Toward Relative Positional Encoding in Spiking Transformers

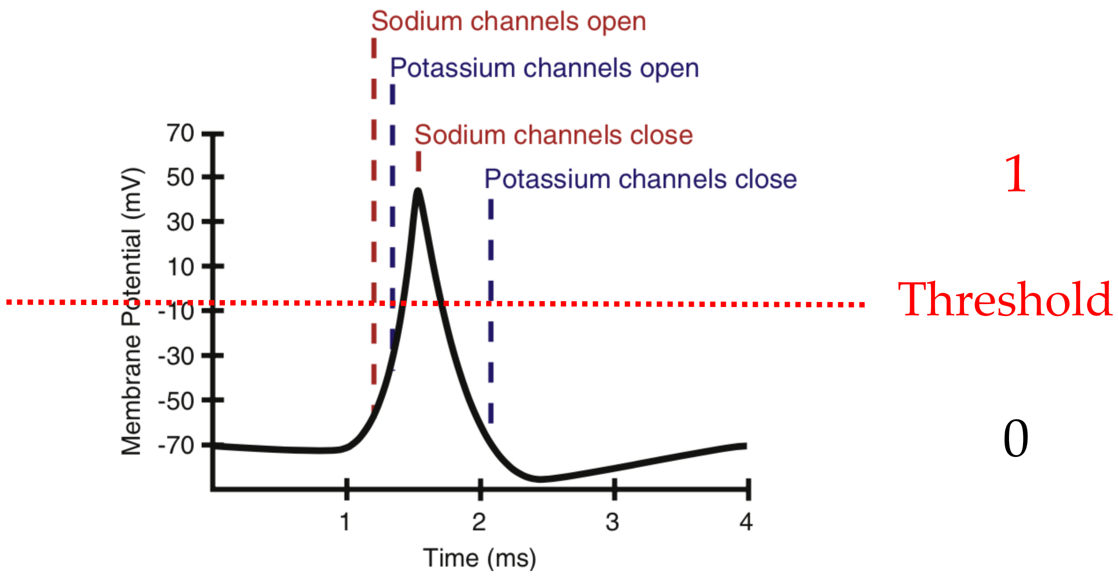
Changze Lv[1], Yansen Wang[2], Dongqi Han[2], Yifei Shen[2], Xiaoqing Zheng[1], Xuanjing Huang[1], Dongsheng Li[2]

[1]College of Computer Science and Artificial Intelligence, Fudan University, [2]Microsoft Research Asia

