Neural Networks 149 (2022) 184–194

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Modeling learnable electrical synapse for high precision spatio-temporal recognition☆

Zhenzhi Wu [a,1], Zhihong Zhang [b,1], Huanhuan Gao [b], Jun Qin [b], Rongzhen Zhao [a], Guangshe Zhao [b], Guoqi Li [c,d,*]

[a] Lynxi Technologies, Beijing 100097, China
[b] School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[c] Institute of Automaton, Chinese Academy of Science, Beijing 100190, China
[d] University of Chinese Academy of Science, Beijing 100190, China

ARTICLE INFO

Article history:
Received 12 June 2021
Received in revised form 30 November 2021
Accepted 6 February 2022
Available online 11 February 2022

Keywords:
Electrical synapse/coupling
Leaky-integrate-and-fire model
Spatio-temporal information
Bio-plausible neuronal dynamics

ABSTRACT

Bio-inspired recipes are being introduced to artificial neural networks for the efficient processing of spatio-temporal tasks. Among them, Leaky Integrate and Fire (LIF) model is the most remarkable one thanks to its temporal processing capability, lightweight model structure, and well investigated direct training methods. However, most learnable LIF networks generally take neurons as independent individuals that communicate via chemical synapses, leaving electrical synapses all behind. On the contrary, it has been well investigated in biological neural networks that the inter-neuron electrical synapse takes a great effect on the coordination and synchronization of generating action potentials. In this work, we are engaged in modeling such electrical synapses in artificial LIF neurons, where membrane potentials propagate to neighbor neurons via convolution operations, and the refined neural model ECLIF is proposed. We then build deep networks using ECLIF and trained them using a back-propagation-through-time algorithm. We found that the proposed network has great accuracy improvement over traditional LIF on five datasets and achieves high accuracy on them. In conclusion, it reveals that the introduction of the electrical synapse is an important factor for achieving high accuracy on realistic spatio-temporal tasks.

1. Introduction

Along with the dividend of Deep Learning (DL) runs out, an increasing number of researchers resort to neuroscience and cognitive science to inject extra vitality into this artificial intelligence wave (Fedus et al., 2020; Lillicrap, Santoro, Marris, Akerman, & Hinton, 2020; Zeng, Chen, Cui, & Yu, 2019; Zhang et al., 2020; Zhao, Zhang, Lu, Cheng, Si, & Feng, 2020). The development of Artificial Neural Networks (ANNs), a key ingredient of DL, is also deriving brand new inspirations from them. Being a bridge of neuroscience and ANNs, Spiking Neural Networks (SNNs), as the brain-inspired neural network models, are valuable of concern. In these models, neurons are sparsely connected and communicate via spike trains (Ghosh-Dastidar & Adeli, 2009; Schliebs & Kasabov, 2013).

Many valuable SNN oriented concepts may contribute to DL, especially including the biological neuronal models with rich temporal dynamics, the bio-plausible localized learning methods, and the low power processing with structural sparsity and event-driven mechanism. As a result, ANNs with features imitated from SNNs improve considerably in computation efficiency of spatio-temporal processing applications, e.g. action recognition and video reconstruction (Han, Ankit, Sengupta, & Roy, 2017; Illing, Gerstner, & Brea, 2019; Pfeiffer & Pfeil, 2018; Tavanaei, Ghodrati, Kheradpisheh, Masquelier, & Maida, 2019; Wozniak, Pantazi, Bohnstingl, & Eleftheriou, 2020).

In SNN, there are many neuron models including LIF (Abbott, 1999; Lapique, 1907), Izhikevich (Izhikevich, 2004) and Hodgkin–Huxley (Hodgkin & Huxley, 1952), in which the LIF model is the most popular one for its simplicity and efficiency. The traditional LIF has many limitations and therefore many generalized LIF models are proposed. Among them, conductance-based (exponential or alpha type) accumulation is introduced in addition

☆ The work was partially supported by National Key R&D program (2020AAA0109100), and National Science Foundation of China (61876215), and Beijing Academy of Artificial Intelligence (BAAI), and Beijing Science and Technology Plan, China (Z191100007519009) , and a grant from the Institute for Guo Qiang of Tsinghua University.

* Corresponding author.
E-mail addresses: zhenzhi.wu@lynxi.com (Z. Wu), zzh159@stu.xjtu.edu.cn (Z. Zhang), gaohuanhuan@stu.xjtu.edu.cn (H. Gao), qinjun@stu.xjtu.edu.cn (J. Qin), rongzhen.zhao@lynxi.com (R. Zhao), zhaogs@mail.xjtu.edu.cn (G. Zhao), guoqi.li@ia.ac.cn (G. Li).

[1] Contributed equally to this work.

https://doi.org/10.1016/j.neunet.2022.02.006
0893-6080/

to the traditional current-based accumulation. For spike initiation, Quadratic Integrate and Fire (QIF) model (Brunel & Latham, 2003) and Exponential Integrate and-Fire (EIF) model (Fourcaud-Trocmé, Hansel, Van Vreeswijk, & Brunel, 2003) are proposed for describing the nonlinear spike-generating currents of biological neurons. The adaptive exponential integrate-and-fire model (aEIF) (Brette & Gerstner, 2005) concludes several theoretical findings including smooth spike initiation zone, subthreshold resonances, and conductance injection to a single model. There are also a few works on electrical synapse modeling, however, most of them are bio-oriented simulations. Among them, Chow et al. (Chow & Kopell, 2000) introduced an electrical coupling LIF model where the coupling is modeled by voltage interchange among cells. Literature (Ferré et al., 2015; Jordan, Helias, Diesmann, & Kunkel, 2020) proposed a gap-junction model in a large SNN network simulation. Some works reveal the evidence of plasticity of electrical synapses (Curti & O'Brien, 2016; O'brien, 2014) in biological nervous systems, wherein the conductance changing of the gap junction is investigated. In these electrical synapse models, the model parameters are measured from biological experiments, and most of the works lack of the investigation of the learning methods and the network capability on spatio-temporal tasks.

High precision spatio-temporal recognition requires for advanced learning approaches. Traditionally, bio-plausible synaptic plasticity approaches are applied for LIF network learning, where Spike-Timing-Dependent-Plasticity (STDP) (Diehl & Cook, 2015) and reward-modulated STDP (Mozafari, Ganjtabesh, Nowzari-Dalini, Thorpe, & Masquelier, 2019) are the most widely used ones. With the rapid development of deep neural network training, several conversion-based methods (Khoei, Yousefzadeh, Pourtaherian, Moreira, & Tapson, 2020; Rueckauer, Lungu, Hu, Pfeiffer, & Liu, 2017) are proposed which can convert trained high accuracy deep neural networks to SNN. In recent years, direct end-to-end training methods become popular. Among them, Spikeprop (Bohte, Kok, & La Poutré, 2000) introduces backpropagation to spiking neural network training. Spatio-Temporal Back Propagation (STBP) (Wu, Deng, Li, Zhu, & Shi, 2018) is proposed for convolutional LIF network training. LIAF-Net (Wu, Zhang, Lin, Li, Wang, & Tang, 2021) introduces analog activation of neurons and enables ANN layers to be introduced to the LIF network. These training methods enable LIF SNN to be applied in many spatio-temporal applications (Cannici, Ciccone, Romanoni, & Matteucci, 2019). However, a limited discussion is revealed for the learning of networks built with electrical synapses.

An electrical synapse is a mechanical and electrically conductive link between two neighboring neurons that is formed at a gap junction, shown in Fig. 1. Formed by gap junction channels between neurons, electrical synapses allow direct transmission of voltage signals between coupled cells (Curti & O'Brien, 2016). Such electrical coupling of neurons receives much attention in recent neuroscience research. It is proved to be responsible for a variety of network effects particularly in networks that generate rhythmic activity such as regulation of phase relationships, synchrony, and pattern formation (Nadim, Li, Gray, & Golowasch, 2017). It has been revealed that the electrical coupling promotes action potential generation and synchronous firing, and electrically coupled interneurons exhibit strong synchronous synaptic activity (Yao et al., 2016).

The electrical synapse is also essential for visual sense, which has been investigated in neuroscience researches (Boron & Boulpaep, 2016a, 2016b, 2016c). The retina neurons receive optical stimuli in a matrix, and through some levels of neural populations, signals are projected to the primary visual cortex point-to-point. In such nerve routes, both chemical and electrical synapses exist universally and play crucial parts in visual signal processing.

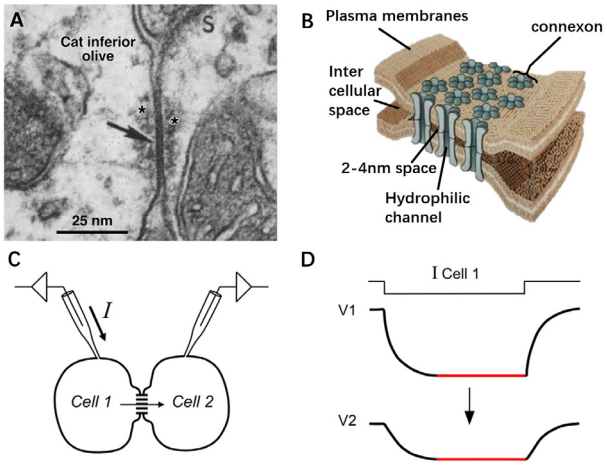

Fig. 1. Biological evidence for electrical coupling. (A) Electron micrograph showing a neuronal gap junction (arrow) between two dendritic profiles in the cat inferior olive, which were located in an extraglomerular position. Modified from Sotelo et al. 1974. (B) Diagram of inter-neuron gap junction where electrical coupling takes place. (C) Experimental design for study electrophysiological properties of electrical synapses. (D) When a hyperpolarizing current pulse is injected to cell 1 (I Cell 1) a voltage deflection is produced in that cell (V1) and also in cell 2 (V2), although voltage change in the latter is of smaller amplitude. Traces are representative drawings. Modified from Curti and O'Brien (2016).

