Cryogenic-DRAM based Memory System for Scalable Quantum Computers: A Feasibility Study

Swamit S. Tannu
Georgia Tech
swamit@gatech.edu

Douglas M. Carmean
Microsoft Research
dcarmean@microsoft.com

Moinuddin K. Qureshi
Georgia Tech
moin@ece.gatech.edu

ABSTRACT

A quantum computer can solve fundamentally difficult problems by utilizing properties of quantum bits (qubits). It consists of a quantum substrate, connected to a conventional computer, termed as control processor. A control processor can manipulate and measure the state of the qubits and act as an interface between qubits and the programmer. Unfortunately, qubits are extremely noise-sensitive, and to minimize the noise; qubits are operated at cryogenic temperatures. To build a scalable quantum computer, a control processor which can work at cryogenic temperatures is essential [3, 14]. In this paper, we focus on the challenges of building a memory system for a cryogenic control processor. A scalable quantum computer will require large memory capacity for storing the program and the data generated by the quantum error correction. To this end, we evaluate the feasibility of cryogenic DRAM-based memory system by characterizing commercial DRAM modules at cryogenic temperatures. In this paper, we report the minimum operational temperature for 55 DIMMs (consisting of a total of 750 DRAM chips) and analyze the error patterns in commodity DRAM devices operated at cryogenic temperatures. Our study shows that a significant fraction of DRAM chips continue to work at temperatures as low as 80K. This study is an initial step towards evaluating the effectiveness of cryogenic DRAM as a main memory for quantum computers.

CCS CONCEPTS

• **Computer systems organization → Quantum computing**;

KEYWORDS

Computer Organization, Quantum Control Processor, Quantum Computer Architecture, Quantum Error Correction

1 INTRODUCTION

Quantum computers can solve commercially and scientifically important problems that are currently unsolvable with conventional computers. For example, accurate simulation of large molecules is a fundamental problem for developing new drugs, materials, and eco-friendly manufacturing processes. Currently, these simulations are intractable on a conventional computer because of the exponentially growing computational complexity. However, a quantum computer, which uses quantum properties itself, to store and process data, can perform these simulations in polynomial time [9]. In addition to quantum simulations, quantum computers can accelerate number theory, algebraic, and optimization problems. For example, Shor's algorithm running on a quantum computer can factorize numbers in polynomial time [25]. The ability to factor numbers in polynomial time makes the existing encryption systems vulnerable, as encryption system such as RSA is based on the hardness of the factoring problem. Recent advances in quantum device fabrication techniques and quantum error correction theory have propelled the idea of quantum computing from experimental physics research to an engineering challenge. To this end, several industry research labs are actively working towards building a quantum computer [3, 4, 17, 26, 27].

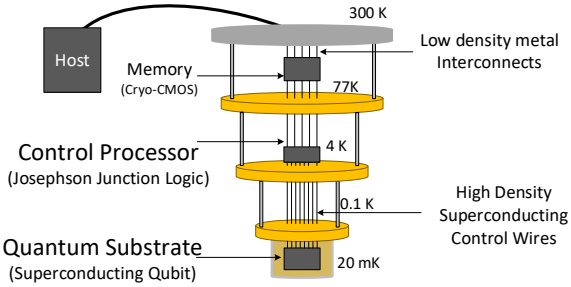

Figure 1: Proposed organization of a scalable superconducting quantum computer [3, 14]. Control processor operating at 4K provides instructions to the quantum substrate operating at 20mK.

Quantum bit (qubit) is a fundamental unit of quantum information and a quantum algorithm solves a problem by manipulating and measuring the states of qubits. A control processor is a conventional computer which facilitates the measurement and the manipulation of the qubit states. It utilizes components like regular compute and memory along with a specialized quantum execution unit. Quantum states are extremely fragile and can quickly decohere (lose its state) due to thermal noise. To protect the quantum states, a quantum substrate is operated at the temperature of a few milli-Kelvins as shown in Figure 1. Moreover, to build a scalable quantum computer, a control processor needs to be placed close to the qubits at cryogenic temperatures because the control processor operated at room-temperature can not scale beyond few hundreds of qubits due to a limited number of wires connecting room-temperature and the quantum substrate. The metal interconnects connecting room-temperature memory, and the cryogenic control processor would have thermal leakage due to the large temperature gradient

MEMSYS 2017, October 2–5, 2017, Alexandria, VA, USA

ACM ISBN 978-1-4503-5335-9/17/10…$15.00
https://doi.org/10.1145/3132402.3132436

between room temperature and control processor (about 300K) that can destabilize the noise-sensitive qubits. In the recent ISCA'16 keynote, several architectural and technological challenges in building a scalable quantum computer were discussed [3]. The keynote highlighted the necessity of the cryogenic control processor and also highlighted key challenges such as the need for memory hierarchy and I/O architectures that can tolerate long latency for data transfer between cryogenic and room temperature. Recently, *Rambus* and *Microsoft* have announced a collaborative effort to investigate the cryogenic memory system placed at 77K temperature for quantum computers [24]. In this paper, we focus on the challenges in building the memory system for a quantum computer using superconducting qubits.

