] [3] [4]. In general, higher threshold voltages, higher mobility and steeper subthreshold swing (SS) can be found in CMOS at cryogenic temperature. In this paper, a commercial 110-nm low threshold voltage MOSFETs (LVT-MOSFETs) technology is experimentally characterized at liquid helium temperature (LHT) and room temperature (RT). Based on that, we have built a BSIM4 compact model which can be used in commercial SPICE or HSPICE simulator for verifying cryo-CMOS analog circuits. Using this accurate model, we have designed a folded cascade gain-boosting operational transconductance amplifier (OTA) by the $g_m/I_d$ method used for a cryogenic 14 bit Pipelined-SAR analog to digital converter (ADC).

$g_m/I_d$ methodology is a powerful philosophy for analog designers especially for the nanoscale MOSFET. The 110nm Cryo-CMOS is complicated and its IV-behavior can’t match the classic square law model, which is based on ideal diode model and applies only near the strong inversion (SI) [5]. For this reason, we use pre-computed lookup tables to maintain a model-agnostic approach. The goal of this work is to design a high performance and low power consumption OTA by the $g_m/I_d$ method based on a 110nm cryo-CMOS BSIM4 model. The fully differential gain boosting-OTA works at 4.2K and is used as a residue amplification in a two-stage Pipelined-SAR ADC.

This paper is divided into five sections. The second section will introduce the characteristics of the 110nm LVT-CMOS at RT and LHT and the third section focuses on the $g_m/I_d$ method, including the establishment of pre-computed lookup tables. The fourth section focuses on the high performance and low power consumption OTA and the fully differential gain boosting structure and analysis of the circuit. The fifth section will give the post-simulation of the OTA, the performance characteristics at RT and LHT.

II. CRYOGENIC CMOS

MOSFETs have specific phenomena at cryogenic temperature, such as kink effect, impurity freeze-out effect and so on. As 110nm model shows, the low-field mobility increases almost ten times for the $10\mu m/10\mu m$ size NMOSFETs at LHT. As illustrated in Fig.1(a,b), the MOSFETs’ $g_m$ increases two or three times at LHT, which means great advantages for analog circuits design comparing to RT. But because of the substrate Fermi potential increasing at cryogenic temperature, a higher threshold voltage $V_{th}$ can be found in Fig.1(c). Also, we find that the ameliorated ON/OFF ratio increases at LHT, a steeper subthreshold swing (SS).

SS is the most important parameter in low-power design and it’s proportional to the slope factor $n$ and the thermal voltage $U_T(kT/q)$.

$$SS = nU_Tln10$$

Usually $n$ lies between 1.2 and 1.5 for bulk technology at RT, and is somewhat smaller in silicon on insulator (SOI) transistors. At 4.2K temperature, SS predicts to be 71 times smaller than RT but because of the change of the density of the interface-traps, SS reduces to only 2-3 times smaller than RT.

We have characterized 110nm LVT-MOSFETs at RT and LHT and have built the BSIM4 model based on these commercial LVT devices. These device under test (DUT) have lower $V_{th}$ than conventional devices, so we can get a fast switching speed even though $V_{th}$ increases at LHT. According
As a beginning, we tabulate some parameters, such as transit frequency $f_T(g_m/2\pi C_{gg})$, intrinsic gain $g_m/g_{ds}$, capacitance scale $C_{gs}/g_{gg}C_{dd}/g_{gg}$ and current density $J_D(I_d/W)$ for different channel lengths and $g_m/I_d$. All these variables are width-independent quantities. In the WI level, $g_m/I_d = 1/nU_T$ and the lowest power consumption can be written as follows:

$$I_{d,min} = \frac{g_m}{(g_m/I_d)_{max}} = g_m \cdot nU_T$$

Fig. 2(a,b) shows that the MOSFETs get a higher $g_m/I_d$ under the same inversion level at 4.2K temperature. In other words, cryogenic analog circuits may consume less static power than room temperature operating at WI region or MI region.

Fig. 3(a,b) shows two most important figures of merit of n-channel device. It shows the relationship of $g_m/g_{ds}$ VS $g_m/I_d$ and $f_T$ VS $g_m/I_d$ at 300K and 4.2K.

According to the unity-gain frequency (GBW) requirement, we get the $g_{m,2} = 2\pi \cdot GBW \cdot C_L$, then pick a reasonable channel length $L$, for a short channel we get the high $f_T$, and for a long channel, we get high intrinsic gain $g_m/g_{ds}$. Then pick the $g_m/I_d$ value, for a large value, we can save the power and the $V_{OV}$ will be small and get large signal swing. According to the requirement of SR and $C_L$, we get the current value, check the lookup table of the $I_d/W$ and $g_m/I_d$, we will get the width of the transistor. In another way, we can determine $I_d$ from $g_m$ and $g_m/I_d$, then determine $W$ from $I_d/W$.

