

Tutorial: High-speed low-power neuromorphic systems based on magnetic Josephson junctions

Cite as: J. Appl. Phys. **124**, 161102 (2018); <https://doi.org/10.1063/1.5042425>

Submitted: 31 May 2018 • Accepted: 28 September 2018 • Published Online: 25 October 2018

Michael L. Schneider, Christine A. Donnelly and  Stephen E. Russek

COLLECTIONS

 This paper was selected as an Editor's Pick



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Tutorial: Brain-inspired computing using phase-change memory devices](#)

Journal of Applied Physics **124**, 111101 (2018); <https://doi.org/10.1063/1.5042413>

[Challenges in materials and devices for resistive-switching-based neuromorphic computing](#)

Journal of Applied Physics **124**, 211101 (2018); <https://doi.org/10.1063/1.5047800>

[Tutorial: Basic principles, limits of detection, and pitfalls of highly sensitive SQUID magnetometry for nanomagnetism and spintronics](#)

Journal of Applied Physics **124**, 161101 (2018); <https://doi.org/10.1063/1.5045299>

Lock-in Amplifiers
up to 600 MHz



Zurich
Instruments



Tutorial: High-speed low-power neuromorphic systems based on magnetic Josephson junctions

Michael L. Schneider,¹ Christine A. Donnelly,² and Stephen E. Russek¹

¹National Institute of Standards and Technology, Boulder, Colorado 80305, USA

²Electrical Engineering, Stanford University, Palo Alto, California 94305, USA

(Received 31 May 2018; accepted 28 September 2018; published online 25 October 2018)

Josephson junctions and single flux quantum (SFQ) circuits form a natural neuromorphic technology with SFQ pulses and superconducting transmission lines simulating action potentials and axons. Josephson junctions consist of superconducting electrodes with nanoscale barriers that modulate the coupling of the complex superconducting order parameter across the junction. When the order parameter undergoes a 2π phase jump, the junction emits a voltage pulse with an integrated amplitude of a flux quantum $\phi_0 = h/(2e) = 2.068 \times 10^{-15}$ V s. The coupling across a junction can be controlled and modulated by incorporating the nanoscale magnetic structure in the barrier. The magnetic state of embedded nanoclusters can be changed by applying small current or field pulses, enabling both unsupervised and supervised learning. The advantage of this magnetic/superconducting technology is that it combines natural spiking behavior and plasticity in a single nanoscale device and is orders of magnitude faster and lower energy than other technologies. Maximum operating frequencies are above 100 GHz, while spiking and training energies are $\sim 10^{-20}$ J and 10^{-18} J, respectively. This technology can operate close to the thermal limit, which at 4 K is considerably lower energy than in a human brain. The transition from deterministic to stochastic behavior can be studied with small temperature modifications. Here, we present a tutorial on the spiking behavior of Josephson junctions; the use of the nanoscale magnetic structure to modulate the coupling across the junction; the design and operation of magnetic Josephson junctions, device models, and simulation of magnetic Josephson junction neuromorphic circuits; and potential neuromorphic architectures based on hybrid superconducting/magnetic technology. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5042425>

I. INTRODUCTION

Deep neural nets have been successful in many tasks,¹ including image recognition/classification, language translation, speech recognition, medical image reconstruction, medical diagnosis, and robotics. Current implementations are mostly software based and are limited by software algorithms for training and the inherent serial nature of the complementary metal-oxide-semiconductor (CMOS) processors used to implement the algorithms. Developing neuromorphic hardware has been a longstanding challenge.^{2,3} Biological neural systems are characterized by complex spike-based communication, co-location of memory and processing, plasticity, and extreme complexity. Extreme complexity includes a large number of neurons (10^{11}) and synapses (10^{15}), large fan-out (>1000), 3-dimensional connectivity, large number of physical parameters and mechanisms working together, and chemical/physics based processing where all components evolve simultaneously.⁴

Many neuromorphic systems have been proposed and developed to various stages of complexity.^{5,6} The most advanced are CMOS based systems, such as True North,⁷ SpiNNaker,⁸ Neurogrid,⁹ and the BrainScaleS machine.¹⁰ Most of these systems imitate neural function with mixed analog and digital

techniques using conventional silicon device structures. Non-CMOS neuromorphic devices have been proposed including memristors,^{11–13} nanowires,¹⁴ photonics,^{15,16} and spin based systems.^{17–20} Most of these technologies do not have low energy plasticity combined with natural spiking at the device level. While there is considerable debate about the added advantage and necessity of spiking systems,^{21,22} when operating at extreme speeds (>100 GHz) and ultra-low energy ($<10^{-20}$ J/spike), systems that use asynchronous spikes for data transmission and unsupervised learning are at present the most accessible.

Here, we focus on hybrid magnetic superconducting devices that naturally output voltage spikes and have plasticity due to embedded nanoscale magnetic structures. Superconducting and magnetic device technologies are relatively mature, and large scale integrated systems have been successfully deployed. On the magnetic device side, large non-volatile memories based on magnetic tunnel junctions, using spin-polarized currents to write, have been commercialized.²³ Magnetic devices are uniquely suited for nanoscale memory applications. On the superconducting device side, high speed microprocessors and communication systems have been developed, based mostly on superconducting tunnel junctions.²⁴ One successful application is the Josephson voltage standard that uses large arrays of Josephson junction oscillators to phase lock on an external microwave signal and output a precise voltage that is related to the flux quantum.²⁵ Superconducting

Note: This paper is part of the Special Topic section “New Physics and Materials for Neuromorphic Computation” published in J. Appl. Phys. **124**(15) (2018).



devices are being used for quantum computation and information processing applications.²⁶ Neuromorphic applications of these devices use completely different design concepts than all of these previous applications. For nanoscale digital memory, systems require high uniformity and discrete well-defined states, while neuromorphic systems embrace nonuniformity and a large number of analog states. In quantum computation and information processing, interactions with the environment are minimized while neuromorphic systems embrace thermal fluctuations and interactions with their environment.

Both the superconducting and the magnetic systems can be described by macroscopic order parameters that characterize a coherent state of electrons, which occurs due to electron-electron interactions. Superconducting systems are characterized, in the simplest case, by a complex order parameter $\psi(\vec{r}, t) = \sqrt{n}e^{i\theta}$, which is derived from a two-particle correlation function that describes a many body coherence of the quantum state. For a standard s-wave superconductor, an attractive interaction leads to correlations of a pair of spin up and spin down electrons near the Fermi surface.^{27,28} The two-body correlation can be thought of as a superconducting electron pair (Cooper pair), which has a wave function similar to positronium and an extent given by the coherence length $\xi_0 = \frac{\hbar v_F}{\pi \Delta_{sc}}$, where v_F is the Fermi velocity and Δ_{sc} is the superconducting energy gap. For Nb, one of the commonly used materials in superconducting circuits, the coherence length is $\xi_0 = 38$ nm.²⁹ The superconducting energy gap, Δ_{sc} , is a measure of the minimum energy to create an electronic excitation and the decrease in energy going into the correlated state $U_{sc} = \frac{1}{2}N(0)\Delta_{sc}^2$, where $N(0)$ is the density of states at the Fermi surface.²⁷ For conventional superconductors, the wave function is built up from momentum states near the Fermi surface and has strong charge modulation at wavelengths comparable to lattice spacings. These charge modulations couple with the lattice to cause distortions (phonons) that lower the overall energy of the electron pair. The superconducting pairs are highly overlapping, so the macroscopic charge density is uniform. The square of the magnitude of the order parameter $n(\vec{r}, t)$ can be viewed as a local density of superconducting pairs, while the phase $\theta(\vec{r}, t)$ is related to motion and flow of the superconducting pairs.

For magnetic systems, there is a strong spin-dependent interaction between electrons due to the antisymmetry of the wave function. This asymmetry requires that two electrons of identical spin cannot occupy the same spatial location, which leads to strong correlations in the spin states of neighboring electrons. For the case where neighboring spins tend to align, the order parameter can be expressed as the local magnetization vector, $\vec{M}(\vec{r})$, which is the local magnetic moment per unit volume. For both superconducting and magnetic systems, especially when they are combined, the order parameters may have a more complex structure.

A. Josephson junctions

The devices considered here, shown in Fig. 1, consist of two superconducting electrodes separated by a barrier that

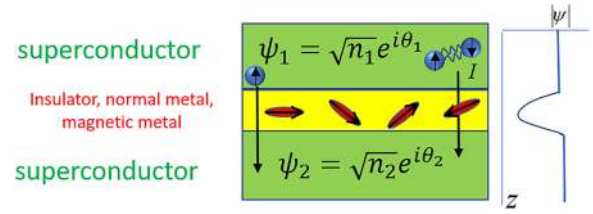


FIG. 1. Josephson junction with two superconducting electrodes separated by an insulating barrier with embedded magnetic nanoclusters (shown in red). The suppression of the superconducting order parameter ψ is shown schematically on the right as a function of vertical z -position. Quasiparticle I_{qp} and superconducting pair I_{sp} currents are shown schematically.

suppresses the superconducting order parameter. The barrier can be an insulator, a normal metal, a magnetic metal or insulator, a constriction, or a nanowire. Figure 1 shows a special type of barrier that has embedded magnetic nanoclusters in a dielectric barrier of thickness ~ 5 nm. If the barrier is sufficiently thin, there will be coupling across the barrier and a dissipationless pair current (shown schematically as the coupled electrons in Fig. 1) can flow up to a critical current I_c . These devices all fall under the category of Josephson junctions, named after Brian Josephson who first predicted the effect in tunnel junctions.³⁰ The dynamics of the superconducting order parameter in a Josephson junction, in the simplest case, can be described by a set of coupled equations³¹

$$i\hbar \frac{\partial \psi_1}{\partial t} = \mu_1 \psi_1 + k \psi_2, \quad (1)$$

$$i\hbar \frac{\partial \psi_2}{\partial t} = \mu_2 \psi_2 + k \psi_1,$$

where $\mu_1 = -\mu_2$ are the pair energy levels on the two sides of the electrodes and k is a coupling energy. The voltage V (half of the energy to transfer a pair) and current I across the device are given by (where n_i is now the number of Cooper pairs on each electrode)

$$V = \frac{1}{2e}(\mu_1 - \mu_2), \quad (2)$$

$$I = \frac{2e\partial n_1}{\partial t} = -\frac{2e\partial n_2}{\partial t}. \quad (3)$$

By equating the real and imaginary parts of these equations, one can derive the Josephson relations which relate the current and voltage to the phase difference $\theta = \theta_2 - \theta_1$ and the time dependence of the phase, respectively,

$$\frac{d\theta}{dt} = \frac{2e}{\hbar} V = \frac{2\pi V}{\phi_0}, \quad (4)$$

$$I = \frac{4ek}{\hbar} \sqrt{n_1 n_2} \sin(\theta) = I_c \sin(\theta), \quad (5)$$

$\phi_0 = \frac{h}{2e} = 2.067833831(13)10^{-15}$ V s is the superconducting magnetic flux quantum and I_c is the junction critical current.

A steady state pair current can flow with no voltage (no dissipation) and is a periodic function of the order parameter phase difference θ , with a maximum value of I_c . I_c is a function of the pair density in the electrodes, which has a strong temperature dependence and goes to zero near the superconducting transition temperature T_c . I_c is also a function of the coupling energy k , which is a sensitive function of the embedded magnetic order. As the pair current and phase difference increase, a voltage is generated proportional to dI/dt and reversible work is done leading to a Josephson energy $U_j = \int IV(t)dt = \int I_c \sin(\theta)(d\theta/dt)(\phi_0/2\pi)dt = (I_c \phi_0/2\pi)[1 - \cos(\theta)]$. Using the standard definition of an inductor $V = L(dI/dt)$, a Josephson device can be considered as a nonlinear inductor with $L_j = \phi_0/2\pi I_c \cos(\theta)$, with the inductance inversely proportional to the critical current. If the phase θ undergoes a phase jump, then a voltage pulse will be generated with an integrated amplitude of $\int V(t)dt = \phi_0$. These pulses are referred to as single flux quantum (SFQ) pulses and have typical amplitudes of 1 mV and duration 2 ps.