A quantum computer will require significantly large capacity memory to store a quantum program. A typical quantum program consists of quantum instructions which manipulate and measure the qubit states according to the quantum algorithms. Quantum algorithms utilize subroutines which generate specific quantum states (states generated by the arbitrary rotations). To generate these states, the subroutine requires a long sequence of quantum instructions (decomposition of arbitrary rotations). This results in a quantum executable with the memory-footprint in the regime of several tens of gigabytes. The problem is even more challenging for the superconducting qubits which have gate latency of the order of few nanoseconds, and generating the instruction sequence at runtime may not be possible due to time constraints. In this scenario, we may have to resort to a static compilation that might result in a large quantum executable [19]. A control processor will require substantial memory capacity to support quantum error correction (QECC). Quantum error correction is essential to maintain the data integrity of quantum data, and the QECC instructions are continuously executed on all the qubits. This generates large amounts of classical data resulting from syndrome measurements every cycle that needs to be buffered for error decoding and qubit tuning [6, 18]. Initial quantum computers will require large storage capacity to log the data for the purpose of debugging and quantum noise modeling. To this end, solid-state quantum computers will require a substantial memory capacity which can be operated at cryogenic temperatures, as keeping the memory at room temperature may not be a scalable solution due to a limited number of wires between room temperature and 20mK [14]. In this paper, we will experimentally evaluate the feasibility of building a cryogenic DRAM based memory system by characterizing commercially available DRAM at cryogenic temperatures. The following are the contributions of this paper:

- We develop an experimental setup and methodology to test the reliability of DRAM chips at cryogenic temperatures. In particular, we design a compact heatsink which can provide localized cooling to the DRAM DIMMs and enable characterization of DRAM chips at cryogenic temperatures using a low-cost, off-the-shelf memory tester and a host machine.
- To understand the effects of cryogenic environment on off-the-shelf DRAM, we characterized DRR4, DDR3, and DDR2 DIMMs at temperatures (80K to 160K). In this study, we have tested 55 DIMMS (with 750 DRAM chips) from six different vendors. We observe 10 out of 55 (about 18%) DIMMs are functional without any errors at 90K, and almost 98% of DIMMs are functional above 150K.
- The characterization shows that most of the errors in DRAM at cryogenic temperatures are localized in a single DIMM, as only 8% of the DRAM chips were faulty. Out of about 17000 reported errors, 99.98% were single-bit errors, and we observe almost no spatial correlation among errors. This observation demonstrates a potential to use traditional forward error correction techniques like block-codes to mitigate the memory errors in cryogenic DRAMs.

As we scale the quantum computers, system and architectural challenges can bottleneck the progress. Designing the cryogenic memory system for quantum computers is one such challenge which will have an impact on future of quantum computing. In this paper, we show that it is possible to operate conventional DRAM chips at about 150K lower than its rated temperature. With this motivating data, we believe DRAM has the potential to be used as a primary memory technology for quantum computers.

2 MOTIVATION

In this section, we will briefly discuss the requirements for memory systems for a control processor by examining the high-level characteristics of the quantum workloads.

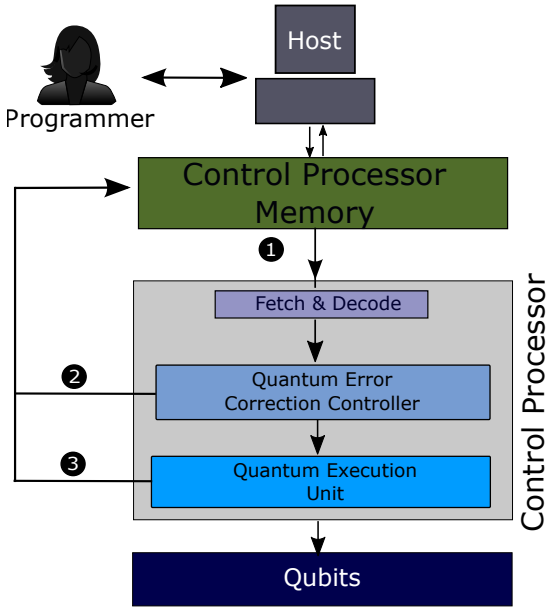

Figure 2: Organization and Programming model of a quantum computer.

2.1 Need for High Capacity Memory

A quantum program can be compiled on a regular host or a server [8] and the quantum executable is placed in the control processor memory. The instructions are fetched from the control processor memory, and the instructions are decoded and executed by the control processor as shown in Figure 2. Control processor memory is also utilized as the data memory for storing the intermediate classical data generated by the quantum execution unit and the quantum error correction controller as shown in Figure 2.

