Spiking World Model with Multi-Compartment Neurons for Model-based Reinforcement Learning

Yinqian Sun, Feifei Zhao, Mingyang Lv, Yi Zeng

Abstract—Brain-inspired spiking neural networks (SNNs) have garnered significant research attention in algorithm design and perception applications. However, their potential in the decisionmaking domain, particularly in model-based reinforcement learning, remains underexplored. The difficulty lies in the need for spiking neurons with long-term temporal memory capabilities, as well as network optimization that can integrate and learn information for accurate predictions. The dynamic dendritic information integration mechanism of biological neurons brings us valuable insights for addressing these challenges. In this study, we propose a multi-compartment neuron model capable of nonlinearly integrating information from multiple dendritic sources to dynamically process long sequential inputs. Based on this model, we construct a Spiking World Model (Spiking-WM), to enable model-based deep reinforcement learning (DRL) with SNNs. We evaluated our model using the DeepMind Control Suite, demonstrating that Spiking-WM outperforms existing SNN-based models and achieves performance comparable to artificial neural network (ANN)-based world models employing Gated Recurrent Units (GRUs). Furthermore, we assess the long-term memory capabilities of the proposed model in speech datasets, including SHD, TIMIT, and LibriSpeech 100h, showing that our multi-compartment neuron model surpasses other SNNbased architectures in processing long sequences. Our findings underscore the critical role of dendritic information integration in shaping neuronal function, emphasizing the importance of cooperative dendritic processing in enhancing neural computation.

Index Terms—Brain-inspired Intelligence, Multi-compartment Neuron Model, Spiking World Model, Model-based Reinforcement Learning

I. INTRODUCTION

In the burgeoning field of neuromorphic computing, the integration of biologically inspired models into artificial neural networks has paved the way for novel computational frameworks that transcend traditional approaches. Inspired by the dynamic processes of neuronal activity and spike-based signal transmission in the brain, spiking neural networks exhibit energy saving and computational efficiency advantages [1],

Manuscript received April 19, 2021; revised August 16, 2021.

Yinqian Sun and Feifei Zhao is with the Brain-inspired Cognitive Intelligence Lab, Institute of Automation, Chinese Academy of Sciences, Beijing, China (e-mail: sunyinqian2018@ia.ac.cn)

Mingyang Lv is with the Brain-inspired Cognitive Intelligence Lab, Institute of Automation, Chinese Academy of Sciences, Beijing, China and School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

Yi Zeng is with the Brain-inspired Cognitive Intelligence Lab, Institute of Automation, Chinese Academy of Sciences, Beijing, China, School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, CAS, Shanghai, China and Center for Long-term Artificial Intelligence, Beijing, China.(e-mail: yi.zeng@ia.ac.cn)

The corresponding author is Yi Zeng.

[2], and have been effectively utilized across a wide range of applications [3]–[6]. Current SNN models predominantly employ Leaky Integrate-and- Fire (LIF) neurons as a trade-off between biological plausibility and computing simplicity. However, recent advancements in neuroscientific research have revealed the significant role of dendritic structures in neuronal information processing [7]–[9]. The multi-compartment neuron (MCN) model refers to a mathematical framework that treats different parts of a biological neuron such as the axon, soma, and dendrites as independent computational units.

The multi-compartment SNNs captures the morphological characteristics of biological neurons, utilizing dynamic equations to model dendritic information processing, membrane potential propagation, and somatic spike generation [9], [10]. Various implementations of this model, such as the two-compartment neuron model [11], [12] and the multi-compartment model with apical and basal dendrites [13], have demonstrated applications in tasks like MNIST classification and target-based learning [14]. Additionally, the MCN models have been integrated into neuromorphic chips for energy-efficient simulation of neural circuits [15], [16]. Recent studies have shown that multi-compartment SNNs outperform LIF-based SNNs in both temporal sequence classification tasks [17] and neuromorphic datasets [18].

Unlike static image recognition tasks, reinforcement learning necessitates that an agent collect data and refine models through dynamic interactions with the environment. In particular, model-based reinforcement learning methods [19]–[21] construct a predictive model based on data gathered from the agent's interactions with the environment. This predictive model is subsequently employed to generate synthetic data through simulated interactions with the agent, thereby optimizing the agent's policy. By reducing the number of direct interactions required with the real environment, these methods enhance sample efficiency. For example, the Dreamer model [22], [23] uses the recurrent state-space model to construct a world model, enabling latent imagination in long-horizon tasks.

Currently, there is a lack of work applying SNNs to model-based reinforcement learning, primarily due to the challenges posed by the need for strong long-term memory capabilities and the ability to effectively learn from the non-stationary distributions of samples in RL environments. Biological studies have demonstrated that the complex computational functions of dendrites enhance neuronal pattern separation [24], significantly improve the temporal sequence memory of neurons [25], [26], facilitate rule learning and the formation of long-term memory [27], and enable neurons to perform more

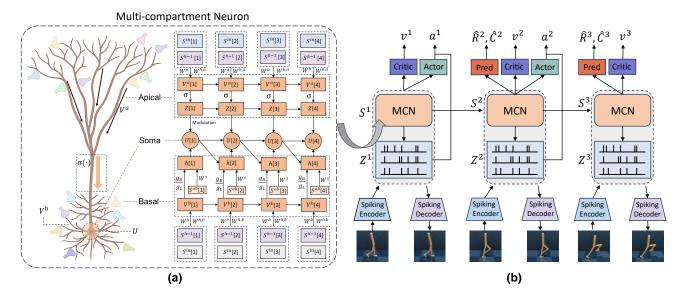


Fig. 1. The spiking world model based on MCN. (a) Biological structure of pyramidal neurons (left) and the multi-compartment neuron model (right). (b) The architecture of spiking world model.

complex perceptual decision-making tasks.

