

Cryogenic PIM: Challenges & Opportunities

Salonik Resch¹,
Husrev Cilasun¹, and Ulya R. Karpuzcu

Abstract—As Moore's Law nears its end, we are searching for alternative technologies and architectures to further increase performance. *Cryogenic computing* has gained considerable attention in the last couple years, due to the highly ideal performance of CMOS circuits at very low temperatures. While cryogenic operation can provide impressive performance benefits, it introduces a new trade-off space which must be examined. Additionally, it does not eliminate current bottlenecks for performance, such as the memory wall. Processing-in-Memory architectures present an interesting opportunity. They are suitable to operate at cryogenic temperatures, enable lower cooling costs via extreme energy efficiency, and enable compute and memory capabilities in this regime with relatively minor adjustments to the architecture.

Index Terms—Cryogenic computing, processing in memory

1 INTRODUCTION

WHILE Moore's Law has lasted longer than expected, nothing lasts forever. Transistor scaling will continue to become more challenging as the years go on. This limitation makes it difficult for computer architects to continue designing systems with higher performance. A possible solution to this problem is *cryogenic computing*, where the processor and supporting memory structures are cooled to very low temperatures. The boiling point of liquid nitrogen (77K) is a common temperature target. This may seem like an extreme solution, but it offers some very attractive advantages. Electrical circuits operate faster and more energy efficiently than at room temperature, enabling further increases in computer performance.

However, a challenge for cryogenic systems is cooling cost. Heat dissipation can result in inordinate cooling costs, unless the architecture is re-designed to reduce power consumption [4]. This raises the question if Processing-in-Memory (PIM) architectures—specifically, architectures which fuse logic and memory functions within an array—are ideal candidates for cryogenic operation. PIM architectures exhibit extreme energy efficiency, due to avoiding energy-costly data transfers between the CPU and memory [6], [20]. Hence, PIM can operate within a very low power budget, which can significantly lower the cooling cost.

PIM can also provide performance improvement. A current limitation for computer performance is the *memory wall*. As CPU performance has increased over recent decades, memory demand has outgrown the increases in memory performance [10]. Combined with recent trends of increased data usage, modern applications are typically *memory bound*, meaning their performance is limited by the memory latency. PIM architectures alleviate the memory wall by performing computation where the data resides, avoiding much of the latency due to data transfer. Cryogenic operation will actually reduce the impact of the memory wall, as cold temperatures significantly improve the performance of memory - DRAM latency is reduced by a factor of 4× at 77K [15]. However, increases

in logic frequency, roughly 30-40 percent [22], and the corresponding increase in CPU frequency at cryogenic temperature, will increase the memory request rate, counteracting some of the benefit. Hence, it likely that cryogenic architectures (consisting of a CPU with supporting memory hierarchy) will still suffer performance loss due the memory wall, leaving room for potential performance increases by utilizing PIM.

Yet another benefit of PIM is that the architecture is easier to adapt to cryogenic temperatures. As they use predominately (slightly modified) memory structures, they are highly homogeneous and simple relative to more traditional systems. This makes any redesign required much easier than for more complex architectures where logic, memory, and their interface need to be optimized separately.

Numerous PIM substrates exist, including in SRAM [1], [2] and DRAM [25]. There is also a number of non-volatile PIM (NV PIM) substrates, including PCRAM [16], RRAM [9], and STT-MRAM [5], which use resistive memory devices. All are suitable candidates for cryogenic operation. More detail is provided in Section 3, but generally speaking SRAM and DRAM will stand to improve the most at cryogenic temperatures, relative to their room temperature performance. However, non-volatile PIM substrates have shown high performance and extreme energy efficiency [16], [24] at room temperature, which makes them even more promising candidates. Here, we discuss how cryogenic operation can improve the performance of PIM. As a case study, we evaluate an STT-MRAM based architecture at room temperature and cryogenic operation.

It is noteworthy that even if the performance benefit from cryogenic operation by itself is insufficient to justify the development of cryogenic architectures, such architectures will be required in order to support emerging technologies. Emerging cryogenic technologies include single flux quantum [18] and quantum computing [21], both of which will require classical hardware support [8].