2.1.1 Program Memory Requirements. Quantum computers will require a significant memory capacity (10 to 100 GB) for program memory due to a large number of arbitrary rotations in a quantum algorithm. An arbitrary rotation is a subroutine which produces a specific quantum state. Unfortunately, this subroutine requires a large number of quantum instructions. Kudrow et. al [19] highlighted the problem of compiling a large sequence of quantum instructions due to arbitrary rotations. To improve the code size, authors propose to use dynamic compilation, where the arbitrary rotations are decomposed into series of regular quantum instructions at run-time. However, the authors highlight the limited benefits of dynamic compilation in the case of superconducting qubits. The decomposition of arbitrary rotation is a computationally-expensive task that control processor may not be able to perform in real-time. Authors recommend using statically-compiled libraries containing decomposed arbitrary rotations. However, this results in an executable with large memory footprint.

2.1.2 Data Memory Requirements. Qubits are extremely noise-sensitive, and even at milli-Kelvin temperatures, a qubit can lose its state. To preserve the data integrity, quantum computers use continuous error correction on all the qubits. To perform quantum error correction, error syndromes are continuously generated and measured. This continuous generation and measurement of syndromes require a large data memory capacity and bandwidth. The error decoding requires storage of classical data generated by the syndrome measurements. As the syndrome measurement data is required to decode the errors. For a quantum computer running a surface code error correction with 'N' superconducting qubits, QECC syndrome generation, and measurement cycle requires about 200ns [6], i.e., every 200ns, about N/2-bits of classical information is generated which needs to be stored in control processor memory. For the initial quantum computers, data logging will be essential as every instruction and measurement will provide valuable information which can be used to build the statistical models for quantum noise and gate fidelity. Also, superconducting quantum computers might require run-time calibration and tuning of qubits [18].

2.2 Need for Cryogenic Memory

In existing lab scale quantum computers, qubits are placed inside a dilution-refrigerator which maintains the temperature of few milli-Kelvins, while the control processor is kept at room temperature (300K). Control processor is connected to the qubits via long co-axial control cables. While this organization works for a small number of qubits, it is not scalable. The number of wires connecting control processor and the qubits is severely limited because of thermal leakage caused by the large thermal gradient across the control wires. To solve this problem, a cryogenic control processor can be connected to qubits using high density superconducting interconnects [29]. The limited number and long lengths of wires would also restrict the data bandwidth between room temperature and the cryogenic control processor. This makes memory placed at cryogenic temperature necessary for a scalable control processor.

As per the recent proposals on a control processor for a quantum computer, a superconducting logic is extremely attractive from a power and a performance perspective [3, 13, 21, 28, 29]. However, the existing superconducting memory technologies can not be used as the main memory as the Josephson-junction (JJ) based memories have limited memory density [13]. Researchers have demonstrations of hybrid JJ-CMOS memories which attempt to improve the memory density [20]. However, the capacity and energy efficiency of hybrid memories placed at 4K is still not enough to support a scalable quantum computer. As architects, we see a possible trade-off, where superconducting memory can be utilized as cache and for main memory– DRAM can be used. DRAM placed at room temperature will have extremely large latency and small bandwidth. The latency of room temperature memory can be amortized by designing a memory hierarchy which utilizes different temperature levels of the cryogenic refrigerator. The typical cryogenic refrigerator has three to four temperature levels- 20mK, 4K, 77K, 120K as shown in Figure 1. Cryogenic stages at 20mK and 4K have extremely small power budget which requires superconducting memories and limits the memory capacity. However, intermediate levels like 77K and 120K can be used to place large capacity CMOS memories. CMOS memories can be operational at 50K to 180K temperatures. In fact, it is an ideal temperature range for CMOS devices. At these operating temperatures leakage current is dramatically reduced and driving capability of an access transistor is improved due to increased carrier mobility [12].

2.3 Prior Work on Cryogenic DRAM

During the late 1980s, IBM demonstrated a specially designed high-performance cryogenic DRAM [12], which was shown to have a 3X performance benefit over DRAM operated at room temperature. The performance benefit resulted from improved carrier mobility. Detailed characterization studies for memory latency, data retention are reported in [2, 10–12]. Several characterization studies have also reported increased threshold voltage and worsened variation and mismatch at cryogenic temperature due to carrier freezout [2]. However, the characterized and designed Cryo-DRAM used a long channel and planar transistor devices. It is important to understand the effect of cryogenic temperatures on today's short-channel and non-planar commodity memory devices. Recently, cold-boot attack demonstrated the retention of data in the DRAM at cryogenic temperatures [7]. However, that study is concerned with recovering the state of some of the cells rather than having a reliable memory to store data. It is important to understand the effect of the cryogenic temperature on the existing DRAM devices. Therefore, we perform a feasibility study for Cryo-DRAM by characterizing commodity DRAMs at cryogenic temperature.

3 EXPERIMENTAL METHODOLOGY

In this section, we will discuss the construction of compact cryogenic heatsink and the Cryo-DRAM characterization setup.

3.1 Compact Cryogenic Heatsink

DRAM timing tests are commonplace in conventional computer systems to detect and track the faulty address regions in the memory. The tests can be performed by a running software testing suites like Memtest86 [23] or using a dedicated memory testers like Ramcheck [15]. Characterizing DRAM at cryogenic temperatures becomes challenging as typical host-systems or specialized memory testers are designed to operate at temperatures- 0^oC to 80^oC.

