Efficient and Effective Time-Series Forecasting with Spiking Neural Networks

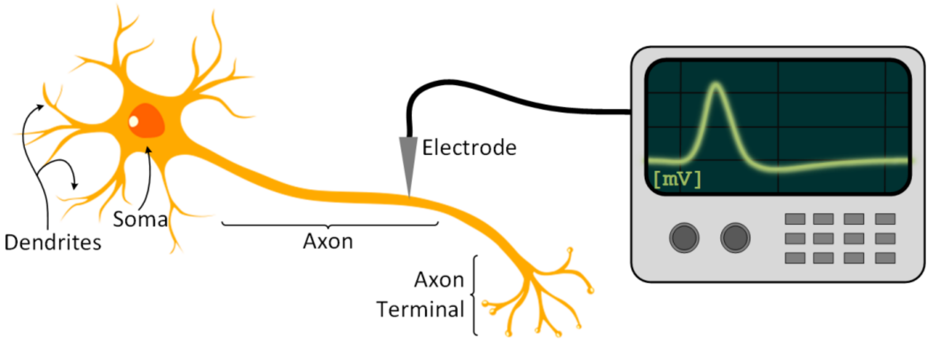
Changze Lv[1], Yansen Wang[2], Dongqi Han[2], Xiaoqing Zheng[1], Xuanjing Huang[1], Dongsheng Li[2]

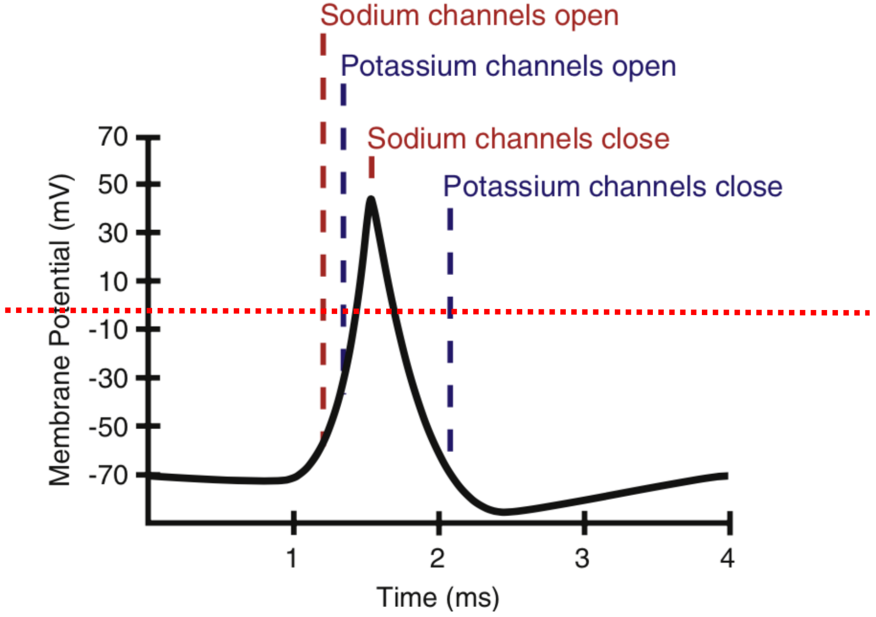
[1]Fudan University, [2]Microsoft Research Asia

Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

Spike Neuron





1

Threshold Potential

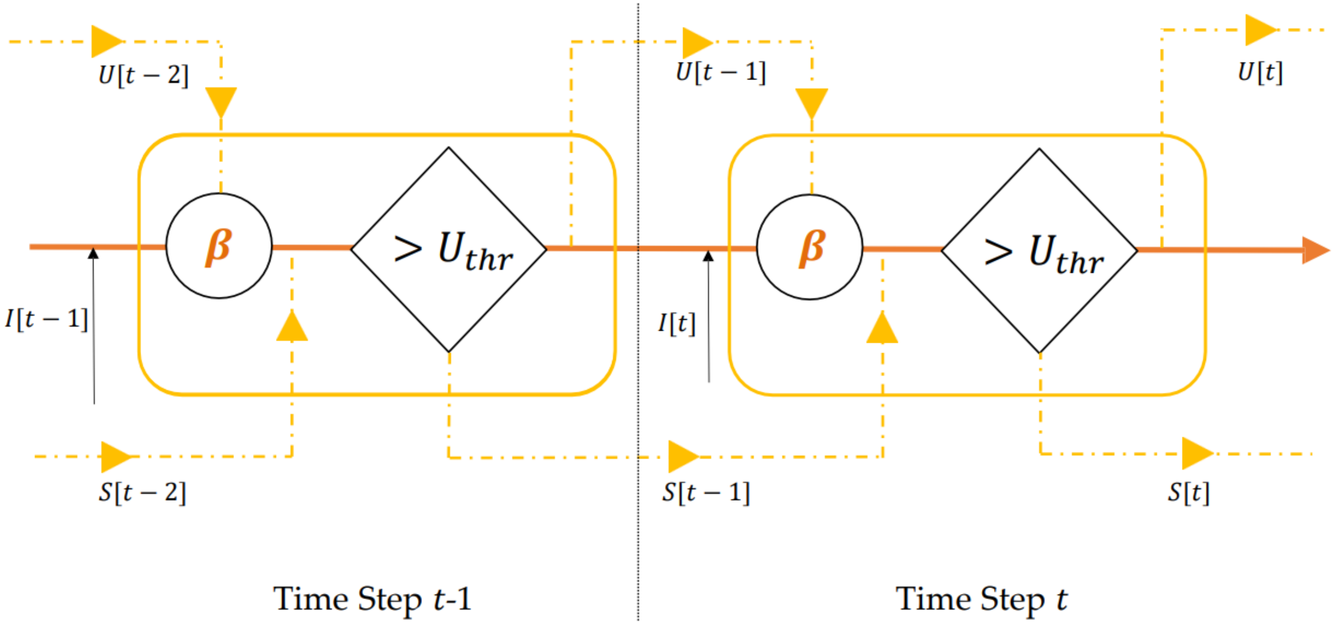
0

Information in Spiking Neurons

- Floating-point Numbers ✗
- Binary Value ✓

"Training Spiking Neural Networks Using Lessons From Deep Learning."
Eshraghian, Jason Kamran et al. 2021

