



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org



www.ijasem.org

A BENCHMARK OF CRYO-CMOS EMBEDDED SRAM/DRAMs IN 40-nm CMOS

M.GAYATHRI¹, G. LAKSHMI DEVI², R.L.B.R. PRASAD REDDY³

¹PG Student, Dept of ECE, SITS, Kadapa.

²Assistant Professor, Dept of ECE, SITS, Kadapa.

³Associate Professor, Dept of ECE, SITS, Kadapa.

Abstract –

The interface electronics needed for quantum processors require cryogenic CMOS (cryo-CMOS) embedded digital memories covering a wide range of specifications. To identify the optimum architecture for each specific application, this article presents a benchmark from room temperature (RT) down to 4.2 K of custom SRAMs/DRAMs in the same 40-nm CMOS process. To deal with the significant variations in device parameters at cryogenic temperatures, such as the increased threshold voltage, lower subthreshold leakage, and increased variability, the feasibility of different memories at cryogenic temperature is assessed and specific guidelines for cryogenic memory design are drafted. Unlike at RT, the 2T low-threshold-voltage (LVT) DRAM at 4.2 K is up to $2\times$ more power efficient than both SRAMs for any access rate above 75 kHz since the lower leakage increases the retention time by $40\,000\times$, thus sharply cutting on the refresh power and showing the potential of cryo-CMOS DRAMs in cryogenic applications.

Keywords – Cryogenic CMOS (cryo-CMOS), DRAM, eDRAM, memory, quantum computing, SRAM.

I. INTRODUCTION

QUANTUM computers (QCs) can deliver an exponential speedup for several computational problems. However, scaling up the number of quantum bits (qubits) to the thousands or millions necessary for useful computations requires an impractical amount of wires connecting the cryogenic qubits to the room-temperature (RT) control electronics. To overcome such an interconnect bottleneck, electronics integrated in commercial CMOS technology but operating at cryogenic temperature, i.e., cryogenic CMOS (cryo-CMOS), has been proposed. As the power consumption of the cryo-CMOS control electronics must be kept below the cooling power of the cryogenic refrigerators adopted in QC applications,

designing power-efficient cryo-CMOS circuits is crucial. The control electronics consist of analog/RF circuits directly interfacing with the qubits to perform operations and measurements, in combination with the digital system-on-chip (SoC) for scheduling the quantum-algorithm execution and processing a large amount of measurement results, e.g., as required for quantum error correction. In modern digital systems, significant fractions of the area and power are consumed by memory, thus making the optimization of cryo-CMOS embedded memories essential. However, accurately estimating the power consumption of a memory at cryogenic temperatures is challenging due to the lack of reliable cryogenic device models. Cooling down to cryogenic temperatures affects the characteristics of short-channel NMOS and PMOS transistors by increasing their threshold voltage V_{th} (100–200 mV), subthreshold slope ($\sim 3\times$ steeper), and carrier mobility ($\sim 2\times$ for low-field mobility). Additionally, the mismatch between devices increases, as shown in and for 40-nm bulk CMOS and 28-nm bulk CMOS, respectively, interconnect resistance drops ($\sim 30\%$), and the capacitance of source/drain junctions decreases due to wider depletion regions due to freeze-out. For analog circuits, this results in an increased bandwidth and reduced power consumption. For full-swing digital circuits, the mobility increase compensates the effects of the larger V_{th} and, together with the reduced resistance and capacitance, results in a speed-up for digital circuits from 10% to 20% for 40-nm bulk CMOS. For more advanced technology nodes, the speed-up from RT to 4.2 K is reduced due to the increased relative importance of interconnect capacitance and lower supplies, enhancing the relative V_{th} increase. However, the speed-up could be recovered for FET technologies by scaling V_{th} . The increased V_{th} and the steeper subthreshold slope lead to severely reduced subthreshold leakage, while gate leakage stays approximately constant ($< 2\times$ smaller). Still, the large variety in memory architectures, adopted CMOS processes, and temperature ranges in those prior works hinders the compilation of a fair comparison. Thus, identifying the best memory design in terms of area and power for each memory application is still a challenging open question. To overcome this issue, this work compares eight different dynamic and static memory cell designs, embedded in identical memory architectures in a nanometer CMOS process (TSMC 40-nm) typically adopted for QC cryo-CMOS interfaces, by comparing the experimental characterization at both RT and 4.2 K. Due to the limited cooling power available in dilution refrigerators, the main focus is on minimizing the memory power consumption. Since the power consumption of the dynamic memories is limited by their refresh power for medium-to-high frequency applications, a detailed characterization of the data-retention time is required for these cells.