NeurIPS 2025 Spotlight

Contents

Spiking Neural Networks

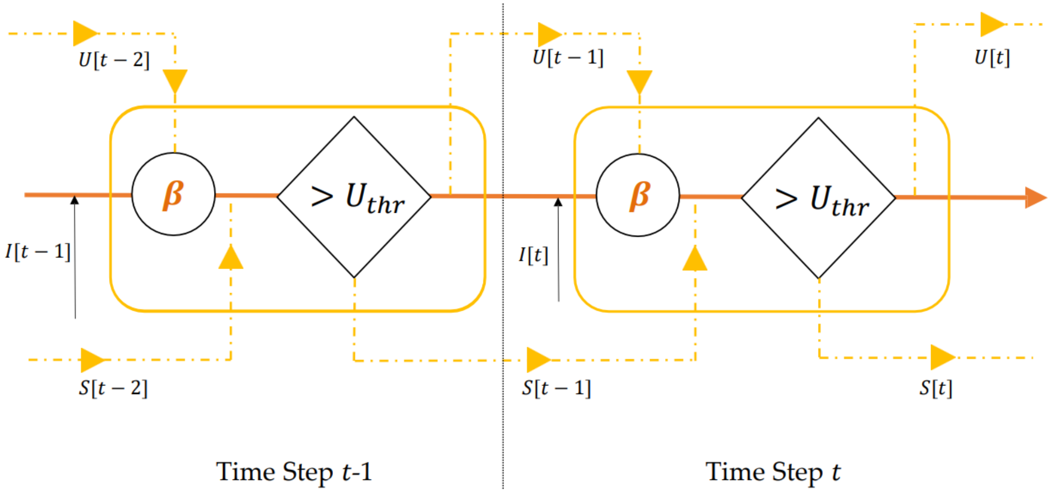


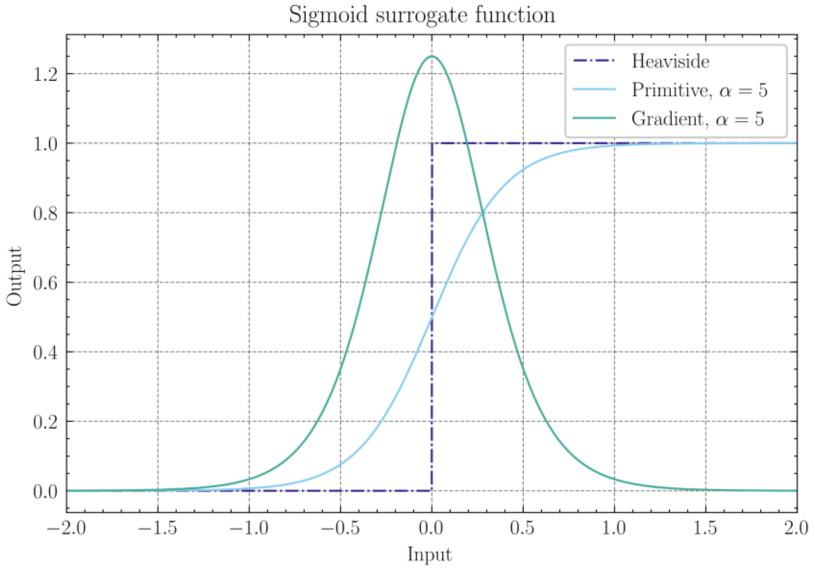
$$U(t) = H(t - \Delta t) + I(t), \quad I(t) = f(\mathbf{x}; \theta),$$

$$H(t) = V_{reset} S(t) + (1 - S(t))\, \beta U(t),$$

$$S(t) = \begin{cases} 1, & \text{if } U(t) \geq U_{\text{thr}} \\ 0, & \text{if } U(t) < U_{\text{thr}} \end{cases},$$

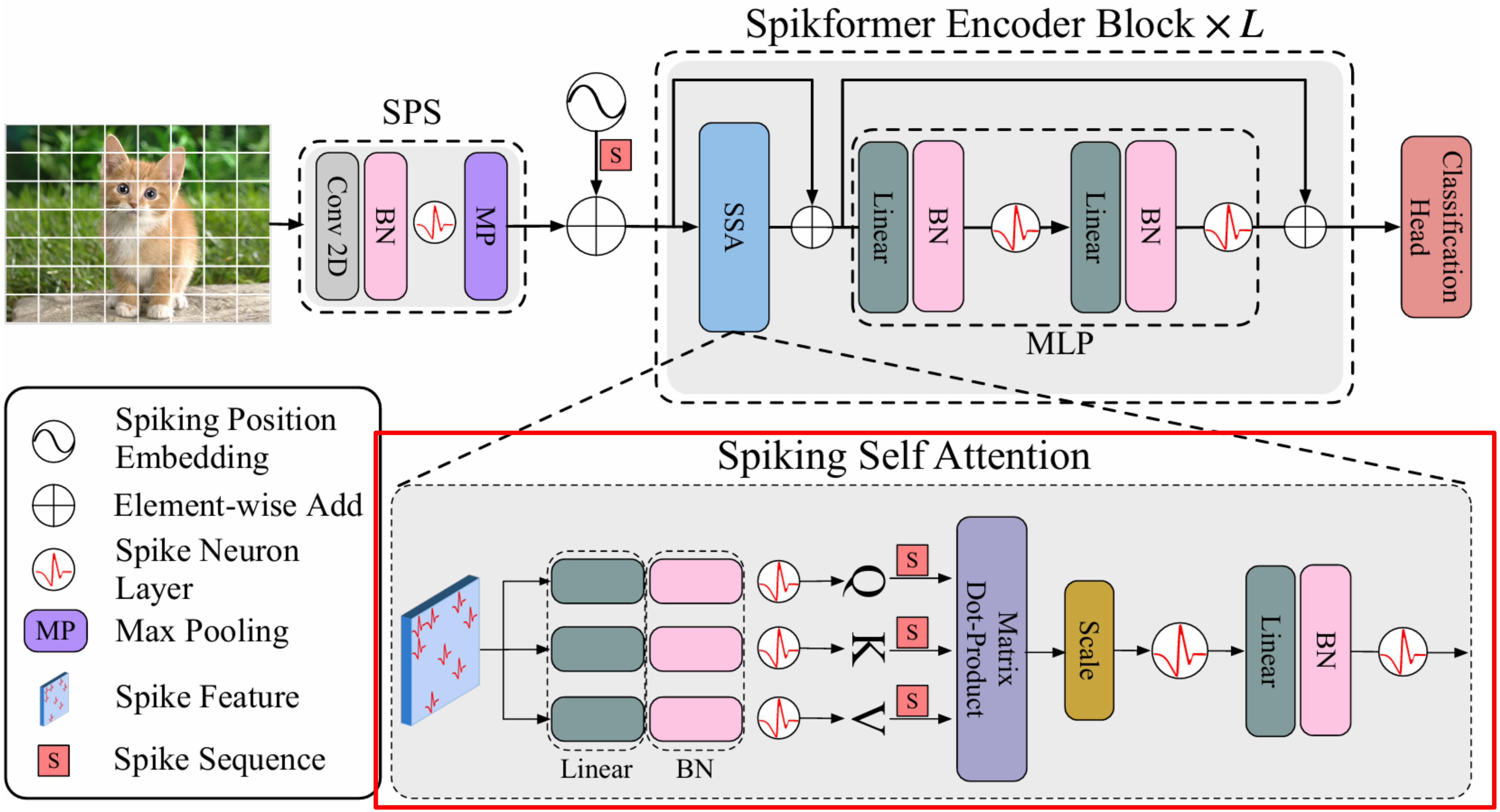
$$S(t) \approx \frac{1}{\pi} \arctan\left(\frac{\pi}{2} \alpha U(t)\right) + \frac{1}{2}$$





