scientific reports

Check for updates

OPEN Addressable superconductor integrated circuit memory from delay lines

Jennifer Volk^{1,2}, Alex Wynn², Evan Golden², Timothy Sherwood³ & Georgios Tzimpragos⁴

Recent advances in logic schemes and fabrication processes have renewed interest in using superconductor electronics for energy-efficient computing and guantum control processors. However, scalable superconducting memory still poses a challenge. To address this issue, we present an alternative to approaches that solely emphasize storage cell miniaturization by exploiting the minimal attenuation and dispersion properties of superconducting passive transmission lines to develop a delay-line memory system. This fully superconducting design operates at speeds between 20 and 100 GHz, with ± 24% and ± 13% bias margins, respectively, and demonstrates data densities in the 10s of Mbit/cm² with the MIT Lincoln Laboratory SC2 fabrication process. Additionally, the circulating nature of this design allows for minimal control circuitry, eliminates the need for data splitting and merging, and enables inexpensive implementations of sequential access and content-addressable memories. Further advances in fabrication processes suggest data densities of 100s of Mbit/cm² and beyond.

Superconductor electronics (SCEs) feature almost zero static power dissipation, speed-of-light energy-efficient interconnects, and clock rates in the 100s of GHz¹. In addition to these characteristics, SCEs can serve as facilitators for integrated classical-quantum computers due to their cryogenic nature^{2,3}. Despite advances in fabrication⁴, tools^{5,6}, and logic schemes^{7,8}, however, the lack of a reliable high-speed and high-density superconducting memory continues to impede the development of practical SCE systems⁹. In this paper, we introduce a scalable superconducting delay line memory that takes advantage of the technology's fast switching, zero resistance, and high-kinetic inductance properties.

Previous research has shown that directly applying single flux quantum (SFQ) principles to memory results in designs with low access latency but insufficient density¹⁰. While arrays of vortex transition (VT) cells have demonstrated the ability to store up to 1 Mbit of data per square centimeter¹¹, they face significant limitations for further advancement due to their reliance on superconducting transformers. On the other hand, hybrid architectures that combine SFQ and complementary metal-oxide semiconductor (CMOS) technologies provide better scalability, albeit with long access latencies¹². CMOS units may scale more effectively than their superconducting counterparts but are slower and usually reside outside of the 4.2 kelvin cryocooler due to their power and thermal footprints. In the search for a viable superconducting memory solution, a considerable effort has also been invested in superconducting memory cells built from novel superconducting-ferromagnetic stack-ups13-17. While promising in many aspects, such designs suffer from complex device structures and thus have their own practical limitations. Lastly, an approach that attempts to find a compromise between the hard-to-scale VT cells and the hard-to-fabricate superconducting-ferromagnetic hybrids is that of superconducting nanowire memory cells^{18,19}. Recent implementation results indicate a bit cell area of 26.5 μ m²²⁰, which is the most compact experimentallyverified superconducting storage element to date. However, despite their advantages, superconducting nanowire memory cells are addressed by hTrons, which bring about relatively slow access times, considerably high power consumption, and high error rates for multi-cell arrays.

In this paper, we forgo traditional array structures and propose a novel superconducting delay-line memory based on Passive Transmission Lines (PTLs). PTLs are, in essence, superconducting wires that can transmit single flux quanta with zero resistance, making them ideal for signal routing in large-scale SFQ designs²¹. If used to form loops, however, they can also be thought of as storage mediums. Figure 1 illustrates one such loop, which transmits SFQ pulses introduced at its input and delivers them after a specific time to its output. The pulses are then picked up and fed back to the loop input, where they are circulated again. Reading from and writing to the

¹Department of Electrical and Computer Engineering, UC Santa Barbara, Santa Barbara, CA 93106, USA. ²Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02420, USA. ³Department of Computer Science, UC Santa Barbara, Santa Barbara, CA 93106, USA. ⁴Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA. Memail: jevolk@ucsb.edu; gtzimpra@umich.edu



Figure 1. High-level view of the proposed memory design. The storage loop is implemented with a passive transmission line (PTL). SFQ pulses introduced at one end of the PTL (*loop_data_in*) travel along it at a controlled speed for a given time and get picked up at the output end of the line (*loop_data_out*). The memory control logic decides whether received pulses will be read out and whether they will be forwarded again to the input of the loop to repeat the cycle.

.....

memory is achieved in a time-serial manner. This enables the time-sharing of control circuitry, eliminates the need for data splitting and merging, and, in doing so, circumvents the obstacles presented by previous approaches.

The idea of using delay lines in memory implementations has roots that go back several decades. One of the earliest examples was the acoustic delay line memory, invented by Eckert and Mauchly²², and published by Auerbach et al.²³. In the mid-1950s, IBM produced the 650 calculator, which relied on a magnetic delay line memory known as drum memory^{24,25}. Approximately a decade later, Oregon State University designed NEBULA, a medium-speed serial digital computer that utilized a content-addressable memory (CAM) made from 35 glass delay lines²⁶. IBM also experimented with a similar concept in the 1970s while constructing non-volatile bubble memory. Bubble memory stores data in small magnetized "bubble" areas, and read and write operations are performed by repositioning these bubbles²⁷. More recently, racetrack memories have been developed, which promise to deliver high performance at a low cost. In this case, data is stored along a series of magnetic domain walls in non-superconducting magnetic nanowires. Reading and writing are accomplished by passing current through the nanowires, which forces domain walls to advance^{28,29}.