2 PIM SUBSTRATES

Cell designs for different PIM-capable memory technologies are shown in Fig. 1. Each technology follows similar operating semantics when it comes to memory access to perform reads and writes.¹ Each technology contains bitlines (BL), which are used to access the memory cell to perform operations. The cells are connected to the BLs via access transistors, which are controlled by rowlines (RL). A second bitline, bitline bar (BLB) is used in SRAM and NV technologies, which is set to the opposite value as BL during write operations. The memory storage device for each technology is different. SRAM uses a latch which is constructed with transistors, shown in Fig. 1a. Due to a circuit feedback loop, there is a stable state when M1 and M4 are ON and M2 and M3 are OFF, and vice versa. To read a state, the access transistors (M5 and M6) connect Q to BL and Q' to BLB. If Q is 1 (0), the cell with pull BL (BLB) up and BLB (BL) down. To write, the same process is performed, except the voltage on the bitline is set by the bitline drivers, which will be strong enough to switch the state of cell. Shown in Fig. 1a is a 6T cell, however 8T is another common design. DRAM cells, shown in Fig. 1b, use capacitors to hold the state. The presence of a charge indicates "1" and the absence indicates "0". The access transistor connects the capacitor to the bitline, allowing the charge on the capacitor to be read or to be overwritten. RRAM and STT-MRAM both use resistive memory devices, which are devices which can have varying levels of resistance. For example, a high resistance can be logic "1" and a low resistance can be logic "0".

1. Non-volatile devices are also commonly used in cross-bar architectures, which have significantly different characteristics. We omit their consideration as they are not random access memory.

• The authors are with the University of Minnesota, Twin Cities, MN 55455 USA.
E-mail: {resc0059, cilas001, ukarpuzc}@umn.edu.

Manuscript received 22 Mar. 2021; revised 25 Apr. 2021; accepted 2 May 2021. Date of publication 4 May 2021; date of current version 9 June 2021.

(Corresponding author: Salonik Resch.)

Recommended for acceptance by Y. Etsion.

Digital Object Identifier no. 10.1109/LCA.2021.3077536

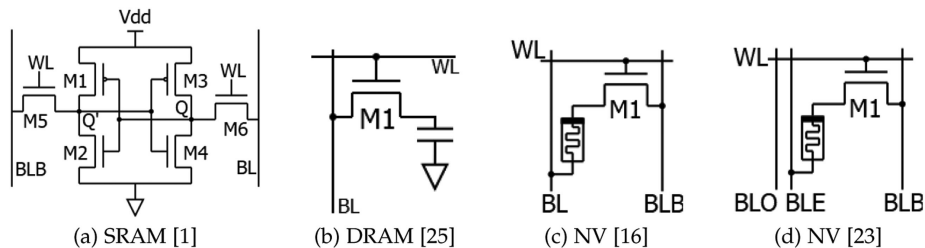


Fig. 1. Representative PIM technologies considered.

For both devices, driving a current of sufficient magnitude will set the state. The direction of the current determines which state it is set to. A typical NV memory cell, such as used in [16], is shown in Fig. 1c. A voltage across BL and BLB will drive a current through the resistive memory device.

The most common method of performing PIM in SRAM and DRAM uses the sense amplifiers and logic in the array periphery. Multiple rows are activated simultaneously, connecting multiple cells to the bitline. The sense amplifiers are then used to sense the voltage and differentiate between different inputs. Immediately after, basic logic circuitry operates on the output of the sense amps, and the result is written back into the array. If a NV PIM substrate uses the cell in Fig. 1c, it must also use this same approach. A voltage applied on BL will drive a current through multiple resistive devices (in parallel) and onto the BLB, which can then be sensed. However, the NV cell in Fig. 1d uses a double bitline, bitline even (BLE) and bitline odd (BLO), to enable computation in the array itself [23], bypassing the sense amps. This technology is called computational RAM (CRAM) [5], [23]. BLE connects to even rows and BLO connects to odd rows (not shown). In this case, voltage is applied across BLE and BLO. Current flows through input memory cells (in parallel) which are in series with an output memory cell (connected over BLB)[23]. The current will set the state of the output cell, depending on the states of the input cells.

