Enhancing SNN Training Performance: A Mixed-Signal Triplet Reconfigurable STDP Circuit with Multiplexing Encoding

Honghao Zheng
Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, USA
zhenghh@vt.edu

Yang Yi
Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, USA
yangyi8@vt.edu

2023 IEEE International Symposium on Circuits and Systems (ISCAS) | 978-1-6654-5109-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISCAS46773.2023.10181729

Abstract—In spike-timing-dependent plasticity (STDP), synaptic weights are modified according to the relative time difference between pre and post-synaptic spikes of spiking neural network (SNN). A triplet STDP model was proposed since this model can better take account of a series of spikes and thus more closely mimic the activity in biological neural systems. Circuit that can switch between different STDP rules was also introduced to improve the range of STDP applications. To apply the advantages of triplet STDP to various tasks, a mixed-signal triplet reconfigurable STDP circuit and its hardware prototype are proposed in this paper. The performance analysis of the STDP training algorithm is carried out with a hardware testbench as well as Pytorch-based SNN. This triplet STDP design achieves 3.28% and 3.63% higher accuracy than the pair STDP learning rule through datasets such as MNIST and CIFAR-10. Our design shows one of the best reconfigurability while keeping a relatively low energy per spike operation (SOP) through the performance comparison with the state of the arts.

Index Terms—Spike-timing-dependent plasticity, triplet STDP, reconfigurable STDP.

I. Introduction

NEUROMORPHIC has shown the potential to improve computing efficiency for data-intensive applications. In-spired by biological neural networks, SNN was proposed as an alternative to the artificial neural network (ANN)s since it can more closely mimic biological neural systems [1]. SNNs only transfer information when the membrane potential exceeds some particular threshold value, and thus the information is transmitted in the system in the form of spikes. Due to its power efficiency and parallel computing property, the SNN has become a promising candidate for data-intensive tasks such as image processing. For example, Loihi, the SNN processor fabricated by Intel, can classify objects in 3D space with only 0.001 times the power of a standard computer [2], [3].

Besides the basic STDP rule [4], another advanced STDP rule is the triplet STDP [5]. It takes account of a series of spikes instead of only a pair of spikes. Thus, it more accurately mimics complex biological neural mechanisms. Even for pair STDP rules, there are rules besides the asymmetric rule. Research has verified that these rules have advantages in various engineering tasks, and researchers have proposed a circuit design that can switch between these algorithms [6]. To integrate the advantage of the triplet STDP rule and the reconfigurable STDP circuits, we have proposed a triplet reconfigurable STDP circuit design. Major contributions of our work are summarized as follows:

- To the best of our knowledge, this proposed work is the first CMOS integrated circuit (IC) design of the triplet reconfigurable STDP circuit.
- Consisting of a hardware testbench, Python simulation, and prototype measurement. The performance analysis of the triplet STDP rule working with multiplexing encoding schemes of the MNIST and CIFAR-10 datasets has shown 3.28% and 3.63% higher accuracy than the pair STDP rule and also 4.9% and 14.05% higher accuracy than working with other temporal encoding schemes such as the time to first spike (TTFS) and Interspike Interval (ISI) encoding.
- The layout design of the introduced work consumes 258μW of power and takes 0.045mm^2 of area. Compared with other state of the arts, it shows the best reconfigurability while keeping reasonable energy per SOP.

II. The Triplet Reconfigurable STDP Circuit Design

A. Multiple STDP Rules and Applications

The asymmetric STDP rule can not fulfill the requirements of various engineering applications. Thus, many other STDP rules were proposed [6]. These STDP algorithms modify the synaptic weights differently with relative time differences. For instance, the anti-STDP algorithm increments the weight when the post-spike arrives before the pre-spike and vice versa. According to their time-weight relation shapes, these

This work was supported in part by the U.S. National Science Foundation (NSF) under Grant CCF-1750450, Grant ECCS-1731928, and Grant CCF-1937487.

