Applied Physics Letters

RESEARCH ARTICLE | MAY 04 2023

Superconducting pulse conserving logic and Josephson-SRAM

Quentin Herr ; Trent Josephsen ; Anna Herr

Appl. Phys. Lett. 122, 182604 (2023)
https://doi.org/10.1063/5.0148235

Applied Physics Letters ARTICLE scitation.org/journal/apl

Superconducting pulse conserving logic and Josephson-SRAM

Cite as: Appl. Phys. Lett. **122**, 182604 (2023); doi: 10.1063/5.0148235
Submitted: 28 February 2023 · Accepted: 20 April 2023 ·
Published Online: 4 May 2023

Quentin Herr,[a] Trent Josephsen, and Anna Herr

AFFILIATIONS

imec USA, Kissimmee, Florida 34744, USA

Note: This paper is part of the APL Special Collection on Advances in Superconducting Logic.
[a]**Author to whom correspondence should be addressed:** quentin.herr@imec-int.com

ABSTRACT

Superconducting digital pulse-conserving logic and Josephson static random access memory (JSRAM) memory together enable scalable circuits with energy efficiency 100× beyond leading-node CMOS. Circuit designs support high throughput and low latency when implemented in an advanced fabrication stack with high-critical-current-density Josephson junctions of 1000 $\mu A/\mu m^2$. Pulse-conserving logic produces one single-flux-quantum output for each input and includes a three-input, three-output gate producing logical or3, majority3, and and3. Gate macros using dual-rail data encoding eliminate inversion latency and produce efficient implementations of all standard logic functions. A full adder using 70 Josephson junctions has a carry-out latency of 5 ps corresponding to an effective 12 levels of logic at 30 GHz. JSRAM memory uses single-flux-quantum signals throughout an active array to achieve throughput at the same clock rate as the logic. The unit cell has eight Josephson junctions, a signal propagation latency of 1 ps, and a footprint of 2 μm^2. Projected density of JSRAM is 4 MB/cm^2, and computational density of pulse-conserving logic is on par with leading node CMOS accounting for power densities and clock rates.

Single flux quantum (SFQ) integrated circuits have shown slow-but-steady progress in clock rates, bit-error rates, interconnects, and energy efficiency, with fundamentals established in small-scale demonstrations. The first demonstrations emphasized high data rates, with a binary counter reported at 100 GHz already in 1982.[1] Measurements of bit error rate (BER) show good performance across the SFQ logic families including rapid-SFQ (RSFQ), quantum flux parametron (QFP), and reciprocal quantum logic (RQL).[2–6] SFQ logic families operate in the thermal limit, meaning that the devices are sized to achieve the desired bit-error rate based on Johnson noise in the Josephson junctions (JJs) at the operating temperature, typically LHe at 4.2 K. Despite the similarities at the gate level, SFQ logic families differ in power distribution and timing characteristics, with widely divergent scalability and performance.

RSFQ has had success in decimation filters for mixed-signal applications[7,8] where scale is modest and bit rates are high, but scale is limited due to DC power delivery to a current draw of 1 A per 1000 gates, and by the timing uncertainty of free-running pulses. Energy efficient SFQ (eSFQ/ERSFQ)[9,10] and dynamic SFQ (DSFQ)[11] variants have these same limitations. Asynchronous approaches[12,13] can guarantee functional circuits, but at a cost to throughput, as a reduced rate of computation is the only solution to timing uncertainty. Partitioning the circuit onto floating ground planes and "recycling," the current has been proposed to reduce current draw, but the interconnect overheads are high[14] and demonstrations limited to 16 partitions with nearest-neighbor interconnect.[15–17] With current recycling, passive transmission line interconnect is unsupported and further scaling is in doubt.[18]

QFP logic benefits from the scalable power distribution and timing stability that comes from AC bias. In fact, a resonant clock distribution network was first put forward in the context of QFPs.[19] Adiabatic QFPs (aQFPs) have a switching energy of 10 zJ at a 1 GHz clock rate,[20,21] but the advantages of the quasi-adiabatic potential are partially offset by the need for larger junctions, so the energy dissipation for a given BER[5] is similar to that of RQL.[6] QFPs are used as auxiliary circuits such as qubit readout,[22,23] where clock rates are low. Demonstrations have advanced to a 4-bit CPU core,[24] but there is a fundamental limitation in scaling to high performance as the technology supports only one level of logic per clock cycle, translating to high latency and a low logic clock rate. QFPs have high dynamic power dissipation despite low switching energy, as every gate switches every cycle. QFPs do not support transmission line interconnect.[25]

