ENERGY-BASED SINGLE-LAYER SPIKING GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spiking graph neural networks have shown strong potential for efficient graph representation learning due to their discrete communication mechanism. However, most existing methods still inherit the hierarchically stacked message-passing paradigm of conventional graph neural networks and rely on multi-step output averaging for prediction, making them vulnerable to over-smoothing, spike information degradation, and high inference latency. To address these issues, we propose an energy-based single-layer spiking graph representation learning model that formulates node representation learning as a finite-horizon spiking dynamical process on graphs. Specifically, we decouple the neuronal dynamics into three components: a constant external current for mitigating over-smoothing, a bidirectional graph attention current for structural interaction, and a Hopfield current for refining node states via prior feature memory. Furthermore, we establish a rigorous theoretical framework based on a global energy functional, mathematically proving that the network's temporal evolution intrinsically executes an energy minimization process, thereby guaranteeing system stability and convergence. Extensive node classification experiments on eight benchmark datasets demonstrate that the proposed method achieves competitive performance while exhibiting clear advantages in dynamical stability, anti-over-smoothing capability, and inference efficiency.

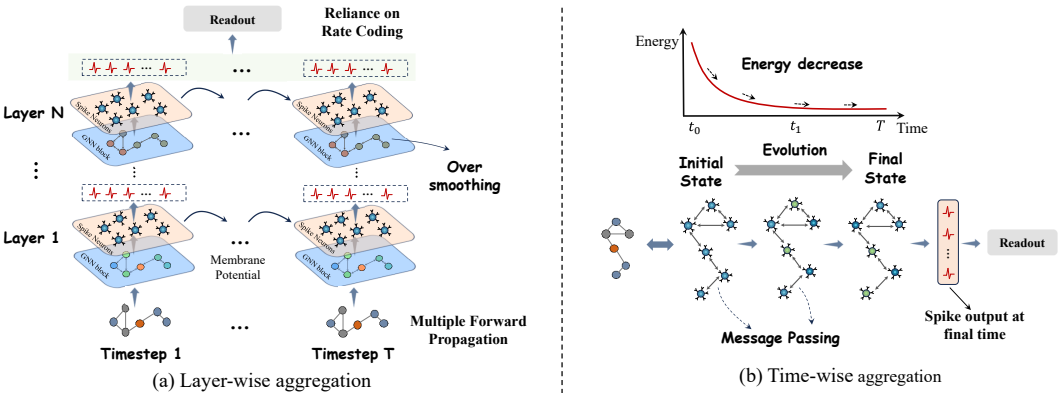
1 INTRODUCTION

Recent years have witnessed graph neural networks (GNNs) (Scarselli et al., 2008) achieving remarkable breakthroughs in various fields (Agrawal et al., 2024; Wu et al., 2025; Deng et al., 2026). However, as GNNs grow deeper and more complex to capture long-range dependencies and intricate graph topologies, they inherently require intensive message-passing operations (Hosseini et al., 2025), thereby limiting their applicability on energy-constrained edge devices (Kose et al., 2024; Zhou et al., 2025). To pursue more energy-efficient and biologically plausible architectures, several studies have integrated spiking neural networks (SNNs) (Maass, 1997) into the GNN paradigm. The event-driven nature of SNNs converts costly dense matrix multiplications into sparse spike additions, which reduces computational complexity and enables ultra-low power consumption on neuromorphic hardware (Zhu et al., 2022; Sun et al., 2024b).

It is noteworthy that most spiking GNNs typically adopt a straightforward adaptation strategy: inheriting the hierarchical spatial stacking architecture of traditional continuous GNNs for graph convolutions, while relying on rate coding mechanisms across multiple time steps to process graph features (Zhu et al., 2022; Li et al., 2023a; Cao et al., 2026; Buschmann et al., 2025). These models suffer from the following two principal limitations (shown in Figure 1 (a)).

- **Spike vanishing and over-smoothing.** The deep stacking architecture exacerbates information loss of discrete binary spike signals during forward propagation across the graph topology, frequently leading to spike vanishing (Zheng et al., 2021). Concurrently, repeated neighborhood aggregation in deep networks readily induces over-smoothing of node features (Oono & Suzuki, 2019).
- **Rate coding constrains theoretical performance and biological plausibility.** Rate coding averages outputs over multiple time steps; more steps make the model a continuous GNN with firing-rate activation (Sengupta et al., 2019). This constrains its theoretical upper bound

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Layer-wise aggregation (b) Time-wise aggregation

Figure 1: (a) Current spiking GNNs predominantly adopt a hierarchical stacking architecture, where graph information is aggregated at each spatial layer. This paradigm inevitably induces the over-smoothing problem and relies heavily on biologically implausible rate coding. (b) Our method naturally models the graph topology as the connections within SNNs and follows a process of continuous energy decrease , intrinsically aggregating graph information as time evolves.

and incurs high computational overhead from multiple forward propagations, undermining SNNs' energy advantage (Cao et al., 2026). Moreover, it conflicts with biological reality, in which the human brain decides efficiently from the final state alone (Beer & Barak, 2024).

To address both challenges, we propose a biologically interpretable, efficient, and theoretically grounded spiking GNN architecture. Departing from the conventional spatial stacking paradigm, we establish an equivalence between graph-structured message-passing and the temporal evolution of SNNs (see Figure 1 (b)). By reformulating the spiking neurons dynamics, the input current of spike neurons is decoupled into an internal interaction current and an external driving current. The internal interaction current comprises two synergistic components: a graph attention current for dynamic local structural aggregation, and a continuous Hopfield network (Hopfield, 1982) current that refines node states via prior feature memory as a global attractor. The external driving current is defined as a constant injection of initial node features throughout the entire temporal evolution, providing a continuous driving force that rigorously mitigates over-smoothing. Notably, our method achieves outstanding performance relying solely on the spike output at the final time step.

Drawing inspiration from the energy functions characterizing network states in modern Hopfield networks, we transcend architectural design to establish a rigorous theoretical framework grounded in an energy functional. Specifically, we formulate a global energy functional comprising three distinct components: membrane recovery potential, internal aggregation potential, and external driving potential. These components effectively capture the states of inputs, outputs, and the interactions among neurons within the network (see Section 3.2). Consequently, the temporal evolution of spiking neurons on graphs is mathematically proven to be equivalent to an energy minimization process. We further prove that, irrespective of whether the SNNs employs a hard step function or a soft surrogate threshold function, the total energy monotonically decreases and remains bounded from below, thereby providing a theoretical guarantee for system stability and convergence.

To validate the effectiveness of the proposed method, we extensively evaluate its performance on eight widely used node classification benchmarks. Our method achieves state-of-the-art performance among spiking GNNs and remains competitive with classical continuous GNNs. Furthermore, it exhibits strong robustness to over-smoothing: under prolonged temporal evolution, neither model performance nor Dirichlet energy shows any degradation. These findings conclusively demonstrate the superiority of our temporal evolution paradigm over conventional deep spatial stacking.

2 Related Work

Spiking GNNs. In recent years, many studies have combined SNNs with GNNs to reduce computational overhead. SpikingGCN (Zhu et al., 2022) proposed an end-to-end fusion framework, validating

the feasibility of neuromorphic graph computing. SpikeNet (Li et al., 2023a) replaced RNNs with SNNs for scalable temporal graph learning. SpikingGCL (Li et al., 2023b) introduced unsupervised graph contrastive learning in the spiking domain. SpikeGT (Sun et al., 2024b) and SGT (Qian & Tao, 2025) integrated Transformer mechanisms with SNNs to build spike-based graph attention. MSG (Sun et al., 2024a) extended SNNs to Riemannian manifold spaces and proposed loop-free manifold differentiation techniques, effectively alleviating the latency issues caused by traditional backpropagation through time. However, these works remain confined to the hierarchical stacking structure of ANNs, which limits the physical interpretability and dynamic flexibility of the network.

Continuous Graph Dynamics. Viewing GNNs as continuous dynamical systems overcomes the limitations of discrete layer stacking. Graph ordinary differential equations (Graph ODEs) (Xhonneux et al., 2020; Chamberlain et al., 2021a) formulate message-passing as continuous-time node state evolution, which mitigates over-smoothing to some extent. Subsequent extensions include coupled oscillatory dynamics (Rusch et al., 2022b), stochastic differential equations (Xing et al., 2024; Cui et al., 2025), and continuous physical diffusion (Choi et al., 2023; Liu et al., 2024). Recent works further capture signed network topologies (Chen et al., 2024), time-varying interaction patterns Eliasof et al. (2024), and convection-diffusion processes Wang et al. (2023; 2025). While these methods draw inspiration from classical physics, our work naturally adopts SNNs dynamics, constraining node states to the spiking domain and equating graph updates to a global energy descent process.

Energy Functionals and Hopfield Networks. Hopfield networks have long served as models of associative memory and energy minimization in biological nervous systems (Hopfield, 1982; 1984; Krotov & Hopfield, 2016). Recent advances in modern Hopfield networks revealed their intrinsic connection to the Transformer attention mechanism (Ramsauer et al., 2020; Krotov & Hopfield, 2020). Building upon this, the energy Transformer (Hoover et al., 2023) innovatively constructs an architecture driven entirely by an energy function, directly translating attention forward propagation into an explicit energy descent process. Beyond architectural designs, the energy-based paradigm provides a rigorous Lyapunov-like framework for analyzing complex network dynamics, guaranteeing natural convergence of iterative forward propagation to an energy minimum (Betteti & Laurenti, 2026; Eliasof et al., 2025; Mali et al., 2024). This offers a highly interpretable mechanism for associative memory retrieval with subsequent extensions to more complex architectures (Marullo et al., 2026). In our work, we draw inspiration from this energy-driven philosophy by decoupling the internal interaction current of spiking neurons into a graph attention component and a local Hopfield component, thereby enhancing biological plausibility while providing a rigorous energy descent guarantee for the discrete forward evolution of SNNs.