In addition to the pair current, there is a current carried by thermal single-particle excitations, quasiparticles, that, at higher energies, are similar to single particle excitations in normal metals. The superconducting energy gap Δ_{sc} defines the minimum energy of a quasiparticle excitation, leading to an exponential decrease in the quasiparticle density at low temperatures and nonlinear conductances across certain junctions, especially those whose transport is dominated by tunneling. The current through the superconducting electrodes can be considered to be the sum of two independent components, the superconducting pair and quasiparticle currents, which constitutes a two-fluid model. In conventional tunnel barriers, the transport through the barrier can also be assumed to be due to independent pair and quasiparticle currents. This may not be the case in more complex barriers that have magnetic structures.

B. Josephson junction dynamics and models

A Josephson junction can be viewed as an ideal Josephson device, which transports superconducting pairs, in parallel with a capacitor and a dissipative element. This model, referred to as the resistively capacitively shunted junction (RCSJ) model,^{29,32} is shown in Fig. 2. The dissipative element can be either the quasiparticle current or a separate resistive shunt. Here, we take R_n to be independent of voltage, which adequately models the junctions used in this work.

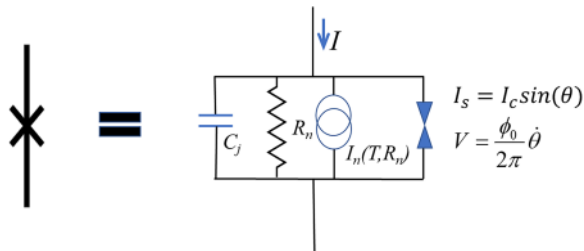


FIG. 2. Resistively capacitively shunted junction (RCSJ) model of a Josephson junction. An \times is used as the symbol for a Josephson junction.

The circuit model describes a resonant LCR circuit with a nonlinear inductor describing the supercurrents. Summing the currents and setting them equal to the applied current yields the RCSJ dynamical equation^{29,32}

$$\frac{\phi_0}{2\pi} C_j \ddot{\theta} + \frac{\phi_0}{2\pi R_n} \dot{\theta} + I_n(t) + I_c \sin(\theta) - I = 0, \quad (6)$$

where current noise $I_n(t)$ is a random variable usually assumed to have a Gaussian distribution with a standard deviation of $I_{nrms} = \sqrt{\frac{4k_B T}{R_n \tau}}$. Here, k_B is Boltzmann's constant, T is the temperature of the junction, and τ is the time over which the random field is applied, which must be much less than dynamical oscillation times.

The dynamics are characterized by two parameters, the plasma frequency, ω_p , defined as the oscillation frequency at the bottom of the potential well and the unitless McCumber parameter, β_c , which determines the strength of the damping.

$$\omega_p = \sqrt{\frac{2\pi I_c}{\phi_0 C_j}}, \quad (7)$$

$$\beta_c = \frac{2\pi I_c R_n^2 C_j}{\phi_0}. \quad (8)$$

Typical values of the plasma frequency for the junctions considered here are 100 to 400 GHz.³³ The McCumber parameter is related to the quality factor of the oscillator $Q = R_n C_j \omega_p = \sqrt{\beta_c}$ and can be chosen to vary over a wide range. Highly damped junctions, $\beta_c \ll 1$, are used for Josephson voltage standard arrays, while underdamped junctions, $\beta_c \gg 1$, have been used for latching logic. Here, we will restrict ourselves to junctions that are moderately damped that have β_c close to 1, which is close to critical damping.

The circuit model yields a dynamical equation that is identical to that of a forced damped pendulum, where the first term gives the kinetic energy with a mass proportional to C_j , the second and third terms are the damping term and associated fluctuations, respectively. The fourth term is the torque due to a “gravitational” potential with the same form as the Josephson energy $U = (-I_c \phi_0/2\pi) \cos(\theta)$. The final term is the external torque corresponding to the applied current.

The dynamics can be visualized using the damped pendulum analogy. When a small torque (or current) is applied, the order parameter phase θ will increase and reach a static value. When the phase gets close to π , there will be a probability that the pendulum will go over the potential energy maximum and will emit a SFQ pulse. When the applied current is larger than I_c , the torque is sufficient to continually drive the pendulum over the potential energy maximum outputting a series of SFQ pulses. At very high applied currents, the pendulum will undergo rapid oscillations with a rate of phase advance proportional to the applied current, corresponding to a linear increase in

the average voltage. Here, the dissipative current is much larger than the supercurrent. This trend is seen in Fig. 3, which is a simulation of the RCSJ equation with a current ramp of $3.0 \mu\text{A}/\text{ns}$, $T = 4.0 \text{ K}$, $I_c = 1.5 \mu\text{A}$, $\beta_c = 0.5$. The critical current is sufficiently low that the thermal energy is comparable to the Josephson energy leading to observed fluctuations.

Experimentally measured time-averaged voltages as a function of applied current, for arrays of Josephson junctions designed for SFQ applications, are shown in Fig. 4. Here, $I_c = 0.7 \text{ mA}$. The three regimes are labeled: the low-current dissipationless regime, the SFQ pulsing regime, and the high-current dissipative regime.

C. Superconducting transmission lines as axons

As important as efficient spiking and synaptic plasticity, a neuromorphic architecture must have the ability to transport spikes with little dispersion over long distances and large fan-out. Superconducting transmission lines have low loss and low dispersion up to frequencies comparable to the gap frequency $hf_g = 2\Delta_{sc}$, above which quasiparticle pairs can be created.³⁴ For Nb transmission lines, which are used in most of the current SFQ circuits, $f_g = 650 \text{ GHz}$.³⁵ 1 ps SFQ pulses can propagate 10 mm without significant loss and dispersion.³⁴

A superconducting transmission line can function as a long dispersionless axon. However, synaptic connections will cause attenuation of the pulse. Several circuits have been demonstrated using Josephson transmission lines [an example is shown in Fig. 5(b)] and SFQ pulse splitters that can be used to add energy and regenerate quantized SFQ pulses.³⁶ Using these techniques, a single spike can be sent to a very large number of synaptic taps.

D. Single flux quantum technology for neuromorphic computing

When operating in a low current regime near I_c , Josephson junctions are naturally spiking devices. These pulses form the basis of several high-speed logic families.³⁷

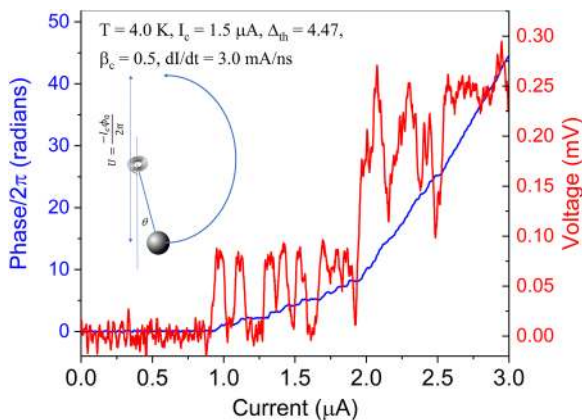


FIG. 3. Stochastic RCSJ simulation for a junction showing the phase and voltage as the current is ramped at $3 \mu\text{A}/\text{ns}$. The temperature is 4.0 K which, for this small critical current, gives rise to significant thermal fluctuations of the phase.

In addition, Josephson junctions have been demonstrated as biologically realistic neurons with two Josephson junctions per cell.^{38,39} They have been proposed for use in stochastic neural networks,⁴⁰ spiking neural networks, and have been used to implement a sigmoid transfer function for rate encoding.⁴¹ Figure 5 shows the correspondence between this superconducting technology and biological neural systems. Figure 5(a) shows a comparison of numerically calculated action potentials using a Hodgkin-Huxley model⁴² and SFQ pulses calculated using the RCSJ model. SFQ pulses are an order of magnitude smaller in voltage and nine orders of magnitude shorter duration. Both neural and SFQ pulses can propagate long distances with little dispersion. SFQ pulses can propagate on passive superconducting transmission lines or active Josephson transmission lines, shown in Fig. 5(b). Neurons also use a combination of regions with the myelinated regions being more passive and the nodes of Ranvier [Fig. 5(b)] corresponding to more active regions. Active lines, where there is energy input along the transmission path, are required for high fan-out to maintain the pulse energy. Biological neural systems have an inherent high-complexity 3-dimensional structure, as indicated schematically by a white matter tractography map for a human brain in Fig. 5(c), while the magnetic/superconducting systems are still limited by current microfabrication technology. Figure 5(c) shows state-of-the-art SFQ technology with 8-metal layers for Josephson junctions, high speed

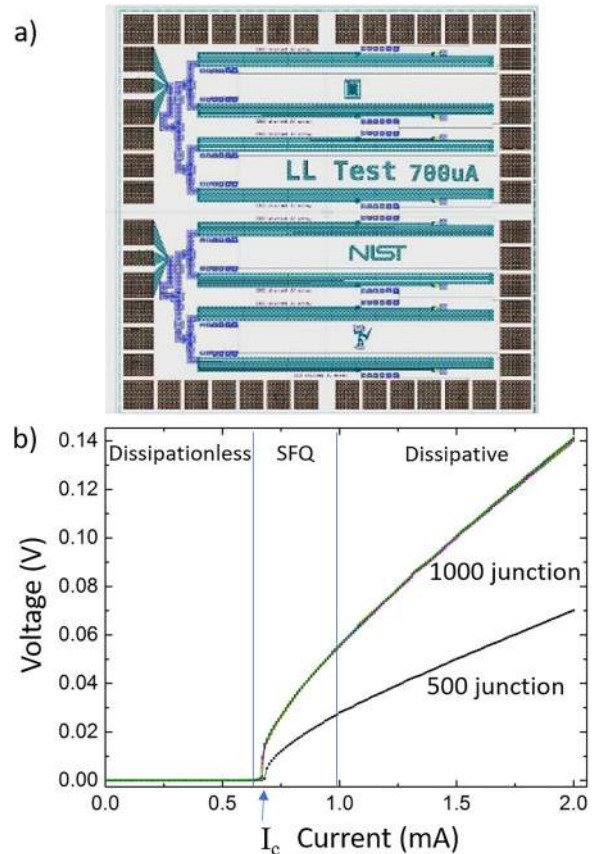


FIG. 4. Measured average voltage vs. applied current for Josephson junction arrays designed for SFQ applications. Figure courtesy of P. Dresselhaus and D. Olaya.