Inspired by the dendritic computational mechanisms of biological neurons, this study presents a multi-compartmental neuron model with sequential nonlinear dendritic integration capabilities. Leveraging this model, we develop a spiking world model to enable a model-based deep reinforcement learning framework using spiking neural networks. The proposed approach achieves performance comparable to conventional RNN-based DRL methods on DeepMind visual control tasks. Moreover, it outperforms existing spiking recurrent neural network models on speech sequence datasets, including SHD, TIMIT, and LibriSpeech 100h. Overall, the main contributions of this work are as follows:

- We propose a multi-compartment neuron model that incorporates dendritic computations, inspired by the structural characteristics of biological neurons. Our spiking MCN model exhibits long-term sequential memory capabilities by leveraging the biophysical dynamics of neuronal compartments, providing enhanced temporal processing capacity for spiking neural networks.
- Building upon the multi-compartment neuron framework, we introduce a spiking world model and develop a fully spiking model-based DRL framework. Specifically, we implement an SNN-based Dreamer model, offering a biologically plausible alternative for model-based reinforcement learning in spiking neural networks.
- We evaluate the proposed Spiking-WM and MCN models on DeepMind visual control tasks and sequential speech datasets, including SHD, TIMIT, and LibriSpeech 100h. Experimental results demonstrate that our models surpass existing state-of-the-art spiking recurrent neural networks with the same parameter budget, establishing a new benchmark for SNN-based sequence modeling and control tasks.

II. RELATED WORK

Most spiking neural network research utilizes point-neuron models to represent biological neurons, such as the widely used LIF [28] neuron model, which computing process relying solely on the somatic membrane potential. The ALIF [29] model proposes an adaptive spiking recurrent neural network to process time-domain datasets classification. The improved recurrent spiking neuron [30] modified LIF neurons to enhance their inherent recurrence for sequential learning, demonstrating that the improved SNN architecture achieves comparable accuracy to LSTMs on speech recognition tasks.

Point neuron models lack structural detail in neuronal modeling and are insufficient for accurately capturing the complex information integration and long-sequence processing capabilities of biological neurons. Consequently, there is growing research interest in computational modeling of dendritic processes and the development of multi-compartmental neuron models. A two-compartment neuron model with distinct soma and dendrite components is introduced by [11], [12], which fine-tunes the dendritic synapse weights to facilitate dendritic predictions of somatic spike firing patterns. Additionally, [13] proposed a three-compartment neuron model that captures the computational processes occurring in apical dendrites, basal dendrites, and the soma. The TC-LIF [17] built a twocompartment spiking neuron model to enhance long-term sequential modeling in SNNs. Target-based learning [14], [31] leverages the apical dendrite to receive the plasticity signal, enabling local modification of neuron weights.

However, the aforementioned multi-compartmental neuron models provide only a simplified representation of dendritic computations, with most implementations incorporating a basic neuronal structure containing a single dendrite. Additionally, they largely overlook the nonlinear interactions and memory storage mechanisms involved in dendrite-to-soma signal transmission. To address these limitations, this study considers the distinct computational roles of dendrites at different loca-

tions within biological neurons, such as the apical and basal dendrites of pyramidal neurons. Furthermore, we model the cooperative computational processes that occur during signal transmission across dendrites at different locations.

Currently, research on brain-inspired SNNs is primarily focused on visual information processing areas. The application of SNNs, particularly those with deep structures, in complex decision-making fields remains relatively limited. One major challenge for implementing RL algorithms using SNNs is the difficulty in training and converging the model, owing to the dynamic nature of RL environments.

Some studies have applied spiking neural networks to reinforcement learning tasks. Using the ANN-to-SNN conversion method, the convolutional layers and fully connected layers in the DQN network are converted to SNN formulation applied to the Atari Breakout game [32], [33].

In the area of continuous control tasks, SNNs based on R-STDP have been applied to intelligent vehicle lane-keeping tasks [34] and training hexapod robots to walk [35]. Pop-SAN [36] uses a population coding method with spiking neurons to implement an SNN-based reinforcement learning Actor-Critic algorithm, which is used for continuous control tasks in MuJoCo simulation environment. A two-module spiking neural network model is introduced that leverages biologically inspired mechanisms, such as "dreaming" (offline learning in a model-based simulated environment) and "planning" (online learning), to enhance reinforcement learning efficiency without requiring the detailed storage of past experiences [37]. The MCS-FQF [38] presents a brain-inspired deep distributional reinforcement learning algorithm based on multi-compartment neuron model SNN and population coding method, which significantly enhances computational power and performance in Atari games.

Nevertheless, existing research on reinforcement learning with spiking neural networks has primarily focused on implementing model-free RL methods, while the application of SNNs in model-based RL remains largely unexplored. In this study, we develop a spiking world model based on a multi-compartmental neuron model with nonlinear dendritic computations, enabling the application of SNNs in model-based deep reinforcement learning.

III. PRELIMINARY AND METHODOLOGY

A. Model-based Reinforcement Learning

Reinforcement learning can be formed as a Markov Decision Process (MDP), which can be described by element matrix: $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}^a_{ss'}, \mathcal{R}^a_{ss'}, \gamma \rangle$. At discrete time step $t \in [1;T]$, the agent takes action with probability $a_t \sim p(a_t|s_t)$ with $a_t \in \mathcal{A}; s_t \in \mathcal{S}$. The state transiting to $s_{t+1} \sim p(s_{t+1} \in \mathcal{S}|s_t, a_t)$ when agent taking action a_t at state s_t with probability $\mathcal{P}^{a_t}_{s_t s_{t+1}}$, and the agent receives the reward $r_t \sim \mathcal{R}^{a_t}_{s_t s_{t+1}}$ generated by the unknown environment. Typically, the main objective of reinforcement learning is to maximize the expected cumulative reward $E_p(\sum_{t=1}^T \gamma \cdot r_t)$.

Model-based reinforcement learning represents the world environment with latent dynamic model, using probability functions $q(s_t|s_{t-1},a_{t-1})$ to fit the state transition process

 $\mathcal{P}^{a_t}_{s_t s_{t+1}}$. Similarly, the rewards from the environment can also be predicted by $q(r_t|s_t)$. Compared to model-free reinforcement learning, model-based reinforcement learning can reduce the number of interactions with the environment by learning policies through the world model. This alleviates the trial-and-error cost of reinforcement learning algorithms and increases the feasibility of deploying these algorithms in real-world environments.

B. Multi-compartment Neuron Model

The spiking neuron model simulates the propagation of electrical potentials and the firing processes of biological neurons. Taking the LIF neuron as an example, which is designed primarily for balancing the intricate complexity of biological neuron details with computational tractability, and is widely used in the construction of spiking neural networks and brain simulations. The membrane potential dynamics of the LIF neuron are described by Equation 1, and the parameter τ is the somatic decay time parameter. When the membrane potential U_t exceeds a threshold V_{th} , the neuron fires a spike S_t and the membrane potential is reset $U_t \leftarrow V_{res}$; $U_t > V_{th}$.

$$\tau \frac{dU_t}{dt} = -U_t + x_t; \quad U_t < V_{th} \tag{1}$$

$$S_t = \begin{cases} 1 & U_t > V_{th} \\ 0 & U_t < V_{th} \end{cases} \tag{2}$$

Although the membrane potential dynamic process of LIF spiking neurons has the inherent ability to process sequential information, it only considers the passive updating of membrane potential and spike firing process. This approach ignores the morphological and structural characteristics of neurons and models neurons as point-like structures without taking into account the local computational capabilities and functional enhancements provided by different parts of biological neurons, such as dendrites and soma.