3 METHODOLOGY

Formally, let $\mathcal{G} = (\mathcal{V}, \mathcal{R}, X)$ be a directed graph with node set $\mathcal{V} = \{v_1, v_2, \cdots, v_N\}$, edge relations $\mathcal{R}$, and initial node features X. For any node $v_i \in \mathcal{V}$, denote its in-neighbors and out-neighbors as $\mathcal{N}_i^{in} = \{v_j \in \mathcal{V} \mid (v_j, v_i) \in \mathcal{R}\}$ and $\mathcal{N}_i^{out} = \{v_j \in \mathcal{V} \mid (v_i, v_j) \in \mathcal{R}\}$, respectively. The combined neighborhood $\mathcal{N}_i = \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}$ comprehensively captures bidirectional interactions. Existing literature distinguishes two firing mechanisms in SNNs: hard threshold and soft threshold (Hunsberger & Eliasmith, 2015; Neftci et al., 2019). To evaluate the generalizability of our method, we investigate both mechanisms. Specifically, we associate each node v_i with a membrane potential $u_i \in \mathbb{R}^D$ and a spike emission s_i, where D denotes the hidden feature dimension. Under hard threshold mapping, $s_i = \Theta(u_i - u_{th}) \in \{0, 1\}^D$ with spike firing threshold u_{th}, where $\Theta(\cdot)$ is the Heaviside step function. Under soft threshold mapping, $s_i = \sigma(u_i) \in (0, 1)^D$ represents a continuous analog state, where $\sigma(\cdot)$ is a differentiable soft threshold function.

3.1 SPIKING NEURON DYNAMICS EQUATION

The leaky integrate-and-fire (LIF) (Gerstner & Kistler, 2002) is the most widely used spiking neuron model. In conventional spiking GNNs, the input current to a LIF neuron is typically defined as the aggregated information from topological neighbors in the preceding spatial layer. This layer-wise dependency limits the physical interpretability and dynamic flexibility of the network. As illustrated in Figure 2, we depart from this spatial paradigm by modeling the graph topology directly within the temporal evolution of SNNs. Building upon the classic LIF model, we decouple the input current

of node v_i into an **external driving current** $I_{ext,i}$ and an **internal interaction current** $I_{int,i}$, and formulate its dynamical update equation via

$$\tau \frac{du_i}{dt} = -u_i + I_{ext,i} + I_{int,i}\left(s_i, \{s_j\}_{j\in\mathcal{N}_i}\right),\ i = 1, 2, \cdots, N, \tag{1}$$

where τ is the membrane time constant; $I_{ext,i}$ encodes the initial features of node v_i and continuously injects them into the network, given by $I_{ext,i} = W_{ext}X_i$ with learnable parameter W_{ext}. $I_{int,i}$ consists of the graph attention current $I_{ATT,i}$ and the Hopfield current $I_{HN,i}$, *i.e.*,

$$I_{int,i} = I_{ATT,i} + I_{HN,i}, \tag{2}$$

where $I_{ATT,i}$ aggregates the incoming and outgoing neighbors of node v_i to capture the structural topological dependencies of the network, defined as

$$I_{ATT,i} = \underbrace{W^T \sum_{v_j\in\mathcal{N}_i^{in}} \alpha_{ij}(Ws_j)}_{\text{Forward Current}} + \underbrace{W^T \sum_{v_k\in\mathcal{N}_i^{out}} \alpha_{ki}(Ws_k)}_{\text{Backward Current}}, \tag{3}$$

where $\alpha_{ij} = \exp\left(\beta s_i^T W^T W s_j\right) / \left(\sum_{m\in\mathcal{N}_i^{in}} \exp\left(\beta s_i^T W^T W s_m\right)\right)$ denotes the attention coefficient between nodes v_i and v_j, with learnable parameter W and scaling factor β. The Hopfield current $I_{HN,i}$ models the intrinsic associative memory of nodes v_i, calculated by

$$I_{HN,i} = (P^T P)s_i, \tag{4}$$

where $P \in \mathbb{R}^{M\times D}$ represents the memory matrix stored internally within the network with M memory patterns. The spike state s_i emitted by the spiking neuron at node v_i activates the associative memory matrix $P^T P$, which then converts it into the feedback drive current $I_{HN,i}$.

Over-smoothing Resistance. For node membrane potentials $\mathbf{U} = [u_1, u_2, \cdots, u_N]^T$, Dirichlet energy $E(\mathbf{U}) = \frac{1}{N}\mathrm{Tr}(\mathbf{U}^T\mathbf{L}\mathbf{U})$ is typically leveraged to quantify the degree of over-smoothing in GNNs (Cai & Wang, 2020; Chamberlain et al., 2021b; Rusch et al., 2022a), where $\mathbf{L}$ is the graph Laplacian matrix. The external driving current drives network dynamics and the following theorem shows that it can effectively mitigate the over-smoothing problem in GNNs.

Theorem 1 (Resistance to Over-Smoothing)**.** *Let $H = I_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$. For a fixed graph, assume that the dynamics in Eq. (1) converge to an equilibrium (U^*, S^*). If the external-current matrix $I_{\text{ext}} \in \mathbb{R}^{N\times D}$ satisfies $\sigma_2(HI_{\text{ext}}) > 0$, where σ_2 denotes the second-largest singular value, then*

$$\lim_{t\to\infty} E(U(t)) = E(U^*) \geq \frac{\lambda_2(L)}{N}\|HU^*\|_F^2 > 0, \tag{5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Intuitively, when all nodes share the same spike state, the centered internal current has rank at most one, whereas $\sigma_2(HI_{\text{ext}}) > 0$ guarantees that the centered external drive contains richer non-uniform structure that cannot be fully canceled by the internal dynamics. Hence, the equilibrium membrane potentials cannot collapse to an identical node representation, and their Dirichlet energy remains strictly positive.

3.2 Energy Functional

To ensure convergence of the spiking dynamical system, we construct a global energy functional, rendering the network update process equivalent to optimization over this energy surface. For the spike states of all nodes $\mathbf{S} = [s_1, s_2, \cdots, s_N]^T$, we define energy functional $\mathcal{E}(\mathbf{S})$ as comprising three terms: membrane recovery potential, internal aggregation potential, and external driving potential

$$\mathcal{E}(\mathbf{S}) = \frac{1}{N}\sum_{i\in\mathcal{V}} \left[\mathcal{E}_{mem}(s_i) - \mathcal{E}_{int}(s_i, \{s_j\}_{j\in\mathcal{N}_i}) - \mathcal{E}_{ext}(s_i)\right], \tag{6}$$

where $\mathcal{E}_{ext}(s_i) = s_i^T I_{ext,i}$ represents the work performed by spiking neurons under external current, while the remaining two terms govern the complex internal dynamics.

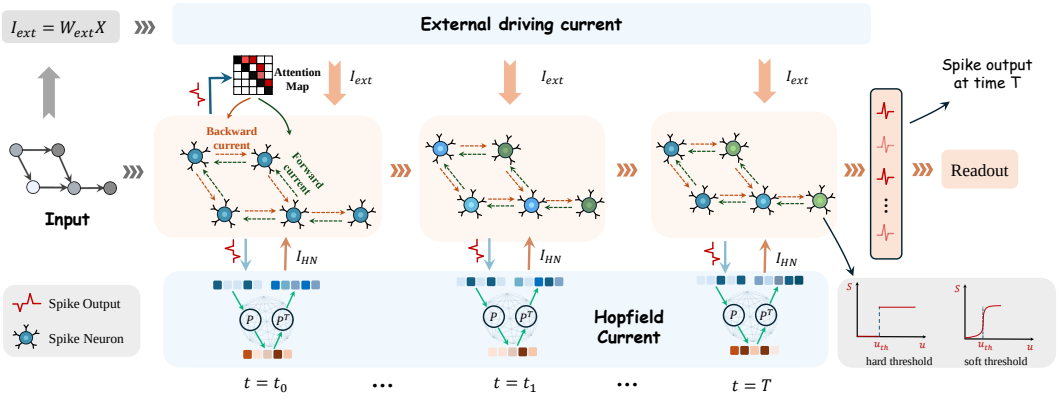


Figure 2: Overall architecture of the proposed temporal evolution paradigm for spiking GNNs. The initial node features X are mapped to constant external driving currents I_{ext} continuously injected into the network. Driven by the external current, the internal graph attention current and Hopfield current, spiking neurons dynamically update their states. Bypassing traditional rate-coding average, the model applies a readout head solely to the sparse spike output at the final time.

Membrane Recovery Potential $\mathcal{E}_{mem}$. The membrane recovery potential quantifies the endogenous tension a spiking neurons must overcome to leave its resting state and sustain excitation. Its definition depends on the firing function. Under the soft-threshold function, the surrogate mapping $s_i = \sigma(u_i) \in (0,1)^D$ is monotonically increasing, thus its inverse function $\sigma^{-1}(\cdot)$ exists. The membrane recovery potential for node v_i is then defined as the integral of the inverse function

$$\mathcal{E}_{mem}(s_i) = \sum_{d=1}^{D} \int_0^{s_{i,d}} \sigma^{-1}(v)dv,\ d = 1, 2, \cdots, D. \tag{7}$$

Under the hard-threshold function, this integral reduces to a linear threshold energy barrier

$$\mathcal{E}_{mem}(s_i) = u_{th}\|s_i\|_1. \tag{8}$$

Internal Aggregation Potential $\mathcal{E}_{int}$. The internal aggregation potential comprises the graph attention potential energy $\mathcal{E}_{ATT}$ and the local Hopfield energy $\mathcal{E}_{HN}$, *i.e.*,

$$\mathcal{E}_{int}(s_i, \{s_j\}_{j\in\mathcal{N}_i}) = \mathcal{E}_{ATT,i} + \mathcal{E}_{HN,i} = \frac{1}{\beta}\log \sum_{j\in\mathcal{N}_i^{in}} \exp\left(\beta s_i^T W^T W s_j\right) + \frac{1}{2}\|Ps_i\|_2^2, \tag{9}$$

where $\mathcal{E}_{ATT}$ measures consistency between adjacent nodes. Using the LogSumExp function as a smooth maximum approximation, it creates an energy landscape that encourages nodes to align their spiking states with their strongly correlated neighbors. $\mathcal{E}_{HN}$ serves as an associative memory reservoir, measuring the match between current spiking states and the memory patterns encoded P.

3.3 Energy Evolution Analysis

To theoretically guarantee convergence, we analyze the energy evolution of the network and provide separate stability analyses for the hard-threshold and soft-threshold firing function.

Theorem 2 (Continuous System Stability). *Under soft-threshold firing $s_i = \sigma(u_i) \in (0,1)^D$, global energy functional $\mathcal{E}(\mathbf{S})$ is bounded from below, i.e., $\inf_{\mathbf{S}} \mathcal{E}(\mathbf{S}) > -\infty$; total energy monotonically decreases, i.e., $\frac{d\mathcal{E}}{dt} \leq 0$, as spiking neurons evolve according to differential Equation* (1).