RQL has stable timing, multiple levels of logic per pipeline stage, low power dissipation with low BER, and low junction count.[4,6,26]

Demonstrations of superconducting transmission line interconnect include synchronous data links chip-to-chip[27] and isochronous links for board-to-board.[28,29] Highly functional 8- and 16-bit CPUs have been presented[30] with AC power distribution at scale.[31] Because SFQ logics are pulse-based, inversion is expensive. The inverting RQL "AnotB" gate[4] has setup time between the inputs, which adds latency in proportion to the local timing uncertainly despite stable timing on a global scale. Phase mode logic (PML)[32] encodes logical inversion as a signal-polarity inversion in analogy to CMOS, but the PML data rate is limited to half the resonator clock rate, and inversion has half a cycle of latency.

In this Letter, we present pulse conserving logic (PCL) and Josephson static random access memory (JSRAM) memory that have the potential to compete with conventional CMOS in integration scale and performance. These SFQ designs fully exploit the resources of the fabrication process proposed in a companion paper.[33] The fabrication stack and resonant power network apply equally to logic and memory and support up to 400M AC-biased JJs per cm^2. Fabrication stack resources were codesigned with the physical layouts of the circuits and carefully balanced to provide dramatic improvements in density relative to legacy designs.

PCL uses dual rail signals. Like CMOS, energy is dissipated only on transitions, and there is no penalty in clock rate or latency for inversion. Our logic gates have the same number of inputs as outputs and preserve pulses, meaning each input SFQ is routed to an output. Gate outputs have symmetric double-well potentials, improving parametric margins and power dissipation. Conservation of pulses enables complex logical functions at reduced latency. The primary set of PCL gates is shown in Fig. 1(a). The Josephson transmission line (JTL) is a buffer that passes the pulse to a single output while providing isolation and transimpedance gain.[34] The oa2 gate passes the first pulse to an or2 output, and the second pulse to an and2. The three-input, three-output oma3 gate passes the first pulse to or3, the second to maj3 (majority function), and the third to and3. Fanout of the buf is two, and fanout of the oa2 is one-half, meaning a buf stage is required between interconnected gates. Fanout of the oma3 is one-quarter, meaning two buf stages are required between gates, with the second stage sized up by a factor of two.

Compared to an earlier implementation, the oa2 logic gate does not require transformers between the signal inductors[4] and does not have inductors between outputs or from an output to ground. These advantages permit generalization to oma3. All inputs connect to all outputs, and the applied flux bias on the interconnect distributes the output junctions along a Josephson phase wheel.[35] Negative input pulses reverse the rotation of the phase wheel. Extension to a four input, four-output gate is possible but would require four flux biases and high drive strength on the inputs. To produce the standard logic functions, such as or4 and and4, a two-level cascade of two-input gates is more efficient.

As with the other ac powered logics, the SFQ signals are bipolar, with the positive pulse representing a transition to logical "1" and a negative pulse to logical "0." Waveforms of superconducting phase as shown in Fig. 1(b) resemble conventional level-based logic. Both low and high levels can persist for multiple clock cycles.

Circuit simulations use a Josephson junction model with a nominal critical current, I_c, of 100 μA, an internal capacitance of 3.5 fH, and a voltage gap of 2.75 mV. The bias tap for a standard JTL junction is modeled as a sinusoidal voltage source in series with a 6.2 fF capacitor. Physical implementation connects, e.g., 50 such capacitors in parallel to a single inductor of size 90 pH that is driven at the LC resonance frequency of 30 GHz. Scaling to a chip-level 2D grid of bias resonators is described in Ref. 33. Both internal and bias capacitors shunt the device, as the bias tap is connected to signal ground, and both scale with device critical current. The junctions are self-shunted, meaning that the combined shunt capacitance is low enough that the gap voltage of itself prevents hysteresis in the I–V curve. The junction internal resistance has only a secondary effect on switching speed. We include an external shunt resistance, R, such that $I_cR = 4.4$ mV, which allows voltage sub-gap plasmon oscillations to ringdown between SFQ switching events.

Gate delays, shown in Fig. 1(c), include the four serial junctions of JTL interconnect needed to connect gate to gate. Latencies are only a few picoseconds using a model for 1 mA/μm^2 Si-barrier junctions, which have a 10× the critical current density at 4× thicker barrier relative to Nb/AlOx/Nb.[36] Materials and models are discussed in greater detail in Ref. 33. Latencies depend strongly on an initial state of the gate and the number of simultaneous inputs, which can be understood qualitatively in terms of potential energy as shown in Fig. 1(d). In real designs, intermediate pulse timings may occur that are neither well-separated nor simultaneous. The curves in Fig. 1(c) should be interpreted as the best-case and worst-case timing extremes that envelope possible gate delays.