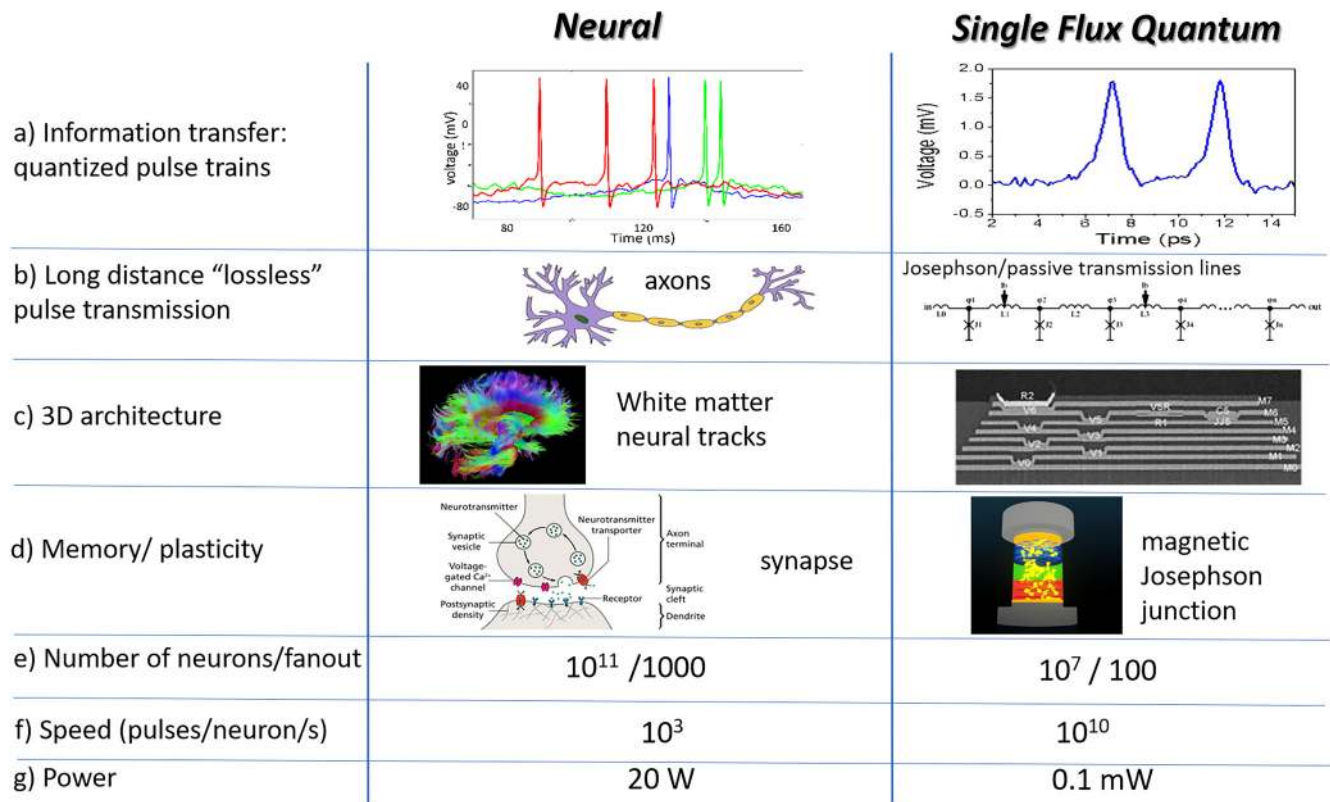


FIG. 5. Comparison of biological and magnetic Josephson junction neural systems. (a) Numerical simulations of action potentials using the Hodgkin–Huxley model and SFQ pulses using the RCSJ model. (b) Schematic of an axon which transmits action-potential pulses long distances without degradation and a Josephson transmission line which transmits SFQ pulses long distances without degradation. The axon schematic is from <https://en.wikipedia.org/wiki/Axon>. (c) A diffusion MRI-based tractography map of a human brain from www.humanconnectomeproject.org illustrating complex 3-dimensional connectivity and state-of-the-art 8-layer SFQ circuit showing present day connectivity in microfabricated circuits.²⁴ (d) Illustrates a biological synapse (<https://en.wikipedia.org/wiki/Synapse>) with complex neurotransmitters and ion channels, and a magnetic Josephson junction synapses with connectivity modulated by local spin clusters. [(e), (f), and (g)] Number of neurons and fan-out possible at present, speed of the neural systems measured in pulses per neuron per second, and the power consumed by a system with the above parameters. The power listed for a typical SFQ neuromorphic circuit does not include cooling overhead.

interconnects, resistive, capacitive, and inductive components.²⁴ While technically advanced, this microfabricated technology cannot match the fan-out possible in biological systems. In SPICE (SPICE refers to *Simulation Programs with Integrated Circuit Emphasis*, which is used to simulate many types of electronic devices and circuits) simulations, we have been able to implement fan-out of 1 to 3 without additional junctions. If we use this as a worst case, implementing all-to-all connections in a feed forward between two 100 node layers would require roughly 3900 additional junctions. This is a major challenge and any larger fan out is difficult without a further advance in the microfabrication technology. Fan-in of 9 to 1 has been simulated and would follow a similar approach, which is more promising. We have thus indicated 100 fan-out as the maximum feasible number with the current technology in Fig. 5(c), though even this level will clearly be challenging. The speed of Josephson based neurons is much faster than biological, with $\approx 10^{10}$ pulses/neuron/s possible. The neuromorphic synapse being discussed in this paper is shown schematically in Fig. 5(d) along with a biological synapse. The biological synapse is exceedingly complex relying on many different neurotransmitters, ion transport mechanisms, and intracellular structures. The neuromorphic synapse relies on the complex interaction of the superconducting order parameter with the

embedded magnetic nanostructure. The magnetic nanostructure forms a memory that can modulate the superconducting spiking behavior. The modulation depends on the relative orientations of the spin clusters, which can be modified using low energy current pulses. Finally, the human brain shows remarkable energy efficiency with operation at 20 W. The superconducting circuits take considerably less energy, excluding the costs for the cryogenic system. In large scale applications where the power consumption is already greater than 10 kW, the cooling overhead is roughly 1000 W of wall power to cool 1 W at 4 K.⁴³

Neuromorphic technologies have many potential advantages over corresponding digital technologies when going to extreme speeds and low energies. Digital technologies require global high-speed clocks for synchronization, which is challenging above 10 GHz, while neuromorphic systems are asynchronous. Digital SFQ systems require precise pulse timing and synchronization of pulse arrival at subsequent gates, while neuromorphic systems embrace pulse time variations as an important form of information transfer. Digital systems require high uniformity of devices, while neuromorphic systems, given that the devices are inherently plastic, embraces nonuniformity. Digital systems require a memory hierarchy with memory located at sequentially farther distances from the information processing core. This separation

leads to both higher power and lower speed. Neuromorphic systems, such as the one discussed here, have the memory intrinsically embedded in the information processing core, collocated at the nanoscale. Digital systems require deterministic response and cannot tolerate stochasticity, which necessarily occurs when switching energies are close to the thermal energy. Neuromorphic systems embrace stochasticity⁴⁴ as an important method of information transfer and as a method to explore options that have not been explicitly preprogrammed.

II. MODULATING COUPLING ACROSS A JOSEPHSON JUNCTION WITH THE NANOSCALE MAGNETIC STRUCTURE

A. Magnetic Josephson junctions

Superconductivity and ferromagnetism have a fundamental competition in their ground states, which makes the combination of the systems quite interesting. The singlet ground state in a superconductor has Cooper pairs with opposite spin orientation, whereas the ferromagnetic exchange interaction tends to align spins parallel. Vitaly Ginzburg in 1956 was the first to posit the coexistence problem of singlet superconductivity with ferromagnetism from the viewpoint of orbital interactions.⁴⁵ The advent of BCS theory²⁷ soon after led to the realization that the problem could also be explained in terms of a competition between the opposite spin orientation in superconducting pairs and the parallel spin orientation favored by the ferromagnetic exchange interaction.⁴⁶ The effect of the exchange field can lead to a nonuniform superconducting state with the pairs acquiring a non-zero center of mass momentum, which was predicted in 1964 by Fulde, Ferrell, Larkin, and Ovchinnikov (FFLO).^{47,48} While these theoretical works were done with respect to continuous ferromagnetic superconductors, the behavior that is observed in superconducting-ferromagnetic heterostructures is often analogous. This analogous behavior has led to many exciting recent developments in nanometer scale superconducting ferromagnetic hybrid devices.^{49–53}

Among the most promising devices for high performance computing applications are the magnetic Josephson junctions (MJJs). In these devices, the Cooper pair wavefunction extends into the ferromagnetic layer with a damped oscillatory behavior. Leveraging the work of the hard disk

drive industry, the magnetic random access memory industry, and the great progress in superconductive thin films, hybrid superconducting-ferromagnetic devices can be made with precise control of the ferromagnetic layer. This has led to the development of many interesting devices that exploit the physics of the interacting order parameters. Among these are π Josephson junctions,⁵⁴ spin triplet Josephson junctions,^{49,55} and pseudo spin-valve Josephson junctions.⁵⁶ What has become clear is that there is a strong interaction between the state of the ferromagnetic barrier in an MJJ and the superconducting order parameter. These effects can be used to modulate the phase and/or amplitude of the Josephson critical current I_c . More specifically, the exchange field F results in a superconducting spin-pair $\uparrow\downarrow$ that occupies a spin-split Fermi level in the ferromagnet. The resultant center of mass momentum of the spin-pair is nonzero $|k_\uparrow - k_\downarrow| \equiv \delta k$ and is directed perpendicular to the interface. In a typical superconductor, the ground state is occupied by Cooper pairs in the singlet state $\uparrow\downarrow - \downarrow\uparrow$. Once the spin splitting from the exchange field is introduced, the pair amplitude is proportional to $\uparrow\downarrow e^{i\delta k \cdot R} - \downarrow\uparrow e^{-i\delta k \cdot R}$ in analogy to the FFLO phase in a ferromagnetic superconductor. This leads to an oscillatory pair-phase $\varphi(x)$ that modulates with distance from the superconductor-ferromagnet interface that is also accompanied by decoherence as the pairs move away from the superconducting interface.^{47,48,51,57} This effect can be observed in the Josephson coupling, which leads to an I_c that oscillates and decays as a function of ferromagnetic thickness. The oscillation period with ferromagnetic thickness is set by the spin splitting at the Fermi level and is $\sim \pi/\delta k$.^{49,58–60} The ability to tune phase and the phase dependent Josephson critical current by changing the magnetic properties allows one to exploit the work that has gone into a large range of spintronics devices to control the superconducting properties of a hybrid MJJ.

B. Manipulating magnetic structure to perform synaptic functions

Significant work has gone into exploring ferromagnetic semiconductors for spintronics applications.^{61,62} One of the unintended effects that was discovered was the tendency for Mn to cluster during annealing when doping semiconductors.^{63,64} The size and energy barrier of the Mn clusters can be tuned by varying these annealing conditions leading to a highly tunable materials system.⁶⁵ This is particularly useful in Nb based Josephson junctions, which enables easy access to clusters that are thermally stable at operating temperature of 4 K. In addition, the clusters can be physically very small with cluster sizes being estimated from magnetometry to be around 4 nm in diameter and having a density of roughly 20 000 clusters per square micron. Being able to make MJJs with these large number of clusters allows us to control the critical current of the junction in a near analog way over two orders of magnitude. This flexible tunability is ideal for enacting the synaptic functions that are needed in SFQ neuro-morphic computing.³³

Figure 6 shows voltage versus applied current data from a $2.5 \times 5 \mu\text{m}$ elliptical MJJ with the SiMn cluster barrier. In

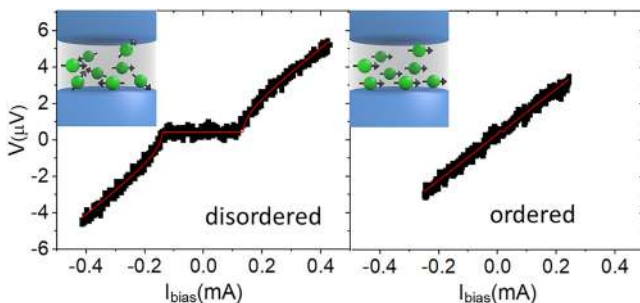


FIG. 6. Voltage versus applied bias current for $2.5 \times 5 \mu\text{m}$ elliptical clustered MJJ in the magnetically disordered (left) and magnetically ordered (right) states.

the magnetically disordered state, where there is dissipationless transport, the critical current is roughly $130\mu\text{A}$. In the magnetically ordered state, the critical current is near the noise limit of this measurement at roughly $10\mu\text{A}$ or less. The magnetic structure in these junctions can be toggled between ordered and disordered with $50 \pm 5\text{ fJ}$, $300 \pm 20\text{ ps}$ electrical pulses. In the presence of a 20 mT external magnetic field, this causes the clusters to order, and in the absence of an applied field, the pulse causes the clusters to disorder. If the field direction is reversed and the junction is pulsed with lower amplitude pulses, the MJJ will transition first to a higher I_c state and then back down to a lower I_c state, indicating that the ordered MJJ clusters point along the direction of the field.

