SF-PE: A SYNERGISTIC FUSION OF ABSOLUTE AND RELATIVE POSITIONAL ENCODING FOR SPIKING TRANSFORMERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Positional signals in spiking neural networks (SNNs) suffer distortion due to spike binarization and the nonlinear dynamics of Leaky Integrate-and-Fire (LIF) neurons, which compromises self-attention mechanisms. We introduce **Spiking-RoPE**, a spiking-friendly relative rotary positional encoding that applies two-dimensional spatiotemporal position-dependent rotations to queries/keys prior to binarization, ensuring that relative phase kernels are preserved in statistical expectation under LIF dynamics while maintaining content integrity. Building on this core, we propose **Spiking Fused-PE (SF-PE)**, a scheme that fuses absolute CPG-based spikes with Spiking-RoPE. The resulting attention score decomposes into complementary row/column (absolute) and diagonal (relative) structures, thereby expanding the representable function space. We validate our method across two diverse domains (time-series forecasting and text classification) on Spikformer, Spike-driven Transformer, and QKFormer backbones. SF-PE consistently improves accuracy and enhances length extrapolation capabilities. Ablations on rotation bases and 1D vs. 2D variants support the design. These results establish rotary encoding as an effective, spiking-friendly relative PE for SNNs and demonstrate that fusing absolute and relative signals yields synergistic benefits under spiking constraints. Code: `https://anonymous.4open.science/r/SNN-RoPE-F6DE`.

1 INTRODUCTION

Spiking neural networks (SNNs) transmit information via discrete spikes that emulate biological firing, enabling event-driven computation with low energy and neuromorphic compatibility (Maass, 1997; Davies et al., 2018; Roy et al., 2019). Recent work has transplanted key Transformer components to SNNs, including spike-friendly self-attention (Zhou et al., 2022; Yao et al., 2023; Song et al., 2024; Zhou et al., 2024a). A persistent bottleneck, however, is positional encoding (PE). While self-attention inherently lacks order awareness and therefore requires PE (Vaswani et al., 2017), PE signals in SNNs suffer distortion through spike binarization and the nonlinear dynamics of Leaky Integrate-and-Fire (LIF) neurons.

Conventional continuous PEs (e.g., sinusoidal) differentiate positions through subtle embedding changes. However, thresholding operations distort this information by either nullifying these subtle differences (when inputs remain subthreshold) or drastically amplifying them (upon crossing the threshold). This fundamental incompatibility motivates the development of spiking-friendly PEs that can survive both binarization and LIF dynamics.

Through a systematic analysis of current approaches, we identify three critical gaps in the existing SNN positional encoding landscape. **Gap 1 (Theory):** Existing SNN Transformers predominantly rely on implicit, weight-based position learning and lack rigorous analysis of how positional information is preserved through binarization. **Gap 2 (Single-paradigm limits):** Absolute PE, such as CPG-PE (Lv et al., 2024), exhibits sensitivity to shifts and suffers from aliasing on long sequences, whereas relative PE (e.g., Gray/Log-PE (Lv et al., 2025)) encounters capacity constraints and distance-resolution limitations. **Gap 3 (Spatiotemporal modeling):** SNN data are inherently spatiotemporal in nature, yet most PEs treat position as one-dimensional, thereby neglecting the separable time and sequence axes.

To systematically address these identified gaps, we propose **Spiking-RoPE**, a comprehensive solution that begins with redesigning rotary positional encoding specifically for SNNs. Spiking-RoPE applies position-dependent rotations to queries/keys prior to spike binarization, yielding relative phase kernels that are preserved in statistical expectation under LIF dynamics while maintaining content integrity. We further decouple rotations along sequence (length) and time axes to obtain 2D Spiking-RoPE, which explicitly models spatiotemporal relations. Finally, we integrate absolute and relative signals in **Spiking Fused-PE (SF-PE)** by combining absolute CPG-based spikes at the input with Spiking-RoPE within blocks. This fused scheme activates complementary row/column (absolute) and diagonal (relative) structures in attention maps (See Fig. 3 in the Appendix), thereby expanding representable function space.

To demonstrate the effectiveness and generalizability of our approach, we conduct extensive validation across two diverse domains (i.e., time-series forecasting and text classification) and three established spiking backbones (i.e., Spikformer, Spike-driven Transformer, QKFormer), supplemented by comprehensive ablations on rotation bases and 1D vs. 2D variants. Our experimental results show that SF-PE consistently improves accuracy and strengthens length extrapolation capabilities across all evaluated scenarios.

Contributions.

- **C1, Theoretical foundation (Gap 1):** We rigorously prove that pre-spike rotary phases preserve relative phase kernels in statistical expectation under LIF (See Appendix A), thereby explaining why rotation-based PEs are inherently compatible with spike dynamics.
- **C2, Fused absolute-relative PE (Gap 2):** SF-PE systematically integrates CPG-PE (absolute) with Spiking-RoPE (relative), jointly inducing complementary row/column and diagonal attention structures.
- **C3, Native spatiotemporal PE (Gap 3):** Spiking-RoPE independently rotates along sequence and time axes to capture spatiotemporal relations that are inaccessible to 1D designs.
- **C4, Cross-domain evidence:** Consistent gains across backbones and tasks, plus ablations, establish robustness and generality across different domains (i.e., time series forecasting, text classification).

