

Deep Neuromorphic Networks with Superconducting Single Flux Quanta

Alexander J. Edwards¹, Gleb Krylov², Joseph S. Friedman¹, and Eby G. Friedman²

¹The University of Texas at Dallas, Richardson, Texas, USA

²University of Rochester, Rochester, New York, USA

Abstract—Conventional semiconductor-based integrated circuits are gradually approaching fundamental scaling limits. Many prospective solutions have recently emerged to supplement or replace both the technology on which basic devices are built and the architecture of data processing. Neuromorphic circuits are a promising approach to computing where techniques used by the brain to achieve high efficiency are exploited. Many existing neuromorphic circuits rely on unconventional and useful properties of novel technologies to better mimic the operation of the brain. One such technology is single flux quantum (SFQ) logic – a cryogenic superconductive technology in which the data are represented by quanta of magnetic flux (fluxons) produced and processed by Josephson junctions embedded within inductive loops. The movement of a fluxon within a circuit produces a quantized voltage pulse (SFQ pulse), resembling a neuronal spiking event. These circuits routinely operate at clock frequencies of tens to hundreds of gigahertz, making SFQ a natural technology for processing high frequency pulse trains.

The similarities between SFQ and neuronal spiking has previously been observed; however, prior proposals for SFQ neural networks often require energy-expensive fluxon conversions, involve heterogeneous technologies, or exclusively focus on device level behavior. In this paper, a design methodology for deep single flux quantum neuromorphic networks is presented. Synaptic and neuronal circuits based on SFQ technology are presented and characterized. Based on these primitives, a deep neuromorphic XOR network is evaluated as a case study, both at the architectural and circuit levels, achieving wide classification margins. The proposed methodology does not employ unconventional superconductive devices or semiconductor transistors. The resulting networks are tunable by an external current, making this proposed system an effective approach for scalable cryogenic neuromorphic computing.

Index Terms—Neuromorphic computing, single flux quantum logic, superconductor electronics

I. INTRODUCTION

Conventional level-based artificial neural networks – while useful in a large number of applications – suffer from high computational costs that may be mitigated by using alternative biomimetic architectures such as spiking neuromorphic networks (SNNs) [1]. As the human brain can seemingly compute similar computation at a fraction of the energy, recent attention has gravitated towards biomimetic hardware, emulating biology in both phenomenological behavior and

emergent computation. Mimicking biological neuronal spiking behavior, SNNs are a class of neural networks in which data are encoded in a sequence of temporal spikes as opposed to a single real-valued signal (illustrated in Fig. 1); SNNs are therefore highly attractive for low power neuromorphic hardware.

Operating on minimal voltage pulses that function like neuronal spiking events, single flux quantum (SFQ) systems are particularly attractive for developing biomimetic SNNs. Superconducting Josephson junctions (JJs) inherently react to and regenerate fluxons, enabling extremely energy efficient neuronal and synaptic circuits. Whereas modern large scale SNNs utilize inefficient packet-based routing networks [2]–[4], SFQ may be ideally suited for ultra-fast, ultra-low power SNN systems. While the similarities between SFQ pulses and neuronal spiking have been explored in the literature [5]–[15], a multilayer fully-SFQ neuromorphic network has yet to be demonstrated.

We propose the first multi-layer fully-SFQ neuromorphic network, enabling ultra-low-power neuromorphic computation. The SFQ network comprises neuronal and synaptic SFQ primitives that may be cascaded with minimal layer-to-layer circuit design to construct several SNN architectures. SFQ synapses emulating the stochasticity in the brain [16], [17] are described along with SFQ leaky-integrate-and-fire (LIF) neurons. These primitive synaptic and neuronal circuits directly cascade with each other permitting the construction of deep neuromorphic networks. A multilayer network computing the XOR functionality is constructed and simulated showing high non-linearity and input separation.

II. BACKGROUND

SNNs and SFQ circuits have received significant attention in the literature including proposals for SNN primitives implemented with SFQ, as we summarize in the following subsections.

