High-Density and High-Speed 4T FinFET SRAM for Cryogenic Computing

Vita Pi-Ho Hu[1]*, Chang-Ju Liu[2], Hung-Li Chiang[3], Jer-Fu Wang[1,3], Chao-Ching Cheng[3], Tzu-Chiang Chen[3], and Meng-Fan Chang[3]

[1]Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan, *E-mail: vitahu@ntu.edu.tw

[2]Department of Electrical Engineering, National Central University, Taoyuan, Taiwan

[3]Corporate Research, Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan

Abstract—Cryogenic on-chip memory requires low energy consumption for enabling cryogenic and exascale computing. With FinFET Si data calibration, this work clearly demonstrates that the 4T cryogenic SRAM surpasses various kinds of the 6T SRAM in area, speed, stability, and energy efficiency at both 77K and 300K. Compared to the 6T SRAM with regular threshold voltage in high-density design (6T-RVt-HD) at 300K, the 4T cryogenic SRAM at 77K shows 20.3% cell area reduction, 44% reduction in read access time, 46% improvement in write time, 2.3× improvement in write stability, and 53% reduction in energy-delay product (EDP). Considering the cooling energy consumption at 77K, the 4T cryogenic SRAMs exhibit 33% EDP reduction compared to the 6T-RVt-HD SRAMs at 77K. The proposed high-density and high-speed 4T cryogenic SRAMs with low access energy and superior stability could be promising candidates for cryogenic high-performance-computing (HPC) applications.

I. Introduction

Cryogenic computing is an up-and-coming solution to dramatically improve the computer's performance and power efficiency for high-performance-computing (HPC) data centers and control electronics in quantum processors. The cryogenic characteristics have been extensively studied on planar devices [1]-[4] and FinFET technology at 77K [5] to improve the power, performance, and reliability. At cryogenic temperature, un-doped or lightly-doped FinFET with superior gate control and steep subthreshold slope [5] reduces the leakage power significantly. Moreover, the BEOL metal wire resistance decreases at low temperatures, improving performance and power efficiency. Cryogenic SRAM with an optimized threshold voltage (Vt) and supply voltage (V_{DD}) had been studied for recurrent neural network-based quantum error correction circuitry [6]. As exascale computing is on the horizon, minimizing energy consumption is essential to meet the needs of HPC applications. A cryogenic on-chip memory system operating at 77K bridges the temperature gap between the 4K cryogenic core and the current room-temperature supercomputers. Computing cores more constantly access on-chip SRAM caches. Therefore, it is crucial to improve the density, speed, and energy efficiency of the SRAM caches for cryogenic computing applications.

In this work, high-density 4T FinFET SRAM cells are studied as cryogenic memories compared to 6T SRAM cells at 77K for the first time. Fig. 1 describes the advantages of 4T SRAM cells compared with the 6T SRAM cells for cryogenic computing applications. The static noise margin (SNM) and performance of 6T SRAM cells are sensitive to the Vt design. Reducing V_{DD} decreases the dynamic energy. The 6T SRAM cell with a low Vt design and low V_{DD} operation improves the energy-saving at 77K. However, the 6T SRAM cell designed with a lower Vt may not provide sufficient stability. This work demonstrates that the 4T cryogenic SRAM with a smaller cell area and lower metal wire resistance and capacitance enhances the energy efficiency and the read/write stability simultaneously. Moreover, the 4T cryogenic SRAM improves the read/write speed and EDP compared to the various 6T SRAM cells at 77K.

II. Device Design and Simulation Methodology

Fig. 2 shows the measured I_D-V_{GS} characteristics of 10nm FinFETs [5] fitted with TCAD simulations at 300K and 77K. The temperature-dependent models, including band-gap widening, intrinsic carrier concentration, mobility, saturation velocity, and metal wire resistance, are calibrated in the TCAD simulations. As temperature reduces to 77K, Vt is increased, and the subthreshold swing is significantly improved for nFET (23mV/dec) and pFET (26mV/dec), as shown in Fig. 2. TACD mixed-mode simulations are then performed to analyze the stability and performance of 4T and 6T SRAM cells. For a fair comparison to examine the 6T SRAM cells among various temperatures, all devices at multiple temperatures are re-engineered to have the same Vt at $|V_{DS}|$ = 0.05V. Two Vt scenarios, including regular Vt (RVt, |Vt| = 0.5V) and low Vt (LVt, |Vt| = 0.2V), are considered to investigate the impact of Vt design on the stability and performance of 6T SRAM cells from 300K to 77K. For the high-density 6T SRAM cell (6T-HD), the fin numbers of the pull-up (PU), pull-down (PD), and pass-gate (PG) devices are equal to 1.