2 Related Work

2.1 SNN Transformer Architectures

The adaptation of Transformers to the SNN domain has gained significant momentum in recent years. Notable contributions include Spikformer (Zhou et al., 2022; 2024b), which pioneered the integration of LIF neurons into vanilla Transformers to create spiking self-attention mechanisms. Building on this foundation, Spike-driven Transformer (Yao et al., 2023; 2024) advanced the field by proposing more computationally efficient spike-based MatMul operations. Spikingformer (Zhou et al., 2023) proposed a spike-based residual learning framework. QKFormer (Zhou et al., 2024a) improved the binarization process of queries and keys to reduce information loss. However, all of these adopt approaches where weights indirectly learn positions without explicit PE, thereby exemplifying **Gap 1**.

2.2 Absolute PE for SNN

Among absolute PE methods designed specifically for SNNs, CPG-PE (Lv et al., 2024) represents the current state-of-the-art approach. This method leverages central pattern generator properties to assign distinct binary spike patterns to each position through frequency channel thresholding. While demonstrating spike consistency and neuromorphic compatibility, CPG-PE exhibits fundamental limitations that our work addresses: (1) **Translation sensitivity**, absolute coordinate terms in attention render it vulnerable to sequence shifts, and (2) **Long sequence aliasing**, finite period synthesis causes pattern collisions that become increasingly severe in extended sequences, directly contributing to **Gap 2** identified in our analysis.

2.3 Relative PE for SNN

In contrast to the absolute PE method, relative PE approaches for SNNs focus on encoding positional relationships rather than absolute positions. Gray-PE and Log-PE (Lv et al., 2025) represent the most advanced relative PE approaches currently available for SNNs. Gray-PE approximates relative distances using Hamming distance-based discrete codes, while Log-PE employs log-scale distance buckets. However, both methods encounter critical limitations that reinforce **Gap 2**: Gray-PE suffers from (1) representation upper bounds due to bit capacity constraints 2^b, and (2) distance ordering violations where Hamming distance fails to preserve actual distance relationships. Log-PE encounters (1) coarse distance resolution due to distant interaction binning, and (2) spatiotemporal instability during 2D extension, thereby highlighting the necessity for our **Gap 3** solution.

3 Preliminary

3.1 Notation

T denotes the number of time steps, L is the sequence length (number of tokens/patches), and D is the feature dimension. Bold uppercase letters denote tensors, and operations apply to the last dimension unless otherwise specified. $\mathrm{BN}(\cdot)$ denotes batch normalization, and $\mathrm{SN}(\cdot)$ denotes the spike operation induced by LIF in Eq. 1.

3.2 Leaky Integrate-and-Fire (LIF) Neuron

In this study, we use LIF neurons (Maass, 1997) for spike binarization in SNNs. At discrete time t, the membrane potential update $H(t)$, spike $S(t)$, and post-reset potential $U(t)$ for input current $I(t)$ are as follows:

$$\begin{aligned} H(t) &= U(t-1) + \tfrac{1}{\tau}\big(I(t) - \big(U(t-1) - U_{\text{reset}}\big)\big)\,, \\ S(t) &= \Theta\big(H(t) - U_{\text{thr}}\big), \\ U(t) &= \big(1 - S(t)\big)\, H(t) + S(t)\, U_{\text{reset}}, \end{aligned} \tag{1}$$

where τ is the leak time constant, U_{thr} is the threshold, U_{reset} is the reset potential, and $\Theta(\cdot)$ is the Heaviside step function. In this paper, $\mathrm{SN}(z)$ refers to spike output under LIF dynamics (e.g., $\Theta(z - U_{\text{thr}})$).

3.3 Spiking Self-Attention

Spiking Self-Attention (SSA) is a transformation of self-attention adapted for spike representations following Spikformer (Zhou et al., 2022). For spike tensor $X \in \{0,1\}^{T\times L\times D}$:

$$\mathbf{Q}_c = \mathrm{BN}(X)\, W_Q, \quad \mathbf{K}_c = \mathrm{BN}(X)\, W_K, \quad \mathbf{V}_c = \mathrm{BN}(X)\, W_V, \tag{2}$$

where $W_{\{\cdot\}}$ are learnable linear mappings. The corresponding spike embeddings are:

$$\mathbf{Q}_s = \mathrm{SN}(\mathbf{Q}_c), \quad \mathbf{K}_s = \mathrm{SN}(\mathbf{K}_c), \quad \mathbf{V}_s = \mathrm{SN}(\mathbf{V}_c). \tag{3}$$

Time indices are omitted for notational simplicity, and attention is computed at each time step as follows, where $\mathrm{AttnMap}$ is an integer matrix reflecting spike co-occurrence over feature dimensions:

$$\mathrm{AttnMap} = \mathbf{Q}_s\, \mathbf{K}_s^{\top} \;\in\; \mathbb{N}_0^{L\times L}, \qquad \mathbf{SSA} = \mathrm{SN}\big(\mathrm{AttnMap}\cdot \mathbf{V}_s\big). \tag{4}$$

