

OPEN

Cryogenic Memory Architecture Integrating Spin Hall Effect based Magnetic Memory and Superconductive Cryotron Devices

Minh-Hai Nguyen^{1*}, Guilhem J. Ribeill¹, Martin V. Gustafsson¹, Shengjie Shi², Sriharsha V. Aradhya^{2,5}, Andrew P. Wagner¹, Leonardo M. Ranzani¹, Lijun Zhu², Reza Baghdadi³, Brenden Butters³, Emily Toomey³, Marco Colangelo³, Patrick A. Truitt^{4,6}, Amir Jafari-Salim^{4,6}, David McAllister^{4,7}, Daniel Yohannes^{4,6}, Sean R. Cheng^{1,8}, Rich Lazarus¹, Oleg Mukhanov^{4,6}, Karl K. Berggren³, Robert A. Buhrman², Graham E. Rowlands¹ & Thomas A. Ohki¹

One of the most challenging obstacles to realizing exascale computing is minimizing the energy consumption of L2 cache, main memory, and interconnects to that memory. For promising cryogenic computing schemes utilizing Josephson junction superconducting logic, this obstacle is exacerbated by the cryogenic system requirements that expose the technology's lack of high-density, high-speed and power-efficient memory. Here we demonstrate an array of cryogenic memory cells consisting of a non-volatile three-terminal magnetic tunnel junction element driven by the spin Hall effect, combined with a superconducting heater-cryotron bit-select element. The write energy of these memory elements is roughly 8 pJ with a bit-select element, designed to achieve a minimum overhead power consumption of about 30%. Individual magnetic memory cells measured at 4 K show reliable switching with write error rates below 10^{-6} , and a 4×4 array can be fully addressed with bit select error rates of 10^{-6} . This demonstration is a first step towards a full cryogenic memory architecture targeting energy and performance specifications appropriate for applications in superconducting high performance and quantum computing control systems, which require significant memory resources operating at 4 K.

Since the invention of the integrated circuit, the exponential increase in computing predicted by Moore's law has made high performance computing (HPC) an essential driver for technological and scientific advances. With exascale computers on the horizon, it has become increasingly important to minimize the energy consumption of computation to meet the needs of HPC applications in a reasonable power envelope. However, the end of Dennard scaling and the difficulty of scaling down conventional CMOS technology made it important to look to alternative technologies that can continue to advance the performance of HPC systems. While many new computing paradigms have recently been proposed, such as in-memory and neuromorphic^{1,2} computing, it is still critical to improve the computational efficiency of conventional architectures.

Superconducting computers have long been a strong candidate for such “more than Moore” computing with the potential of two orders of magnitude lower power consumption than CMOS technology, even including the cost of cryogenic cooling power³. This large advantage is due to near-lossless signal propagation along superconducting wires, and the picosecond delay and atto-Joule switching energy of Josephson junctions (JJs). In recent years, Energy-efficient Rapid Single Flux Quantum⁴ and Reciprocal Quantum Logic circuits⁵ have demonstrated operation with zero static power dissipation. Practical implementations of superconducting logic are expected to operate at clock frequencies of up to 80 GHz^{5,6}, potentially an order of magnitude faster than their CMOS

¹Raytheon BBN Technologies, Cambridge, MA, 02138, USA. ²Cornell University, Ithaca, NY, 14850, USA.

³Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. ⁴HYPRES, Inc., Elmsford, NY, 10523, USA.

⁵Present address: Western Digital Corp., Fremont, CA, 94539, USA. ⁶Present address: SeeQC Inc., Elmsford, NY, 10523, USA. ⁷Present address: Bluemont Technology, Luray, VA, 22835, USA. ⁸Present address: Harvard University, Cambridge, MA, 02138, USA. *email: minh-hai.nguyen@raytheon.com

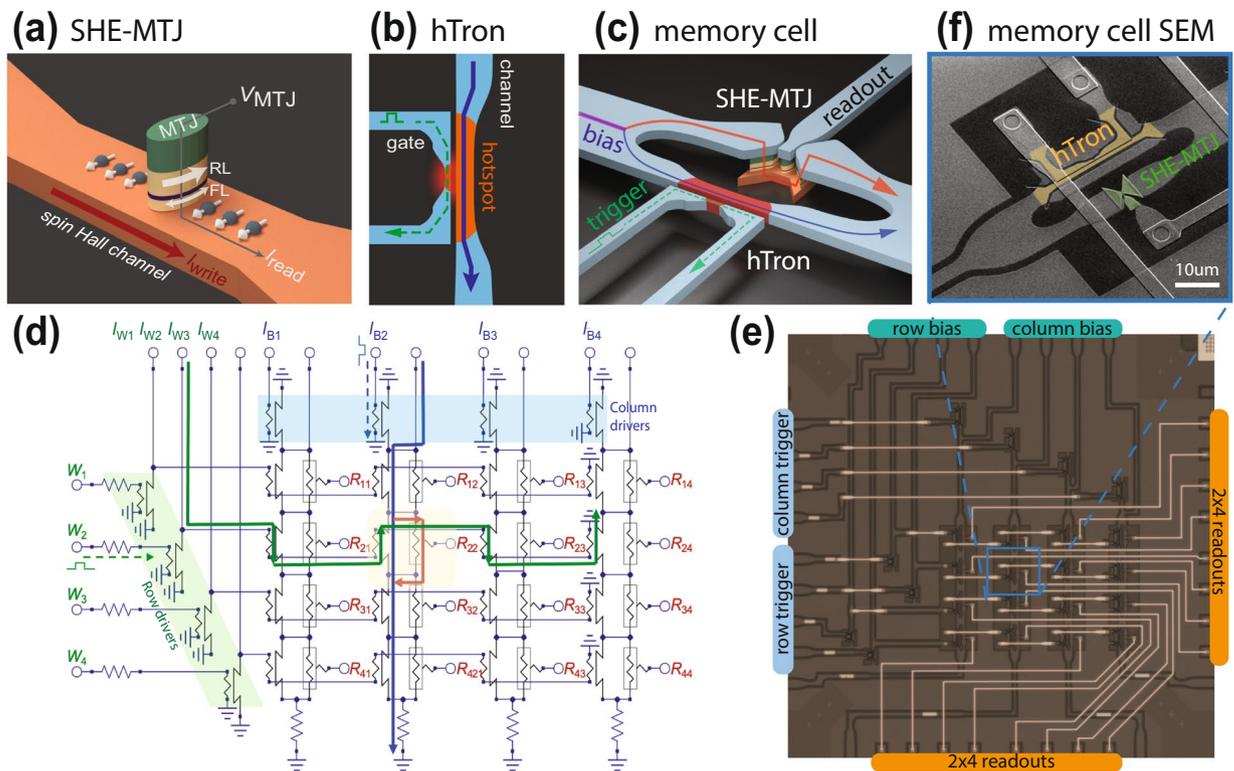


Figure 1. Proposed cryogenic memory architecture. Not-to-scale illustrations of (a) a SHE-MTJ device, (b) a hTron device and (c) a memory cell. (d) Schematic of the 4×4 memory array interfaced with external control circuitries by column and row drivers. The arrows indicate the current paths during the writing process of cell (2,2). (e) Micrograph of the memory array. (f) SEM image of cell (2,2) in the array.

counterparts. Furthermore, the energy efficiency of superconducting electronics makes them an attractive technology for use as an interface to quantum computing systems that operate at millikelvin temperatures.