Spiking Neural Networks

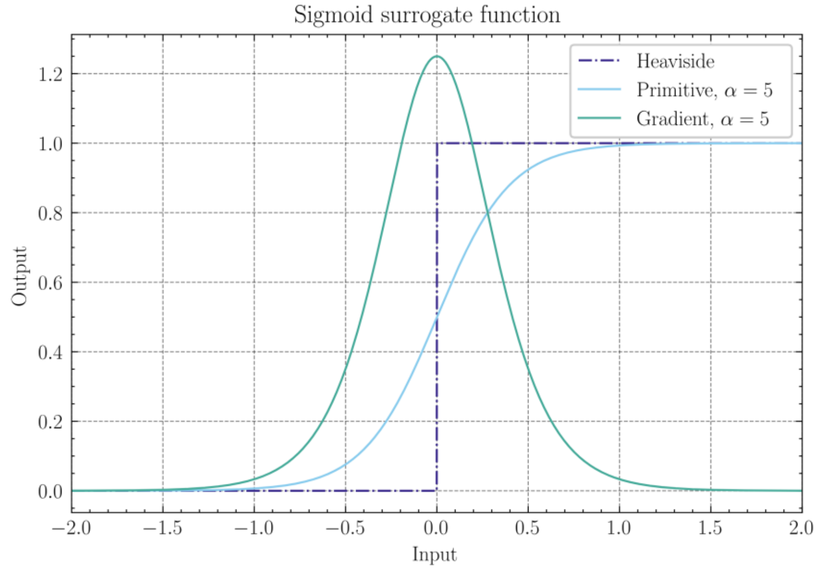


$$U(t) = H(t - \Delta t) + I(t), \quad I(t) = f(\mathbf{x};\theta),$$

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

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

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



https://spikingjelly.readthedocs.io/

Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

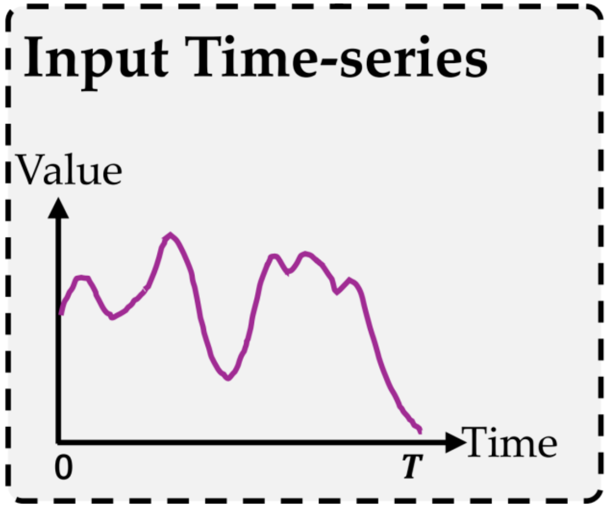
Time-series Forecasting

Observation

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\} \in \mathbb{R}^{T \times C}$$

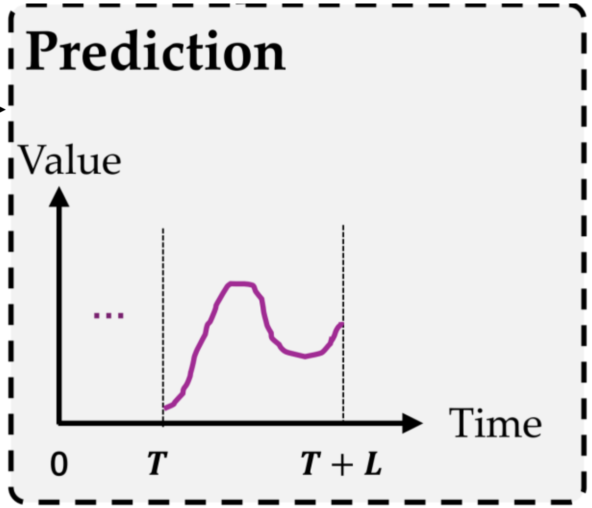
Horizon

$$\mathbf{Y} = \{\mathbf{x}_{T+1}, \mathbf{x}_{T+2}, \ldots, \mathbf{x}_{T+L}\} \in \mathbb{R}^{L \times C}$$