3.4 CPG-PE

CPG-PE (Lv et al., 2024) is an absolute PE that borrows the periodic firing principles of central pattern generators. For K channels with different periods, at position $i \in \{0, \ldots, L-1\}$:

$$u_k(i) = \cos\big(\omega_k i + \phi_k\big), \qquad k = 1, \ldots, K, \tag{5}$$

$$g_k(i) = \Theta\big(u_k(i) - \tau_k\big) \;\in \{0, 1\}, \tag{6}$$

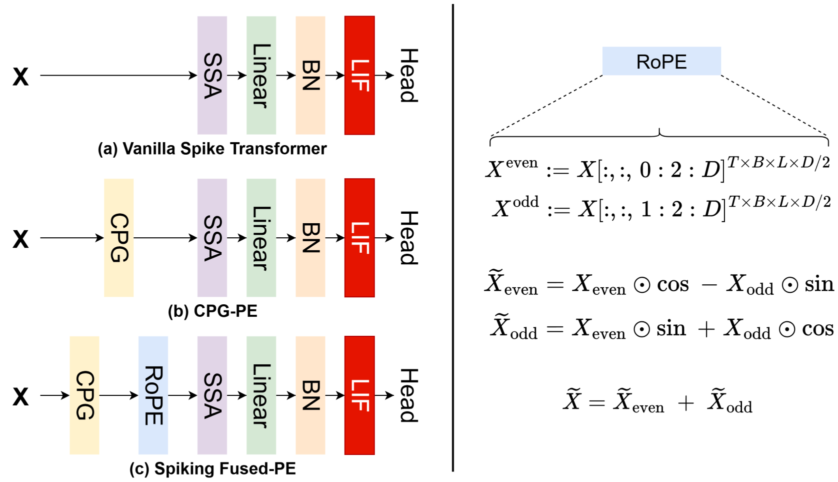


Figure 1: SFPE architecture. The diagram illustrates the integration of CPG-PE and RoPE with Spiking Neural Network, showing the flow from input spike trains through the fused PE to the attention computation in spiking transformers.

binary signals are synthesized to create:

$$\mathbf{p}_i^{\text{CPG}} = \left[g_1(i),\, g_2(i),\, \ldots,\, g_K(i)\right]^\top \in \{0,1\}^K, \tag{7}$$

where ω_k is the angular frequency ($T_k = 2\pi/\omega_k$), ϕ_k is the phase, τ_k is the threshold, and Θ is the Heaviside step function.

3.5 Rotary Positional Embedding (RoPE)

RoPE (Su et al., 2024) encodes positional information by rotating $(2i-1, 2i)$ channel pairs of embeddings at position m with position-dependent angles. For per-head dimension d (even), the frequencies are set as $\theta_i = B^{-2(i-1)/d}$ $(i = 1, \ldots, d/2)$, where base $B > 1$ determines the rotation frequency. The block diagonal rotation matrix at position $m \in \{0, \ldots, L-1\}$ is:

$$R_m = \operatorname{diag}\left(\begin{bmatrix}\cos(m\theta_1) & -\sin(m\theta_1)\\ \sin(m\theta_1) & \cos(m\theta_1)\end{bmatrix}, \ldots, \begin{bmatrix}\cos(m\theta_{d/2}) & -\sin(m\theta_{d/2})\\ \sin(m\theta_{d/2}) & \cos(m\theta_{d/2})\end{bmatrix}\right) \in \mathbb{R}^{d\times d}. \tag{8}$$

For query/key $\mathbf{Q}_c, \mathbf{K}_c \in \mathbb{R}^{L\times d}$, RoPE is applied position-wise as:

$$(\tilde{\mathbf{Q}}_c)_m = R_m\,(\mathbf{Q}_c)_m, \qquad (\tilde{\mathbf{K}}_c)_m = R_m\,(\mathbf{K}_c)_m, \tag{9}$$

while values remain unchanged.

4 Methodology

4.1 Overview

Positional encoding (PE) in SNN-based transformers has evolved through the development process shown in Fig. 1. Early SNN transformers (a) suffered from performance degradation by learning positional information implicitly. CPG-PE (b) was introduced as the first explicit, absolute positional encoding (PE), but it revealed new limitations, such as pattern collisions in long sequences.

Building on this, our proposed **Spiking Fused-PE** (c) is a fusion approach that combines CPG-PE for learning absolute positions with our Spiking-RoPE for injecting relative relationships. This dual-stage design leverages both types of information to expand the representation space and enhance its performance.

4.2 SPIKING-ROPE

RoPE, originally designed for continuous neural networks, decomposes attention inner products into kernels that depend only on relative phase differences Δ through position-dependent rotations. Building on this foundation, we propose Spiking-RoPE by adapting this rotation mechanism to the pre-spike binarization stage, where relative phase kernels are maintained from a statistical expectation perspective even under LIF leakage and threshold conditions. (See Fig. 4 in the Appendix for Spiking-RoPE transformation steps; See Appendix C for the detailed implementation).

Theoretical Foundation (Gap 1 Resolution): The critical insight enabling Spiking-RoPE is our theoretical proof that phase rotation preserves relative positional information statistically even after spike binarization. This addresses Gap 1 by providing the first rigorous analysis of positional information preservation in SNNs. The complete theoretical analysis, including expectation preservation proofs and LIF dynamics interaction, is presented in Appendix A.

4.2.1 1D SPIKING-ROPE

1D Spiking-RoPE encodes relative positional information along a single axis. For this purpose, rotation matrices $R_{\varphi(i)}$ are applied to $(2r-1, 2r)$ channel pairs in even dimension d.

$$\tilde{q}_i = R_{\varphi(i)}\, q_i^{(c)}, \quad \tilde{k}_j = R_{\varphi(j)}\, k_j^{(c)}, \quad \Delta_{ij} = \varphi(i) - \varphi(j). \tag{10}$$

As a result, the relative relationship between two positions i, j depends only on the phase difference Δ_{ij}.