Dual rail encoding solves the fundamental timing problems of inversion in SFQ pulse-based logics. While global timing is stable as signals travel through the wave pipeline at the rate determined by the ac clock, significant local timing uncertainty arises from thermal noise, parameter spread, and relative pulse arrival time. This translates to latency in SFQ inverting logic due to race conditions. Dual rail inversion is "free" in analogy to CMOS and enables conventional combinational logic. The cost of duplication in dual rail is partially recouped when mapped onto our PCL gate primitives as the multiple outputs are all available. The dual rail macros, shown in Fig. 1(e), produce all the logical functions of a standard CMOS gate library. The xor3 macro implements a full adder with the sum generated by xor3 and the carry generated by maj3 from the first stage. The full adder has 70 JJs and carry-out latency of 5 ps corresponding to 12 levels of two-input-gate logic per pipeline stage at 30 GHz. Junction count is about twice the number of transistors in a CMOS implementation.[37]

SFQ-compatible memory has historically demonstrated quite modest density despite fundamental SFQ state retention in superconducting loops. The difficulty in scaling superconducting memory centers on the addressing, which requires fanout. Approaches can be categorized as active, which has the same signal levels and power distribution as logic, i.e., nondestructive read-out (NDRO);[38,39] passive, meaning the array must be driven with current levels at the perimeter, e.g., vortex-transitional;[40–43] and magnetic, which requires hysteretic tunnel devices.[44,45] Passive and magnetic approaches require drive levels corresponding to hundreds of SFQ pulses, which constrains density, power dissipation, throughput, and latency. Drivers of current must be able to target and tune the levels to accommodate analog addition of word and bit line signals in the unit cell. The vortex memory has transformers in the unit cell that cannot be miniaturized below 0.25 μm with practical drive levels.[43] Magnetic memory devices would require even larger currents to produce the demonstrated switching fields.[46,47]