Biological neurons can utilize different parts of their dendrites to receive input signals from various sources, enabling local independent computations. In contrast, the LIF neuron model only considers somatic computation. This paper proposes a multi-compartment neuron model that incorporates the collaborative computation of multiple dendrites and the soma. Specifically, pyramidal neurons, which are widely present in the hippocampus and neocortex, are modeled as computational units consisting of three parts: apical dendrites, basal dendrites, and the soma, shown as Figure 1(a). The dynamic processing of the proposed MCN model is as follows:

$$\tau \frac{dU_t}{dt} = -U_t + \frac{g_B}{g_L} (V_t^b - U_t) + \frac{g_A}{g_L} (V_t^a - U_t)$$
 (3)

where V^b and V^a are the basal and apical dendritic membrane potentials, respectively. The g_B , and g_A are conductance parameters for basal and apical dendrites, and g_L is membrane leaky conductance. Same as LIF model, the MCN neuron fires a spike when the somatic potential U_t exceeds the threshold.

C. Sequential Nonlinear Dendritic Integration for MCN

Spiking neurons, by modeling the dynamic biophysical processes of biological neurons, possess the inherent ability to process sequential information due to the mathematical form of their dynamic differential equations. In this work, we propose a method for sequential information processing based on multi-compartment neurons and apply it to model-based deep reinforcement learning. Inspired by the brain mechanism where information from different parts of a biological neuron's dendrites has varying modulatory effects on the somatic membrane potential, we formulate the MCN model with local membrane potential computation processes of basal and apical dendrites as

$$\begin{cases}
\tau_b \frac{dV_t^b}{dt} = -V_t^b + x_t^b \\
x_t^b = W^b S_t^{in} + W^{h,b} S_t^{h-1}
\end{cases}$$
(4)

$$\begin{cases} \tau_a \frac{dV_t^a}{dt} = -V_t^a + x_t^a \\ x_t^a = W^a S_t^{in} + W^{h,a} S_t^{h-1} \end{cases}$$
 (5)

$$\tau \frac{dU_t}{dt} = -U_t + \sigma(V_t^a) (\frac{g_B}{g_L} (V_t^b - U_t) + W^s S_t^{in})$$
 (6)

where the V^b_t and V^a_t is basal and apical dendritic potentials, respectively. Both change dynamically based on the external sources of spike signals S^{in}_t and internal signals S^h_t representing hidden states. The basal and apical dendrites are equipped with synaptic weights $[W^b,W^{h,b}]$ and $[W^a,W^{h,a}]$, which are adjustable parameters providing the dendrites with local learning capabilities.

This work primarily explores the potential recurrent memory function inherent in pyramidal neurons. One important source of these internal signals S_t^h can be the recurrent and feedback loop connections that are abundantly present in the hippocampus and neocortex. Similar to the multi-compartment neuron model described in Equation 3, the computational results of the dendrites are integrated at the soma in the form of membrane potential. However, in this work, we explore the nonlinear modulatory effects of apical dendrites on the somatic potential U_t , where the potential of the apical dendrites V_t^a determines the contribution of the information transmitted by the basal dendrites V_t^b and the soma's input information S_t^{in} which is processed by W^s the learnable somatic synapse parameter.

D. Spiking World Model

The membrane potential dynamics of the spiking neuron model make it suitable for processing sequential signals. Based on the proposed MCN model, we built a spiking State Space Model (SSM) as in Dreamer work [23]. Furthermore, we use LIF neurons to construct the model's encoder, decoder, and policy network, thereby implementing a fully spiking world model. In this work, we handle continuous control tasks with visual inputs using LIF neurons to construct a spiking convolutional neural network (SCNN), which extracts features from image inputs and performs spiking encoding as Equation 7. The O[t] is an observation input at time t and the SCNN encodes it to spikes $S^{ob}[t]$. A fully connected LIF-based SNN model is used to integrate the previous latent state

 $Z^{h-1}[t]$ and action $A^{h-1}[t]$ to generate sequential model spike input $S^{in}[t]$.

$$S^{ob}[t] = SCNN(O[t]); S^{in}[t] = SNN(Z^{h-1}[t], A^{h-1}[t])$$
 (7)

We used the Euler method to numerically expand the Equation (4–6) MCN formulation, proposing the following sequential model:

$$V^{b}[t] = V^{b}[t-1] + (W^{b}S^{in}[t] + W^{h,b}S^{h-1}[t] - V^{b}[t-1])/\tau_{b}$$
(8)

$$V^{a}[t] = V^{a}[t-1] + (W^{a}S^{in}[t] + W^{h,a}S^{h-1}[t] - V^{a}[t-1])/\tau_{a}$$
(9)

$$h[t] = \frac{g_B}{g_L}(V^b[t] - U[t-1]) + W^s S^{ob}[t]; \ z[t] = \sigma(V^a[t])$$
 (10)

$$U[t] = U[t-1] + (z[t] \odot h[t] - U[t-1])/\tau$$
 (11)

$$S^{h}[t] = \Theta(U[t] - V_{th}) \tag{12}$$

where $\Theta(\cdot)$ is the Heaviside step function. When firing a spike, the MCN resets its membrane potential as $U[t] = U[t](1 - S^h[t])$. We use the Sigmoid function with parameter β as the gating function for the apical dendritic membrane potential with

$$\sigma(x) = \frac{1}{1 + e^{-\beta x}} \tag{13}$$

The state spiking encoding $S^h[t]$ of the multi-compartment neuron model, along with the dynamic updates of the neuron's dendritic and somatic membrane potentials, endows it with long-short term memory capabilities.