4.2.2 2D SPIKING-ROPE

To explicitly model the spatiotemporal characteristics of SNN data, 2D Spiking-RoPE encodes positional information by separating it into sequence length axis l and time axis t. First, the embedding is divided into two equal-dimensional blocks, and independent 1D Spiking-RoPE is applied to each block. One block rotates based on sequence position (i), while the other rotates based on time step (t_i).

$$\tilde{q}_i = \left[\, R_{\varphi_l(i)}\, q_i^{(c,l)} \;;\; R_{\varphi_t(t_i)}\, q_i^{(c,t)} \,\right], \qquad \tilde{k}_j = \left[\, R_{\varphi_l(j)}\, k_j^{(c,l)} \;;\; R_{\varphi_t(t_j)}\, k_j^{(c,t)} \,\right]. \tag{11}$$

When computing query/key inner products, letting $\Delta_l = \varphi_l(i) - \varphi_l(j)$ and $\Delta_t = \varphi_t(t_i) - \varphi_t(t_j)$, the inner product operation naturally separates into the sum of relative phase kernel Δ_l along the length axis and relative phase kernel Δ_t along the time axis as shown below. This allows the model to consider both spatiotemporal relative distances.

$$\begin{aligned} \langle \tilde{q}_i, \tilde{k}_j \rangle = & \left(\langle q_i^{(c,l)}, k_j^{(c,l)} \rangle \cos \Delta_l + \langle q_i^{(c,l)}, J\, k_j^{(c,l)} \rangle \sin \Delta_l \right) \\ & + \left(\langle q_i^{(c,t)}, k_j^{(c,t)} \rangle \cos \Delta_t + \langle q_i^{(c,t)}, J\, k_j^{(c,t)} \rangle \sin \Delta_t \right). \end{aligned} \tag{12}$$

4.2.3 PHASE PRESERVATION UNDER LIF

LIF performs nonlinear transformations (Eq. 1), and there is a risk of losing positional information encoded as continuous values during this process. We show that PE applied at the pre-spike stage through Spiking-RoPE is preserved from a statistical expectation perspective even after LIF's spiking transformation.

This proof involves approximating the firing probability function of LIF neurons as a linear function. Batch Normalization in SSA (Eq. 2) stabilizes the distribution of pre-spike inputs to mean 0 and variance 1. Assuming that input currents are distributed in a narrow region around the mean (0) such that the firing probability function operates almost linearly, we can show that the probability of query and key firing simultaneously in a specific dimension is approximately proportional to the product of pre-spike values $\tilde{q}_{id}, \tilde{k}_{jd}$. Under this linear approximation, the expectation of inner products over all dimensions is derived as follows.

$$\mathbb{E}[q_i^\top k_j] \approx \alpha \langle \tilde{q}_i, \tilde{k}_j \rangle, \tag{13}$$

where $\alpha > 0$ is a scaling constant depending on neuron sensitivity and input distribution, and attention scores are determined by the inner product value $\langle \tilde{q}_i, \tilde{k}_j \rangle$. When Spiking-RoPE is applied, $\langle \tilde{q}_i, \tilde{k}_j \rangle$ is expanded as a function of relative phase $\Delta_{ij} = i - j$, resulting in the final attention score having the following relationship:

$$\mathbf{SA}(q_i, k_j) \propto \mathbb{E}[q_i^\top k_j] \propto \langle q_i^{(c)}, k_j^{(c)} \rangle \cos \Delta_{ij} + \langle q_i^{(c)}, J k_j^{(c)} \rangle \sin \Delta_{ij}. \tag{14}$$

In conclusion, Spiking-RoPE preserves phase kernels containing relative positional information under nonlinear LIF dynamics, enabling the utilization of positional information in SNNs.

4.3 Fused PE

In Transformer-based models, positional encoding (PE) injects order and dependencies between tokens. Existing research has independently used either absolute or relative PE. We propose fused PE, which combines both approaches to enhance positional representation power. To simultaneously reflect absolute PE p^{abs} and relative PE $R_{\varphi(\cdot)}$ in input x, query/key are defined as follows:

$$q_i \;=\; R_{\varphi(i)}\, W_Q\big(x_i + p_i^{\text{abs}}\big), \qquad k_j \;=\; R_{\varphi(j)}\, W_K\big(x_j + p_j^{\text{abs}}\big), \tag{15}$$

where p_i^{abs} is the absolute PE, and $R_{\varphi(i)} \in \mathbb{R}^{d \times d}$ is the RoPE-style block rotation at position i. This configuration organically fuses two information sources within a single vector by projecting content+absolute information through linear mapping, then injecting relative information through phase rotation. The row/column structure created by absolute PE and the diagonal structure created by relative PE are simultaneously activated, providing richer representation power compared to single PE (See Fig. 3 in the Appendix). Subsequently, in continuous (pre-spike) space, letting $q_i^{(c)} = W_Q(x_i + p_i^{\text{abs}})$ and $k_j^{(c)} = W_K(x_j + p_j^{\text{abs}})$,

$$\tilde{q}_i = R_{\varphi(i)} q_i^{(c)}, \quad \tilde{k}_j = R_{\varphi(j)} k_j^{(c)}, \quad \Delta_{ij} = \varphi(i) - \varphi(j).$$

With the 90° block rotation operator J for even channel pairs, the inner product is as follows.

$$\langle \tilde{q}_i, \tilde{k}_j \rangle \;=\; \langle q_i^{(c)}, k_j^{(c)} \rangle \cos \Delta_{ij} \;+\; \langle q_i^{(c)},\, J\, k_j^{(c)} \rangle \sin \Delta_{ij}. \tag{16}$$

According to Appendix A, this inner product approximately preserves the relative phase kernel form of the above equation in the expectation $\mathbb{E}[q_i^\top k_j]$ even after spike binarization.