While superconducting logic has been extensively developed and demonstrated experimentally, fast and power-efficient cryogenic memory has been one of the main obstacles in realizing fully functional superconducting computers^{3,7}. Memories based on single-flux-quantum (SFQ) technology⁸ have large cell sizes and power dissipation³. CMOS-based memories^{9–11} provide better scalability but also suffer from their high dynamic power consumption at cryogenic temperatures, in the range of 10–100 pJ/bit. Although there is an active pursuit of other types of superconducting memories^{12–15}, a great deal of effort has been invested in magnetic^{16–18} and magnetic-superconducting hybrid memories^{19,20} which have the potential to offer high speed, low power, non-volatility, and scalability. Early achievements in cryogenic spin-valves^{16,17} and magnetic tunnel junctions¹⁸ (MTJs) have recently been reported with energy consumption per switching of a few tens of fJ. Integrating these successfully demonstrated magnetic elements with superconducting circuitry becomes an appealing route in the pursuit of larger scale cryogenic memory.

We report on the design and experimental characterization of a prototype cryogenic memory array comprising two key technologies: magnetic tunnel junctions based on the spin-Hall effect (SHE-MTJs)²¹ as storage elements, and superconductive heater-cryotrons (hTrons)¹² as control elements, as illustrated in Fig. 1(a–c). The SHE-MTJs are three-terminal devices consisting of an MTJ atop a metallic spin Hall channel. Nanosecond magnetic reversal time and write error rates $< 10^{-5}$ for writing current densities as low as 8×10^{11} A/m² at room temperature (RT) have been reported for this class of device²². Our recent work confirmed similar performance at 4 K¹⁸ providing a high-speed and low-power solution for cryogenic memory. hTrons are a variant of the nano-cryotron (nTron)²³ logic family, which can be driven by SFQ logic and support sufficient current to switch the magnetic element. In this work, we successfully integrate the two components (SHE-MTJs and hTrons) into a 4×4 cryogenic memory array in which the SHE-MTJ memory elements are addressed via hTron bit-select, row and column drivers.

Architecture Description

The cryogenic memory system we developed includes a memory array and superconducting control circuits. The latter are currently fabricated on separate chips which can be connected to the memory array by wire-bonding or Multi-Chip Module (MCM) technology. The control circuits, designed with SFQ technologies, can supply triggering signals of ~ 100 μ A into a load of a few Ohms. Here we limit our discussion to the design and performance of the main memory array which will target compatibility with the specifications of these control chips.

Memory array architecture. As illustrated in Fig. 1(a–c), our proposed memory cell consists of a SHE-MTJ device in parallel with a bit-select hTron device. The addressing architecture was based on a design described in ref. ²⁴. The SHE-MTJ, as depicted in Fig. 1(a), is a three-terminal device consisting of an MTJ patterned on top of a spin Hall channel. The MTJ itself is an elliptical nanopillar consisting of two ferromagnetic layers separated by a thin tunnel barrier: the reference layer's (RL) magnetization is rigidly pinned, while the free layer's (FL) magnetization can be switched parallel (P) or antiparallel (AP) to that of RL. One bit of information is encoded in this non-volatile relative orientation: the tunneling magnetoresistance (TMR) effect gives rise to two easily distinguishable resistance states (R_P or R_{AP}) in either case. In our measurements, memory readout is performed by monitoring the MTJ voltage V_{MTJ} while applying a small $5 \mu\text{A}$ sense current through the MTJ.

To switch the magnetization of the FL, i.e., to change the MTJ state, a current is applied in the spin Hall channel. The SHE^{25–27} of the channel induces spin accumulations on the surfaces of the channel, in turn enacting magnetization reversal of the FL via the spin (transfer) torque^{21,28,29} as spins diffuse into the ferromagnet. The low switching current and confinement of the spin-torque switching mechanism within the device (no external magnetic fields, with their detrimental impact on superconductive circuits, are required) make the SHE-MTJ suitable for integration with superconducting structures.

An issue with the SHE-MTJ, however, is that its characteristic impedance and switching currents are too large to be directly compatible with our separately fabricated SFQ control circuits. For example, a 300 nm wide, 5 nm thick spin Hall channel requires a switching current of roughly 1 mA into a 0.5 k Ω load, which is incompatible with typical SFQ circuit output impedance of a few Ohms.

To resolve this problem, we use a heater-cryotron (hTron) bit-select element connected in parallel with the SHE-MTJ channel which acts as a three terminal switch and also can be engineered to have better impedance match to the SHE-MTJ. The hTron¹² is a non-contact variation of the nano-cryotron²³ in which heat generated in a gate nanowire to suppress the critical current of an adjacent superconducting channel below its nominal operating current, thereby switching it into the normal state. Our implementation uses a nanowire gate constriction made from the same material and placed in the same plane and very near to the main channel, as illustrated in Fig. 1(b). A small triggering pulse (green arrow) is applied to the gate constriction, sending it into the normal state and creating a hot spot because of dissipative current flow. The lateral heat flowing away from this hotspot suppresses the critical current of the nearby channel and causes the channel to enter the normal state if it is biased (blue arrow). The resistance of the channel in the normal state can be engineered to reach multiple kilo-Ohms while carrying in excess of 1 mA. Our primitive memory cell is formed by placing the hTron in parallel with a SHE-MTJ channel, as depicted in Fig. 1(c). Triggering the hTron presents a large enough impedance that current redirects through the SHE channel and switches the MTJ's state. In this manner, the memory element can be controlled by a small hTron gate pulse (100 μA) sourced by a line-driver hTron which can be interfaced directly with the SFQ peripheral circuitry. The use of an hTron or an equivalently high isolation device is particularly important in the interior of the memory arrays: sneak currents or leakage current from writing to a cell can otherwise cause redistribution of current within the surrounding circuitry, since in the inactive state all lines are superconducting. This can create multiple current loops for leakage current to accumulate. This effect is mitigated by inclusion of a bit select element that has high isolation. In addition, when an hTron is inactive, current simply flows through the superconducting path in the memory cell to subsequent cells in the columns and is recycled elsewhere with no concerns of other leakage paths.

Scaling to a device array is accomplished by chaining these memory cells into many parallel columns. The schematic of a 4×4 memory array is shown in Fig. 1(d). To be able to address the individual memory cells, the array is connected to peripheral row and column drivers, which are also hTron devices, to address the individual memory cells. The writing procedure is as follows: All row and column driver hTrons are biased with currents, which are initially shunted to ground through the driver channels. To write to a cell, a triggering pulse is applied to the corresponding column driver hTron (dashed blue arrow applied at I_{B2}). The column driver becomes normal and consequently the bias current is diverted down the column of cells and flows through all bit-select hTrons in the column (solid blue arrow), which are still superconducting. Any combination of columns can be activated, with each column constituting a bit in the 4-bit word to be written. Next, one row driver hTron is triggered (dashed green arrow), diverting its bias current through the gates of all bit-select hTrons in that row (solid green arrow). For those bit-select hTrons which have both the current from the column driver flowing through their channel and receive a gate current from the row driver, the triggering condition is fulfilled, diverting sufficient current to the SHE-MTJ channel to produce a change in magnetic state (orange arrow).