The observation encoding and sequence model output are used to generate the model state representation $Z^h[t]$:

$$S^{h}[t] = MCN(S^{in}[t], S^{h-1}[t])$$
(14)

$$Z^{h}[t] \sim q_{\theta}(Z^{h}[t]|S^{ob}[t], S^{h}[t])$$
 (15)

Similar to the Dreamer model, we use the SNNs to output distributional state parameters instead of the direct values of the state representation. The model state $Z^h[t]$ and the spiking information output $S^h[t]$ from the MCN are used to predict the reward $\hat{R}[t]$ and the continuous signal $\hat{C}[t]$ for each episode. The decoder uses a spiking deconvolutional network to obtain predictions of observation $\hat{O}[t]$ from the model's latent state. The predictive outputs of the world model are used to select actions and evaluate states.

$$\hat{Z}^h[t] \sim p_\theta(\hat{Z}^h[t]|S^h[t]) \tag{16}$$

$$\hat{R}^{h}[t] \sim p_{\theta}(\hat{R}^{h}[t]|S^{h}[t], Z^{h}[t])$$
 (17)

$$\hat{C}^{h}[t] \sim p_{\theta}(\hat{C}^{h}[t]|S^{h}[t], Z^{h}[t])$$
 (18)

$$\hat{O}[t] \sim p_{\theta}(\hat{O}[t]|S^h[t], Z^h[t]) \tag{19}$$

Figure 1(b) shows the spiking world model and action selection process. We use an SNN model to construct the

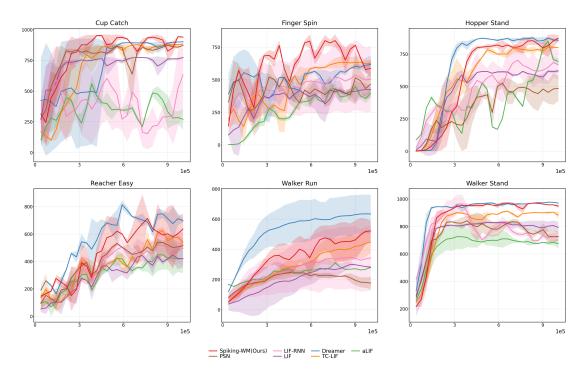


Fig. 2. Evaluation scores for DeepMind visual control experiments.

policy model, which includes both actor and critic networks. The actor network integrates the spiking sequence representation of the world model state $[S_t^h, Z_t^h]$ at spiking time step t to determine the action selection a^h at the h-th step within the environmental context. The critic network generates environmental state values v^h based on the spiking sequence representation of the model as an estimate of the expected returns obtainable from the current state.

Actor:
$$a^h \sim SNN(S_t^h, Z_t^h)$$

Critic: $v^h \sim SNN(S_t^h, Z_t^h)$ (20)

We train the actor and critic using the bootstrapped method. Within the predicted horizon of length H, the λ -returns are calculated based on the current reward predictions $\hat{R}^{1:H}$ and environmental condition predictions $\hat{C}^{1:H}$, as shown by Equation 21. As in Dreamer method, SNN based actor and critic are trained by discrete regression and maximum returns.

$$\mathcal{R}^{\lambda,h} = \hat{R}^h + \gamma \hat{C}^h[t]((1-\lambda)v^{h+1} + \lambda \mathcal{R}^{\lambda,h+1}) \quad \lambda \mathcal{R}^{\lambda,H} \doteq v^H$$
(21

The loss function of the world model consists of the model prediction loss $\mathbb{L}_{pred}(\theta)$ and the posterior estimation loss $\mathbb{L}_{model}(\theta)$, as shown in Equation 22. The objective of the model prediction loss $\mathbb{L}_{pred}(\theta)$ is to optimize the maximum likelihood of the decoder, reward predictor, and continue predictor under the true observations, rewards, and continue condition values of the environment. The model dynamics prediction loss $\mathbb{L}_{model}(\theta)$ aims to improve the sequence model MCN's prediction of the next state by minimizing the KL

divergence between the prior predictions and the posterior representations using Bayesian estimation methods.

$$\mathbb{L}_{model}(\theta) = E_{q_{\theta}} \left[\sum_{h=1}^{H} (\mathbb{L}_{pred}(\theta) + \mathbb{L}_{model}(\theta)) \right]$$
 (22)

$$\mathbb{L}_{pred}(\theta) = \sum_{x^h} -lnp_{\theta}(x^h|, S^h[t], Z^h[t]) \quad x^h \in \{O^h, r^h, c^h\}$$
(23)

$$\mathbb{L}(\theta) = KL[q_{\theta}(Z^{h}|S^{ob}[t], S^{h}[t])||p_{\theta}(\hat{Z}^{h}[t]|S^{h}[t])]$$
 (24)

Considering that the long-short term memory processes of SNNs and MCN involve the temporal integration of spike sequences, we optimize the SNNs' parameters by using the Spatio-temporal Backpropagation (STBP) method [38]. At the discontinuous points of spike firing, we use the a surrogate function g'(x) as the gradient of spike firing moments:

$$\frac{\partial S^h[t]}{\partial U[t]} \longleftarrow g'(x)(U[t] - V_{th}) \tag{25}$$

$$g'(x) = \begin{cases} 0, & |x| > \frac{1}{\alpha} \\ -\alpha^2 |x| + \alpha, & |x| \le \frac{1}{\alpha} \end{cases}$$
 (26)

IV. EXPERIMENTS

To evaluate the performance of the proposed spiking world model, we conducted experiments on the DeepMind Visual Control Suite. Additionally, to investigate the long-sequence



Fig. 3. Comparison of visual state predictions for different neuron models.

memory capability of the MCN model, we tested the performance of different neurons on the speech datasets SHD, TIMIT, and LibriSpeech 100h. Furthermore, to assess the impact of MCN parameters on model performance, we employed grid search and adaptive parameter learning methods to analyze the task performance and long-sequence memory capacity of the Spiking-WM model under different dendritic parameter configurations of the MCN.

A. DeepMind Visual Control Experiments

We conducted experiments on visual input control tasks in DeepMind Control (DMC) Suite, comparing the performance of MCN-based spiking world model across 19 continuous control tasks. For comparison, we evaluated the capabilities of the traditional GRU-based world model algorithm Dreamer [23]. In addition, to demonstrate the critical role of the MCN model in the spiking world model and its advantages over other pointneuron models and simple two-compartment neuron models, we constructed equivalent spiking world models based on several alternatives. These included the aLIF [29] model, formed by augmenting the baseline LIF [39] neuron for processing long-sequence spiking data, the LIF-RNN [30] model, which enhances internal recurrence, the parallelized spiking neuron (PSN) [40] and the two-compartment neuron (TC-LIF) [17] model. Experiments were conducted on the DeepMind Visual Control suite using identical model parameters and training settings. Detailed model and experimental parameter settings are shown in Table II.