When it comes to superconductor electronics, the concept of realizing memory through data circulation is nearly unexplored. Hattori et al.³⁰ made an early attempt to utilize a YBa₂Cu₃O_{7- δ} (YBCO) coplanar delay line in the late 1990s for constructing high-speed cell buffer storage for asynchronous transfer mode switching systems. Another effort was made in 2016 by Ishida et al.³¹, who proposed an SFQ cache architecture that relies on circular shift registers built from synchronous destructive read-out (DRO) cells. Although these approaches provide evidence of the feasibility and speed of superconducting delay line memories, their underlying designs are limited in capacity or addressing capability. For example, in the study conducted by Hattori et al.³⁰, a GaAs 2 × 2 crossbar switch allowed for an interface to be established with the YBCO coplanar delay line. But this crossbar lacks support for addressing, non-destructive readouts, or speeds exceeding 10 GHz. In addition, the estimated signal travel speed in the YBCO line is approximately 0.4*c*, where *c* is the speed of light. This implies that the minimum spacing between two subsequent data signals in the YBCO line would be about 12 mm. Hence, both the controller design and the line material impose significant constraints on the memory density. Regarding the DRO-based cache architecture³¹, synchronous cells are used to form a circular shift register, which hampers energy and area efficiency. Furthermore, shifting is controlled by a sequence of clock pulses whose length is determined by the provided address and the register's current position, thereby resulting in additional overhead.

The proposed PTL-based memory design, in contrast, employs primarily passive components; is fully superconducting; encompasses all typical memory functionalities, such as addressing, data overwrite, and non-destructive readout; and achieves interface speeds of up to 100 GHz in simulation with satisfactory bias margins. The high controller (pulse injection) speed, coupled with the SFQ slow-down caused by high-kinetic inductance materials, boosts memory density by reducing the minimum spacing between subsequent pulses in PTLs. For example, the minimum spacing in NbN nanowires is about 570 times shorter than that of YBCO lines³⁰. To validate these hypotheses and quantify the projected gains, we conduct detailed analog simulations and formulate models that establish the relationship between memory density and various factors such as the operating frequency of the interface circuitry, pulse travel speed in the PTL, and line dimensions. For the latter two, we investigate PTL designs with different topologies, material compositions, and fabrication processes.

Results

Architecture description and functional evaluation

The block diagram of the proposed memory is shown in Fig. 1. The design consists of a PTL-based delay line and a control logic block. The delay line serves as the circulating loop storage and delays any data that arrives at its input (*loop_data_in*). The delay introduced by the loop depends on the line's length and the pulse travel speed in the line. At the end of each round trip, the data at the output of the line (*loop_data_out*) enters the controller, which serves as a memory interface. The controller is responsible for deciding whether signals from the feedback path (*loop_data_out*) or the input (*write_data*) will be forwarded to the delay line (*loop_data_in*) for another round and whether it will be copied and forwarded to the readout port (*read_data*).

Figure 2 illustrates the schematic of the controller. Temporally-encoded SFQ signals, generated by comparing the value of an address counter with a target address, are used for addressing. The Merger cell, denoted by the letter *m*, stitches together and forwards all signals that appear on its two input lines to its single output line. When no pulse arrives on the *write_address* signal line within the designated interval, a pulse appears on its complementary signal line, $\neg write_address$. A pulse on the *loop_data_out* line then flows from the DRO2R (DRO with two outputs) on the left into the delay line input (*loop_data_in*) on the right, without waiting for other signals. Inversely, when a pulse arrives on the *write_address* line, a pulse on the *loop_data_out* line input and readout circuitry. The use of differential signaling for write addressing enables the correction of potential data timing distortions incurred in the control circuitry and the storage loop. The readout circuitry on the first DRO2R, there are two cases: either a pulse arrives through the *loop_data_in* line and pushes the stored value to the Q0 output port (*read_data*), or a pulse on the complementary *¬read_address* line clears the cell, flushing the stored value.

Simulation results for this design are provided in Fig. 3. In both cases, the controller operates at 100 GHz, and three full rotations, or trips, are shown. Each trip consists of four intervals, with three of them corresponding to the number of supported memory addresses and one serving as the header, denoted by *h*. More specifically, in Fig. 3a, a pulse is provided on the *write_data* line during the header interval of the first rotation, trip 0. In the third interval of the same trip, a pulse arrives on each of the *write_address* and *read_address* lines, denoting a write to and read from address 1. Upon the arrival of the *write_address* pulse, a pulse appears on the *loop_data_in* line, which demonstrates a successful memory write operation. The subsequent appearance of a pulse on the *read_data* line after the arrival of the pulse on the *read_address* line indicates that write operations have higher priority than read. To illustrate the non-destructive nature of readout, in trip 1 of Fig. 3a, a pulse is asserted on the *read_data* line, which evidences the desired behavior. To demonstrate data overwrite, the same pulse ordering as before is used to set up the memory in Fig. 3b. However, in this case, a pulse is asserted on the *write_address* line agrees on the *loop_data_in* line after this operation, nucleating the first pulse. As anticipated, no pulse appears on the *loop_data_in* line after this operation, he write_address line during trip 1 without being accompanied by a second *write_data* pulse. As anticipated, no pulse appears on the *loop_data_in* line after this operation, indicating that the first pulse was successfully overwritten.

Note that the presented memory system allows one to search and operate on all of the memory contents while waiting for the entire circulation time to pass, thereby eliminating the need to broadcast to or continuously poll individual cells. The design's rotating nature not only circumvents classic fan-in and fan-out limitations of superconductor electronics but also supports the addition of multiple write and read ports and the inexpensive implementation of content-addressable memories²⁶.

Circuit descriptions and performance evaluation

To evaluate the performance and feasibility of the proposed memory design, we first provide schematics and simulation results for the memory controller's main components; next, we analyze their latency; lastly, we perform a voltage bias margin analysis for the entire system, including all loading effects due to the control logic,



Figure 2. Block diagram of a controller enabling sequential-access addressing. A destructive readout (DRO) cell stores a *write_data* pulse until *write_address* arrives. A DRO cell with two readout ports (DRO2R) is used to re-synchronize data on *loop_data_out* with the arrival of a pulse on the ¬*write_address* line; or, conversely, *loop_data_out* is filtered out upon the arrival of *write_address*. For reading, another DRO2R cell is used. In this case, a pulse on the *read_address* line loads this DRO2R cell, and a subsequent *loop_data_in* or ¬*read_address* pulse triggers it. If *loop_data_in* arrives first, the stored pulse is forwarded to the *read_data* line; otherwise, it is filtered out. A Merger cell, denoted by *m*, joins two lines: the first from the output of the DRO for a write operation, and the second from the DRO2R for memory loop re-circulation. A fan-out of two is required after the Merger; cell *I_C* ranking is adopted to maintain signal fidelity around this point⁵⁹. The final design uses serial resistors at the boundaries of the PTL to eliminate stray DC currents caused by trapped fluxons and to provide damping for possible reflections due to impedance mismatches.