Especially, chemical synapses exchange information between different levels of populations; electrical synapses, forming among somas very close to one another and exchange information locally via gap junctions, which is responsible for a variety of effects including action potential generation and synchronous firing (Bennett, 2000; Pereda, 2014).

In this work, we try to build an Electrical Coupling LIF (ECLIF) model and correspond to deep neural networks, wherein the densely allocated neurons communicate with each other via both chemical synapses and electrical synapses. For the electrical synapse, the electrical coupling effect is modeled that the membrane potentials of one neuron will take effect on the membrane potentials of its neighbor neurons.

The main contribution of this paper is two folds:

(1) We bring in the bio-plausible electrical synapse to learnable LIF, and a newly revised LIF model called ECLIF is proposed. Since the traditional EC model is established in neural dynamic equations and defined for an independent neuron, we found out that convolution operation can be applied to simplify representing the electrical coupling effect for a group of neurons in the tensor shape. Such a model obeys the original biological dynamic expression, whereas is more friendly for deep neural network construction and training. Hence the coupling weights now can be learned via the BPTT algorithm.

(2) We enhance existing deep ANNs with ECLIF, and a deep neural network named ECLIF-Net is built with such ECLIF layers for spatio-temporal processing. For enhancing the network accuracy, we applied several key features including analog action potential, neural homeostasis (via batch normalization), and pooling to ECLIF-Net, for achieving a better network accuracy on realistic tasks. For revealing the efficiency of EC, we conducted five experiments with the proposed ECLIF-Net, which shows obvious performance gain is achieved on the mainstream datasets. We also compared the computational efficiency of the proposed model with the traditional LIF model, 3D convolutional model, and ConvLSTM model which reveals the computational efficiency of ECLIF-Net.

The paper is organized as follows. In Section 2, the model description of the proposed ECLIF is introduced. In Section 3,

we evaluated the performance of ECLIF-Net with datasets and compared it with other traditional networks. The experiment proved that ECLIF has excellent spatio-temporal processing capabilities. The computational complexity is analyzed for each task, which also reveals the efficiency of ECLIF-Net. Finally, Section 4 concludes the paper.

2. Model description

2.1. Traditional LIF

The original LIF model is described in a differential function (Ferré, Mamalet, & Thorpe, 2018; Roy, Jaiswal, & Panda, 2019) to reveal the neuronal dynamic, following equation

$$\tau \frac{dV_j(t)}{dt} = -(V_j(t) - V_{rest}) + RI_j(t). \tag{1}$$

where τ is the timing factor of the neuron, j represents the current neuron, V_{rest} is the rest potential. When $V_j(t)$ reaches a certain threshold V_{th}, a spike is emitted, and the $V_j(t)$ is reset to its initial value V_{reset}. R is the resistance. $I_j(t)$ is the injected current, which can be described by $I_j(t) = \sum_{i=1}^{n} W_{i,j} \cdot X_i(t)$, $X_i(t)$ is the input signal (spike or none) from the ith neuron connecting to the dendrite of current neuron through a synapse with strength $W_{i,j}$.

2.2. Electrical coupling LIF

We introduce the electrical coupling effect illustrated by Nadim et al. (2017) to the LIF model. The effect is observed in vivo that the action potential originates in a neuron cell and propagates through the electrically coupled neurons. The coupling is obliged for the voltage change and the new action potentials in the coupled neighboring cells.

We define $N\{j\}$ as the set of densely allocated neighbor neurons to neuron j, indexed by $k \in N\{j\}$ (excluding neuron j itself), their membrane potentials denoted by $V_k(t)$, and the membrane potential impact they have taken on the current neuron j is related by weights $W^R_{k,j}$ and membrane potential difference $V_k(t) - V_j(t)$ (Ferré et al., 2015). After electrical coupling is introduced, Eq. (1) is refined as

$$\begin{aligned}\tau \frac{dV_j(t)}{dt} =& -(V_j(t) - V_{rest}) + \\ & \sum_{k\in N\{j\}} W^R_{k,j} \cdot (V_k(t) - V_j(t)) + RI_j(t)\end{aligned} \tag{2}$$

For maintaining the spatial invariance, we use a convolution operation to better represent the coupling effect, following

$$Conv_j(V_k(t), W^R_{k,j}) = \sum_{k\in N\{j\}} W^R_{k,j} \cdot (V_k(t) - V_j(t)) \tag{3}$$

2.3. Detail derivation of the discrete form of ECLIF

Since the BPTT training needs a discrete-time iterative representation of ECLIF, we introduce Euler method (Neftci, Mostafa, & Zenke, 2019; Wu et al., 2018) to obtain it.

From Eq. (2), we define Δt as the sampling duration which is a small fraction of time, with the Euler method the equation can be solved numerically

$$\begin{aligned}V_j(t + \Delta t) =& V_j(t) + \\ & \frac{\Delta t}{\tau}\left(-V_j(t) + V_{rest} + RI_j(t)\right) + \\ & \frac{\Delta t}{\tau}\left(\sum_{k\in N\{j\}} W^R_{k,j} \cdot \left(V_k(t) - V_j(t)\right)\right)\end{aligned} \tag{4}$$

After sampling with sampling rate $1/\Delta t$, and we denote the time step as n, and $t = n\Delta t$ then we have

$$\begin{aligned}V_j((n+1)\Delta t) =& V_j(n\Delta t)(1 - \frac{\Delta t}{\tau}) + \\ & \frac{\Delta t}{\tau} V_{rest} + \\ & \frac{\Delta t}{\tau} RI_j(n\Delta t) + \\ & \frac{\Delta t}{\tau} \sum_{k\in N\{j\}} W^R_{k,j} \cdot \left(V_k(n\Delta t) - V_j(n\Delta t)\right)\end{aligned} \tag{5}$$

For simplicity, we further define $\alpha = 1 - \frac{\Delta t}{\tau}$, $\beta = \frac{\Delta t}{\tau} V_{rest}$, $r = R\frac{\Delta t}{\tau}$, $W^r_{k,j} = \frac{\Delta t}{\tau} W^R_{k,j}$, then Eq. (5) can be written as

$$\begin{aligned}V_j((n+1)\Delta t) =& \alpha V_j(n\Delta t) + \beta + rI_j(n\Delta t) + \\ & \sum_{k\in N\{j\}} W^r_{k,j} \cdot \left(V_k(n\Delta t) - V_j(n\Delta t)\right)\end{aligned} \tag{6}$$

In the discrete form, we skip the notation Δt, therefore

$$V^{n+1}_j = \alpha V^n_j + \beta + rI^n_j + \sum_{k\in N\{j\}} W^r_{k,j} \cdot (V^n_k - V^n_j) \tag{7}$$

we further define

$$V^n_{Lj} = \alpha V^n_j + \beta \tag{8}$$

$$V^n_{ECj} = V^n_{Lj} + \sum_{k\in N\{j\}} W^r_{k,j} \cdot (V^n_{Lk} - V^n_{Lj}) \tag{9}$$

$$V^n_{Ij} = V^n_{ECj} + rI^n_j \tag{10}$$

From Eq. (9) we have

$$\begin{aligned}V^n_{ECj} =& (1 - \sum_{k\in N\{j\}} W^r_{k,j}) V^n_{Lj} + \sum_{k\in N\{j\}} W^r_{k,j} \cdot V^n_{Lk} \\ =& W^r_{j,j} V^n_{Lj} + \sum_{k\in N\{j\}} W^r_{k,j} \cdot V^n_{Lk} \\ =& \sum_{k\in N\{j\},j} W^r_{k,j} \cdot V^n_{Lk}\end{aligned} \tag{11}$$

where $W^r_{j,j} = 1 - \sum_{k\in N\{j\}} W^r_{k,j}$ is defined.

For maintaining the spatial invariance, the weight only relates to the spatial location relationship between these two neurons, and only the adjacent neurons (within the coupling window for neuron j) are considered, therefore we can use a convolutional kernel to simplify Eq. (11), following

$$\begin{aligned}V^n_{ECj} =& \sum_{k\in\{N\{j\},j\}} W^r_{k,j} \cdot V^n_{Lk} \\ =& Conv_j(V^n_{Lk}, W^r_{k,j})\end{aligned} \tag{12}$$

In the next section, We will use corresponding vector (or tensor) representations, i.e. $\boldsymbol{V}^n_L$, $\boldsymbol{V}^n_{EC}$ and $\boldsymbol{V}^n_I$, to denote group of neurons, then $\boldsymbol{V}^n_{EC} = Conv(\boldsymbol{V}^n_L, \boldsymbol{W}^r)$.

Note that the $Conv_j$ here is slightly different to Eq. (3). For a compatible version to Eq. (3), we can define $W^{r'}_{j,j} = -\sum_{k\in N\{j\}} W^r_{k,j}$, then Eq. (11) can also be rewritten by $V^n_{ECj} = V^n_{Lj} + Conv_j(V^n_{Lk}, W^{r'}_{k,j})$.

2.4. Electrical coupling LIF for 1D signal processing

Form the deviation in Section 2.3, the discrete ECLIF model can be described in an iterative process consisting of 5 steps as follows. Here, we define the input and output signal in each time step as a 1D vector signal, and the number of ECLIF neurons is L.