III. 6T SRAM Cells at 77K with Two Vt Scenarios

Fig. 3(a) shows that at 300K, the 6T-RVt-HD SRAM shows RSNM = 152mV at V_{DD} = 0.75V and 121mV at V_{DD} = 0.6V, respectively. A high-density 7nm 6T FinFET SRAM had been demonstrated to show RSNM = 102mV at V_{DD} = 0.6V [7], and a 6T double-gate SRAM showed RSNM = 124mV at V_{DD} = 0.8V [8]. In other words, the 6T-RVt-HD SRAM cell in this work exhibits good stability. Fig. 3(a) and 3(b) compare the read static noise margin (RSNM) for 6T-RVt-HD and 6T-LVt-HD SRAM cells. At V_{DD} = 0.75V, as temperature changes from 300K to 77K, the 6T-RVt-HD SRAM shows 6% RSNM increments, while the 6T-LVt-HD SRAM shows 8% RSNM reduction. For 6T SRAM with RVt design (Fig. 3(a)), the benefit of a low subthreshold swing at 77K becomes more significant, which makes the voltage transfer curves sharper during high to low transition, and improves the RSNM at 77K. For the 6T-RVt-HD SRAM cell, the enhanced RSNM due to a steeper subthreshold swing at 77K can be observed for a wide range of V_{DD} (0.4V ~ 0.75 V).

Fig. 3(b) shows that as temperature decreases to 77K, the RSNM of 6T-LVt-HD SRAM decreases at V_{DD} = 0.75V and increases at V_{DD} = 0.4V, respectively. As temperature reduces to 77K, reduced phonon scattering improves mobility [5]. At V_{DD} = 0.75V, the pass-gate operates at the saturation region during reading "0" at the storage node. The output conductance of pass-gate is increased at low temperature [4], thus increasing the read disturb voltage and degrading the RSNM at 77K at V_{DD} = 0.75V. As V_{DD} scales to 0.5V and 0.4V, the benefits of low subthreshold swing become more significant and improves the RSNM of 6T-LVt-HD SRAM at 77K. Fig. 3(c) shows that at V_{DD} =

2021 IEEE International Electron Devices Meeting (IEDM) | 978-1-6654-2572-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEDM19574.2021.9720511

0.75V, the RSNM degradations due to LVt design are more significant at 77K (−55%) than 300K (−40%). Therefore, 6T cryogenic SRAM with LVt design may not provide sufficient RSNM at 77K. Reducing V_{DD} for 6T-LVt-HD SRAM cells improves the energy consumption while lowering V_{DD} will further degrade the read stability. Fig. 4(a) shows that the write static noise margin (WSNM) slightly degrades at 77K. Due to improved subthreshold swing at 77K, the hold static noise margin (HSNM) improves at 77K for both 6T-RVt-HD and 6T-LVt-HD SRAMs, as shown in Fig. 4(b).

IV. 4T Cryogenic SRAM

A. Reduced Cell Area and Lower BEOL Wire RC

For cryogenic SRAM, low-temperature operations can significantly reduce the static energy consumption due to leakage current. However, the dynamic energy consumed due to the charging and de-charging of capacitance during read and write operations remains the same at low temperatures. Decreasing Vt and V_{DD} for 6T cryogenic SRAM reduces the dynamic energy consumption while degrading the read stability simultaneously. We propose 4T load-less cryogenic SRAM (Fig. 5(a)) with a smaller cell area in this work. The 4T cryogenic SRAM with reduced metal line capacitance decreases the dynamic energy consumption while improving the read/write stability and speed. Table I describes the 7nm design rules [9] used in this work. The layout designs of 6T-HD (PU:PG:PD = 1:1:1) and 4T SRAM cells are shown in Fig. 1. Fig. 5(b) shows that the 4T SRAM reduces the cell area by 20.3% compared to the 6T-HD cell. Increasing cell ratio (PU:PG:PD = 1:2:2) is often used to improve the performance while the cell area is increased by 22.3% due to the increased fin numbers, as shown in Fig. 5(b).