https://spikingjelly.readthedocs.io/

Spiking Transformers



Contents

- **Background**
  - Spiking Transformers
  - Relative Positional Encoding
- **Methods**
  - XNOR-based Spiking Self-Attention
  - Gray-PE and Log-PE
- **Experiments**
  - Main Results
  - Analysis

Relative Positional Encoding

Two Forms of RPE

1. ALiBi-like:

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \underbrace{\text{Softmax}\left(\frac{\mathbf{Q}\cdot\mathbf{K}^{\mathrm{T}}}{\sqrt{d_k}} + \mathbf{R}_{i,j}\right)}_{\textbf{AttnMap}} \cdot \mathbf{V}$$

2. RoPE-like:

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \underbrace{\text{Softmax}\left(\frac{(\mathbf{Q}\mathbf{R}_i)\cdot(\mathbf{K}\mathbf{R}_j)^{\mathrm{T}}}{\sqrt{d_k}}\right)}_{\textbf{AttnMap}} \cdot \mathbf{V}$$

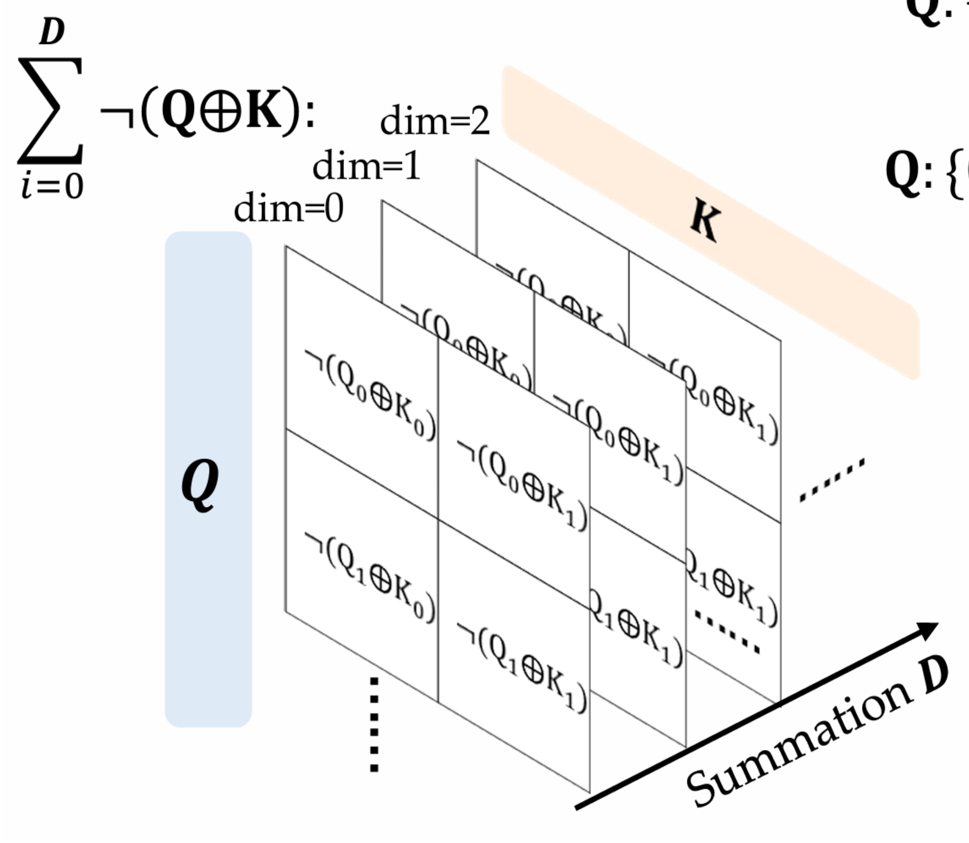
Distance Consistency:

The magnitude of $R_{i,j}$ is determined exclusively by the relative positional offset a $|i-j|$ ensuring that the model systematically differentiates between proximal and distant tokens.

Contents

XNOR Spiking Self Attention

XNOR-Based SSA



$\mathbf{Q}: \{0, 1\}^{T\times L\times D}$ $\quad$ $\mathbf{K}: \{0, 1\}^{T\times L\times D}$

unsqueeze $\quad$ unsqueeze

$\mathbf{Q}: \{0, 1\}^{T\times L\times 1\times D}$ $\quad$ $\mathbf{K}: \{0, 1\}^{T\times 1\times L\times D}$

$\neg(\mathbf{Q}\oplus\mathbf{K})$ $\quad$ XNOR (Broadcast)

$\mathbb{N}_0^{T\times L\times L\times D}$

$\sum_{i=0}^{D}$ $\quad$ Summation Along Dimension D

$\mathbf{AttnMap} \in \mathbb{N}_0^{T\times L\times L}$

Attention map is a similarity map.

For binary matrices, we can use **Hamming Distance** as similarity

Hamming Distance

A	1	0	1	$\boxed{1}$	0	0	$\boxed{1}$	0	$\boxed{0}$	1
B	1	0	1	$\boxed{0}$	0	0	$\boxed{0}$	0	$\boxed{1}$	1

$d_H(A, B) = 3$

XNOR == Hamming Distance

$$\mathbf{AttnMap} = \sum_{i=0}^{D} \neg(\mathbf{Q} \oplus \mathbf{K})$$

Contents

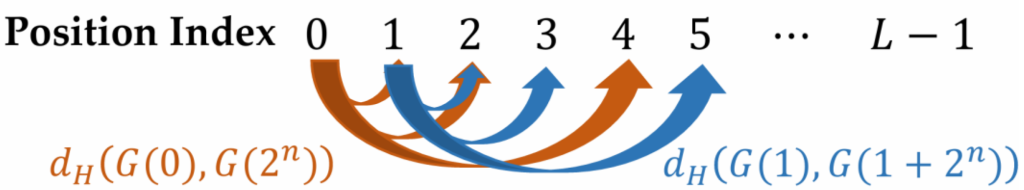
Gray-PE

Theorem 1. *(Proof in Appendix A) For two position indices differing by $2^n (n \geq 0)$, their Gray Code representations have a consistent Hamming distance. Specifically, $\forall$ position i, we have:*

$$d_H(G(i), G(i+2^n)) = \begin{cases} 1 & \textit{if } n = 0, \\ 2 & \textit{if } n \geq 1. \end{cases}$$

Gray-PE

$$d_H(G(i), G(i+2^n)) = \begin{cases} 1, & \textit{if } n = 0 \\ 2, & \textit{if } n \geq 1 \end{cases}, \text{ for any position } i$$