To characterize the DRAM at cryogenic temperatures, we build a compact cryogenic heatsink (CCH) which can cool only the DRAM chips so that the hardware driving the DIMM (CPU, FPGA, or specialized memory tester) is not subjected to cryogenic temperatures. The isolation of memory controller and the Cryo-DRAM enables conventional means of DRAM testing at cryogenic temperatures.

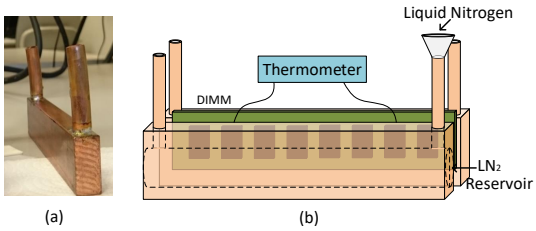

Figure 3: (a) Compact Cryogenic Heatsink (CCH) is built using a rectangular copper bar. (b) To cool down the DIMM, it is placed between pair of heatsinks.

As shown in Figure 3, a compact heatsink is designed using a rectangular copper bar such that a DIMM can be clamped between two copper bars. The heatsink is designed to enclose the DIMM such that it will be in contact with all the DRAM chips. The temperature of DRAM can be reduced if both heatsinks are maintained at cryogenic temperatures. To cool the compact heatsinks, we use Liquid Nitrogen (LN_2) as a cryogenic coolant. LN_2 boils at 77K (-196^oC or -321^oF) which can be used to reduce the temperature of a heatsink. We used LN_2 due to its inert nature, sufficiently large heat of vaporization, low-cost, and ease of handling.

To construct CCH, a horizontal through hole across the length of the copper bar was drilled. Then two vertical drills at the corners were drilled across the width such that vertical drills connect the horizontal through hole. The through hole was sealed, and two protruding copper pipes were soldered into vertical drills as shown in Figure 3. With this design, Liquid Nitrogen can be poured into the copper bar, and a small reservoir of LN2 can be retained to cool the heatsink and the attached DRAM chips.

The heatsinks are clamped on a DIMM while it is plugged into the DRAM slot of the CPU or the memory tester as shown in Figure 4 and Figure 5. The temperature of a DIMM can be lowered by pouring Liquid Nitrogen inside the heatsink using flask attached to the copper pipes. The cryogenic temperature can be maintained by continuously pouring Liquid Nitrogen inside the heatsink. The designed heatsink can lower the temperature of the DRAM chips, and it can achieve the minimum temperature of 80K. We use thermocouples to monitor the temperature of DRAM chips and the heatsink. The size of the reservoir and the thickness of copper bar were chosen to minimize the temperature fluctuations.

3.2 Experimental Setup

We used two DRAM test setups: Setup-A, and Setup-B for characterizing the cryogenic DRAMs. Both of the test setups use identical heatsink assembly. Setup-A utilizes Ramcheck DRAM tester [15] to characterize DDR2 and DDR3 DIMMs. Whereas Setup-B uses a host machine running Memtest86, a DRAM testing application [23] to characterize DDR4 DIMMs.

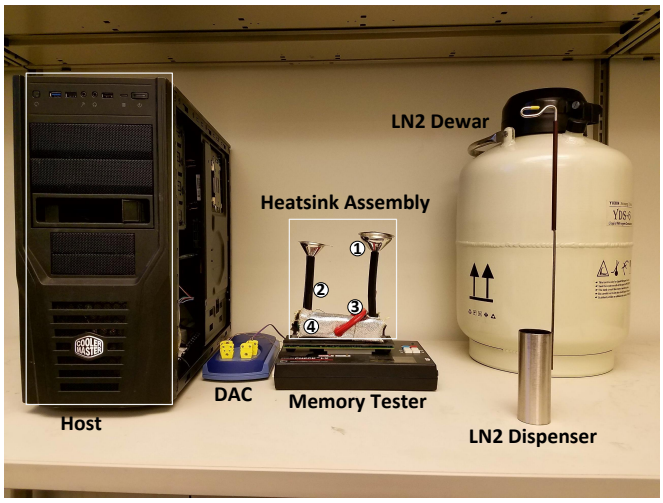

Figure 4: Characterization Setup-A utilizes Ramcheck memory tester [15] and the compact heatsink to characterize DDR2 and DDR3 DIMMs.

Design of Setup-A: Figure 4 shows the experimental test Setup-A. It consists of Ramcheck DRAM tester, CCH assembly, thermocouple and temperature data logger. The heatsink assembly is mounted on a DIMM while it is plugged in the Ramcheck tester. CCH assembly consists of ① flask, ② insulated copper pipes, ③ C-clamp and ④ a pair of insulated heatsink. Ramcheck tester is connected to the host computer to log the DRAM test results and to control the test sequences. The host computer also logs the surface temperature of the DRAM chips and the heatsinks using a thermocouple and a data-logger. Table 1 summarizes the details of the characterized DIMM using Setup-A. Ramcheck performs the March tests, voltage bounce test, and other specialized data patterns to test DIMMs and determines the locations of faulty memory cells.