II. SRAM AND DRAM DESIGN SYSTEM

The memory cells in this work have been mainly optimized for maximum density, and, where possible, for optimum (expected) performance at cryogenic temperature. All memory cells are implemented in two versions, using either standard-threshold-voltage (SVT) or low-threshold-voltage (LVT) devices. LVT cells are expected to perform worse at RT since their higher sub-threshold leakage reduces the retention time of dynamic memories and increases the static power consumption. At cryogenic temperatures, however, the V_{th} increase may cause SVT designs to fail due to the insufficient overdrive voltage limiting the readout currents. Although forward-biasing the bulk–source voltage could help circumvent the cryogenic V_{th} increase, no individual bulk contacts have been employed to avoid an excessive increase in the design effort and the area of the memory cells. The memory peripherals always use LVT devices, unless otherwise noted, to ensure functionality at cryogenic temperatures and minimize their effect on memory performance, while the synthesized digital circuits, e.g., the controllers, adopt SVT devices with extra hold margin to anticipate the cryogenic logic speed- up.

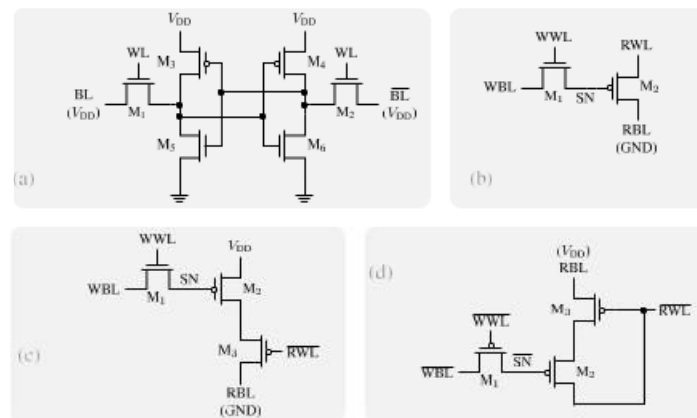


Fig.1. Schematics of the four cell designs: (a) 6T static cell; (b) 2T NW-PR dynamic gain cell; (c) 3T NW-PR dynamic gain cell; and (d) 3T PW-PR preferentially boosted dynamic gain cell.

As the most commonly used embedded-memory cell, the conventional six-transistor static cell [6T, Fig. 1(a)] represents a good reference for comparison with alternative designs. It consists of a latch formed by two inverters ($M3-6$) and two access transistors ($M1,2$) that connect the latch nodes to the differential bit lines (BLs) (BL and $\overline{BL}$). The latch state is written by differentially driving the BLs and pulling the word line (WL) high. To read the state, both BLs are first pre charged to V_{DD} before enabling the WL. Then, the BL connected to the low side of the latch will be discharged by one of the pull-down transistors ($M5,6$). To

minimize the cell area, most transistors have minimum size ($W/L = 120 \text{ nm}/40 \text{ nm}$). Since the cell design is ratio, the pull-down transistors ($M5,6$) are sized $1.5\times$ larger ($W/L = 180 \text{ nm}/40 \text{ nm}$) to ensure writing and reading under device mismatch. For a fair comparison with the other cells, the static cell is manually implemented using the logic design rule check (DRC) rule set and occupies $0.435 \mu\text{m}^2$ using a lithographically symmetrical layout. This is 80% larger than the foundry-offered cells ($0.242 \mu\text{m}^2$) that violate several logic DRC rules.