The detailed proof is given in Appendix B.2. Theorem 2 demonstrates that under soft-threshold firing, the network's temporal evolution performs gradient descent on the global energy landscape, and boundedness guarantees convergence to a stable fixed point [1].

[1] In practice, we discretize the dynamics using the explicit Euler method. The corresponding finite-step analysis is provided in Appendix C.

Table 1: Performance comparison. The best results are highlighted in **bold**, and the second-best are underlined. '-' indicates out-of-memory or results not reported in the original paper.

Methods	Cora	CiteSeer	PubMed	Computers	Photo	CS	Physics	arXiv
GCN	86.1 ± 0.2	75.9 ± 0.4	88.2 ± 0.5	86.5 ± 0.5	92.4 ± 0.2	92.5 ± 0.4	95.7 ± 0.5	70.4 ± 0.3
GAT	86.7 ± 0.7	78.5 ± 0.4	86.8 ± 0.3	86.9 ± 0.2	92.5 ± 0.3	92.3 ± 0.2	95.4 ± 0.3	70.6 ± 0.3
GGD	86.2 ± 0.2	75.5 ± 0.1	84.2 ± 0.1	88.0 ± 0.1	92.9 ± 0.2	93.1 ± 0.1	95.3 ± 0.0	$\mathbf{71.6 \pm 0.5}$
DGI	86.3 ± 0.2	78.9 ± 0.2	86.2 ± 0.1	84.0 ± 0.5	91.6 ± 0.2	92.2 ± 0.6	94.5 ± 0.5	65.1 ± 0.4
AS-GCL	-	-	-	$\mathbf{89.9 \pm 0.3}$	93.5 ± 0.2	93.5 ± 0.3	95.2 ± 0.2	-
SpikeNet	83.5 ± 0.3	71.4 ± 0.5	83.9 ± 0.4	88.0 ± 0.7	92.9 ± 0.1	93.4 ± 0.2	95.8 ± 0.7	66.8 ± 0.1
SpikingGCN	87.7 ± 0.4	77.7 ± 0.3	87.5 ± 0.5	86.9 ± 0.3	92.6 ± 0.7	92.6 ± 0.3	94.3 ± 0.1	55.8 ± 0.7
SpikeGCL	87.4 ± 0.6	77.6 ± 0.6	88.8 ± 0.3	88.9 ± 0.3	93.0 ± 0.1	92.8 ± 0.1	95.2 ± 0.6	70.9 ± 0.1
SpikeGT	-	-	-	81.0 ± 1.1	90.7 ± 0.4	91.9 ± 0.6	94.4 ± 1.6	-
MSG	-	-	-	89.3 ± 0.2	93.1 ± 0.1	92.7 ± 0.0	95.9 ± 0.1	-
GRAND	$\mathbf{88.3 \pm 0.7}$	77.5 ± 0.6	87.6 ± 0.6	88.6 ± 0.3	$\mathbf{94.5 \pm 0.2}$	93.4 ± 0.1	93.9 ± 0.1	68.5 ± 0.2
GRAND++	$\underline{88.0 \pm 0.2}$	78.5 ± 0.6	87.9 ± 0.4	88.6 ± 0.3	$\underline{94.2 \pm 0.1}$	93.5 ± 0.1	94.0 ± 0.1	69.0 ± 0.3
GREAD	87.2 ± 0.9	78.4 ± 0.5	89.3 ± 0.3	89.0 ± 0.3	$\underline{94.2 \pm 0.5}$	93.3 ± 0.3	93.6 ± 0.3	-
TANGO	87.0 ± 0.5	78.6 ± 1.0	87.6 ± 2.5	$\underline{89.8 \pm 0.1}$	93.7 ± 0.3	93.3 ± 0.5	96.0 ± 0.1	-
GraphCON	85.9 ± 0.2	75.8 ± 1.4	87.7 ± 0.3	86.1 ± 0.5	93.1 ± 0.5	90.7 ± 0.2	95.5 ± 0.1	-
Ours(hard)	86.7 ± 0.4	$\mathbf{79.5 \pm 0.4}$	$\mathbf{90.8 \pm 0.6}$	$\mathbf{89.9 \pm 0.2}$	$\underline{94.2 \pm 0.4}$	$\underline{93.7 \pm 0.2}$	$\mathbf{96.3 \pm 0.1}$	$\underline{71.5 \pm 0.0}$
Ours(soft)	86.5 ± 0.2	$\underline{79.0 \pm 0.6}$	$\underline{90.6 \pm 0.2}$	89.7 ± 0.5	$\underline{94.2 \pm 0.3}$	$\mathbf{94.0 \pm 0.3}$	$\underline{96.2 \pm 0.1}$	71.4 ± 0.1

Theorem 3 (Discrete System Convergence)**.** *Under hard threshold firing, when a membrane potential $u_{i,d}$ exceeds u_{th} and triggers a one-step discrete state flip $\Delta s_{i,d}$, discrete global energy strictly decreases, i.e., $\Delta \mathcal{E} < 0$.*

The detailed proof appears in Appendix B.3. Under the hard-threshold firing function, the output space of spiking neurons is finite. Therefore, Theorem 3 guarantees that the energy decreases monotonically and eventually converges to a stable state.

4 Experiments

Datasets. We conduct extensive experiments on the fundamental node classification task across eight widely used graph benchmark datasets, including three citation datasets (Cora, CiteSeer, PubMed) (Sen et al., 2008), two co-purchase datasets (Shchur et al., 2018) (Computers, Photo), two co-author datasets (Shchur et al., 2018) (CS, Physics), and one large-scale dataset (Hu et al., 2020) (arXiv). Our dataset splitting strategy consistently follows SpikeGCL (Li et al., 2023b) to ensure a fair and reliable comparison. Detailed dataset statistics and experimental settings can be found in Appendix A.

Baseline. We compare with three categories of GNN baselines. The first category consists of semi-supervised/unsupervised methods, including GCN (Kipf & Welling, 2016), GAT (Veličković et al., 2017), DGI (Veličković et al., 2018), GGD (Zheng et al., 2022), and AS-GCL (Liu et al., 2025). The second category consists of SNN-based GNN baselines, namely SpikeNet (Li et al., 2023a), SpikingGCN (Zhu et al., 2022), SpikeGCL (Li et al., 2023b), SpikeGT (Sun et al., 2024b), and MSG (Sun et al., 2024a). The third category includes ODE-based baselines GRAND (Chamberlain et al., 2021a), GRAND++ (Thorpe et al., 2022), GraphCON (Rusch et al., 2022c), GREAD (Choi et al., 2023),, and TANGO (Eliasof et al., 2025).

4.1 Model Performance

The comprehensive results over all eight datasets are summarized in Table 1. Overall, the accuracy of both hard and soft threshold variants substantially surpasses that of existing SNNs baselines on the majority of evaluated datasets. [2] This improvement empirically validates our temporal evolution paradigm, guided by a global energy functional, effectively overcomes the representation bottlenecks

[2]Since the soft and hard threshold models exhibit highly similar performance, we default to the hard-threshold model in subsequent experiments to prevent empirical redundancy.

Table 2: Ablation study of different current components in our method. Each variant removes one current component from the full model while keeping other settings unchanged.

Variants	Cora	CiteSeer	PubMed	Computers	Photo	CS	Physics	arXiv
w/o I_{HN}	86.2 ± 0.5	78.6 ± 1.0	90.2 ± 0.2	89.7 ± 0.3	94.0 ± 0.6	93.5 ± 0.3	96.1 ± 0.1	70.5 ± 0.4
w/o I_{ATT}	75.7 ± 1.3	75.4 ± 0.6	89.2 ± 0.4	81.4 ± 0.4	87.1 ± 0.6	91.5 ± 0.3	94.8 ± 0.1	53.8 ± 0.3
w/o I_{ext}	31.9 ± 0.0	20.5 ± 2.6	41.0 ± 0.3	28.7 ± 10.7	25.4 ± 0.0	15.9 ± 5.5	42.9 ± 15.3	15.4 ± 7.8
w/o I_{int}	75.7 ± 0.8	75.7 ± 0.8	89.1 ± 0.5	81.3 ± 0.3	87.2 ± 0.4	91.4 ± 0.2	94.7 ± 0.2	53.8 ± 0.1
Full Model	$\mathbf{86.7 \pm 0.4}$	$\mathbf{79.5 \pm 0.4}$	$\mathbf{90.8 \pm 0.6}$	$\mathbf{89.9 \pm 0.2}$	$\mathbf{94.2 \pm 0.4}$	$\mathbf{93.7 \pm 0.2}$	$\mathbf{96.3 \pm 0.1}$	$\mathbf{71.5 \pm 0.0}$

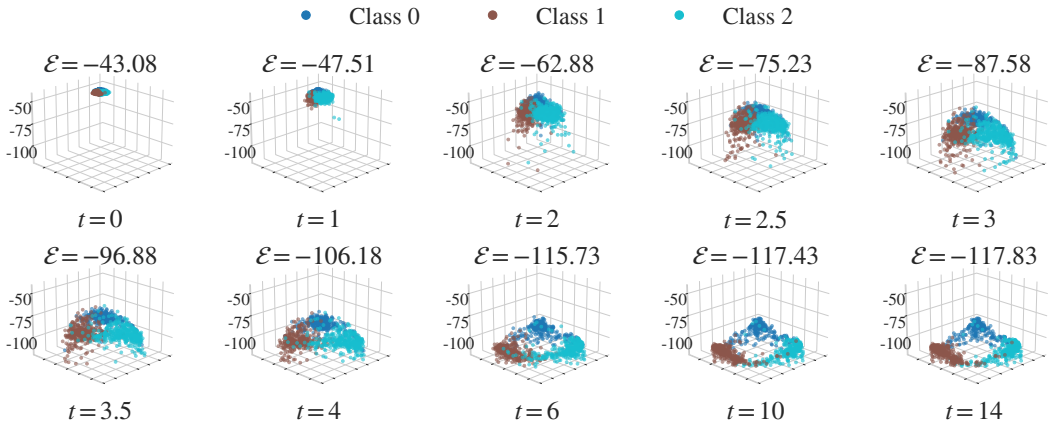


Figure 3: Visualization of node representations and energy evolution on PubMed. As the dynamics evolve from $t = 0$ to $t = 14$, the global energy decreases monotonically while the node embeddings separate into class-specific clusters. Results for other datasets are provided in the Appendix F.

of prior spiking GNNs. Moreover, our method achieves performance competitive with or superior to continuous GNNs and ODE-based models. Despite operating in the discrete spike domain, it ranks first on CiteSeer, PubMed, CS and Physics, and second on Computers, Photo, and arXiv. These results show that our biologically interpretable architecture bridges the performance gap between discrete spiking models and traditional continuous GNNs without sacrificing expressive power [3].