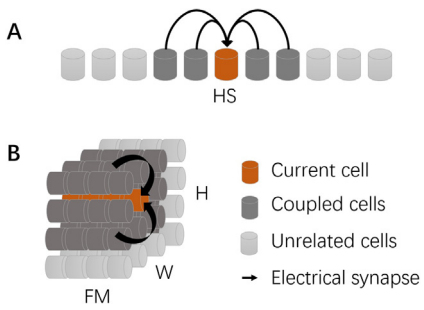

Fig. 2. Proposed biological inspired ECLIF model. (A) Proposed 1D ECLIF, wherein the membrane potential of a yellow color neuron is impacted by the neurons in gray, modeled by 1D convolution. (B) Proposed 2D ECLIF, wherein the nearby $K \cdot K \cdot FM$ neurons (FM is the feature maps) will influence the membrane potential of the current neuron, modeled by 2D convolution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(a) Perform leakage

$$\boldsymbol{V}_{\boldsymbol{L}}^{n} = \boldsymbol{\alpha} \cdot \boldsymbol{V}^{n} + \boldsymbol{\beta} \tag{13}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent the multiplicative decay and additive decay respectively.

(b) Perform electrical coupling on inter-neurons for ECLIF

$$\boldsymbol{V}_{EC}^{n} = Conv1D(\boldsymbol{V}_{L}^{n}, \boldsymbol{W}^{r}) \tag{14}$$

where the $Conv1D$ is 1D Convolution with kernel size $1 \times K$. We name $\boldsymbol{W}^{r}$ as coupling kernel, which is learnable. For traditional LIF, we have $\boldsymbol{V}_{EC}^{n} = \boldsymbol{V}_{L}^{n}$.

(c) Accumulates with previous membrane potential

$$\boldsymbol{V}_{I}^{n} = r\boldsymbol{I}^{n} + \boldsymbol{V}_{EC}^{n} \tag{15}$$

where $\boldsymbol{V}_{EC}^{n}$ and $\boldsymbol{V}_{\boldsymbol{I}}^{n}$ refer to the previous and current membrane potential respectively.

(d) Compare with the threshold and fire

$$\boldsymbol{F}^{n} = \boldsymbol{V}_{I}^{n} \geq \boldsymbol{V}_{th} \tag{16}$$

where $\boldsymbol{F}^{n}$ is the fire signal. For each F_{j}^{n} in $\boldsymbol{F}^{n}$, $F_{j}^{n} = 1$ indicates a firing event, otherwise $F_{j}^{n} = 0$.

(e) Reset the membrane potential when fired

$$\boldsymbol{V}^{n+1} = \boldsymbol{F}^{n} \cdot \boldsymbol{V}_{reset} + (\boldsymbol{1} - \boldsymbol{F}^{n}) \cdot \boldsymbol{V}_{I}^{n} \tag{17}$$

2.5. ConvECLIF2D for spatio-temporal processing

For spatio-temporal processing, where the input is a sequence of frames, and the spatial domain is 2D images, therefore the 1D ECLIF model proposed above needs to be evolved to 2D format.

Firstly, the synaptic integration for $I_j(n)$ is described by 2D convolution, following

$$\boldsymbol{I}^{n} = Conv2D(\boldsymbol{X}^{n}, \boldsymbol{W}) \tag{18}$$

where $\boldsymbol{X}^{n}$ stands for the activations from the presynaptic neurons, $\boldsymbol{W}$ refers to the chemical synaptic weights. The synaptic integration can exist in a fully connected format or convolutional format. Both $\boldsymbol{X}^{n}$ and $\boldsymbol{I}^{n}$ are 3-dimensional tensors, e.g. $\boldsymbol{X}^{n}$ has the shape of (H, W, FM_{in}), and $\boldsymbol{I}^{n}$ has the shape of (H, W, FM) where H and W represent the height and width of the feature maps, FM_{in} is the number of input feature maps. FM is the number of feature maps of output activations. Accordingly the membrane potential V_{I}^{n}, firing signal F^{n}, Electrical Coupling (EC) effect V_{EC}^{n} are all 3D tensors with shape (H, W, FM_{in}). The electrical coupling effect described in (14) is substituted by

$$\boldsymbol{V}_{EC}^{n} = Conv2D(\boldsymbol{V}_{L}^{n}, \boldsymbol{W}^{r}) \tag{19}$$

where the $Conv2D$ is 2D Convolution with a coupling kernel size of $K \times K$.

The ECLIF1D and ECLIF2D are depicted and compared in Fig. 2.

2.6. ECLIF with analog activation

Borrowed the ideas from Wu et al. (2021), an alternative model of ECLIF by extending spiking to an arbitrary activation function is proposed. Instead of output $\boldsymbol{F}^{n}$ to succeeding layer, this model output $\boldsymbol{Y}^{n}$ defined as an activation function of membrane potential $\boldsymbol{V}_{I}^{n}$, following

$$Y^{n} = f(\boldsymbol{V}_{I}^{n}, \boldsymbol{V}_{th}) \tag{20}$$

where $f(x, V_{th})$ is the activation function. It can be any threshold-related function or threshold unrelated function, and the output can be in analog or spiking format. We term this model as ECLIF with Activation function (ECLIF-A) model. Traditional spiking ECLIF (termed as ECLIF-S) can be viewed as a special case of ECLIF-A, where $f(x, V_{th}) = x \geq V_{th} ? 1 : 0$.

In conclusion, Alg. 1 summarizes the ECLIF execution flow in pseudo-code. It is worth noting that there are several sub-types of ECLIF, listed in Table 1. In the table, "Direct-" indicates that no synaptic integration is performed, i.e. $\boldsymbol{I}^{n} = \boldsymbol{X}^{n}$; "FC-" means a fully connected layer is used for synaptic integration, i.e. $\boldsymbol{I}^{n} = FullyConnected(\boldsymbol{X}^{n}, \boldsymbol{W})$. For saving parameters, we apply sharing mechanism for $\boldsymbol{V}_{th}$, $\boldsymbol{V}_{reset}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The units for each of them may vary for each FM (the neurons within an FM share a unique value), or the same for all neurons. We termed these modes as Channel-Sharing ("-CS") mode, and All-Sharing ("-AS") mode, respectively. We use Non-Sharing ("-NS") to represent that there is no sharing on parameters. As a result, there are many combinations of these sub-types, and we can denote each model configuration as a "prefix-model-suffix" notation. For example "ConvECLIF2D-S" means an electrical coupling LIF model with 2D convolutional synaptic integration and outputs with spiking format.

Algorithm 1 Pseudo code for one time step (n) of ECLIF.

Input: Synaptic integration current $\boldsymbol{I}^{n}$
Output: $\boldsymbol{Y}^{n}$
$\boldsymbol{V}_{L}^{n} \leftarrow \boldsymbol{\alpha} \cdot \boldsymbol{V}^{n} + \boldsymbol{\beta}$
$\boldsymbol{V}_{EC}^{n} \leftarrow Conv(\boldsymbol{V}_{L}^{n}, \boldsymbol{W}^{r})$
$\boldsymbol{V}_{I}^{n} \leftarrow r\boldsymbol{I}^{n} + \boldsymbol{V}_{EC}^{n}$
$\boldsymbol{F}^{n} \leftarrow \boldsymbol{V}_{I}^{n} \geq \boldsymbol{V}_{th}$
$\boldsymbol{Y}^{n} \leftarrow f(\boldsymbol{V}_{I}^{n}, \boldsymbol{V}_{th})$
$\boldsymbol{V}^{n+1} \leftarrow \boldsymbol{F}^{n} \cdot \boldsymbol{V}_{reset} + (1 - \boldsymbol{F}^{n}) \cdot \boldsymbol{V}_{I}^{n}$

2.7. Biological similarities

The biological evidence of electrical synapse is shown in Fig. 1. Different from biological coupling strength which varies for all pairs of neurons, in our model, the synaptic weights $\boldsymbol{W}^{r}$ are shared for all the neurons due to convolution. Such sharing enables the in-variance of position and reduces the number of overall parameters, which is widely applied in deep neural networks.

2.8. Computational point of view

The membrane potential is in high accuracy and contains additional information than the threshold-controlled firing signals. By introducing the convolution operation, additional computation and storage are applied for enhancing information exchange capability between neurons. It can be viewed as a balance of

Table 1
Configurations for ECLIF model family.

