5nm FinFET Cryogenic SRAM Evaluation for Quantum Computing

Shivendra Singh Parihar[1,2], Simon Thomann[1], Girish Pahwa[3], Yogesh Singh Chauhan[2], and Hussam Amrouch[1]

[1]*Chair of Semiconductor Test and Reliability (STAR), University of Stuttgart, Stuttgart, 70569 Germany*
[2]*Indian Institute of Technology Kanpur, Kanpur, 208016, India*
[3]*University of California, Berkeley, Berkeley, CA 94720 USA*
Email: parihasa@iti.uni-stuttgart.de

Abstract: The tight power and timing constraints at cryogenic temperatures necessitate ultra low power on-chip memory with minimal latency for efficient scaled-up quantum and exascale computing. We present the 5 nm technology-based 6T Static Random Access Memory (SRAM) performance evaluation for quantum computers. We also demonstrate that an SRAM designed for iso-I_{OFF} operation results in 17 % to 21 % delay improvement at a constant supply voltage.

Introduction: The recent demonstrations of state-of-the-art Quantum Computers (QC) highlight the need for on-chip SRAM [1]. An on-chip memory at 10 K can significantly reduce the thermal gradient between qubits operating at sub-mK and control circuitry operating at 300 K [2]. Hence, enhancing the energy efficiency and operating speed of the SRAM for QC is crucial. State-of-the-art CMOS technologies deliver the best power and performance till date [3, 4, 5]. However, CMOS circuit efficiency is mainly limited by the leakage current and supply voltage (V_{DD}). Cryogenic operation delivers lower leakage, higher performance and energy efficiency [2, 5]. In this study, we explore the cryogenic performance of 5 nm SRAM, which is not available in contemporary literature.

Experimental Setup: A minimum channel length multi–fin, multi–finger n-FinFET, and p-FinFET from a commercial 5 nm technology were characterized at room and cryogenic temperatures. A cryogenic aware BSIM-CMG model [6] is calibrated for accurate SPICE simulations, as shown in Figs. 1(a) and 1(b). In the SRAM array, we utilize 32 cells (rows) in a column, a pre-charge circuitry, write driver, sense amplifier, and a latch per column. At cryogenic temperatures, the metal wire resistance decreased significantly (33 % at 77 K); however, the gate capacitance remains constant [7]. The metal wire resistance and capacitance values presented in [7] are utilized in this study.

Performance Evaluation: At 10 K, threshold voltage (V_{TH}) increases by 92 mV and 87 mV, and sub-threshold swing (SS) improves by 78 % and 83 % (16.7 and 12.7 mV/decade at 10 K) for p-FinFET and n-FinFET, respectively (Fig. 1(a)). Fig. 1(c) shows the 40 % ON current (I_{ON}) improvement under the iso-I_{OFF} operation (i.e., I_{OFF} at 10 K is the same as I_{OFF} at 300 K), which can be obtained by gate material work function engineering. The V_{DD} can be reduced by 200 mV from nominal V_{DD} ($V_{DD,nom}$ = 0.75 V) to achieve the iso-I_{OFF}, and iso-performance (similar I_{ON} at 10 K for both cases, with and without work function engineering). Figs. 2(a) to 2(c) show that with the temperature reduction from 300 K to 10 K, Hold Static Noise Margin (HSNM) and Read Static Noise Margin (RSNM) increase by 21.40 % and 27.96 %, respectively and Write Static Noise Margin (WSNM) decreases by 8.42 % at $V_{DD,nom}$. The High-Density Cell (HDC), Low-Voltage Cell (LVC), and High-Performance Cell (HPC) exhibit the highest HSNM, RSNM, and WSNM, respectively. Fig. 2(d) shows that in the HDC, reducing the V_{DD} for iso-performance at iso-I_{OFF} condition degrades the HSNM and RSNM by 50 % and 87 %, respectively, and improves the WSNM by 10.87 %. Although both n-FinFET and p-FinFET have nearly identical current at 300 K and 10 K (Figs. 1(a) and 1(b)), a reduced Bit Line (BL) resistance results in lower read delay at 10 K (Fig. 3(a)). The write performance is hindered by the write driver performance resulting in almost no improvement in write delay at 10 K (Fig. 3(b)). Reduced parasitics also help in lowering the power consumption for both read and write operations (Figs. 3(c) and 3(d)). However, an iso-I_{OFF} operation significantly increases the I_{ON} and hence the dynamic read and write power consumption at a constant V_{DD} (Figs. 3(c) and 3(d)). A lower I_{OFF} at cryogenic temperature reduces the leakage power by six orders of magnitude (Fig. 4(a)). Figs. 4(a) to 4(c) show that an iso-I_{OFF} operation at 10 K with reduced V_{DD} decreases leakage power by ~44 %, read and write power by ~59 % compared to the operation at $V_{DD,nom}$ due to the reduction in both static and dynamic power. Decrease in the BL resistance from 20 Ω/cell (at 300 K) to 13.4 Ω/cell (at 10 K) along with the improvement in I_{ON} due to V_{TH} engineering lower the read and write delay by ~16 % at V_{DD} = 0.55 V as shown in Fig. 4(d). A smaller delay and lower operating voltage of 0.55 V significantly improve the overall energy consumption of the SRAM array and result in a minimum of 19 %, and 35 % improved read and write energy efficiency (Fig. 5(a)).