A. Spiking Neuromorphic Networks

SNNs attempt to replicate biological spiking behavior so as to unlock the ultra-low-power computation observed in biological systems. Biological neural networks encode information based on the frequency and relative timing of neuronal spiking

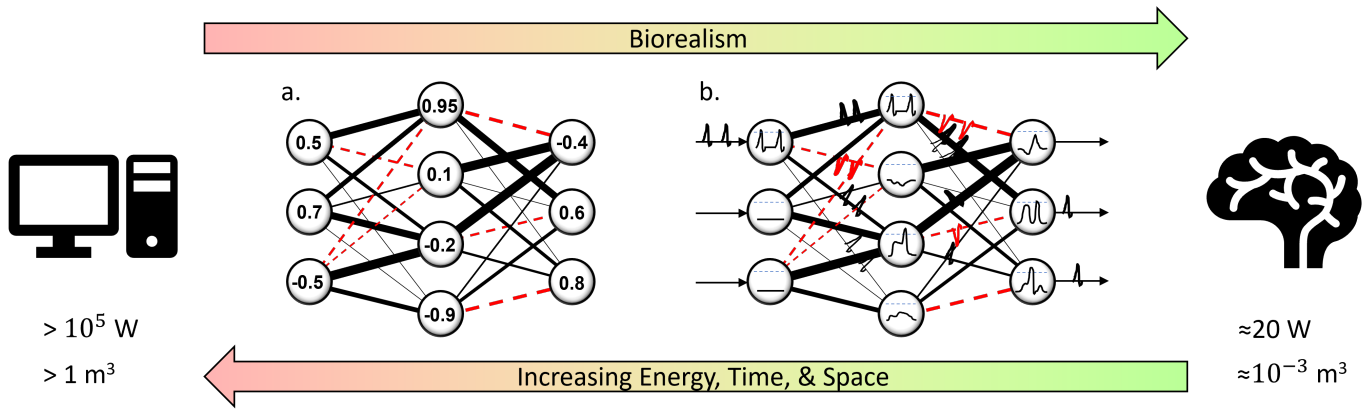


Fig. 1: Biological inspiration for spiking neuromorphic networks with SFQ. **a.** Conventional artificial neural network. Information propagates through the network from left to right. The neuron states (circles) are encoded by a single real-valued number which is modulated by downstream synapses (black and red lines) which are collected by the successive (post-synaptic) neurons according to an activation function. **b.** Spiking neural network (SNN). Rather than representing information as a single number, spiking neurons mimic biology by retransmitting spikes downstream, which are modulated by the synapse strength. The information representation can therefore be much richer, enabling systems approaching the incredible efficiency observed in biology.

events. This behavior is counter to conventional software neural networks, which represent information as single-valued numbers. Spiking neuromorphic networks are an attempt to more richly encode information with equivalent or cheaper hardware for more efficient, biomimetic computation.

Neurons in SNNs mimic biology by aggregating spiking activity from upstream neurons and firing when sufficiently stimulated, creating a spike event that is propagated downstream. A common neuronal model, the leaky-integrate-and-fire (LIF) neuron, integrates input spikes into an internal potential which leaks over time if not stimulated. When the potential crosses the action potential threshold, the neuron fires, resets, and begins integrating again. The LIF model is a very common neuronal model for SNNs [1].

Synapses in SNNs mimic biology by modulating the strengths of the connections between upstream and downstream neurons. Most SNNs employ synapses that modulate the *amplitude* of spiking events, whereas alternative approaches may modulate *spike rate* or – inspired by the prevalent stochasticity observed in biological systems [16], [17] – stochastically gate individual synaptic spiking events, the method used here.

B. Single Flux Quanta Circuitry

Single flux quantum logic is an emerging cryogenic technology for highly energy efficient computing [18]. SFQ circuits are based on Josephson junctions (JJs) and superconducting quantum interference devices (SQUIDs), and operate with magnetic flux quanta. The quanta are typically represented by voltage pulses of quantized area equal to the magnetic flux quantum ($\Phi_0 \sim 2.07 \text{ mV} \cdot \text{ps}$) [19]. These pulses are generated in a process often referred to as JJ switching – a shift of superconducting phase between the terminals of the JJ by 2π .

The primary advantage of SFQ circuits for digital logic is the unparalleled energy efficiency. Each 2π transition of a

typical $100 \mu\text{A}$ JJ dissipates energy on the order of $2 \times 10^{-19} \text{ J}$ [20]. Although a logic operation requires several switches, the energy per operation is several orders of magnitude lower than state-of-the-art CMOS logic even accounting for the cryogenic cooling to 4.2 K (liquid helium temperature) [21].

Whereas in conventional SFQ logic, information is encoded as the presence or absence of an SFQ pulse within a specific time period, alternative JJ-based circuits can generate and operate on more complex SFQ pulse sequences. This includes the generation of pulse sequences with a controllable frequency and stochastic switching induced by thermal noise. Both properties are exploited in this work.