Signal type (shape)	EC		Synaptic integration				Activation		Parameter sharing		
	With EC	Without EC	Direct	Dense	Conv1D	Conv2D	Spiking	Act. fun.	None	Channel	All
1D signal $[T, L]$	ECLIF1D	LIF1D	Direct-	FC-	Conv-	/	-S	-A	-NS	-CS	-AS
2D signal $[T, H, W, FM]$	ECLIF2D	LIF2D	Direct-	/	/	Conv-	-S	-A	-NS	-CS	-AS

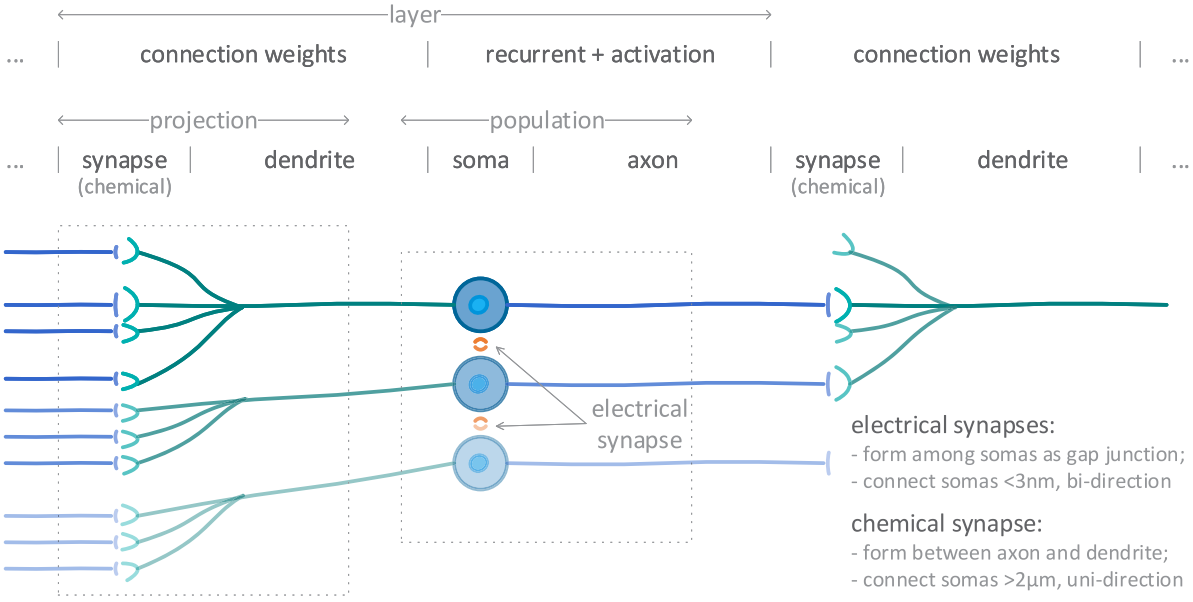

Fig. 3. Modeling neurons as an ANN layer. A neuronal cell consists of the synapse, dendrite, soma, and axon, where synapses can be categorized into electrical and chemical. SNN defines the network projection as a group of synaptic connections and the neural population as a group of cells (each cell includes dendrite, soma, and axon). Previous works revealed that LIF can be used to enhance ANN by replacing the activation function (for temporal free ANN models) and recurrent neural cells (for temporal or spatio-temporal ANN models) with SNN cell models (Deng et al., 2020; Wu et al., 2021), but neglecting electrical synapses, which are as universal and crucial as chemical synapses in information exchange and processing. Our work fixes such defects.

computational cost and recognition accuracy. Although it requires more computations and weight storage, it still shows better computational efficiency than traditional ANNs (Conv3D or ConvLSTM), which will be discussed in Section 3.

2.9. Plasticity and learning

Regarding plasticity, there are at least three categories of mechanisms for changing the strength of electrical coupling between two neurons (O'brien, 2014) in biological nervous systems: (1) by altering membrane properties of the communicating cells, (2) by changing the conductance of the gap junction, and 3) by changing the expression level gap junction proteins. In our simplified model, like most simplified neural models, the details of membrane properties and proteins are not included. The strength of electrical coupling between two neurons is modeled simply by a single value represented as one of the weight parameter $\boldsymbol{W}^r$. Therefore we mainly focus on adjusting the synaptic weights, which are equivalents to an integrated effect of the three categories of plasticity mentioned above. The weights are learned via BPTT in a supervised manner. Training ECLIF is similar to the training of RNN and LIF. The training of LIF-SNN with BPTT with gradient approximation is well discussed in literature (Hong, Wei, Wang, Deng, Yu, & Che, 2019; Lee, Delbruck, & Pfeiffer, 2016; Neftci et al., 2019; Pineda, 1987). The additional EC operation (convolution operation) to LIF-SNN is also differentiable.

2.10. Deep neural network with ECLIF

In this section, we compare ECLIF with typical deep neural network (DNN) layers and explain the way for integrating ECLIF into the DNN training framework.

As for neuroscience, biological brains are described as Spiking Neural Networks (SNNs), where neurons are sparsely connected and activated, and information is carried in binary spiking trains (Ghosh-Dastidar & Adeli, 2009; Schliebs & Kasabov, 2013). As shown in Fig. 3, the synapses in the ECLIF cell model can be treated as the ANN connection weights in the fully connected layers of convolution layers, and the soma in ECLIF can be roughly equivalent to the ANN recurrent cell updates and activation. Hence it is admissible to replace ANN activation functions with ECLIF soma for a temporal-free case. For temporal or spatio-temporal cases, The soma neural dynamics are similar to the recurrent updates held by the "cell" in the recurrent ANN models, such as Recurrent Neural Networks (RNN). Building deep neural networks with ECLIF is as easy as building networks with RNN layers or traditional activation layers.

We further build DNN with ECLIF as the basic building block called ECLIF-Net. In such a network, we use the phrase "ECLIF layer" instead of "ECLIF model". In ECLIF-Net, multiple ECLIF layers may exist. Besides, the DNN layers such as batch normalization, pooling, and fully connected perceptron layers are introduced for the network construction, as we will illustrate in our experiment. The ECLIF-Net can be trained end-to-end with BPTT, and we trained our models with either Tensorflow or Pytorch.

3. Spatio-temporal tasks with ECLIF-net

In this section, we evaluate the performance of ECLIF-Net on several spatio-temporal tasks including two 1D signal processing tasks and three event-driven spatio-temporal tasks. These

Table 2
Performance comparison of temporal models on BABI QA tasks.

Temporal Layer	RNN	LSTM	GRU	FCLIF1D-S	FCECLIF1D-S	FCECLIF1D-A
Avg acc.	54.24%	56.98%	56.93%	49.73%	51.98%	53.93%

Table 3
Efficiency and cost comparison for the candidate temporal models on BABI dataset.

Layer	Test	ADDs	MULs	Weights
LIF	49.73%	0.4M	7.7K	5.5K
ECLIF-S (3)	51.98%	0.4M	31K	5.5K
ECLIF-A (3)	53.93%	0.4M	0.4M	5.5K
RNN	54.24%	1.2M	1.2M	15K
GRU	56.93%	3.5M	3.5M	45K
LSTM	56.98%	4.6M	4.6M	60K

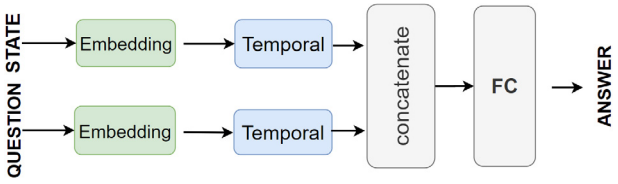

Fig. 4. The network structure designed for the bAbI QA task. One of RNN/LSTM/GRU/LIF/ECLIF is selected as the temporal layer.

experiments cover many areas such as Natural Language Processing (NLP), speech recognition, gesture recognition, moving object classification, and large-scale image classification, which fully demonstrate the spatio-temporal processing capabilities of ECLIF-Net.

3.1. Simple natural language processing with ECLIF

We first tested ECLIF-Net and traditional temporal models on a Question Answering (QA) task on bAbI QA dataset (Weston et al., 2015). bAbI QA contains 20 subtasks, each of which is a group of QA, and each QA contains a statement, a question, and a label (answer). In this work, we used a unified network framework for all temporal models, shown in Fig. 4. The temporal layer is selected from one of GRU (Cho, van Merrienboer, Gülçehre, Bougares, Schwenk, & Bengio, 2014)/LSTM (Hochreiter & Schmidhuber, 1997)/RNN (Mikolov, Karafiát, Burget, Černockỳ, & Khudanpur, 2010)/FCLIF1D (Abbott, 1999)/FCECLIF1D. We need to first encode the states and questions into a one-hot vector with a dimension of 50 and then fed them into the network. The vectors output by the temporal layer are concatenated and then send to a fully connected layer. For comparison fairly, we applied the same parameters (e.g. cell numbers) for all the temporal layer alternatives. We set the coupling kernel size of 1x3. The results are shown in Table 2.

It can be revealed from the table that the accuracy of ECLIF-S and ECLIF-A are better than LIF by 2.25% and 4.20% respectively. This also illustrates the importance of the electrical coupling effect in neuronal information interaction. The performance of ECLIF-S is not comparable with LSTM and GRU currently whereas ECLIF-A reaches similar performance to GRU and LSTM. It is noticeable that ECLIF is a lightweight structure that consumes much less computation and the number of weights compared to them, revealed in Table 3. It can be seen that the parameter amount of ECLIF-A is 63.33% less than that of RNN, and the performance is similar. Compared with LSTM/GRU, it consumes 90.83% / 87.78% less weight parameters respectively. Besides, the computational cost of ECLIF-A is 88.57% less than that of the GRU layer and 91.30% than the LSTM layer.

3.2. Speech recognition on spiking heidelberg digits

3.2.1. Dataset and data pre-processing

Spiking Heidelberg Digits (SHD) (Cramer, Stradmann, Schemmel, & Zenke, 2020) is the first event-driven dataset for speech classification. They are inspired by neuro-physiology and made by converting general audio into spike sequences. It contains 10K high-quality digital recordings of English and German voices of number 0 to 9, resulting in a 20-classification task. Each sample in this dataset is an event stream and each event is represented by a [time stamp, spike location] pair, where the spike location ranges from 0 to 699. In order to facilitate input into the network, we use continuous 600 ms spike events to generate 10-time steps of spike data in each voice sample. Finally, the data format of the input network is [Batchsize, 10, 700], and Batchsize $= 100$.