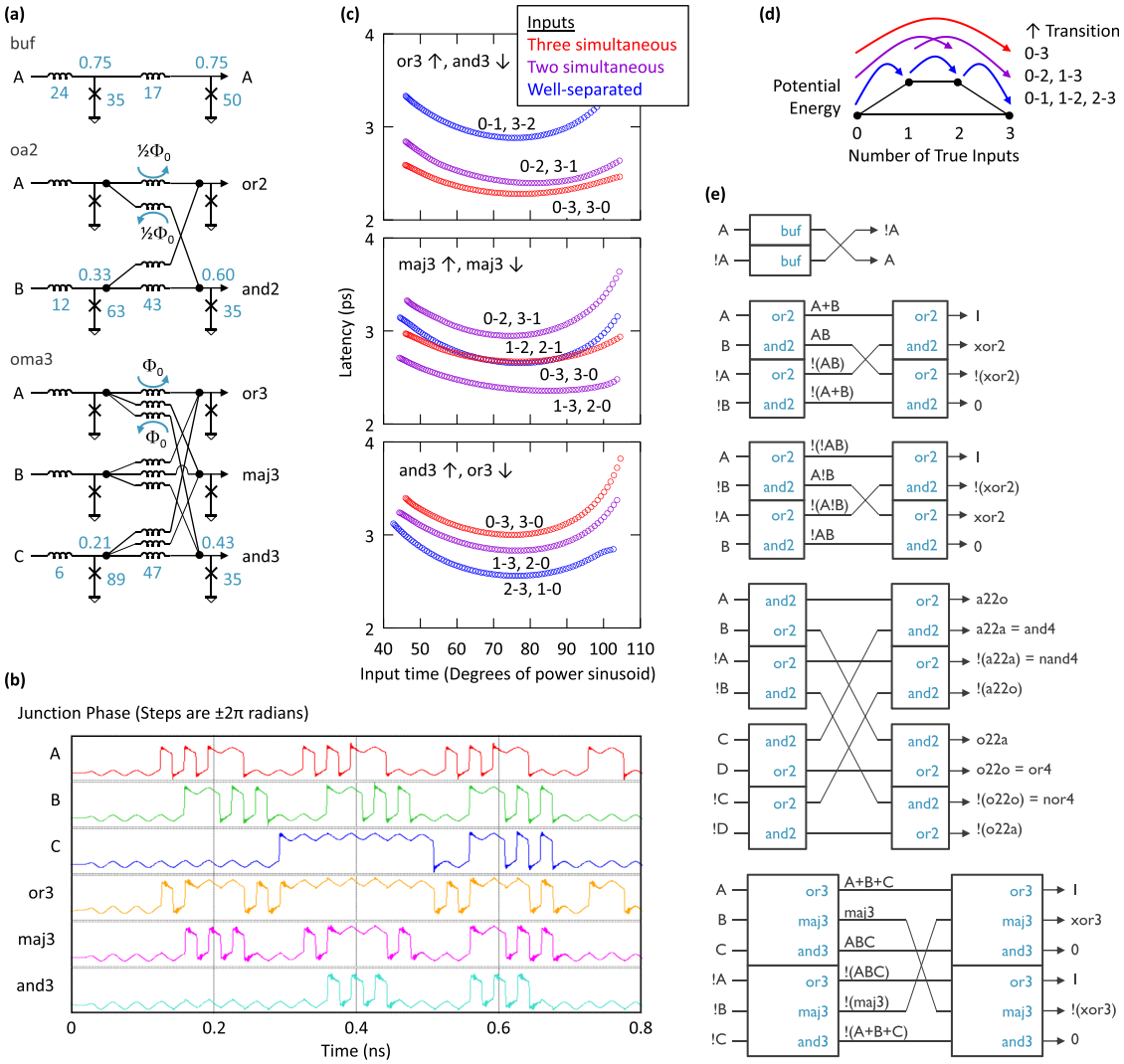

FIG. 1. PCL logic schematics, modeling, and macros. (a) Gate schematics show the single-input JTL buf (buffer) used for isolation and gain, the two-input oa2 gate that produces logical or2 and and2, and the three-input oma3 gate that produces or3, maj3 (majority), and and3. All junctions are AC biased (not shown). Values for inductors are shown in pH, junction critical currents in μA, and peak-amplitude AC bias current as a fraction of junction critical current. Blue arrows indicate flux bias on inductors where Φ_0 is the single flux quantum. (b) WRSpice simulation of the oma3 gate showing waveforms of junction phase, with each positive SFQ pulse equal to a rising edge and each negative pulse equal to a falling edge. (c) Simulations of oma3 gate delay as a function of input time shown for different initial states and relative input pulse timings. Positive-pulse delays on Or correspond to negative-pulse delays on And, and vise versa. (d) A diagram of the potential energy of SFQ stored within the gate, and trajectories between states that correspond to the gate delays in (c). For clarity, only positive-pulse transitions are shown. (e) PCL dual rail macros produce a complete set of standard logic functions using only positive-logic primitives.

Figure 2(a) shows the JSRAM unit cell, an active array that can be understood as an optimized version of NDRO with about 3× fewer JJs. The unit cell has four JJs for storage and readout of state and four AC-biased JJs that are JTL-based active address lines. All signals and internal states are SFQ. Resistive coupling to the address lines is used instead of the "escape" junctions used in RSFQ logic. This is feasible as the clock period is long compared to the SFQ pulse width. The current scale in the LR loops is set by the inductance, and the relaxation time is then set by the resistance to be small compared to the clock period. Simulated bias current margins of the unit cell embedded in an array exceed ±30% for global bias current and flux bias amplitudes, which affords tolerance of crosstalk and mistargeting across chip. Simulations involved a 5 × 5 array at 30 GHz with all devices on the same clock phase. Larger arrays would require an advance in a clock phase in both the X and Y directions to wave-pipeline the signals with a latency of 1 ps per stage.

Physical layout of the unit cell, shown in Fig. 2(b), achieves densities for active memory even higher than contemplated for passive arrays. The fabrication stackup proposed in our companion paper[33] enables 4 MB/cm^2 density, which represents a 600× increase in