Forecast

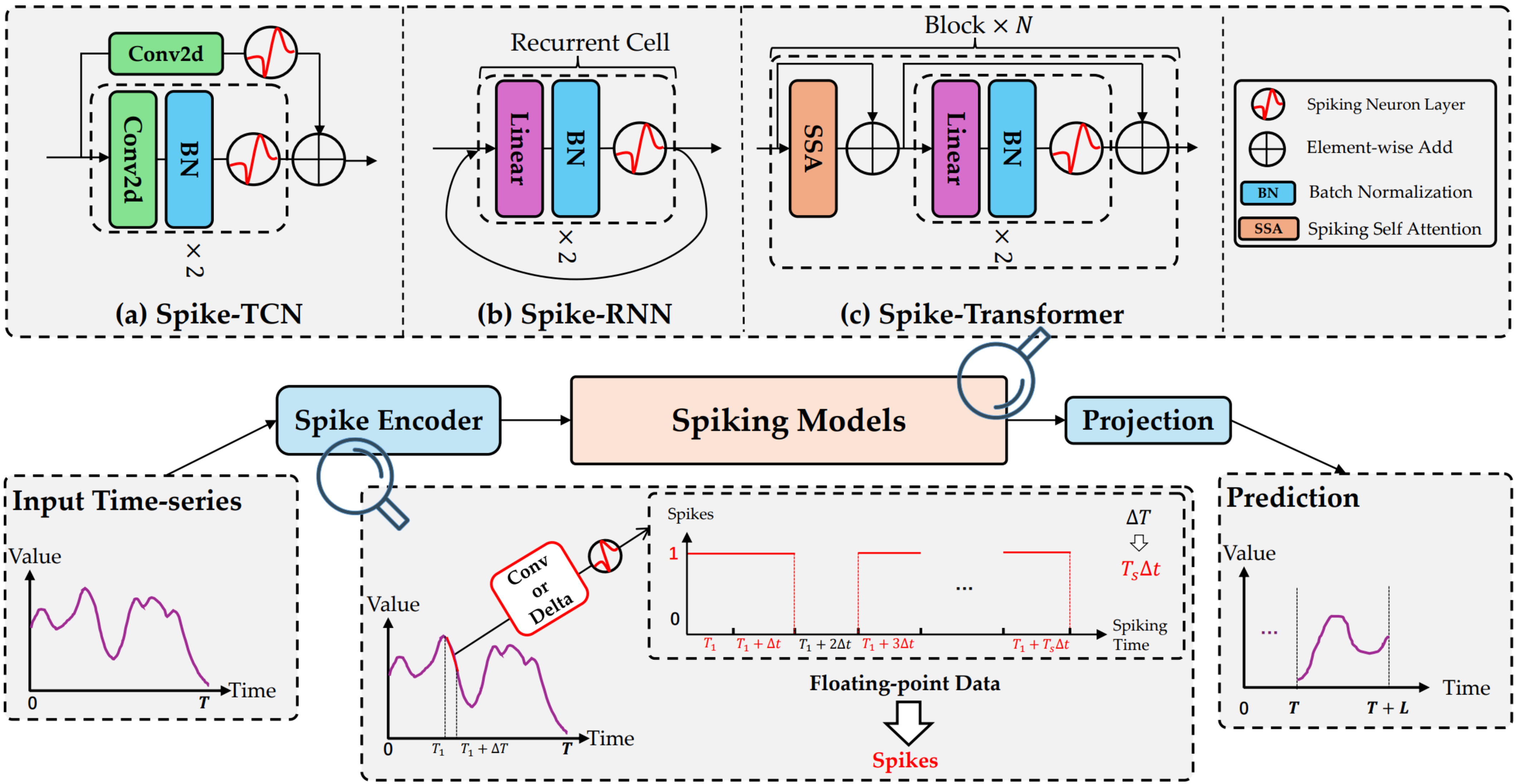
- Statistics-base
- CNN-base
- RNN-base
- Transformer-base
- …



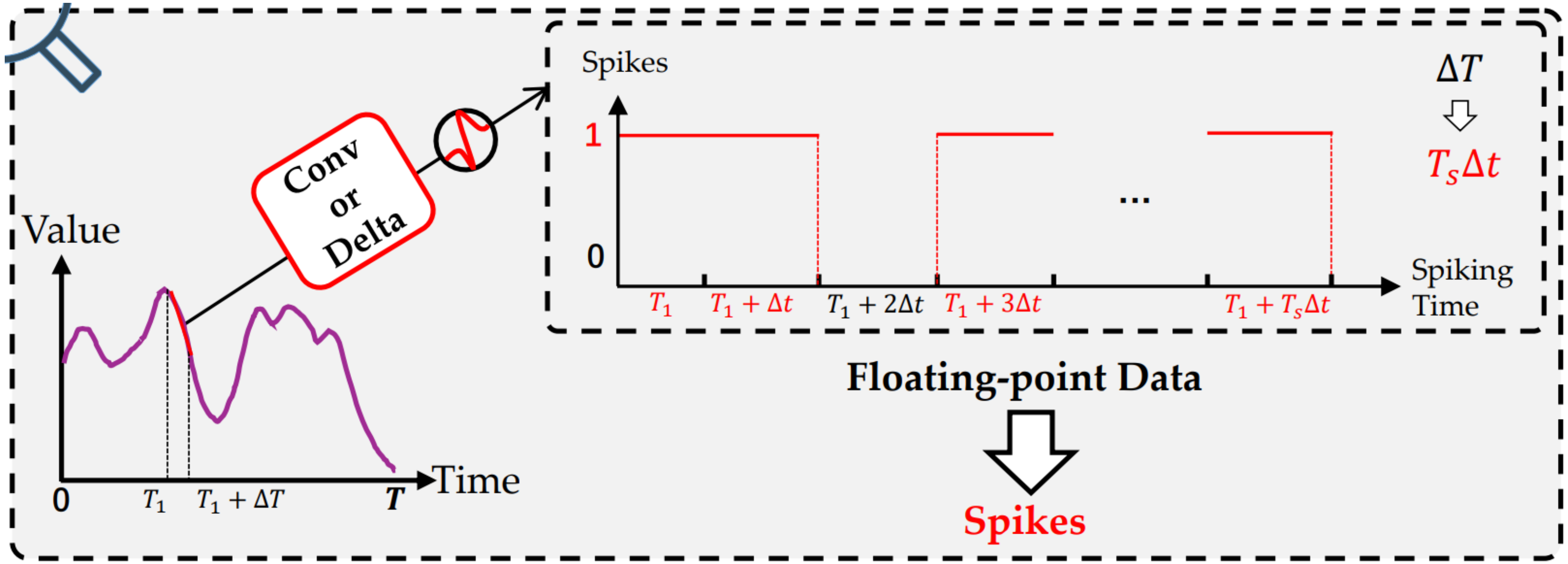
Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