Several approaches for neuromorphic computing with SFQ circuits have been proposed, however a fully-SFQ multilayer neuromorphic network remains to be demonstrated. Several proposals demonstrate individual SFQ gates for neuronal [5]–[7], [15], synaptic [7]–[9], [22], or interconnect [10] functionality. Two-neuron oscillatory Hopfield networks [11], [12] and single layer feed-forward Spiking networks [13], [23] have been proposed, and a multi-layer feed-forward spiking network with heterogeneous CMOS-SFQ circuitry has been demonstrated in simulation [14]. Some of the proposed approaches utilize magnetic Josephson junctions to gradually modify the internal state of the gates, enabling online learning [8], [13]. Fabrication processes used to manufacture these devices are, however, not well established, and the resulting circuits are limited in scale. In [24], an SFQ-based methodology for building neuromorphic networks is proposed, where a bipolar current is used to represent a logic state, which is converted into a train of SFQ pulses for transmission.

III. DEEP NEUROMORPHIC NETWORKS WITH SFQ PRIMITIVE CIRCUITS

The first fully SFQ deep neuromorphic network is described here; primitive SFQ synaptic and neuronal circuits digitally

manipulating SFQ spike sequences are described. The stochastic synapse is based on a Josephson balanced comparator [25], and the pulse trains are briefly converted into magnetic flux within neurons before undergoing a threshold to produce an output pulse train. The resulting network uses established superconductive fabrication processes [26] and is tuned by an external current, facilitating large scale integration. All of the circuits presented here are simulated in WRspice [27] based on the state-of-the-art MIT LL SFQ5ee fabrication process [26].

A. Data encoding with SFQ

Because of the phenomenological similarity between neuronal spikes and SFQ pulses, information in the network is represented in the timing and frequencies of SFQ pulses. Upon sufficiently stimulating a JJ with voltage and/or bias current, a 2π phase shift is created around a superconducting loop, producing a corresponding SFQ voltage pulse across the JJ, which can be propagated to stimulate downstream JJs. As a single flux quantum is the smallest possible non-zero voltage pulse, fluxons are ideal information carriers for SNNs, enabling extremely low power neuromorphic processing. Furthermore, information in a sequence of fluxons is not encoded in the magnitude of the spike but rather in the spike timing as is the case in SNNs and, more notably, the brain. Fluxons have non-volatile attributes as well, enabling short-term memory and the construction of low power LIF neurons.

A wide range of SNN topologies are available, all of which are amenable to SFQ implementations. Among these topologies are perceptron networks and feed-forward deep neural networks; both of which can employ stationary input spike rates for classification tasks. Recurrent neural networks may be constructed as well, enabled by internal time delays in fluxon propagation. Information coding in recurrent neuromorphic networks is dependent upon the relative timing between spiking events. Spike-rate-coded networks – with information encoded in the mean spike frequency – may be trained with backpropagation similar to a level-based ANN. Online learning using spike timing – such as spike-time-dependent plasticity [1] – is also amenable to SFQ networks [22] and may similarly ease training costs.

B. Stochastic-Pass Synapses

Inspired by the stochasticity in biological systems, the stochastic-pass synapse modulates the connection between upstream and downstream neurons by stochastically gating fluxons. Whereas conventional synapse approaches modulate the effective amplitude of spiking events, due to the quantized nature of fluxons, it is difficult to accomplish this without conversion between the fluxons and analog current, thereby decreasing system efficiency. Inspired by the pervasive stochasticity in biological systems [16], [17], the proposed synapse stochastically gates incoming spikes, encoding the synaptic weight in the probability that an incoming fluxon will propagate through the synapse.

The proposed synapse is constructed from a Josephson balanced comparator [25], [28], harnessing thermal noise for

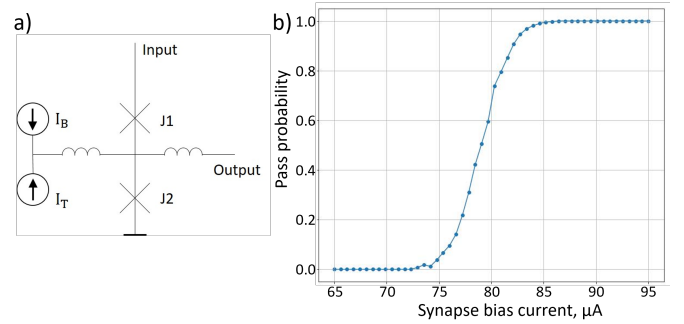


Fig. 2: Synaptic circuit, a) schematic, and b) graph of pass-probability with bias current. The data are presented for $I_C(J_1, J_2) = 150 \mu\text{A}$, input SFQ pulse frequency of 50 GHz, and $T = 4.2 \text{ K}$.