STDP algorithms are named DPD, PP, DD, PD... where D means depression and P means potentiation. These rules have advantages for various applications, respectively. The basic asymmetric STDP rule, DP, is used in unsupervised learning. In order to carry out supervised learning in SNN, the DP learning rule and the anti-STDP learning rule, PD, are both utilized in applications. For associative learning, the symmetric STDP rule, DPD, is proven effective in [7]. [8] has used three different STDP rules in the SNN for the classification tasks. As for the symmetric rules such as potentiating rule, PP, and depressive rule, DD, although they only have one side of weight adjustment, they are still very effective in liquid state machine (LSM) and some classification projects [9].

B. Triplet STDP Rule and Advantages

As mentioned in Section.I, the triplet STDP rule (TSTDP) takes account of three spikes. The combination of the spikes could be pre-post-pre or post-pre-post. These combinations of spikes describe how the time difference of spikes is utilized to modify synaptic weights. Compared with the pair-based STDP (PSTDP) rules, the TSTDP can mimic and realize higher order of spiking patterns. The mathematical model of the basic asymmetric TSTDP learning rule [5] can be written as:

$$\Delta W = \begin{cases} A_1^+ e^{\left(\frac{-\Delta t_1}{\tau_1^+}\right)} + A_2^+ e^{\left(\frac{-\Delta t_2}{\tau_2^+}\right)} e^{\left(\frac{-\Delta t_1}{\tau_1^+}\right)} \\ -A_1^- e^{\left(\frac{\Delta t_1}{\tau_1^-}\right)} - A_2^- e^{\left(\frac{-\Delta t_3}{\tau_2^-}\right)} e^{\left(\frac{\Delta t_1}{\tau_1^-}\right)} \end{cases} \quad (1)$$

where A_1 and A_2 are the potentiation and depression parameters and Δt_1 represents the time difference between the pre-neuron and post-neuron spikes. As for Δt_2, it equals to $t_{post}(n)$–$t_{post}(n-1)$, the time gap of two immediate successive post spikes. n means one certain time step and $n-1$ represents its immediate last post spike. Similarly, Δt_3 represents the time gap between two pre-spikes. It is noticeable that the TSTDP has a higher-order term in both depression and potentiation formulas.

Physiological experiments have demonstrated that the TSTDP can mimic the biological mechanism more accurately. At first, researchers have found that the PSTDP rule could not explain the biological synaptic weight changing result. Moreover, the TSTDP can reproduce the frequency effects that happened in the experiments where the potentiation amplitude increases with the spike firing rate [10]. Besides these advantages, the most critical advantage of TSTDP is the capability to reproduce Bienenstock–Cooper–Munro (BCM) model behavior [5]. With that, neurons will enable input selectivity when receiving multiple inputs [10].

C. Triplet Reconfigurable STDP Circuit Design

Intending to apply the advantages of the TSTDP learning rule to various applications, we proposed an STDP learning circuit that applies different shapes of STDP learning rules to the triplet STDP. The triplet part of the triplet reconfigurable STDP circuit is inspired by [11] and the idea of switching between different rule is inspired by [12]. With such configuration, the merits of TSTDP, such as frequency effects and BCM model behavior, will be utilized in different applications. Similar as the PSTDP learning rule, the TSTDP can be applied to unsupervised learning, supervised learning, associative learning and recognition tasks and improve the performance of neural networks. To the best of our knowledge, it is the first IC implementation of the triplet reconfigurable STDP circuit.

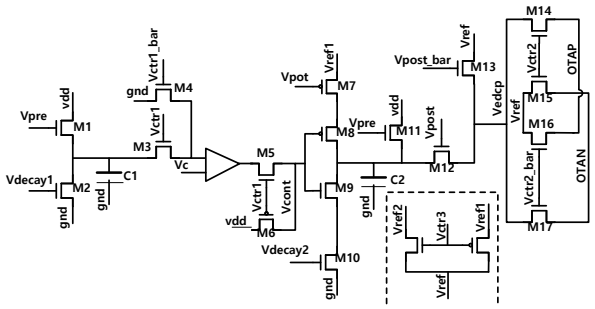

Fig. 1. Circuit schematic of the first-order time window generator.

The triplet reconfigurable STDP circuit, is controlled by six digital signals to switch between different shapes of learning rules. These six digital signals are separated into pre-spike and post-spike control signals. The circuit implementation can be divided into five parts. Two are the time window generators for first-order potentiation and depression, and the other two are for second-order potentiation and depression. The fifth part of the circuit is the synapse core. It uses time window signals from the other four parts to adjust the weight of the synapse.