Overview


Conv2d
Conv2d
BN
× 2
(a) Spike-TCN
Recurrent Cell
Linear
BN
× 2
(b) Spike-RNN
Block × N
SSA
Linear
BN
× 2
(c) Spike-Transformer
Spiking Neuron Layer
Element-wise Add
BN Batch Normalization
SSA Spiking Self Attention
Spike Encoder
Spiking Models
Projection
Input Time-series
Value
0
T
Time
Conv or Delta
Value
0
T_1
$T_1 + \Delta T$
T
Time
Spikes
1
0
T_1
$T_1 + \Delta t$
$T_1 + 2\Delta t$
$T_1 + 3\Delta t$
...
$T_1 + T_s\Delta t$
Spiking Time
ΔT
$T_s\Delta t$
Floating-point Data
Spikes
Prediction
Value
...
0
T
T + L
Time

Temporal Alignment



Temporal Alignment

- Divide a time step ΔT of the time series into T_s segments.

$$\Delta T = T_s \Delta t$$

- Target: Floating-point Numbers → Spike Trains

$$\mathbb{R}^{T\times C} \rightarrow \{0,1\}^{T_s\times T\times C}$$

Spike Encoders

Target: $\mathbb{R}^{T \times C} \rightarrow \{0,1\}^{T_s \times T \times C}$

For previous studies on image processing:

Repetition for T_s times is widely-used.

Disrupt the continuous nature of time series !!!

- Delta-base

$$\mathbf{S} = \mathcal{SN}\left(\mathrm{BN}\left(\mathrm{Linear}\left(\mathbf{x}_t - \mathbf{x}_{t-1}\right)\right)\right)$$

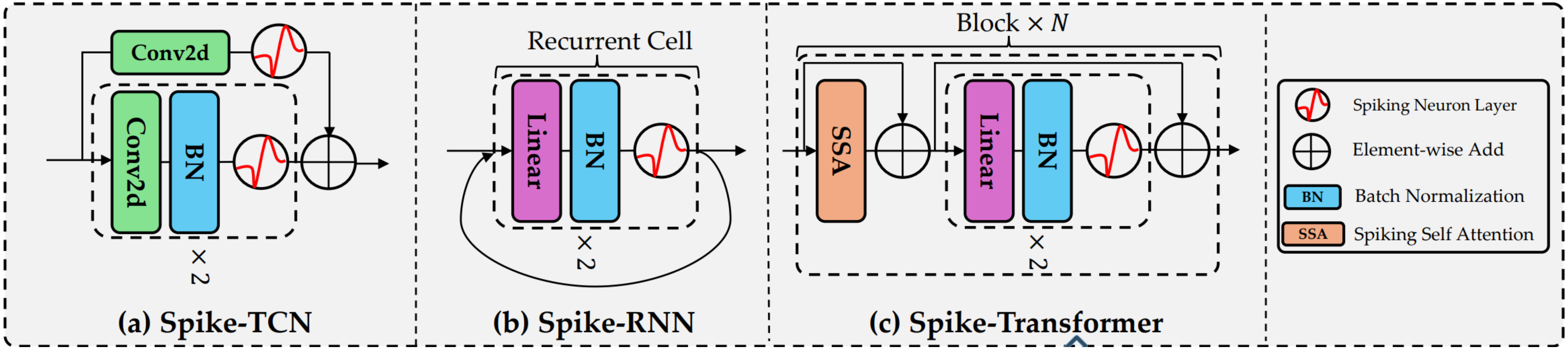
- Convolution-base

$$\mathbf{S} = \mathcal{SN}\left(\mathrm{BN}\left(\mathrm{Conv}\left(\mathbf{X}\right)\right)\right)$$

Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

Spiking Model Architecture



Model Architectures:

✓TCN → Spike-TCN

✓RNN → Spike-RNN

✓iTransformer → iSpikformer

Rules:

✓Replace ReLU as Leaky-IF

✓SEW Residual Module

✓Softmax-free Spiking Self-attention

Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

Main Experiments

Datasets:

- **Metr-la**: Average traffic speed measured on the highways of Los Angeles County;
- **Pems-bay**: Average traffic speed in the Bay Area;
- **Solar**: Records of solar power production.
- **Electricity**: Hourly electricity consumption measured in kWh;

Dataset	Samples	Variables	Length	Train-Valid-Test Ratio
Metr-la	$34,272$	207	12	$(0.7, 0.2, 0.1)$
Pems-bay	$52,116$	325	12	$(0.7, 0.2, 0.1)$
Solar-energy	$52,560$	137	168	$(0.6, 0.2, 0.2)$
Electricity	$26,304$	321	168	$(0.6, 0.2, 0.2)$

Evaluation Metrics:

$$\mathrm{RSE} = \sqrt{\frac{\sum_{m=1}^{M} ||\mathbf{Y}^m - \hat{\mathbf{Y}}^m||^2}{\sum_{m=1}^{M} ||\mathbf{Y}^m - \bar{\mathbf{Y}}||^2}},$$

$$R^2 = \frac{1}{MCL} \sum_{m=1}^{M} \sum_{c=1}^{C} \sum_{l=1}^{L} \left[1 - \frac{(Y_{c,l}^m - \hat{Y}_{c,l}^m)^2}{(Y_{c,l}^m - \bar{Y}_{c,l})^2} \right]$$

Main Experiments

Table 1. Experimental results of time-series forecasting on 4 benchmarks with different prediction lengths (horizons) L. The best and the second-placed results are formatted in bold font and underlined format. ↑ (↓) indicates the higher (lower) the better. All SNNs are equipped with a convolutional spike encoder in this table. The numbers in the **Avg. Rank** column indicate the average ranking of the current row's models within each specific setting. Numbers in the **Avg.** column with * indicate that a model significantly ($p < 0.05$) outperforms its counterpart. All results are averaged across 3 random seeds.