The strong modulation of the Josephson critical current at zero field, seen in Fig. 6, suggests that the effect is due to the interacting order parameters. However, it is also known that the magnetically ordered state will produce a dipole field that can shift the maximum critical current away from zero applied field.⁶⁶ To confirm that the observed change in I_c is not due to the dipolar field effect, we can measure I_c as a function of the applied magnetic field. Figure 7 shows the characteristic voltage $I_c R_n$ for a $10\mu\text{m}$ circular junction in the magnetically ordered (blue) and disordered states (red). The classic Fraunhofer pattern²⁹ is observed for the magnetically disordered state and is partially fit by the Airy function. The discrepancy of the Airy function fit on the right-hand lobe illustrates the complexity of the interaction of the superconducting order parameter with the magnetic clusters; the details of which are currently under investigation. In the magnetically ordered state (red; Fig. 7), we did not fit the data to the Airy function because I_c was at the noise level of the measurement and no clear pattern was observed for I_c as a function of applied field.

Figure 8 shows the Josephson I_c as a function of a number of electrical pulses used to order the Josephson

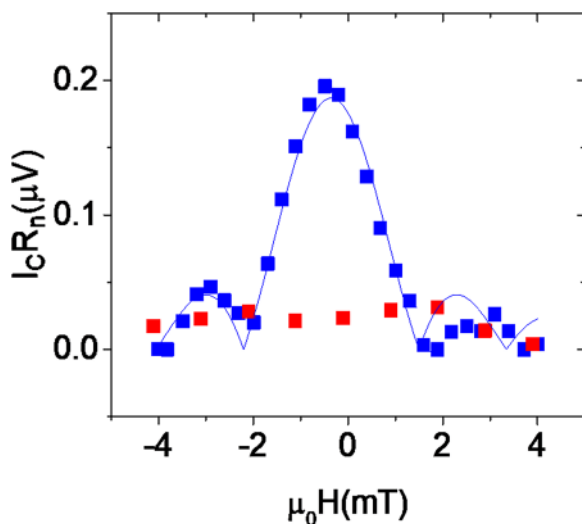


FIG. 7. Josephson I_c as a function of applied field for a $10\mu\text{m}$ circular clustered MJJ. Blue data are taken in the magnetically disordered state and fit to the Airy function, and red data are in the magnetically ordered state and were at the instrument resolution and therefore not fit to the Airy function.

junction.³³ The data demonstrate the ability to tune the order in a quasi-analog way between the fully magnetically ordered and disordered states, which is critical for synaptic functionality. The ability to reset the clusters into the disordered state is easily achieved by pulsing the MJJ in the absence of an external magnetic field. These data were taken on a $10\mu\text{m}$ circular clustered MJJ. The electrical pulses were applied in a 20 mT external field. The measurements of I_c were taken in zero applied field and it was confirmed that sweeping the field between zero and 20 mT had no effect on I_c without the additional electrical pulses. The pulse voltage was chosen to be a factor of 10 less than the voltage required to fully order the clusters with a single pulse. This indicates that the amount of order induced with the electrical pulse can be modulated by changing the pulse amplitude. This is an important tunability that can be used in the design of self-training synaptic circuits.

Figure 9 shows the energy scaling well with the device size.³³ The red squares are the energy in the electrical pulse that was required to fully order the junction in an applied 20 mT external magnetic field. It was confirmed that the magnetic field alone was sufficiently small that it did not affect the magnetic order without the addition of an electrical pulse. It is worth noting that pulses many times below this energy also affected the magnetic order, as seen in Fig. 8, and allows for the near analog control of the magnetic order. It is also worth noting that the minimum training spike voltage was roughly $700\mu\text{V}$. The blue circles are the approximate spiking energy for the MJJ as a function of size. The energy value is calculated as $I_c \Phi_0$ and is close to the energy dissipated during a spiking event in the MJJ synapse during inference. If the observed scaling trends continue, then junctions with approximately $1\mu\text{m}^2$ area will be in the self-training regime. In this limit, the magnetic order can interact directly with the SFQ pulses and self-learning circuits based on Hebbian learning rules can be designed. If the size scaling fails to continue, we should also be able to reduce the training energy required by reducing the Mn cluster size. This can likely be achieved with a lower annealing temperature.

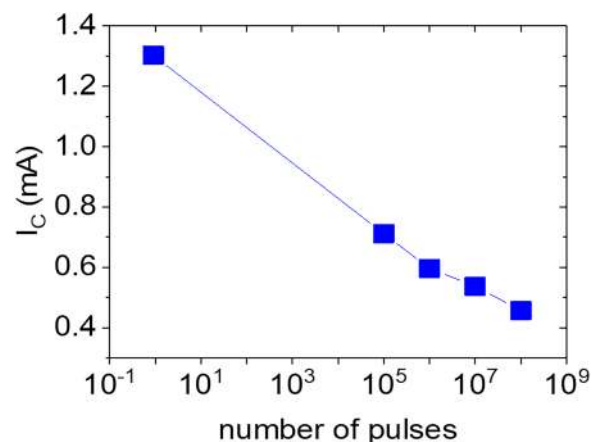


FIG. 8. Josephson critical current vs. the number of applied electrical pulses demonstrating the ability to tune order in small increments for synaptic functionality.

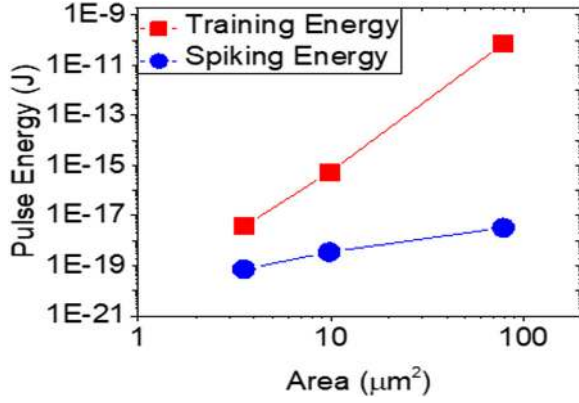


FIG. 9. Pulse energy scaling as a function of the MJJ device area. The red squares represent the electrical pulse energy required to fully order the magnetic state in a 20 mT applied external field. The blue circles are the approximate energy dissipated during a 2π spiking event of the Josephson junction and are calculated as $I_c \Phi_0$.

III. MODELING NEUROMORPHIC SPIKING OF JOSEPHSON JUNCTIONS AND MAGNETIC JOSEPHSON JUNCTIONS

In this section, we present simulations of spiking Josephson junctions based on the RCSJ models. Sections III A and III B present stochastic models of a single Josephson junction driven by thermal noise. Section III C extends the RCSJ model to include a magnetic degree of freedom that can serve as a synaptic weight.

A. Modeling stochastic dynamics of Josephson junction

When the Josephson energy becomes comparable to $k_B T$, fluctuations become important. The thermal stability factor is given by

$$\Delta_{th} = \frac{I_c \phi_0}{2\pi k_B T}. \quad (9)$$

Here, k_B is the Boltzmann constant and T is the temperature in kelvin. When Δ_{th} is large, the dynamics are deterministic, whereas when $\Delta_{th} \lesssim 6$, there is a significant stochastic component. Given the large nonlinear variation of I_c , with temperature near the superconducting critical temperature T_c , we can vary Δ_{th} over a wide range by adjusting the temperature over just a few degrees, thereby controlling the amount of stochasticity in the neural circuit.

Figure 10 shows a numerical simulation of the RCSJ equation for junctions with different critical currents and thermal stability factors showing a transition from small to extensive stochastic spiking as the critical current is lowered from $1.0 \mu\text{A}$ to $0.5 \mu\text{A}$. Here, a small external current bias, $I_b = 0.3 \mu\text{A}$, is applied, which sets the polarity of the spiking and provides energy input. The energy input is given by the product of I_b and the time-integrated voltage, which is then $n I_b \phi_0$, where n is the number of 2π phase slips that occur. The external energy required to generate a spike is $11 k_B T$. This low spiking energy is an important advantage to operating in the stochastic regime. A second advantage is the exponential dependence of spiking rate on device parameters. By

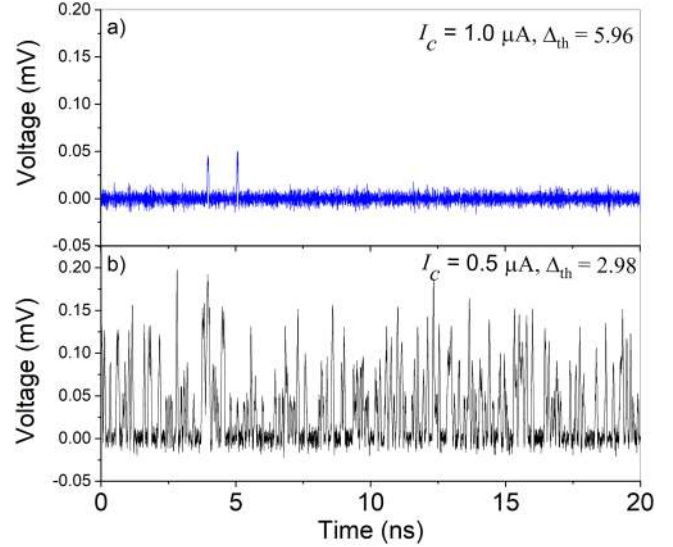


FIG. 10. Simulations of spiking of a $0.5 \mu\text{m}$ diameter 400Ω junction at 4.2 K for critical currents of (a) $1.0 \mu\text{A}$ and (b) $0.5 \mu\text{A}$ and a bias current of $0.3 \mu\text{A}$. The junction with $I_c = 0.5 \mu\text{A}$ shows spikes corresponding to 1, 2, and 3 flux quanta.

dynamically changing the critical current by a factor of 2, the physics of which was described in Sec. II, a large increase in spiking can be obtained. Without any external energy input in the stochastic limit, the system will undergo random positive and negative spikes as energy is transferred in and out of the thermal bath, which is an inherent component of the dissipative element.

B. Modeling spike timing effects

All neuromorphic systems have spike timing effects due to their intrinsic dynamics. For SFQ systems, the dynamics are based on ω_p and β_c , as well as on stochastic effects caused by temperature. Figure 11 shows an RCSJ simulation for a series of current-pulse-pairs input into a $1 \mu\text{m}$, 200Ω junction. The current-pulse pairs have temporal separations ranging from 100 ps to 0 ps and are each 5 ps in duration. For pulse separation ≥ 50 ps, no SFQ pulses are emitted. As the pulse separation is reduced below 20 ps, increasing number of SFQ pulses are emitted, showing a strong spike timing dependence. While there is a transient response for pulse separations of 50 ps and 100 ps, the output voltage pulses integrate precisely to zero. The integrated amplitudes of all output voltage pulses are quantized in units of ϕ_0 and range from $0 \phi_0$ to $8 \phi_0$. This spike timing dependence is due to the energy relaxation time of the junction, which is dependent on β_c . As seen in the inset to Fig. 11(c), the relaxation time, for this junction, is on the order of 20 ps. The oscillations in the junction, which occur at the plasma frequency ω_p , will lead to fine-scale structure in the spiking probabilities on short time scales if $\beta_c > 1$.