DMC tasks based on visual input require the model to accurately represent the features of the current observation images and make correct control actions based on visual perception. In the experiment, we use a 4-layer spiking convolutional network to process image observations from the environment, encoding the image observation into the spiking sequence space. All compared models use the same SCNN model structure and parameter settings to process visual inputs and employ the same SNN-based Actor-Critic policy learning.

In the DMC experiments, the sequence prediction model spiking-WM consists of 512 multi-compartment neurons. For the sake of fairness in comparison, we increased the number of neurons and the dimensionality of weights in the other baseline models to ensure consistency in the total number of parameters. Detailed parameter statistics and experimental results are provided in Table I. The decoder uses a 4-layer spiking transposed convolutional network, and the predictor is a fully connected LIF neuron SNN.

The final evaluation results were obtained after training for one million (1M) frames. The results in Figure 2 and Table I demonstrate that our proposed Spiking-WM model outperforms other spiking neuron baselines in terms of learning speed and final scores, achieving results comparable to the ANN-based Dreamer model. On the average score across 19 tasks, the Spiking-WM model achieves 90.4% of the performance of the ANN-based Dreamer(GRU) model, surpassing the TC-LIF model, which reached only 81.9% of Dreamer(GRU) scores, and outperforming point-neuron models like LIF and its derivative models by 31%~38%.

Task	Dreamer(GRU) [23]	LIF [39]	aLIF [29]	PSN [40]	LIF-RNN [30]	TC-LIF [17]	Spiking-WM(Ours)
Parameters (M)	19.10	19.11	18.85	18.84	19.11	18.84	18.84
Env Steps (M)	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Acrobot Swingup	219.1	104.0	79.1	95.8	125.5	106.5	113.7
Cartpole Balance	970.7	663.0	502.4	788.6	957.7	965.3	973.3
Cartpole Balance Sparse	991.2	991.5	748.3	832.6	718.2	992.4	989.6
Cartpole Swingup	810.1	774.1	614.0	575.9	772.6	667.7	791.0
Cartpole Swingup Sparse	480.2	261.3	217.1	240.6	211.1	265.5	420.1
Cheetah Run	720.1	515.5	558.1	283.9	510.8	517.8	577.2
Cup Catch	958.2	768.2	494.2	915.8	523.0	938.7	960.3
Finger Spin	637.3	416.2	433.5	472.4	509.3	657.8	682.0
Finger Turn Easy	763.5	489.2	542.9	592.7	706.1	688.5	727.5
Finger Turn Hard	397.2	193.6	201.3	289.1	269.4	369.1	406.3
Hopper Hop	373.4	104.9	145.7	114.4	215.6	205.4	234.9
Hopper Stand	913.6	650.6	704.1	493.1	715.1	864.2	905.3
Pendulum Swingup	800.4	660.9	665.0	556.4	669.2	746.2	796.5
Quadruped Run	690.4	440.7	425.3	303.9	485.9	372.1	515.3
Quadruped Walk	387.7	259.7	260.6	280.3	320.3	302.2	350.7
Reacher Easy	736.2	403.4	422.0	511.5	503.1	613.2	701.3
Walker Run	705.5	320.4	325.8	264.1	352.2	460.8	541.6
Walker Stand	965.9	799.8	724.6	790.2	784.6	916.7	960.5
Walker Walk	955.1	795.2	871.5	706.0	924.0	916.1	938.5
Mean	709.2	505.9	470.2	479.3	540.7	608.7	662.4
GRU%	-	67.0	63.2	64.3	73.8	81.9	90.4

TABLE I

SCORES FOR DMC VISUAL INPUTS CONTROL EXPERIMENTS AT 1M FRAMES.

Parameter	Value	Description	Parameter	Value	Description
τ V_{th}	2.0 1.0	Soma decay parameter Threshold potential	g _L g _B	1.0 1.0	Leaky conductance Basal conductance
V_{reset}	0.0	Reset potential	β	1.0	Gating parameter
T_s τ_A τ_B	8 ms 2.0 2.0	Simulation Times Apical decay parameter Basal decay parameter	lr	$1.0e^{-4}$	Learning rate

TABLE II
PARAMETER SETTINGS FOR EXPERIMENTS.

Additionally, the Spiking-WM model exceeds the performance of Dreamer(GRU) in four tasks: Cartpole Balance, Cup Catch, Finger Spin, Finger Turn Hard. In contrast, other spiking neuron-based world models generally do not outperform Dreamer(GRU) in terms of task performance, with the exception of TC-LIF on the Cartpole Balance Sparse task. These results demonstrate the advantages of our proposed MCN over existing simple neuron models in implementing the Spiking-WM model and highlight the potential of biologically plausible brain-inspired spiking neural networks for model-based reinforcement learning algorithms.

Using the trained model, we sampled the predictions of world models with different neurons for the first 50 steps of the Walker Run task, as shown in Figure 3. Here, T=0 represents the input state, and T=5-50 represents the model's Decoder prediction outputs. The results show that our proposed MCN-based world model, Spiking-WM, achieves better predictive capabilities, leading to higher final task scores compared to other neural world models.

B. Sequential Speech Data Experiments

To validate the potential of the proposed multi-compartment neuron model in handling sequence models, we conducted tests on the speech datasets SHD, TIMIT and LibriSpeech, the resuts are shown in Table III. Different neural models share an

TABLE III

COMPARISON OF DIFFERENT NEURON TYPES ON LONG-SEQUENCE DATASETS. IN THE "NEURON TYPE" CATEGORY, "FF" DENOTES NEURONS THAT PRIMARILY COMPUTE INFORMATION IN A FEEDFORWARD MANNER, WHILE "REC" REPRESENTS NEURONS WITH AN ENHANCED DESIGN FOR RECURRENT INFORMATION PROCESSING WITHIN THE NEURON ITSELF.