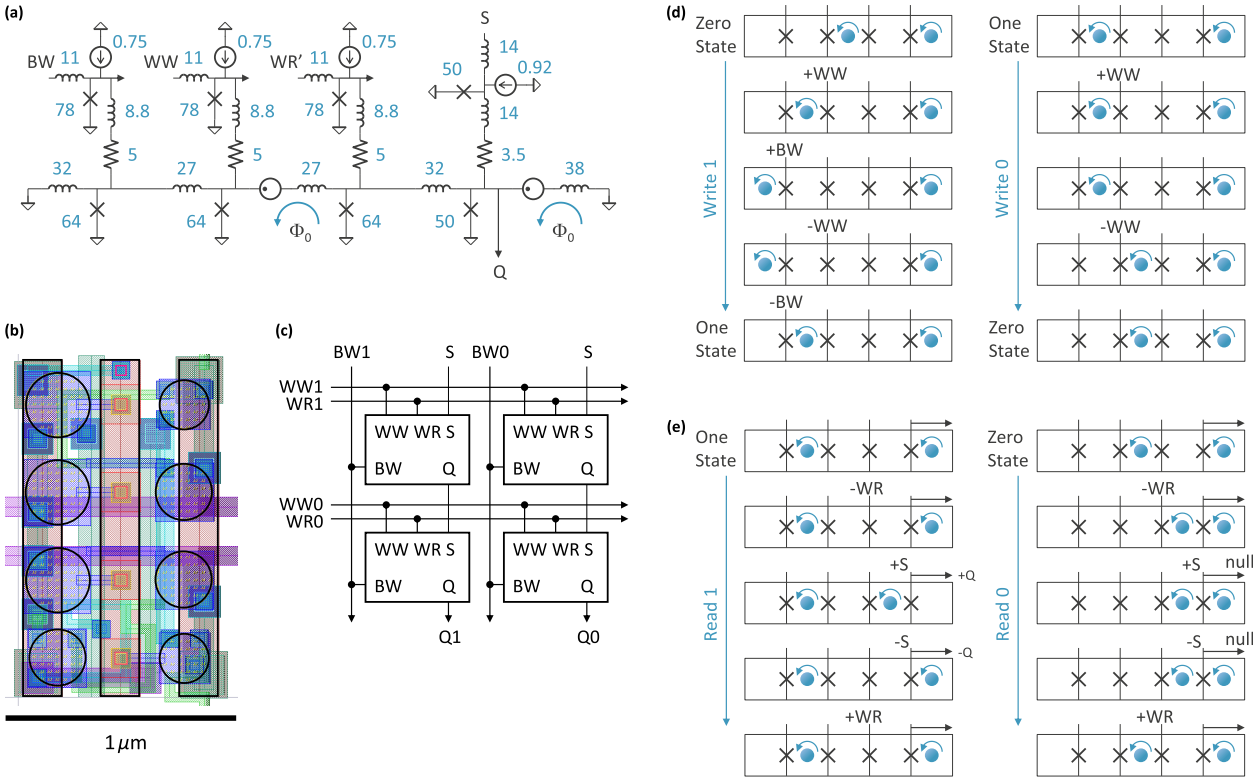

FIG. 2. The JSRAM memory design. (a) The unit cell schematic shows inductors in units of pH, JJ critical current in μA, peak AC bias as a fraction of junction critical current, and resistors in Ω. Blue arrows indicate flux bias to initialize two storage loops. This is accomplished with pulsed voltage sources in the schematic but with magnetic flux bias in the physical implementation. (b) Physical layout of the unit cell has dimensions of 1×2 μm^2. High critical current density and high fill density of about 25% accommodates the junctions (circular outlines). The flux bias is implemented in a single metal layer with a central primary and secondaries on each side (rectangular outlines). (c) The memory fabric is a systolic array of unit cells. Word write (WW), bit write (BW), and word read (WR) signals fanout to all cells in a vine configuration, whereas the strobe-to-output (S-to-Q) bit read is a daisy chain through the array. (d) Operation of the unit cell is illustrated as time sequences of initial state, event, and final state. Stored SFQ are shown as filled blue circles with current polarity indicated by blue arrows. If the current of a stored SFQ interferes constructively with the current associated with an event in any JJ, the JJ critical current will be exceeded with the effect of moving the stored SFQ to the adjacent loop. Otherwise the event will be lost across the coupling resistor with no change in state. While the readout of state involves conditional propagation of the bit read, a cell that is not selected with word read will perform pass-through of the bit read irrespective of state.

density relative to legacy NDRO circuits implemented in a 250 nm Nb fabrication process. Increased density of fabrication resources is achieved across the stack including bias taps, junctions, and inductive wiring. Moving from 250 to 50 nm critical dimension in the wiring layers increases inductance density by a factor of 25 and moving to high-kinetic-inductance materials increases density by another 10×. The unit cell uses only the first four wiring layers (M1–M4) for inductors, and a single metal layer to implement the two SFQ flux biases, with mutual inductance of 0.5 pH and 4 mA current in the primary, which is consistent with the critical current of low-resistivity NbTiN with a penetration depth of 200 nm and a 200×200 nm^2 cross section.[48] The remaining top two metal layers are available for passive transmission line wireup across subarrays.

The array, shown in Fig. 2(c), is wave pipelined with throughput at 30 GHz, making it the functional equivalent of CMOS SRAM. Read and Write operations are illustrated in Fig. 2(d). Important features incorporated by the design have been published, including SFQ readout on the bit line,[49] and equal-energy levels for the different flux states in the memory cell based on moving an SFQ between superconducting loops.[50,51]

Table I compares the size and performance of a PCL bfloat16 multiply-accumulate unit (MAC) to the TPUv4 MAC implemented in leading-node CMOS.[52] We use a carry-save architecture with ripple-carry for the final add, consistent with energy-efficient CMOS design.