C. Magnetic Josephson junction model

In addition to the standard RCSJ model that is available as open source in WRspice, we have developed a Verilog model of our magnetic Josephson junctions and integrated it

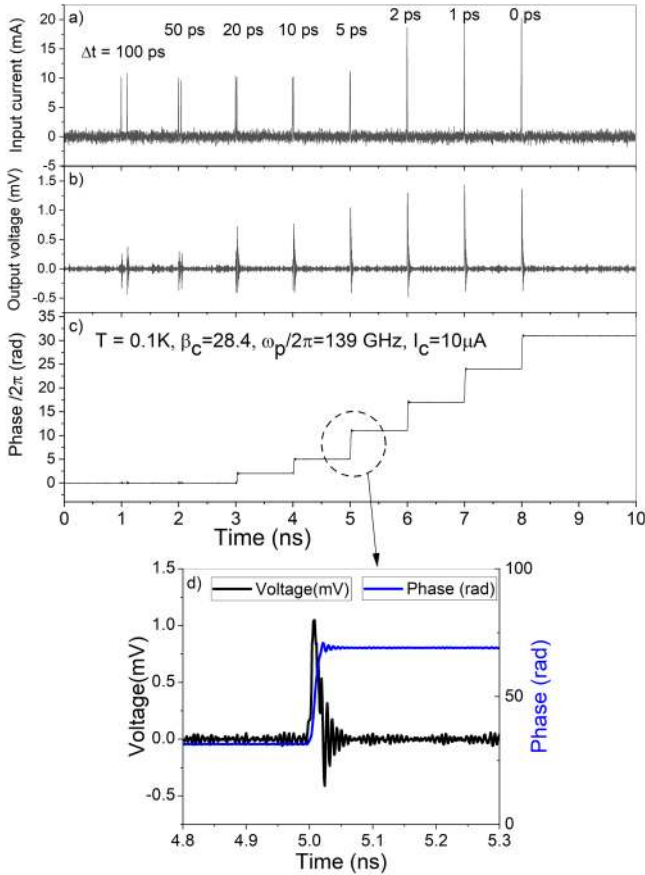


FIG. 11. Simulations of spike timing effects. A pair of 5 ps wide 10 μA amplitude Gaussian input pulses are applied to a 1 μm diameter 200 Ω junction with different pulse separations ranging from 100 ps to 0 ps. (a) Input current spikes, (b) output voltage spikes, (c) phase progression with the left axis being the number of 2π phase jumps. (d) A close up of a voltage spike and phase jump where the inherent dynamics corresponding to a damped oscillator can be seen.

into WRspice. In this model, the magnetic order parameter m is expressed by its effect on the junction critical current (I_C), according to the relationship

$$I_C(m, T) = [(1 - m)I_{CV} + I_{CM}] \left[1 - \left(\frac{T}{T_C} \right)^2 \right], \quad (10)$$

where T is the temperature, T_C is the critical temperature of the superconducting material, m is the magnetic order parameter (the normalized total magnetic moment in the barrier), and the sum $I_{CV} + I_{CM}$ is the maximum critical current, achieved when $m=0$ (the magnetically disordered state) and $T=0$. Equation (10) is a simple phenomenological model which coarsely mimics the observed behavior that the I_C changes with each voltage pulse and that I_C varies with the observed power law behavior $1 - (T/T_C)^2$. The magnetic order parameter varies with the integrated junction voltage according to $dm = kV(t)dt$ between $m=0$ and $m=1$; in other words, a voltage pulse across the junction causes the order parameter to increase and critical current to decrease, until saturation at $m=1$ and $I_C = I_{CM}$. The proportionality constant k is a simulation input parameter; physically, this parameter varies based on fabrication parameters and external field

strength. To simulate the neural network *in situ* self-training mode, k can be set so that m varies from 0 to 1 within ~ 100 SFQ voltage pulses. For simulation of network operation after MJJ, I_C 's have converged to their optimal values and k is set so that m is essentially invariant in normal circuit operation.

IV. NEUROMORPHIC CIRCUIT MODELS INCORPORATING JOSEPHSON JUNCTIONS AND MAGNETIC JOSEPHSON JUNCTIONS

Josephson junctions and MJJs can be built into simple superconducting circuit elements that emulate the synapses and neurons of a neural network. A simple 3-layer feed forward network is shown schematically in Fig. 12, with neurons and synapses labeled n and s , respectively. These elements can be used to realize the layers of a standard feed-forward neural network and perform, in hardware, the computation

$$\vec{y} = f(W\vec{x} + \vec{b}), \quad (11)$$

where \vec{x} and \vec{y} are the vectors of network layer inputs and outputs, W is a matrix of weight factors, \vec{b} is a vector of offset bias values, and $f(\vec{z})$ is a nonlinear function, applied element-wise to each element of \vec{z} , which is the weighted sum of the inputs and biases. W and \vec{b} are variables and these are modified, in either a supervised or unsupervised manner, to train the circuit for a particular function.

For a software based neural net, \vec{x} , \vec{y} , W , and \vec{b} would consist of integers or floating-point numbers. For a neural net implemented in analog, silicon-based circuitry \vec{x} , \vec{y} , and \vec{b} would be continuous voltages and W would be linear amplification stages. For neuromorphic spiking systems, being discussed here, \vec{x} and \vec{y} represent complex functions of the input and output current spike trains (e.g., the number of spikes or spike rate), W characterizes the complex processing of the incoming spikes by the Josephson circuit elements, and \vec{b} are external bias currents. Hence, in our discussion, we refer to “synapses” as the circuit elements that implement the weight matrix elements w_{ij} and “neurons” as the circuit elements that perform the summation of multiple synaptic inputs, add a bias, apply a nonlinear function, and produce an output spike train.

A. Synapses

Synapses are represented numerically by the elements w_{ij} of a matrix of weights W . In a superconducting neuromorphic circuit with tunable I_C 's, the synapse circuit elements are designed such that the I_C of the (i,j) th synaptic MJJ maps to the weight element w_{ij} . Ideally, the mapping from I_C to weight w_{ij} would be unique and monotonic. In addition, due to the high volume of synapses in a feed-forward network, a simple synapse architecture with small circuit area is ideal.

A proposed synapse circuit structure is shown in Fig. 13. This structure uses a single MJJ with tunable I_C , in parallel with a fixed inductor L_{fix} . In a small-signal linear model, the Josephson junction is modeled as an inductor L_J with inductance $L_J \approx \Phi_0/2\pi I_C$, in parallel with a resistance R_n . For pulses much slower than the plasma frequency of the MJJ

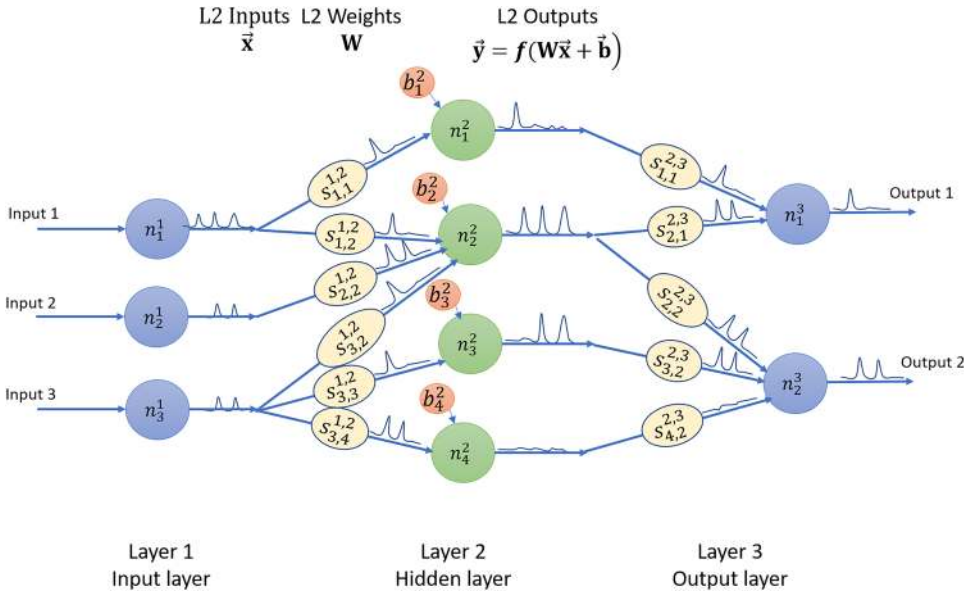


FIG. 12. Schematic of a simple spiking feed forward network. n , s , and b represent neurons, synapses, and biases, respectively. Not all connections are shown.

(100 GHz as demonstrated³³), the input current splits between the two inductive branches in a ratio that depends on the MJJ I_C . The current through L_{fix} is used as the synapse output signal. Given an input current I_{in} to this synapse structure, the current through L_{fix} is given by

$$I_{L_{fix}} = w_{ij}I_{in}, \quad (12)$$

where

$$w_{ij} = \frac{L_J}{L_J + L_{fix}}.$$

A SPICE simulation of the I_C versus w_{ij} relationship is shown in Fig. 14. This simulation uses circuit parameters $L_{fix} = 4$ pH, $R_n = 1 \Omega$, and I_C varying from $20 \mu\text{A}$ to $200 \mu\text{A}$ (corresponding to junction characteristic frequency varying from 9.6 GHz to 96 GHz). In this simulation, the input current I_{in} is ramped from 0 to $150 \mu\text{A}$, and back to $0 \mu\text{A}$, over a range of ramp times. The synapse weight w_{ij} is calculated by the ratio of the peak current through L_{ext} to the peak input current of $150 \mu\text{A}$. The solid curve corresponds to the slowly varying input case (ramp time >150 ps). The value of

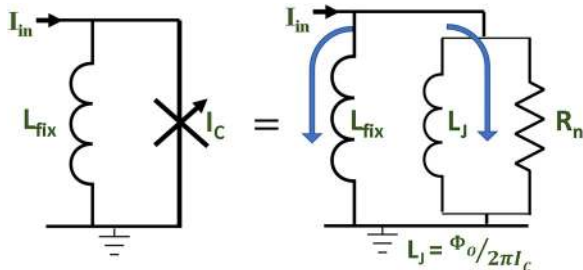


FIG. 13. Synapse circuit element consisting of a Josephson junction in parallel with a fixed inductor, L_{fix} , which is inductively coupled to the output. The symbol for a MJJ is similar to that of a Josephson junction, except there is an additional arrow to indicate that the junction properties can be varied during circuit operation.

w_{ij} varies from 0.95 to 0.3 over the range of I_C 's simulated. To achieve smaller magnitude of weights w_{ij} , the maximum MJJ I_C could be further increased. For example, increasing I_C to 2 mA would decrease the synaptic weight factor to ~ 0.04 in the slowly varying input case. The fixed inductance L_{fix} could also be increased to achieve smaller w_{ij} , although this would result in larger circuit areas (with current NIST fabrication technology, $L_{fix} = 4$ pH corresponds to a synapse circuit area of $\sim 10 \mu\text{m} \times 15 \mu\text{m}$). Alternatively, each synaptic weight w_{ij} could be physically represented by two synapse structures with independent MJJ I_C 's, one with a positive contribution and one with a negative contribution so that $w_{ij}(net) = w_{ij}^+ - w_{ij}^-$. This scheme will be further explained in the discussion of output neurons.