### III. $g_m/I_d$ DESIGN METHOD

In the WI region, the current is due to diffusion and mostly due to drift in SI region; $V_{OV}$ has played a central role in classical square-law design, it determines which region the MOSFETs in. But this simple model is obsolete with modern nanoMOSFET transistors. Modern EKV and BSIM transistor models are extremely complicated, they usually have thousands of parameters and the basic first order IV-behaviour can’t match the simulation results.

Some researchers present that analog circuit designers can use the inversion coefficient IC for classifying operation regions of a MOSFET [4].

$$ \begin{cases} IC < 0.1 & \text{weak inversion (WI)} \\ 0.1 < IC < 10 & \text{moderate inversion (MI)} \\ IC > 10 & \text{strong inversion (SI)} \end{cases} $$

and

$$ IC = \frac{I_D}{I_{spec}} \bigg|_{saturation} = \frac{I_D}{2nU_T^2 \mu C_{ox} W/L} = \frac{L}{2nU_T^2 \mu C_{ox}} J_D $$

and $I_{spec}$ is the product of specific currents per square $I_{spec||}$ and $W/L$. But the IC parameter is still connected with the technology variables. It’s not the best solution for cryogenic analog design to maintain a model-agnostic approach.

We use $g_m/I_d$ as a proxy, it determines the MOSFET’s inversion level, it also controls some most important Figures-of-Merit that analog circuit designers care about.

$$ \frac{g_m}{I_d} = \frac{1}{I_D} \frac{\partial D}{\partial V_G} = \frac{2}{V_{OV}} $$

For long-channel MOSFETs, the $g_m/I_d$ is invariant of the technology. Short channel effect (SCE) degrade $g_m/I_d$ because of the velocity saturation [6].

### IV. PROPOSED OTA

Based on the 110nm cryo-CMOS BSIM4 model, we have designed a high performance and low power consumption...
OTA by the $g_m/I_d$ method used for a 14-bit Pipelined-SAR ADC. This ADC is used as a readout subsystem for the semiconductor QC platform, which includes a 7-bit SAR ADC, a residue amplifier and a 8-bit SdAR ADC. As illustrate in Fig. 4, the first 7b ADC quantify the input analog data as the first seven bits, and the residue is amplified by the closed-loop residue amplifier to the second 8-bit SAR ADC. Then the residue is quantified as the second eight bits. A digital error correction (DEC) module converts the 15-bit digital data to the final 14-bit digital codes as final output.

This OTA is based on a 110nm BSIM4 model and need to satisfy the following specifications:

- power supply $AVDD=1.5V$
- operating temperature 4.2K
- open-loop gain $A_{V0}>84.5$dB
- unity gain frequency $GBW >993$MHz
- capacitive load $C_L>2.3pF$

The nondominant poles lie in so high frequency that telescopic and folded-cascode topologies have faster speed than two-stage structure, and they can be just considered as core candidates. Folded-cascode accommodate a wider input common mode voltage range, take into consideration of the high GBW, we choose folded-cascode as the main OTA structure. In this project, we use NMOS as input transistor, though it has lower second nondominant pole than PMOS input, it has a higher input $g_m$ than PMOS under same current, which is the key solution for saving power consumption. The nondominating pole $P_2$ can be write as:

$$g_{m5,6}/2\pi(C_{DD1,2} + C_{DD3,4} + C_{GS5,6} + C_{parP-Booster})$$

$C_{parP-Booster}$ means the P-booster amplifier’s input parasite capacitance. Fig. 5 show the structure of the main fully differential folded-cascode OTA circuit. Using this gain-enhancement technique we can get a even bigger $Z_{out}$ and the total gain is

$$g_{m1,2} \{ (1+A_{bp})g_{m5,6}r_{o5,6}r_{ao3,4}) \{ (1+A_{bn})g_{m7,8}r_{o7,8}r_{ao9,10} \}$$

$A_{bp}$ and $A_{bn}$ means the open-loop gain of p-booster and n-booster amplifiers.