3 ADAPTATION TO CRYOGENIC TEMPERATURES

Cryogenic operation affects how the memory devices and supporting circuitry perform. In this section, we go over each of these changes and how they impact the performance of PIM. We highlight the differences between NV PIM and the more traditional PIM based on SRAM and DRAM.

3.1 Low Level Impact of Cryogenic Temperatures

Wire resistance and capacitance both decrease. Specifically, wire resistance is a linear function of temperature [15]. As this has the most impact on the bitline, which all technologies share, this provides a universal benefit to all PIM. For SRAM and DRAM, the benefit is reduced latency for the bitline pre-charge [1], [2], [25]. This will significantly reduce latency for both memory and logic operations, as wire latency dominates memory access time [15]. The main benefit for NV PIM is energy savings. When performing a read, write or logic operation, the wire resistance is in series with the resistance of the memory devices. A low bitline resistance will decrease unwanted energy dissipation. An additional benefit for NV PIM is increased reliability. This is because NV PIM uses resistive memory devices. To perform operations, a specified voltage needs to be applied across a connected set of resistive memory devices [23], [35]. Some of the applied voltage will drop over the bitline, which reduces the margin of error. Hence, cryogenic NV PIM will be more resilient to voltage fluctuations and process variation.

CMOS transistors perform better at 77K in a number of ways. The one drawback is that the threshold voltage increases slightly with decreasing temperature [3]. Otherwise, an increase in the charge carrier mobility results in a higher ON current [33], and

both the transconductance and the sub-threshold slope are higher [26]. The steep sub-threshold slope drastically lowers leakage current. Logic built from transistors has a lower latency, roughly 30-40 percent [22]. These improvements benefit every aspect of SRAM and DRAM PIM. Logic performed by CMOS in the array periphery will be faster and more energy efficient. The reduction in leakage current will nearly eliminate static power for SRAM and reduce the refresh overhead for DRAM [30]. The benefits are less for NV PIM, which already has near zero leakage current and no refresh overhead. The main benefit for NV PIM will be the impact on the row-decoder, which is CMOS based. The row-decoder has considerable latency overhead, as it needs to be activated 1-3 times for every operation [16], [24]. Hence, superior transistor performance will have a lesser, but noticeable positive impact on NV PIM.

Resistive Memory Devices. Apply only to NV PIM, and have a number of changes at cryogenic temperature, some positive and others negative. Magnetic Tunnel Junctions (MTJs) are widely used and are the basis of STT-MRAM. MTJs have a higher endurance at cryogenic temperature [14]. As noted in Section 2, resistive memory devices store logic values in their resistivity, having both a high and low resistance state. The resistance ratio is the relative difference of resistance of the two states. MTJs have a higher ratio at cryogenic temperatures [32], [34]. This increase can be quite significant, greater than 30 percent relative to room temperature in some cases [34]. A high ratio is desirable, as it makes them easier to discern during read operations. The high ratio also increases the robustness of logic operations, making them less susceptible to process variation and voltage fluctuations [23]. A negative impact is that the absolute resistance for both states increases with lower temperature as well [32], [34]. Different fabrication processes can be used to create MTJs with varying parameters. The increase in resistance can be different for different types of MTJs and can also be different for each state of the MTJ. The resistance can increase anywhere from approximately 10 percent up to 40 percent [34]. This is undesirable, as a higher resistance requires a higher voltage to perform the same write or logic operation [14], leading to more energy consumption. RRAM is an alternative resistive technology which has a significantly higher ratio than MTJs but also a lower endurance. RRAM will also function properly at 77K [28], however it will have a further reduced endurance [31]. For example, an endurance of 10^{10} write cycles at 298K was reduced to 10^8 at 100K [11]. Additionally, it was demonstrated that they have a slightly higher operating voltage and a narrower switching voltage window [28], making them more susceptible to voltage fluctuations.