III. EXISITING SYSTEM

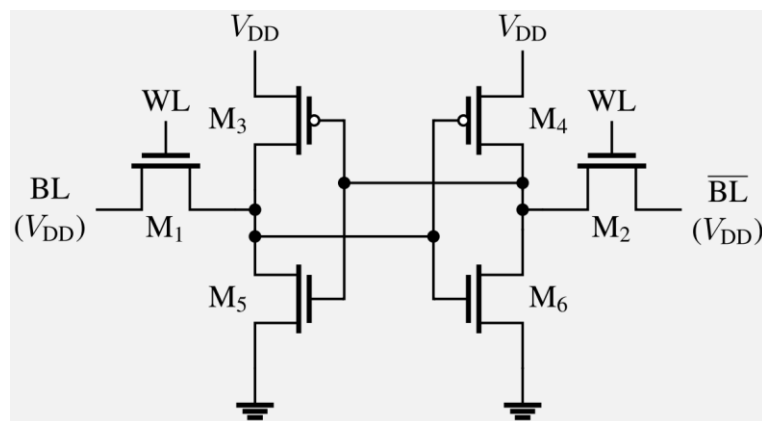


Fig.2. Schematic Diagram of 6T SRAM

In the existing system technique with configurable multiple boost planes to implement low-power 6T SRAM. Measurement results show that the proposed 6T SRAM demonstrated stable performance from room temperature to 6 K, achieving extremely low minimum operating voltages of 0.23V and 0.31V at room temperature and 6 K, respectively. Despite the optimization of power consumption, more research works focused on improving memory density. Presenting a 6T SRAM Cell Analysis in Various Technologies to analyse the Static Noise margin of the cell. Together with the read channel's isolation from the real internal storage nodes, this takes out the read-disturbance. Additionally, it uses a write-assist mechanism to carry out its write operation in pseudo differential form using a write bit line and control signal. In this study, a 6T SRAM cell for 90nm and 180nm technologies has been created. The Cadence Virtuoso tool has been applied to modelling and design. Using nmos1V and nmos2V cells in 180nm and 90nm technologies, the static noise margin for SRAM has been determined. SNM decreases with shrinking technology, as expected, and CMOS 1v transistors have been found to have better SNM than CMOS 2v transistors. The static random-access memories (SRAM) are most widely used, due to their high performance: microprocessors may contain up to 70% of SRAMs in transistor count or area. The trend in

the semiconductor market is to push for more integration and more sized reduction: the development and optimization of a technological node is more and more difficult and expensive. The reduction in size of a SRAM circuit in coming nodes is nonetheless complex and it faces several limitations. The reliability of the SRAM bit-cell is degraded with ever smaller technologies and the device functionality is endangered. Designing SRAM circuits in CMOS 65nm requires technical and technological solutions to overcome the size reduction limitations, while ensuring satisfactory functionality, with a guaranteed reliability so that it can be economically fabricated. The manufacturing of a standard SRAM is fully compatible with CMOS core processes. The standard SRAM bit-cell is based on a 6-transistor arrangement: it is called a 6T1SRAM. Two CMOS inverters, formed by PU and PD transistors, are connected, one opposite to the other, and two access transistors, PG transistors, are added (figure 2). Three possible operations are: writing a bit data, retaining the bit data and reading the bit data. The operation is controlled through the word-lines that activate or block the access transistors PG, so that there is, or not, a connection to the bit-lines BLT and BLF that propagate the bit value from or to the bit-cell.