Figure 5: Setup-B utilizes the host machine with CCH assembly mounted on DDR4 DIMM. Memtest86 is used to characterize the DIMMs.

Design of Setup-B: Figure 5 shows the Setup-B. It consists of a desktop machine and a compact heatsink assembly with the attached thermocouple and temperature data logger. The heatsink

assembly is mounted on a DIMM while it is plugged in the desktop machine running Memtest86 to log the DRAM tests. The details of the experiment setup are summarized in Table 2.

3.3 Test Methodology

The primary objective of the characterization is to understand the functionality of DRAM at cryogenic temperature. In this paper, we focus on determining the Minimum Operational Temperature (MOT) of the DIMMs, and the nature of errors in the Cryo-DRAM. Cryogenic DRAM testing requires two steps: cooling of DIMM and timing and data integrity tests for the DIMM. The temperature of the DIMM is gradually reduced at the rate of 10K to 25K per minute. This rate is derived empirically as we observed permanent failures with the higher rate of change of temperature (40 K/min to 50 K/min). The failures can result from rapid cooling as it can cause thermal stress in the package and solder.

To determine the nature of errors and the functionality of DDR2 and DDR3 DIMMs, we make use of the basic and advanced test sequences in the Ramcheck-LX memory tester (Setup-A). The details of the test sequences can be found in the Ramcheck-LX manual [15]. For DDR4 characterization, we use five memory tests from Memtest86 test suit: Test-0, Test-1, Test-2, Test-3, and Test-4 [23]. We term these tests as basic tests of Memtest-86 for to simplify the explanation in further sections. We limit the number of tests in both Setup-A and Setup-B, so that duration of the test sequence does not exceed more than 10 minutes. We use this empirical limit because, beyond about 30 minutes, the setup starts accumulating ice particles which disrupt the tests and causes connectivity failures.

To determine the functionality of a DIMM we use the basic tests in the Ramcheck and Memtest86. If a DIMM passes the basic tests consistently (4 out of 5 attempts), we declare the DIMM to be functional. To find the minimum operational temperature for a DIMM, we start cooling the DIMM temperature (20K/min) while running the basic tests in parallel until DIMM fails. In the second pass, we cool the DIMM around 20K higher than the recorded temperature. Once the temperature of the DIMM is stable, we start the basic tests while maintaining the temperature by continuously pouring adequate amounts of Liquid Nitrogen. If the test is passed consistently, then we reduce the temperature by 5K and repeat the process. Otherwise, we increase the temperature and repeat the test. This process gives us a rough estimate of minimum operational temperature. We utilize basic and advanced test patterns of Ramcheck memory tester to characterize the MOT, and type of errors for DDR2 and DDR3 DIMMs.

Table 1: Specifications for Characterized DIMMS

Type	DDR2	DDR3	DDR4
No. DIMMS	24	19	12
No.Chips	384	272	96
No.Vendors	5	5	3

Table 2: Specifications of Characterization Setups

Motherboard	Rampage-V Extreme [1]
CPU	Intel-core i7-5960X [16]
Data logger	Omega TC-8 [22]
Thermocouples	Omega 5TC-TT-K-30-36

3.4 Limitations of The Experimental Setup

In this section, we will briefly discuss the limitations of the compact cryogenic heatsink and the characterization of Setup-A and Setup-B. For the existing compact cryogenic heatsink, Liquid Nitrogen is manually dispensed using 100ml canister as shown in Figure 4. This limits the precise LN2 flow-rate and the rate of temperature reduction. Manual dispensing also limits the scalability of the characterization setup as only one DIMM can be characterized at a time. A single run of the characterization experiment requires approximately 20 minutes to reduce the temperature from 300K to 80K maintain the 15-20K/min rate of change of temperature. However, these problems can be fixed by building a scaled setup which has the automated LN2 dispensing.

For Setup-A, Ramcheck can only report a single error for every experiment as it stops the tests once it encounters an error. This limits our ability to profile the errors. To solve this problem, we initiate a new test sequence from a distant randomly chosen address every time there is an error. However, this makes the tests less exhaustive for the purpose of investigating error patterns in a DIMM. Even so, Setup-A can still determine the MOT as we just need one error to declare a DIMM non-functional when performing MOT experiment. In Setup-B, the regular desktop machine could not be completely isolated from the cryogenic temperatures due to mechanical challenges. Both the Setup-A and Setup-B give some flexibility regarding changing DRAM timings. However, certain DRAM timing parameters such as refresh interval, supply voltage can not be changed beyond a limit. This severely limited our ability to tests the retention time of the functional cryogenic DRAM. Prior studies show 8 hours retention time for the IBM's Cryo-DRAM [12]. However, we could not verify retention time of DRAM due to limited reconfigurability of DRAM testers.