3.2 System Level Impact and New Challenges

Overall, the impact of cryogenic conditions is generally positive for PIM technologies, as summarized in Table 1. That said, cryogenic operation also introduces correctness concerns which must be addressed.

SRAM and DRAM PIM. Both off the shelf DRAM [27], [30] and SRAM [19] have been demonstrated to work at cryogenic temperatures. However, these did not consider PIM. The change in relative transistor strengths and timing of the analog circuitry may affect the correctness of logic operations. The impact will depend heavily

TABLE 1
Positive and Negative Effects of Cryogenic Operation on Different Technologies

Effect	DRAM	SRAM	MTJ	RRAM
Positive	(-)Bitline R (-)Peripheral L (-)CMOS Logic L (-)Refresh Overhead	(-)Bitline R (-)Peripheral L (-)CMOS Logic L (-)Static Power	(-)Bitline R (-)Peripheral L (+)Logic Robustness (+)Endurance	(-)Bitline R (-)Peripheral L (-)CMOS Logic L
Negative	Timing Changes	Timing Changes	(+)Write E	(-)Logic Robustness (-)Endurance

+ = increased, - = decreased, L=Latency, E=Energy, R=Resistance

on the specific architecture, but we provide a few illustrative examples. An example of 8T SRAM PIM is X-SRAM [2], where logic is performed by pre-charging the read bitline and then connecting multiple cells to the read path simultaneously. Depending on the value stored in the SRAM cells, the voltage on the bitline decays at a known rate. At a specified time, a logic buffer reads the value of the read bitline and then writes it to the write bitline. The delay between the read and the write determines the type of logic that is implemented. Cryogenic temperatures change this timing. Compute Cache [1] uses a similar approach, but with 6T SRAM cells. 6T introduces the concern that SRAM cells may destructively interact, since the read and write paths share the same bitlines. This can be avoided by lowering the wordline voltage [1], [12] at room temperature. However, at cryogenic temperatures the transistors connecting the bitline to both the supply voltage and ground will be stronger, and this may increase the susceptibility. Ambit [25] uses a sense amplifier to read multiple DRAM cells simultaneously to differentiate between different input combinations. The amount of current that is drawn from each DRAM cell through the access transistor could be significantly different from that at room temperature.

NV PIM. While NV PIM has some correctness concerns as well, these are easier to account for. The voltages applied to perform writes and logic operations change, due changes in the MTJ resistance, and the peripheral circuitry latency decreases. Changing the supply voltage and the frequency can account for this, no circuit re-design is required. The main disadvantages for cryogenic NV PIM are the non-ideal device characteristics, reduced energy efficiency for MTJs and reduced endurance for RRAM. A reduction in energy efficiency may be tolerable, given NV PIM demonstrates extreme energy efficiency at room temperature.

4 QUANTITATIVE ANALYSIS

To determine the efficacy of cryogenic PIM we evaluate SRAM, DRAM, and NV-PIM both at 300K and at 77K. To estimate the performance of SRAM PIM we take data from Min *et al.* [19] and for DRAM PIM we take data from Lee *et al.* [15]. These studies provide the latency and energy of memory accesses at 77K relative to 300K. We assume that a single memory access and logic operation have the same overhead. For NV PIM, we use MTJs as a representative case study, as they have a higher endurance at cryogenic temperatures. We take peripheral circuitry estimates from NVSIM [7] and match them with MTJ parameters from [35]. This provides the latency and energy for each PIM operation at 300K. To estimate performance at cryogenic temperatures, we modify the peripheral circuitry latency and energy based on experimental data for CMOS operation [33], and MTJ latency and energy based on experimental data from [32], [34]. The peripheral circuitry and CMOS logic latency reduces by 35 percent and the MTJ write energy increases by 15 percent at 77K relative to 300K.

Machine-learning inference has been demonstrated efficiently in memory [13], [23], [29]. It is also ubiquitous in the cloud/server environment, making it a notable candidate for cryogenic acceleration. Hence, we use neural network inference as a case study with

the CIFAR-10 image recognition dataset as the input. We take the network configurations provided by [17] and map them by hand to run on the PIM substrate.