4.4 Final Incorporation

Following fused PE, we propose Spiking Fused-PE (SF-PE), a fused method that combines CPG-PE for absolute PE and Spiking-RoPE for relative PE. First, after injecting CPG-PE from Eq. 7 into the embedding,

$$q_i^{(c)} = W_Q\big(x_i + E\, \mathbf{p}_i^{\text{CPG}}\big), \qquad k_j^{(c)} = W_K\big(x_j + E\, \mathbf{p}_j^{\text{CPG}}\big), \qquad E \in \mathbb{R}^{d \times K}, \tag{17}$$

which is rotated with 2D Spiking-RoPE. Then, the continuous inner product is decomposed with respect to the relative phases of the two axes as follows:

$$\langle \tilde{q}_i, \tilde{k}_j \rangle = \sum_{r=1}^{d_l/2} \Big(A_{ij,r}^{(l)} \cos \Delta_r^{(l)} + B_{ij,r}^{(l)} \sin \Delta_r^{(l)} \Big) + \sum_{r=1}^{d_t/2} \Big(A_{ij,r}^{(t)} \cos \Delta_r^{(t)} + B_{ij,r}^{(t)} \sin \Delta_r^{(t)} \Big), \tag{18}$$

where d_l and d_t denote per-axis even dimensions, and the amplitude terms are

$$A_{ij}^{(l)} = \langle q_i^{(c,l)}, k_j^{(c,l)} \rangle, \quad B_{ij}^{(l)} = \langle q_i^{(c,l)}, J\, k_j^{(c,l)} \rangle,$$

and similarly $A_{ij}^{(t)}, B_{ij}^{(t)}$ are defined for the time axis. Consequently, the attention score of SF-PE is structured as a sum of contributions from the length l and time t axes, each of which is itself a sum over individual rotational frequency channel pairs. This decomposition shows how two types of positional information are complementarily combined for each channel pair: absolute information (amplitudes A, B) and relative information (trigonometric kernels, Δ_l, Δ_t), allowing the model to capture richer and more granular positional details. This structure is preserved from a statistical expectation perspective even after the spike binarization process, enabling SNNs to effectively learn complex spatiotemporal patterns.

Table 1: Performance comparison on time series forecasting on 4 benchmarks with various prediction lengths 6, 24, 48, 96. The best results are shown in **bold**. PE types: A = absolute, F = Fused (absolute + relative). Metrics: higher R^2 and lower RSE indicate better performance. All results are averaged across 3 random seeds.