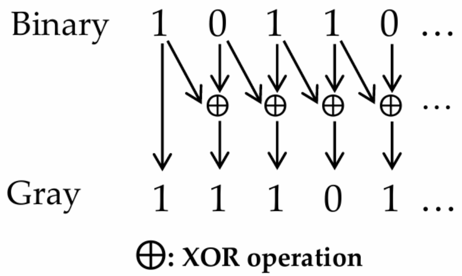


Implementation on XNOR-Based SSA:

$$\neg([\,Q \mid G(l)\,] \oplus [\,K \mid G(l)\,])$$

Concatenate on Dimension D

Gray Code

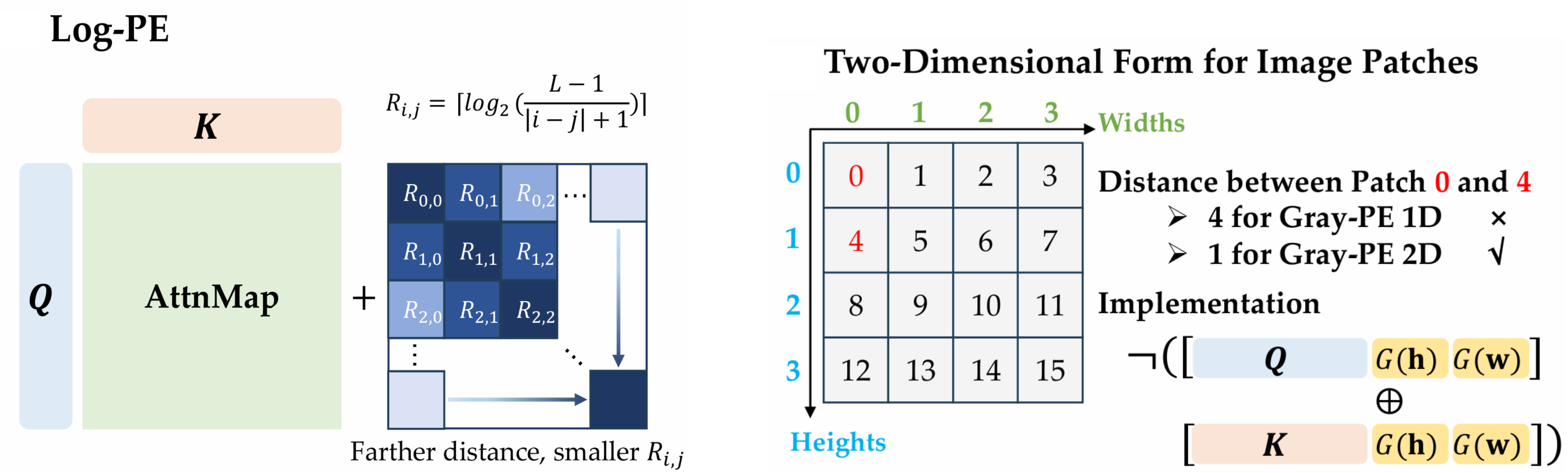


Decimal	Binary	Gray
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110

$$\textbf{AttnMap} = \sum_{i=0}^{D} \neg([\mathbf{Q} \parallel G(\boldsymbol{l})] \oplus [\mathbf{K} \parallel G(\boldsymbol{l})])$$

Log-PE

$$\mathbf{AttnMap} = \left(\sum_{i=0}^{D} \neg(\mathbf{Q} \oplus \mathbf{K})\right) + \mathbf{R}_{i,j}, \ \text{where} \ \mathbf{R}_{i,j} = [R_{i,j}] = \left[\left\lceil \log_2\left(\frac{L-1}{|i-j|+1}\right)\right\rceil\right]$$