4.2 Ablation Study

To evaluate the contribution of each core component in our method, we systematically remove specific currents and compare the complete model with four variants: namely w/o I_{HN}, w/o I_{ATT}, w/o I_{ext}, and w/o I_{int}. The results are summarized in Table 2. Note that the complete model consistently maintains the best performance. Specifically, the removal of I_{ATT} leads to a dramatic performance drop across the board (*e.g.*, from 86.7% to 75.7% on Cora), highlighting its critical role in capturing structural dependencies. Furthermore, removing I_{ext} leads to a catastrophic performance collapse (*e.g.*, to 15.9% on CS), as it is the sole mechanism injecting input features into the dynamics, rendering the network effectively feature-blind. Finally, discarding I_{HN} results in a slight but consistent performance degradation, confirming its effectiveness in refining node representations and optimizing the energy landscape. These results verify that the three current components jointly contribute to a balanced energy landscape for discriminative representation learning.

4.3 Visualization of Energy-Driven Evolution

To intuitively understand the energy-driven representation learning of our model, we visualize the temporal evolution of node embeddings on PubMed in Figure 3. Initially ($t = 0$), node representations

[3]The code is available at `https://anonymous.4open.science/r/Energy-Based-Single-Layer-Spiking-Graph-Representation-Learning-2235/`

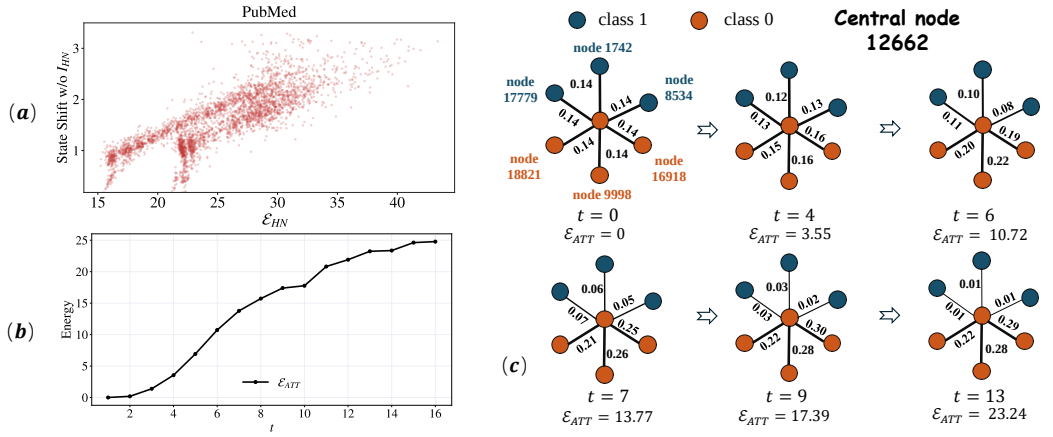
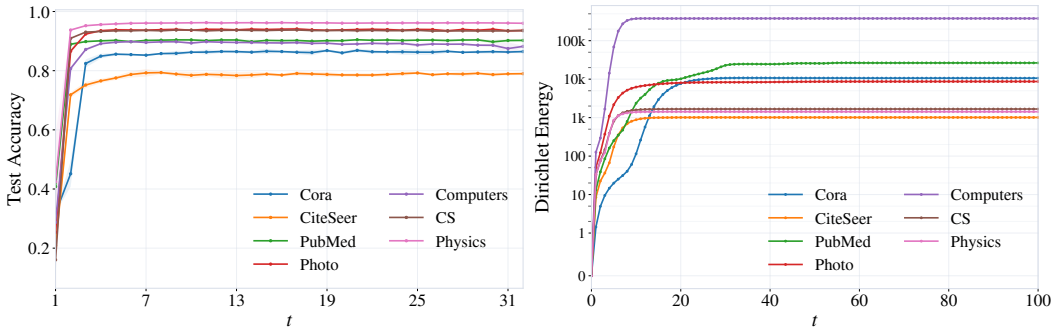


Figure 4: (a) Hopfield branch analysis on PubMed. Nodes with higher Hopfield energy undergo larger state changes when the Hopfield current is removed. (b) (c) Evolution of energy and neighbor attention scores for node 12662. As time evolves, attention energy continuously increases, while scores increase for intra-class neighbors and decrease for inter-class neighbors.



(a) ACC (b) Dirichlet energy

Figure 5: Accuracy and Dirichlet energy under long-horizon evolution. (a) Accuracy converges rapidly and remains stable for large T. (b) Dirichlet energy remains positive, indicating strong resistance to over-smoothing.

are highly entangled and indistinguishable. As dynamics evolve under the synergistic effect of decoupled internal and external currents, global energy strictly decreases from -43.08 to -117.83, and node embeddings progressively separate into class-specific clusters. This visual evidence empirically validates our theoretical proof of strict energy descent and demonstrate that energy minimization intrinsically optimizes representations for classification. In addition, we provide the energy descent curves, with detailed results reported in the Appendix D.

4.4 Dynamics Roles of the Attention and Hopfield Branches

To distinguish the roles of internal energy terms, we analyzed Hopfield energy and attention energy separately. For Hopfield energy (Figure 4(a)), an ablation study removes I_{HN} and measures deviation in each node's final state. Nodes with higher energy exhibit greater state deviation, indicating stronger reliance on Hopfield correction. For attention energy (Figure 4(b)(c)), we track its temporal evolution of node 12662 in PubMed whose local neighborhood initially contains a complex mix of both intra-class and inter-class neighbors, making it an ideal candidate to observe the attention shift. As the dynamics evolve, attention energy continuously increases, with intra-class neighbor scores increasing (*e.g.*, node 18821 from 0.14 to 0.22) and inter-class scores decreasing (*e.g.*, node 1742 from 0.14 to 0.01). Thus, attention energy drives nodes to maintain consistency with highly correlated neighbors while preserving adaptive structural discrimination during the entire temporal evolution process.

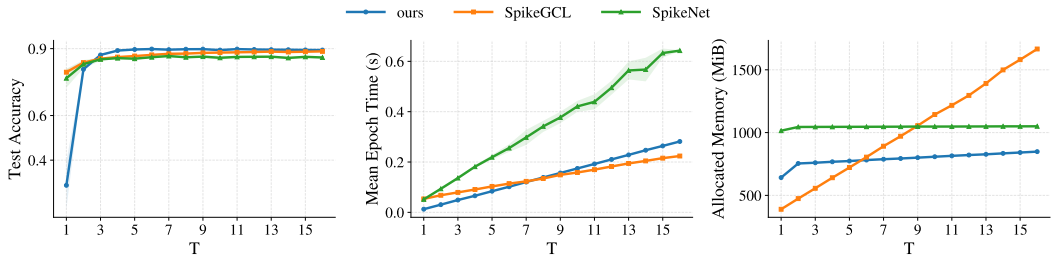


Figure 6: Accuracy, training time per epoch, and memory consumption over time on Computers. SpikeGCL incurs a higher memory footprint and SpikeNet a longer execution time, whereas our method achieves excellent efficiency in both metrics while maintaining accuracy.

4.5 Anti-Oversmoothing and Stability Analysis

To evaluate robustness over extended temporal evolution, we track test accuracy and Dirichlet energy as T increases, and the result is shown in Figure 5. For the test datasets, the accuracy rises rapidly in early steps and gradually converges to a stable, high-performance plateau (Figure 5a). Remarkably, even when the temporal depth extends to large values, the model exhibits no signs of performance degradation and preserves consistent predictive behavior throughout the evolution process. This is further corroborated by Dirichlet energy evolution (Figure 5b), which remains stable even when T reaches 100. These results demonstrate that the energy-driven dynamics effectively preserve high-frequency separability of node features, avoiding the over-smoothing issue and maintaining stable accuracy across extended propagation steps.

4.6 Hardware Efficiency

To demonstrate hardware efficiency, we evaluated the memory footprint and execution time of our method against two representative spiking graph models, SpikeGCL and SpikeNet, each with a time-step hyperparameter T. Results in Figure 6 show that our method achieves significant advantages in both metrics, corroborating the inherent efficiency of the single-layer spiking architecture. Notably, our method attains stable accuracy at $T = 5$, whereas SpikeGCL requires a substantially larger T to reach higher accuracy, inevitably increasing its computational burden. Furthermore, we evaluate the theoretical inference cost of our method and observe consistent advantages over existing approaches. Detailed analysis and comparisons are provided in Appendix E.

5 Conclusion

We propose a novel Spiking GNN paradigm that replaces spatial stacking with temporal evolution, grounded in an explicit energy functional. By decoupling the neuronal dynamics into internal interactions and external driven currents, the proposed framework effectively mitigates the pervasive issues of spike vanishing and over-smoothing while providing an interpretable dynamical perspective on graph representation learning. Mathematical analysis further characterizes the energy-descent behavior of the system and establishes corresponding stability properties under the stated conditions. Empirically, our method achieves state-of-the-art performance among spiking GNNs and remains competitive with classical continuous GNNs, while preserving high Dirichlet energy and significantly reducing computational cost by eliminating the overhead associated with deep spatial stacking.

Limitations and Future Work. A limitation of our work is that the theoretical convergence analysis is established under specific assumptions on the firing dynamics and energy landscape, while the behavior of more complex spiking mechanisms, such as adaptive neuronal models and stochastic spike dynamics, remains unexplored. Extending the energy formulation to a broader class of biologically plausible neurons is an interesting direction for future research. In future work, we plan to explore adaptive energy-based spiking graph models with learnable dynamics, extend the theoretical analysis to more general update rules and continuous-time settings, and investigate their deployment on neuromorphic hardware to fully exploit the event-driven computation advantages of SNNs.

AI USE STATEMENT

Generative AI tools were used only for language polishing and improving the readability of the manuscript. All AI-assisted edits were reviewed by the authors, who take full responsibility for the final content.