The micrograph of the 4×4 memory array is shown in Fig. 1(e). A scanning electron microscopy (SEM) image of cell (2,2) in the array is shown in Fig. 1(f). See Methods for detail on sample multilayer and fabrication process.

Optimizing for power efficiency. Because current is shared between the channels of the SHE-MTJ and hTron bit-select elements in the memory cell during writing events, it is important to match their impedances to minimize power consumption. For fixed SHE-MTJ parameters, the problem becomes determining an hTron design which minimizes dissipation while still providing enough current to switch the MTJ. This means that the normal-state resistance of the hTron should be maximized, so that most of the current will be diverted into the memory cell, with the following two constraints:

- (i) The column bias current must be lower than the hTron critical current of the hTron channel, to ensure that the bit-select hTron channel is superconducting before its gate is triggered.
- (ii) The current through the hTron channel during the writing event must be large enough to sustain the hot-spot in the channel³⁰.

Solving straightforward inequalities (see Methods), the optimal width and length of the hTron are found to be:

$$W_0 = \frac{I_{\text{SH}}}{J^c d - \sqrt{2\alpha\Delta T/R_{\text{sq}}}}, \quad (1)$$

$$L_0 = I_{\text{SH}} \cdot R_{\text{SH}} / \sqrt{2\alpha\Delta T R_{\text{sq}}}, \quad (2)$$

where I_{SH} is the SHE-MTJ switching current, R_{SH} its channel resistance which is roughly 600Ω for our devices; J^c , d , R_{sq} , α are hTron channel's critical current density, thickness, sheet resistance, and specific heat per unit area, respectively; and $\Delta T = T_c - T_s$ is the difference between the hTron critical temperature T_c and the sample temperature T_s .

In this optimal design, the overhead power consumption added by the hTron is then (see Methods):

$$P_{\text{overhead}} = \frac{\sqrt{2\alpha\Delta T/R_{\text{sq}}}}{J^c d - \sqrt{2\alpha\Delta T/R_{\text{sq}}}}, \quad (3)$$

which is fixed by the hTron's material properties and system temperature, independent of SHE-MTJ parameters. This result implies that we can always design the bit-select hTron to accommodate the impedance of a particular SHE-MTJ device with a fixed overhead power. Using typical values for our NbN film (not fully optimized for this application), $\alpha = 220 \text{ kW/m}^2\text{K}$, $R_{\text{sq}} = 100\text{--}140 \Omega/\text{sq}$, $T_c = 12.8 \text{ K}$, $T_s = 3.6 \text{ K}$, $J_c = 25\text{--}30 \text{ GA/m}^2$, the minimal energy overhead P_{overhead} is 30(6)%. Estimated using the results from nanosecond pulse switching of a standalone SHE-MTJ device, described in the next section, the typical switching energy for a SHE-MTJ device is roughly 6 pJ and thus that of a memory cell is about 8 pJ per switching.

Measurements and Results

Standalone SHE-MTJ device. As described above, our memory element, the spin-Hall-effect based magnetic tunnel junction (SHE-MTJ), consists of an MTJ atop a spin Hall channel, as illustrated in Fig. 1(a). A writing (electrical) current in the channel induces a transverse spin accumulation that exerts torque on the free layer (FL) via the SHE, thereby switching its magnetization back and forth. The conventional figure of merit representing the strength of the SHE is the spin Hall angle, which is the ratio of the induced spin current to the applied electrical current. Energy-efficient switching devices require metallic materials possessing very strong SHE such as Pt³¹, β -Ta²¹, β -W³² and their alloys^{33–35} whose (absolute) spin Hall angles are in the range of 0.1–0.35.

SHE-MTJ devices in our memory arrays employ a Pt₈₅Hf₁₅ alloy channel with spin Hall angle ~ 0.2 . This was an early design decision made with fabrication stability in mind, and without the benefit of subsequent materials advances. As seen in the SHE-MTJ stack illustrated in Fig. 2(a), a thin Hf layer of 0.5 nm thick is inserted between the channel and the MTJ to suppress the Gilbert magnetic damping caused by spin pumping³⁶. Further device details are given in the Methods section.

We find robust SHE-MTJs response seen in other devices of similar composition³⁷. The tunneling magnetoresistance (TMR) response to an easy-axis magnetic field is shown in Fig. 2(b). The MTJ switches between the parallel (P, low resistance) and antiparallel (AP, high resistance) states with a 20 mT coercivity (loop half-width) and a small offset field of -5 mT . Bi-stability at zero external field is thus achieved: a mandatory criterion given the flux-trapping propensity of the superconducting peripheral circuitry. Another criterion is satisfied in that the AP and P resistance states are easily discernible using a simple superconducting comparator.

The nanosecond pulse switching performance of the SHE-MTJ device is gauged using the same measurement setup and technique outlined in our previous reports^{18,37}. The switching probability is shown in Fig. 2(c) for the two switching polarities as a function of pulse duration and amplitude. We fit the pulse amplitudes $I_{50\%}$ and pulse durations $t_{50\%}$ along the 50% probability boundary (blue dots) with the macrospin model for ballistic reversal (blue curves)^{38,39} to obtain the critical current I_0 and characteristic reversal time t_0 . For P to AP switching we find $I_0 = 0.9 \text{ mA}$ and $t_0 = 0.7 \text{ ns}$ while for AP to P switching we find $I_0 = 0.9 \text{ mA}$ and $t_0 = 1.5 \text{ ns}$. These t_0 values are similar to previous measurements (at 4 K) on Pt channel devices¹⁸, while the I_0 values are about two times lower, which is consistent with the two times higher spin Hall angle of Pt₈₅Hf₁₅ when compared to pure Pt³⁷. With the performance of the memory element itself confirmed, we verify the ability of our hTron select element to redirect its channel current across the SHE-MTJ to actuate switching.

Standalone hTron device. The heater cryotron (hTron) devices shown in Fig. 1(b) are arranged in parallel with a SHE-MTJ devices as shown in Fig. 1(c). A current pulse of $100 \mu\text{A}$ is supplied the hTron gate, ultimately sourced by a SFQ-DC or SFQ-RO (relaxation oscillator) driver, sending the gate into the normal state and producing sufficient Ohmic heating to trigger hotspot formation in the hTron's current carrying channel situated 100 nm away. The supercurrent is impeded by the emergence of a $\sim 1 \text{ k}\Omega$ dissipating impedance of the normal state NbN, and a large fraction of the current is redirected into the SHE-MTJ channel where it switches the memory element. To minimize the overhead power consumption of the bit-select hTron (see analysis in the previous section), the hTron channel was designed to be $1.7 \mu\text{m} \times 25.5 \mu\text{m}$, which gives the optimal overhead energy consumption about 31%.