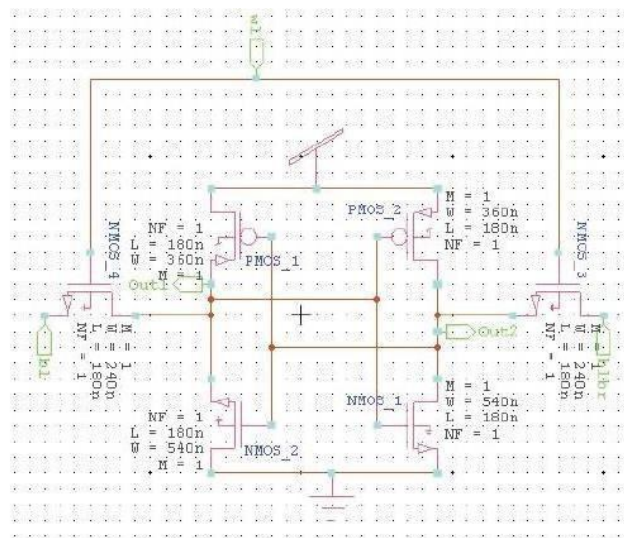


Fig.3. Existing System Implementation in SPICE Simulation Tool.

IV. PROPOSED SYSTEM

In the proposed system Memory Circuits Volatile and non-volatile memories are two major categories of CMOS-based memory. Volatile memories, such as static random access memory (SRAM) and dynamic RAM (DRAM), exhibit improved performance metrics at cryogenic temperatures due to reduced leakage currents and enhanced carrier mobility. Additionally, the high compatibility of volatile memory with silicon-based CMOS processes

facilitates achieving high levels of integration. Due to its data stability and high-speed advantages, SRAM has become the most widely used memory topology.

4T SRAM Circuit

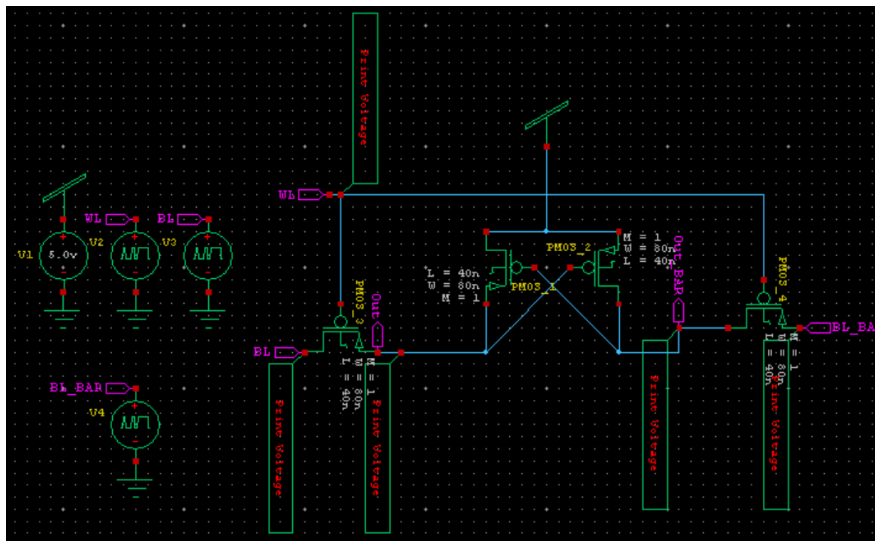


Fig.4. schematic of 4T SRAM

Above is the schematic of the proposed 4T SRAM cell, which is implemented using SPICE simulation tool. The SPICE implementation serves as a powerful tool for simulating the behaviour of the SRAM cell under various operating conditions, enabling through analysis and optimization of its performance characteristics.