Figure 3. WRSPICE⁴³ simulation results of the presented superconducting delay-line memory using MIT Lincoln Laboratory's $100 \,\mu$ A/ μ m² SFQ5ee process parameters. The operating frequency is set to 100 GHz, and the number of supported addresses is set to three. A header interval, *h*, is needed at the beginning of each memory rotation, or trip, to load data for writing. (a) The loop starts empty and a pulse on the *write_data* signal line is provided in the header interval of trip 0. During the third interval of the same trip, a pulse is observed on both the *write_address* and *read_address* lines, indicating a write operation to and read operation from address 1. After the *write_address* pulse arrival, a pulse appears on the *loop_data_in* line, demonstrating a successful memory write operation. After the *read_address* pulse arrives, a pulse is observed on the *read_data* line, indicating a successful memory read operation. During the following two trips, the pulse stored in address 1 is successfully retrieved again, which provides evidence of the memory's non-destructive readout capability. (b) The first rotation, trip 0, is identical to that of (a). However, in trip 1, the pulse stored in address 1 is successfully overwritten, as indicated by the absence of pulses in the *loop_data_in* and *loop_data_out* lines after the overwrite.

PTL, and accompanying driver and receiver circuitry. The controller, shown in Fig. 2, consists of a DRO cell, two DRO2R cells, and a Merger cell (m). Figures 4, 5, and 6 provide schematics for each cell as well as corresponding simulation waveforms to demonstrate cell function.

As is the case for any system, the electrical and timing properties of these cells affect both the performance and functionality of the proposed memory. In particular, electrical issues, typically brought on by susceptibility to parametric variation, can lead to fatally under- or over-biased Josephson Junctions (JJs), which in turn can lead to circuit dysfunction, or delayed or early switching times. To avoid erroneous behavior and ensure correct system timing, the effects of under- and over-biasing are first examined at the cell level. Performing this bias analysis for cells in isolation, however, is not sufficient because it excludes the loading effects that are present in a system setting. To account for loading, iterative measurements and component tuning are carried out in an in-situ approach to achieve the desired timing. Accordingly, each cell is fully loaded by the remaining components in the memory controller. The results of this process are shown in Fig. 7. Nominal cell delays are indicated in red. DRO and DRO2R delays are measured as clock-to-Q delays, while the Merger delay is measured as the propagation delay from either of the inputs to the output. Delays in each cell increase as bias decreases, and decrease as bias



time (ps)

Figure 4. Symbol (**a**), schematic (**b**), and simulation results (**c**) of a destructive readout (DRO) cell. An incoming *data_in* SFQ pulse is stored in the superconducting quantum interference device (SQUID) formed by B2-L2-B3 until a *clock* pulse arrives. The arrival of a *clock* pulse switches B3 and releases an SFQ pulse on *data_out*. The corresponding circuit parameters are: L0 = 0.7 pH; L1 = 1.3 pH; L2 = 5 pH; L3 = 1.6 pH; B0, B3 = 227 μ A; B1, B2 = 188 μ A; and I0 = 125 μ A. Shunt resistors for B0 and B3 are 3.3 Ω and for B1 and B2 are 3.4 Ω .



Figure 5. Symbol (**a**), schematic (**b**), and simulation results (**c**) of a DRO cell with two readout ports (DRO2R). The DRO2R cell performs largely the same operation as the DRO—but in this case, the storage element is shared between two parallel loops: B4-L5-B2-L2-B3 and B4-L5-B5-L8-B8⁴⁶. A pulse appearing on either clock input will clear the stored SFQ and push it to the respective output line. The corresponding circuit parameters are: L0, L6 = 0.3 pH; L1, L3 = 2 pH; L2, L8 = 0.12 pH; L4 = 0.5 pH; L5 = 5 pH; L7 = 3 pH; L9 = 5.5 pH; L10 = 0.52 pH; B0, B7 = 325 μ A; B1, B6, B8 = 162 μ A; B2, B5 = 197 μ A; B3 = 152 μ A; B4 = 237 μ A; B9 = 223 μ A; and I0, I1, I2, I3 = 176 μ A. Shunt resistors for B0 and B7 are 2 Ω ; for B1, B3, B6, and B8, 4.2 Ω ; for B2 and B5, 3.4 Ω ; for B4, 2.8 Ω ; and for B9, 3 Ω .



Figure 6. Symbol (**a**), schematic (**b**), and simulation results (**c**) of a Merger cell. As its name implies, this design passes incoming SFQ pulses from either of its two input ports to its output. To prevent backward propagation of an input SFQ from the opposite line, two blocking Josephson junctions, B0 and B3, are used. The corresponding circuit parameters are: L0 = 1 pH; L1 = 0.46 pH; L2 = 0.2 pH; L3 = 1.3 pH; L4 = 0.8 pH; B0, B3 = 197 µA; B1, B2, B4 = 220 µA; and I0 = 500 µA. Shunt resistors for B1, B2, and B4 are all 3Ω .



Figure 7. Bias versus propagation delay for DRO (**a**), DRO2R (**b**), and Merger (**c**) cells. Red markings indicate nominal values. To achieve symmetric bias margins, the nominal delay of each cell is centered between its upper and lower time bounds.

increases. To make bias margins symmetric, a set of component parameters is chosen that centers the nominal delay of each cell between the upper and lower time bounds of the controller.