REFERENCES

Nimesh Agrawal, Anuj Kumar Sirohi, Sandeep Kumar, et al. No prejudice! fair federated graph neural networks for personalized recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 10775–10783, 2024.

Chen Beer and Omri Barak. Revealing and reshaping attractor dynamics in large networks of cortical neurons. *PLOS Computational Biology*, 20(1):e1011784, 2024.

Simone Betteti and Luca Laurenti. Hybrid energy-based models for physical ai: Provably stable identification of port-hamiltonian dynamics. *arXiv preprint arXiv:2604.00277*, 2026.

Fernando Vera Buschmann, Horacio Rotstein, and Vincent Oria. neuro-inspired enhancing spiking graph convolutional networks. In *2025 IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1–7. IEEE, 2025.

Chen Cai and Yusu Wang. A note on over-smoothing for graph neural networks. *arXiv preprint arXiv:2006.13318*, 2020.

Buqing Cao, Qian Peng, Xiang Xie, Liang Chen, Min Shi, and Jianxun Liu. Spiking heterogeneous graph attention networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 19853–19861, 2026.

Ben Chamberlain, James Rowbottom, Maria I Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. Grand: Graph neural diffusion. In *International Conference on Machine Learning*, pp. 1407–1418. PMLR, 2021a.

Ben Chamberlain, James Rowbottom, Maria I Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. Grand: Graph neural diffusion. In *International Conference on Machine Learning*, pp. 1407–1418. PMLR, 2021b.

Lanlan Chen, Kai Wu, Jian Lou, and Jing Liu. Signed graph neural ordinary differential equation for modeling continuous-time dynamics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8292–8301, 2024.

Jeongwhan Choi, Seoyoung Hong, Noseong Park, and Sung-Bae Cho. Gread: Graph neural reaction-diffusion networks. In *International Conference on Machine Learning*, pp. 5722–5747. PMLR, 2023.

Wenjun Cui, Qiyu Kang, Xuhao Li, Kai Zhao, Wee Peng Tay, Weihua Deng, and Yidong Li. Neural variable-order fractional differential equation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 16109–16117, 2025.

Lijia Deng, Ziyang Dong, Zhengling Yang, Bo Gong, and Le Zhang. Graph learning in bioinformatics: A survey of graph neural network architectures, biological graph construction and bioinformatics applications. *Biomolecules*, 16(2):333, 2026.

Moshe Eliasof, Eldad Haber, Eran Treister, and Carola-Bibiane B Schönlieb. On the temporal domain of differential equation inspired graph neural networks. In *International Conference on Artificial Intelligence and Statistics*, pp. 1792–1800. PMLR, 2024.

Moshe Eliasof, Eldad Haber, and Carola-Bibiane Schönlieb. Tango: Graph neural dynamics via learned energy and tangential flows. *arXiv preprint arXiv:2508.05070*, 2025.

Wulfram Gerstner and Werner M Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

Benjamin Hoover, Yuchen Liang, Bao Pham, Rameswar Panda, Hendrik Strobelt, Duen Horng Chau, Mohammed Zaki, and Dmitry Krotov. Energy transformer. *Advances in Neural Information Processing Systems*, 36:27532–27559, 2023.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, 1984.

Ryien Hosseini, Filippo Simini, Venkatram Vishwanath, Rebecca Willett, and Henry Hoffmann. Sketch-augmented features improve learning long-range dependencies in graph neural networks. *arXiv preprint arXiv:2511.03824*, 2025.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in Neural Information Processing Systems*, 33:22118–22133, 2020.

Eric Hunsberger and Chris Eliasmith. Spiking deep networks with lif neurons. *arXiv preprint arXiv:1510.08829*, 2015.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Habib Taha Kose, Jose Nunez-Yanez, Robert Piechocki, and James Pope. A survey of computationally efficient graph neural networks for reconfigurable systems. *Information*, 15(7):377, 2024.

Dmitry Krotov and John Hopfield. Large associative memory problem in neurobiology and machine learning. *arXiv preprint arXiv:2008.06996*, 2020.

Dmitry Krotov and John J Hopfield. Dense associative memory for pattern recognition. *Advances in Neural Information Processing Systems*, 29, 2016.

Jintang Li, Zhouxin Yu, Zulun Zhu, Liang Chen, Qi Yu, Zibin Zheng, Sheng Tian, Ruofan Wu, and Changhua Meng. Scaling up dynamic graph representation learning via spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8588–8596, 2023a.

Jintang Li, Huizhe Zhang, Ruofan Wu, Zulun Zhu, Baokun Wang, Changhua Meng, Zibin Zheng, and Liang Chen. A graph is worth 1-bit spikes: When graph contrastive learning meets spiking neural networks. *arXiv preprint arXiv:2305.19306*, 2023b.

Jiaxu Liu, Xinping Yi, Sihao Wu, Xiangyu Yin, Tianle Zhang, Xiaowei Huang, and Shi Jin. Continuous geometry-aware graph diffusion via hyperbolic neural pde. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 231–249. Springer, 2024.

Ruyue Liu, Rong Yin, Yong Liu, Xiaoshuai Hao, Haichao Shi, Can Ma, and Weiping Wang. Asgcl: Asymmetric spectral augmentation on graph contrastive learning. *IEEE Transactions on Multimedia*, 2025.

Wolfgang Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.

Ankur Mali, Tommaso Salvatori, and Alexander Ororbia. Tight stability, convergence, and robustness bounds for predictive coding networks. *arXiv preprint arXiv:2410.04708*, 2024.

Chiara Marullo, Giuseppe Buonaiuto, Francesco Gargiulo, and Massimo Esposito. Et-kan: an energy-based transformer model with kolmogorov–arnold network for image reconstruction. *Complex & Intelligent Systems*, 2026.

Emre O Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, 2019.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. *arXiv preprint arXiv:1905.10947*, 2019.

Jun Qian and Haotian Tao. Sgt: an energy-efficient spiking graph transformer for graph representation learning. In *Fifth International Conference on Image Processing and Intelligent Control (IPIC 2025)*, volume 13782, pp. 360–366. SPIE, 2025.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*, 2020.

T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *International Conference on Machine Learning*, pp. 18888–18909. PMLR, 2022a.

T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *International Conference on Machine Learning*, pp. 18888–18909. PMLR, 2022b.

T. Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18888–18909. PMLR, 2022c.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–93, 2008.

Abhronil Sengupta, Yuting Ye, Robert Wang, Chiao Liu, and Kaushik Roy. Going deeper in spiking neural networks: Vgg and residual architectures. *Frontiers in Neuroscience*, 13:95, 2019.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

Li Sun, Zhenhao Huang, Qiqi Wan, Hao Peng, and Philip S Yu. Spiking graph neural network on riemannian manifolds. *Advances in Neural Information Processing Systems*, 37:34025–34055, 2024a.

Yundong Sun, Dongjie Zhu, Yansong Wang, Zhaoshuo Tian, Ning Cao, and Gregory O'Hared. Spikegraphormer: A high-performance graph transformer with spiking graph attention. *arXiv preprint arXiv:2403.15480*, 2024b.

Matthew Thorpe, Tan Minh Nguyen, Hedi Xia, Thomas Strohmer, Andrea Bertozzi, Stanley Osher, and Bao Wang. Grand++: Graph neural diffusion with a source term. In *International Conference on Learning Representations*, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. *arXiv preprint arXiv:1809.10341*, 2018.

Tangjun Wang, Zehao Dou, Chenglong Bao, and Zuoqiang Shi. Diffusion mechanism in residual neural network: theory and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2):667–680, 2023.

Zhiqiang Wang, Xiaoyi Wang, and Jianqing Liang. Csg-ode: Controlsynth graph ode for modeling complex evolution of dynamic graphs. In *Forty-second International Conference on Machine Learning*, 2025.

Bin Wu, Yihang Wang, Yuanhao Zeng, Jiawei Liu, Jiashu Zhao, Cheng Yang, Yawen Li, Long Xia, Dawei Yin, and Chuan Shi. Graph foundation models for recommendation: A comprehensive survey. *arXiv preprint arXiv:2502.08346*, 2025.

Louis-Pascal Xhonneux, Meng Qu, and Jian Tang. Continuous graph neural networks. In *International Conference on Machine Learning*, pp. 10432–10441. PMLR, 2020.

Yucheng Xing, Jacqueline Wu, Yingru Liu, Xuewen Yang, and Xin Wang. Evolved differential model for sporadic graph time-series prediction. *Intelligent and Converged Networks*, 5(3):237–247, 2024.

Man Yao, Jiakui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. *Advances in Neural Information Processing Systems*, 36:64043–64058, 2023.

Hanle Zheng, Yujie Wu, Lei Deng, Yifan Hu, and Guoqi Li. Going deeper with directly-trained larger spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11062–11070, 2021.

Yizhen Zheng, Shirui Pan, Vincent Cs Lee, Yu Zheng, and Philip S Yu. Rethinking and scaling up graph contrastive learning: An extremely efficient approach with group discrimination. In *Advances in Neural Information Processing Systems*, 2022.

Ao Zhou, Jianlei Yang, Tong Qiao, Yingjie Qi, Zhi Yang, Weisheng Zhao, and Chunming Hu. Gcode: Efficient device-edge co-inference for gnns via architecture-mapping co-search. *IEEE Transactions on Computers*, 75(1):58–72, 2025.

Zulun Zhu, Jiaying Peng, Jintang Li, Liang Chen, Qi Yu, and Siqiang Luo. Spiking graph convolutional networks. *arXiv preprint arXiv:2205.02767*, 2022.

A Datasets

We conduct extensive experiments on eight widely used benchmark datasets for node classification. Detailed statistics of these datasets, including the number of nodes, edges, feature dimensions, and classes, are summarized in Table 3. Specifically, for the citation network datasets, we adopt the full split. For the co-author and co-purchase datasets, we randomly split the nodes into training, validation, and testing sets with a ratio of 1:1:8. For the ogbn-arxiv dataset, we use the official public split. The hidden dimension is set to 64 for all datasets except arXiv, which is set to 256. All experiments were conducted on an NVIDIA RTX 3090 GPU with 24 GB of memory.

Table 3: Statistics of the benchmark datasets used in our experiments.