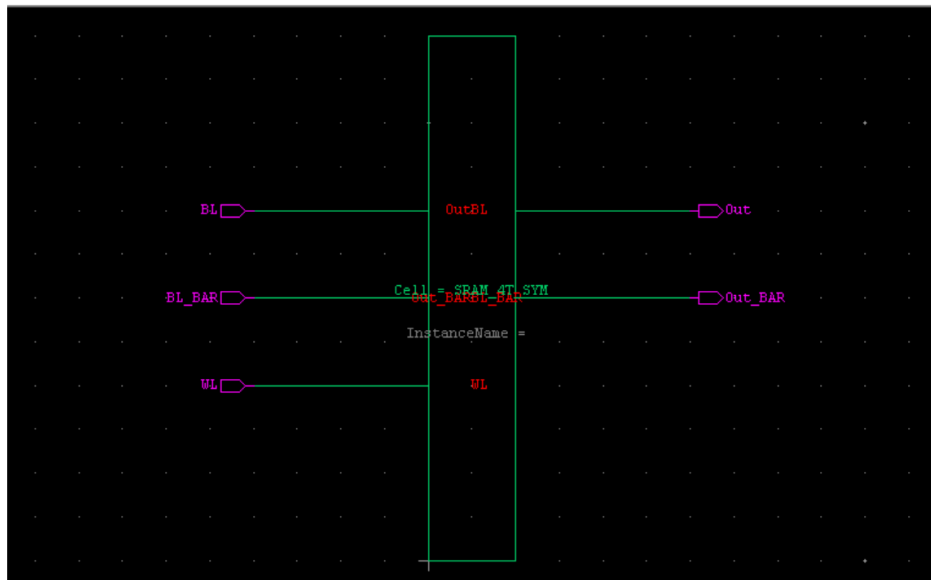


Fig.5. Block Diagram of 4T SRAM

The proposed 4T SRAM structure optimized for operation at 77 K by eliminating two pMOS transistors from the pull-up network. The 4T SRAM achieved a reduction of cell

areaby 20.3%, compared to the standard 6T SRAM structure. Moreover, it also provided faster readand write operations. Compared to SRAM, DRAM can achieve higher density due to its compact bit cell layout. However, DRAM requires frequent refresh operations to maintain datacorrectness, resulting in extra area and power overhead. The additional costs associated with DRAM make it less attractive for memory implementation at room temperature. Benefitingfrom the optimized leakage current and carrier mobility at cryogenic temperature, the data retention time and read/write operations of DRAM can be further improved.

V. RESULTS

Proposed System Results:

The simulation results of proposed 4TSRAM cell using CMOS is shown below. When WL (write line) = 0 the PMOS_3 and PMOS_4 turn ON, when the input BL(bit line) = 1, BL_BAR = 0 the output of PMOS_3 Out = 1, the output of PMOS_4 Out_BAR = 0.

By giving input to the circuit we are performing write operation or writing of data into SRAM cell. While, by retrieving data from output portwe are performing read operation in SRAM cell.

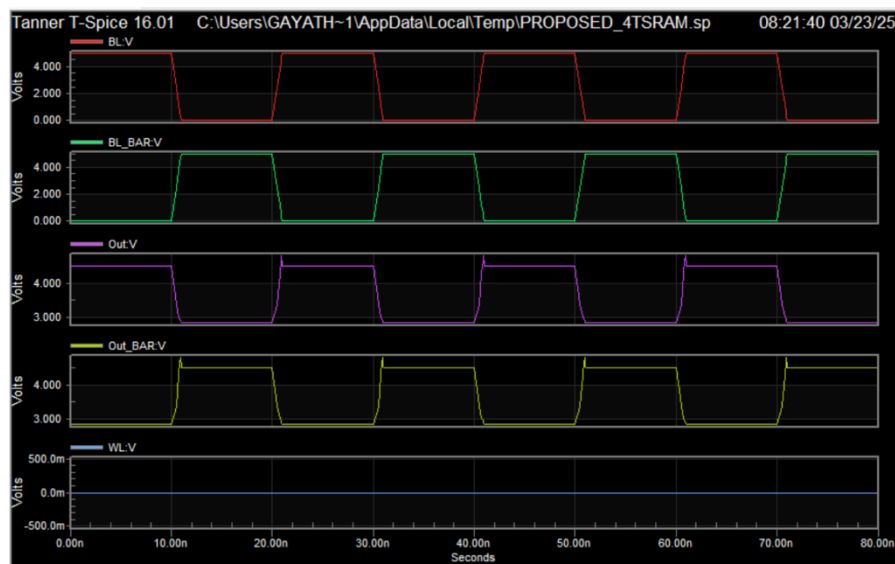


Fig.6. 4T SRAM Wave forms

This analysis provides a visual representation of the dynamic behavior of the 4T SRAM cell during write and read operations, aiding in understanding its functionality and performance characteristics.