Before studying integrated devices, we performed quasi-static switching measurements on one standalone hTron device. As shown in Fig. 3(a), we biased the hTron channel with a square pulse (blue) and the gate with a triangular pulse (red), while monitoring their respective voltages (right axis). The voltages remain at the noise floor until the gate current reaches the critical value of $87 \mu\text{A}$, beyond which the gate becomes normal (resistive). If the channel current is high enough, as is the case in Fig. 3(a), the channel also becomes normal. Beyond t

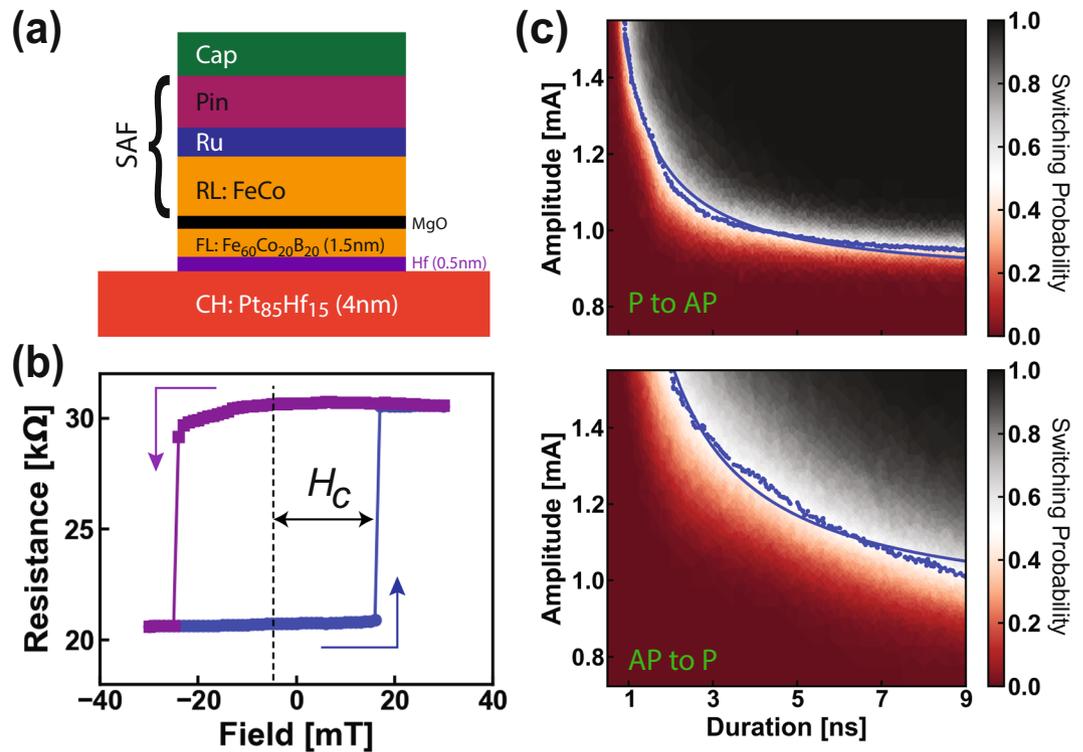


Figure 2. Cryogenic characterization of a standalone SHE-MTJ device. (a) SHE-MTJ multilayer consisting of a spin Hall channel, Hf spacer and MTJ stack pinned by a synthetic antiferromagnetic (SAF) structure. (b) Magnetic minor loop of the MTJ with a $3\ \mu\text{A}$ sense current. The dotted line shows the center of the loop (stray field). (c) Switching probability (encoded by the color) as a function of pulse amplitude and duration for P to AP (top) and AP to P (bottom). Blue dots are points at 50% switching probability which are fitted by the macrospin model (blue curves).

= 0.3 ms, the resistive state of the channel is maintained by the channel current (below its critical value) even though the gate current has been turned off. This important latching behavior is a result of thermal run-away in the channel as its resistance increases and allows the triggering pulse to the gate to be short and of low amplitude.

We repeated the measurement in Fig. 3(a) 100 times and measured the distribution of current required to turn the gate resistive, shown in Fig. 3(b) along with the cumulative switching probability. We acquired repeated histograms of this type, for a range of channel currents, and the resulting switching probability is shown in the left panel of Fig. 3(c). The clear and orthogonal switching boundaries shows that the gate and channel of the hTron are galvanically isolated, as expected.

We next performed a similar experiment but this time monitored the switching (to normal state) of the hTron channel, and obtained the results shown in the right panel of Fig. 3(c). Comparing the two panels of Fig. 3(c) reveals an interesting correlation: For $|I_{\text{gate}}| > 90\ \mu\text{A}$, the gate always turns normal, but the channel state does not necessarily follow. Instead, the boundaries around $I_{\text{channel}} = 0.2\ \text{mA}$ and $|I_{\text{gate}}| = 90\ \mu\text{A}$ are curved. This difference from the orthogonal boundaries seen in the left panel is due to the relative sizes of the hotspots created in the channel and the gate.

Memory cell. To perform switching experiments on a memory cell — as shown in Fig. 1(c) — we used external pulse generators to provide bias and triggering current pulses via reference resistances of 5–10 kΩ, as shown in Fig. 4(a). To measure the MTJ resistance, we applied a DC bias current of $5\ \mu\text{A}$ into the top of the MTJ and monitored its voltage drop.

We performed microsecond pulse switching measurement of the representative memory cell (2,2) in the array at 4 K without an external magnetic field (see Fig. 1(d)). In these experiments, we biased the 2nd row driver hTron with a 1 mA pulse and applied a trigger pulse of $100\ \mu\text{A}$ to the gate. The gate trigger level of the 2nd column driver hTron was also $100\ \mu\text{A}$, but its channel bias was varied between -1.5 and $1.5\ \text{mA}$. Figure 4(b) shows the traces of the input and sample voltages read from an oscilloscope (for column bias = $-1.5\ \text{mA}$). The MTJ state is determined by the sign of ΔV_{MTJ} which is the MTJ drop voltage subtracted by its median value of 53 mV.

As shown in Fig. 4(b), for time $t < 2\ \mu\text{s}$, the column bias current flows through the superconductive hTron channels of the 2nd column and a small sensing resistance at the bottom, resulting in a small voltage drop. At $t = 2\ \mu\text{s}$, the row driver is triggered, and after a $2.5\ \mu\text{s}$ delay (due to the length of the coaxial cables of a few meters from the chip package in the cryostat to the RT electronics) at $t = 4.5\ \mu\text{s}$ the row driver channel builds up enough current to trigger the bit-select hTron, resulting in a rapid rise in the column voltage. This result confirms the intended operation of the cell and driver hTrons.