We assume the PIM substrates consist of 1024x1024 memory arrays. Each memory array can perform logic operations within the columns, i.e., they have *column-level parallelism*. Such PIM architectures have been used with SRAM [2], DRAM [25], and NV technologies [16], [23]. Operations are driven by an external controller. We use a data layout similar to that used in PIMBALL [23]. Multiplications, additions, and subtractions are performed in a bit-serial manner within the columns of the memory arrays. Data is moved between columns and arrays with read and write operations (orchestrated by the external controller). All network parameters, including weights and thresholds, are stored in the memory prior to inference and kept constant. Input data (images) and the neurons of hidden layers are moved between memory arrays as needed throughout the run of the program.

To estimate the total latency and energy, we sum the latency and energy of all logic operations, reads, and writes required to perform the program. For every operation, we account for the overhead due to the peripheral circuitry and row decoder activation, which vary depending on the operations and the sequence they are performed in.

Results are shown for latency in Fig. 2a and for energy in Fig. 2b. The energy consumption reported does not include cooling costs, which can be significant [4]. Cryogenic operation provides a latency advantage for all technologies due to improvements of the peripheral circuitry. Latency is reduced to 50 percent of the room temperature counterpart for SRAM, to 32 percent for DRAM, and to 89 percent for NV PIM, respectively. NV PIM's latency does not improve as much because the MTJ write latency remains the same at cryogenic temperatures. SRAM and DRAM PIM feature significantly reduced energy consumption, as well, down to 38 percent for SRAM and to 48 percent for DRAM PIM, respectively, of the room temperature counterparts. This is largely due to a reduction in leakage current, which enables a lower operating voltage. NV PIM actually is less efficient at cryogenic temperatures, with an energy consumption of 109 percent relative to the room temperature counterpart. This is due to the increase in the MTJ write energy, which dominates energy consumption. Additionally, NV

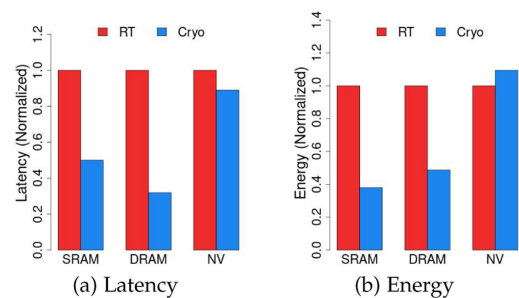


Fig. 2. Benchmark characterization at room temperature (RT) versus cryogenic temperature (Cryo). Values are normalized relative to the RT performance. Authorized licensed use limited to: University of Minnesota. Downloaded on October 09, 2023 at 19:23:29 UTC from IEEE Xplore. Restrictions apply.

PIM already has near zero leakage current at room temperature, and hence this is not an added benefit at cryogenic temperatures. While the energy increase relative to room temperature is a detriment, it may be tolerable. At room temperature, NV PIM has demonstrated superior energy efficiency to more traditional architectures. For example, a Xeon E5-2640 CPU consumes 129J to perform the CIFAR-10 benchmark [17], where an NV PIM solution only consumes 31.9 μ J [23], superior even to specialized FPGAs consuming 299 μ J [17]. Hence, a reasonable increase in its energy consumption will still result in an overall energy efficient operation. Note that MTJs optimized for room temperature were used in this analysis; MTJs designed specifically for cryogenic temperatures may display superior efficiency.

5 CONCLUSION

PIM technologies are well suited for the cryogenic domain. Their modular architectures facilitate ease of transition and their energy efficiency should enable lower cooling budgets. SRAM and DRAM based PIM show significant improvements in both performance and energy efficiency (when compared to their room temperature counterparts). NV PIM also exhibits increased performance but suffer from reduced energy efficiency when compared to its room temperature counterpart. This does not rule out NV technologies at cryogenic temperatures, however, as at room temperatures NV PIM tends to be typically much more energy efficient than SRAM or DRAM based PIM.