Dataset	**Cora**	**CiteSeer**	**PubMed**	**Computers**	**Photo**	**CS**	**Physics**	**ogbn-arXiv**
Nodes	2,708	3,327	19,717	13,752	7,650	18,333	34,493	169,343
Edges	5,429	4,732	44,338	245,861	119,081	81,894	247,962	1,166,243
Features	1,433	3,703	500	767	745	6,805	8,415	128
Classes	7	6	3	10	8	15	5	40

B Proof

B.1 Proof of Theorem 1

Proof. Let $H = I_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$ and define the node-wise constant representation subspace as $\mathcal{M} = \{\mathbf{1}a^\top : a \in \mathbb{R}^D\}$.

For the combinatorial Laplacian $L = L^\top \succeq 0$ of a connected undirected support graph, the graph Poincaré inequality gives

$$E(U) = \frac{1}{N}\operatorname{Tr}(U^\top L U) \geq \frac{\lambda_2(L)}{N}\|HU\|_F^2,$$

where $\|HU\|_F = \operatorname{dist}_F(U, \mathcal{M})$.

Assume that the dynamics in Eq. (1) converge to an equilibrium (U^*, S^*). At equilibrium, $U^* = I_{\text{ext}} + I_{\text{int}}(S^*)$. Define

$$\delta_* = \|H\left[I_{\text{ext}} + I_{\text{int}}(S^*)\right]\|_F = \|HU^*\|_F.$$

We prove $\delta_* > 0$ by contradiction. Suppose $\delta_* = 0$. Then $HU^* = 0$, and hence all nodes share the same equilibrium membrane potential, i.e., $U^* = \mathbf{1}(u^*)^\top$. Since all nodes use the same element-wise threshold mapping, their equilibrium spike states are also identical: $S^* = \mathbf{1}(s^*)^\top$.

Let $A = W^\top W$ and $Q = P^\top P$. Under the uniform spike state, all attention logits within the in-neighborhood of a node are identical, and thus $\alpha_{ij} = 1/d_i^{\text{in}}$, where $d_i^{\text{in}} = |\mathcal{N}_i^{\text{in}}|$. Therefore, the forward attention current becomes $I_{\text{fwd},i} = As^*$.

Define $\rho_i = \sum_{k \in \mathcal{N}_i^{\text{out}}} 1/d_k^{\text{in}}$. Then the backward attention current is $I_{\text{bwd},i} = \rho_i As^*$, while the Hopfield current is $I_{\text{HN},i} = Qs^*$. Hence,

$$I_{\text{int},i}(S^*) = Qs^* + (1 + \rho_i)As^*.$$

Let $\rho = [\rho_1, \ldots, \rho_N]^\top$. Stacking all internal currents row-wise yields

$$I_{\text{int}}(S^*) = \mathbf{1}(Qs^* + As^*)^\top + \rho(As^*)^\top.$$

Since $H\mathbf{1} = 0$, $HI_{\text{int}}(S^*) = H\rho(As^*)^\top$. The right-hand side is an outer product, so $\text{rank}(HI_{\text{int}}(S^*)) \leq 1$.

On the other hand, $HU^* = 0$ and the equilibrium condition imply $HI_{\text{ext}} = -HI_{\text{int}}(S^*)$. Thus, $\text{rank}(HI_{\text{ext}}) \leq 1$, which contradicts $\sigma_2(HI_{\text{ext}}) > 0$. Therefore, $\delta_* = \|HU^*\|_F > 0$.

Applying the Poincaré inequality at equilibrium gives

$$E(U^*) \geq \frac{\lambda_2(L)}{N}\|HU^*\|_F^2 = \frac{\lambda_2(L)}{N}\delta_*^2 > 0.$$

Finally, since $U(t) \to U^*$ and $E(U)$ is continuous,

$$\lim_{t\to\infty} E(U(t)) = E(U^*) \geq \frac{\lambda_2(L)}{N}\delta_*^2 > 0.$$

□

B.2 Proof of Theorem 2

Proof. First, we establish the monotonic decrease of the global energy. Taking the gradient of the total continuous energy functional $\mathcal{E}(\mathbf{S})$ with respect to the spiking state vector $s_i \in (0,1)^D$ of node v_i, we obtain:

$$\begin{aligned}
-N\nabla_{s_i}\mathcal{E}(\mathbf{S}) &= -\nabla_{s_i}\left(\sum_{d=1}^{D}\int_0^{s_{i,d}}\sigma^{-1}(v)dv\right) + \nabla_{s_i}\sum_{m\in\mathcal{V}}\mathcal{E}_{int}(s_m, \{s_j\}_{j\in\mathcal{N}_m}) + \nabla_{s_i}\left(s_i^T I_{ext,i}\right) \\
&= -\sigma^{-1}(s_i) + \nabla_{s_i}\sum_{m\in\mathcal{V}}\Big(\mathcal{E}_{ATT,m} + \mathcal{E}_{HN,m}\Big) + I_{ext,i} \\
&= -u_i + \left[W^T\sum_{v_j\in\mathcal{N}_i^{in}}\alpha_{ij}(Ws_j) + W^T\sum_{v_k\in\mathcal{N}_i^{out}}\alpha_{ki}(Ws_k)\right] + (P^TP)s_i + I_{ext,i} \\
&= -u_i + I_{int,i} + I_{ext,i} \\
&= \tau\frac{du_i}{dt}.
\end{aligned} \tag{10}$$

By differentiating the global energy functional with respect to time using the chain rule, the rate of change is given by:

$$\begin{aligned}\frac{d\mathcal{E}}{dt} &= \sum_{i\in\mathcal{V}} \left(\nabla_{s_i}\mathcal{E}\right)^T \frac{ds_i}{dt} \\ &= \sum_{i\in\mathcal{V}} \left(-\frac{\tau}{N}\frac{du_i}{dt}\right)^T \left(\mathrm{diag}(\sigma'(u_i))\frac{du_i}{dt}\right) \\ &= -\frac{\tau}{N}\sum_{i\in\mathcal{V}}\sum_{d=1}^{D} \sigma'(u_{i,d}) \left(\frac{du_{i,d}}{dt}\right)^2 .\end{aligned} \tag{11}$$

Given that the membrane time constant $\tau > 0$ and the surrogate derivative $\sigma'(u_{i,d}) > 0$, it strictly follows that $\frac{d\mathcal{E}}{dt} \leq 0$.

Next, we establish that the energy functional is bounded from below. Since the continuous surrogate mapping bounds the state space $\mathbf{S} \in (0,1)^{N\times D}$, the integral of the membrane recovery potential is strictly bounded. Furthermore, the external drive constitutes a linear product with bounded states, and the internal aggregation potential (comprising a bounded LogSumExp function and a positive semi-definite Hopfield quadratic form) is intrinsically bounded. Therefore, $\inf_{\mathbf{S}} \mathcal{E}(\mathbf{S}) > -\infty$. This bounded monotonic descent guarantees convergence to a stable fixed point. □

B.3 Proof of Theorem 3

Proof. The discrete global energy functional $\mathcal{E}(\mathbf{S})$ is formulated as:

$$\mathcal{E}(\mathbf{S}) = \frac{1}{N}\left[u_{th}\sum_{i\in\mathcal{V}}(\mathbf{1}^T s_i) - \sum_{i\in\mathcal{V}}\mathcal{E}_{int}(s_i) - \sum_{i\in\mathcal{V}} s_i^T I_{ext,i}\right]. \tag{12}$$

First, we establish the convexity of the total internal energy $\sum_{m\in\mathcal{V}}\mathcal{E}_{int,m}$ with respect to the state vector s_i. For the Hopfield energy $\mathcal{E}_{HN} = \frac{1}{2}s_i^T(P^TP)s_i$, its Hessian matrix is $\nabla^2\mathcal{E}_{HN} = P^TP \succeq 0$, guaranteeing positive semi-definiteness and convexity. The graph attention energy $\mathcal{E}_{ATT}$ is formulated as a LogSumExp (LSE) function. For cross-node interactions ($j \neq i$), its argument is an affine mapping of s_i (i.e., $s_i \mapsto \beta s_i^T W^T W s_j$). For the self-loop case ($j = i$), the argument becomes a quadratic form $\beta s_i^T W^T W s_i$. Because $W^TW \succeq 0$, this quadratic form is inherently convex. Since the LSE function is standardly convex and non-decreasing in each argument, its composition with these affine and convex quadratic mappings preserves overall convexity. Consequently, their sum is inherently convex.

Consider an asynchronous one-step discrete state flip occurring at a specific feature dimension d of node v_i, denoted by the scalar jump $\Delta s_{i,d} \in \{1, -1\}$. According to the first-order Taylor condition for convex functions ($f(y) \geq f(x) + f'(x)(y-x)$), the change in the total internal energy satisfies:

$$\sum_{m\in\mathcal{V}}\mathcal{E}_{int,m}(\mathbf{S}+\Delta\mathbf{S}) - \sum_{m\in\mathcal{V}}\mathcal{E}_{int,m}(\mathbf{S}) \geq \frac{\partial\sum_{m\in\mathcal{V}}\mathcal{E}_{int,m}}{\partial s_{i,d}}\Delta s_{i,d} = I_{int,i,d}\Delta s_{i,d}. \tag{13}$$

Let the total input current at this dimension be $I_{total,i,d} = I_{int,i,d} + I_{ext,i,d}$. Substituting the upper bound of the negative internal energy change into the exact global energy difference yields a strictly scalar expression:

$$\begin{aligned}\Delta\mathcal{E} &\leq \frac{1}{N}\left[u_{th}\Delta s_{i,d} - I_{int,i,d}\Delta s_{i,d} - I_{ext,i,d}\Delta s_{i,d}\right] \\ &= -\frac{1}{N}(I_{total,i,d} - u_{th})\Delta s_{i,d}.\end{aligned} \tag{14}$$

We analyze this scalar state jump in two physical scenarios. Because the discrete states change instantaneously at the exact crossing moment t^*, we rigorously evaluate the continuous temporal derivative at the left-sided limit t^{*-} (i.e., the state right before the flip) to justify the energy change:

- **Spike Firing ($\Delta s_{i,d} = 1$):** The membrane potential $u_{i,d}$ crosses u_{th} from below at time t^*, requiring the left-sided continuous temporal derivative $\left.\frac{du_{i,d}}{dt}\right|_{t^{*-}} > 0$. At this exact moment

evaluated from the left ($u_{i,d}(t^{*-}) = u_{th}$), the dynamics equation dictates $\tau \frac{du_{i,d}}{dt}\Big|_{t^{*-}} = -u_{th} + I_{total,i,d} > 0$, implying the driving current must satisfy $I_{total,i,d} > u_{th}$. Substituting this yields the energy change $\Delta\mathcal{E} \leq -\frac{1}{N}(I_{total,i,d} - u_{th})(1) < 0$.