In different temperatures how 4T SRAM works:

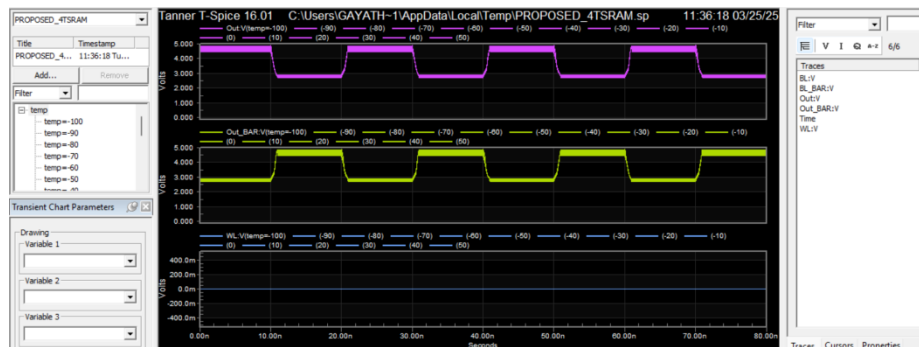


Fig.7. 4T SRAM waveforms at different temperatures.

At lower temperatures like -100°C the average power consumed by 4T SRAM is less i.e 28.42mw.

At higher temperatures like 10°C the average power consumed by 4T SRAM is more i.e 90.96mw.

Below figure shows the power results of 4T SRAM at different temperatures.

```
Power Results: temp=-100

VV3 from time 0 to 8e-008
Average power consumed -> 2.888787e-002 watts
Max power 6.076169e-002 at time 4.1e-008
Min power 0.000000e+000 at time 1.1e-008

VV4 from time 0 to 8e-008
Average power consumed -> 2.796152e-002 watts
Max power 6.076257e-002 at time 7.1e-008
Min power 0.000000e+000 at time 0

Power Results: temp=-90

VV3 from time 0 to 8e-008
Average power consumed -> 3.516462e-002 watts
Max power 7.394075e-002 at time 7e-008
Min power 0.000000e+000 at time 1.1e-008

VV4 from time 0 to 8e-008
Average power consumed -> 3.403743e-002 watts
Max power 7.394073e-002 at time 2e-008
Min power 0.000000e+000 at time 0

Power Results: temp=10

VV3 from time 0 to 8e-008
Average power consumed -> 9.244691e-002 watts
Max power 1.942317e-001 at time 7e-008
Min power 0.000000e+000 at time 1.1e-008

VV4 from time 0 to 8e-008
Average power consumed -> 8.948661e-002 watts
Max power 1.942328e-001 at time 2e-008
Min power 0.000000e+000 at time 0
```

Fig.8. Power Results of 4T SRAM at different temperatures.

VI. CONCLUSION

By comparing single-bank static and dynamic memories at cryogenic temperature, these article showsthat well-designed dynamic memories can outperform static memories for middle-to-high frequencyapplications in terms of area and power. The 4T SRAM cell has

been explored and simulated using Tanner Tools at 40nm CMOS. The results showcase its robustness and remarkably low power consumption. Still, adopting dynamic cells with enhanced resistance to gate leakage and cryogenic V_{th} shifts can significantly increase retention time, thus lowering the refresh power. Embracing the design guidelines outlined here for cryogenic embedded memories will facilitate the adoption of dynamic-memory cells for high-density low-power cryogenic memories, thereby enabling the complex cryo-CMOS SoCs needed in future QCs.