Cryogenic characterization is challenging due to condensation of ice and water. As the temperature of the heatsink drops, ice starts to condense on the DIMM and the tester as shown in Figure 5. The rate of condensation is directly proportional to the humidity. To solve this problem, we utilized an isolation chamber with the continuous supply of dry air and passive dehumidifiers to minimize the condensation. Also, after every experiment run, complete defrosting of the heatsink is required. Another challenge with Cryo-DRAM characterization is the failure of non-DRAM components on the DIMM. For example, we observe the failure of the serial presence detect (SPD) chip below 120K to 100K. As SPD is responsible for storing DRAM timings in a non-volatile memory, this resulted in an SPD error. We insulate SPD devices with a layer of insulation foam and glue to solve most of the SPD errors. SPD errors can also be avoided by forcing the DRAM timings.

4 EVALUATIONS

In this section, we present the findings of cryogenic DRAM characterization.

4.1 Minimum Operational Temperature

As discussed in Section 3, we characterize the Minimum Operating Temperature (MOT) using both the setup-A and the setup-B. Figure 6 shows the MOTs for DDR4 DIMMs. All the 12 characterized DIMMs use chips with the capacity of 512MB and each reported MOT value is averaged over ten trials. Figure 7 shows MOT for DDR3 DIMMs. We have characterized 19 DDR3 DIMMs with 64MB, 128MB, 256MB and 512MB chips from five vendors A, B, C, D, and E. All the reported MOT values are averaged over 20 trials. We characterize 24 DDR2 DIMMs with the chip size of 32MB, 64MB, 128MB and 256MB and evaluate their MOTs as shown in Figure 8. The reported MOTs have been averaged over 20 experimental trials. Almost 18% of DIMMs are operational at 80K to 90K temperature, and about 98% of DIMMs are operational at 150K.

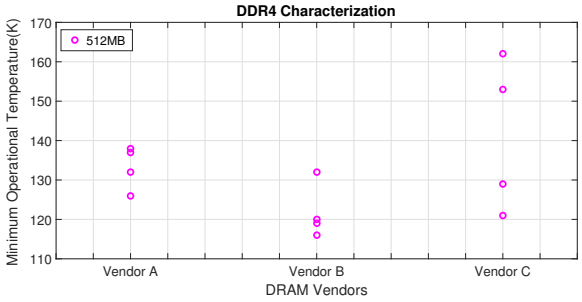

Figure 6: Minimum Operating Temperature for DDR4

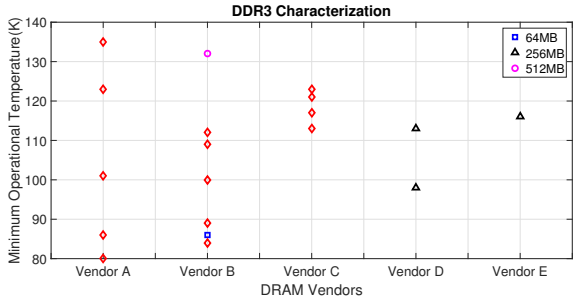

Figure 7: Minimum Operating Temperature for DDR3

Although 18% working DIMMs seems a pessimistic result, it is important to further analyze the nature of the faults in the DRAM chips. Table 3 shows a number of faulty chips in a DIMM. In case of DDR2 DIMMs, out of 24 DIMMs, 22 DIMMs show a single faulty chip, one DIMM shows two faulty chips and no DIMM shows more than three faulty chips. Moreover, on an average only 8% of the DRAM chips are reported faulty. Hence the majority of the DIMMs have a single chip failing consistently across all the experiments. This suggests a possibility of weak and strong chips wherein strong chips are resilient to the cryogenic environment. This fact can be leveraged to build DIMMs with only strong chips. Also, this

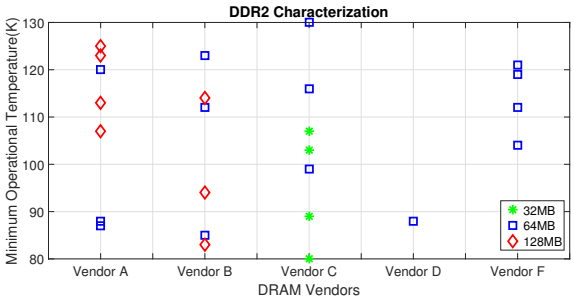

Figure 8: Minimum Operating Temperature for DDR2

opens up an opportunity to build an error correction scheme like chipkill [5] wherein a single chip failure can be tolerated.

Table 3: Number of faulty chips

No. Faulty Chips in a DIMM	No. DDR2 DIMMs	No. DDR3 DIMMs	No. DDR4 DIMMs
1	22	18	8
2	1	0	1
3	1	1	1
4	0	0	2
Total Faulty Chips	27/384=7%	21/272=8%	13/96=13.5%

Table 4: Correlation between capacity and MOT

Pearson Coefficient	0.65
Spearman Coefficient	0.67

We observe a weak correlation between the capacity of the DRAM chips and the MOT. Table 4 shows the Pearson and Spearman correlation coefficients between the MOT and the capacity of a DRAM. This result shows a correlation between technology node and the MOT as the capacity of the chip is correlated with the technology node.