Models	PE Type	Metric	Metr-la (L = 12)				Pems-bay (L = 12)				Solar (L = 168)				Electricity (L = 168)				Avg.
			6	24	48	96	6	24	48	96	6	24	48	96	6	24	48	96	
Transformer w/Sin-PE (Upper bound)	A	$\mathbf{R}^2\uparrow$	.727	.554	.413	.284	.785	.734	.688	.673	.953	.858	.759	.718	.978	.975	.972	.964	.733
		RSE$\downarrow$	.551	.704	.808	.895	.502	.558	.610	.618	.223	.377	.504	.545	.260	.277	.347	.425	.512
Spikformer w/Conv-PE Zhou et al. (2022)	A	$\mathbf{R}^2\uparrow$	.713	.527	.399	.267	.773	.697	.686	.667	.929	.828	.744	.674	.959	.955	.955	.954	.733
		RSE$\downarrow$	.565	.725	.818	.903	.514	.594	.606	.621	.272	.426	.519	.586	.373	.371	.379	.382	.541
Spikformer w/CPG-PE Lv et al. (2024)	A	$\mathbf{R}^2\uparrow$	.726	.526	.418	.287	.780	.712	.690	.666	.937	.833	.757	.707	.972	.970	.966	.960	.744
		RSE$\downarrow$	.553	.720	.806	.890	.508	.580	.602	.622	.257	.420	.506	.555	.299	.310	.314	.355	.519
Spikformer w/SF-PE (Ours)	F	$\mathbf{R}^2\uparrow$	**.739**	**.561**	**.432**	**.317**	**.783**	**.713**	**.698**	**.670**	**.939**	**.877**	**.782**	**.752**	**.981**	**.975**	**.972**	**.965**	**.760**
		RSE$\downarrow$	**.538**	**.698**	**.795**	**.871**	**.499**	**.576**	**.593**	**.618**	**.251**	**.362**	**.479**	**.511**	**.240**	**.280**	**.300**	**.336**	**.497**
SDT-V1 w/Conv-PE Yao et al. (2023)	A	$\mathbf{R}^2\uparrow$	.588	.364	.236	.121	.674	.668	.658	.639	.922	.837	.732	.685	.958	.951	.946	.939	.682
		RSE$\downarrow$	.692	.841	.935	.984	.599	.605	.616	.637	.281	.405	.533	.584	.367	.389	.412	.430	.582
SDT-V1 w/CPG-PE Lv et al. (2024)	A	$\mathbf{R}^2\uparrow$	.601	.387	.257	.152	.693	.695	.680	.664	.935	.860	.748	.710	.966	.955	.959	.945	.700
		RSE$\downarrow$	.667	.827	.910	.972	.580	.578	.592	.607	.260	.383	.515	.553	.329	.378	.362	.417	.558
SDT-V1 w/SF-PE (Ours)	F	$\mathbf{R}^2\uparrow$	**.703**	**.470**	**.296**	**.187**	**.741**	**.700**	**.686**	**.679**	**.945**	**.871**	**.794**	**.766**	**.979**	**.971**	**.971**	**.969**	**.733**
		RSE$\downarrow$	**.576**	**.769**	**.886**	**.952**	**.533**	**.573**	**.587**	**.593**	**.242**	**.369**	**.466**	**.496**	**.259**	**.299**	**.302**	**.310**	**.513**
QKFormer w/Conv-PE Zhou et al. (2024a)	A	$\mathbf{R}^2\uparrow$	.706	.509	.411	.275	.735	.671	.667	.663	.927	.841	.737	.689	.966	.961	.958	.955	.729
		RSE$\downarrow$	.577	.743	.816	.901	.557	.621	.625	.629	.275	.402	.527	.569	.302	.324	.340	.358	.535
QKFormer w/CPG-PE Lv et al. (2024)	A	$\mathbf{R}^2\uparrow$	.711	.522	**.423**	.286	.743	.684	.681	**.668**	.930	.856	.755	.732	.977	.968	.966	**.959**	.734
		RSE$\downarrow$	.567	.729	**.801**	.890	.548	.608	.611	**.623**	.271	.389	.508	.531	.264	.289	.307	**.361**	.533
QKFormer w/SF-PE (Ours)	F	$\mathbf{R}^2\uparrow$	**.717**	**.520**	.419	**.292**	**.749**	**.702**	**.698**	**.668**	**.934**	**.868**	**.793**	**.737**	**.981**	**.972**	**.968**	.954	**.748**
		RSE$\downarrow$	**.561**	**.730**	.804	**.887**	**.542**	**.590**	**.594**	**.623**	**.264**	**.372**	**.468**	**.526**	**.244**	**.299**	**.318**	.383	**.513**

5 EXPERIMENTS

We evaluate on two diverse domains, time series forecasting and text classification, to test the modality-agnostic nature of SF-PE. The choice follows directly from the method's characteristics: (1) pre-spike rotary phases preserve relative kernels under LIF (C1; Sec. 4.2, Sec. 4.2.3); (2) the fused absolute–relative scheme induces complementary row/column vs. diagonal attention structure that any ordered data exhibits (C2; Eq. 16); and (3) the 2D variant decouples length and time to model spatiotemporal relations while remaining compatible with 1D sequences (C3; Sec. 4.2.2, Eq. 12). We therefore assess robustness across (a) modalities (continuous signals vs. discrete tokens) and (b) SNN backbones (Spikformer, SDT-V1, QKFormer), and we include length extrapolation to specifically probe relative-position generalization.

Our experimental validation systematically demonstrates how our gap-targeted solutions (C1-C3) translate to performance improvements across diverse domains. For our primary comparisons, we evaluate against two baselines: Conv-PE (Zhou et al., 2022; Yao et al., 2023; Zhou et al., 2024a), where positional information is learned implicitly, and CPG-PE (Lv et al., 2024), the state-of-the-art absolute PE for SNNs. However, there has been no research effort for applying relative PE to SNNs, making a direct comparison with a pre-existing method challenging (See the alternative comparison in Appendix E). Additionally, we utilize our Spiking-RoPE as a relative-only baseline and provide a detailed comparison in the ablation studies. Detailed experimental settings, including datasets, metrics, and hyperparameters, are provided in Appendix B and length extrapolation analysis in Appendix D.

5.1 TIME SERIES FORECASTING

Table 1 shows the performance of SF-PE on four time series forecasting datasets. The results reveal several notable patterns:

Consistent superiority of SF-PE: SF-PE consistently outperforms absolute PE approaches across all backbone models (Spikformer, SDT-V1, and QKFormer). In particular, the average R^2 score improved from 0.744 to 0.760 on Spikformer and showed a substantial improvement from 0.700 to 0.733 on SDT-V1. Similarly, SF-PE achieved a leading average R^2 score of 0.748 on QKFormer.

Robustness in long-term prediction: While the performance degradation occurs as prediction length increases (6 to 96 hours), SF-PE maintains relatively stable performance compared to other methods. Specifically, in the 96-hour prediction on the Metr-la dataset, SF-PE achieves $R^2 = 0.317$, showing a 10.5% improvement over CPG-PE's 0.287. This suggests that our SF-PE is more effective at capturing long-term dependencies.

Dataset-specific characteristic analysis:

Table 2: Performance comparison on six text classification tasks using the Spikformer backbone. The best results are shown in **bold**. PE types: A = absolute, F = Fused (absolute + relative). Metrics: F1 score for MRPC, and accuracy for all other tasks. All results are averaged across 3 random seeds.