TABLE I. Comparison of our proposed SFQ technology to a tensor processor unit (TPU) implemented in leading-node CMOS.

	SFQ	TPUv4i	Ratio
Process node	28 nm	7 nm	4
MAC area (μm^2)	6700[a]	600[b]	11
Clock (GHz)	30	1.05	29
Computational density (M op/s/μm^2)	4.5	1.75	2.6
Energy efficiency (op/pJ)	69[c]	3.1	22

[a]Using a derated 50% critical resource utilization relative to JSRAM.
[b]An estimated 10% of the die area dedicated to MAC units.[52]
[c]Including a 325 W/W cryocooling overhead.

The MAC contains an 8 × 8 array multiplier for the significand and an 8-bit adder for the exponent, with a small additional overhead for alignment and normalization. The total device count is 14,330 JJs, including the systolic array 8 × 8 multiplier and the single-precision floating-point 32-bit accumulator. The multiplier contains 49 and2 gates, 7 half-adders, and 42 full-adders for a total of 4284 JJs. The accumulator has 4163 JJs. The area is estimated by derating JSRAM design density by a factor of two. As shown in the table, power efficiency at the level of on-chip logic is 20× better than CMOS while including a 325× cooling overhead. Computational density of SFQ at 28 nm is higher than 7 nm CMOS, in part due the high clock rate of 30 GHz and in part due to the high utilization of chip real-estate used for logic.

PCL and JSRAM technology enable even higher energy efficiency on the system level, up to 100× due to superconducting interconnect. Low power dissipation of the JSRAM enables efficient chip stacking providing sufficient SRAM-to-compute. Ultimate system performance is determined by capacity and bandwidth to cryo-DRAM that is an area of active research.[53–58]

In conclusion, we have put forward advances in SFQ logic and memory design that are essential to scaling. AC-biased PCL logic provides economical dual-rail macros with low-latency inversion, solving one of the central problems of pulse driven logics. JSRAM is a superconducting memory providing SRAM capability and high density. Both technologies are enabled by the high-integration-density fabrication stack described in a companion paper. PCL logic projects to 12 levels of logic per stage with a 30 GHz clock, and a computational density on par with 7 nm CMOS. At the gate level, energy efficiency is 20× better than CMOS including cryocooler efficiency. JSRAM memory projects to 4 MB/cm^2 with throughput at the 30 GHz clock rate and a latency of 1 ps per stage in the active array. These technologies enable practical architectures with 100× better power efficiency at the system level and 30× higher clock rate.

The work at imec and imec USA was supported by imec INVEST+ and by Osceola County.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

All authors contributed to the work equally.