Fig. 6(a) compares the metal line capacitance for 6T-HD, 6T(1:2:2), and 4T SRAM cells. The total bitline (BL) capacitance includes the BL wire capacitance and the access transistor junction capacitance. The entire wordline (WL) capacitance consists of the metal wire capacitance and the gate capacitance of access transistors. As temperature changes from 300K to cryogenic temperature, the total gate capacitance of the transistor remains invariant [4]. 4T SRAM with a smaller cell area reduces the WL capacitance by 9.5%. The metal wire resistance reduces by ~33% at 77K [5]. In Fig. 6(b), compared to the 6T-HD SRAM at 300K, the 4T SRAM shows 20% and 47% WL resistance reductions at 300K and 77K, respectively. Fig. 7 shows the I_D-V_{GS} curves for 4T SRAMs at 300K and 77K with dual Vt designs (|Vt| = 0.35V for nFET and 0.2V for pFET). Please be noted that the nFET and pFET for 4T SRAMs need to be re-engineered for stable 4T operations.

B. Static Noise Margin and Speed Analysis

Fig. 8 compares the read, write and hold stabilities among various 6T and 4T SRAM cells at V_{DD} = 0.75V. As temperature changes from 300K to 77K, the 6T-LVt-HD SRAM shows degraded RSNM, while the 4T and 6T-RVt SRAM cells show improved RSNM, as shown in Fig. 8(a). Compared to 6T-RVt-HD SRAM at 300K, the 4T SRAM at 77K shows 5.7% improvement in RSNM. Fig. 8(b) shows that increasing cell ratio (1:2:2) at 77K improves WSNM by 45%. The 4T SRAM at 77K exhibits significant WSNM improvements (2.3×) than the other 6T SRAM cells. Fig. 8(c) shows that the 4T SRAM cell exhibits smaller but adequate HSNM (170mV at 77K).

Fig. 9 illustrates the read access time and write time comparisons. The read access time consists of WL wire delay and BL delay. As temperature reduces to 77K, the read access time and write time improves for all cryogenic SRAM cells thanks to the reduced metal wire resistance. The 6T-RVt-HD SRAM with a lower read current shows more considerable BL delay and read access time than the 6T-LVt-HD counterpart. As the numbers of cells per WL increase from 128 to 256, from Fig. 9(a) to 9(b), WL wire delay dominates the read access time, and therefore the 6T-LVt-HD shows less read access time improvements in 256×128 arrays. Increasing cell ratio (1:2:2) improves the BL discharging delay while degrading the WL wire delay, exhibiting significant read access time degradations than the 6T-HD and 4T SRAM cells. The 4T cryogenic SRAM at 77K with lower WL resistance and capacitance offers substantial improvements in WL wire delay and read access time (−44% for 128×128 and −49% for 256×128 arrays). The read access time reduction of 4T cryogenic SRAM becomes more effective for larger capacity SRAMs (Fig. 9(b)). Besides, the 4T cryogenic SRAM shows 46% reduction in write time, as shown in Fig. 9(c). In other words, the 4T cryogenic SRAM offers superior read/write stability and high speed compared to the other 6T SRAM cells at 77K.