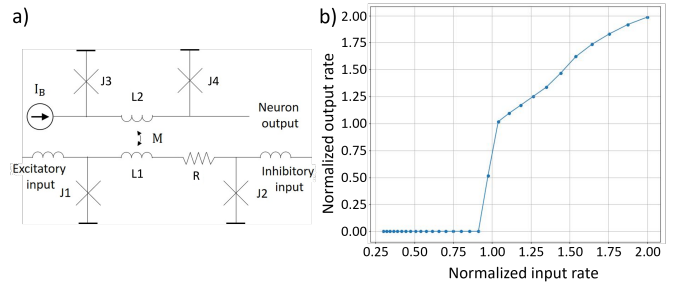


Fig. 3: a) Neuronal circuit. b) Neuron activation function. The input and output rates are normalized to approximately 33.3 GHz.

true stochasticity. A Josephson balanced comparator is a pair of serially connected JJs with the bias current, I_B , applied between these JJs, as shown in Fig. 2a. When an SFQ pulse is applied to the input, one of the JJs within the comparator undergoes a 2π phase shift, depending on the magnitude of I_B . For small I_B , $J1$ switches, absorbing the input pulse, whereas for large I_B , $J2$ switches, propagating the input pulse to the output. In the absence of noise current I_T , the gating functionality depends deterministically on the applied currents, however in the presence of thermal fluctuations in the applied currents, the balanced comparator exhibits stochastic behavior – a “grey zone” [25]. A similar approach has been proposed for SFQ based synapses [24], where a C-SQUID [29] is used rather than a balanced comparator.

The synapse weight is encoded in the probability that an input spike will propagate to the output and can be modulated after fabrication by adjusting I_B . Fig. 2b depicts the probability of passing an incoming SFQ pulse to the output as a function of bias current for specific circuit parameters, demonstrating a continuous swing of weights between 0 (all fluxons blocked) and 1 (all fluxons propagated).

C. Leaky Integrate-and-Fire Neurons

Neuronal LIF circuits – integrating incoming SFQ pulses until an internal threshold is reached and outputting a resultant SFQ spike sequence – are described here. A leaky SQUID

loop inductively coupled with a firing SQUID loop [30] to implement integrating, leaking, and firing behavior.

A conventional SQUID loop with a large inductance is used to integrate incoming SFQ pulses. The input SFQ pulses are applied across $J1$ or $J2$ (indicated in Fig. 3a) as excitatory or inhibitory inputs respectively. As $J1$ switches, the fluxon energy is stored as current in inductor $L1$. This energy is accumulated with additional inputs increasing the flux stored in the loop. A fluxon applied to the inhibitory input switches $J2$ reduces the flux stored in the loop allowing neurons to have both excitatory and inhibitory interactions.

Leaking may be implemented with the addition of a resistive element R dissipating the energy stored in inductor $L1$. A resistance on the order of a few ohms produces a linear leakage characteristic, the timing of which can be tuned by changing the size of the resistor. Utilizing the dynamic SFQ (DSFQ) leakage mechanism, additional JJs may be introduced into the loop to provide a faster reset of the circulating current [31], [32].

When the current in the SQUID is sufficiently large, an inductively coupled SQUID produces a firing pulse. A JJ partially biased by an inductively coupled neuron loop current, is shown in Fig. 3a. Due to the coupling between $L1$ and $L2$, as $L1$ integrates input pulses, current in $L2$ simultaneously increases. Eventually current in $L2$ – with contributions from $L1$ and I_B – is sufficiently large to switch $J4$ producing an output pulse. In the case where inhibitory inputs are more frequent than excitatory inputs, $J3$ will switch instead of $J4$ sans output spike. Bias currents through the JJs help to regenerate fluxons, enable signal fan-out, and improve the cascade characteristics.