Dataset	Method	Neuron Type	Params (M)	Acc(%)
	GRU	Rec	3.5	89.92
	LIF [39]	FF	2.4	72.06
	aLIF [29]	Rec	2.5	88.64
SHD	PSN [40]	FF	2.0	85.86
	LIF-RNN [30]	Rec	2.1	80.34
	TC-LIF [17]	FF	2.1	89.45
	Ours	Rec	2.5	89.57
	GRU	Rec	14.1	83.75
	LIF [39]	FF	13.9	78.35
	aLIF [29]	Rec	11.4	82.10
TIMIT	PSN [40]	FF	11.8	80.05
	LIF-RNN [30]	Rec	19.6	81.28
	TC-LIF [17]	FF	11.9	80.90
	Spiking-WM(Ours)	Rec	14.0	83.01
	GRU	Rec	5.1	90.22
	LIF [39]	FF	3.9	80.16
	aLIF [29]	Rec	3.7	87.08
LibriSpeech	PSN [40]	FF	2.6	87.56
100h	LIF-RNN [30]	Rec	3.8	88.25
	TC-LIF [17]	FF	3.9	88.59
	Spiking-WM(Ours)	Rec	3.9	88.77

identical network architecture and training parameter configurations when applied to the same dataset. For the SHD dataset, the network adopts a fully connected structure comprising two hidden layers, each containing 128 neurons, with the STSC [41] spike processing method implemented at the input layer, which has been demonstrated to significantly enhance the performance of SNNs in handling spike information [42], [43].

We conducted comparative experiments on the TIMIT and

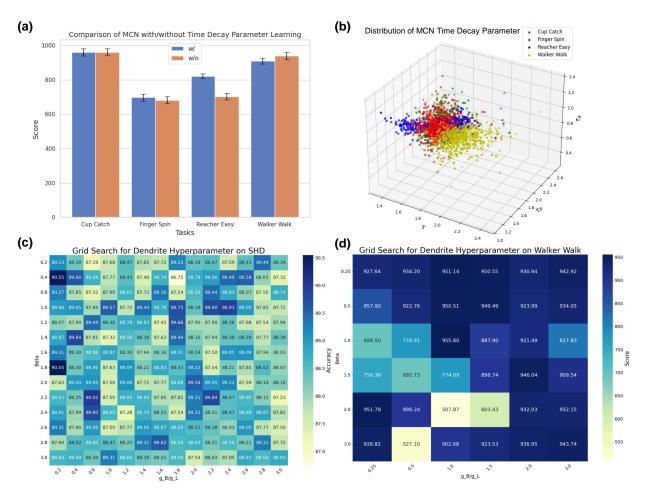


Fig. 4. Experimental results on the impact of MCN parameters on model performance. (a) Comparison results of experiments on MCN with and without learnable time decay parameters. (b) The distibution of soma, basal dendrite and apical dendrite time parameters after learning. (c) Grid search of basal dendrite conductance parameters and apical dendrite gate parameter on SHD dataset. (d) Grid search of basal dendrite conductance parameters and apical dendrite gate parameter on Walker Walk.

LibriSpeech 100h datasets using a two-layer neural network with the same recurrent connection and experimental settings as same in paper [30]. The TIMIT dataset is a widely used benchmark for speech recognition, containing phonetically diverse sentences spoken by multiple speakers, while LibriSpeech 100h is a subset of the larger LibriSpeech corpus, consisting of approximately 100 hours of transcribed English speech derived from audiobook recordings. The network architecture follows a two-layer recurrent structure, maintaining the same parameter configurations as in the referenced study LIF-RNN [30]. Experimental results demonstrate that our proposed spiking-WM model outperforms other spiking neuron models on the SHD, TIMIT and LibriSpeech 100h datasets, achieving performance comparable to that of GRU, which is consistent with the results of the DeepMind visual control experiments.

C. Analysis of MCN Parameters

Compared to the LIF neuron model, the multi-compartment neuron model introduces more parameters through the refinement of dendritic structures, thus providing more adjustable components for the neuron model. In our experiment, we investigated the impact of MCN parameters on model performance. For the membrane time constants of the soma (τ)

and dendritic membrane (τ_a , τ_b), we applied the approach from PLIF [44], treating them as learnable parameters and conducting experiments on the DMC task. The results are shown in Figure 4(a)(b). In the four experimental tasks, the learnable membrane potential decay parameters in the Finger Spin and Reacher Easy tasks led to performance improvements. However, in the Cup Catch task, the learnable decay parameters did not yield higher scores compared to the fixed decay parameters, and even resulted in a negative performance gain in the Walker Walk task.

The statistical distribution of the trained membrane potential decay parameters for dendrites and the soma is presented in Figure 4(b). Although the final distributions of the optimized membrane potential decay parameters exhibit slight variations across the four tasks, a significant portion of the distribution regions remain overlapping. Combined with the results in Figure 4(a), this suggests that adjustments to different membrane time constants in the MCN model do not substantially impact overall model performance. Instead, the key contribution of the MCN model lies in the incorporation of dendritic computational processes.

Furthermore, we analyzed the impact of the basal dendritic conductance parameter $\frac{g_B}{q_L}$ and the apical dendritic nonlinear

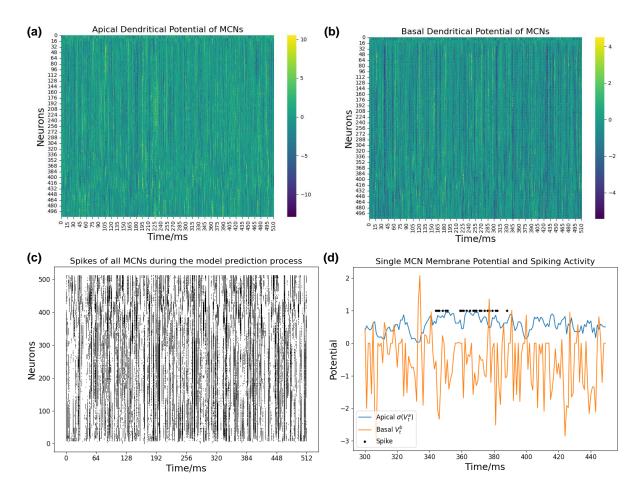


Fig. 5. Statistics of dendritic membrane potential and spike firing activity of all multi-compartment neurons at task Walker Run prediction process. (a) The apical dendritical potential of MCNs. (b) The basal dendritical potential of MCNs. (c) The spike sequential of MCNs during the model prediction process. (d) Details of the dendritic membrane potential and spike firing of the 200th neuron.

function parameter β on the MCN model. Using a grid search approach, we tested the model performance with different values of $\frac{g_B}{g_L}$ and β on the SHD dataset and the DMC Walker Walk task. The results are shown in Figure 4(c)(d).

The parameter search results indicate that the synergistic interaction between apical and basal dendrites significantly impacts the performance of the MCN model. On the SHD dataset, higher accuracy is observed in the upper-left region of Figure 4(c) ($\frac{g_B}{g_L} \leq 1.8, \, \beta \leq 1.8$). In the Walker Walk task, $\frac{g_B}{g_L}$ and β exhibit a more distinct adaptation relationship, with high scores concentrated in the upper-right region of Figure 4(d). Moreover, better performance is more likely to be achieved when β falls within the range of 0.5 to 1.0 and $\frac{g_B}{g_L}$ within the range of 1.0 to 1.5.