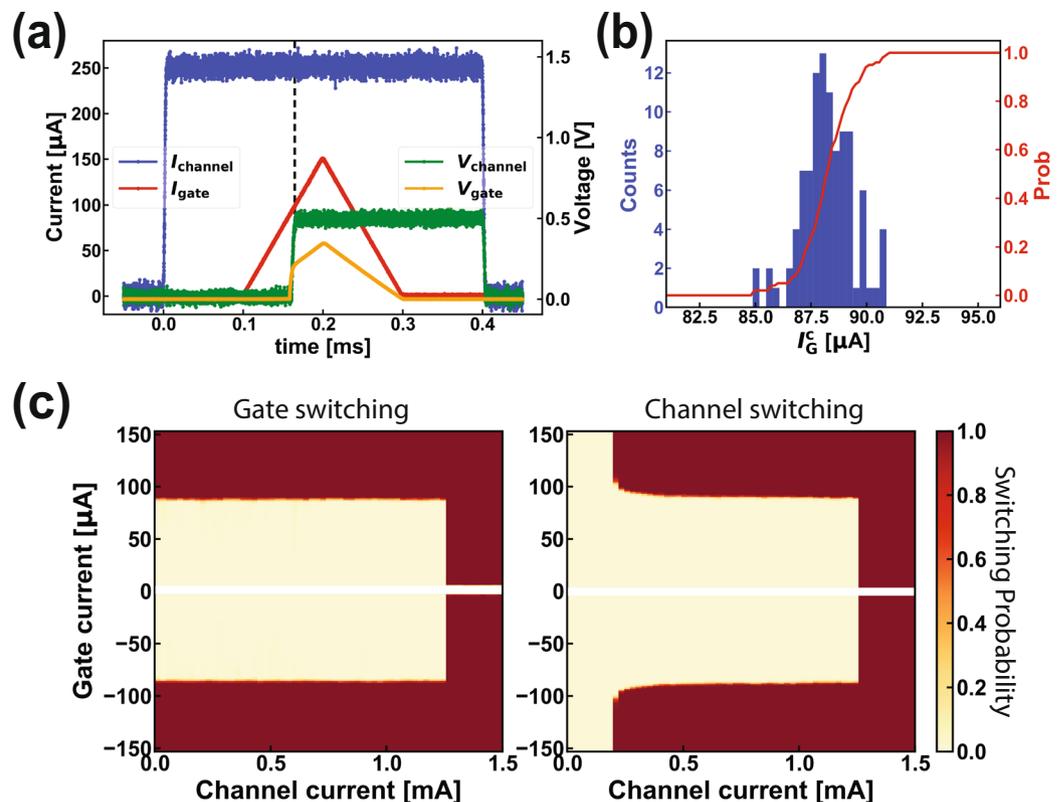


Figure 3. Characterisation of a standalone hTron device. (a) Voltage drops (right axis) over the hTron gate and channel as functions of applied currents (left axis). Switching occurs at $I_G^c = 87 \mu\text{A}$. (b) Distribution of the gate critical current obtained by repeating the bias sequence in panel (a) 100 times, and its cumulative probability of switching to the normal state (right axis). (c) Probabilities for the gate (left) and channel (right) of switching to the normal state as functions of applied gate and channel currents.

The response of ΔV_{MTJ} to the column bias is shown in Fig. 4(c). When the $6 \mu\text{s}$ long column bias current (orange line) is negative, the triggering of the bit-select hTron causes the MTJ voltage to relax to the AP state ($\Delta V_{\text{MTJ}} > 0$). Conversely, when the column bias current is positive, the MTJ voltage relaxes to its P state ($\Delta V_{\text{MTJ}} < 0$). This clearly demonstrates the successful control of the memory cell (2,2) in the array. We note that the long rise and fall times observed in Fig. 4(b,c) result from the use of slow external electronics, and are not characteristic of the underlying devices.

We determined the write-error-rate (WER) of the cell by repeating the switching experiment 10^3 – 10^6 times, with each attempt preceded by resetting the MTJ to its original state by applying a high current bias to the column. The resulting behavior, shown in Fig. 5(a), is very similar in shape to the previously reported WER of single standalone SHE-MTJs^{18,22}. Note that the open symbols near the bottom of the plot indicate no write error was observed in 1 million switching attempts, implying that the WER is below 10^{-6} for bias current of 1.0 mA (P to AP) and 1.5 mA (AP to P).

We characterized the intercell crosstalk by monitoring the MTJ voltage of cell (2,2) while triggering its four neighboring cells. Figure 5(b) shows the response in ΔV_{MTJ} of cell (2,2) when the adjacent cell (2,1) was triggered. A voltage rise of about 3 mV is observed during triggering events, which translates to roughly $10 \mu\text{A}$ of leakage current in the SHE-MTJ channel of cell (2,2). This leakage current is too small to affect the state of the cell (2,2), which remains unchanged in the AP state. We repeated the experiment with all permutations of cell (2,2) original state, bias current polarity, switching each of the four neighboring cells, with 10^4 switching attempts for each case. No unintended switching of cell (2,2) was observed.

Array triggering fidelity. We then investigated the triggering fidelity (successful rate) of the 4×4 memory array by writing all possible 4-bit words to each row and detecting whether the targeted bit-select hTrons triggered. As illustrated in Fig. 6(a), the measurement procedure is as follows: The column and row driver hTron channels are biased with long current pulses. For each bit in the word to be written, the corresponding column driver is activated with a $1 \mu\text{s}$ long pulse on its gate. The voltage at the top of each column, V_{pre} , is measured, to make sure that none of the bit-select hTrons have been accidentally activated before the row driver. One of the four row drivers is then activated with a $1 \mu\text{s}$ long current pulse, and the column voltages V_{1-4} are measured again. The write is considered successful if the following criteria are fulfilled:

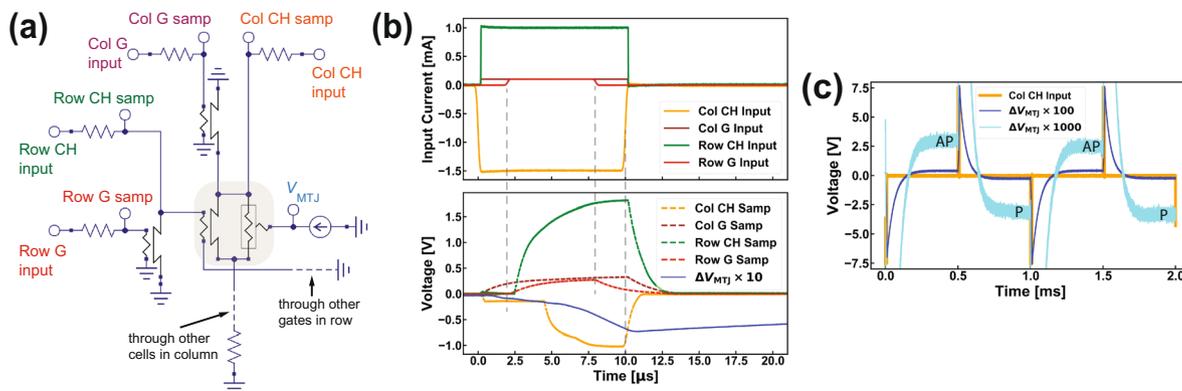


Figure 4. Switching of memory cell (2,2). (a) Schematic of the measurement circuitry on a representative cell (1,1) in the array (other cells not shown). (b) Input current (top) and sample voltage (bottom) traces of column and row channel (CH) and gate (G). Blue trace (bottom) shows the variation of the MTJ voltage ΔV_{MTJ} . (c) Column bias and enlarged ΔV_{MTJ} show the MTJ state's responding to the sign of the column bias.