![Fig. 3. 300k and 4k temperature gm/gds_gm/id and ft_gmid](image)

![Fig. 4. the 14 bit Pipelined-SAR ADC structure and residue amplifier](image)

![Fig. 5. Struture of the gain boosting-OTA](image)

There are two non-overlapping clocks in the design of Pipelined-SAR ADC, so the CMFB modual of the main OTA can use the switch capacitance for convenience. When PHI2=1, the quantity of electric charge is

$$Q_1 = 2C_2(V_{cm} - V_{bias3}) + C_1(V_{outp} + V_{outn} - 2V_{cmfb})$$

and when PHI1=1, the quantity of electric charge is

$$Q_2 = (C_1 + C_2)(V_{outp} + V_{outn} - 2V_{cmfb})$$

Let $Q_1 = Q_2$, we realize the purpose of extracting the common mode voltage $(V_{outp} + V_{outn})/2$ and comparing it with a known $V_{cm}$.

$$V_{cmfb} = \frac{V_{outp} + V_{outn}}{2} - V_{cm} + V_{bias3}$$
As for the structure of booster amplifiers, we can take the p-booster as an example. Set input common mode voltage is $V_{pi}$, and output common mode voltage is $V_{po}$. In the main folded cascode OTA we can get

$$V_{pi} = V_{DD} - V_{DS3,4}$$

$$V_{po} = V_{pi} - V_{GS5,6}$$

and

$$V_{po} = V_{pi} - |V_{thp}| - V_{OV5,6}$$

In view of this limiting condition, the basic 5-T and telescope structure cannot be chosen as the solution for booster amplifiers.

Fig. 6 demonstrates the structure of the p-booster and n-booster amplifier and the CMFB circuits. It’s a fully differential folded cascode as same as the main OTA. Because of the input voltages of the booster amplifiers change little, so we choose the continue time-CMFB circuits. To avoid the doublets slowing the setting time caused by the p-booster and n-booster amplifiers, the unity-gain frequency of the booster amplifiers must be in the range of $f_{-3db}$ of the closed-loop amplifier and the nondominant pole $P_2$ of the main OTA [7].

$$\beta \omega_m < \omega_{booster} < \omega_{P_2}$$

$\beta$ is the feedback factor and $\beta \omega_m$ is the closed-loop $f_{-3db}$ bandwidth. $\omega_{P_2}$ is the first nondominant pole of the main OTA.

V. POST SIMULATION RESULTS

Bode-plot measurements show a open-loop gain of 101dB and a unity-gain frequency $\omega_m$ of 1.23GHz (2.3pF load) in Fig.7. And the power consumption is as low as 2.77mW. Setting measurements with a feedback factor $\beta$ of 1/33 show a fast single-pole settling behavior corresponding to a closed-loop $f_{-3db}$ of 37MHz and a setting accuracy better than 0.01%.

VI. CONCLUSION

REFERENCES

[1] E. Charbon, F. Sebastiano, M. Rabaei, A. Vladimirescu, M. Shahmohammadi, R. B. Staszewski, H. A. Hommel, B. Patra, J. P. Van Dijk, R. M. Incandela et al., “15.5 cryo-cmos circuits and systems for scalable quantum computing,” in 2017 IEEE International Solid-State Circuits Conference (ISSCC). IEEE, 2017, pp. 264–265.

[2] A. Beckers, F. Jazaeri, and C. Enz, “Characterization and modeling of 28-nm bulk cmos technology down to 4.2 k,” IEEE Journal of the Electron Devices Society, vol. 6, pp. 1007–1018, 2018.

[3] Y. Zhang, J. Xu, T.-T. Lu, Y. Zhang, C. Luo, and G. Guo, “Hot carrier degradation in mosfets at cryogenic temperatures down to 4.2 k,” IEEE Transactions on Device and Materials Reliability, vol. 21, no. 4, pp. 620–626, 2021.

[4] A. Beckers, F. Jazaeri, H. Bohuslavskyi, L. Hutin, S. De Franceschi, and C. Enz, “Design-oriented modeling of 28 nm fdsoi cmos technology down to 4.2 k for quantum computing,” in 2018 Joint International EUROSOI-ULIS. IEEE, 2018, pp. 1–4.

[5] P. G. Jespers and B. Murmann, Systematic design of analog CMOS circuits. Cambridge University Press, 2017.

[6] C. Enz, M.-A. Chalkiadaki, and A. Mangla, “Low-power analog/RF circuit design based on the inversion coefficient,” in ESSCIRC Conference 2015-41st European Solid-State Circuits Conference (ESSCIRC). IEEE, 2015, pp. 202–208.

[7] K. Bult and G. J. Geelen, “A fast-settling cmos op amp for sc circuits with 90-db dc gain,” IEEE Journal of Solid-State Circuits, vol. 25, no. 6, pp. 1379–1384, 1990.