REFERENCES

- [1] S. Aga, S. Jeloka, A. Subramaniyan, S. Narayanasamy, D. Blaauw, and R. Das, "Compute caches," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit.*, 2017, pp. 481–492.
- [2] A. Agrawal, A. Jaiswal, C. Lee, and K. Roy, "X-SRAM: Enabling in-memory boolean computations in CMOS static random access memories," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 12, pp. 4219–4232, Dec. 2018.
- [3] A. Beckers, F. Jazaeri, A. Ruffino, C. Bruschini, A. Baschirotto, and C. Enz, "Cryogenic characterization of 28 nm bulk CMOS technology for quantum computing," in *Proc. 47th Eur. Solid-State Device Res. Conf.*, 2017, pp. 62–65.
- [4] I. Byun, D. Min, G. Lee, S. Na, and J. Kim, "CryoCore: A fast and dense processor architecture for cryogenic computing," in *Proc. ACM/IEEE 47th Annu. Int. Symp. Comput. Archit.*, 2020, pp. 335–348.
- [5] Z. Chowdhury et al., "Efficient in-memory processing using spintronics," *IEEE Comput. Archit. Lett.*, vol. 17, no. 1, pp. 42–46, Jan.–Jun. 2018.
- [6] K. De Bosschere, A. Cohen, J. Maebe, and H. Munk, "HiPEAC vision 2015," 2015. [Online]. Available: <http://erict.win.tue.nl/docs/Local/ARTEMIS%202015%20Pre-Brokerage%20-%20Project%20Summaries,%20Pitches%20&%20Posters/General%20Presentations/hipeac-vision-2015.0bde1ff5d0b5.pdf>
- [7] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, "NVSim: A circuit-level performance, energy, and area model for emerging nonvolatile memory," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 31, no. 7, pp. 994–1007, Jul. 2012.
- [8] X. Fu et al., "eQASM: An executable quantum instruction set architecture," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit.*, 2019, pp. 224–237.
- [9] D. Fujiki et al., "In-memory data parallel processor," *ACM SIGPLAN Notices*, vol. 53, no. 2, pp. 1–14, 2018.
- [10] J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*. Amsterdam, The Netherlands: Elsevier, 2011.
- [11] X.-D. Huang, Y. Li, H.-Y. Li, K.-H. Xue, X. Wang, and X.-S. Miao, "Forming-free, fast, uniform, and high endurance resistive switching from cryogenic to high temperatures in W/AlO_x/Al₂O₃/Pt Bilayer memristor," *IEEE Electron Device Lett.*, vol. 41, no. 4, pp. 549–552, Apr. 2020.
- [12] S. Jeloka, N. B. Akes, D. Sylvester, and D. Blaauw, "A 28 nm configurable memory (TCAM/BCAM/SRAM) using push-rule 6T bit cell enabling logic-in-memory," *IEEE J. Solid-State Circuits*, vol. 51, no. 4, pp. 1009–1021, Apr. 2016.
- [13] H. Jia, H. Valavi, Y. Tang, J. Zhang, and N. Verma, "A programmable embedded microprocessor for bit-scalable in-memory computing," in *Proc. IEEE Hot Chips 31 Symp.*, 2019, pp. 1–29.
- [14] L. Lang et al., "A low temperature functioning CoFeB/MgO-based perpendicular magnetic tunnel junction for cryogenic nonvolatile random access memory," *Appl. Phys. Lett.*, vol. 116, 2020, Art. no. 022409.
- [15] G.-H. Lee, D. Min, I. Byun, and J. Kim, "Cryogenic computer architecture modeling with memory-side case studies," in *Proc. ACM/IEEE 46th Annu. Int. Symp. Comput. Archit.*, 2019, pp. 774–787.
- [16] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, "Pinatubo: A processing-in-memory architecture for bulk bitwise operations in emerging non-volatile memories," in *Proc. 53rd Annu. Des. Autom. Conf.*, 2016, pp. 1–6.