C. Energy-Delay Product and Leakage Power Analysis

Compared to the 6T-RVt-HD SRAM at 300K (Fig. 10(a)), the 4T SRAM significantly improves the EDP by 34% at 300K and 53% at 77K. WL and BL switching during read and write operations contribute the dynamic energy, which can be expressed in Fig. 10(b). Fig. 10(c) shows that the EDP gain is mainly contributed by the improved EDP of WL switching during read/write operations. The energy consumption due to WL switching is more significant than the BL switching because the WL capacitance is larger than the BL capacitance. Besides, the WL is charged to full-swing V_{DD} during read and write operations. Therefore, the EDP of WL switching dominates the total EDP. Fig. 10(d) shows that the 4T SRAM with reduced WL switching energy (-9.5%) and improved WL delay (−27.6% at 300K and −49% at 77K) results in significant reductions of EDP during WL switching (−34.5% at 300K and −54% at 77K). For the 6T-RVt-HD SRAM at 77K, the lower WL resistance and improved WL delay contribute to the 30.3% improvement in WL switching EDP.

Fig. 11 compares the leakage power for 6T and 4T SRAM cells at 300K and 77K. At 300K, the 4T and 6T-LVt-HD SRAM cells with lower Vt design show more considerable leakage power than the 6T-RVt SRAMs. However, cryogenic temperature (77K) significantly reduces the leakage power of 6T and 4T SRAM cells. Cryogenic technologies require refrigeration and consume cooling energy to remove the heat dissipated from the electronic devices [10]. Since the static energy consumption is negligible compared to the dynamic energy at 77K, the cooling energy is mainly determined by the dynamic energy of memory cells. Fig. 12 shows that when considering the cooling energy consumption, the 4T cryogenic SRAM exhibits 33% improvements in EDP compared to the 6T-RVt-HD cell at 77K.

V. Conclusion

High-density, high-speed, and low-energy memories are essential to support cryogenic computing efficiently. Table II summarizes the cell area, stability, speed, and EDP benchmark among various 6T and 4T cryogenic SRAM cells. The 4T SRAM at 77K shows superior read and write SNM (161mV and 451mV at V_{DD} = 0.75V). Compared to the 6T-RVt-HD SRAM at 77K, the 4T cryogenic SRAM at 77K offers significant reductions in read access time (−20%), write time (−25%), and EDP (−33%). The 4T cryogenic SRAM achieves excellent performance in all aspects and may provide the path for high-performance computing to break through the exascale barrier.

Acknowledgment

This work was supported by the Young Scholar Fellowship Program by the Ministry of Science and Technology in Taiwan, under Grants MOST110-2636-E-002-027 and MOST109-2622-8-002-003.

References

[1] S. Bonen et al., *IEEE Electron Device Lett.*, vol. 40, vol.1, p.127, 2019. [2] W. Chakraborty et al., *IEDM*, 2019, 39.4. [3] J.Y.C. Sun et al., *IEEE Trans, Electron Dev.*, vol. 34, p.19, 1987. [4] W. Chakraborty et al., *VLSI*, 2021, T5-3. [5] H. L. Chiang et al., *VLSI*, 2020. [6] P. Wang et al., *IEDM*, 2020, 38.5. [7] S.-Y. Wu et al., *IEDM*, 2016, 2.6. [8] F. Andrieu et al., *IEDM*, 2017, 20.3. [9] L. T. Clark et al., *Microelectron. J.* 53, 105–115, 2016. [10] International Roadmap for Devices and Systems, CEQIP Report, 2020.

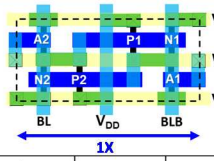
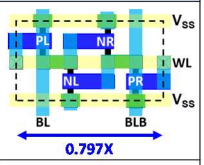
Cryogenic SRAM	6T SRAM			4T SRAM
Layout				
Threshold voltage	Regular Vt	Low Vt	Low Vt	Dual Vt
V_{DD}	0.75	0.75	Low V_{DD}	0.75
Dynamic energy	Large	Large	Small (Good)	Small (Good)
Stability (Read SNM)	Large (Good)	Small	Small	Large (Good)
Cell area	Large	Large	Large	Small (Good)

Fig. 1. Advantages of 4T SRAM compared with 6T SRAM cells for cryogenic computing. Layout designs for 6T-HD and 4T SRAM cells. For 6T-HD cell, the effective width ratio for (PU:PG:PD) = (1:1:1). The 4T SRAM cell shows small dynamic energy due to its lower metal wire capacitance. (Vt: threshold voltage)