Conclusion: Fig. 5(b) summarizes various SRAM benchmark parameters at 300 K and 10 K. The 6T HDC with engineered V_{TH} at optimized V_{DD} exhibits superior delay and energy efficiency at 10 K. These results suggest that 5 nm FinFETs-based SRAM is suitable for cryogenic interface circuits in QC.

Acknowledgement: Cryogenic measurements were facilitated by the CRF IIT Delhi, India.

[1] F. Sebastiano et al., IEDM Tech. Dig., p. 25.2.1, (2020).
[2] P. Wang et al., IEDM Tech. Dig., p.38.5.1, (2020).
[3] G. Yeap et al., IEDM Tech. Dig., p. 36.7.1, (2019).
[4] B. Choi et al., VLSI, p. 1, (2021).
[5] H. L. Chiang et al., VLSI, p. 1, (2020).
[6] G. Pahwa et al., IEEE-TED, vol. 68, p. 4233, (2021).
[7] V. P. -H. Hu et al., IEDM Tech. Dig., p. 8.6.1, (2021).

2023 Device Research Conference (DRC) | 979-8-3503-2310-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/DRC58590.2023.10186959

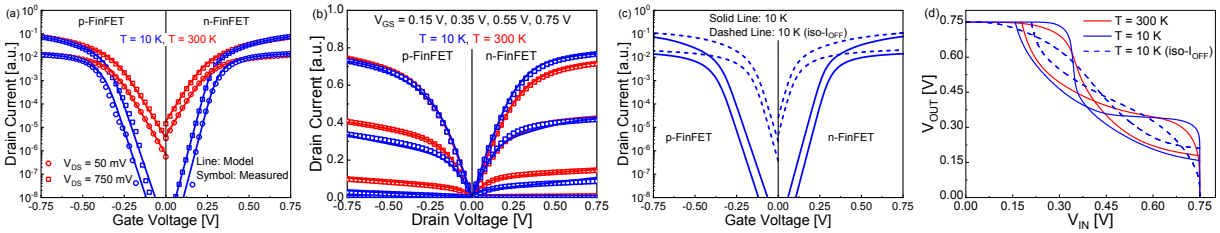


Fig. 1: Transistor characterization at 10 K and 300 K: (a) $I_{DS} - V_{GS}$ of characterized p-FinFET and n-FinFET, (b) $I_{DS} - V_{DS}$ from 150 mV to 750 mV at 200 mV steps. (c) Simulated $I_{DS} - V_{GS}$ at 10 K for iso-I_{OFF} at V_{DS}= 50 mV and V_{DS}= 750 mV. (d) BFC of HDC during read operation down to 10 K. SNM improves at cryogenic temperatures; however, at iso-I_{OFF} condition, reduction in V_{TH} results in the SNM degradation with temperature.

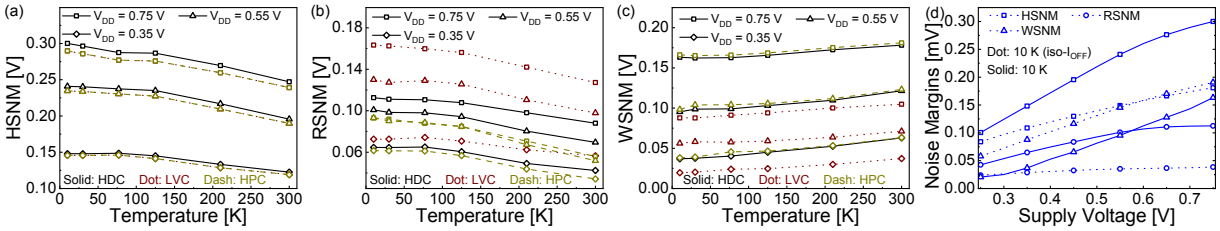


Fig. 2: Impact of V_{DD} on: (a) HSNM, (b) RSNM, and (c) WSNM of different cell types for temperature ranging from 10 K to 300 K, (d) HSNM, RSNM and WSNM of HDC at 10 K with and without V_{TH} engineering.