Method	Spike	Metric	Metr-la				Pems-bay				Solar				Electricity				Avg.	Avg. Rank↓
			6	24	48	96	6	24	48	96	6	24	48	96	6	24	48	96		
ARIMA	✗	R^2↑	.687	.441	.282	.265	.741	.723	.692	.670	.951	.847	.725	.682	.963	.960	.914	.863	.713	7.3
		RSE↓	.575	.742	.889	.902	.532	.548	.562	.612	.202	.365	.588	.589	.522	.534	.564	.599	.583	7.3
GP	✗	R^2↑	.685	.437	.265	.233	.732	.712	.689	.665	.944	.836	.711	.675	.962	.968	.912	.852	.705	8.4
		RSE↓	.572	.738	.912	.925	.544	.532	.577	.592	.225	.388	.612	.575	.603	.612	.633	.642	.605	7.6
TCN	✗	R^2↑	.820	.601	.455	**.330**	.881	**.749**	**.695**	**.689**	.958	.871	.737	.661	.975	.973	.968	.962	.770	3.9
		RSE↓	.446	.665	.778	**.851**	.373	.541	.583	.587	.210	.359	.513	.583	.282	.287	**.319**	.345	.483	3.6
Spike-TCN	✓	R^2↑	.783	.603	**.468**	.326	.811	.729	.662	.633	.937	.840	.708	.650	.970	.963	.958	.953	.750	7.0
		RSE↓	.491	.665	.769	.865	.469	.541	.625	.635	.259	.401	.541	.596	.333	.342	.368	.389	.518	6.1
GRU	✗	R^2↑	.759	.429	.301	.194	.747	.703	.691	.665	.950	.875	.781	.737	**.981**	.972	.971	**.964**	.733	5.8
		RSE↓	.517	.797	.882	.947	.529	.573	.584	.608	.219	.355	.476	.522	.506	.598	.537	.587	.573	7.1
Spike-GRU	✓	R^2↑	**.846**	.615	.427	.275	.864	.741	.688	.657	.912	.822	.771	.668	.978	.964	.962	.959	.759	6.2
		RSE↓	.414	.663	.827	.943	.398	.535	.601	.621	.299	.430	.485	.629	.280	.317	.338	.484	.517*	6.0
Spike-RNN	✓	R^2↑	**.846**	.622	.433	.283	.872	.745	.685	.654	.923	.820	**.812**	.714	.977	.972	.962	.960	.768*	5.2
		RSE↓	**.412**	**.648**	.794	.935	.387	**.528**	.588	.634	.278	.425	**.435**	.586	.267	.296	.346	.481	.503*	4.8
Autoformer	✗	R^2↑	.762	.548	.411	.282	.782	.711	.689	.668	.960	.852	.791	.701	.980	**.977**	**.975**	.963	.753	4.6
		RSE↓	.565	.692	.785	.872	.452	.543	.577	**.565**	.212	.432	.622	.685	.481	.506	.566	.548	.569	6.6
iTransformer	✗	R^2↑	.829	**.623**	.439	.285	**.887**	.719	.685	.668	**.964**	**.879**	.799	**.738**	.979	**.977**	**.975**	**.964**	**.776**	**2.9**
		RSE↓	.436	**.648**	.780	.878	**.362**	.547	**.561**	.584	**.191**	.348	.448	.563	**.259**	.305	.335	.427	.480	**2.8**
iSpikformer	✓	R^2↑	.817	.618	.440	.279	.879	.744	.687	.674	.961	.876	.795	**.738**	.977	.974	.972	.963	.775	3.5
		RSE↓	.475	.668	**.752**	.905	.376	.536	.569	.580	.204	**.333**	.465	**.521**	.263	**.284**	.338	**.348**	**.476**	2.9

Main Experiments

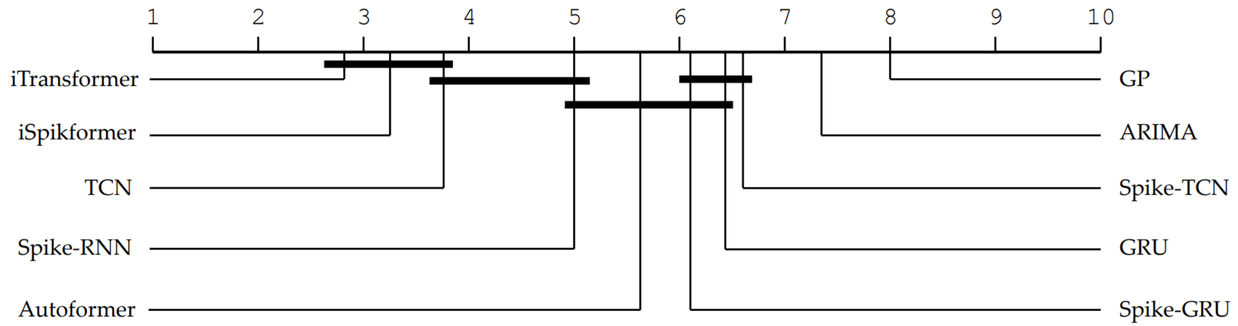


Figure 3. Critical Difference (CD) diagram of all methods in Table 1 on time series forecasting tasks with a confidence level of 95%.

Findings:

✓SNNs succeed when temporal dynamics are properly preserved and handled.

✓ SNNs with our temporal modeling can be further improved by advanced spatial modeling techniques.

Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

Model Analysis

Method	Encoder	Metric	Metr-la 6	Metr-la 24	Metr-la 48	Metr-la 96	Electricity 6	Electricity 24	Electricity 48	Electricity 96
Spike-TCN	Convolutional	$R^2\uparrow$	**.783**	**.603**	**.468**	**.326**	**.970**	**.963**	**.958**	**.953**
		RSE↓	**.491**	**.664**	.769	.935	**.333**	**.342**	**.368**	**.389**
	Delta	$R^2\uparrow$	.751	.582	.458	.317	.963	.956	.948	.942
		RSE↓	.525	.676	**.768**	**.871**	.344	.371	.432	.460
	Repetition	$R^2\uparrow$	.024∗	.024∗	.022∗	.020∗	.878	.710∗	.710∗	.710∗
		RSE↓	1.05∗	1.04∗	1.04∗	1.05∗	.662	1.03∗	1.03∗	1.03∗
Spike-RNN	Convolutional	$R^2\uparrow$	**.846**	**.622**	**.433**	**.283**	**.977**	**.972**	**.962**	**.960**
		RSE↓	**.412**	**.648**	**.794**	**.935**	**.267**	**.296**	.346	**.481**
	Delta	$R^2\uparrow$	.839	.616	.430	.277	.969	.966	.962	.876
		RSE↓	.420	.652	.799	.938	.301	.318	**.344**	.685
	Repetition	$R^2\uparrow$	.817	.578	.021∗	.021∗	.901	.816	.710∗	.710∗
		RSE↓	.481	.684	1.04∗	1.04∗	.592	.766	1.03∗	1.04∗
iSpikformer	Convolutional	$R^2\uparrow$	**.817**	**.618**	**.440**	**.279**	**.977**	**.974**	**.972**	**.963**
		RSE↓	**.475**	.668	**.752**	**.905**	**.263**	**.284**	**.338**	**.348**
	Delta	$R^2\uparrow$	.804	.601	.434	.272	.972	.969	.960	.944
		RSE↓	.496	**.666**	.759	.910	.274	.302	.391	.455
	Repetition	$R^2\uparrow$	.692	.548	.238	.021∗	.962	.953	.849	.710∗
		RSE↓	.573	.708	.847	1.04∗	.289	.557	.705	1.03∗

- SNNs with repetition encoders may struggle to converge;
- Both convolutional and delta spike encoder are effective event-driven spike generators;
- Shape-based encoder which takes a wide scope of sequence into consideration is more effective than a change-based encoder.

Model Analysis

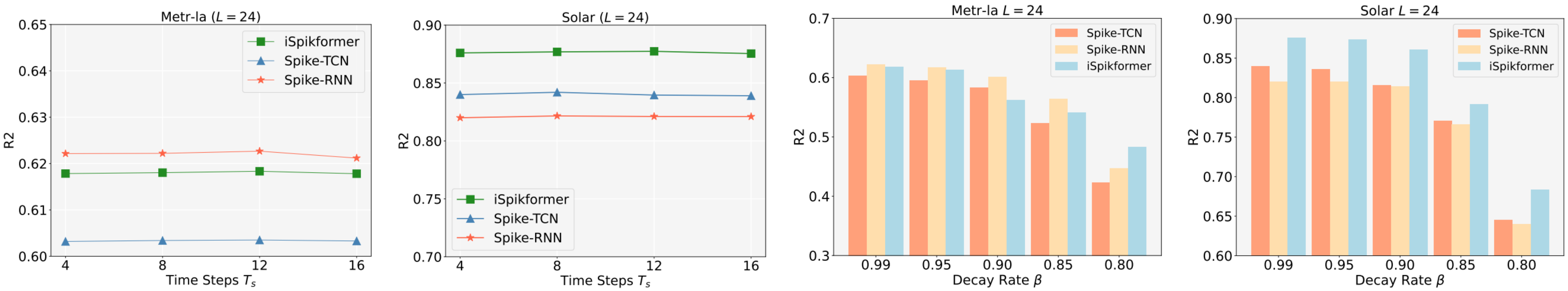


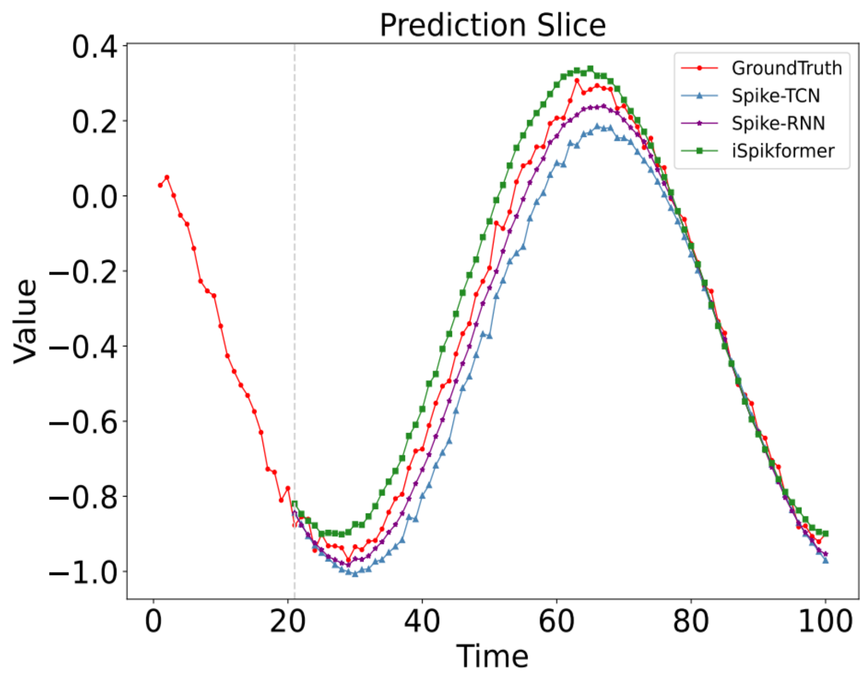
Figure 4. The impact of two crucial hyper-parameters in SNNs: time Steps T_s and the decay rate β. (a) and (b): R^2 versus T_s on Metr-la and Solar respectively. (c) and (d): R^2 versus β on Metr-la and Solar respectively. The horizon L of these experiments is set to 24.