- **Spike Reset ($\Delta s_{i,d} = -1$):** The membrane potential $u_{i,d}$ drops below u_{th} from above, requiring the left-sided derivative $\frac{du_{i,d}}{dt}\Big|_{t^{*-}} < 0$. Symmetrically, at this moment evaluated from the left, $\tau \frac{du_{i,d}}{dt}\Big|_{t^{*-}} = -u_{th} + I_{total,i,d} < 0$, which strictly implies $I_{total,i,d} < u_{th}$. Substituting this yields $\Delta\mathcal{E} \leq -\frac{1}{N}(I_{total,i,d} - u_{th})(-1) < 0$.

In both valid biological state transitions, the global discrete energy strictly decreases. □

C Energy Descent under Euler Discretization

The theoretical analysis in the main text characterizes the continuous dynamics and the discrete hard-threshold state transitions. In practice, however, the membrane dynamics are numerically integrated with a finite step size using the explicit Euler method. To connect the continuous formulation with its numerical implementation, we first derive sufficient step-size conditions for energy descent under soft-threshold and asynchronous hard-threshold updates. We then empirically examine the effect of the Euler step size and further evaluate the asynchronous hard-threshold variant covered by our theoretical analysis.

C.1 Soft-Threshold Updates

Let $\widetilde{\mathcal{E}} = N\mathcal{E}$, $h = \Delta t/\tau$, and $g_k = \nabla_S \widetilde{\mathcal{E}}(S_k)$. According to Eq. (15), the Euler update is

$$U_{k+1} = U_k - hg_k.$$

Since $S_k = \sigma(U_k)$, the element-wise mean value theorem gives a diagonal matrix $D_k = \mathrm{diag}(d_{k,1}, \ldots, d_{k,ND})$ such that

$$S_{k+1} - S_k = -hD_k g_k, \qquad 0 \preceq D_k \preceq M_\sigma I,$$

where $M_\sigma = \sup_u \sigma'(u)$.

Assume that $\nabla\widetilde{\mathcal{E}}$ is $L_{\widetilde{\mathcal{E}}}$-Lipschitz continuous over a bounded convex region containing S_k, S_{k+1}, and the line segment between them. By the descent lemma,

$$\widetilde{\mathcal{E}}(S_{k+1}) - \widetilde{\mathcal{E}}(S_k) \leq -h \sum_r d_{k,r} \left(1 - \frac{hL_{\widetilde{\mathcal{E}}} d_{k,r}}{2}\right) g_{k,r}^2.$$

Therefore, a sufficient condition for non-increasing discrete energy is $0 < h \leq 2/(M_\sigma L_{\widetilde{\mathcal{E}}})$, or equivalently,

$$0 < \Delta t \leq \frac{2\tau}{M_\sigma L_{\widetilde{\mathcal{E}}}}.$$

For the standard logistic sigmoid, $M_\sigma = 1/4$, and thus $0 < \Delta t \leq 8\tau/L_{\widetilde{\mathcal{E}}}$. This condition provides a sufficient, rather than necessary, bound for energy descent. We next consider the hard-threshold dynamics, for which the discontinuous spike mapping requires a separate analysis.

C.2 Asynchronous Hard-Threshold Updates

Consider one asynchronous Euler update of node i:

$$u_i^+ = (1-h)u_i + hI_{\mathrm{total},i}, \qquad h = \frac{\Delta t}{\tau},$$

followed by $s_i^+ = \Theta(u_i^+ - u_{\mathrm{th}})$, while all other node states remain fixed.

Table 4: Average minimum normalized step size $h = \Delta t/\tau$ at which energy fluctuations are first observed for the soft-threshold model over ten random seeds.

Dataset	Minimum h
Cora	1.24
CiteSeer	1.45
PubMed	1.18
Computers	1.34
Photo	1.28
CS	1.90
Physics	1.84

When $0 < h < 1$, u_i^+ is a strict convex combination of u_i and $I_{\text{total},i}$. Excluding exact threshold ties, a transition $0 \to 1$ implies $I_{\text{total},i,d} > u_{\text{th}}$, whereas a transition $1 \to 0$ implies $I_{\text{total},i,d} < u_{\text{th}}$.

Hence, for every changed dimension,

$$(I_{\text{total},i,d} - u_{\text{th}})\,\Delta s_{i,d} > 0.$$

Using the node-wise convexity condition in Theorem 3, we obtain

$$N\left[\mathcal{E}(S^+) - \mathcal{E}(S)\right] \le -\sum_d (I_{\text{total},i,d} - u_{\text{th}})\,\Delta s_{i,d} \le 0.$$

Therefore, the energy is non-increasing when $0 < \Delta t/\tau < 1$, and decreases strictly whenever at least one spike dimension changes without an exact threshold tie.

C.3 Empirical Validation of the Euler Step Size

We evaluate the sensitivity of the energy evolution to the normalized Euler step size $h = \Delta t/\tau$. For the soft-threshold model, we gradually increase h and record the smallest value at which an energy fluctuation is first observed. The reported results are averaged over ten random seeds. As shown in Table 4, the first observed energy fluctuation occurs at $h = 1.18$–1.90 across the evaluated datasets, whereas the normalized step sizes used in our main experiments lie in the substantially smaller range $h \in [0.15, 0.55]$. This empirical observation indicates that the configurations used in our experiments operate well within the stable regime observed in the step-size sensitivity study. For the asynchronous hard-threshold dynamics, the analysis above gives the explicit sufficient condition $0 < h < 1$. Consistent with this condition, we observe small energy fluctuations once the step size is increased to $h = 1.01$, while the step sizes used in our experiments remain below one.

D Empirical Validation of Energy Descent

To validate our theoretical proof of the global energy monotonic decrease, we track the dynamic changes of different energy components during discrete forward evolution. As illustrated in Figure 7, the total energy $\mathcal{E}$ (black curve) exhibits a strict monotonic decrease across the eight datasets , swiftly converging to a stable lower bound within a limited number of time steps. This empirical observation perfectly aligns with our mathematical stability analysis. Furthermore, decomposing the total energy provides profound insights into the underlying physical dynamics of the spiking graph networ. The membrane recovery potential $\mathcal{E}_{mem}$ (red curve) gradually increases, acting as an endogenous tension that resists unbounded excitation. Concurrently, the negative internal aggregation potential $-\mathcal{E}_{int}$ (green curve) and negative external driving potential $-\mathcal{E}_{ext}$ (blue curve) continuously descend, collectively providing the essential driving forces that guide the system toward an optimal state. The synergistic convergence of these competing components strongly confirms that our proposed temporal evolution paradigm intrinsically operates as a continuous energy descent process over the global energy landscape, thereby securing a robust and stable equilibrium for final readout.

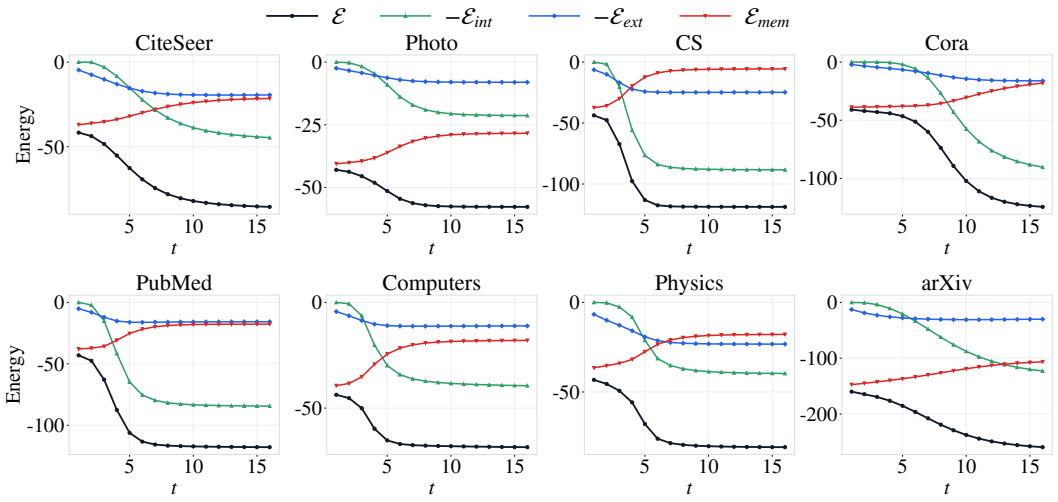


Figure 7: Evolution of different energy components on the eight datasets. The total energy decreases as t increases, eventually converging to stable.

E Computational Cost

A fundamental advantage of SNNs lies in their event-driven nature, which promises significant efficiency on neuromorphic platforms. To validate this inherent advantage, we systematically evaluate the computational cost of our proposed model. The computational cost calculation in SpikeGCL solely considers the costs associated with spike encoding and spike firing. However, beyond these costs, the model involves a massive amount of dense floating-point operations, which account for the majority of the total computational cost. Therefore, following the approach in (Yao et al., 2023), we take all computational operations executed during the forward propagation process into account. The results are presented in Table 5. Compared to both traditional and spiking GNNs, our method achieves the lowest computational cost. This demonstrates that our single-layer architecture effectively reduces the computational cost of the model.

The dimensionality of input features fundamentally dictates the model's parameter size. Interestingly, for SNNs, the raw node features across these datasets are intrinsically binary and can be naturally treated as spike trains. This mechanism reveals why spiking GNNs achieve a lower computational cost than traditional GNNs on parameter-heavy datasets like Physics. Furthermore, a compelling observation in our method emerges when comparing the Photo and CS datasets: although the parameter size of Photo is drastically smaller than that of CS (69.0K vs. 457.3K), its computational cost is paradoxically higher (1.12 mJ vs. 1.06 mJ). This phenomenon is directly attributed to the input feature density. Specifically, the proportion of active features (*i.e.* 1) in the raw input is 34.74% for Photo, compared to extreme sparsity of 0.88% for CS. As a result, the high sparsity of the CS dataset drives its overall computational cost slightly lower.

Table 5: Parameter count and theoretical neuromorphic inference computational cost across multiple datasets. Each dataset reports parameter count (#Param, K) and computational cost (Cost, mJ).