The relationship between the neuron input and output rates is non-linear, a necessary condition for proper neural network functionality. Fig. 3b depicts the relationship between input and output spike rates (equivalent to the activation function of a conventional level-based neuron if spike-rate-coding is employed). Note the distinct non-linearity akin to a rectified linear unit (RELU), scaled exponential linear unit (SELU), or sigmoidal activation function. The width and slope of the threshold region as well as the saturation characteristics are adjusted by tuning I_B , $L1$, $L2$, R , and $I_C(J1)$ (see Fig. 3a). The threshold input rate may be tuned after fabrication through the application of bias input sequences. At higher frequencies, output rate saturates as $J4$ moves toward the resistive regime adding additional non-linearity.

The input fan-in is realized using conventional RSFQ confluence buffers (pulse mergers) [19]. These buffers exhibit a saturating spike rate as multiple input pulses arriving in close succession produce only one output pulse. This property assists in the saturation of the neuronal output spike rate.

Negative weights are applied as inhibitory inputs to the neuron. A single synapse can therefore be implemented as a differential pair of synapse circuits connected to the excitatory and inhibitory inputs of a neuron. Inhibitory or excitatory bias input spike sequences may be added to adjust the neuron threshold.

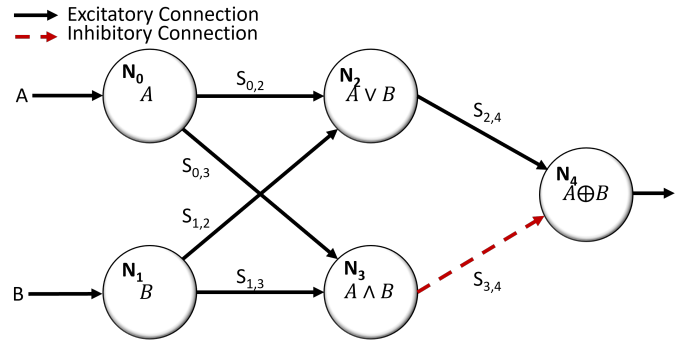


Fig. 4: XOR spiking neural network. The primary inputs, bias input, and primary outputs are represented as spike frequencies in proportion to the saturation output neuronal rate.

Fan-out is managed by splitter trees, which could incur well-known area and delay overhead [10]. Multiple techniques exist to reduce the overhead of the signal fanout in large scale SFQ circuits [33], [34]. These techniques are primarily based on utilizing a splitter with more than two outputs at the cost of reduced parameter margins.

D. Cascading Synapses and Neurons for Deep Networks

The SFQ neuromorphic primitives can be cascaded to construct deep neuromorphic networks enabling straightforward circuit design with minimal layer-to-layer tuning. As inputs and outputs from synapses, neurons, fan-in, and fan-out circuits are all SFQ pulses there is no need for costly signal conversion to construct large multilayer networks.

To ensure that spiking activity is similar from layer to layer, spiking activity can be regenerated through the use of signal confluence and additional input spike sequences. When an RSFQ confluence buffer merges two or more spike sequences, the output spike rate is the combined rate of the input sequences, increasing signal activity. Furthermore, additional bias spike sequences may be incorporated to further regenerate signal activity.

Layer-to-layer spiking activity can be tuned at a network-architecture level, mitigating the need to individually tune circuit parameters for each layer or for a specific network architecture. Specifically, synapse weights may be tuned to ensure that spiking activity remains within proper regions of operation for the circuits. Furthermore, the rates of the bias input spike sequences may be tuned through the use of synapse circuits.

IV. CASE STUDY: DEMONSTRATION OF TWO-LAYER XOR NEURAL NETWORK

The first demonstration of a multi-layer fully SFQ neuromorphic network is described here. The network is trained to compute the XOR functionality, and demonstrates large non-linearity and input separation.

A. Training and Weight Mapping of XOR Network

A trained neuromorphic network to compute the two bit XOR function requires at least two layers in an ANN along with negative weights and bias inputs. The chosen network