- Time Step T_s: the R^2 values remain relatively stable with minimal variation as T_s increases.
- Decay Rate β: a higher β makes the SNN more persistent in its internal state, which is beneficial for retaining long-term information.

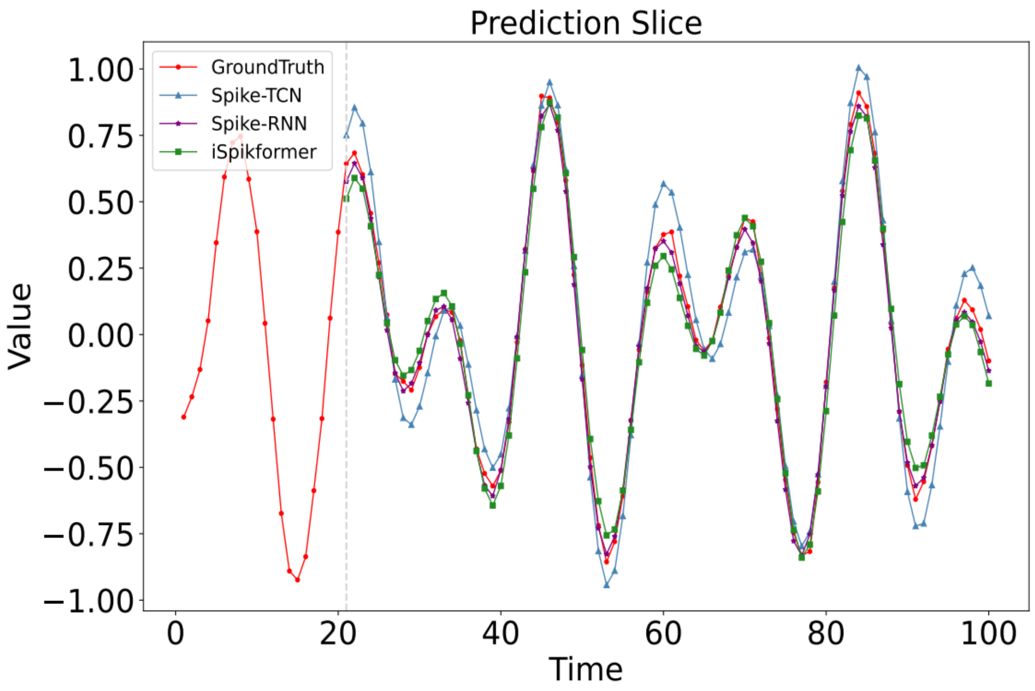
Temporal Analysis

Evaluation on Synthetic Time-series Data:

$$\mathcal{X}(t) = A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t + \phi) + \mathcal{N}(0, \sigma)$$



Low-frequency



High-frequency

Contents

- **Background**
  - Spiking Neural Networks; Time-series Forecasting
- **Methods**
  - Temporal Alignment; Spiking Model Architecture
- **Experiments**
  - Main Experiments; Model and Temporal Analysis; Energy Reduction

Energy Reduction

Energy Estimation

ANNs: For layer *b*,

$$Energy(b) = \overset{4.6\ pJ}{\boxed{E_{MAC}}} \times FLOPs(b)$$

SNNs: For layer *l*,

$$Energy(l) = \overset{0.9\ pJ}{\boxed{E_{AC}}} \times SOPs(l)$$

$$SOPs(l) = T \times \gamma \times FLOPs(l)$$

T is the time step, γ is the firing rate of the input spikes

Model	Param(M)	OPs (G)	Energy (mJ)	Energy Reduction	R^2
TCN	0.460	0.14	0.64	**63.60%** $\downarrow$	.973
Spike-TCN	0.461	0.15	0.23		.963
GRU	1.288	1.32	6.07	**75.05%** $\downarrow$	.972
Spike-GRU	1.289	1.63	1.51		.964
iTransformer	1.634	2.05	9.47	**66.30%** $\downarrow$	.977
iSpikformer	1.634	3.55	3.19		.974

Efficient and Effective Time-Series Forecasting with Spiking Neural Networks

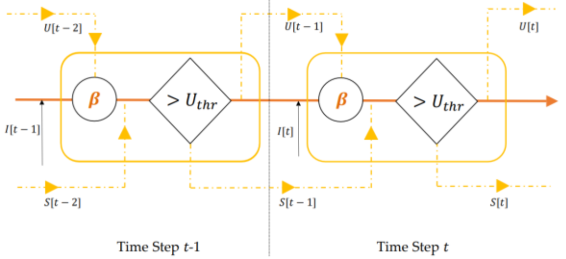
Changze Lv, Yansen Wang, Dongqi Han, Xiaoqing Zheng, Xuanjing Huang, Dongsheng Li
Correspondence to: yansenwang@microsoft.com, zhengxq@fudan.edu.cn

ICML International Conference On Machine Learning

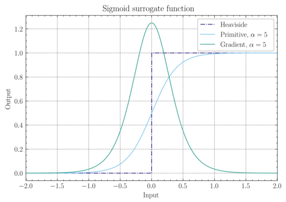
FUDAN UNIVERSITY Microsoft

Project Homepage | Changze Lv (WeChat) | Yansen Wang (WeChat) | Xiaoqing Zheng (WeChat)

Spiking Neural Networks



$$U(t) = H(t - \Delta t) + I(t), \quad I(t) = f(\mathbf{x}; \theta),$$
$$H(t) = V_{reset} S(t) + (1 - S(t))\, \beta U(t),$$
$$S(t) = \begin{cases} 1, & \text{if } U(t) \geq U_{\text{thr}} \\ 0, & \text{if } U(t) < U_{\text{thr}} \end{cases},$$
$$S(t) \approx \frac{1}{\pi} \arctan\left(\frac{\pi}{2} \alpha U(t)\right) + \frac{1}{2}$$