Quentin Herr: Conceptualization (equal); Investigation (equal); Writing – original draft (equal). **Trent Josephsen:** Conceptualization (equal); Investigation (equal); Writing – original draft (equal). **Anna Herr:** Conceptualization (equal); Investigation (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1]C. A. Hamilton and F. L. Lloyd, IEEE Electron Device Lett. **3**, 335 (1982).
[2]Q. P. Herr and M. J. Feldman, Appl. Phys. Lett. **69**, 694 (1996).
[3]Q. P. Herr, M. W. Johnson, and M. J. Feldman, IEEE Trans. Appl. Supercond. **9**, 3594 (1999).
[4]Q. P. Herr, A. Y. Herr, O. T. Oberg, and A. G. Ioannidis, J. Appl. Phys. **109**, 103903 (2011).
[5]N. Takeuchi, H. Suzuki, and N. Yoshikawa, Appl. Phys. Lett. **110**, 202601 (2017).
[6]Q. Herr, A. Braun, A. Brownfield, E. Rudman, D. Dosch, T. Josephsen, and A. Herr, Supercond. Sci. Technol. **35**, 025017 (2022).
[7]O. A. Mukhanov, D. Gupta, A. M. Kadin, and V. K. Semenov, Proc. IEEE **92**, 1564 (2004).
[8]R. Kashima, I. Nagaoka, M. Tanaka, T. Yamashita, and A. Fujimaki, IEEE Trans. Appl. Supercond. **31**, 1301006 (2021).
[9]O. A. Mukhanov, IEEE Trans. Appl. Supercond. **21**, 760 (2011).
[10]A. F. Kirichenko, I. V. Vernik, M. Y. Kamkar, J. Walter, M. Miller, L. R. Albu, and O. A. Mukhanov, IEEE Trans. Appl. Supercond. **29**, 1302407 (2019).
[11]S. V. Rylov, IEEE Trans. Appl. Supercond. **29**, 1300805 (2019).
[12]Z. J. Deng, N. Yoshikawa, S. Whiteley, and T. Van Duzer, IEEE Trans. Appl. Supercond. **7**, 3634 (1997).
[13]G. Tzimpragos, J. Volk, D. Vasudevan, N. Tsiskaridze, G. Michelogiannakis, A. Madhavan, J. Shalf, and T. Sherwood, IEEE Micro **41**, 71 (2021).
[14]M. W. Johnson, Q. P. Herr, D. J. Durand, and L. A. Abelson, IEEE Trans. Appl. Supercond. **13**, 507 (2003).
[15]T. V. Filippov, A. Sahu, S. Sarwana, D. Gupta, and V. K. Semenov, IEEE Trans. Appl. Supercond. **19**, 580 (2009).
[16]S. B. Kaplan, IEEE Trans. Appl. Supercond. **22**, 1300103 (2012).
[17]K. Sano, T. Shimoda, Y. Abe, Y. Yamanashi, N. Yoshikawa, N. Zen, and M. Ohkubo, IEEE Trans. Appl. Supercond. **27**, 1300305 (2017).
[18]V. K. Semenov and Y. Polyakov, IEEE Trans. Appl. Supercond. **29**, 1302304 (2019).
[19]M. Hosoya, W. Hioe, J. Casas, R. Kamikawai, Y. Harada, Y. Wada, H. Nakane, R. Suda, and E. Goto, IEEE Trans. Appl. Supercond. **1**, 77 (1991).
[20]N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, Appl. Phys. Lett. **102**, 052602 (2013).
[21]T. Yamae, N. Takeuchi, and N. Yoshikawa, J. Appl. Phys. **126**, 173903 (2019).
[22]M. H. Volkmann, L. J. Swenson, P. Spear, B. Bumble, P. K. Day, B. H. Eom, J. D. Whittaker, F. Altomare, A. J. Berkley, A. Kleinsasser *et al.*, in *15th International Superconductive Electronics Conference (ISEC)* (IEEE, 2015), pp. 1–3.
[23]J. A. Grover, J. I. Basham, A. Marakov, S. M. Disseler, R. T. Hinkey, M. Khalil, Z. A. Stegen, T. Chamberlin, W. DeGottardi, D. J. Clarke *et al.*, PRX Quantum **1**, 020314 (2020).
[24]C. L. Ayala, T. Tanaka, R. Saito, M. Nozoe, N. Takeuchi, and N. Yoshikawa, IEEE J. Solid-State Circuits **56**, 1152 (2021).
[25]F. China, N. Tsuji, T. Narama, N. Takeuchi, T. Ortlepp, Y. Yamanashi, and N. Yoshikawa, IEEE Trans. Appl. Supercond. **27**, 1300205 (2016).
[26]A. Y. Herr, Q. P. Herr, O. T. Oberg, O. Naaman, J. X. Przybysz, P. Borodulin, and S. B. Shauck, J. Appl. Phys. **113**, 033911 (2013).
[27]J. Egan, M. Nielsen, J. Strong, V. Talanov, E. Rudman, B. Song, Q. Herr, and A. Herr, Supercond. Sci. Technol. **35**, 105010 (2022).
[28]V. V. Talanov, D. Knee, D. Harms, K. Perkins, A. Urbanas, J. Egan, Q. Herr, and A. Herr, Supercond. Sci. Technol. **35**, 055011 (2022).
[29]H. Dai, C. Kegerreis, D. W. Gamage, J. Egan, M. Nielsen, Y. Chen, D. Tuckerman, S. E. Peek, B. Yelamanchili, M. Hamilton *et al.*, Supercond. Sci. Technol. **35**, 045014 (2022).
[30]M. Vesely, *paper presented at the Applied Superconductivity Conference*, Washington State Convention Center, Seattle, WA, October 29, 2018.
[31]J. A. Strong, V. V. Talanov, M. E. Nielsen, A. C. Brownfield, N. Bailey, Q. P. Herr, and A. Y. Herr, Nat. Electron. **5**, 171 (2022).
[32]D. Carmean, A. Braun, A. Y. Herr, and Q. P. Herr, "Phase-mode based superconducting logic," US patent 9,543,959 (2017).
[33]A. Herr, Q. Herr, S. Brebels, M.-S. Kim, A. Pokhrel, B. Hodges, T. Josephsen, S. ONeal, R. Bai, K. Nowack *et al.*, arXiv:2303.16792 (2023).
[34]K. K. Likharev and V. K. Semenov, IEEE Trans. Appl. Supercond. **1**, 3 (1991).
[35]G. S. Lee and H. L. Ko, IEEE Trans. Appl. Supercond. **3**, 3001 (1993).
[36]B. Baek, P. D. Dresselhaus, and S. P. Benz, IEEE Trans. Appl. Supercond. **16**, 1966 (2006).
[37]N. Zhuang and H. Wu, IEEE J. Solid-State Circuits **27**, 840 (1992).
[38]M. Dorojevets, P. Bunyk, and D. Zinoviev, IEEE Trans. Appl. Supercond. **11**, 326 (2001).