Using interval analysis and the above delays, the maximum operating frequency of the controller is estimated, and the cell tuning and bias margin measurements are repeated. Another round of bias margin measurements is conducted, wherein the bias of the entire design is varied instead of individual cells. Figure 8 illustrates our results for frequencies ranging from 20 to 100 GHz. We notice that electrical issues—caused by, for example, Josephson junctions that are subject to the above biasing concerns and that switch too frequently or not at all—drive limitations in bias margin width at lower frequencies, while timing issues are the limiter at higher



Figure 8. Bias margins of the complete memory design for a variety of operating frequencies.

.....

frequencies. This happens because timing constraints get tighter as the address timing interval is reduced. For example, at 100 GHz, the address timing interval is just 10 ps, which leaves little room for the same variations in propagation delay that we observed in Fig. 7. Our SPICE simulations show bias margins ranging from \pm 24% (at 20 GHz) to \pm 13% (at 100 GHz), which surpass the widely-accepted \pm 10% threshold³².

Data density estimation

The physical storage density—that is, bits per area—of the proposed memory depends on (1) the PTL's linewidth and spacing requirements, set by the fabrication process; (2) the travel speed of SFQ pulses in the PTL, set by the material of choice and the line topology; (3) the relative timing between two adjacent SFQ pulses, set by the controller's operating frequency; and (4) the number of PTL memory routing layers. We estimate the density of the proposed PTL-based superconducting delay line memory by choosing various settings for each of these free variables and summarize our results in Table 1.

A typical Nb stripline of 250 nm linewidth with a minimum spacing of 250 nm⁴ propagates SFQ pulses at a speed of 0.3*c*. This leads to data densities of up to 0.9 Mbit/cm² at 100 GHz, if four metal routing layers are used. Reducing the Nb stripline linewidth and minimum spacing from 250 to 120 nm is a possible but more aggressive design choice³³ and results in densities of up to 1.9 Mbit/cm². By switching device material and topology to an MoN kinetic inductor microstrip with the same dimensions, available on just one layer within MIT Lincoln Laboratory's SFQ5ee³⁴ and SC2³³ processes, the travel speed of pulses in the line falls by about 6×. This slowdown yields densities of up to 1.4 and 4.0 Mbit/cm² for 250 nm and 120 nm linewidths, respectively, at 100 GHz. By default, the SFQ5ee and SC2 fabrication processes allow for one MoN high-kinetic inductance layer, four Nb signal-routing layers, and three ground planes, based on the typical allocation. An increase in the number of the MoN high-kinetic inductance layers from one to four transforms the line topology into that of a stripline and increases the data density to 3.2 Mbit/cm² at 20 GHz and 19 Mbit/cm² at 100 GHz for a 120 nm linewidth and 120 nm spacing.

At this point, it is evident that the use of materials with increasingly high kinetic inductance is conducive to higher densities. To this end, we explore the potential of NbTiN striplines that exhibit approximately an order of magnitude higher inductance than their MoN counterparts and propagate SFQ pulses at a speed of $0.011c^{35}$. Our results indicate that for a NbTiN stripline with 100 nm width, 120 nm spacing, four metal routing layers, and controller frequencies between 20 and 100 GHz, the estimated data densities range from 10.7 to 53.3 Mbit/cm².

A more forward-looking approach comes from the use of NbN kinetic inductor nanowires. In the case of an experimentally-tested NbN nanowire with 40 nm linewidth, the inductance scales to 2050 pH/ μ m³⁶. A roughly proportional drop in capacitance keeps the pulse travel speed the same, 0.011*c*, but the reduced linewidth pushes the maximum data density to 75.4 Mbit/cm² at 100 GHz. A reduction in linewidth to 15 nm causes the inductance to increase to 5467 pH/ μ m³⁶ based on a factor of 82 pH/ \Box , which drops the travel speed to 0.007*c* and increases data density to 140.3 Mbit/cm² at 100 GHz. This is equivalent to a physical pulse spacing of 21 μ m, about 600× shorter than that in the YBCO line³⁰. Moreover, assuming that CMOS layer stacking techniques—such as those used to create 100-layer stacks in V-NANDs—could be applied to superconductor technology, this NbN nanowire technology could provide a memory density of 3507 Mbit/cm² at 100 GHz of Nbit/cm² at 100 GHz.

Discussion

The microarchitectural features of our most prominent computer systems have historically co-evolved with the technology and devices that embody them. For instance, in the case of CMOS, the combination of cheap transistors and power-hungry interconnects³⁷ has motivated the development of designs that trade off resource redundancy for reduced data movement. By contrast, data movement in superconductor electronics is relatively cheaper than switching due to the nearly lossless nature of superconducting interconnects. Therefore, the critical question—which the presented research addresses—is whether, in the case of superconductor electronics, we can capitalize on the unique properties of superconducting interconnects to improve hardware efficiency, simplicity, and density.

Device	Linewidth (nm)	Spacing (nm)	Fabrication process	Memory layers	Process maturity	Capacitance (fF/µm)	Inductance (pH/µm)	Travel speed (xc)	Frequency (GHz)	Density (Mbit/cm ²)
Nb stripline	250	250	SFQ5ee	4	Mature	0.25	0.50	0.298	20	0.2
									50	0.4
									75	0.7
									100	0.9
	120	120	SC2	4	Aggressive	0.19	0.65	0.296	20	0.4
									50	0.9
									75	1.4
									100	1.9
MoN kinetic inductor microstrip	250	250	SFQ5ee	1	Mature	0.16	32	0.047	20	0.3
									50	0.7
									75	1.1
									100	1.4
	120	120	SC2	1	Aggressive	0.14	66.70	0.034	20	0.8
									50	2.0
									75	3.0
									100	4.0
MoN kinetic inductor stripline	120	120	SC2	4	Aggressive	0.19	66.70	0.029	20	3.2
									50	8.1
									75	12.1
									100	19.0
NbTiN kinetic inductor stripline	100	120	Not estab- lished	4	Academic	0.17	490.5	0.011	20	10.7
									50	26.6
									75	40.0
									100	53.3
NbN kinetic inductor nanowire	40	120	Not estab- lished	4	Academic	0.04	2050	0.011	20	15.1
									50	37.7
									75	56.6
									100	75.4
	15	120	Not estab- lished	4	Academic	0.04	5467	0.007	20	28.1
									50	70.1
									75	105.2
									100	140.3
	15	120	Not estab- lished	100	Academic	0.04	5467	0.007	20	701.4
									50	1753
									75	2630
									100	3507

Table 1. Memory density estimates for a variety of mature, aggressive, and academic fabrication processes. The MIT Lincoln Laboratory SFQ5ee $100 \,\mu A/^2 m^2$ process has served as the de facto standard for fabrication since its introduction in 2016, and thus is considered mature³⁴. The MIT Lincoln Laboratory SC2³³ is considered aggressive, as it has yielded successful experimental circuits, but pushes the bounds on common design rules. The term academic refers to processes that have been purpose-built and evaluated either experimentally or theoretically.