The first-order time window generator is demonstrated in Fig. 1. The V decay signals are used to control the decaying rate of voltages across capacitors after being charged to supply voltage. As for V pot, this signal controls the rising rate of voltage across C2 when V cont is at digital low. When V pre spikes come, the voltage across C1 will increase and decrease to 0 exponentially due to RC effect. When V ctr1 is at digital high, the voltage is compared with V c. Before the voltage across C1 becomes lower than V c, V cont will uncharge C2, and when C1 voltage is smaller than V c, V cont will charge C2. When V ctr1 is set to 0, V cont will always uncharge C2. After that, when pre spikes come, the voltage across C2 will increase to the supply voltage and exponentially decrease to 0. After V cont is set to 0, C2 voltage will increase back to V ref1. The post spikes will trigger M12 and make V edcp equal to C2 voltage. V ref will be compared with C2 voltage and produce current for weight changing with an operational transconductance amplifier (OTA). V ctr2 controls which voltage is fed into the positive input of the OTA and the other to the negative input. V ctr3 selects which one of V ref1 and V ref2 is fed to V ref.

As shown in Fig. 2, the second-order time window generators are similar to the first-order ones. Only after C2 the post spike controlled switch is replaced by two transmission gates and one capacitor to delay the signal to the next coming spike.

The fifth part, the synapse core, is demonstrated in Fig. 3. It is composed of 2 OTAs, one for first-order terms and the

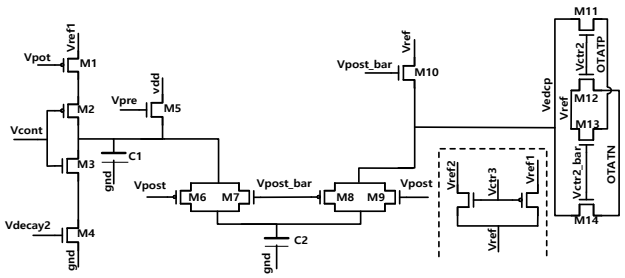

Fig. 2. Circuit schematic of the second-order time window generator.

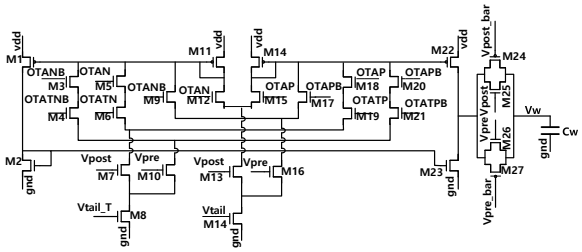

Fig. 3. Circuit schematic of the synapse core.

other for second-order terms. The differential voltage input will be compared and converted to current output. Thus, Vctr2 controls whether the current flow into or out of the synapse weight capacitor Cw. For the second-order terms, both the first-order and second-order time window signals are compared simultaneously to provide high-order potentiation or depression. Thus, the exponential operation from the first- and second-order time window signals are combined. The two transmission gates ensure the weight adjustment only happens when spikes arrive.

With the help of V decays, the circuit can adjust the weight with different decaying rates. Higher V decay leads to smaller weight changes with the same time difference. What is more, different V r e f s enable the adjustment of time constants, and V tail in synapse core helps with the reconfigurability of A_1 and A_2 in formula (1). Thus, besides the switching between different TSTDP types, the reconfigurability is also shown in the adjustment of decaying rate, weight changing rate, and weight changing amplitude.

III. Simulation and Testing Results of the Triplet STDP Circuit

A. Performance of STDP training with encoding schemes

To verify the training accuracy of the STDP rules, neu-ral networks with different encoding schemes are imple-mented with Pytorch and hardware testbench. The MNIST and CIFAR-10 datasets are utilized to verify that the STDP training can work with various encoding schemes and provide good performance. The datasets are first transmitted into the microcontroller from a personal computer (PC), then converted to electric signals and inputted into a chip with an encoder block, as shown in Fig. 4(a). After the encoding process, the encoded spike signal will be transferred back to the PC through the microcontroller for the following training