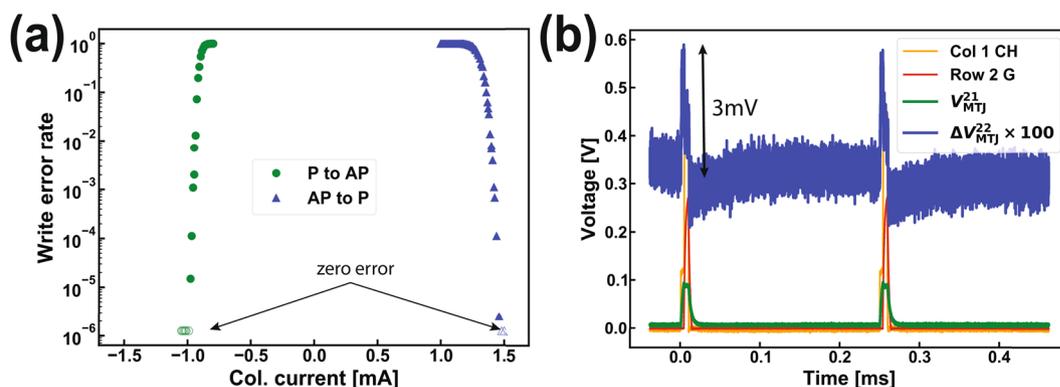


Figure 5. Write error rate and crosstalk of memory cell (2,2). (a) Memory write error rate (log scale) as a function of column bias current. The open symbols indicate zero error observed in 10^6 switching attempts. (b) Enlarged ΔV_{MTJ} of cell (2,2) when triggering cell (2,1) shows a leakage voltage of 3 mV.

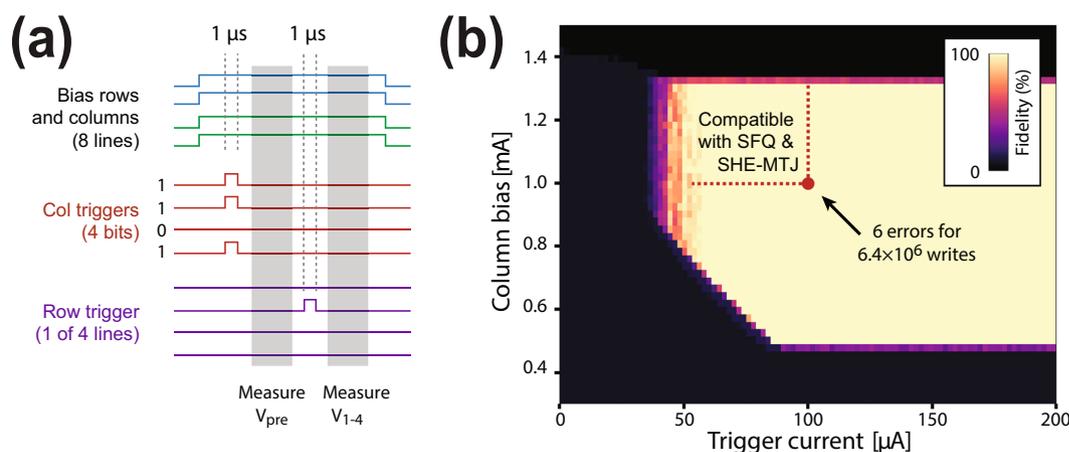


Figure 6. Fidelity of the array triggering. (a) Illustration of a pulse sequence of the fidelity measurements. The column voltages are measured before and after row triggering. (b) Triggering fidelity versus column bias current and trigger current with 640 writing sequences for each pixel. The red dotted lines indicate the constraints set by the memory element and SFQ control circuits.

- (i) All column voltages fall below a threshold before the row is triggered: $V_{\text{pre}} < V_{\text{min}}$.
- (ii) All column voltages V_{1-4} for unwritten bits fall below a threshold value after the row is triggered: $V_{1-4} \text{ (unwritten)} < V_{\text{min}}$.
- (iii) All column voltages V_{1-4} for written bits fall within a specified interval after the row is triggered: $V_{\text{min}} < V_{1-4} \text{ (written)} < V_{\text{max}}$.

Figure 6(b) shows the measured triggering fidelity with respect to column bias current and the trigger current used to activate both row and column drivers. The fidelity is assessed from 640 written words per pixel, encompassing all 16 possible bit patterns, written ten times to each row. We found that the fidelity is nearly 100% for a wide range of parameter values. Here we used $V_{\text{min}} = 400$ mV, $V_{\text{max}} = 1$ V, and a row bias current of 0.8 mA, but found that the measured fidelity did not depend critically on any of these values.

As previously determined, a column current of ~ 1 mA suffices to reliably change the state of a SHE-MTJ element, and SFQ circuitry is able to provide trigger pulses of ~ 100 μ A. For this critical combination of parameters, highlighted with a red spot in Fig. 6(b), we benchmarked the writing fidelity in detail with 6.4 million writing cycles (distributed over all rows and bit patterns as above) and yielded a total of 6 errors. The bright area confined by the dotted line in Fig. 6(b) indicates the parameter space suitable for the operation of the SHE-MTJ elements and SFQ control pulses.

Discussion

We have experimentally demonstrated the successful integration of SHE-based MRAM technology and the superconductive hTron bit-selects and drivers in a 4×4 memory array architecture which can be triggered by current signals as low as 100 μ A, compatible with SFQ DC/RO control circuits. The array is fully addressed with an error rate as low as 1×10^{-6} , while the write-error-rate of the individual memory cells is shown to be below 10^{-6} . The bit-select hTron was designed to match the impedance of the SHE-MTJ channel, adding a minimal energy overhead of 31% to the switching energy which is currently in the order of 6 pJ per switch.

Beyond the proof-of-concept demonstration of the cryogenic memory array architecture presented in this Report, we anticipate to make significant improvements to array speed and energy consumption. Because the minimum energy overhead added by the bit-select hTron is independent of SHE-MTJ parameters, the overall energy consumption of the memory cell scales with that of the SHE-MTJ element which can be further decreased using newly discovered spin Hall materials^{35,40}, device structures⁴¹ and better lithography technologies. For example, the SHE-MTJ channel can be replaced with materials such as $\text{Au}_{1-x}\text{Pt}_x$ alloys whose spin Hall angles approach 0.35³⁵ for a four-fold energy reduction. The SHE-MTJ channel resistance, currently about 0.6 k Ω , whose major contribution comes from the two vias providing electrical contact to the SHE channel (see Fig. 1(c)) can be reduced 2-3 times by replacing them with metallic vias. Another obvious improvement in design is scaling down the SHE-MTJ channel to reduce the switching current. Combining all these modifications in materials and design, we estimate the cell's switching energies could be as low as 0.1 pJ, more than 10 times less than that of cryogenic DRAM¹⁰.

Besides SHE-MTJ, the same memory architecture can be applied for other types of MRAM, including the 2-terminal spin-transfer-torque (STT) MRAM or cryogenic spinvalves^{16,17}. Albeit more complexity in fabrication, the 3-terminal structure of SHE-MTJs provides several advantages over STT-MRAMs: the read and write paths are separate thus can be optimized independently; read-disturb-error (accidental writing when reading) is negligible (has not been observed so far); superior durability (no known wear-out mechanism). On the other hand, in terms of scalability and energy efficiency, spinvalves can be the best option due to their all-metallic structure hence very low impedance. However, their typically small magnetoresistance variation of less than 1% causes difficulty for reliable read-out. We anticipate that this problem may soon be resolved with future material development, such as the application of Heusler compounds⁴².