[39]R. Burnett, R. Clarke, T. Lee, H. Hearne, J. Vogel, in Q. Herr, and A. Herr, in *Proceedings of the International Symposium on Memory Systems* (IEEE, 2018), pp. 321–323.
[40]S. Tahara, I. Ishida, Y. Ajisawa, and Y. Wada, J. Appl. Phys. **65**, 851 (1989).
[41]S. Nagasawa, Y. Hashimoto, H. Numata, and S. Tahara, IEEE Trans. Appl. Supercond. **5**, 2447 (1995).
[42]S. Nagasawa, H. Numata, Y. Hashimoto, and S. Tahara, IEEE Trans. Appl. Supercond. **9**, 3708 (1999).
[43]V. K. Semenov, Y. A. Polyakov, and S. K. Tolpygo, IEEE Trans. Appl. Supercond. **29**, 1302809 (2019).
[44]I. M. Dayton, T. Sage, E. C. Gingrich, M. G. Loving, T. F. Ambrose, N. P. Siwak, S. Keebaugh, C. Kirby, D. L. Miller, A. Y. Herr *et al.*, IEEE Magn. Lett. **9**, 3301905 (2018).
[45]M.-H. Nguyen, G. J. Ribeill, M. V. Gustafsson, S. Shi, S. V. Aradhya, A. P. Wagner, L. M. Ranzani, L. Zhu, R. Baghdadi, B. Butters *et al.*, Sci. Rep. **10**(1), 248 (2020).
[46]E. Gingrich, B. M. Niedzielski, J. A. Glick, Y. Wang, D. Miller, R. Loloee, W. Pratt, Jr., and N. O. Birge, Nat. Phys. **12**, 564 (2016).
[47]N. O. Birge, A. E. Madden, and O. Naaman, Proc. SPIE **10732**, 107321M (2018).
[48]A. Pokhrel, A. Guptal, M.-S. Kim, J.-P. Soulie, S. K. Sarkar, Y. Canvel, V. Renaud, B. Kenens, B. Hodges, S. O'Neal, Q. Herr, A. Herr, and Z. Tökei, in *IEEE International Interconnect Technology Conference* (IEEE, 2023).
[49]S. Polonsky, A. Kirichenko, V. Semenov, and K. Likharev, IEEE Trans. Appl. Supercond. **5**, 3000 (1995).
[50]N. Nair and Y. Braiman, Supercond. Sci. Technol. **31**, 115012 (2018).
[51]E. Toomey, M. Onen, M. Colangelo, B. Butters, A. McCaughan, and K. Berggren, Phys. Rev. Appl. **11**, 034006 (2019).
[52]N. P. Jouppi, D. H. Yoon, M. Ashcraft, M. Gottscho, T. B. Jablin, G. Kurian, J. Laudon, S. Li, P. Ma, X. Ma *et al.*, in *ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA)* (IEEE, 2021), pp. 1–14.
[53]F. Ware, L. Gopalakrishnan, E. Linstadt, S. A. McKee, T. Vogelsang, in K. L. Wright, C. Hampel, and G. Bronner, in *Proceedings of the International Symposium on Memory Systems* (ACM, 2017), pp. 183–188.
[54]F. Wang, T. Vogelsang, B. Haukness, and S. C. Magee, in *IEEE International Memory Workshop (IMW)* (IEEE, 2018), pp. 1–4.
[55]J.-H. Bae, J.-W. Back, M.-W. Kwon, J. H. Seo, K. Yoo, S. Y. Woo, K. Park, B.-G. Park, and J.-H. Lee, IEEE Electron Device Lett. **40**, 1614 (2019).
[56]G-h Lee, D. Min, I. Byun, and J. Kim, in *Proceedings of the 46th International Symposium on Computer Architecture* (IEEE, 2019), pp. 774–787.
[57]G.-H. Lee, S. Na, I. Byun, D. Min, and J. Kim, in *ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA)* (IEEE, 2021), pp. 637–650.
[58]E. Garzón, Y. Greenblatt, O. Harel, M. Lanuzza, and A. Teman, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (IEEE, 2021).