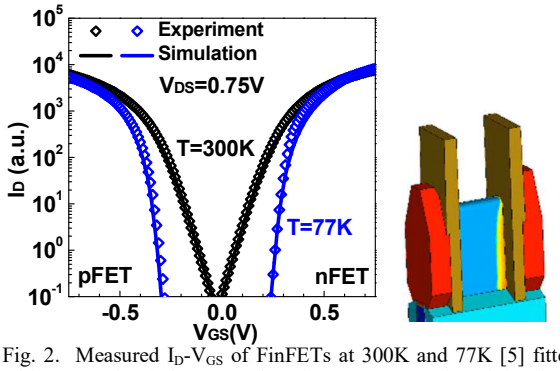


Fig. 2. Measured I_D-V_{GS} of FinFETs at 300K and 77K [5] fitted with TCAD simulations. Gate length (L_G) = 23nm. The subthreshold swing is improved at 77K (23mV/dec and 26mV/dec for nFET and pFET). (Right)Schematic of FinFET device in TCAD.

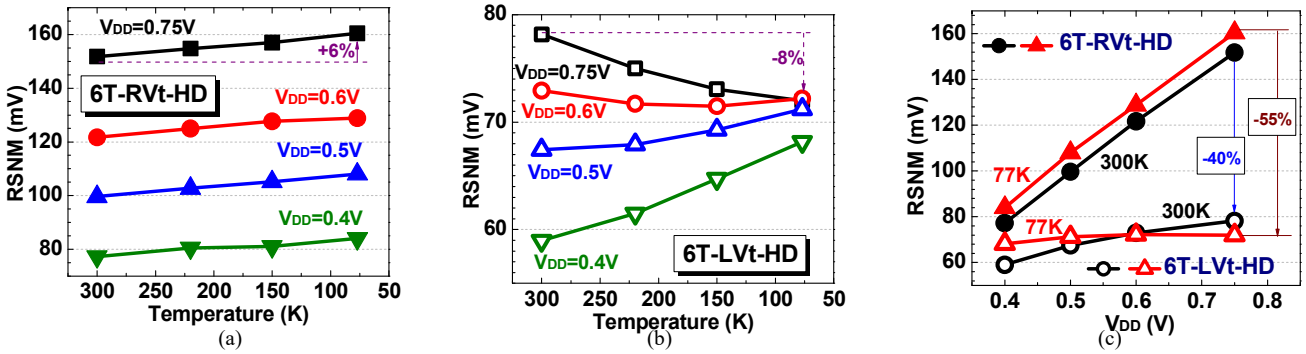
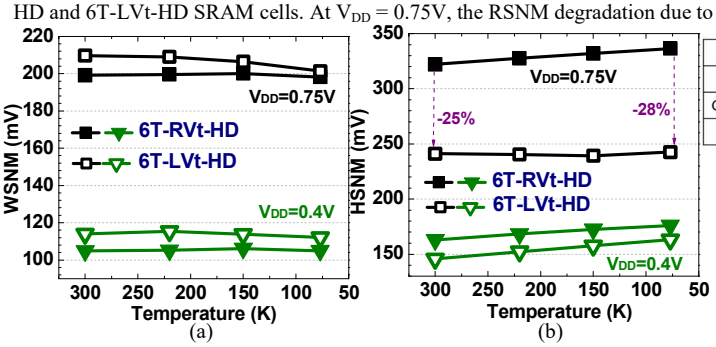
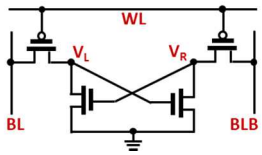


Fig. 3. RSNM comparisons for (a) 6T-RVt-HD and (b) 6T-LVt-HD SRAM cells at various V_{DD} and temperatures. At V_{DD} = 0.75V, as temperature reduces to 77K, the RSNM of 6T-RVt-HD increases (+6%), while the RSNM of 6T-LVt-HD decreases (−8%). (c) RSNM vs. V_{DD} comparisons between 6T-RVt-HD and 6T-LVt-HD SRAM cells. At V_{DD} = 0.75V, the RSNM degradation due to LVt design is more significant at 77K than 300K.