The mapping between I_C and w_{ij} deviates from the outlined model when either the input current changes too quickly or the magnitude of input current becomes too high. In particular, as the inverse of the input signal ramp time approaches the junction characteristic frequency, the junction

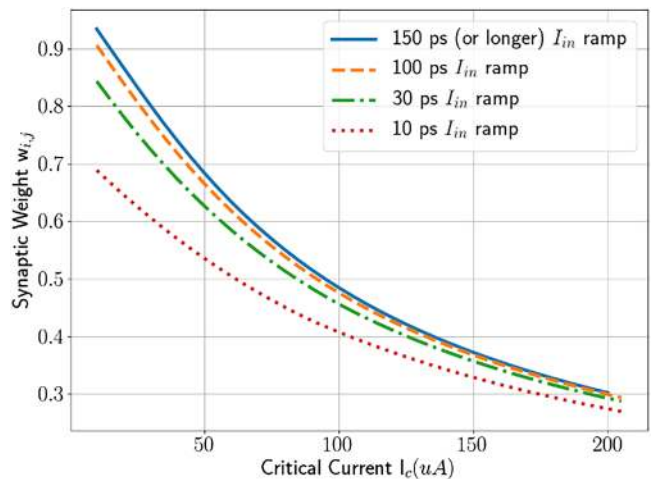


FIG. 14. SPICE simulation of synaptic weight vs. Josephson junction critical current for varying ramp times.

resistance R_n shunts an increasing portion of the input current, lowering the current through L_{fix} and limiting the range of possible variation in w_{ij} . This effect is shown by the dashed curves in Fig. 14. For the junction parameters simulated in this example, the range of possible w_{ij} begins to drop significantly for input ramp times <30 ps. To achieve a higher maximum operating speed while maintaining full sensitivity of w_{ij} to I_C , the MJJ R_n could be increased. Finally, when the input current magnitude increases such that $I_{in}L_{fix} \sim 0.5 \Phi_0$ or larger, the small-signal model for the Josephson inductance L_J is no longer valid. Instead, the MJJ outputs an SFQ pulse for sufficiently large I_{in} , causing an abrupt change in the effective w_{ij} . The proposed synapses are not intended for operation in this large-signal regime.

The input current signals to the synapses could be either analog (i.e., a current pulse from an external source with programmable amplitude or the scaled output from a previous network layer) or digital and binary (a “1” or “0” represented by the current, or lack thereof, from an SFQ pulse). An advantage of SFQ pulses is that the input bits could be stored prior to operation in an on-chip SFQ memory bank, and then clocked at speed to the network input with a single universal clock signal. For long input vectors \vec{x} and networks that require synchronous operation, this operation scheme is more realistic than separately driving each input channel in real time.

B. Output neurons

1. Overview

We envision using standard Josephson junctions, e.g., Nb/AIOx/Nb, to provide the functionality of neurons. These will serve as the spiking elements and have already been well developed for digital SFQ circuits. As previously described, the role of output neuron j is to sum the synaptic inputs and apply a nonlinear function

$$I_{out,j} = f\left(\sum_{i=0}^n I_{in,i}w_{ij} + b_j\right). \quad (13)$$

One proposed circuit structure that implements this functionality is shown in Fig. 15. The current through L_{fix} of each synapse is inductively coupled to an output loop that consists of a large inductor in series with two Josephson junctions. This output circuit structure is a superconducting quantum interference device (SQUID).^{29,67} The SQUID is characterized by its total inductance L_{SQ} , the coupled magnetic flux through the SQUID loop Φ_{cpl} , the DC bias current I_{bias} , and

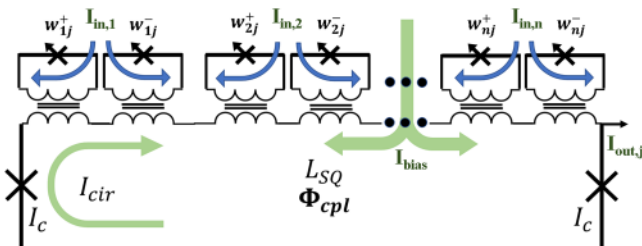


FIG. 15. Schematic of multiple synapses inductively coupled to the “ j th” output neuron.

the I_C ’s of the component Josephson junctions (which could be fixed or tunable if MJJs are used). Here, it is assumed that Φ_{cpl} arises from current through the inductively coupled synaptic inputs.

The total magnetic flux through the SQUID ($\Phi_{tot} = \Phi_{cpl} + L_{SQ}I_{cir}$) is proportional to the difference in the gauge-invariant phase φ between the two Josephson junctions; $\Phi_{tot} = \Phi_0(\varphi_1 - \varphi_2)/2\pi$. A further constraint on φ_1 and φ_2 is that the total current through the Josephson junctions must sum to I_{bias} in the absence of any coupled flux. Whenever I_{bias} is not split evenly between the two Josephson junctions, a nonzero I_{cir} is present. Depending on all SQUID parameters and input values, a steady-state solution for the distribution of I_{bias} between the two Josephson junctions may exist (meaning that $d\varphi/dt = 0$ for both Josephson junctions), and the SQUID remains in the “zero-voltage state;” if a solution does not exist and $d\varphi/dt \neq 0$, the SQUID is in the “voltage state.”

In the zero-voltage state, as the coupled flux Φ_{cpl} increases, the current through each Josephson junction (and value of I_{cir}) varies smoothly until the current through one of the Josephson junctions reaches its I_C threshold. At this point, the Josephson junction at threshold outputs an SFQ pulse, the circulating current I_{cir} increments or decrements by a discrete value, and the SQUID again evolves smoothly along a new solution space as Φ_{cpl} increases further. If Φ_{cpl} subsequently decreases to zero or its original value, the second Josephson junction outputs an SFQ pulse, returning I_{cir} to its original value.

Figure 16 shows the results of a SPICE simulation of a SQUID while ramping the coupled flux. One of the SQUID Josephson junctions outputs an SFQ pulse when Φ_{cpl} increases above a threshold, and the second outputs an SFQ pulse as Φ_{cpl} decreases. In the proposed neuromorphic SFQ circuit architecture, Φ_{cpl} is provided by the summation of the coupled flux from all input synapses; $\Phi_{cpl,j} = \sum_i MI_{L_{fix},ij}$, where M is the mutual inductance between the synapses and

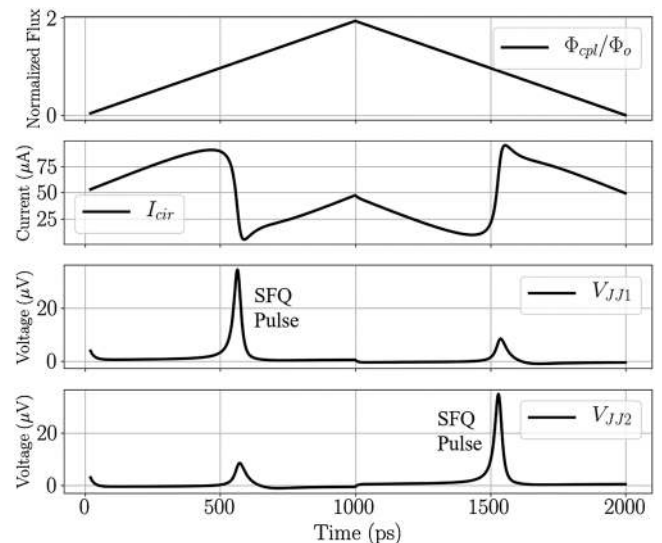


FIG. 16. SPICE simulations of a SQUID with coupled flux ramp. When the coupled flux reaches a certain threshold, one of the SQUID Josephson junctions outputs an SFQ pulse and the circulating current jumps in magnitude.

neuron and $I_{L_{fix},ij} = w_{ij}I_{in,i}$. As shown in Fig. 15, each synapse is physically represented by two loops that couple flux of opposite polarity to the output SQUID, corresponding to the weight factors w_{ij}^+ and w_{ij}^- .

The inductively coupled scheme for transmitting synapse signals to the output neuron was chosen to minimize unwanted cross-talk between circuit structures. The Josephson junctions are two-terminal devices with low output impedance, which makes it difficult to achieve isolation between inputs and output, especially when designing directly coupled circuits with large fan-in. For example, consider N MJJ synapses that are directly connected to the bias input of a neuron SQUID loop. An SFQ voltage pulse from one synapse could split and travel back through the other $N - 1$ synapses rather than forward to the output neuron. Additional circuitry can be added to block this back-propagation,³⁷ but our simulations of directly coupled neuromorphic circuits still reveal limited scalability and poor circuit operating margins, which is why we propose an inductively coupled scheme.

Cross-talk between synapses in the inductively coupled scheme only results from second order effects, as long as the parasitic mutual inductance between nearby synapses is kept small. The second-order effect arises because incremental addition to Φ_{cpl} by one synapse slightly changes the output neuron I_{cir} , which in turn changes the amount of magnetic flux back-coupled from the output neuron to the other synapses. The back-coupled flux through the synapse loops slightly changes the current through L_{fix} and effective synapse weight w_{ij} . However, these incremental changes are negligible and even the large-signal flux back-coupled to the synapses from the total I_{cir} in the neuron is small. For example, assuming a synapse L_{fix} of 4 pH, with each synapse coupled to a ~ 1 pH section of the output neuron L_{SQ} with coupling coefficient ~ 0.3 , and an I_{cir} magnitude of $80 \mu\text{A}$ or less, the magnetic flux back-coupled to the synapses is at most $0.02\Phi_0$. Simulations show that $\Phi_{cpl} = 0.02\Phi_0$ through the synapses has negligible impact on w_{ij} . As network fan-in grows, the neuron L_{SQ} will increase to accommodate the larger number of synapse inputs, while the neuron I_{cir} 's and I_{cir} will decrease. Therefore, unlike the directly coupled

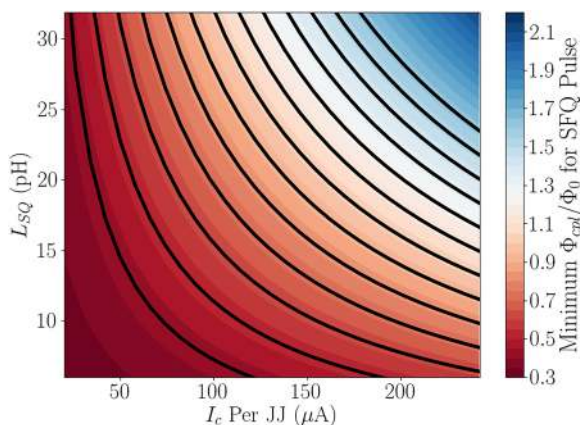


FIG. 17. Results of SPICE simulations showing coupled flux Φ_{cpl}^T/Φ_0 to create a neuron SQUID output pulse as a function of I_C and L_{SQ} . Lines of constant $I_C L_{SQ}$ are shown.

architecture, unwanted cross-talk between circuit structures will decrease with larger fan-in for the inductively coupled architecture.

2. Neuron nonlinearity implementation

To fully implement the feed-forward neural network layer model in Eq. (11), the output SQUID neuron should apply a nonlinear function $f(z)$ to the summation of coupled synapse fluxes.

In a binary-output neuron model, the nonlinear function $f(z)$ is simply a threshold function. The SQUID Josephson junctions each output an SFQ pulse if the value of Φ_{cpl} exceeds a threshold value and do not output an SFQ pulse if Φ_{cpl} does not reach the threshold. This threshold (Φ_{cpl}^T/Φ_0) is tunable based on the parameters of the output SQUID, including the I_C and L_{SQ} . The value of Φ_{cpl}^T/Φ_0 is shown in Fig. 17 as a function of I_C and L_{SQ} ; this figure shows that Φ_{cpl}^T/Φ_0 mainly depends on the product $I_C L_{SQ}$. Figure 17 is generated from a SPICE simulation in which I_{bias} was initially $0.5I_C$ for each Josephson junction in the SQUID. The effective neural network offset bias b_j of neuron j can thus be set by tuning the $I_C L_{SQ}$ value, which will tune the amount of coupled Φ_{cpl} needed to trigger a binary “1” output. The binary-output neuron scheme is attractive from an energy and speed standpoint, because the output from each neuron is encoded in a single pulse with sub-attojoule energy and only ~ 10 s of ps duration. The tradeoff is that more information can be encoded per neuron if analog output signals are used.