	Computers		CiteSeer		Photo		CS		Physics	
Method	#Param↓	Cost↓	#Param↓	Cost↓	#Param↓	Cost↓	#Param↓	Cost↓	#Param↓	Cost↓
GCN	66.4	4.80	254.1	3.90	64.9	2.57	453.2	2.01	555.6	3.62
GAT	67.0	6.04	254.6	3.93	65.4	3.17	453.7	2.47	556.1	4.90
SpikeGCL	57.6	4.62	253.4	0.57	54.6	2.40	456.1	3.67	561.3	7.75
SpikeNet	113.0	3.91	486.3	0.19	108.9	2.11	889.1	1.80	1088.8	3.77
Ours	70.5	**2.24**	258.2	**0.10**	69.0	**1.12**	457.3	**1.06**	559.7	**2.53**

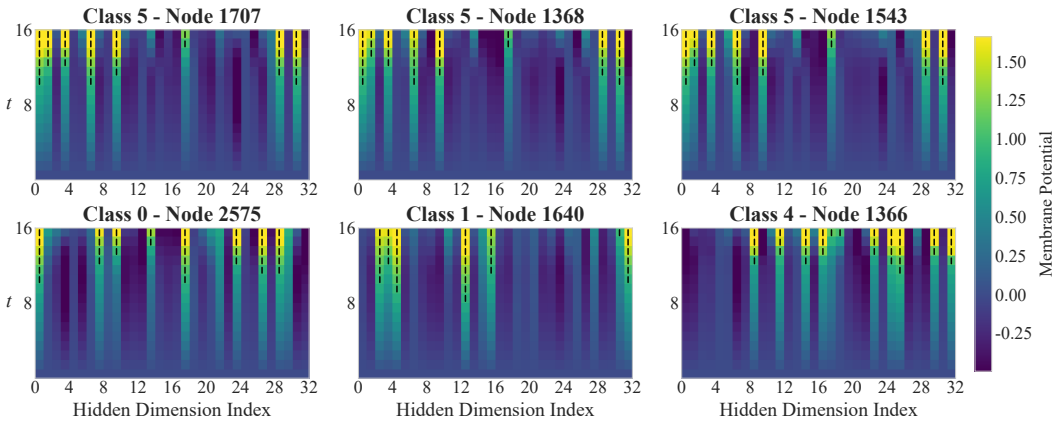


Figure 8: Temporal evolution of membrane potentials and spike firing (black dashed lines) on Cora. The membrane potentials and firing patterns of intra-class nodes gradually converge over time, while the evolutionary trajectories of inter-class nodes continuously diverge, exhibiting strong separability.

E.1 The Spike Firing on Different Nodes

Figure 8 illustrates the spatiotemporal dynamics of membrane potentials and spikes for six nodes on Cora that serve as representative comparative cases. The x-axis is intentionally truncated at 32 for better visualization. Initially, membrane potential variations are minimal. As t progresses, the membrane potentials and spike firing of nodes within the same class become highly synchronized, generating identical activations in specific hidden dimensions. In contrast, nodes from different classes activate almost entirely disjoint feature subspaces. This demonstrates that our method drives intra-class convergence and inter-class divergence, providing robustly discriminative features for classification.

F Visualization of Energy-Driven Evolution on Other Datasets

In this section, we present the energy descent trajectories of individual nodes across additional datasets. To maintain clarity of visualization, we restrict our presentation to datasets with fewer than 10 classes. The result is shown in Figure 9. Similar to the PubMed dataset, during the energy descent process of these four datasets, the node embeddings progressively separate into class-specific clusters.

G Algorithm

In our implementation, the soft-threshold mapping is specifically instantiated as a standard sigmoid function. To enable end-to-end training of the discrete SNNs, we employ the surrogate gradient method (Neftci et al., 2019) to overcome the non-differentiability of the hard-threshold firing function during backpropagation. Regarding the default hyperparameter configurations, the scaling factor in the attention mechanism is set to $\beta = 1/\sqrt{D}$, the membrane time constant is set to $\tau = 2.0$, and the number of memory patterns M in the Hopfield memory matrix P is default to 256. The forward temporal evolution of our method is shown in Algorithm 1. All reported results are averaged over ten independent runs with different random seeds.

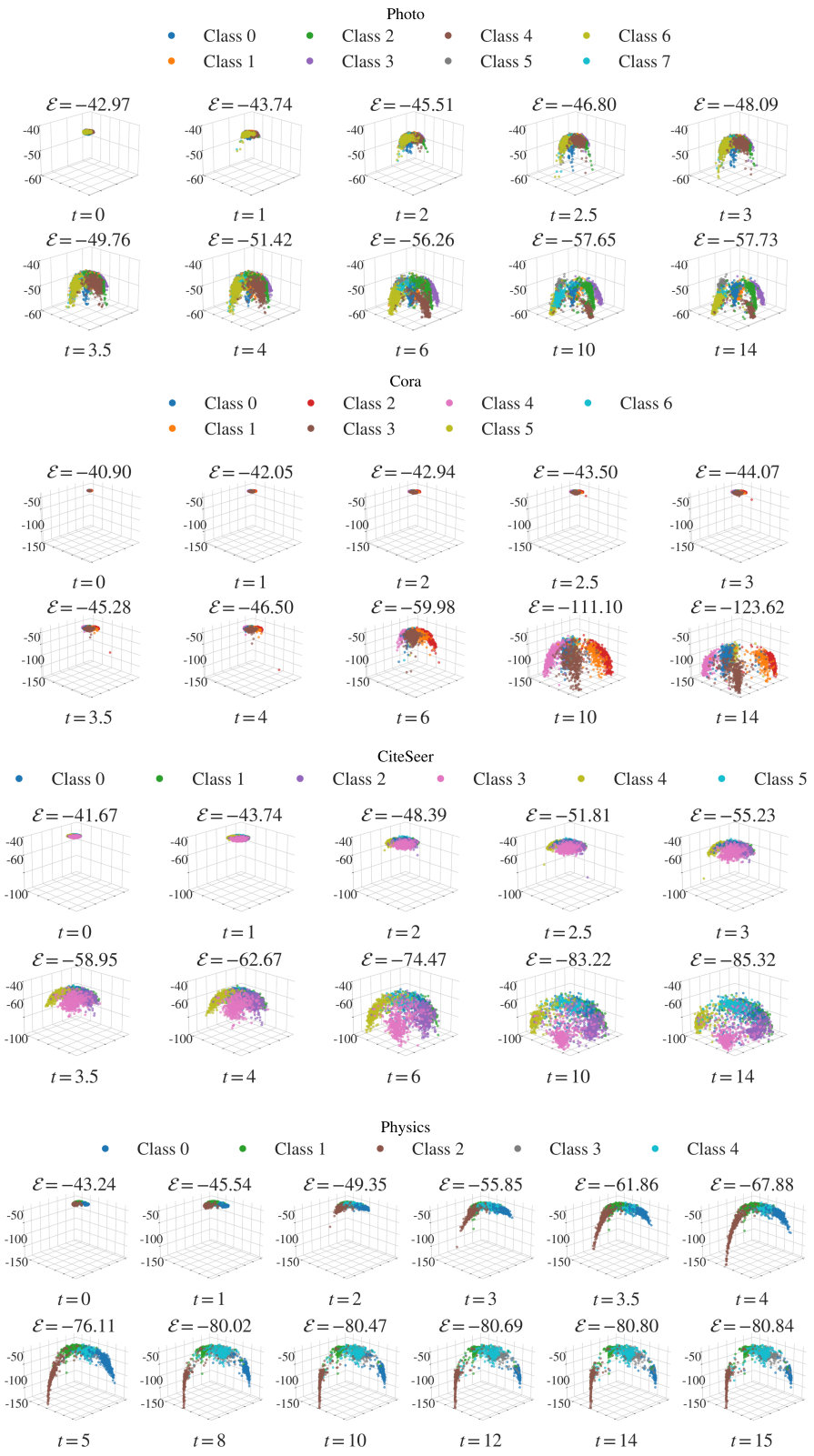


Figure 9: Evolution of the energy on four datasets (Photo, Cora, CiteSeer, and Physics).

Algorithm 1 The forward temporal evolution of our method.

Require: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{R}, X)$, Total time steps T, Integration step dt, Membrane time constant τ, β. Learnable parameters W_{ext}, W, P, Threshold u_{th}.

Ensure: Final node classification predictions $\hat{\mathbf{Y}}$.

Initialization:
Initialize membrane potentials $\mathbf{U} \leftarrow \mathbf{0}$, spike states $\mathbf{S} \leftarrow \mathbf{0}$
Compute constant external driving current: $I_{ext,i} \leftarrow W_{ext} X_i, \quad \forall v_i \in \mathcal{V}$
for $t = 1, 2, \ldots, T$ **do**
  # 1. Compute Graph Attention Current (Eq. 3)
  $I_{fwd,i} \leftarrow W^T \sum_{v_j \in \mathcal{N}_i^{in}} \alpha_{ij} (W s_j)$ {*Forward current*}
  $I_{bwd,i} \leftarrow W^T \sum_{v_k \in \mathcal{N}_i^{out}} \alpha_{ki} (W s_k)$ {*Backward current*}
  $I_{ATT,i} \leftarrow I_{fwd,i} + I_{bwd,i}$
  # 2. Compute Hopfield Current (Eq. 4)
  $I_{HN,i} \leftarrow (P^T P) s_i$
  # 3. Aggregate Internal Current
  $I_{int,i} \leftarrow I_{ATT,i} + I_{HN,i}$
  # 4. Update Membrane Potential via Euler Integration (Eq. 1)
  $u_i \leftarrow u_i + \frac{dt}{\tau} \left(-u_i + I_{ext,i} + I_{int,i} \right), \quad \forall v_i \in \mathcal{V}$
  # 5. Update Spike State
  if using hard-threshold **then**
    $s_i \leftarrow \Theta(u_i - u_{th}), \quad \forall v_i \in \mathcal{V}$
  else
    $s_i \leftarrow \sigma(u_i), \quad \forall v_i \in \mathcal{V}$ {*Soft-threshold mapping*}
  end if
end for
 # 6. Readout Classification
Compute predictions $\hat{\mathbf{Y}} \leftarrow \text{Readout}(\mathbf{S})$ {*Using solely the final spike state*}
Return $\hat{\mathbf{Y}}$