Contents

Time-series Forecasting

Models	PE Spike	PE Type	Metric	Metr-la ($L=12$) 6	24	48	96	Pems-bay ($L=12$) 6	24	48	96	Solar ($L=168$) 6	24	48	96	Electricity ($L=168$) 6	24	48	96	Avg.
Transformer w/ RoPE	✗	R	R^2↑	**.729**	**.560**	**.416**	**.306**	**.787**	.730	**.694**	.676	.951	.854	**.763**	**.720**	**.984**	.978	.974	**.968**	**.756**
			RSE↓	**.548**	**.696**	**.802**	**.878**	**.499**	.563	**.600**	.617	.225	**.373**	**.492**	**.539**	.251	.274	.341	**.420**	**.507**
Transformer w/ ALiBi	✗	R	R^2↑	.725	.558	.409	.293	.782	.727	.690	**.677**	.924	.845	.741	.665	**.984**	**.980**	**.976**	**.968**	.747
			RSE↓	.556	.700	.814	.885	.507	.569	.606	**.615**	.281	.393	.527	.602	**.250**	**.271**	**.339**	.422	.521
Transformer w/ Sin-PE	✗	A	R^2↑	.727	.554	.413	.284	.785	**.734**	.688	.673	**.953**	**.858**	.759	.718	.978	.975	.972	.964	.752
			RSE↓	.551	.704	.808	.895	.502	**.558**	.610	.618	**.223**	.377	.504	.545	.260	.277	.347	.425	.512
Spikformer w/ Conv-PE (Original)	✓	A	R^2↑	.713	.527	.399	.267	.773	.697	.686	.667	.929	.828	.744	.674	.959	.955	.955	.954	.733
			RSE↓	.565	.725	.818	.903	.514	.594	.606	.621	.272	.426	.519	.586	.373	.371	.379	.382	.541
Spikformer w/ ALiBi	✗	R	R^2↑	.665	.483	.380	.104	.760	.644	.348	.064	.080	.080	.080	.080	.710	.710	.710	.710	.413
			RSE↓	.622	.768	.833	1.02	.529	.709	.870	1.04	1.01	1.01	1.01	1.01	1.03	1.03	1.03	1.03	.909
Spikformer w/ RoPE	✗	R	R^2↑	.699	.493	.390	.243	.768	.699	.680	.664	.911	.820	.714	.644	.954	.951	.949	.940	.720
			RSE↓	.584	.757	.835	.920	.519	.591	.614	.625	.294	.441	.550	.633	.375	.383	.384	.454	.559
Spikformer w/ CPG-PE	✓	A	R^2↑	.726	.526	.419	.287	.780	.712	.690	.666	.937	.833	.757	.707	.972	.970	.966	.960	.744
			RSE↓	.553	.720	.806	.890	.508	.580	.602	.622	.257	.420	.506	.555	.299	.310	.314	.355	.519
Spikformer-XNOR w/ Conv-PE	✓	A	R^2↑	.718	.531	.405	.269	.771	.693	.690	.665	.928	.829	.740	.669	.960	.957	.955	.953	.733
			RSE↓	.559	.721	.813	.910	.518	.599	.613	.628	.273	.421	.527	.595	.365	.371	.376	.384	.542
Spikformer-XNOR w/ Gray-PE	✓	R	R^2↑	.728	**.544**	.414	**.295**	.782	**.724**	**.694**	**.673**	**.936**	.840	.756	.710	.974	.972	.966	.962	.748
			RSE↓	.546	**.706**	.806	.885	.506	.578	**.597**	**.618**	**.257**	.409	.507	.546	.276	.304	.320	.342	.513
Spikformer-XNOR w/ Log-PE	✓	R	R^2↑	**.735**	.535	**.424**	.290	**.789**	.717	.691	.670	.933	**.841**	**.758**	**.734**	**.978**	**.974**	**.968**	**.964**	**.750**
			RSE↓	**.543**	.719	**.799**	**.876**	**.496**	**.575**	.601	.620	.265	**.408**	**.504**	**.525**	**.272**	**.300**	**.314**	**.340**	**.509**
Spikingformer w/o PE (Original)	–	–	R^2↑	.717	.530	.362	.212	.800	.704	.681	.629	.934	.751	.518	.381	.973	.971	.967	.964	.693
			RSE↓	.560	.720	.842	.936	.483	.587	.611	.659	.258	.500	.694	.788	.299	.305	.325	.340	.557
Spikingformer-XNOR w/ Gray-PE	✓	R	R^2↑	.720	**.537**	.396	**.260**	**.820**	.714	.681	**.646**	.934	.832	.535	.420	.970	.973	**.973**	.965	.711
			RSE↓	.558	**.712**	.819	.907	**.459**	.578	.610	**.643**	.257	.421	.663	.768	.305	.293	.294	.338	.539
Spikingformer-XNOR w/ Log-PE	✓	R	R^2↑	**.737**	.535	**.403**	**.260**	.816	**.719**	**.682**	.640	**.939**	**.854**	**.544**	**.434**	**.977**	**.974**	.972	**.967**	**.716**
			RSE↓	**.540**	.714	**.814**	**.906**	.463	**.573**	**.609**	.652	**.246**	**.382**	**.651**	**.759**	**.270**	**.292**	**.293**	**.336**	**.531**
SDT-V1 w/ Conv-PE (Original)	✓	A	R^2↑	.689	.517	.409	.253	.769	.700	.647	.630	.917	.819	.723	.655	.956	.952	.949	.950	.721
			RSE↓	.604	.735	.811	.915	.522	.596	.665	.673	.286	.439	.538	.602	.371	.376	.388	.386	.557
SDT-V1 w/ CPG-PE	✓	A	R^2↑	.701	.525	**.418**	.257	.778	**.716**	.660	**.656**	.919	**.820**	.710	.644	.963	.960	.958	.952	.727
			RSE↓	.585	.724	**.799**	.920	.515	**.578**	.633	.642	.285	.439	.558	.637	.361	.368	.370	.376	.548
SDT-V1 w/ Log-PE	✓	R	R^2↑	**.714**	**.531**	.415	**.265**	**.784**	.709	**.672**	.654	**.921**	**.820**	**.730**	**.674**	**.972**	**.968**	**.963**	**.957**	**.734**
			RSE↓	**.554**	**.713**	.807	**.904**	**.502**	.585	**.629**	**.641**	**.280**	**.437**	**.527**	**.598**	**.353**	**.356**	**.360**	**.366**	**.538**
QKFormer w/ Conv-PE (Original)	✓	A	R^2↑	.717	.513	.376	.246	.767	.706	.681	.654	.920	.748	.512	.416	.970	.967	.963	.958	.695
			RSE↓	.561	.735	.832	.917	.521	.586	.609	.635	.289	.515	.716	.784	.306	.319	.355	.367	.565
QKFormer w/ CPG-PE	✓	A	R^2↑	.740	.554	.419	.276	.783	.714	.702	.660	.922	.754	.702	.604	.977	.969	.968	.963	.732
			RSE↓	.536	.704	.803	.896	.503	.578	.589	.633	.285	.520	.581	.645	.266	.312	.315	.332	.531
QKFormer-XNOR w/ Gray-PE	✓	R	R^2↑	**.742**	**.551**	**.418**	**.274**	.799	**.715**	.691	**.674**	.927	.817	.710	.691	.974	.970	.968	.965	.742
			RSE↓	**.534**	**.711**	**.804**	**.898**	.484	**.577**	.601	**.616**	.276	.438	.556	.570	.277	.310	.314	.331	.519
QKFormer-XNOR w/ Log-PE	✓	R	R^2↑	**.742**	.541	.416	.265	**.801**	.710	**.707**	.661	**.928**	**.818**	**.748**	**.698**	**.978**	**.974**	**.972**	**.966**	**.746**
			RSE↓	.535	.715	.805	.903	**.482**	.581	**.585**	.629	**.274**	**.437**	**.515**	**.564**	**.264**	**.285**	**.296**	**.328**	**.514**