3.2.2. Network structure

The network structure is shown in Fig. 6. We only use three temporal layers, a linear layer and a Temporal Aggregation Layer (TAL). TAL is applied for integrating temporal information. In order to perform a comparative evaluation, we used FCLIF1D and FCECLIF1D for the temporal layer. According to the format of the dataset after processing, we use FCECLIF1D with a coupling range of 3 (coupling kernel 1x3). CrossEntropy is chosen as the loss function. The network is trained for 100 epochs. The initial learning rate is 0.03, and the learning rate decays to 1/10 at epoch 50 and epoch 80.

3.2.3. Performance analysis

We conducted experiments with varieties of LIF neuron models under the same network structure on this dataset. It can be revealed from Fig. 5 that whether the activation format is spike or analog, the electrical coupling effect takes an obvious effect on accuracy, improved by 11.91% and 8.41% respectively. As a result, the test accuracy of ECLIF-A and ECLIF-S reach up to 91.82% and 90.50% respectively, which are both higher than the existing state-of-the-art work, shown in Table 4.

3.3. Dynamic gesture recognition on DVS128 GESTURE dataset

3.3.1. Dataset and preparation

The DVS128 Gesture (Amir et al., 2017) dataset is recorded by a Dynamic vision sensor (DVS). For DVS, only the intensity change is recorded, which differs from traditional frame-based cameras that are recorded at a fixed frame rate. The dataset includes a set of 11 hand and arm postures with 1342 instances, and 29 subjects were collected under 3 different lighting conditions. Each collection lasts about 6 s. The spike is represented by a quad such as (x, y, ts, pol), where x and y are the spatial coordinates of the spike, ts is the time-stamp of the event, and pol represents the type of light intensity change (lighten to 1 or darken to -1). We have collected events within 25 ms as a frame, and a frame includes only two channels, which represent whether the intensity is enhanced or weakened. The visualization of the captured data is shown in Fig. 7. It is worth noting that there are two classification options for this dataset, and we choose the configuration of 11 classifications.

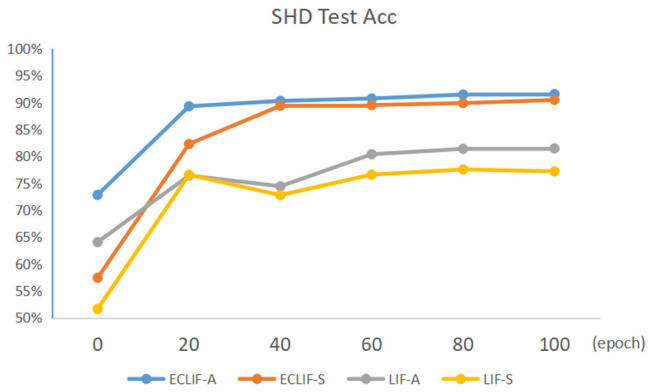

Fig. 5. The test accuracy of the speech recognition experiment on Spiking Heidelberg Digits, which shows the accuracy gain when introducing the electrical coupling effect to LIF.

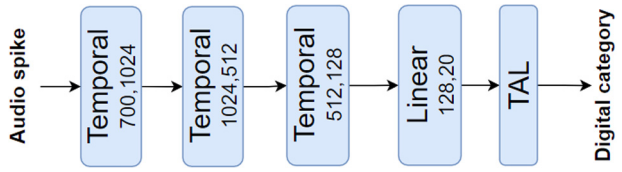

Fig. 6. The evaluation network for the SHD speech recognition task. We performed four experiments with the temporal layer being configured as FCLIF1D-S/FCLIF1D-A/FCECLIF1D-S/FCECLIF1D-A respectively.

Table 4
Speech recognition on SHD dataset.

Proposals	Methods	Acc.
Cramer et al. (2020)	LIF RSNN	71.40%
Yin, Corradi, and Bohté (2020)	RELU SRNN	88.93%
Zenke and Vogels (2021)	SG-base SNN	84.00%
This work	FCECLIF1D-S	**90.50%**
This work	FCECLIF1D-A	**91.82%**

Table 5
Network structure for the DVS128 GESTURE recognition experiment.

Layer	Out channel	Kernel	Use pool	Pool size
ConvECLIF2D	64	(5,5)	False	
	128	(3,3)	True	(2,2)
	128	(3,3)	True	(2,2)
FCLIF1D	256		False	
	11		False	

3.3.2. Network structure

The network structure is composed of three layers of EC neurons and two FCLIF1D layers. The EC block mentioned here is developed based on ConvECLIF2D, which reflects the ability of neurons to obtain spatio-temporal information. The network structure is shown in Table 5. To improve the accuracy of the network during the modeling process, we also borrowed some key features of deep neural networks, including simulated action potentials, neural homeostasis (through batch normalization), and pooling is used to simulate the biological characteristics of neuronal cell synaptic structure.

3.3.3. Network parameter settings

Several hyperparameters are set in this experiment. The time window is set to 60, (α, β) is assigned to $(0.3, 0)$ respectively. V_{th} is set to 0.5. The coupling kernel size can be set to 1x1, 3x3, 5x5. We choose ReLU as the activation function and use Max pooling to extract feature information in the EC block. We choose CrossEntropyLoss as the loss function and trained the networks via Adam optimizer with a learning rate of 10^{-3} and the weight decay of 10^{-4}.

Table 6
Accuracy of solutions for the DVS128 GESTURE dataset (11 classes).

Proposals	Methods	Act.	Acc.
Massa, Marchisio, Martina, and Shafique (2020)	SNN converted from CNN on Loihi	Spike	89.64%
Amir et al. (2017)	CNN on TrueNorth	Spike	94.59%
Kugele, Pfeil, Pfeiffer, and Chicca (2020)	SNN converted from ANN	Spike	95.56%
Wu et al. (2021)	SNN(ConvLIF)	Spike	94.10%
This work	ConvECLIF2D-S	Spike	**97.22%**
Khoei et al., SpArNet 2020 (Khoei et al., 2020)	Converted CNN	Analog	95.10%
Wang, Zhang, Yuan, and Lu (2019)	PointNet++	Analog	95.32%
Bi (2020)	Residual graph CNN + Res.3D	Analog	97.20%
Wu et al. (2021)	LIAF-Net(ConvLIAF)	Analog	97.56%
This work	ConvECLIF2D-A	Analog	**99.65%**

3.3.4. Performance analysis

We evaluate the top 1 accuracy on the test set. The accuracy of ECLIF-A-Net reaches 99.65%, and ECLIF-S-Net reaches 97.22%. We compared this approach with related work on the same dataset, and the accuracy results are listed in Table 6. It reveals that the proposed method achieves the best accuracy. The accuracy achieved by the analog value network has reached the best accuracy, which illustrates the importance of both biological electrical coupling effects and analog activation in neural network transmission.

3.3.5. Efficiency analysis

This section investigates the performance against the power consumption when adjusting the coupling kernel size. The results are listed in Table 7. For the coupling kernel size, we have selected three different configurations of 1x1, 3x3, and 5×5 representing 1/9/25 neurons in the spatial domain that have participated in the EC convolution. The computational cost and the accuracy are listed in Table 7. It can be observed that the best accuracy is achieved when the coupling kernel size is 3×3. However, a better trade-off of performance and computational cost is attained when the coupling kernel size is set to 1×1 (which indicates that all the neurons that at the same spatial position but in different feature maps are participated in the EC convolution), which is an economical choice. Further increasing the coupling kernel size to 5×5 does not contribute to the performance gain. Besides, we also conducted a performance investigation of the ECLIF 1×1 kernel size version with other network structures in Table 8. This network can still be accurate and precise among various structures.

3.4. Moving object classification on CIFAR10-DVS dataset

3.4.1. Dataset and preparation

We conducted our experiment on another neuromorphic dataset CIFAR10-DVS (Li, Liu, Ji, Li, & Shi, 2017) to verify the generality of ECLIF-NET. A sample in this event-based dataset is recorded by DVS as spatio-temporal spike train which contains abundant temporal information. The spike train is represented by a quad such as (x, y, ts, pol). The CIFAR10-DVS dataset consists of

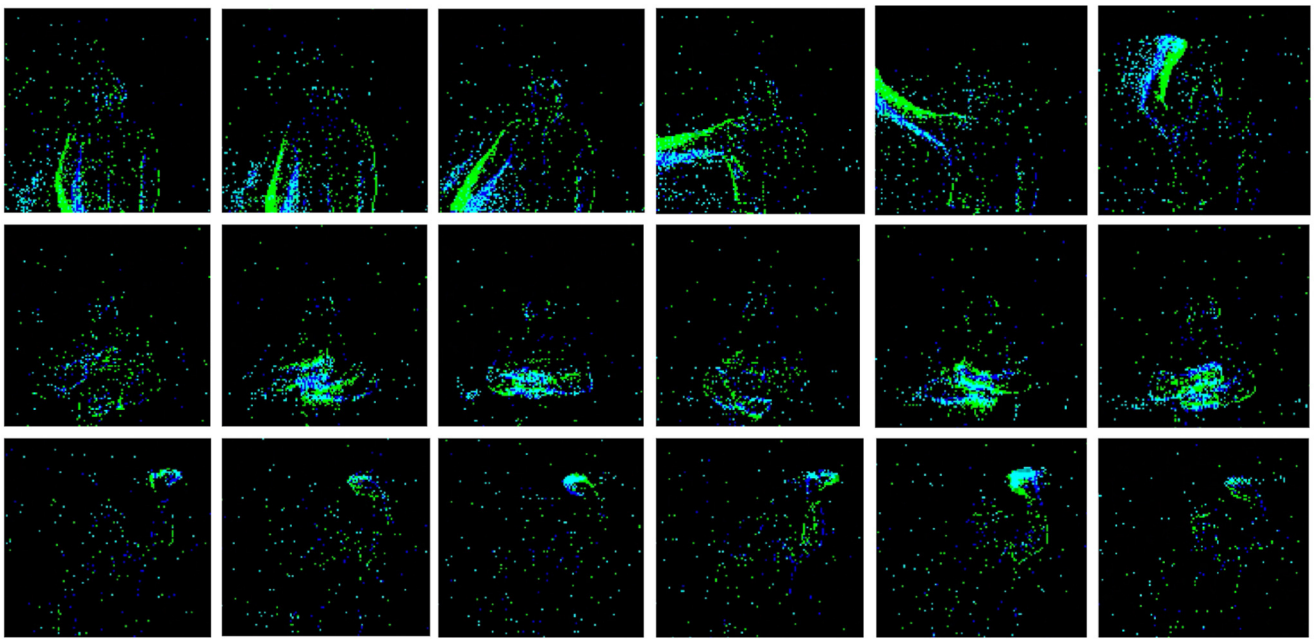

Fig. 7. Since the original event data are extremely sparse, the events within 25 ms are collected into one frame, and each frame is divided into two channels, i.e. ON and OFF, according to the changes in intensity. The three different actions from top to bottom are (*a*) right hand counterclockwise (*b*) arm roll (*c*) left hand wave.