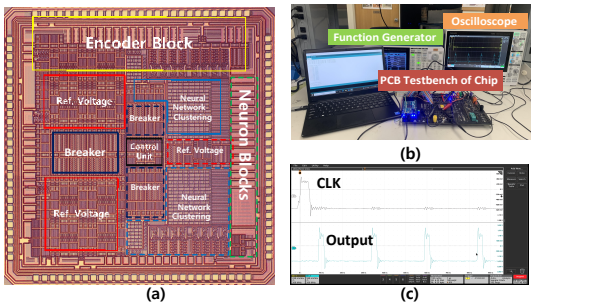

Fig. 4. (a) Die micrograph of the encoder chip fabricated in 180nm. (b) Hardware part of the STDP training and encoding testbench. (c) The measurement result of the hardware testbench.

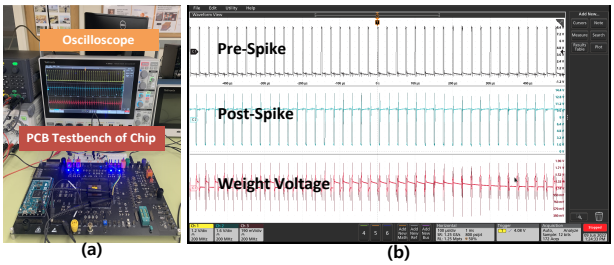

Fig. 5. (a) Prototype of the TSTDP training circuit. (b) Measurement result of the TSTDP prototype.

process. The hardware testbench is demonstrated in Fig. 4(b). In Fig. 4(c), the measurement results of the encoder block on the chip are illustrated. It shows the spike train in one sampling window. What is more, a hardware testbench of the TSTDP prototype is also implemented. As shown in Fig. 5(a), it utilizes the neuron blocks on the chip and realizes the weight-changing function with appropriate components such as capacitors and transistors. The measurement result of the prototype is depicted in Fig. 5(b), where the first two channels show the pre-and post-neuron spike, and the third channel shows the corresponding weight voltage. As shown in Fig. 5(b), the weigh-changing frequency can be as high as 40kHz.

After the encoded spikes are transmitted back to the PC, spikes are utilized in various encoding schemes. The TTFS and ISI encoding schemes are multiplexed with a larger timescale phase of firing encoding scheme, forming the multiplexing TTFS-phase and ISI-phase encoding schemes. Those encoded datasets are then fed into deep spiking neural networks with pair and triplet STDP training methods based on SpykeTorch [13]. The hardware and software hybrid testbench is then being operated with MNIST and CIFAR-10 datasets. Table I demonstrates the accuracy of PSTDP and TSTDP working with different encoding schemes when evaluated with the MNIST and CIFAR-10 datasets. The TSTDP training method achieves 96.5% accuracy for MNIST when working with the ISI-phase encoding scheme and demonstrates 3.28% more accuracy than PSTDP and at most 4.9% more accuracy than other codes (TTFS code). As for the CIFAR-10 dataset, the TSTDP algorithm gets 84.35% of classification accuracy and thus achieves 3.63% of more accuracy than PSTDP and at most 14.05% of more accuracy than other schemes (TTFS

TABLE I
ACCURACY OF PSTDP AND TSTDP TRAINING WITH DIFFERENT ENCODING SCHEME FOR MNIST AND CIFAR-10 DATASETS

PSTDP	TTFS	TTFS-phase	ISI	ISI-phase
MNIST	88.88%	92.5%	91.8%	93.78%
CIFAR-10	67.3%	74.45%	75.3%	80.72%
TSTDP	**TTFS**	**TTFS-phase**	**ISI**	**ISI-phase**
MNIST	91.1%	93.2%	92.8%	96.5%
CIFAR-10	70.3%	74.11%	78.2%	84.35%

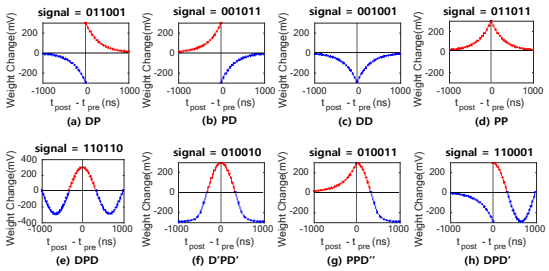

Fig. 6. Simulation results of different STDP shapes with different control signals

encoding). In conclusion, the TSTDP learning rule working with the multiplexing encoding scheme provides better performance than the PSTDP learning rule and other basic temporal encoding schemes.