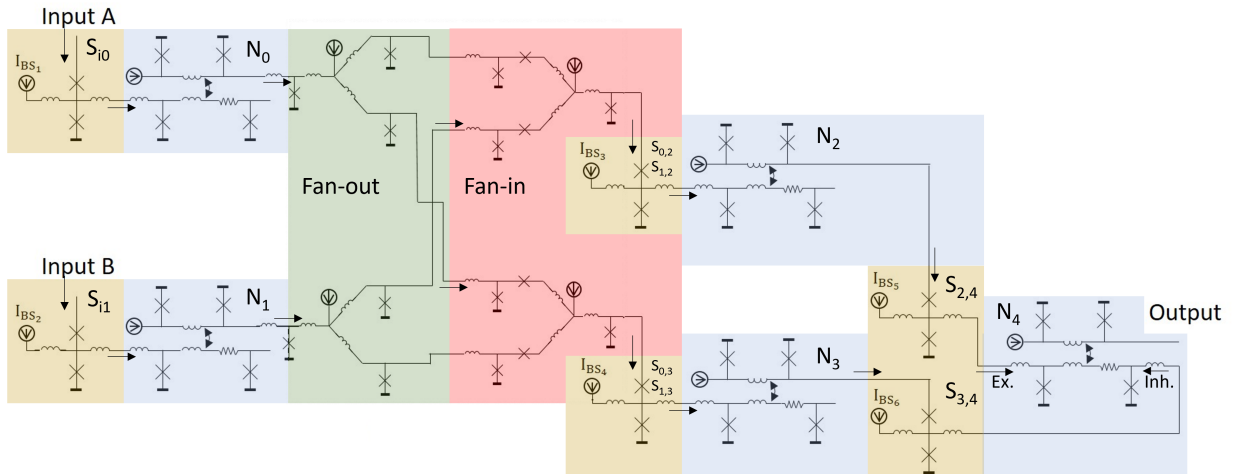


Fig. 5: Neuromorphic SFQ network computing a two bit XOR function with synapses in yellow, neurons in blue, fan-out in green, and fan-in in red.

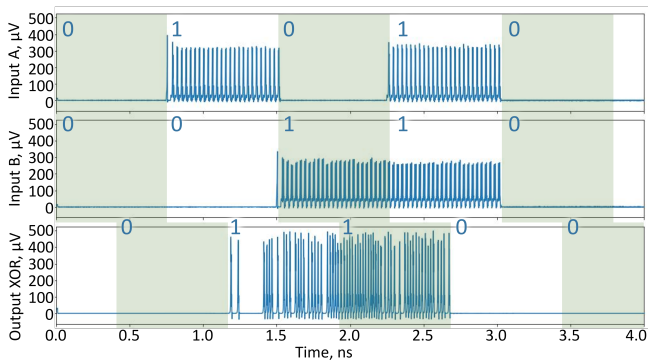


Fig. 6: Spiking behavior with different input stimuli for XOR network.

is depicted in Fig. 4. Input neurons 0 and 1 normalize the network inputs rates to the neuronal output levels. Due to trained neuronal biases, neuron 2 will have a high output when one of the two inputs is on while neuron 3 requires both inputs to be on in order to activate. Neuron 4 uses inhibitory weights to compute the XOR functionality, activating only if neuron 2 is active while neuron 3 is not. Neuron 4 will therefore only turn on when exactly one of the inputs is on: the XOR functionality.

Spike-rate-coding was chosen to map the SFQ neuromorphic primitives to the chosen network. As described in Section III-A, spike rate coding supports training similar to a level-based ANN and allows direct mapping of architecture and trained weights between the level-based ANN and the spike-rate-coded SNN. The two network inputs are stationary pulse trains with rates close to neuronal saturating rates, encoding a logic 0 (1) as a low (high) spike rate. The synapse weights are encoded in the probability of the spike propagation. The bias input sequences are trained to make the neurons sensitive to different combinations of inputs, tuning the threshold input rate of Fig. 3b to ensure large separability between the input patterns. The network outputs may be interpreted by the average spike rate in proportion to the saturating spike rate of the neurons.

B. Circuit topology of multilayer network

The XOR network was directly implemented with the SFQ primitives described in Section III. A schematic of the neuromorphic network based on these components, computing a two bit XOR function (presented in Fig. 4) is shown in Fig. 5. The synaptic weights and neuronal biases are tuned by changing the bias current.

C. Network Results

As shown in Fig. 6, the network correctly computes the XOR function. Additionally the network is robust to variations in the input rates, showing a large separation between classes. A phase diagram of classification for the XOR network is shown in Fig. 7, where the dependence of the network output rate is shown as a function of the two input rates. It is desirable for the high output rates corresponding to the output of logic 1 (shown in Fig. 7 in red) to map to the input rates corresponding to logic 10 and 01. The input rates corresponding to logic 00 and 11 should produce low spike activity (shown in blue). This classification diagram displays good separation of the output states with respect to the input spike rates. The proposed network is therefore robust to variations in the input rates.