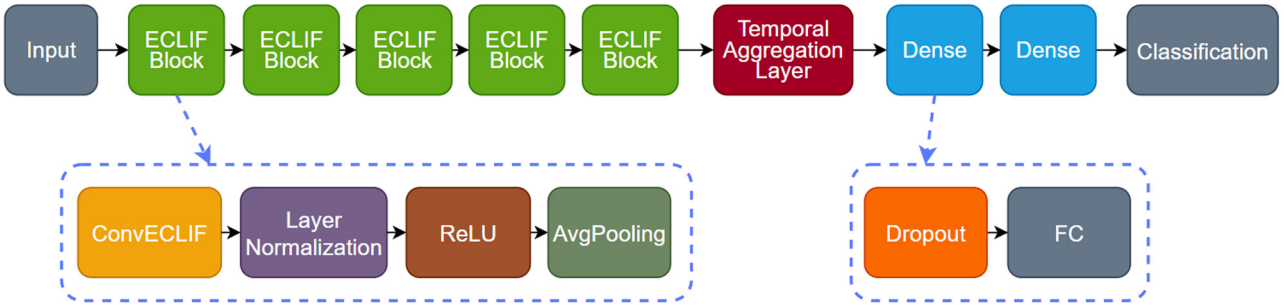

Fig. 8. Illustration of network structure used for CIFAR10-DVS dataset classification tasks.

Table 7
Comparison of different electrical coupling ranges.

Mode	Coupling kernel	Test acc.	ADDs	MULs	Weights
Analog	0 (ConvLIF)	97.56%	6.2 G	6.2 G	0.22 M
	1 × 1	99.31%	7.4 G	7.4 G	0.26 M
	3 × 3	**99.65%**	18 G	18 G	0.55 M
	5 × 5	97.57%	39 G	39 G	1.1 M
Spike	0 (ConvLIF)	94.10%	6.2 G	17 M	0.22 M
	1 × 1	97.22%	7.4 G	1.3 G	0.26 M
	3 × 3	**97.22%**	18 G	12 G	0.55 M
	5 × 5	94.79%	39 G	33 G	1.1 M

Table 8
Comparison of power consumption calculations with other network structures.

Network	Test acc	ADDs	MULs	Weight
ConvLSTM	94.10%	87 G	87 G	2.5 M
RGCNN+Res3D (Bi, 2020)	97.20%	–	14 G	12 M
SNN converted from CNN (Massa et al., 2020)	89.64%	0.44G	0.44G	0.35 M
ConvECLIF2D-S (1 × 1)	**97.22%**	7.4 G	1.3 G	0.26 M
ConvECLIF2D-A (1 × 1)	**99.31%**	7.4 G	7.4 G	0.26 M

Table 9
The network configuration for the CIFAR10-DVS classification experiment.

Layer	Out FM	Kernel	Pool
ECLIF block	32	(3,3)	(2,2)
	64	(3,3)	(2,2)
	128	(3,3)	(2,2)
	256	(3,3)	(2,2)
	512	(3,3)	(4,4)
Dense	512	–	–
	10	–	–

6,000 examples in 10 classes, with 600 examples per class. For data pre-processing, we treat every 5 ms spike train as a separate event-frame and select 10 event-frames as a CIFAR10-DVS sample, then a sample is derived in size [10, 128, 128, 2].

3.4.2. Network structure

We chose the network structure shown in Fig. 8 for the CIFAR10-DVS classification task. The network consists of five ECLIF blocks and two Dense blocks, each ECLIF-A block consists of ConvECLIF2D-A, TD-layer-normalization, TD-ReLU, and TD-AvgPooling, where TD refers to the time-distributed syntax in Tensorflow Keras which copies the temporal-free operation (such as ReLU) for all time steps; the Dense blocks consist of a dropout

Table 10
Comparison with state-of-the-art results on CIFAR10-DVS.

Proposals	Methods	Acc.
Lagorce, Orchard, Galluppi, Shi, and Benosman (2017)	HOTS	27.10%
Shi, Li, Wang, and Luo (2018)	Lightweight Statistical	31.20%
Cannici, Ciccone, Romanoni, and Matteucci (2019)	Attention Mechanisms	44.10%
Sironi, Brambilla, Bourdis, Lagorce, and Benosman (2018)	HATS	52.40%
Wu, Deng, Li, Zhu, Xie, and Shi (2019)	–	63.53%
Wu et al. (2021)	LIAF-Net(ConvLIAF)	70.40%
This work	ConvECLIF2D-S	**67.37%**
This work	ConvECLIF2D-A	**74.08%**

Table 11
Performance and resource consumption comparison among spatio-temporal models on CIFAR10-DVS.

Network	Coupling kernel	Test acc.	ADDs	MULs	Weights
Conv2D		67.80%	3.8*G*	3.3*G*	1.5*M*
Conv3D		71.70%	10.0*G*	9.5*G*	4.7*M*
ConvLSTM		70.80%	43.0*G*	42.0*G*	18.0*M*
LIF-S	0×0	63.53%	3.8*G*	0.21*G*	1.5*M*
LIF-A	0×0	70.40%	3.8*G*	3.3*G*	1.5*M*
ECLIF-S	1×1	67.18%	3.9*G*	0.38*G*	1.8*M*
	3×3	**67.37%**	5.4*G*	1.8*G*	4.6*M*
ECLIF-A	1×1	72.97%	4.6*G*	4.1*G*	1.8*M*
	3×3	**74.08%**	11.0*G*	10.0*G*	4.6*M*

layer and a dense layer. The TAL is applied for integrating temporal information. More specifically, the TAL is obliged for summing up the tensor of all time steps of the time dimension. The specific parameter settings of the network are shown in Table 9. Variables V_{th}, V_{reset}, α, β are all trained by BPTT and use Channel-Sharing mode. For spike mode, because the data transferred between layers are spike trains, we replaced the ECLIF-A blocks with the blocks ECLIF-S consisting of TD-Conv2D, TD-layer-normalization, TD-ReLU, TD-AvgPooling, and DirectECLIF2D-S.

3.4.3. Performance analysis

As shown in Table 10, we tested the proposed network and other existing state-of-the-art networks on the CIFAR10-DVS dataset. The accuracy of the network with ConvECLIF2D-A blocks reaches 74.08%, and the network with ECLIF-S blocks reaches 67.37%. On the other hand, we separately tested the performance and the costs of our proposed model and traditional spatio-temporal networks built by time-distributed Conv2D, Conv3D, and ConvLSTM on the CIFAR10-DVS dataset. They have the same network structure, as shown in Fig. 8, and their parameters are shown in Table 9. It can be seen from Table 11 that our proposed ECLIF-A-Net achieves the highest accuracy among them. Besides, we further verified the different coupling kernel sizes of ECLIF-Net on the CIFAR10-DVS dataset. It shows that the network with a 1×1 coupling kernel size still achieves satisfying performance and better efficiency.

3.5. Large-scale image classification on ES-IMAGENET

3.5.1. Dataset

ES-Imagenet is currently the most challenging event-stream dataset (Lin, Ding, Qiang, Deng, & Li, 2021). It is converted from the popular computer vision dataset ILSVRC2012 (Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009). It generates about 1.3 million frames of images, divided into 1000 categories, including 1257K training and 50K test samples. It is currently the largest ES-dataset for object classification, which is dozens of times larger than other neuromorphic datasets. The brightness information is firstly obtained by converting the RGB (Red–Green–Blue) color space to the HSV (Hue-Saturation-Value), and then an event-stream is generated by the Omnidirectional Discrete Gradient (ODG) algorithm. This algorithm is to imitate the neuron cells with biological characteristics to obtain the necessary information for object recognition. Finally, the events are accumulated through the time axis, and 8 event frames are generated. The size of an event frame is 224×224. The frame contains two channels, indicating the event polarity of each coordinate position. The converted dataset has the characteristics of rich spatio-temporal information in the pulse event stream. It can well verify the model's spatio-temporal information extraction ability.