Layout Design Rules			
Fin Pitch (nm)	27	Contact CD Width (nm)	18
Gate Pitch (nm)	54	Contact CD Pitch (nm)	25
M1 Pitch (nm)	36	M1 Width (nm)	18

Table I



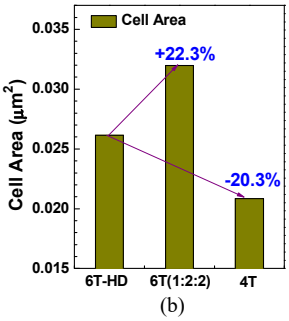


Fig. 4. (a) WSNM and (b) HSNM comparisons for 6T-RVt-HD and 6T-LVt-HD SRAM cells at various V_{DD} and temperatures. As temperature reduces from 300K to 77K, the WSNM slightly decreases for both 6T-RVt-HD and 6T-LVt-HD SRAM cells. However, the HSNM increases as temperature reduce to 77K. At V_{DD} = 0.75V, LVt cell shows 28% HSNM degradation at 77K.

Table I. Layout design rules for 7nm technology node [9] used in this work.

Fig. 5. (a) Schematic of 4T load-less SRAM cell, which consists of two nFETs and two pFETs. The 6T(1:2:2) indicates the effective width for (PU:PG:PD) = (1:2:2), which increases the cell area by 22.3%. The 4T cell improves the cell area by 20.3% compared to the 6T-HD (1:1:1) design.

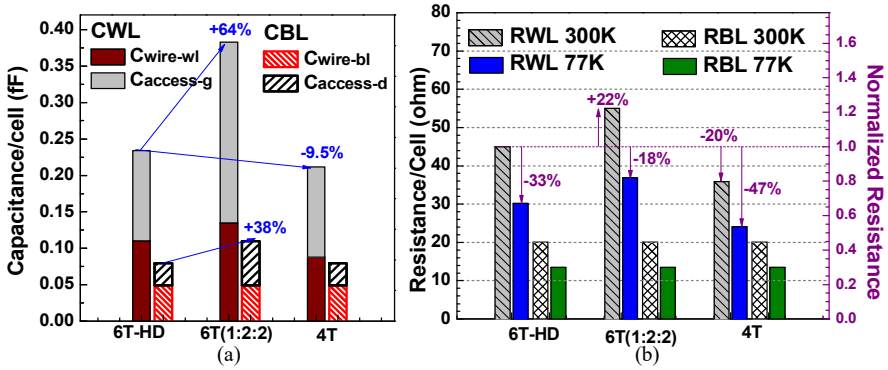


Fig. 6. (a) SRAM wordline (CWL) and bitline capacitance (CBL), including the metal wire capacitance ($C_{wire\text{-}wl}$, $C_{wire\text{-}bl}$) and access device capacitance ($C_{access\text{-}g}$, $C_{access\text{-}d}$). (b) SRAM wordline (RWL) and bitline resistance (RBL) comparisons at 300K and 77K. Compared to 300K, the BEOL metal wire resistance reduces by ~33% at 77K [5].

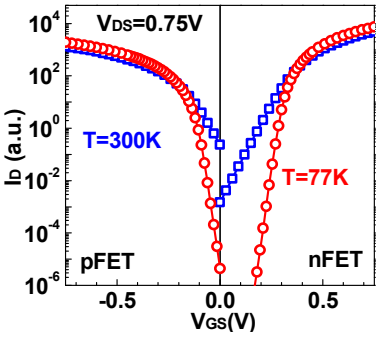


Fig. 7. The I_D-V_{GS} curves of FinFETs for 4T SRAM cells with dual-Vt design. 4T SRAMs are designed with high-Vt/low-leakage nFET and low-Vt pFET. (|Vt| = 0.35V for nFET and 0.2V for pFET at V_{DS} = 0.05V)