V. CONCLUSIONS

The first demonstration of a fully SFQ multi-layer neuromorphic network is presented here along with SFQ neuromorphic primitive circuits that can be directly cascaded to construct a broad range of network architectures. While the demonstration shows one feed-forward network using spike rate coding, the circuits are amenable to richer representations of information in spiking neuromorphic networks including recurrent neural networks, spike-time-dependent networks, and online learning. As network layers can regenerate spiking behavior, deep network architectures are readily attainable through natural cascading of successive layers.

The proposed network – entirely based in available SFQ technologies – has the advantage of being extremely energy efficient as compared with conventional CMOS technologies

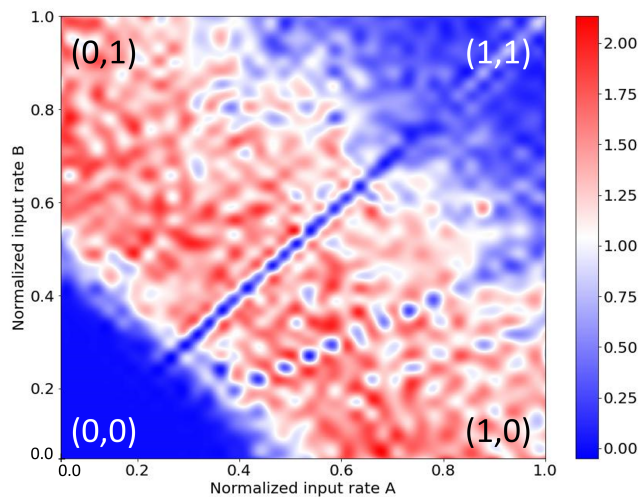


Fig. 7: Phase diagram of classification for XOR network output. Red represents high output rates (between 1 and 2), blue represents output rate between 0 and 1 with the colors in between representing the domain boundaries. The input and output rates are normalized to a pulse rate of approximately 33.3 GHz.

[21], [14]. Furthermore, as conversions between fluxons and analog currents are constrained within each individual neuron, the network is compact and scalable. Additionally, the proposed scheme does not require unconventional devices or complex 2.5-D or 3-D integration, and can be produced using standard niobium fabrication processes.