Previous attempts to develop cryogenic memory technology, which functions at temperatures of 4 kelvin or below, have predominantly centered on creating individual storage cells that can subsequently be arranged in a standard grid configuration³⁸. While such configurations have been effective in CMOS, they cannot fully leverage the advantage that superconductor electronics provide in terms of low-cost signal transmission. Additionally, they often disregard the forthcoming obstacles that arise from the limitations of inductor scaling and fan-in/ out¹¹, as well as the significant power consumption and long access times¹², high device complexity¹³⁻¹⁷, and

increased bit-error rate¹⁸⁻²⁰ that come with the suggested structures.

The presented approach departs from this tradition and exploits superconducting passive transmission lines for the construction of a delay-line memory system. From a device perspective, delay lines are easy to construct and offer high signal fidelity. For instance, experimental results indicate that SFQ pulses can travel over 7 mm-long PTLs without re-amplification³⁹. Moreover, the footprint and kinetic inductance of PTLs can be substantially modified by their topology, material composition, and fabrication process. By harnessing these characteristics in conjunction with our high-speed SFQ control logic circuitry, we anticipate achieving data density improvements of two orders of magnitude compared to the current state of the art. This can be accomplished while retaining the same number of metal (memory) layers as existing fabrication processes. The method of vertical

growth—common in technologies like CMOS V-NANDs—is also anticipated to increase data density gains by an additional order of magnitude.

Looking forward, although there are no fundamental limitations to the development of scalable superconducting delay-line memories, a direction that invites further exploration is that of impedance-matching transformers, such as tapers⁴⁰. Alternatively, in order to address any potential impedance-mismatch challenges, exploring methods to boost the transmission line capacitance and utilizing higher-permittivity dielectrics, like NbOx⁴¹, between the signal and ground planes could prove to be a promising approach. Lastly, there are several interesting questions pertaining to the microarchitecture-compiler relationship. Specifically, optimizations in hierarchical design, instruction scheduling, and data placement can significantly reduce the memory readout time and improve its efficiency in cases involving both sequential access and content-based addressing⁴².

Methods

Memory control circuit design and analysis

Simulation setup

The simulations were conducted using WRSPICE⁴³, a version of SPICE designed for superconductor electronic designs. The resistively and capacitively-shunted JJ model employed in the simulations is based on MIT Lincoln Laboratory's SFQ5ee $100 \,\mu A/\mu m^2$ fabrication process³⁴ with an I_CR_n value of 1.65 mV. Bias sources consisted of resistors tied to a voltage bias line, with resistor values chosen for the equivalent current listed. Bias was ramped up from zero to the nominal value. To generate SFQ input pulses from an input DC current stimulation, DC-to-SFQ converters were utilized. The subsequent analysis does not consider parasitic inductances associated with shunt resistors, as their impact on circuit performance has been determined to be negligible⁴⁴.

Circuit design

For the design of the memory controller's key components, as illustrated in Figs. 4, 5, and 6, we began with publicly accessible schematics^{45,46}. To achieve a higher operational frequency and wider bias margins, the number of JJs on the critical path of the corresponding circuit was minimized. To accomplish this, we designed a DRO cell with 4 JJs that matched the critical currents of JJs in the neighboring cells, resulting in improved signal quality. Additionally, the original DRO2R design features output paths that are symmetrical. However, as depicted in Fig. 2, only the Q0 output (*data_out0*) is connected to the delay line loop. To meet the timing demands of the loop, the internal path from *data_in* to *data_out0* was shortened. The serial inductors in the Merger design were also reduced from their initial values to improve critical path latency. Finally, we eliminated active Splitter cells⁴⁵, which are commonly used to provide fan-out for shared nodes but incur an overhead of 3 JJs each, by adopting a recently-proposed I_C ranking technique⁴⁷. The combined modifications resulted in a memory controller circuit that comprises 29 JJs and a logic path delay of ~ 10 ps, which is determined by the sum of the DRO2R cell's setup and propagation delays, averaged over its bias margins.

Analysis

To evaluate performance, the timing resolution was set to 0.5 ps, which was interpolated from the internal step size used by WRSPICE, and delays were measured as peak-to-peak values. Specifically, the DRO and DRO2R propagation delays were measured as the clock-to-Q delay, and the Merger propagation delay was measured as the delay from either input to the output. To determine the upper and lower time bounds of each cell, detailed bias margin analyses were conducted. For bias margin measurements, we started at the nominal voltage and decremented by steps of 1% of the nominal voltage over the bias range where the circuit operates, to locate the lower limit. We then incremented the nominal voltage to locate the upper limit. Static timing analysis using minimum and maximum intervals was performed to ensure that the timing at the limits of the bias margins meets the hold, setup, and propagation time requirements at different target frequencies.

Memory density analysis

The memory density of the PTL delay line was calculated using the equation:

$$Density = \frac{f}{w \times v},\tag{1}$$

where *f* represents the operating frequency of the memory controller in Hz, *w* is the PTL pitch in m, and *v* is the travel speed of single flux quanta in m/s.