REFERENCES

1. P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in Proc. 35th Annu. Symp. Found. Comput. Sci., Nov. 1994, pp. 124–134.
2. L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack," Phys. Rev. Lett., vol. 79, no. 2, p. 325, 1997.
3. A. Montanaro, "Quantum algorithms: An overview," npj Quant. Inf., vol. 2, no. 1, pp. 1–8, 2016.
4. A. W. Harrow and A. Montanaro, "Quantum computational supremacy," Nature, vol. 549, no. 7671, pp. 203–209, Sep. 2017.
5. D. J. Egger et al., "Quantum computing for finance: State-of-the-art and future prospects," IEEE Trans. Quantum Eng., vol. 1, pp. 1–24, 2020.
6. F. Bova, A. Goldfarb, and R. G. Melko, "Commercial applications of quantum computing," EPJ Quantum Technol., vol. 8, no. 1, p. 2, Dec. 2021.
7. F. Sebastiano et al., "Cryo-CMOS electronic control for scalable quantum computing," in Proc. 54th ACM/EDAC/IEEE Design Autom. Conf. (DAC), Jun. 2017, pp. 1–6.
8. B. Patra et al., "Cryo-CMOS circuits and systems for quantum computing applications," IEEE J. Solid-State Circuits, vol. 53, no. 1, pp. 309–321, Jan. 2018.
9. X. Fu, L. Lao, K. Bertels, and C. G. Almudever, "A control microarchitecture for fault-tolerant quantum computing," Microprocess. Microsyst., vol. 70, pp. 21–30, Oct. 2019.
10. P. Wang, X. Peng, W. Chakraborty, A. I. Khan, S. Datta, and S. Yu, "Cryogenic benchmarks of embedded memory technologies for recurrent neural network based quantum error correction," in IEDM Tech. Dig., Dec. 2020, pp. 38.5.1–38.5.4.
11. R. W. J. Overwater, M. Babaie, and F. Sebastiano, "Neural-network decoders for quantum error correction using surface codes: A space exploration of the hardware cost-performance tradeoffs," IEEE Trans. Quantum Eng., vol. 3, pp. 1–19, 2022.

12. P. Das, A. Locharla, and C. Jones, "LILLIPUT: A lightweight low- latency lookup-table decoderfor near-term quantum error correction," in Proc. 27th ACM Int. Conf. Architectural SupportProgram. Lang. Operating Syst. New York, NY, USA: Association for Computing Machinery, Feb.2022, pp. 541–553, doi: 10.1145/3503222.3507707.
13. P. Das et al., "AFS: Accurate, fast, and scalable error-decoding for fault- tolerant quantumcomputers," in Proc. IEEE Int. Symp. High-Perform. Comput. Archit. (HPCA), Apr. 2022, pp. 259–273.
14. F. Battistel et al., "Real-time decoding for fault-tolerant quantum computing: Towards higherdecoding speed and lower communication latency," Bull. Amer. Phys. Soc., vol. 7, no. 3,Aug. 2023, Art. no. 032003, doi: 10.1088/2399-1984/aceba6.
15. J. P. G. van Dijk et al., "A scalable cryo-CMOS controller for the wideband frequency-multiplexedcontrol of spin qubits and transmons," IEEE Sensors J. Solid-State Circuits, vol. 55, no. 11, pp.2930–2946, Nov. 2020.
16. M. Prathapan et al., "A cryogenic SRAM based arbitrary waveform generator in 14 nm for spinqubit control," in Proc. IEEE 48th Eur. Solid State Circuits Conf. (ESSCIRC), Sep. 2022, pp. 57–60.
17. S. Chakraborty et al., "A cryo-CMOS low-power semi-autonomous transmon qubit state controller in 14-nm FinFET technology," IEEE J. Solid-State Circuits, vol. 57, no. 11, pp. 3258–3273, Nov.2022.
18. R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenicictemperatures," IEEE J. Electron Devices Soc., vol. 6, pp. 996–1006, 2018.