Model	PE Type	Param(M)	Sentiment Analysis				Inference	Similarity	Avg.
			MR	SST-2	Subj	SST-5	RTE	MRPC	
Fine-tuned BERT (Upper bound)	A	109.8	86.39	92.01	95.43	49.87	69.42	89.75	80.48
Conv-PE Zhou et al. (2022)	A	109.8	71.84 ± 0.14	80.17 ± 0.08	88.35 ± 0.12	38.69 ± 0.10	**52.71** ± 0.00	68.38 ± 0.03	66.69
CPG-PE Lv et al. (2024)	A	110.4	72.73 ± 0.09	81.77 ± 0.04	88.97 ± 0.09	39.15 ± 0.12	**52.71** ± 0.00	70.10 ± 0.04	67.57
SF-PE (Ours)	F	110.4	**73.57** ± 0.10	**81.83** ± 0.05	**89.70** ± 0.08	**40.05** ± 0.09	**52.71** ± 0.00	**70.59** ± 0.03	**68.07**

- **Solar dataset:** The strong periodic patterns in this dataset appear well-suited for spatiotemporal PE, as all models achieved their highest performance on this task.
- **Electricity dataset:** Despite high dimensionality (321 customers), SF-PE shows particularly strong results, indicating that our fused PE approach can effectively capture complex multivariate relationships.
- **Traffic data (Metr-la, Pems-bay):** SF-PE consistently maintains its performance advantage on the more volatile and inherently challenging traffic datasets, even with lower absolute scores.

Comparison with upper bound: Vanilla transformer used in SNNs is considered the performance upper bound, as the binarization process in spiking models can cause information loss compared to the continuous values used in standard transformers. Notably, in some instances, the addition of SF-PE enables spiking models to outperform the upper bound.

5.2 Text Classification

Table 2 presents the performance on six text classification tasks.

Improvement in sentiment analysis: Our SF-PE consistently surpasses both the CPG-PE baseline and the model without PE across all sentiment analysis tasks. Specifically, it achieves 73.57% accuracy on the MR task, an improvement of 0.84% over CPG-PE 72.73%. It also attains the highest performance of 40.05% in the fine-grained sentiment classification of SST-5.

Improvement in other tasks: On the MRPC, SF-PE improves the F1 score by 0.49% to 70.59%, compared to CPG-PE. However, for the RTE, neither CPG-PE nor SF-PE provides a performance benefit over the Spikformer baseline.

Comparison with upper bound: While the BERT model provides upper bounds, SF-PE demonstrates substantial performance. Particularly, on the Subj (subjectivity classification) task, our method achieves 89.70%, narrowing the performance gap to BERT's 95.43%.

Performance vs. model complexity: SF-PE shows consistent performance improvements while maintaining nearly identical parameter count (110.4M) as existing CPG-PE. This means SF-PE improves representation power without additional overhead.

5.3 Ablation Study

5.3.1 1D vs. 2D in Spiking-RoPE

We proposed Spiking-RoPE, which independently encodes positional information along the temporal t and spatial l axes. Tab. 3 validates this design by comparing the performance of 1D and 2D in Spiking-RoPE on the Electricity dataset.

The results show a clear progression in performance. While both 1D Spiking-RoPE variants show strong performance, 2D Spiking-RoPE further improves the average R^2 score to 0.971. The complete SF-PE model ultimately attains

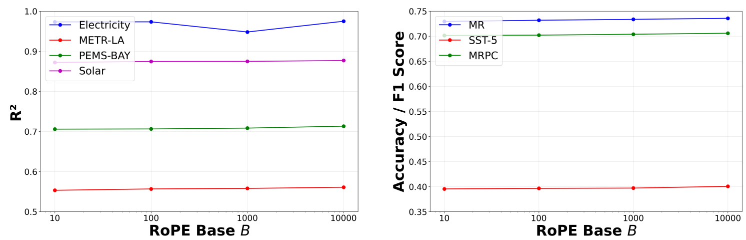
Table 3: 1D vs. 2D SNN RoPE performance on Electricity dataset using Spikformer backbone. The best results are shown in **bold** and the second highest results are <u>underlined</u>. Metrics: a higher R^2 indicates better performance.

Models	PE Type	Electricity				Avg.
		6	24	48	96	
Conv-PE	A	.959	.955	.955	.954	.956
CPG-PE	A	.972	.970	.966	.960	.967
1D-Spatial RoPE	R	.975	.972	.966	.960	.968
1D-Temporal RoPE	R	.976	.973	.968	.962	.970
2D-RoPE	R	<u>.978</u>	<u>.973</u>	<u>.969</u>	<u>.963</u>	<u>.971</u>
Spiking Fused-PE	F	**.981**	**.975**	**.972**	**.965**	**.973**

the highest score of 0.973. This demonstrates a clear synergy between separating spatiotemporal features and fusing absolute with relative positional information.

5.3.2 Rotation Frequency Variation

The base value B is a core hyperparameter of Spiking-RoPE, which determines the rotation frequency. Appropriate frequency selection enables the model to effectively distinguish relative distances of various lengths. Too high frequencies may focus excessively on short-range relationships and miss long-range dependencies, while too low frequencies make fine positional distinctions difficult.



(a) Time series forecasting (b) Text classification

Figure 2: Performance versus RoPE base parameter B.