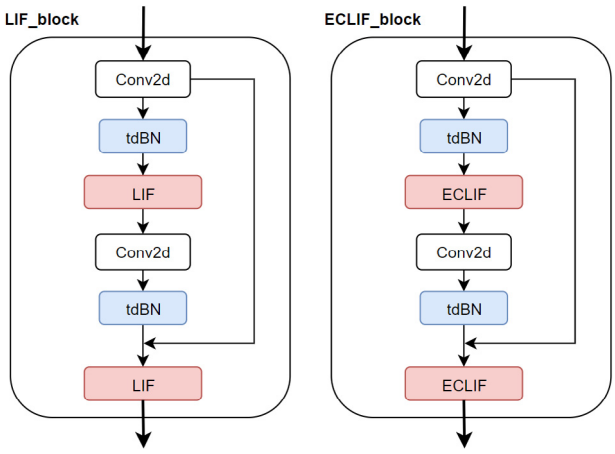

Fig. 9. The SNN neuron replaces the activation function in the ResNet block. The two methods do not perform convolution operations when receiving the pulse input of the previous layer.

Table 12
Verify model validity on ES-IMAGENET.

Backbone	Layer	Test acc.	Parameters
ResNet-18	Conv2D	41.03%	11.68*M*
	ConvLIF2D-S	39.89%	11.69*M*
	ConvLIF2D-A	42.54%	11.69*M*
	ConvECLIF2D-A	44.25%	17.99*M*

3.5.2. Network structure

Due to the relatively large scale of this dataset, we used a deep neural network. The experiment uses a ResNet-18 (He, Zhang, Ren, & Sun, 2016) as the network topology. Three different networks are designed and compared as shown in Fig. 9. The first is a traditional 2D CNN. We compress and reconstruct the 4D event-data in ES-imagenet into a 2D grayscale image, and then obtain the classification via the 2D CNN. The second network is obtained by replacing the activation with DirectLIF2D-S and DirectLIF2D-A in two experiments respectively and replacing the Batchnorm layer with tdBN (Zheng, Wu, Deng, Hu, & Li, 2020). The third network is obtained by replacing DirectLIF2D with DirectECLIF2D-A. The LIF-Net and ECLIF-Net also include the

Table 13
Performance comparison of temporal layers on BABI QA tasks.

	RNN	LSTM	GRU	LIF	ECLIF-S		ECLIF-A	
					last	best	last	best
QA1-Single Supporting Fact	47.80%	46.80%	49.00%	52.70%	47.60%	52.70%	49.40%	52.70%
QA2-Two Supporting Facts	27.50%	33.60%	29.40%	27.50%	19.00%	19.00%	19.00%	29.90%
QA3-Three Supporting Facts	22.40%	22.80%	27.40%	20.20%	22.00%	22.30%	19.40%	21.80%
QA4-Two Arg. Relations	71.30%	70.50%	52.90%	31.10%	70.30%	70.30%	72.40%	73.50%
QA5-Tree Arg. Relations	39.30%	73.50%	73.00%	32.20%	33.60%	55.70%	33.10%	55.40%
QA6-Yes/No Questions	49.50%	50.80%	51.20%	52.60%	48.70%	52.90%	47.70%	52.00%
QA7-Counting	79.30%	79.00%	76.40%	48.80%	75.70%	78.80%	77.80%	78.00%
QA8-Lists/Sets	53.80%	75.70%	73.40%	33.60%	33.60%	33.60%	68.70%	69.50%
QA9-Simple Negation	61.00%	63.80%	62.40%	61.20%	59.00%	62.20%	59.50%	64.30%
QA10-Indefinite Knowledge	45.10%	46.80%	46.70%	46.40%	47.40%	48.00%	45.80%	48.00%
QA11-Basic Coreference	69.70%	65.50%	67.10%	75.10%	71.20%	75.10%	71.50%	75.10%
QA12-Conjunction	64.80%	64.50%	62.70%	77.20%	72.50%	77.20%	70.20%	77.20%
QA13-Compound Coreference	93.60%	92.00%	91.40%	94.40%	93.50%	94.40%	91.30%	94.40%
QA14-Time Reasoning	27.40%	38.40%	39.70%	29.10%	27.80%	30.30%	30.70%	31.60%
QA15-Basic Deduction	45.30%	24.30%	45.90%	25.70%	25.70%	28.20%	28.50%	29.50%
QA16-Basic Induction	44.90%	46.70%	44.40%	45.40%	45.80%	47.60%	47.30%	50.00%
QA17-Positional Reasoning	48.00%	48.00%	48.80%	49.60%	48.70%	52.10%	50.90%	54.00%
QA18-Size Reasoning	90.40%	90.30%	91.20%	90.60%	92.60%	92.60%	91.00%	92.10%
QA19-Path Finding	10.20%	9.60%	8.80%	8.60%	9.90%	9.90%	9.90%	11.30%
QA20-Agent's Motivations	93.40%	97.10%	96.70%	92.60%	95.00%	95.20%	94.40%	95.40%
Average performance on all tasks	**54.24%**	**56.98%**	**56.93%**	**49.73%**	**51.98%**	**54.56%**	**53.93%**	**57.79%**

replacement of the down-sampling convolution of the first layer. The network topology and hyper-parameter settings of the three experiments remain the same. Totally 25 epochs were used for training, with a learning rate of 0.03 and a batch size of 150. The network is trained end-to-end with no data argumentation held on the input frame and no pre-training is involved.

3.5.3. Performance analysis

On this dataset, we are mainly to verify the effectiveness of EC neurons in more complex and high interference situations. As shown in Table 12, it can be seen that under the experimental comparison of the backbone network and the same parameters, the original LIF-Net is underperformed to 2D-CNN, however, after introducing EC and analog activation, the modified ECLIF-Net has reached the highest accuracy of 44.25%.

4. Conclusion

In this work, we modeled the electrical synapses among the densely located neurons and proposed an improved neural model ECLIF. In this model, the coupling effect is modeled as convolutional operations on the membrane potential among neurons. With the electrical coupling modeled in convolutions, our ECLIF integrates the typical inter-layer communication via chemical synapses, as well as intra-layer communication directly in local via electrical synapses. We built ECLIF-NET based on the ECLIF model and trained by BPTT end-to-end. The ECLIF-Net achieved obvious accuracy gain on several datasets over the traditional LIF model. In addition, the accuracy gain is received under limited computational cost when using a smaller coupling kernel size. ECLIF is a representative model inspired by neuroscience discovery and contributes to real task performance. It may provide new insight on the future brain-inspired neural model design and application.

CRediT authorship contribution statement

Zhenzhi Wu: Conceptualization, Methodology, Writing. **Zhihong Zhang:** Conceptualization, Methodology, Software, Writing. **Huanhuan Gao:** Software. **Jun Qin:** Software. **Rongzhen Zhao:** Writing. **Guangshe Zhao:** Methodology. **Guoqi Li:** Conceptualization, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. bAbI QA subtask scores

The bAbI QA dataset contains 20 individual sub-tasks. Here we listed the detailed bAbI QA validation accuracy for each sub-task in Table 13. The 20 sub-tasks are independent, and we calculated the average accuracy to represent the overall performance. From the table, it is revealed that ECLIF-S with the best accuracy is still weaker than LSTM and GRU. Compared with ECLIF-A, the binary activations in ECLIF-S may cause loss of information. Further study also reveals that tasks have their characteristics. For some simple reasoning logic tasks (QA11/QA12), ECLIF performs better. For complex relational reasoning tasks (QA8/QA14), LSTM/GRU performs better than LIF/ECLIF-S.

References

Abbott, L. F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, *50*(5–6), 303–304.

Amir, A., Taba, B., Berg, D., Melano, T., McKinstry, J., Di Nolfo, C., et al. (2017). A low power, fully event-based gesture recognition system. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7243–7252).

Bennett, M. (2000). Seeing is relieving: electrical synapses between visualized neurons. *Nature Neuroscience*, *3*, 7–9.

Bi, Y. (2020). *Graph-based feature learning for neuromorphic vision sensing* (Ph.D. thesis), UCL (University College London).

Bohte, S. M., Kok, J. N., & La Poutré, J. A. (2000). SpikeProp: backpropagation for networks of spiking neurons. In *ESANN. Vol. 48* (pp. 419–424). Bruges.

Boron, W., & Boulpaep, E. (2016a). Sensory transduction. In *Medical physiology* (3rd ed.). (pp. 353–389). Elsevier.

Boron, W., & Boulpaep, E. (2016b). Signal transduction. In *Medical physiology* (3rd ed.). (pp. 47–72). Elsevier.

Boron, W., & Boulpaep, E. (2016c). Synaptic transmission in the nervous system. In *Medical physiology* (3rd ed.). (pp. 307–333). Elsevier.

Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal Of Neurophysiology*, *94*(5), 3637–3642.

Brunel, N., & Latham, P. E. (2003). Firing rate of the noisy quadratic integrate-and-fire neuron. *Neural Computation*, *15*(10), 2281–2306.

Cannici, M., Ciccone, M., Romanoni, A., & Matteucci, M. (2019). Asynchronous convolutional networks for object detection in neuromorphic cameras. In *IEEE/CVF conference on computer vision and pattern recognition workshops*.

Cannici, M., Ciccone, M., Romanoni, A., & Matteucci, M. (2019). Attention mechanisms for object recognition with event-based cameras. In *2019 IEEE winter conference on applications of computer vision* (pp. 1127–1136). http://dx.doi.org/10.1109/WACV.2019.00125.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. CoRR. arXiv:1406.1078.

Chow, C. C., & Kopell, N. (2000). Dynamics of spiking neurons with electrical coupling. *Neural Computation*, *12*(7), 1643–1678.