Controller operational frequency and PTL pitch

The delay results shown in Fig. 7 indicate a maximum controller operational frequency of 100 GHz. In the conducted memory density analysis, as summarized in Table 1, the value of *f* is selected to span from this value down to 20 GHz to highlight the trade-off between density versus bias margins, as suggested by the findings presented in Fig. 8. Regarding the PTL pitch—the sum of the PTL linewidth and minimum spacing requirement—this is typically determined by the fabrication process. For example, the well-established SFQ5ee³⁴ process allows for the reliable fabrication of Nb and MoN microstrips and striplines with a 250 nm linewidth and 250 nm spacing, or a pitch of 500 nm. The more advanced SC2³³ process can reduce the pitch to 240 nm. Further miniaturization down to a pitch of 135 nm and smaller is possible using e-beam lithography³⁶.

Travel speed

The velocity factor equation was used to determine the travel speed (ν) of single flux quanta in PTLs constructed with different materials. The equation is given by:

$$v = \frac{1}{c\sqrt{L \times C}}.$$
(2)

Here, *c* is the speed of light in m/s, *L* is the inductance per unit length in H/m, and *C* is the capacitance per unit length in F/m. The capacitance and inductance values used in the calculations were sourced from literature and are shown in Table 1. In cases where geometric inductance is significant, such as in Nb striplines, the inductance scales non-linearly and must be evaluated at each linewidth individually. Based on available experimental measurements³³, a 250 nm-wide Nb stripline has 0.5 pH/ μ m, while a 120 nm-wide line has 0.65 pH/ μ m. For high-kinetic inductors like MoN and NbTiN striplines and MoN microstrips, the kinetic inductance is directly proportional to the length of the inductor and inversely proportional to its width. This can be calculated by multiplying the known inductance per square values of 8 pH/ \Box for MoN striplines and microstrips⁴⁸ and 49 pH/ \Box for NbTiN striplines³⁵ by the unit length divided by the linewidth. Regarding their capacitances, we relied on recently-published fabrication results³³. For instance, a stripline having a linewidth of 120 nm has a capacitance of 0.19 fF/µm, which is solely determined by the dielectric material and geometry. Similarly, a microstrip having a linewidth of 120 nm has a capacitance of 0.14 fF/µm. For the 40 nm NbN nanowires, experimental measurements yielded an inductance of 82 pH/ \Box or 2050 pH/ μ m for this width, and a capacitance of 0.044 fF/ μ m³⁶. Lastly, inductance and capacitance values for 15 nm NbN nanowires were obtained from simulation results³⁶. Effects of coherent quantum phase slips (CQPS) have been observed in NbN films with comparable crosssectional dimensions⁴⁹. Nevertheless, these effects do not present an issue here, as the kinetic inductance of the films, even with the most extreme linewidths considered, is still one order of magnitude lower than what has been used to observe CQPS. Consequently, the phase-slip amplitude in our case will be e^{-9} times lower than in systems in which CQPS has been shown to be a significant contributor to conductivity⁵⁰, and for this reason, we neglect it in our analysis.

Layer stacking

An approach to further enhance memory density is by vertically stacking multiple layers of PTLs. In this case, the memory density increases linearly with the number of PTL layers, denoted by *N* in the equation that follows:

$$Density = \frac{f}{w \times v} \times N.$$
(3)

In this regard, a stackup of Nb striplines with nine planarized superconducting layers, stackable stud vias, selfshunted Nb/AlOx–Al/Nb Josephson junctions, and a single layer of MoN kinetic inductors has been successfully developed and currently serves as the de facto fabrication process^{4,51}. This stackup, MIT Lincoln Laboratory's state-of-the-art SFQ5ee process, allows for four memory layers based on the typical allocation of ground planes and signal planes. Several advancements to this process are under development, including the addition of selfshunted junctions⁵², which use higher critical current density and can reduce the chip area of digital hardware by 50%, and the introduction of two additional routing layers⁵³. Looking forward, foundries have recently announced plans to scale up to sixteen routing layers and one layer for NbTiN kinetic inductors⁵⁴; to the best of our knowledge, however, there is no fundamental limitation that constrains vertical growth^{45,55-58}.

Data availability

All data relevant to the study are available from the corresponding authors upon request.

Received: 20 May 2023; Accepted: 21 September 2023 Published online: 03 October 2023

References

- Holmes, D. S., Kadin, A. M. & Johnson, M. W. Superconducting computing in large-scale hybrid systems. *Computer* 48, 34–42. https://doi.org/10.1109/MC.2015.375 (2015).
- Li, K., McDermott, R. & Vavilov, M. G. Hardware-efficient qubit control with single-flux-quantum pulse sequences. *Phys. Rev. Appl.* 12, 014044. https://doi.org/10.1103/PhysRevApplied.12.014044 (2019).
- Holmes, A. et al. NISQ+: Boosting quantum computing power by approximating quantum error correction. In 2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA) 556–569. https://doi.org/10.1109/ISCA45697.2020.00053 (2020).
- Tolpygo, S. K. et al. Developments toward a 250-nm, fully planarized fabrication process with ten superconducting layers and self-shunted Josephson junctions. In 2017 16th International Superconductive Electronics Conference (ISEC) 1–3. https://doi.org/ 10.1109/ISEC.2017.8314189 (2017).
- Fourie, C. J. *et al.* ColdFlux superconducting EDA and TCAD tools project: Overview and progress. *IEEE Trans. Appl. Supercond.* 29, 1–7. https://doi.org/10.1109/TASC.2019.2892115 (2019).
- Christensen, M. et al. PyLSE: A pulse-transfer level language for superconductor electronics. In Proc. 43rd ACM SIGPLAN International Conference on Programming Language Design and Implementation PLDI 2022 (Association for Computing Machinery, 2022).
- Tzimpragos, G. et al. A computational temporal logic for superconducting accelerators. In Proc. Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS 20 435–448. https://doi.org/10.1145/ 3373376.3378517 (Association for Computing Machinery, 2020).