Both Gray-PE and Log-PE are effective on various spiking Transformers, including Spikformer、Spikingformer、SDT-V1、QKFormer.

Other Tasks

Text Classification

Model	PE Spike	PE Type	Param(M)	English Dataset (Length = 128) MR	SST-2	Subj	SST-5	Chinese Dataset (Length = 32) ChnSenti	Waimai	Avg.
Fine-tuned BERT	✗	A	109.8	**87.63**±0.18	**92.31**±0.17	**95.90**±0.16	**50.41**±0.13	**89.48**±0.16	**90.27**±0.13	**84.33**
Spikformer w/o PE	–	–	109.8	75.87±0.35	81.71±0.31	91.60±0.30	41.84±0.39	85.62±0.25	86.87±0.28	77.25
Spikformer w/ CPG-PE	✓	A	110.4	82.42±0.42	82.90±0.33	92.50±0.25	43.62±0.36	86.54±0.26	**88.49**±0.29	79.41
Spikformer-XNOR w/o PE	–	–	109.8	75.80±0.40	81.74±0.40	91.50±0.29	41.88±0.38	85.64±0.31	86.66±0.33	77.20
Spikformer-XNOR w/ Gray-PE	✓	R	109.8	83.73±0.45	84.52±0.39	92.50±0.33	44.06±0.48	87.41±0.36	88.40±0.30	80.11
Spikformer-XNOR w/ Log-PE	✓	R	109.8	**83.88**±0.40	**84.64**±0.37	**92.80**±0.30	**44.52**±0.43	**87.95**±0.34	88.46±0.28	**80.38**