B. Performance Analysis of the Triplet Reconfigurable STDP Circuit.

In this section, the result of the STDP circuit, the silicon area and the power will be simulated and summarized based on the post-layout simulation. As mentioned in Section.II.C, the learning rule of the reconfigurable STDP circuit can be controlled by digital signals. As depicted in Fig. 6, the post-layout simulation shows several STDP curve shapes with different control signals. For instance, with signal 011001, the circuit provides the basic asymmetric STDP learning rule, also called DP. With signal 011011, the STDP rule implemented is the symmetric rule, the PP. The x- and y-axis represent the time difference between spikes and weight changes.

Fig. 7 demonstrates the layout of the triplet reconfigurable STDP circuit. As depicted in this figure, the circuit consists of four time window generators and one synapse core. Each time window generator takes approximately $132 \times 48\mu m^2$ of area. It is noticeable that the comparators take most of the area in the time window generators. As for the synapse core, it takes less area than other parts, only $24\times86\mu m^2$. With these blocks, the total area of the STDP circuit is $281 \times 162\mu m^2$, which is around $0.045mm^2$.

Similarly, the static power consumption of the circuit can also be divided into five parts. Each time window generator consumes $61\mu W$ of power. With $14\mu W$ of the synapse core power consumption, the total power of the triplet reconfigurable STDP circuit is $258\mu W$. What is more, the comparison of the proposed work and other state of the arts of the STDP

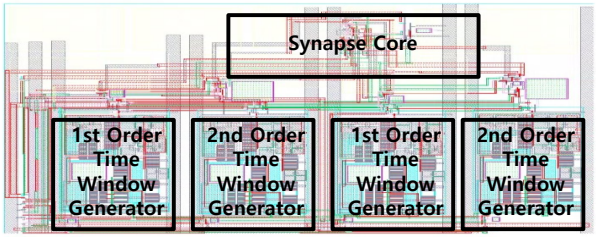

Fig. 7. Layout of the triplet reconfigurable STDP circuit.

TABLE II
COMPARISON OF THIS WORK WITH OTHER STATE OF THE ART WORKS

	[14]	[6]	[8]	This work
Technology	180nm	65nm	10nm	180nm
Supply Voltage (V)	1.8	1.2	0.525	1.8
STDP Types	2	8	1	16
Area (mm^2)	0.006	N/A	0.003	0.045
Energy/SOP (fJ)	1.2×10^7	400	3.8×10^3	775

circuit is illustrated in Table II. The STDP types used in [14] are basic DP and DPD training algorithm. [6] has better reconfigurability. It can switch between DP, DPD, DPD', DD, PD, PPD, DPD" and PP algorithms. As for [8], it only utilized the basic DP algorithm in the training process. In terms of learning performance, only [8] has reported the training accuracy. It achieved 97.9% of training accuracy on MNIST dataset with much better technology node and way higher energy per spike operation (SOP) while this work has achieved 96.5% with lower energy/SOP. The Energy/SOP represents the dynamic energy that is consumed for one spike operation for these works. Since this design is based on SNN, most components are triggered only when spikes come. What is more, the minimum transistor sizes are selected that still keep the circuit functional also help with power efficiency. Therefore, the comparison result shows that even with not that advanced technology, this introduced work still provides one of the best reconfigurability while keeping a relatively low energy per SOP.

IV. CONCLUSION

In this paper, we presented a novel design of the triplet reconfigurable STDP circuit that integrates various STDP types with the triplet STDP rule to take account of three spikes and adjust the synaptic weight. We have also implemented the hardware prototype of the proposed circuit. With performance analysis implemented in hardware testbench and Pytorch training, TSTDP training rule working with multiplexing encoding schemes of MNIST and CIFAR-10 datasets has shown 3.28% and 3.63% higher accuracy than PSTDP training rule and 4.9% and 14.05% higher accuracy than working with other encoding schemes. To the best of our knowledge, this circuit is the first IC implementation of the triplet reconfigurable STDP synapse circuit with $0.045mm^2$ of area and $258\mu W$ of power consumption. Compared with other state of the arts, it shows the best reconfigurability while keeping $775fJ$, a reasonable energy per SOP.