D. Analysis of MCN Spiking Activity

We further examined the membrane potentials and spike output of MCN neurons during the model prediction phase of the Spiking-WM in the DMC Walker Run task, as shown in Figure 5. During this process, the dendritic membrane potentials of MCN neurons exhibit synchronized activity over time, as illustrated in Figure 5(a) and Figure 5(b), which leads to a burst firing pattern of the neuron, as shown in

Figure 5(c). Furthermore, the frequency of apical dendritic membrane potential activity is higher than that of the basal dendrites, as indicated by the larger areas of alternating light and dark regions in the statistical plot.

The Figure 5(d) presents the recorded dendritic membrane potential, somatic membrane potential, and spike firing of a single neuron (Neuron 200) within the 300 ms to 450 ms interval. To facilitate observation of the regulatory effect of apical dendrites on neuronal activity, we have included the gating signal $z[t] = \sigma(V^a[t])$ (the blue line) of the apical dendrite in the figure. The neuron exhibits significant spike firing activity during the 350 ms to 380 ms period, during which the gating signal of the apical dendritic membrane potential remains within the range of $0.5 \sim 1.0$ for an extended duration. At this time, the acpial dendritic membrane potential is greater than zero, corresponding to the bright regions in Figure 5(a). In contrast, although the basal dendritic membrane potential exhibits large fluctuations at other recorded time points, the corresponding apical dendrite remains in a scaled inhibition state (with the gating signal mostly below 0.5), preventing the neuron from generating spike firing. These findings highlight the modulatory effect of apical dendritic activity on neuronal spike output in the proposed MCN model.

V. CONCLUSION

This work introduces a multi-compartment neuron-based spiking world model, demonstrating its effectiveness in modelbased SNN deep reinforcement learning and long-term memory tasks. Through comprehensive evaluations on Deep-Mind Control suits and large-scale speech datasets, our approach consistently outperforms conventional SNN models and achieves competitive performance relative to ANN-based world models. Moreover, our analysis of neuronal membrane potentials, spiking behavior, and dendritic parameter optimization underscores the significance of dendritic information processing. We show that coordinated dendritic activity plays a crucial role in shaping neuronal responses and improving overall model performance. These findings provide new insights into the functional importance of dendritic computation in SNNs, paving the way for more biologically inspired and efficient neural architectures in future research.

A. Data availability

The model of this research is one of the core and part of BrainCog Embot [45]. BrainCog Embot is an Embodied AI platform under the Brain-inspired Cognitive Intelligence Engine (BrainCog) framework, which is an open-source Brain-inspired AI platform based on Spiking Neural Network.

ACKNOWLEDGMENTS

This study is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB1010302), the funding from Institute of Automation, Chinese Academy of Sciences (Grant No. E411230101), Postdoctoral Fellowship Program of CPSF (Grant No.GZC20232994) the National Natural Science Foundation of China (Grant No. 62406325).

REFERENCES

- K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing," *Nature*, vol. 575, no. 7784, pp. 607–617, 2019.
- [2] N. Rathi, I. Chakraborty, A. Kosta, A. Sengupta, A. Ankit, P. Panda, and K. Roy, "Exploring neuromorphic computing based on spiking neural networks: Algorithms to hardware," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–49, 2023.
- [3] Y. Li and Y. Zeng, "Efficient and accurate conversion of spiking neural network with burst spikes," arXiv preprint arXiv:2204.13271, 2022.
- [4] Y. Luo, M. Xu, C. Yuan, X. Cao, L. Zhang, Y. Xu, T. Wang, and Q. Feng, "Siamsnn: siamese spiking neural networks for energy-efficient object tracking," in *International Conference on Artificial Neural Networks*. Springer, 2021, pp. 182–194.
- [5] Y. Sun, Y. Zeng, and Y. Li, "Solving the spike feature information vanishing problem in spiking deep q network with potential based normalization," Frontiers in Neuroscience, vol. 16, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins. 2022.953368
- [6] Z. Zhou, Y. Zhu, C. He, Y. Wang, S. Yan, Y. Tian, and L. Yuan, "Spikformer: When spiking neural network meets transformer," arXiv preprint arXiv:2209.15425, 2022.
- [7] M. E. Larkum, J. J. Zhu, and B. Sakmann, "A new cellular mechanism for coupling inputs arriving at different cortical layers," *Nature*, vol. 398, no. 6725, pp. 338–341, 1999.
- [8] S. L. Smith, I. T. Smith, T. Branco, and M. Häusser, "Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo," *Nature*, vol. 503, no. 7474, pp. 115–120, 2013.