4.2 Memory Errors at Cryogenic Temperatures

The objective of the error characterization of Cryo-DRAM is twofold– first, to understand if the errors are systematic or random in nature; second, to understand if the error model of uniform error distribution still holds for cryo-DRAM or not. To evaluate the error patterns, we decrease the temperature of a DIMM below MOT and record the errors for 10 trials. Table 5 shows the total number of recorded errors and percentage of two bit and single bit errors. We observe single bit errors account for more than 99.9% of the errors and the worst case errors were two-bit errors. The data suggests almost no spatial correlation within a single burst of 8 bytes. This observation implies that error correction techniques such as linear block codes (e.g. SECDED) used in the existing DRAM-based memory systems can still be used to mitigate the errors.

To investigate the possibility of systematic errors due to weak cells, we evaluate the fraction of permanent and transient errors. To this end, error on a repeated physical address is categorized as a permanent error, whereas errors with unique addresses across 10

Table 5: Number of single and double bit errors

Parameter	DDR2	DDR3	DDR4
Total faults	360	285	16543
Single Bit Faults	360	285	16539
Two Bit Faults	0	0	4

trials are declared as transient errors. Table 6 reports the fraction of transient and permanent errors. Even though our numbers of experimental trials and errors are small, we still see repeated erroneous addresses. This implies, we can statically or dynamically spare the cells to improve reliability by tolerating permanent faults.

Table 6: Fraction of transient and permanent faults

Type of Error	DDR2	DDR3	DDR4
Transient Error	36%	41%	52.5%
Permanent Error	64%	59%	47.5%

5 CONCLUSION

As quantum computers will scale from few tens of qubits to thousands of qubits, system and architectural challenges will need to be addressed. In this paper, we focus on the challenge of building a memory system for quantum computers. We highlight the need for high capacity memory which can work at cryogenic temperatures. In this paper, we evaluate the feasibility of Cryo-DRAM by characterizing off the shelf DRAM chips at the temperature range of 80K to 160K. We show that a significant fraction of the DRAM chips continue to operate at cryogenic temperatures (around 80K). Furthermore, the error patterns that are encountered at cryogenic temperatures initially tend to be uncorrelated errors, which can be mitigated with conventional schemes such as forward error correction Chipkill, and sparing. Our studies show that Cryo-DRAM can be an attractive option for building main-memory for future quantum computers.

ACKNOWLEDGEMENTS

We would like to thank Chia-Chen Chou, and Vinson Young for technical discussions, comments, and feedback. We would also like to thank the anonymous reviewers for their feedback and comments. We thank Microsoft Research for the equipment donation for this study. This work was supported in part by NSF grant 1526798.

REFERENCES

[1] ASUS. 2015. Rampage V Extreme. https://www.asus.com/us/Motherboards/RAMPAGE_V_EXTREME/. (2015). [Online; last accessed 5-May-2017].

[2] Francis Balestra and Gérard Ghibaudo. 2001. *Device and circuit cryogenic operation for low temperature electronics.* Springer.

[3] Douglas M. Carmean. 2017. Quantum and Cryo and DNA, oh my! sights along the new yellow brick road. ISCA 2016 Keynote, http://dcarmean.azurewebsites.net/ISCA2016.pdf. (2017). [Online; accessed 7-June-2017].

[4] International Business Machines Corporation. 2017. Universal Quantum Computer Development at IBM:. http://research.ibm.com/ibm-q/research/. (2017). [Online; accessed 3-April-2017].

[5] Timothy J Dell. 1997. A white paper on the benefits of chipkill-correct ECC for PC server main memory. *IBM Microelectronics Division* 11 (1997).

[6] Austin G Fowler, Adam C Whiteside, and Lloyd CL Hollenberg. 2012. Towards practical classical processing for the surface code. *Physical review letters* 108, 18 (2012), 180501.

[7] J Alex Halderman, Seth D Schoen, Nadia Heninger, William Clarkson, William Paul, Joseph A Calandrino, Ariel J Feldman, Jacob Appelbaum, and Edward W Felten. 2009. Lest we remember: cold-boot attacks on encryption keys. *Commun. ACM* 52, 5 (2009), 91–98.

[8] Thomas Häner, Damian S Steiger, Krysta Svore, and Matthias Troyer. 2016. A software methodology for compiling quantum programs. *arXiv preprint arXiv:1604.01401* (2016).

[9] Matthew B. Hastings, Dave Wecker, Bela Bauer, and Matthias Troyer. 2015. Improving Quantum Algorithms for Quantum Chemistry. *Quantum Info. Comput.* 15, 1-2 (Jan. 2015).

[10] WH Henkels, NCC Lu, W Hwang, TV Rajeevakumar, RL Franch, KA Jenkins, TJ Bucelot, DF Heidel, and MJ Immediato. 1989. A low temperature 12 ns DRAM. In *VLSI Technology, Systems and Applications, 1989. Proceedings of Technical Papers. 1989 International Symposium on*. IEEE, 32–35.

[11] WH Henkels, NCC Lu, W Hwang, TV Rajeevakumar, RL Franch, KA Jenkins, TJ Bucelot, DF Heidel, and MJ Immediato. 1989. Low temperature SER and noise in a high speed DRAM. In *Low Temperature Semiconductor Electronics, 1989., Proceedings of the Workshop on*. IEEE, 5–9.