Image Classification

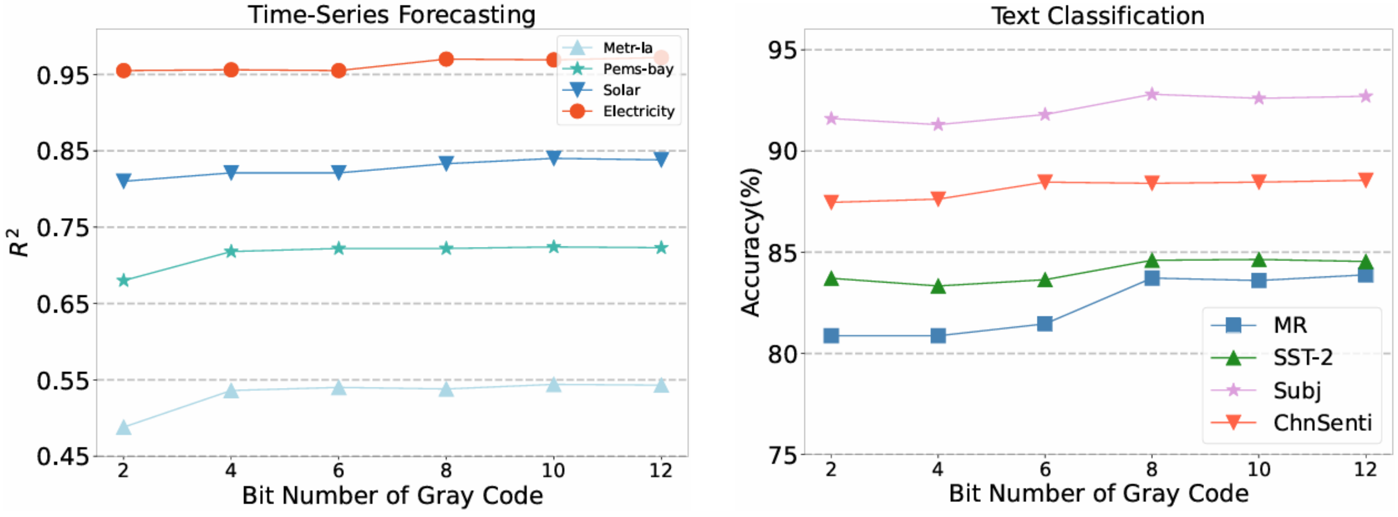
Model	PE Spike	PE Type	CIFAR10 Param (M)	CIFAR10 Acc	CIFAR10-DVS Param (M)	CIFAR10-DVS Acc	CIFAR100 Param (M)	CIFAR100 Acc	Tiny-ImageNet Param (M)	Tiny-ImageNet Acc	Avg.
Vision-Transformer	✗	A	9.32	**96.73**	–	–	9.36	**81.02**	9.40	**62.18**	–
Spikformer w/ Conv-PE (Original)	✓	A	9.32	94.80*	2.57	78.10*	9.36	77.04*	9.40	48.10*	74.51
Spikformer w/ CPG-PE	✓	A	8.17	95.06	2.06	78.40	8.20	77.82	8.24	48.52*	74.95
Spikformer-XNOR w/ Gray-PE 1D	✓	R	8.00	95.46	1.99	77.90	8.04	78.12	8.08	48.33	74.95
Spikformer-XNOR w/ Gray-PE 2D	✓	R	8.00	**95.66**	1.99	**78.70**	8.04	**78.45**	8.08	**48.74**	**75.39**

Analysis

Bits of Gray Code

Pigeonhole principle

$L - 1 > 2^b$



(a) (b)

Long Sequence Analysis:

AGNEWS, Length 1024

IMDB, Length 2048

Model	PE Spike	PE Type	AGNEWS	IMDB	Avg.
Fine-tuned BERT	✗	A	**94.50**	**92.10**	**93.30**
Spikformer w/ Conv-PE (Original)	✓	A	83.84	79.08	81.46
Spikformer w/ CPG-PE	✓	A	84.70	79.47	82.09
Spikformer-XNOR w/ Gray-PE	✓	R	84.92	79.79	82.36
Spikformer-XNOR w/ Log-PE	✓	R	**86.77**	**80.46**	**83.62**

THANKS!

Contact us:

czlv24@m.fudan.edu.cn