REFERENCES

[1] W. Maass, "Networks of spiking neurons: the third generation of neural network models," Neural networks, vol. 10, no. 9, pp. 1659–1671, 1997.

[2] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain et al., "Loihi: A neuromorphic manycore processor with on-chip learning," Ieee Micro, vol. 38, no. 1, pp. 82–99, 2018.

[3] R. Gaurav, T. C. Stewart, and Y. C. Yi, "Spiking reservoir computing for temporal edge intelligence on loihi," in 2022 IEEE/ACM 7th Symposium on Edge Computing (SEC). IEEE, 2022, pp. 526–530.

[4] C. Zhao, Q. An, K. Bai, B. Wysocki, C. Thiem, L. Liu, and Y. Yi, "Energy efficient temporal spatial information processing circuits based on stdp and spike iteration," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 67, no. 10, pp. 1715–1719, 2019.

[5] J. Gjorgjieva, C. Clopath, J. Audet, and J.-P. Pfister, "A triplet spike-timing–dependent plasticity model generalizes the bienenstock–cooper–munro rule to higher-order spatiotemporal correlations," Proceedings of the National Academy of Sciences, vol. 108, no. 48, pp. 19 383–19 388, 2011.

[6] C. Huan, Y. Wang, X. Cui, and X. Zhang, "A reconfigurable mixed signal cmos design for multiple stdp learning rules," in 2020 IEEE 3rd International Conference on Electronics Technology (ICET). IEEE, 2020, pp. 639–643.

[7] B. Guo, Y. Cai, Y. Pan, Z. Zhang, Y. Fang, and R. Huang, "Associative learning based on symmetric spike time dependent plasticity," in 2014 12th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). IEEE, 2014, pp. 1–3.

[8] G. K. Chen, R. Kumar, H. E. Sumbul, P. C. Knag, and R. K. Krishnamurthy, "A 4096-neuron 1m-synapse 3.8-pj/sop spiking neural network with on-chip stdp learning and sparse weights in 10-nm finfet cmos," IEEE Journal of Solid-State Circuits, vol. 54, no. 4, pp. 992–1002, 2018.

[9] Y. Hao, X. Huang, M. Dong, and B. Xu, "A biologically plausible supervised learning method for spiking neural networks using the symmetric stdp rule," Neural Networks, vol. 121, pp. 387–395, 2020.

[10] J.-P. Pfister and W. Gerstner, "Triplets of spikes in a model of spike timing-dependent plasticity," Journal of Neuroscience, vol. 26, no. 38, pp. 9673–9682, 2006.

[11] B. J. Kailath et al., "Bistable-triplet stdp circuit without external memory for integrating with silicon neurons," in 2021 IEEE World AI IoT Congress (AIIoT). IEEE, 2021, pp. 0297–0302.

[12] T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri, and B. Linares-Barranco, "Stdp and stdp variations with memristors for spiking neuromorphic learning systems," Frontiers in neuroscience, vol. 7, p. 2, 2013.

[13] M. Mozafari, M. Ganjtabesh, A. Nowzari-Dalini, and T. Masquelier, "Spyketorch: Efficient simulation of convolutional spiking neural networks with at most one spike per neuron," Frontiers in neuroscience, p. 625, 2019.

[14] F. L. M. Huayaney, S. Nease, and E. Chicca, "Learning in silicon beyond stdp: a neuromorphic implementation of multi-factor synaptic plasticity with calcium-based dynamics," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 63, no. 12, pp. 2189–2199, 2016.

[15] F. Nowshin and Y. Yi, "Memristor-based deep spiking neural network with a computing-in-memory architecture," in 2022 23rd International Symposium on Quality Electronic Design (ISQED). IEEE, 2022, pp. 1–6.

[16] H. Zheng, K. J. Bai, and Y. Yi, "Enabling a new methodology of neural coding: Multiplexing temporal encoding in neuromorphic computing," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2023.

[17] K. Bai, L. Liu, and Y. Yi, "Spatial-temporal hybrid neural network with computing-in-memory architecture," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 7, pp. 2850–2862, 2021.