Cramer, B., Stradmann, Y., Schemmel, J., & Zenke, F. (2020). The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions On Neural Networks And Learning Systems*.

Curti, S., & O'Brien, J. (2016). Characteristics and plasticity of electrical synaptic transmission. *BMC Cell Biology*, *17*(1), 59–70.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.

Deng, L., Wu, Y., Hu, X., Liang, L., Ding, Y., Li, G., et al. (2020). Rethinking the performance comparison between SNNS and ANNS. *Neural Networks*, *121*, 294–307.

Diehl, P. U., & Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers In Computational Neuroscience*, *9*, 99.

Fedus, W., Ramachandran, P., Agarwal, R., Bengio, Y., Larochelle, H., Rowland, M., et al. (2020). Revisiting fundamentals of experience replay. In *International conference on machine learning. Vol. 119* (pp. 3061–3071).

Ferré, P., Mamalet, F., & Thorpe, S. J. (2018). Unsupervised feature learning with winner-takes-all based STDP. *Frontiers In Computational Neuroscience*, *12*, 24.

Fourcaud-Trocmé, N., Hansel, D., Van Vreeswijk, C., & Brunel, N. (2003). How spike generation mechanisms determine the neuronal response to fluctuating inputs. *Journal Of Neuroscience*, *23*(37), 11628–11640.

Ghosh-Dastidar, S., & Adeli, H. (2009). Spiking neural networks. *International Journal Of Neural Systems*, *19*, 295–308.

Hahne, J., Helias, M., Kunkel, S., Igarashi, J., Bolten, M., Frommer, A., et al. (2015). A unified framework for spiking and gap-junction interactions in distributed neuronal network simulations. *Frontiers In Neuroinformatics*, *9*, 22.

Han, B., Ankit, A., Sengupta, A., & Roy, K. (2017). Cross-layer design exploration for energy-quality tradeoffs in spiking and non-spiking deep artificial neural networks. *IEEE Transactions On Multi-Scale Computing Systems*, *4*, 613–623.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hodgkin, A. L., & Huxley, A. F. (1952). Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *The Journal Of Physiology*, *116*(4), 449–472.

Hong, C., Wei, X., Wang, J., Deng, B., Yu, H., & Che, Y. (2019). Training spiking neural networks for cognitive tasks: A versatile framework compatible with various temporal codes. *IEEE Transactions On Neural Networks And Learning Systems*, 1–12. http://dx.doi.org/10.1109/TNNLS.2019.2919662.

Illing, B., Gerstner, W., & Brea, J. (2019). Biologically plausible deep learning - but how far can we go with shallow networks? *Neural Networks*, *118*, 90–101.

Izhikevich, E. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions On Neural Networks*, *15*, 1063–1070.

Jordan, J., Helias, M., Diesmann, M., & Kunkel, S. (2020). Efficient communication in distributed simulations of spiking neuronal networks with gap junctions. *Frontiers In Neuroinformatics*, *14*, 12.

Khoei, M. A., Yousefzadeh, A., Pourtaherian, A., Moreira, O., & Tapson, J. (2020). SpArNet: Sparse asynchronous neural network execution for energy efficient inference. In *IEEE international conference on artificial intelligence circuits and systems* (pp. 256–260).

Kugele, A., Pfeil, T., Pfeiffer, M., & Chicca, E. (2020). Efficient processing of spatio-temporal data streams with spiking neural networks. *Frontiers In Neuroscience*, *14*, 439.

Lagorce, X., Orchard, G., Galluppi, F., Shi, B. E., & Benosman, R. B. (2017). HOTS: A hierarchy of event-based time-surfaces for pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(7), 1346–1359. http://dx.doi.org/10.1109/TPAMI.2016.2574707.

Lapique, L. (1907). Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. *Journal Of Physiology And Patholgy*, *9*, 620–635.

Lee, J. H., Delbruck, T., & Pfeiffer, M. (2016). Training deep spiking neural networks using backpropagation. *Frontiers In Neuroscience*, *10*, 508.

Li, H., Liu, H., Ji, X., Li, G., & Shi, L. (2017). Cifar10-DVS: an event-stream dataset for object classification. *Frontiers In Neuroscience*, *11*, 309.

Lillicrap, T., Santoro, A., Marris, L., Akerman, C., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*, 335–346.

Lin, Y., Ding, W., Qiang, S., Deng, L., & Li, G. (2021). Es-imagenet: a million event-stream classification dataset for spiking neural networks. *Frontiers in Neuroscience*, 1546.

Massa, R., Marchisio, A., Martina, M., & Shafique, M. (2020). An efficient spiking neural network for recognizing gestures with a DVS camera on the loihi neuromorphic processor. arXiv preprint arXiv:2006.09985.

Mikolov, T., Karafiát, M., Burget, L., Černockỳ, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*.

Mozafari, M., Ganjtabesh, M., Nowzari-Dalini, A., Thorpe, S. J., & Masquelier, T. (2019). Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks. *Pattern Recognition*, *94*, 87–95.

Nadim, F., Li, X., Gray, M., & Golowasch, J. (2017). The role of electrical coupling in rhythm generation in small networks. In *Network functions and plasticity* (pp. 51–78). Elsevier.

Neftci, E. O., Mostafa, H., & Zenke, F. (2019). Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, *36*(6), 51–63.

O'brien, J. (2014). The ever-changing electrical synapse. *Current Opinion In Neurobiology*, *29*, 64–72.

Pereda, A. E. (2014). Electrical synapses and their functional interactions with chemical synapses. *Nature Reviews Neuroscience*, *15*, 250–263.

Pfeiffer, M., & Pfeil, T. (2018). Deep learning with spiking neurons: Opportunities and challenges. *Frontiers In Neuroscience*, *12*, 774.

Pineda, F. J. (1987). Generalization of back-propagation to recurrent neural networks. *Physical Review Letters*, *59*(19), 2229.

Roy, K., Jaiswal, A., & Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature*, *575*(7784), 607–617.

Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., & Liu, S.-C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Frontiers In Neuroscience*, *11*, 682.

Schliebs, S., & Kasabov, N. (2013). *Evolving spiking neural network - a survey. Vol. 4* (pp. 87–98).

Shi, C., Li, J., Wang, Y., & Luo, G. (2018). Exploiting lightweight statistical learning for event-based vision processing. *IEEE Access*, *6*, 19396–19406. http://dx.doi.org/10.1109/ACCESS.2018.2823260.

Sironi, A., Brambilla, M., Bourdis, N., Lagorce, X., & Benosman, R. (2018). HATS: Histograms of averaged time surfaces for robust event-based object classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Tavanaei, A., Ghodrati, M., Kheradpisheh, S., Masquelier, T., & Maida, A. (2019). Deep learning in spiking neural networks. *Neural Networks*, *111*, 47–63.

Wang, Q., Zhang, Y., Yuan, J., & Lu, Y. (2019). Space-time event clouds for gesture recognition: From RGB cameras to event cameras. In *2019 IEEE winter conference on applications of computer vision* (pp. 1826–1835). http://dx.doi.org/10.1109/WACV.2019.00199.

Weston, J., Bordes, A., Chopra, S., Rush, A. M., van Merriënboer, B., Joulin, A., et al. (2015). Towards AI-complete question answering: A set of prerequisite toy tasks. arXiv:1502.05698.

Wozniak, S., Pantazi, A., Bohnstingl, T., & Eleftheriou, E. (2020). Deep learning incorporating biologically inspired neural dynamics and in-memory computing. *Nature Machine Intelligence*, *2*, 325–336.

Wu, Y., Deng, L., Li, G., Zhu, J., & Shi, L. (2018). Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers In Neuroscience*, *12*, 331.

Wu, Y., Deng, L., Li, G., Zhu, J., Xie, Y., & Shi, L. (2019). Direct training for spiking neural networks: Faster, larger, better. *Proceedings Of The AAAI Conference On Artificial Intelligence*, *33*(01), 1311–1318. http://dx.doi.org/10.1609/aaai.v33i01.33011311.

Wu, Z., Zhang, H., Lin, Y., Li, G., Wang, M., & Tang, Y. (2021). LIAF-Net: Leaky integrate and analog fire network for lightweight and efficient spatiotemporal information processing. *IEEE Transactions On Neural Networks And Learning Systems*.

Yao, X.-H., Wang, M., He, X.-N., He, F., Zhang, S.-Q., Lu, W., et al. (2016). Electrical coupling regulates layer 1 interneuron microcircuit formation in the neocortex. *Nature communications*, *7*(1), 1–15.

Yin, B., Corradi, F., & Bohté, S. M. (2020). Effective and efficient computation with multiple-timescale spiking recurrent neural networks. In *International conference on neuromorphic systems 2020* (pp. 1–8).

Zeng, G., Chen, Y., Cui, B., & Yu, S. (2019). Continual learning of context-dependent processing in neural networks. *Nature Machine Intelligence*, *1*, 364–372.

Zenke, F., & Vogels, T. P. (2021). The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural Computation*, *33*(4), 899–925.

Zhang, Y., Qu, P., Ji, Y., Zhang, W., Gao, G., Wang, G., et al. (2020). A system hierarchy for brain-inspired computing. *Nature*, *586*, 378–384.

Zhao, D., Zhang, Z., Lu, H., Cheng, S., Si, B., & Feng, X. (2020). Learning cognitive map representations for navigation by sensory-motor integration. *IEEE Transactions On Cybernetics*, 1–14.

Zheng, H., Wu, Y., Deng, L., Hu, Y., & Li, G. (2020). Going deeper with directly-trained larger spiking neural networks. arXiv preprint arXiv:2011.05280.