[12] W. H. Henkels, N. C. C. Lu, W. Hwang, T. V. Rajeevakumar, R. L. Franch, K. A. Jenkins, T. J. Bucelot, D. F. Heidel, and M. J. Immediato. 1989. A low temperature 12 ns DRAM. In *VLSI Technology, Systems and Applications, 1989. Proceedings of Technical Papers. 1989 International Symposium on*. 32–35. https://doi.org/10.1109/VTSA.1989.68576

[13] D. S. Holmes, A. L. Ripple, and M. A. Manheimer. 2013. Energy-Efficient Superconducting Computing x2014;Power Budgets and Requirements. *IEEE Transactions on Applied Superconductivity* 23, 3 (June 2013), 1701610–1701610. https://doi.org/10.1109/TASC.2013.2244634

[14] J. M. Hornibrook, J. I. Colless, I. D. Conway Lamb, S. J. Pauka, H. Lu, A. C. Gossard, J. D. Watson, G. C. Gardner, S. Fallahi, M. J. Manfra, and D. J. Reilly. 2015. Cryogenic Control Architecture for Large-Scale Quantum Computing. *Phys. Rev. Applied* 3 (Feb 2015), 024010. Issue 2.

[15] Inc. Innovations. 2012. Ramcheck-LXI Owners Manual. http://www.memorytesters.com/ramcheck_lx/ramcheck_lx_manual.pdf. (2012). [Online; last accessed 5-May-2017].

[16] Intel. 2015. Intel-Core-i7-5960X. http://ark.intel.com/products/82930/Intel-Core-i7-5960X-Processor-Extreme-Edition-20M-Cache-up-to-GHz. (2015). [Online; last accessed 5-May-2017].

[17] Intel. 2015. Intel Invests US$50 Million to Advance Quantum Computing. https://newsroom.intel.com/news-releases/intel-invests-us50-million-to-advance-quantum-computing/. (2015). [Online; accessed 2-November-2016].

[18] J Kelly, R Barends, AG Fowler, A Megrant, E Jeffrey, TC White, D Sank, JY Mutus, B Campbell, Yu Chen, et al. 2016. Scalable in situ qubit calibration during repetitive error detection. *Physical Review A* 94, 3 (2016), 032321.

[19] Daniel Kudrow, Kenneth Bier, Zhaoxia Deng, Diana Franklin, Yu Tomita, Kenneth R Brown, and Frederic T Chong. 2013. Quantum rotations: a case study in static and dynamic machine-code generation for quantum computers. In *ACM SIGARCH Computer Architecture News*, Vol. 41. ACM, 166–176.

[20] Q Liu, K Fujiwara, X Meng, SR Whiteley, T Van Duzer, Nobuyuki Yoshikawa, Y Thakahashi, T Hikida, and N Kawai. 2007. Latency and power measurements on a 64-kb hybrid Josephson-CMOS memory. *IEEE transactions on applied superconductivity* 17, 2 (2007), 526–529.

[21] J. Medford, M. Stoutimore, Q. Herr, O. Naaman, H. Hearne, J. Strand, A. Przybysz, A. Pesetski, and J. Przybysz. 2015. Demonstrated control of a Transmon using a Reciprocal Quantum Logic digital circuit - Part 2. In *APS Meeting Abstracts*.

[22] Omega. 2017. TC08 Manual. https://www.omega.com/manuals/manualpdf/M4525.pdf. (2017). [Online; last accessed 5-May-2017].

[23] PassMark. 2014. Memtest86 Technical Information. http://www.memtest86.com/technical.html. (2014). [Online; last accessed 5-May-2017].

[24] Rambus. 2017. Cryogenic Memory. https://www.rambus.com/emerging-solutions/cryogenic-memory. (2017). [Online; last accessed 5-May-2017].

[25] Peter W. Shor. 1995. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM J.Sci.Statist.Comput. 26 (1997) 1484. (1995). arXiv:arXiv:quant-ph/9508027

[26] Tom Simonite. 2015. Google Quantum Dream Machine. https://www.technologyreview.com/s/544421/googles-quantum-dream-machine/?state=email-verified. (2015). [Online; accessed 2-November-2016].

[27] Tom Simonite. 2016. The Tiny Startup Racing Google to Build a Quantum Computing Chip. https://www.technologyreview.com/s/600711/the-tiny-startup-racing-google-to-build-a-quantum-computing-chip/. (2016). [Online; accessed 2-November-2016].

[28] M. Stoutimore, J. Medford, Q. Herr, O. Naaman, H. Hearne, J. Strand, A. Przybysz, A. Pesetski, and J. Przybysz. 2015. Demonstrated control of a Transmon using a Reciprocal Quantum Logic digital circuit - Part 1. In *APS Meeting Abstracts*.

[29] David B Tuckerman, Michael C Hamilton, David J Reilly, Rujun Bai, George A Hernandez, John M Hornibrook, John A Sellers, and Charles D Ellis. 2016. Flexible superconducting Nb transmission lines on thin film polyimide for quantum computing applications. *Superconductor Science and Technology* 29, 8 (2016), 084007.