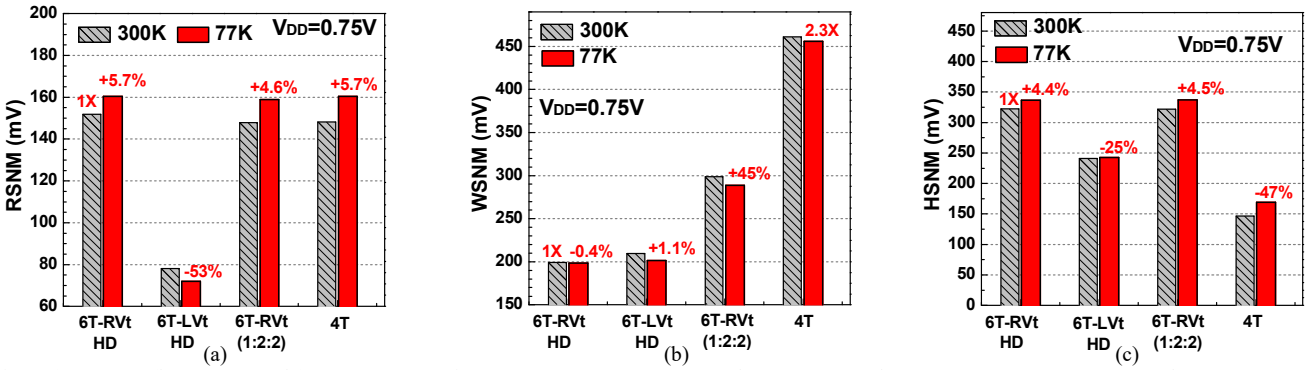


Fig. 8. (a) RSNM, (b) WSNM, and (c) HSNM comparisons at V_{DD} = 0.75V at 300K and 77K. Compared to 6T-RVt-HD SRAM at 300K, the 4T SRAM cell at 77K shows larger RSNM (+5.7%) and superior WSNM (2.3×). Although the 4T SRAM exhibits lower HSNM than the 6T SRAM cells, the 4T SRAM cell still shows adequate HSNM (170mV at 77K and 147mV at 300K). The improvement percentage are normalized to the 6T-RVt-HD cell at 300K.

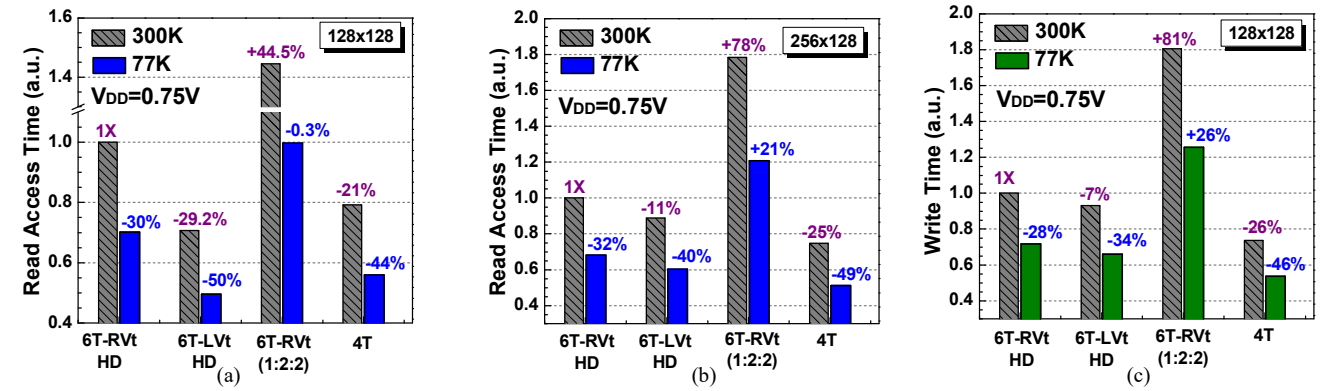


Fig. 9. Read access time comparisons for (a) 128/WL×128/BL, and (b) 256/WL×128/BL SRAM arrays at V_{DD} = 0.75V. The read access time consists of WL delay (D_{WL}), and BL delay (D_{BL}). (c) SRAM write time comparisons for 128×128 SRAM arrays at V_{DD} = 0.75V.