- [9] A. Gidon, T. A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, and M. E. Larkum, "Dendritic action potentials and computation in human layer 2/3 cortical neurons," *Science*, vol. 367, no. 6473, pp. 83–87, 2020.
- [10] P. Poirazi, T. Brannon, and B. W. Mel, "Pyramidal neuron as two-layer neural network," *Neuron*, vol. 37, no. 6, pp. 989–999, 2003.
- [11] R. Urbanczik and W. Senn, "Learning by the dendritic prediction of somatic spiking," *Neuron*, vol. 81, no. 3, pp. 521–528, 2014.
- [12] J. Sacramento, R. Ponte Costa, Y. Bengio, and W. Senn, "Dendritic cortical microcircuits approximate the backpropagation algorithm," *Advances in neural information processing systems*, vol. 31, 2018.
- [13] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli et al., "A deep learning framework for neuroscience," *Nature neuroscience*, vol. 22, no. 11, pp. 1761–1770, 2019.
- [14] C. Capone, C. Lupo, P. Muratore, and P. S. Paolucci, "Burst-dependent plasticity and dendritic amplification support target-based learning and hierarchical imitation learning," arXiv preprint arXiv:2201.11717, 2022.
- [15] J. Schemmel, L. Kriener, P. Müller, and K. Meier, "An accelerated analog neuromorphic hardware system emulating nmda-and calciumbased non-linear dendrites," in 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 2217–2226.
- [16] S. Yang, B. Deng, J. Wang, H. Li, M. Lu, Y. Che, X. Wei, and K. A. Loparo, "Scalable digital neuromorphic architecture for largescale biophysically meaningful neural network with multi-compartment neurons," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 1, pp. 148–162, 2019.
- [17] S. Zhang, Q. Yang, C. Ma, J. Wu, H. Li, and K. C. Tan, "Tc-lif: A two-compartment spiking neuron model for long-term sequential modelling," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 15, 2024, pp. 16838–16847.
- [18] X. Li, J. Tang, and J. Lai, "Learning high-performance spiking neural networks with multi-compartment spiking neurons," in *International Conference on Image and Graphics*. Springer, 2023, pp. 91–102.
- [19] K. Gregor, D. Jimenez Rezende, F. Besse, Y. Wu, H. Merzic, and A. van den Oord, "Shaping belief states with generative environment models for rl," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [20] T. M. Moerland, J. Broekens, A. Plaat, C. M. Jonker et al., "Model-based reinforcement learning: A survey," Foundations and Trends® in Machine Learning, vol. 16, no. 1, pp. 1–118, 2023.
- [21] Q. Liu, T. Yu, Y. Bai, and C. Jin, "A sharp analysis of model-based reinforcement learning with self-play," in *International Conference on Machine Learning*. PMLR, 2021, pp. 7001–7010.
- [22] D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi, "Dream to control: Learning behaviors by latent imagination," arXiv preprint arXiv:1912.01603, 2019.
- [23] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, "Mastering diverse domains through world models," arXiv preprint arXiv:2301.04104, 2023.
- [24] B. F. Behabadi, A. Polsky, M. Jadi, J. Schiller, and B. W. Mel, "Location-dependent excitatory synaptic interactions in pyramidal neuron dendrites," *PLoS computational biology*, vol. 8, no. 7, p. e1002599, 2012.
- [25] M. Häusser and B. Mel, "Dendrites: bug or feature?" Current opinion in neurobiology, vol. 13, no. 3, pp. 372–383, 2003.
- [26] T. Branco and M. Häusser, "The single dendritic branch as a fundamental functional unit in the nervous system," *Current opinion in neurobiology*, vol. 20, no. 4, pp. 494–502, 2010.
- [27] T. Jarsky, A. Roxin, W. L. Kath, and N. Spruston, "Conditional dendritic spike propagation following distal synaptic activation of hippocampal ca1 pyramidal neurons," *Nature neuroscience*, vol. 8, no. 12, pp. 1667– 1676, 2005.
- [28] W. Gerstner and W. M. Kistler, Spiking neuron models: Single neurons, populations, plasticity. Cambridge university press, 2002.
- [29] B. Yin, F. Corradi, and S. M. Bohté, "Accurate and efficient time-domain classification with adaptive spiking recurrent neural networks," *Nature Machine Intelligence*, vol. 3, no. 10, pp. 905–913, 2021.
- [30] W. Ponghiran and K. Roy, "Spiking neural networks with improved inherent recurrence dynamics for sequential learning," in *Proceedings* of the AAAI Conference on Artificial Intelligence, vol. 36, no. 7, 2022, pp. 8001–8008.
- [31] B. J. Lansdell, P. R. Prakash, and K. P. Kording, "Learning to solve the credit assignment problem," arXiv preprint arXiv:1906.00889, 2019.
- [32] D. Patel, H. Hazan, D. J. Saunders, H. T. Siegelmann, and R. Kozma, "Improved robustness of reinforcement learning policies upon conversion to spiking neuronal network platforms applied to atari breakout game," *Neural Networks*, vol. 120, pp. 108–115, 2019.

- [33] W. Tan, D. Patel, and R. Kozma, "Strategy and benchmark for converting deep q-networks to event-driven spiking neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 11, 2021, pp. 9816–9824.
- [34] Z. Bing, C. Meschede, G. Chen, A. Knoll, and K. Huang, "Indirect and direct training of spiking neural networks for end-to-end control of a lane-keeping vehicle," *Neural Networks*, vol. 121, pp. 21–36, 2020.
- [35] A. S. Lele, Y. Fang, J. Ting, and A. Raychowdhury, "Learning to walk: Spike based reinforcement learning for hexapod robot central pattern generation," in 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS). IEEE, 2020, pp. 208–212.
- [36] G. Tang, N. Kumar, R. Yoo, and K. Michmizos, "Deep reinforcement learning with population-coded spiking neural network for continuous control," in *Conference on Robot Learning*. PMLR, 2021, pp. 2016– 2029
- [37] C. Capone and P. S. Paolucci, "Towards biologically plausible dreaming and planning in recurrent spiking networks," arXiv preprint arXiv:2205.10044, 2022.
- [38] Y. Sun, F. Zhao, Z. Zhao, and Y. Zeng, "Multi-compartment neuron and population encoding powered spiking neural network for deep distributional reinforcement learning," *Neural Networks*, vol. 182, p. 106898, 2025. [Online]. Available: https://www.sciencedirect.com/ science/article/pii/S089360802400827X
- [39] B. Cramer, Y. Stradmann, J. Schemmel, and F. Zenke, "The heidelberg spiking data sets for the systematic evaluation of spiking neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 2744–2757, 2020.
- [40] W. Fang, Z. Yu, Z. Zhou, D. Chen, Y. Chen, Z. Ma, T. Masquelier, and Y. Tian, "Parallel spiking neurons with high efficiency and ability to learn long-term dependencies," Advances in Neural Information Processing Systems, vol. 36, 2024.
- [41] C. Yu, Z. Gu, D. Li, G. Wang, A. Wang, and E. Li, "Stsc-snn: Spatio-temporal synaptic connection with temporal convolution and attention for spiking neural networks," *Frontiers in Neuroscience*, vol. 16, p. 1079357, 2022.
- [42] G. Shen, D. Zhao, and Y. Zeng, "Eventmix: An efficient data augmentation strategy for event-based learning," *Information Sciences*, vol. 644, p. 119170, 2023.
- [43] Y. Li, Y. Sun, X. He, Y. Dong, D. Zhao, and Y. Zeng, "Parallel spiking unit for efficient training of spiking neural networks," in 2024 International Joint Conference on Neural Networks (IJCNN). IEEE, 2024, pp. 1–8.
- [44] W. Fang, Z. Yu, Y. Chen, T. Masquelier, T. Huang, and Y. Tian, "Incorporating learnable membrane time constant to enhance learning of spiking neural networks," in *Proceedings of the IEEE/CVF international* conference on computer vision, 2021, pp. 2661–2671.
- [45] "The braincog embot engine," https://www.brain-cog.network/embot.