- 8. Tzimpragos, G., Volk, J., Wynn, A., Smith, J. E. & Sherwood, T. Superconducting computing with alternating logic elements. In 2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA) 651-664. https://doi.org/10.1109/ISCA5 2012.2021.00057 (2021).
- 9. Holmes, D. S., Ripple, A. L. & Manheimer, M. A. Energy-efficient superconducting computing-power budgets and requirements. IEEE Trans. Appl. Supercond. 23, 1701610. https://doi.org/10.1109/TASC.2013.2244634 (2013).
- 10. Nagasawa, S., Numata, H., Hashimoto, Y. & Tahara, S. High-frequency clock operation of Josephson 256-word/spl times/16-bit RAMs. IEEE Trans. Appl. Supercond. 9, 3708-3713. https://doi.org/10.1109/77.783834 (1999).
- 11. Semenov, V. K., Polyakov, Y. A. & Tolpygo, S. K. Very large scale integration of Josephson-junction-based superconductor random access memories. IEEE Trans. Appl. Supercond. 29, 1-9. https://doi.org/10.1109/TASC.2019.2904971 (2019)
- 12. Van Duzer, T. et al. 64-kb hybrid Josephson-CMOS 4 Kelvin RAM with 400 ps access time and 12 MW read power. IEEE Trans. Appl. Supercond. 23, 1700504. https://doi.org/10.1109/TASC.2012.2230294 (2013).
- 13. Feofanov, A. K. et al. Implementation of superconductor/ferromagnet/superconductor *π*-shifters in superconducting digital and quantum circuits. Nat. Phys. 6, 593-597. https://doi.org/10.1038/nphys1700 (2010).
- Vernik, I. V. et al. Magnetic Josephson junctions with superconducting interlayer for cryogenic memory. IEEE Trans. Appl. Super*cond.* 23, 1701208. https://doi.org/10.1109/TASC.2012.2233270 (2013).
 Baek, B., Rippard, W. H., Benz, S. P., Russek, S. E. & Dresselhaus, P. D. Hybrid superconducting-magnetic memory device using
- competing order parameters. Nat. Commun. 5, 4888. https://doi.org/10.1038/ncomms4888 (2014).
- Gingrich, E. C. et al. Controllable 0-π Josephson junctions containing a ferromagnetic spin valve. Nat. Phys. 12, 564–567. https:// 16 doi.org/10.1038/nphys3681 (2016).
- 17. Nguyen, M.-H. et al. Cryogenic memory architecture integrating spin hall effect based magnetic memory and superconductive cryotron devices. Sci. Rep. 10, 9. https://doi.org/10.1038/s41598-019-57137-9 (2019).
- 18. Murphy, A., Averin, D. V. & Bezryadin, A. Nanoscale superconducting memory based on the kinetic inductance of asymmetric nanowire loops. New J. Phys. 19, 063015. https://doi.org/10.1088/1367-2630/aa7331 (2017).
- 19 Zhao, Q.-Y. et al. A compact superconducting nanowire memory element operated by nanowire cryotrons. Supercond. Sci. Technol. 31, 035009. https://doi.org/10.1088/1361-6668/aaa820 (2018).
- 20. Butters, B. A. et al. A scalable superconducting nanowire memory cell and preliminary array test. Supercond. Sci. Technol. 34, 035003. https://doi.org/10.1088/1361-6668/abd14e (2021).
- 21. Kameda, Y., Yorozu, S. & Hashimoto, Y. A new design methodology for single-flux-quantum (SFQ) logic circuits using passivetransmission-line (PTL) wiring. IEEE Trans. Appl. Supercond. 17, 508-511. https://doi.org/10.1109/TASC.2007.898718 (2007).
- 22. Eckert, J. J. P. & Mauchly, J. W. Memory System. US Patent 2,629,827 (1953).
- Auerbach, I., Eckert, J., Shaw, R. & Sheppard, C. Mercury delay line memory using a pulse rate of several megacycles. Proc. IRE 23 37, 855-861. https://doi.org/10.1109/JRPROC.1949.229683 (1949).
- 24. Hamilton, F. E. & Kubie, E. C. The IBM magnetic drum calculator type 650. J. ACM 1, 13-20. https://doi.org/10.1145/320764. 320768 (1954).
- 25. Frankel, S. Useful applications of a magnetic-drum computer. Electr. Eng. 75, 634-639. https://doi.org/10.1109/EE.1956.6442018 (1956)
- 26. Rux, P. A glass delay line content-addressed memory system. IEEE Trans. Comput. 18, 512-520. https://doi.org/10.1109/T-C.1969. 222703 (1969).
- 27 Bonyhard, P. et al. Magnetic bubble memory chip design. IEEE Trans. Magn. 9, 433-436. https://doi.org/10.1109/TMAG.1973. 1067599 (1973).
- 28. Parkin, S. S. P., Hayashi, M. & Thomas, L. Magnetic domain-wall racetrack memory. Science 320, 190-194. https://doi.org/10.1126/ science.1145799 (2008).
- 29. Parkin, S. & Yang, S.-H. Memory on the racetrack. Nat. Nanotechnol. 10, 195–198. https://doi.org/10.1038/nnano.2015.41 (2015).
- 30. Hattori, W., Yoshitake, T. & Tahara, S. A reentrant delay-line memory using a YBa2Cu3O7-6 coplanar delay-line. IEEE Trans. Appl. Supercond. 9, 3829-3832. https://doi.org/10.1109/77.783862 (1999).
- Ishida, K., Tanaka, M., Ono, T. & Inoue, K. Single-flux-quantum cache memory architecture. In 2016 International SoC Design Conference (ISOCC) 105-106. https://doi.org/10.1109/ISOCC.2016.7799755 (2016).
- 32. Hashimoto, Y. et al. Design and investigation of gate-to-gate passive interconnections for SFQ logic circuits. IEEE Trans. Appl. Supercond. 15, 3814-3820. https://doi.org/10.1109/TASC.2005.847487 (2005).
- Tolpygo, S. K., Golden, E. B., Weir, T. J. & Bolkhovsky, V. Inductance of superconductor integrated circuit features with sizes down 33. to 120 nm. Supercond. Sci. Technol. 34, 085005. https://doi.org/10.1088/1361-6668/ac04b9 (2021).
- 34. Tolpygo, S. K. et al. Advanced fabrication processes for superconducting very large-scale integrated circuits. IEEE Trans. Appl. Supercond. 26, 1-10. https://doi.org/10.1109/TASC.2016.2519388 (2016).
- 35. Hazard, T. M. et al. Nanowire superinductance fluxonium qubit. Phys. Rev. Lett. 122, 010504. https://doi.org/10.1103/PhysRevLett. 122.010504 (2019).
- Niepce, D., Burnett, J. & Bylander, J. High kinetic inductance NbN nanowire superinductors. Phys. Rev. Appl. 11, 044014. https:// 36. doi.org/10.1103/PhysRevApplied.11.044014 (2019).
- Magen, N., Kolodny, A., Weiser, U. & Shamir, N. Interconnect-power dissipation in a microprocessor. In Proc. 2004 International 37. Workshop on System Level Interconnect Prediction, SLIP 04 7–13. https://doi.org/10.1145/966747.966750 (Association for Computing Machinery, 2004).
- Alam, S., Hossain, M. S., Srinivasa, S. R. & Aziz, A. Cryogenic memory technologies. Nat. Electron. 6, 1-14 (2023).
- Talanov, V. V. et al. Propagation of picosecond pulses on superconducting transmission line interconnects. Supercond. Sci. Technol. 39. 35, 055011 (2022)
- 40. Klopfenstein, R. W. A transmission line taper of improved design. Proc. IRE 44, 31-35. https://doi.org/10.1109/JRPROC.1956. 274847 (1956).
- 41. Zhao, Y., Zhang, Z. & Lin, Y. Optical and dielectric properties of a nanostructured NbO2 thin film prepared by thermal oxidation. J. Phys. D Appl. Phys. 37, 3392 (2004).
- 42. Crofut, W. A. & Sottile, M. R. Design techniques of a delay-line content-addressed memory. IEEE Trans. Electron. Comput. 15(529-534), 1966. https://doi.org/10.1109/PGEC.1966.264360 (1966).
- 43. Incorporated, W. R. WRspice Reference Manual. Tech. Rep. (2019).
- 44. Maezawa, M. Numerical study of the effect of parasitic inductance on rsfq circuits. IEICE Trans. Electron. 84, 20-28 (2001).
- 45. Likharev, K. & Semenov, V. RSFQ logic/memory family: A new Josephson-junction technology for sub-terahertz-clock-frequency digital systems. IEEE Trans. Appl. Supercond. 1, 3-28. https://doi.org/10.1109/77.80745 (1991).
- Zinoviev, D. Design and Partial Implementation of RSFQ-Based Batcher-Banyan Switch and Support Tools (Lambert Academic 46. Publishing, 1997).
- 47. Volk, J., Tzimpragos, G., Wynn, A., Golden, E. & Sherwood, T. Low-cost superconducting fan-out with cell I C ranking. IEEE Trans. Appl. Supercond. 33, 1-12. https://doi.org/10.1109/TASC.2023.3256797 (2023).
- Tolpygo, S. K. et al. Superconductor electronics fabrication process with MoNx kinetic inductors and self-shunted Josephson 48. junctions. IEEE Trans. Appl. Supercond. 28, 1-12. https://doi.org/10.1109/TASC.2018.2809442 (2018).
- Shaikhaidarov, R. et al. Quantized current steps due to the AC coherent quantum phase-slip effect. Nature 608, 45-49 (2022).
- 50. Astafiev, O. et al. Coherent quantum phase slip. Nature 484, 355-358 (2012).