In one scheme to achieve analog neuron output signals, the SQUID bias current I_{bias} is set at a sufficiently high value that when Φ_{cpl} exceeds a threshold Φ_{cpl}^T/Φ_0 , there is no possible zero-voltage SQUID configuration, and the SQUID enters the voltage state. The time-average voltage (measured across either Josephson junctions) depends on Φ_{cpl} , as shown in Fig. 18. For this figure, SQUID output voltage was averaged over 1 ns. Although the time-average voltage does not depend monotonically on Φ_{cpl} , there is a regime of Φ_{cpl} in which the SQUID behaves approximately as a Rectified Linear Unit (ReLU) neuron, with zero output up to an input threshold Φ_{cpl}^T/Φ_0 and steadily increasing output above this threshold.

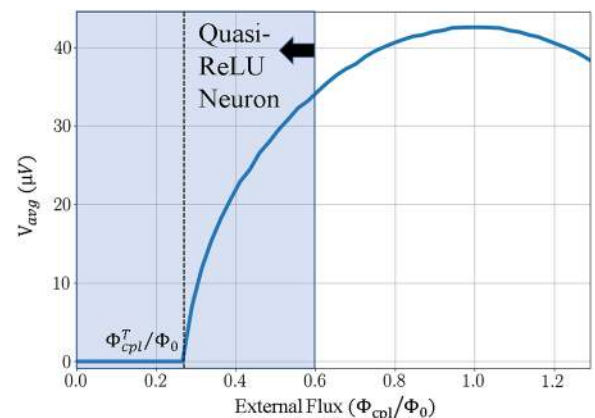


FIG. 18. Simulation of a neuron SQUID operated as an analog-output neuron, with 0 output voltage up to a threshold Φ_{cpl}^T/Φ_0 and subsequent monotonic increase in V_{avg} with Φ_{cpl} , within a limited range of Φ_{cpl} .

Unlike a software-defined ReLU neuron, the output eventually saturates and then decreases with increasing input.

3. Trainability

The I_C 's of the MJJs can be tuned in either direction depending on the presence or absence of a small magnetic field ~ 1 mT. The magnetic field can be generated either globally or locally by a write line. In the absence of any other stimulus, the magnetic clusters in the MJJs remain stable in this small magnetic field and do not change. However, any of the MJJs that receive a voltage spike with the applied magnetic field will increase in magnetic order. This reduces the I_C value of these MJJs. In the absence of an applied magnetic field, a voltage spike will reduce the magnetic order of any MJJ that receives one of these training spikes. The details of the implementation of such training methods are an area of active investigation.

In one example of a simple training algorithm for a supervised learning task, consider a training set of binary-valued input vectors \vec{x} and their corresponding binary-valued known outputs \vec{y} . For each training example, the activated neurons in output vector \vec{y}_m are biased so that their circulating current, I_{cir} , couples significant magnetic flux through the input synapses ($\sim 0.3\Phi_0$), but does not induce an SFQ pulse from the synapse MJJs. The non-activated output neurons are biased so that I_{cir} is negative, as is the magnetic flux coupled to the input synapses. A positive I_{in} is then applied to the activated inputs of \vec{x}_m . When bias and input amplitudes are chosen correctly, this scheme will cause an SFQ pulse only from the MJJs in w_{ij}^+ synapses that connect an activated input to an activated output, or in w_{ij}^- synapses that connect an activated input to a non-activated output. A schematic of this supervised learning scheme is shown for a simple 3-to-2 network in Fig. 19. It is desirable to be able to supply a localized magnetic field to target MJJ synapses, which can be achieved with a dedicated field write line. It should be noted that in this case, we would need to reduce the required field for ordering.

V. TOWARD LARGE SCALE NEUROMORPHIC ARCHITECTURES

The combination of clustered MJJs that can mimic the functionality of the synapse and normal Josephson junctions that can mimic the functionality of a neuron allows for a powerful device set for implementing neuromorphic circuits. Neuromorphic systems based on this hardware platform have clear advantages in speed and power consumption compared to other systems. In addition, the substantial amount of work that has been done for digital superconducting circuit design and fabrication can also be leveraged to greatly accelerate the time from proof of concept to functioning circuits. The previous modeling section has illustrated the ability to model the functionality of the various circuit elements that will be needed to construct a larger scale neuromorphic system. The fact that the Josephson junction models used in the SPICE simulations have already been proven to be accurate for digital circuit design is a substantial benefit.

Building a truly neuromorphic system that can take advantage of the spike time dependent dynamics and the plasticity demonstrated by the clustered MJJs is the ultimate goal. However, the demonstration of functional circuits composed of these devices is the first step in this direction. We can leverage the already existing simulation tools to design various circuit architectures much more rapidly than testing the circuit design only after fabrication. Given that we have circuit elements that mimic the neuron and synapse, we should be able to make circuits that perform the general artificial neural network functionality of Eq. (11). This functionality is at the heart of software based neural networks that are now the best in class for applications such as image recognition,¹ internet search,⁶⁸ and are starting to show great promise for medical diagnostics.⁶⁹ We have shown that we can implement a feed forward neural network directly with these devices with an efficiently compact design in SPICE simulations.⁷⁰

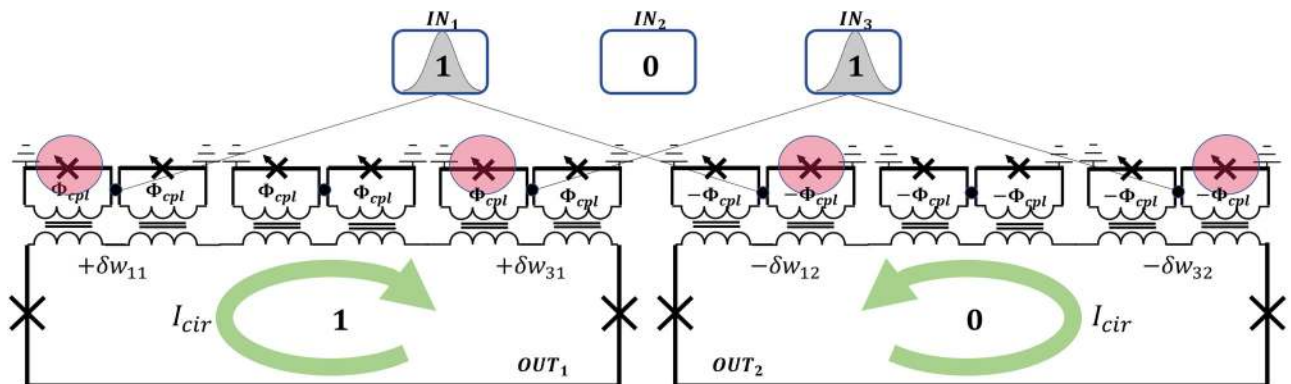


FIG. 19. Proposed on a chip supervised learning scheme for MJJ neural network. For each training example, input pulses shown as 1s on top of figure are applied to the activated input neurons, which will produce currents in the MJJs in the synapses connected to the active inputs. In addition, a positive training current, shown as clockwise green arrow, is applied to the desired active output neuron, labeled as 1, and negative training current is applied to the desired non-activated output neurons, labeled as 0. These applied currents at the outputs will induce currents in the MJJs of the connected synapses. The sum of the active input and training currents will add together and exceed the MJJ cluster ordering threshold and, in the presence of an applied field, result in a strengthening of the weight of the connected excitatory part of the synapse. The sum of the active input and training current will tend to cancel each other in the MJJ of the inhibitory part of the synapse. Similarly, the current in an active input and non-active output neuron will add in the inhibitory part of the synapse (strengthening the inhibitory weight) and will tend to cancel each other in the excitatory part of the connected synapses. The MJJs that will have their weight strengthened are circled in red.

In the example network shown in Fig. 20, we are able to implement a 9 pixel 3 class feed forward network.⁷¹ We trained this network on the example problem previously demonstrated with a memristor implementation that was used to recognize three distinct letters (z, v, or n) with any one pixel of input noise.¹² Since there is no separate training and test data in the example, 100% accuracy is easily achievable. To test our implementation, we first trained a standard software neural network in Python using back propagation techniques.⁷² We were then able to use a linear mapping of the weight values in the Python network to the I_c values of the MJJs in our SPICE simulation. With this mapping, we were able to achieve 100% accuracy on the full dataset of noisy 9 pixel letters. Furthermore, because the propagation speed of SFQ pulses is roughly 1/3 the speed of light in vacuum, the network can respond very quickly. In our unoptimized simulations, we could input a new image vector every 3 ns.⁷⁰

Because of the structure used, the output neurons of the model can be used as inputs for a subsequent layer. In general, we can directly implement conventional deep learning architectures in our hardware platform. The key problem of mapping software neural networks into this hardware will be the large fan-in and fan-out that is often required. However, the structures are naturally separated into layers in a feed forward architecture, which greatly simplifies the issue of connectivity. The fan-out in these circuits can be repeated in architectures where the neurons have a fixed spike amplitude. We have simulated fan-out of 1 to 3, which is typical even in digital circuits. As discussed above, nesting such a fan-out would require a substantial overhead. We have simulated fan-in of 9 to 1, which is quite promising, and imply that this side of connectivity between layers will be less of an issue. This is important as the fan-in from synaptic layers will

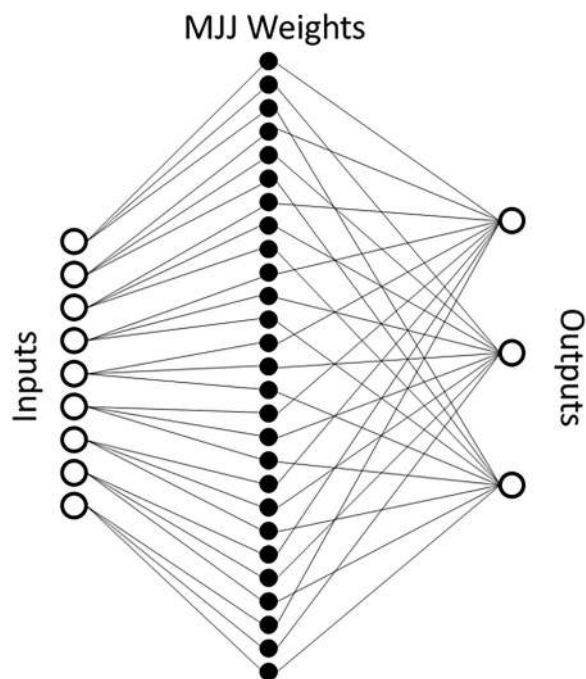


FIG. 20. Schematic of the feed forward neural network that was implemented in SPICE simulations to solve a 9-pixel three class example; this network requires 12 Josephson junctions and 27 MJJs.

contain weighted spikes that would potentially require weighted repeating nodes. Given the structure of modern feed forward neural networks, the additional overhead in Josephson junction count and wiring complexity will be one of the main challenges for this technology for the foreseeable future.

VI. SUMMARY AND PERSPECTIVE

Deep neural nets have been successful in many tasks. However, the current implementations are largely limited by software algorithms for training the CMOS processors used to implement the algorithms. Here, we focus on hybrid magnetic superconducting devices that naturally output voltage spikes and have plasticity due to embedded nanoscale magnetic structures. These devices have the advantage of very low energy ($<10^{-20}$ J/spike) and very fast operating speeds (>100 GHz) while naturally utilizing low loss (negligible dispersion up to 1 cm) high speed superconducting communication ($\sim 10^8$ m/s).

Both the magnetic and superconducting technologies have already been proven to scale well by leveraging standard semiconductor fabrication techniques. On the magnetic device side, large non-volatile memories based on magnetic tunnel junctions have been commercialized. On the superconducting device side, high speed microprocessors and communication systems have been developed, and large arrays of Josephson junctions are used to create the Josephson voltage standard. The scalability in size and energy and the proven manufacturability of these devices are a large advantage compared to many other emerging devices.