- [17] S. Liang et al., "FP-BNN: Binarized neural network on FPGA," *Neurocomputing*, vol. 275, pp. 1072–1086, 2018.
- [18] M. A. Manheimer, "Cryogenic computing complexity program: Phase 1 introduction," *IEEE Trans. Appl. Supercond.*, vol. 25, no. 3, Jun. 2015, Art. no. 1301704.
- [19] D. Min et al., "CryoCache: A fast, large, and cost-effective cache architecture for cryogenic computing," in *Proc. 25th Int. Conf. Archit. Support Program. Lang. Oper. Syst.*, 2020, pp. 449–464.
- [20] O. Mutlu, "Intelligent architectures for intelligent machines," in *Proc. Int. Symp. VLSI Des. Autom. Test*, 2020, pp. 1–4.
- [21] National Academies of Sciences, Engineering, and Medicine and others, *Quantum Computing: Progress and Prospects*. Washington, D.C., USA: National Academies Press, 2019.
- [22] 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.
- [23] S. Resch et al., "PIMBALL: Binary neural networks in spintronic memory," *ACM Trans. Archit. Code Optim.*, vol. 16, no. 4, 2019, Art. no. 41.
- [24] S. Resch et al., "MOUSE: Inference in non-volatile memory for energy harvesting applications," in *Proc. 53rd Annu. IEEE/ACM Int. Symp. Microarchit.*, 2020, pp. 400–414.
- [25] V. Seshadri et al., "Ambit: In-memory accelerator for bulk bitwise operations using commodity DRAM technology," in *Proc. 50th Annu. IEEE/ACM Int. Symp. Microarchit.*, 2017, pp. 273–287.
- [26] M. Shin et al., "Low temperature characterization of 14nm FDSOI CMOS devices," in *Proc. 11th Int. Workshop Low Temp. Electron.*, 2014, pp. 29–32.
- [27] S. S. Tannu et al., "Cryogenic-DRAM based memory system for scalable quantum computers: A feasibility study," in *Proc. Int. Symp. Memory Syst.*, 2017, pp. 189–195.
- [28] C. Vaca et al., "Study from cryogenic to high temperatures of the high-and low-resistance-state currents of ReRAM Ni-HfO₂-Si capacitors," *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1877–1883, May 2016.
- [29] H. Valavi, P. J. Ramadge, E. Nestler, and N. Verma, "A 64-tile 2.4-Mb in-memory-computing CNN accelerator employing charge-domain compute," *IEEE J. Solid-State Circuits*, vol. 54, no. 6, pp. 1789–1799, Jun. 2019.
- [30] F. Wang, T. Vogelsang, B. Haukness, and S. C. Magee, "DRAM retention at cryogenic temperatures," in *Proc. IEEE Int. Memory Workshop*, 2018, pp. 1–4.
- [31] F. Ware et al., "Do superconducting processors really need cryogenic memories? The case for cold DRAM," in *Proc. Int. Symp. Memory Syst.*, 2017, pp. 183–188.
- [32] J.-B. Yau, Y.-K.-K. Fung, and G. W. Gibson, "Hybrid cryogenic memory cells for superconducting computing applications," in *Proc. IEEE Int. Conf. Rebooting Comput.*, 2017, pp. 1–3.
- [33] M. B. Yelten, "Cryogenic DC characteristics of low threshold voltage (VTH) n-channel MOSFETs," *Balkan J. Elect. Comput. Eng.*, vol. 7, no. 3, pp. 362–365, 2019.
- [34] L. Yuan et al., "Temperature dependence of magnetoresistance in magnetic tunnel junctions with different free layer structures," *Physical Rev. B*, vol. 73, no. 13, 2006, Art. no. 134403.
- [35] M. Zabihi, Z. I. Chowdhury, Z. Zhao, U. R. Karpuzcu, J.-P. Wang, and S. S. Sapatnekar, "In-memory processing on the spintronic CRAM: From hardware design to application mapping," *IEEE Trans. Comput.*, vol. 68, no. 8, pp. 1159–1173, Aug. 2019.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.