Our experiments, illustrated in Figure 2, demonstrate that the model is robust to the parameter B, with performance consistently peaking at $B = 10000$ across all tested datasets. We therefore adopt this value as the default for B in all our experiments.

5.4 Results Discussion and Analysis

The experimental results demonstrate the effective design of SF-PE from multiple angles:

Empirical validation of theoretical predictions: The phase preservation theory presented in Section 4.2.3 has been confirmed in actual experiments. Despite nonlinear transformations of LIF neurons, the consistently improved performance of models with Spiking-RoPE suggests that phase kernel preservation under linear approximation is indeed effective.

Synergistic effect of Fused PE: The combination of absolute PE (CPG-PE) and relative PE (Spiking-RoPE) creates synergy beyond simple performance summation. As shown in Eq. 16, this is because absolute information (amplitudes A, B) and relative information (trigonometric kernels $\cos\Delta, \sin\Delta$) work complementarily to expand the representation space.

Task-specific adaptability:

- **Time series forecasting:** The spatiotemporal separation approach of Spiking-RoPE is particularly effective for tasks where periodic patterns and long-term dependencies are important.
- **Text classification:** Relative positional information contributes to performance improvement even in natural language tasks where contextual understanding is crucial.
- **Length extrapolation:** Shows stable performance even on inputs longer than training sequences, confirming generalization ability (Appendix D).

6 Conclusion

We presented **Spiking Fused-PE (SF-PE)**, a spiking-friendly positional encoding that fuses absolute CPG codes with pre-spike rotary phases. Built on Spiking-RoPE and its 2D extension, our design preserves relative phase kernels under LIF dynamics while injecting complementary absolute information. Across time-series and text tasks on Spikformer, SDT-V1, and QKFormer backbones, SF-PE delivers consistent accuracy gains and stronger length extrapolation without an increase in model parameters. These results validate that absolute and relative encodings are synergistic in spiking transformers and provide a principled approach for spatiotemporal PE under spiking constraints.

REFERENCES

Alexis Conneau and Douwe Kiela. Senteval: An evaluation toolkit for universal sentence representations. *arXiv preprint arXiv:1803.05449*, 2018.

Mike Davies, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Nabil Imam, Shweta Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee Micro*, 38(1):82–99, 2018.

Bill Dolan and Chris Brockett. Automatically constructing a corpus of sentential paraphrases. In *Third international workshop on paraphrasing (IWP2005)*, 2005.

R Bar Haim, Ido Dagan, Bill Dolan, Lisa Ferro, Danilo Giampiccolo, Bernardo Magnini, and Idan Szpektor. The second pascal recognising textual entailment challenge. In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, volume 7, pp. 785–794, 2006.

Hosagrahar V Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M Patel, Raghu Ramakrishnan, and Cyrus Shahabi. Big data and its technical challenges. *Communications of the ACM*, 57(7):86–94, 2014.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pp. 95–104, 2018.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018.

Changze Lv, Dongqi Han, Yansen Wang, Xiaoqing Zheng, Xuanjing Huang, and Dongsheng Li. Advancing spiking neural networks for sequential modeling with central pattern generators. *Advances in Neural Information Processing Systems*, 37:26915–26940, 2024.

Changze Lv, Yansen Wang, Dongqi Han, Yifei Shen, Xiaoqing Zheng, Xuanjing Huang, and Dongsheng Li. Toward relative positional encoding in spiking transformers. *arXiv preprint arXiv:2501.16745*, 2025.

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

Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Kevin Knight, Hwee Tou Ng, and Kemal Oflazer (eds.), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pp. 115–124, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics. doi: 10.3115/1219840.1219855. URL `https://aclanthology.org/P05-1015/`.

Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784):607–617, 2019.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1631–1642, 2013.

Xiaotian Song, Andy Song, Rong Xiao, and Yanan Sun. One-step spiking transformer with a linear complexity. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 3142–3150, 2024.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

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.

Man Yao, JiaKui Hu, Tianxiang Hu, Yifan Xu, Zhaokun Zhou, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer v2: Meta spiking neural network architecture inspiring the design of next-generation neuromorphic chips. *arXiv preprint arXiv:2404.03663*, 2024.

Chenlin Zhou, Liutao Yu, Zhaokun Zhou, Zhengyu Ma, Han Zhang, Huihui Zhou, and Yonghong Tian. Spikingformer: Spike-driven residual learning for transformer-based spiking neural network. *arXiv preprint arXiv:2304.11954*, 2023.

Chenlin Zhou, Han Zhang, Zhaokun Zhou, Liutao Yu, Liwei Huang, Xiaopeng Fan, Li Yuan, Zhengyu Ma, Huihui Zhou, and Yonghong Tian. Qkformer: Hierarchical spiking transformer using qk attention. *Advances in Neural Information Processing Systems*, 37:13074–13098, 2024a.

Zhaokun Zhou, Yuesheng Zhu, Chao He, Yaowei Wang, Shuicheng Yan, Yonghong Tian, and Li Yuan. Spikformer: When spiking neural network meets transformer. *arXiv preprint arXiv:2209.15425*, 2022.

Zhaokun Zhou, Kaiwei Che, Wei Fang, Keyu Tian, Yuesheng Zhu, Shuicheng Yan, Yonghong Tian, and Li Yuan. Spikformer v2: Join the high accuracy club on imagenet with an snn ticket. *arXiv preprint arXiv:2401.02020*, 2024b.