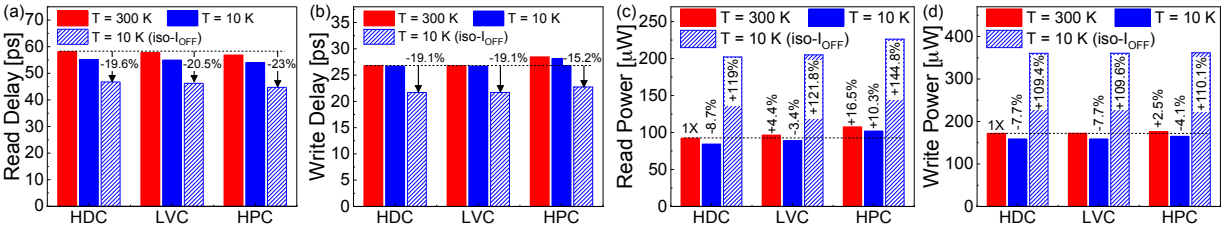


Fig. 3: Impact of cryogenic temperature on (a) read delay, (b) write delay, (c) read power and (d) write power. Read delay is the time between WL activation and when latch output reaches 90 % of target potential. Write delay measurement starts at write driver activation and stops at the 90 % of the target voltage of respective bit line.

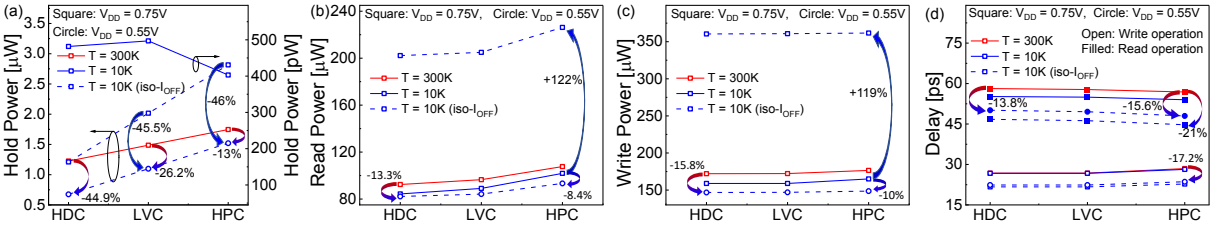
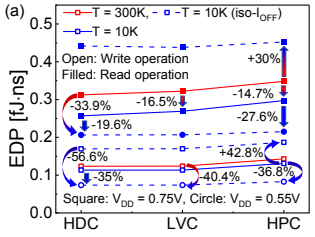


Fig. 4: Power consumption during the (a) hold, (b) read, and (c) write operation. (d) Delay during the read and write operation. The iso-I_{OFF} operation at lower V_{DD} (0.55 V) offers ~10 % power and 16 % speed improvement for HPC.



(b) SRAM

Temperature	HDC				LVC				HPC			
	300 K	10 K	10 K (iso-I_{OFF})	10 K (iso-I_{OFF})	300 K	10 K	10 K (iso-I_{OFF})	10 K (iso-I_{OFF})	300 K	10 K	10 K (iso-I_{OFF})	10 K (iso-I_{OFF})
Supply Voltage	0.75 V	0.75 V	0.75 V	**0.55 V**	0.75 V	0.75 V	0.75 V	**0.55 V**	0.75 V	0.75 V	0.75 V	**0.55 V**
Read Energy [fJ]	5.37	4.65	9.45	4.12	5.57	4.90	9.48	4.17	6.12	5.50	10.10	4.48
Write Energy [fJ]	4.61	4.24	7.82	3.29	4.62	4.24	7.83	3.29	5.02	4.65	8.22	3.51
Read Delay [ps]	58.11	55.19	46.73	50.11	57.77	54.96	46.23	49.54	56.87	54.00	44.71	47.99
Write Delay [ps]	26.79	26.71	21.69	22.39	26.79	26.71	21.69	22.39	28.46	28.16	22.73	23.58
Read EDP [fJ * ns]	0.312	0.256	0.441	0.206	0.322	0.269	0.438	0.207	0.348	0.297	0.452	0.215
Write EDP [fJ * ns]	0.123	0.113	0.170	0.073	0.123	0.113	0.170	0.073	0.143	0.130	0.187	0.082
Leakage Power [μW]	1.22	4.82e-4	1.21	0.675	1.49	4.97e-4	2.02	1.10	1.75	4.03e-4	2.82	1.52

Fig. 5: (a) The iso-I_{OFF} operation at 10 K with reduced V_{DD} results in a minimum 19 % and 35 % improvement in EDP during the read and write operation, respectively for the HDC. (b) Cryogenic benchmark results of HDC, LVC, and HPC. The HDC provides lowest EDP and leakage power at 10 K followed by LVC and HPC, respectively.

979-8-3503-2310-8/23/$31.00 ©2023 IEEE