REFERENCES

- [1] C. D. Schuman *et al.*, “Opportunities for neuromorphic computing algorithms and applications,” *Nature Computational Science*, vol. 2, no. 1, pp. 10–19, Jan 2022.
- [2] S. B. Furber *et al.*, “Overview of the spinnaker system architecture,” *IEEE Transactions on Computers*, vol. 62, no. 12, pp. 2454–2467, 2013.
- [3] F. Akopyan *et al.*, “Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1537–1557, 2015.
- [4] M. Davies *et al.*, “Loihi: A neuromorphic manycore processor with on-chip learning,” *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.
- [5] U. S. Goteti and R. C. Dynes, “Superconducting neural networks with disordered josephson junction array synaptic networks and leaky integrate-and-fire loop neurons,” *Journal of Applied Physics*, vol. 129, no. 7, 2021.
- [6] Y. Yamanashi *et al.*, “Pseudo sigmoid function generator for a superconductive neural network,” *IEEE Transactions on Applied Superconductivity*, vol. 23, no. 3, pp. 1 701 004–1 701 004, 2013.
- [7] P. Crotty *et al.*, “Josephson junction simulation of neurons,” *Phys. Rev. E*, vol. 82, p. 011914, Jul 2010.
- [8] M. L. Schneider *et al.*, “Synaptic weighting in single flux quantum neuromorphic computing,” *Scientific Reports*, vol. 10, no. 1, p. 934, Jan 2020.
- [9] —, “Ultralow power artificial synapses using nanotextured magnetic josephson junctions,” *Science Advances*, vol. 4, no. 1, p. e1701329, 2018.
- [10] M. L. Schneider and K. Segall, “Fan-out and fan-in properties of superconducting neuromorphic circuits,” *Journal of Applied Physics*, vol. 128, no. 21, p. 214903, 2020.
- [11] K. Segall *et al.*, “Synchronization dynamics on the picosecond time scale in coupled josephson junction neurons,” *Phys. Rev. E*, vol. 95, p. 032220, Mar 2017.
- [12] —, “Phase-flip bifurcation in a coupled josephson junction neuron system,” *Physica B: Condensed Matter*, vol. 455, pp. 71–75, 2014, 21st Latin American Symposium on Solid State Physics - SLAFES 2013.
- [13] M. L. Schneider *et al.*, “Tutorial: High-speed low-power neuromorphic systems based on magnetic josephson junctions,” *Journal of Applied Physics*, vol. 124, no. 16, p. 161102, 2018.
- [14] A. Bozbey *et al.*, “Single flux quantum based ultrahigh speed spiking neuromorphic processor architecture,” *arXiv:1812.10354 [cs.ET]*, 2020.
- [15] M. A. Karamuftuoglu *et al.*, “J-soma: Towards a spiking neuromorphic processor architecture,” *IEEE Transactions on Applied Superconductivity*, pp. 1–7, 2023.
- [16] W. Maass, “To spike or not to spike: That is the question,” *Proceedings of the IEEE*, vol. 103, no. 12, pp. 2219–2224, 2015.
- [17] H. A. Braun, “Stochasticity versus determinacy in neurobiology: From ion channels to the question of the “free will”,” *Frontiers in Systems Neuroscience*, vol. 15, 2021.
- [18] G. Krylov and E. G. Friedman, *Single Flux Quantum Integrated Circuit Design*, Springer, ISBN # 978-3-030-76884-3, 2022.
- [19] K. K. Likharev and V. K. Semenov, “RSFQ Logic/Memory Family: a New Josephson-Junction Technology for Sub-Terahertz-Clock-Frequency Digital Systems,” *IEEE Transactions on Applied Superconductivity*, vol. 1, no. 1, pp. 3–28, March 1991.
- [20] O. A. Mukhanov, “Energy-Efficient Single Flux Quantum Technology,” *IEEE Transactions on Applied Superconductivity*, vol. 21, no. 3, pp. 760–769, June 2011.
- [21] D. S. Holmes *et al.*, “Energy-Efficient Superconducting Computing—Power Budgets and Requirements,” *IEEE Transactions on Applied Superconductivity*, vol. 23, no. 3, p. 1701610, June 2013.
- [22] K. Segall *et al.*, “A superconducting synapse exhibiting spike-timing dependent plasticity,” *Applied Physics Letters*, vol. 122, no. 24, p. 242601, 06 2023.
- [23] M. Schneider *et al.*, “Supermind: a survey of the potential of superconducting electronics for neuromorphic computing,” *Superconductor Science and Technology*, vol. 35, no. 5, p. 053001, mar 2022.
- [24] V. K. Semenov *et al.*, “A new family of biosfq logic & memory cells,” *IEEE Transactions on Applied Superconductivity*, vol. 32, no. 4, pp. 1–5, 2022.
- [25] T. Filippov and V. Kornev, “Sensitivity of the balanced josephson-junction comparator,” *IEEE Transactions on Magnetics*, vol. 27, no. 2, pp. 2452–2455, 1991.
- [26] S. K. Tolpygo *et al.*, “Advanced Fabrication Processes for Superconductor Electronics: Current Status and New Developments,” *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 5, pp. 1–13, August 2019.
- [27] S. R. Whiteley. WRspice Reference Manual. Whiteley Research Inc. [Online]. Available: <http://www.wrcad.com/manual/wrsmmanual.pdf>
- [28] T. V. Filippov *et al.*, “Gray zone and threshold current measurements of the josephson balanced comparator,” *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 5, pp. 1–7, 2021.
- [29] V. Semenov, “Digital SQUIDS: New Definitions and Results,” *IEEE Transactions on Applied Superconductivity*, vol. 13, no. 2, pp. 747–750, June 2003.
- [30] G. Krylov and E. G. Friedman, “Design for Testability of SFQ Circuits,” *IEEE Transactions on Applied Superconductivity*, vol. 27, no. 8, pp. 1–7, Art no. 1 302 307, December 2017.
- [31] S. V. Rylov, “Clockless Dynamic SFQ and Gate With High Input Skew Tolerance,” *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 5, pp. 1–5, August 2019.
- [32] G. Krylov and E. G. Friedman, “Asynchronous Dynamic Single Flux Quantum Majority Gates,” *IEEE Transactions on Applied Superconductivity*, vol. 30, no. 5, pp. 1–7, Art no. 1 300 907, August 2020.
- [33] T. Jabbari *et al.*, “Splitter trees in single flux quantum circuits,” *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 5, pp. 1–6, 2021.
- [34] N. Katam *et al.*, “Design of multiple fanout clock distribution network for rapid single flux quantum technology,” *Proceedings of the Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 384–389, January 2017.