- Tolpygo, S. K. et al. Deep sub-micron stud-via technology of superconductor VLSI circuits. Supercond. Sci. Technol. 27, 025016. https://doi.org/10.1088/0953-2048/27/2/025016 (2014).
- Tolpygo, S. K. & Semenov, V. K. Increasing integration scale of superconductor electronics beyond one million Josephson junctions. J. Phys. Conf. Ser. 1559, 012002. https://doi.org/10.1088/1742-6596/1559/1/012002 (2020).
- Tolpygo, S. K. et al. Progress toward superconductor electronics fabrication process with planarized nbn and nbn/nb layers. *IEEE Trans. Appl. Supercond.* 33, 1–12. https://doi.org/10.1109/TASC.2023.3246430 (2023).
- 54. Herr, A. et al. Scaling nbtin-based ac-powered Josephson digital to 400 m devices/cm² (2023).
- 55. Radparvar, M. Superconducting niobium and niobium nitride processes for medium-scale integration applications. *Cryogenics* 35, 535-540 (1995).
- 56. Villegirr, J.-C. et al. NbN multilayer technology on R-plane sapphire. IEEE Trans. Appl. Supercond. 11, 68-71 (2001).
- Baggetta, E., Ebert, B., Hadacek, N., Villegier, J.-C. & Maignan, M. New design and implementation of a fast modulator in NbN technology. *IEEE Trans. Appl. Supercond.* 15, 453–456 (2005).
- 58. Villegier, J.-C. *et al.* Extraction of material parameters in NbN multilayer technology for RSFQ circuits. *Physica C* **326**, 133–143 (1999).
- Volk, J., Tzimpragos, G., Wynn, A., Golden, E. & Sherwood, T. Low-cost superconducting fan-out with cell I_C ranking. *IEEE Trans. Appl. Supercond*. https://doi.org/10.1109/TASC.2023.3256797 (2023).

Acknowledgements

The authors would like to thank D. Scott Holmes for his helpful comments and valuable discussions.

Author contributions

J.V. and G.T. conceived the idea, designed the proposed memory system, and performed the timing analysis. J.V. designed, simulated, and analyzed the corresponding circuits. J.V., A.W., E.V., and G.T. contributed to the memory density analysis and the writing of the manuscript. G.T. and T.S. oversaw the project. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.V. or G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023