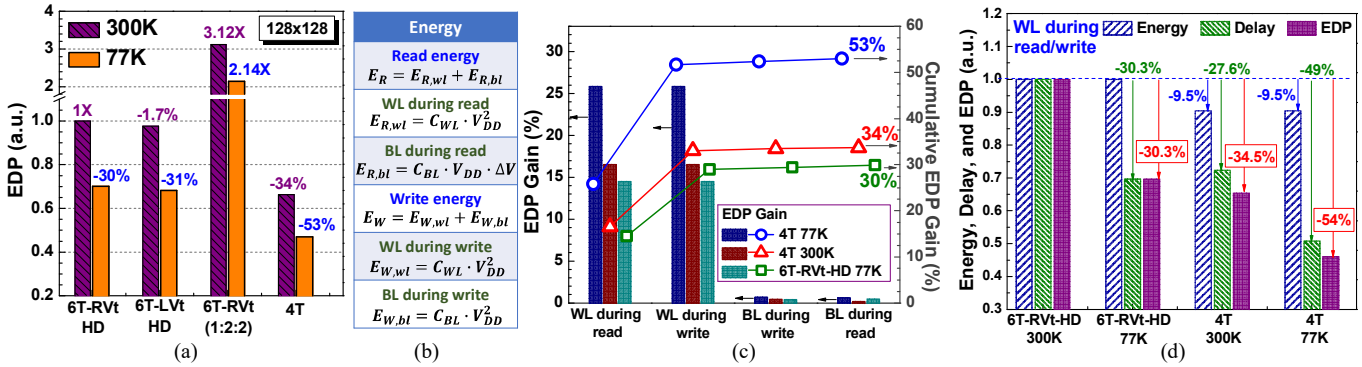


Fig. 10. (a) EDP comparisons at V_{DD} = 0.75V at 300K and 77K without considering the cooling energy. (b) Equations for switching energy consumption of read and write operations. (c) Pareto chart with contributors to the EDP gain during read and write operations compared to the 6T-RVt-HD SRAM at 300K. WL switching during read and write dominates the total EDP gain. The 4T SRAM at 77K shows the most significant EDP gain due to its improved energy (-9.5%) and delay (-49%) during WL switching, as shown in Fig.10(d). (d) The energy, delay, and EDP for WL switching during read and write operations.

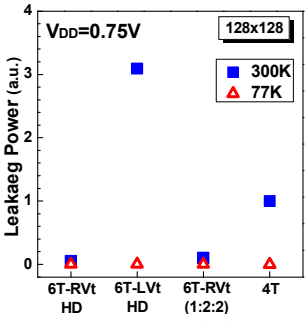


Fig. 11. SRAM static leakage power comparisons at V_{DD} = 0.75V. The leakage power of 6T and 4T SRAM cells at 77K are significantly reduced.

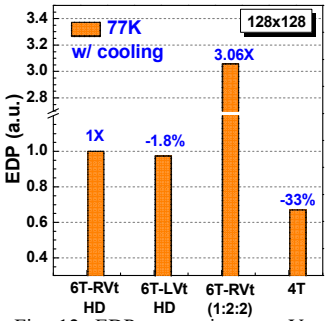


Fig. 12. EDP comparisons at V_{DD} = 0.75V for various SRAM cells at 77K considering the cooling energy consumption.

SRAM 128x128	6T-RVt HD	6T-RVt HD	6T-LVt HD	6T-RVt (1:2:2)	4T
Temperature (K)	300	77	77	77	77
Cell Area	1x	1x	1x	22.3%	-20.3%
RSNM (mV)	152	160	72	159	161
WSNM (mV)	199	198	201	289	456
HSNM (mV)	322	337	243	337	170
Read access time	1	-30%	-50%	-0.3%	-44%
Write time	1	-28%	-34%	+26%	-46%
EDP w/o Cooling	1	-30%	-31%	2.14x	-53%
EDP with Cooling	--	1	-1.8%	3.06x	-33%

Table II. Cell area, SNM, speed, and EDP benchmark for 6T and 4T SRAM cells at V_{DD} = 0.75V for cryogenic computing applications. 4T cryogenic SRAM at 77K achieves excellent performance in all aspects.