We have demonstrated nanoclustered magnetic Josephson junctions that have low-energy high-speed plasticity, which, in turn, enables a Josephson-junction-based native-neuromorphic technology. We have used proven simulation tools to model networks of these devices and have mapped them onto a standard neural network architecture. In this context, they have the potential to be faster and more energy efficient than the same network implemented in software even when run on specialized CMOS hardware such as a tensor processing unit. In addition, because of the high speed and low energy consumption, extending these devices to large, truly complex spiking networks has a promising outlook. In the next generation spiking neural networks, we can take advantage of spike timing dependence and the dynamics of the plasticity, which may enable all new functionality beyond that currently associated with software neural networks. This neuromorphic technology may be well suited for much more complex highly recurrent networks that can have learned intelligence that goes well beyond currently utilized feed forward networks.

ACKNOWLEDGMENTS

We acknowledge the technical and intellectual contributions from B. Baek, M. R. Pufall, P. F. Hopkins, P. D. Dresselhaus, S. P. Benz, and W. H. Rippard. We acknowledge NIST for funding this research and the IARPA C3 program for initial funding of the clustered MJJ work. Contribution of the National Institute of Standards and Technology, not subject to copyright.

- ¹Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521**(7553), 436–444 (2015).
- ²C. Mead, *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, MA, 1989).
- ³S. Furber, *J. Neural Eng.* **13**(5), 051001 (2016).
- ⁴P. Sterling and S. Laughlin, *Principles of Neural Design* (MIT Press, Cambridge, MA, 2015).
- ⁵C. Diorio, P. Hasler, B. A. Minch, and C. Mead, in *Neuromorphic Systems Engineering: Neural Networks in Silicon*, edited by T. S. Lande (Springer US, Boston, MA, 1998), pp. 315–337.
- ⁶D. B. Frédéric, J. Siddharth, W. Jun, and C. Gert, *J. Neural Eng.* **14**(4), 041002 (2017).
- ⁷P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, *Science* **345**(6197), 668 (2014).
- ⁸S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, *Proc. IEEE* **102**(5), 652–665 (2014).
- ⁹B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J. M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, *Proc. IEEE* **102**(5), 699–716 (2014).
- ¹⁰J. Schemmel, D. Brüderle, A. Grübl, M. Hock, K. Meier, and S. Millner, in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS), 30 May–2 June 2010* (IEEE, Paris, France), pp. 1947–1950.
- ¹¹M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, *Nat. Mater.* **12**(2), 114–117 (2013).
- ¹²M. Prezioso, F. Merrih-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, *Nature* **521**(7550), 61–64 (2015).
- ¹³J. J. Yang, M. X. Zhang, J. P. Strachan, F. Miao, M. D. Pickett, R. D. Kelley, G. Medeiros-Ribeiro, and R. S. Williams, *Appl. Phys. Lett.* **97**(23), 232102 (2010).
- ¹⁴W. Xu, S. Y. Min, H. Hwang, and T. W. Lee, *Sci. Adv.* **2**(6), e1501326 (2016).
- ¹⁵Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, *Nat. Photonics* **11**, 441 (2017).
- ¹⁶A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, *Sci. Rep.* **7**(1), 7430 (2017).
- ¹⁷K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, *Phys. Rev. X* **7**(3), 031014 (2017).
- ¹⁸J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, *Nature* **547**(7664), 428–431 (2017).
- ¹⁹A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier, and D. Querlioz, *Nat. Commun.* **9**(1), 1533 (2018).
- ²⁰A. Sengupta and K. Roy, *Appl. Phys. Express* **11**(3), 030101 (2018).
- ²¹C. D. Schuman, in *2017 International Joint Conference on Neural Networks (IJCNN), April 19, 2017* (IEEE, Anchorage, AK, USA), 2017), pp. 2636–2643.
- ²²M. Davies, N. Srinivasa, T. H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. H. Weng, A. Wild, Y. Yang, and H. Wang, *IEEE Micro*, **38**(1), 82–99 (2018).
- ²³D. Apalkov, B. Diény, and J. M. Slaughter, *Proc. IEEE* **104**(10), 1796–1830 (2016).
- ²⁴S. K. Tolpygo, *Low Temp. Phys.* **42**(5), 361–379 (2016).
- ²⁵S. P. Benz and C. A. Hamilton, *Proc. IEEE* **92**(10), 1617–1629 (2004).
- ²⁶G. Wendin, *Rep. Prog. Phys.* **80**(10), 106001 (2017).
- ²⁷J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**(5), 1175–1204 (1957).
- ²⁸M. Tinkham, *Introduction to Superconductivity* (Dover, New York, 1996).
- ²⁹T. Van Duzer and C. W. Turner, *Principles of Superconductive Devices and Circuits*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ, 1999).
- ³⁰B. D. Josephson, *Phys. Lett.* **1**(7), 251–253 (1962).
- ³¹R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman Lectures on Physics, Volume III: Quantum Mechanics* (Addison-Wesley, New York, 1965), Vol. 3.
- ³²J. A. Blackburn, M. Cirillo, and N. Grønbech-Jensen, *Phys. Rep.* **611**, 1–33 (2016).
- ³³M. L. Schneider, C. A. Donnelly, S. E. Russek, B. Baek, M. R. Pufall, P. F. Hopkins, P. D. Dresselhaus, S. P. Benz, and W. H. Rippard, *Sci. Adv.* **4**(1), e1701329 (2018).
- ³⁴R. L. Kautz, *J. Appl. Phys.* **49**(1), 308–314 (1978).
- ³⁵M. M. T. M. Dierichs, B. J. Feenstra, A. Skalare, C. E. Honingh, J. Mees, H. V. D. Stadt, and T. D. Grauw, *Appl. Phys. Lett.* **63**(2), 249–251 (1993).
- ³⁶P. Bunyk, K. Likharev, and D. Zinoviev, *Int. J. High Speed Electron. Syst.* **11**(1), 257–305 (2001).
- ³⁷K. K. Likharev and V. K. Semenov, *IEEE Trans. Appl. Supercond.* **1**(1), 3–28 (1991).
- ³⁸K. Segall, M. LeGro, S. Kaplan, O. Svitelskiy, S. Khadka, P. Crotty, and D. Schult, *Phys. Rev. E* **95**(3), 032220 (2017).
- ³⁹P. Crotty, D. Schult, and K. Segall, *Phys. Rev. E* **82**(1), 8 (2010).
- ⁴⁰T. Onomi and K. Nakajima, in *11th European Conference on Applied Superconductivity*, edited by S. Farinon, I. Pallecchi, A. Malagoli, and G. Lamura (IOP Publishing Ltd, Bristol, 2014), Vol. 507.
- ⁴¹Y. Yamanashi, K. Umeda, and N. Yoshikawa, *IEEE Trans. Appl. Supercond.* **23**(3), 1701004–1701004 (2013).
- ⁴²A. L. Hodgkin and A. F. Huxley, *J. Physiol.* **117**(4), 500–544 (1952).
- ⁴³S. Holmes, A. L. Ripple, and M. A. Manheimer, *IEEE Trans. Appl. Supercond.* **23**(3), 10 (2013).
- ⁴⁴E. O. Neftci, B. U. Pedroni, S. Joshi, M. Al-Shedivat, and G. Cauwenberghs, *Front. Neurosci.* **10**, 241 (2016).
- ⁴⁵V. Ginzburg, *Sov. Phys. JETP* **4**(2), 153–160 (1957) [*Zh. Eksp. Teor. Fiz.* **31**, 202–214 (1956)].
- ⁴⁶B. Matthias, H. Suhl, and E. Corenzwit, *Phys. Rev. Lett.* **1**(3), 92 (1958).
- ⁴⁷P. Fulde and R. A. Ferrell, *Phys. Rev.* **135**(3A), A550 (1964).
- ⁴⁸A. Larkin and Y. N. Ovchinnikov, *Sov. Phys. JETP* **20**, 762–769 (1965) [*Zh. Eksp. Teor. Fiz.* **47**, 1136–1146 (1964)].
- ⁴⁹A. I. Buzdin, *Rev. Mod. Phys.* **77**(3), 935–976 (2005).
- ⁵⁰F. Bergeret, A. L. Yeyati, and A. Martin-Rodero, *Phys. Rev. B* **72**(6), 064524 (2005).
- ⁵¹M. Eschrig, *Phys. Today* **64**(1), 43 (2011).
- ⁵²M. G. Blamire and J. W. Robinson, *J. Phys.: Condens. Matter.* **26**(45), 453201 (2014).
- ⁵³J. Linder and J. W. Robinson, *Nat. Phys.* **11**(4), 307 (2015).
- ⁵⁴A. Feofanov, V. Oboznov, V. Bol’ginov, J. Lisenfeld, S. Poletto, V. Ryazanov, A. Rossolenko, M. Khabipov, D. Balashov, and A. Zorin, *Nat. Phys.* **6**(8), 593 (2010).
- ⁵⁵F. Bergeret, A. F. Volkov, and K. B. Efetov, *Rev. Mod. Phys.* **77**(4), 1321 (2005).
- ⁵⁶C. Bell, G. Burnell, C. W. Leung, E. J. Tarte, D. J. Kang, and M. G. Blamire, *Appl. Phys. Lett.* **84**(7), 1153–1155 (2004).
- ⁵⁷E. A. Demler, G. B. Arnold, and M. R. Beasley, *Phys. Rev. B* **55**(22), 15174–15182 (1997).
- ⁵⁸A. I. Buzdin, L. Bulaevskii, and S. Panyukov, *Pis’ma Zh. Eksp. Teor. Fiz.* **35**(4), 147 (1982) [*JETP Lett.* **35**, 178–180 (1982)].
- ⁵⁹V. V. Ryazanov, V. A. Oboznov, A. Y. Rusanov, A. V. Veretennikov, A. A. Golubov, and J. Aarts, *Phys. Rev. Lett.* **86**(11), 2427–2430 (2001).
- ⁶⁰T. Kontos, M. Aprili, J. Lesueur, F. Genet, B. Stephanidis, and R. Boursier, *Phys. Rev. Lett.* **89**(13), 137007 (2002).
- ⁶¹I. Žutić, J. Fabian, and S. Das Sarma, *Rev. Mod. Phys.* **76**(2), 323–410 (2004).
- ⁶²T. Dietl and H. Ohno, *Rev. Mod. Phys.* **86**(1), 187–251 (2014).
- ⁶³T. Dietl and H. Ohno, *Mater. Today* **9**(11), 18–26 (2006).
- ⁶⁴V. Ko, K. Teo, T. Liew, T. Chong, M. MacKenzie, I. MacLaren, and J. Chapman, *J. Appl. Phys.* **104**(3), 033912 (2008).
- ⁶⁵S. E. Russek, C. A. Donnelly, M. L. Schneider, B. Baek, M. R. Pufall, W. H. Rippard, P. F. Hopkins, P. D. Dresselhaus, and S. P. Benz, in *2016 IEEE International Conference on Rebooting Computing (ICRC), 17–19 Oct. 2016* (IEEE, San Diego, CA, 2016), pp. 1–5.
- ⁶⁶B. Baek, W. H. Rippard, S. P. Benz, S. E. Russek, and P. D. Dresselhaus, *Nat. Commun.* **5**, 3888 (2014).
- ⁶⁷C. D. Tesche and J. Clarke, *J. Low Temp. Phys.* **29**(3–4), 301–331 (1977).
- ⁶⁸J. Clark, Bloomberg News (2015).
- ⁶⁹A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, *Nature* **542**(7639), 115 (2017).
- ⁷⁰M. L. Schneider, S. E. Russek, W. H. Rippard, and M. R. Pufall, U.S. patent application 15/722,508 (2017).
- ⁷¹M. L. Schneider, C. A. Donnelly, S. E. Russek, B. Baek, M. R. Pufall, P. F. Hopkins, and W. H. Rippard, in *2017 IEEE International Conference on Rebooting Computing (ICRC), 809 Nov. 2017* (IEEE, Washington, DC, 2017), pp. 1–4.
- ⁷²M. A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, 2015).