There are also opportunities to improve the performance of the control elements. The hTron column and row drivers used in our prototype array, whose triggering delay is roughly 8 ns¹² due to the heat propagation across the 100 nm gap in the film plane, can be replaced with nTrons²³ with vanishing triggering delay owing to the galvanic connectivity of their gate and channel. Our hTron bit-selects can also be replaced with a stacked variant¹⁵ in which the gate crosses over the channel thereby providing more efficient heating and smaller triggering current. These enhancements will result in a faster, more energy efficient, and more compact bit-select and driver elements.

By combining the high speed and non-volatility of the MRAM, the superconductivity of the select elements, and the compatibility with SFQ circuitry and fabrication, our cryogenic memory architecture is proven to be a fast and power efficient candidate for exascale superconducting computing and quantum control systems.

Methods

Samples. As illustrated in Fig. 2(a), the SHE-MTJ stack is as follows (thicknesses in nanometer): *substrate*/Pt₈₅Hf₁₅(4)/Hf(0.5)/FL/MgO/RL/Ru/PL/cap, where FL = 1.5 nm Fe₆₀Co₂₀B₂₀, RL is the magnetic fixed layer pinned by the pinning layer PL in the synthetic antiferromagnetic structure RL/Ru/PL which is tuned to minimize the stray dipole field at the FL so that the magnetization of the FL is bistable at zero bias field. The stack was patterned into SHE-MTJ devices with 300 nm wide channel and 75 nm \times 190 nm nanopillar MTJ by deep-UV and e-beam lithography and Argon ion milling.

Upon the fabrication of SHE-MTJ devices, a NbN layer of 30 nm thick was deposited at ambient temperature and patterned by photo- and electron-beam lithography. Superconducting structures are designed using conformal mapping curves⁴³ to prevent current crowding effect at corners. The bit-select hTron channel size was 1.7 μ m \times 25.5 μ m, which was optimized to achieve the minimal energy overhead of 31%. A second level of metallic interconnect was provided by e-beam evaporated aluminum wires and vias and PECVD SiN dielectric. Micrographs of the array and a representative cell are shown in Fig. 1(e,f).

Impedance matching. To minimize the overhead power consumption added to the SHE-MTJ switching by the bit-select hTron, the normal-state resistance of the hTron should be maximized, within the following two constraints:

- (i) The column bias current must be lower than the hTron critical current of the hTron channel: $I_{bias} < J^c Wd$ where J^c , W , d are the critical current density, width and thickness of the hTron channel.
- (ii) The current through the hTron channel during the writing event must be large enough to maintain the hotspot in the channel²⁵: $J^2 d^2 R_{sq} > 2\alpha\Delta T$ where J is the current density in the hTron channel during writing, R_{sq} and α are its sheet resistance and specific heat per unit area in normal state, and $\Delta T = T_c - T_s$ is the difference between the hTron critical temperature T_c and the sample temperature T_s .

Combining the above two conditions and the relation where I_{SH} is the SHE-MTJ switching current, we obtain:

$$J^c Wd > \sqrt{2\alpha\Delta T/R_{sq}} \cdot W + I_{SH}. \quad (M1)$$

Solving for W , we get

$$W > \frac{I_{SH}}{J^c d - \sqrt{2\alpha\Delta T/R_{sq}}} \quad (M2)$$

which implies Eq. (1) in the main text. Upon determining the minimum width, to find the optimal hTron channel length L , we use the condition (ii) and notice that $JWd = I_{SH}R_{SH}/(R_{sq}L/W)$ where R_{SH} is the SHE-MTJ channel resistance, then obtain

$$L < I_{SH} \cdot R_{SH} / \sqrt{2\alpha\Delta TR_{sq}} \quad (M3)$$

which gives Eq. (2). The overhead power consumption added by the hTron is then (using (ii) and (M2)):

$$P_{\text{overhead}} = \frac{I_{\text{bias}}}{I_{SH}} - 1 = \frac{JWd}{I_{SH}} > \frac{W\sqrt{2\alpha\Delta T/R_{sq}}}{W(J^c d - \sqrt{2\alpha\Delta T/R_{sq}})} \quad (M4)$$

which yields Eq. (3), independent of SHE-MTJ parameters.

Measurements. All measurements presented in this Report were performed at the base temperature of 3.6 K in cryogen-free cryostats. Characterizations of stand-alone SHE-MTJ devices (shown in Fig. 2) were done using the same HPD cryostat and control electronics as described in our previous report¹⁸. Measurements and data collection procedures were performed using our custom-built Python package⁷, Auspex, available online at <https://github.com/BBN-Q/Auspex>.

Received: 1 July 2019; Accepted: 8 November 2019;

Published online: 14 January 2020

References

1. Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668 (2014).
2. Davies, M. *et al.* Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).
3. Holmes, D. S., Ripple, A. L. & Manheimer, M. A. Energy-Efficient Superconducting Computing—Power Budgets and Requirements. *IEEE Trans. Appl. Supercond.* **23**, 1701610 (2013).
4. Mukhanov, O. A. Energy-efficient single flux quantum technology. *IEEE Trans. Appl. Supercond.* **21**, 760 (2011).
5. Herr, Q. P., Herr, A. Y., Oberg, O. T. & Ioannidis, A. G. Ultra-low-power superconductor logic. *J. Appl. Phys.* **109**, 103903 (2011).
6. Tolpygo, S. K. *et al.* 20 kA/cm² Process development for superconducting integrated circuits with 80 GHz clock frequency. *IEEE Trans. Appl. Supercond.* **17**, 946 (2007).
7. Manheimer, M. A. Cryogenic computing complexity program: Phase 1 introduction. *IEEE Trans. Appl. Supercond.* **25**, 1301704 (2015).
8. Nagasawa, S., Hinode, K., Satoh, T., Kitagawa, Y. & Hidaka, M. Design of all-dc-powered high-speed single flux quantum random access memory based on a pipeline structure for memory cell arrays. *Supercond. Sci. Technol.* **19**, S325 (2006).
9. Liu, Q. *et al.* Latency and power measurements on a 64-kb hybrid Josephson-CMOS memory. *IEEE Trans. Appl. Supercond.* **17**, 526 (2007).
10. Ware, F. *et al.* Do superconducting processors really need cryogenic memories? The case for cold DRAM. In *Proc. Int'l Sym. On Mem. Sys.* **183**, <https://doi.org/10.1145/3132402.3132424> (2017).
11. Tanaka, M. *et al.* Josephson-CMOS Hybrid Memory with Nanocryotrons. *IEEE Trans. Appl. Supercond.* **27**, 1800904 (2017).
12. Zhao, Q. Y. *et al.* A compact superconducting nanowire memory element operated by nanowire cryotrons. *Supercond. Sci. Technol.* **31**, 035009 (2018).
13. McCaughan, A. N., Toomey, E. A., Schneider, M. L., Berggren, K. K. & Nam, S. W. A kinetic-inductance-based superconducting memory element with shunting and sub-nanosecond write times. *Supercond. Sci. Technol.* **32**, 015005 (2018).
14. Semenov, V. K., Polyakov, Y. A. & Tolpygo, S. K. & Member, S. Very Large Scale Integration of Josephson-Junction- Based Superconductor Random Access Memories. *IEEE Trans. Appl. Supercond.* **29**, 1302809 (2019).
15. Butters, B. A. Development of a scalable superconducting memory. (Massachusetts Institute of Technology, 2018).
16. Rowlands, G. E. *et al.* A cryogenic spin-torque memory element with precessional magnetization dynamics. *Sci. Rep.* **9**, 803 (2019).
17. Rehm, L. *et al.* Sub-nanosecond switching in a cryogenic spin-torque spin-valve memory element with a dilute permalloy free layer. *Appl. Phys. Lett.* **114**, 212402 (2019).

18. Rowlands, G. E. *et al.* Nanosecond reversal of three-terminal spin Hall effect memories sustained at cryogenic temperatures. *arXiv:1909.10613* (2019).
19. Vernik, I. V. *et al.* Magnetic Josephson junctions with superconducting interlayer for cryogenic memory. *IEEE Trans. Appl. Supercond.* **23**, 1701208 (2013).
20. Dayton, I. M. *et al.* Experimental Demonstration of a Josephson Magnetic Memory Cell with a Programmable pi-Junction. *IEEE Magn. Lett.* **9**, 3301905 (2018).
21. Liu, L. *et al.* Spin-torque switching with the giant spin Hall effect of tantalum. *Science* **336**, 555 (2012).
22. Aradhya, S. V., Rowlands, G. E., Oh, J., Ralph, D. C. & Buhrman, R. A. Nanosecond-Timescale Low Energy Switching of In-Plane Magnetic Tunnel Junctions through Dynamic Oersted-Field-Assisted Spin Hall Effect. *Nano Lett.* **16**, 5987 (2016).
23. McCaughan, A. N. & Berggren, K. K. A superconducting-nanowire three-terminal electrothermal device. *Nano Lett.* **14**, 5748 (2014).
24. Ohki, T. A., Mukhanov, O. & Kirichenko, A. Magnetic RAM array architecture. U.S. Patent. **US 9,747,9**, (2017).
25. Dyakonov, M. I. & Perel, V. I. Current-induced spin orientation of electrons in semiconductors. *Phys. Letters* **35A**, 459 (1971).
26. Hirsch, J. Spin Hall Effect. *Phys. Rev. Lett.* **83**, 1834 (1999).
27. Zhang, S. Spin hall effect in the presence of spin diffusion. *Phys. Rev. Lett.* **85**, 393 (2000).
28. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
29. Mihai Miron, I. *et al.* Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. *Nat. Mater.* **9**, 230 (2010).
30. Gurevich, A. V. & Mints, R. G. Self-heating in normal metals and superconductors. *Rev. Mod. Phys.* **59**, 941–999 (1987).
31. Liu, L., Moriyama, T., Ralph, D. C. & Buhrman, R. A. Spin-Torque Ferromagnetic Resonance Induced by the Spin Hall Effect. *Phys. Rev. Lett.* **106**, 036601 (2011).
32. Pai, C.-F. *et al.* Spin transfer torque devices utilizing the giant spin Hall effect of tungsten. *Appl. Phys. Lett.* **101**, 122404 (2012).
33. Nguyen, M.-H., Zhao, M., Ralph, D. C. & Buhrman, R. A. Enhanced spin Hall torque efficiency in Pt100–xAlx and Pt100–xHf alloys arising from the intrinsic spin Hall effect. *Appl. Phys. Lett.* **108**, 242407 (2016).
34. Laczkowski, P. *et al.* Experimental evidences of a large extrinsic spin Hall effect in AuW alloy. *Appl. Phys. Lett.* **104**, 142403 (2014).
35. Zhu, L., Ralph, D. C. & Buhrman, R. A. Highly Efficient Spin-Current Generation by the Spin Hall Effect in Au_{1-x}Pt_x. *Phys. Rev. Appl.* **10**, 031001 (2018).
36. Nguyen, M.-H. *et al.* Enhancement of the anti-damping spin torque efficacy of platinum by interface modification. *Appl. Phys. Lett.* **106**, 222402 (2015).
37. Nguyen, M. H. *et al.* Efficient switching of 3-terminal magnetic tunnel junctions by the giant spin Hall effect of Pt85Hf15alloy. *Appl. Phys. Lett.* **112**, 062404 (2018).
38. Sun, J. Z. Spin-current interaction with a monodomain magnetic body: A model study. *Phys. Rev. B* **62**, 570 (2000).
39. Koch, R. H., Katine, J. A. & Sun, J. Z. Time-resolved reversal of spin-transfer switching in a nanomagnet. *Phys. Rev. Lett.* **92**, 088302 (2004).
40. Han, J. *et al.* Room-Temperature Spin-Orbit Torque Switching Induced by a Topological Insulator. *Phys. Rev. Lett.* **119**, 077702 (2017).
41. Shi, S., Ou, Y., Aradhya, S. V., Ralph, D. C. & Buhrman, R. A. Fast, low-current spin-orbit torque switching of magnetic tunnel junctions through atomic modifications of the free layer interfaces. *Phys. Rev. Appl.* **9**, 011002 (2018).
42. Sukegawa, H., Kasai, S., Furubayashi, T., Mitani, S. & Inomata, K. Spin-transfer switching in an epitaxial spin-valve nanopillar with a full-Heusler Co₂FeAl_{0.5}Si_{0.5} alloy. *Appl. Phys. Lett.* **96**, 042508 (2010).
43. Clem, J. R. & Berggren, K. K. Geometry-dependent critical currents in superconducting nanocircuits. *Phys. Rev. B* **84**, 174510 (2011).

Acknowledgements

The authors thank B. Hassick and A. Kreider of Raytheon BBN; M. Onen, I. Charaev, and A. McCaughan of MIT for technical assistance; J. Walter, M. Kamkar, A. D'Addario, I. Vernik, J. Vivalda and M. Renzullo of HYPRES; and Synopsis, Inc. for technical assistance. The magnetic films for our SHE-MTJ devices were deposited by Canon ANELVA, Inc. The majority of nanofabrication in this work was performed at Cornell NanoScale Science and Technology Facility (NSF Grant NNCI-1542081) and Harvard Center for Nanoscale Systems (NSF Grant No. 1541959), members of National Nanotechnology Coordinated Infrastructure Network. The research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), under the Cryogenic Computing Complexity program via contract W911NF-14-C0089. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

Author contributions

Supervision: R.L., G.E.R., T.A.O., R.A.B., O.M. and K.K.B. Device and chip design: M.-H.N., G.J.R., M.G., A.P.W., L.M.R., G.E.R., P.A.T., A.J.-S., D.M. and B.B. Fabrication: G.J.R., A.P.W., S.S., S.V.A., L.Z., P.A.T., D.Y., E.T., M.C. and R.B. Measurements and analyses: M.-H.N., M.G., S.R.C., S.S., A.J.-S., R.B. and B.B. Manuscript composition: All authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.-H.N.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020