Temporal Alignment and Spike Encoder

Temporal Alignment

- Divide a time step ΔT of the time series into T_s segments. $\Delta T = T_s \Delta t$
- Floating-point Numbers → Spike Trains $\mathbb{R}^{T \times C} \rightarrow \{0,1\}^{T_s \times T \times C}$

Spike Encoder

- Delta-base

$$\mathbf{S} = \mathcal{SN}\left(\mathrm{BN}\left(\mathrm{Linear}\left(\mathbf{x}_t - \mathbf{x}_{t-1}\right)\right)\right)$$

- Convolution-base

$$\mathbf{S} = \mathcal{SN}\left(\mathrm{BN}\left(\mathrm{Conv}\left(\mathbf{X}\right)\right)\right)$$

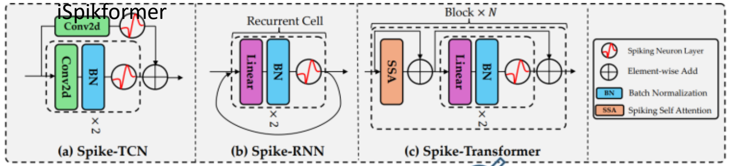
A Unified Framework

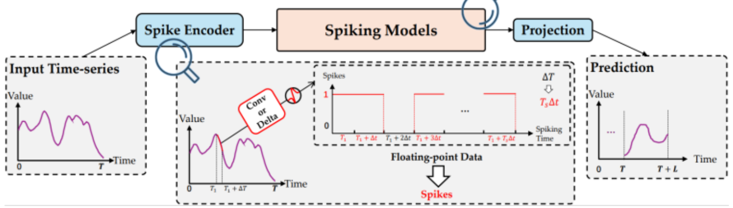
Model Architectures:

- ✓ TCN → Spike-TCN
- ✓ RNN → Spike-RNN
- ✓ iTransformer → iSpikformer

Attention:

- ✓ Replace ReLU as Leaky-IF
- ✓ SEW Residual Module
- ✓ Softmax-free Self-attention

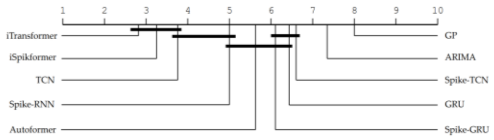




Experimental Results

Real-world Datasets:

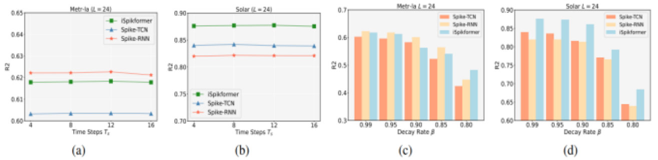
- ✓ Metr-la, Pems-bay
- ✓ Solar-Energy
- ✓ Electricity



Method	Spike	Metric	Metr-la 6	24	48	96	Pems-bay 6	24	48	96	Solar 6	24	48	96	Electricity 6	24	48	96	Avg.
ARIMA	✗	R^2↑	.687	.441	.282	.265	.741	.723	.692	.670	.951	.847	.725	.682	.963	.960	.914	.863	.713
		RSE↓	.575	.742	.889	.902	.532	.548	.562	.612	.202	.365	.588	.589	.522	.534	.564	.599	.583
GP	✗	R^2↑	.685	.437	.265	.233	.732	.712	.689	.665	.944	.836	.711	.675	.962	.968	.912	.852	.705
		RSE↓	.572	.738	.912	.925	.544	.532	.577	.592	.225	.388	.612	.575	.603	.612	.633	.642	.605
TCN	✗	R^2↑	.820	.601	.455	.330	.881	.749	.695	.689	.958	.871	.737	.661	.975	.973	.968	.962	.770
		RSE↓	.446	.665	.778	.851	.373	.541	.583	.587	.210	.359	.513	.583	.282	.287	.319	.345	.483
Spike-TCN	✓	R^2↑	.783	.603	.468	.326	.811	.729	.662	.633	.937	.840	.708	.650	.970	.963	.958	.953	.750
		RSE↓	.491	.665	.769	.865	.469	.541	.625	.635	.259	.401	.541	.596	.333	.342	.368	.389	.518
GRU	✗	R^2↑	.759	.429	.301	.194	.747	.703	.691	.665	.950	.875	.781	.737	.981	.972	.971	.964	.733
		RSE↓	.517	.797	.882	.947	.529	.573	.584	.608	.219	.355	.476	.522	.506	.598	.537	.587	.573
Spike-GRU	✓	R^2↑	.846	.615	.427	.275	.864	.741	.688	.657	.912	.822	.771	.668	.978	.964	.962	.959	.759
		RSE↓	.414	.663	.827	.943	.398	.535	.601	.621	.299	.430	.485	.629	.280	.317	.338	.484	.517*
Spike-RNN	✓	R^2↑	.846	.622	.433	.283	.872	.745	.685	.654	.923	.820	.812	.714	.977	.972	.962	.960	.768*
		RSE↓	.412	.648	.794	.935	.387	.528	.588	.634	.278	.425	.435	.586	.267	.296	.346	.481	.503*
Autoformer	✗	R^2↑	.762	.548	.411	.282	.782	.711	.689	.668	.960	.852	.791	.701	.980	.977	.975	.963	.753
		RSE↓	.565	.692	.785	.872	.452	.543	.577	.565	.212	.432	.622	.685	.481	.506	.566	.548	.569
iTransformer	✗	R^2↑	.829	.623	.439	.285	.887	.719	.685	.668	.964	.879	.799	.738	.979	.977	.975	.964	.776
		RSE↓	.436	.648	.780	.878	.362	.547	.561	.584	.191	.348	.448	.563	.259	.305	.335	.427	.480
iSpikformer	✓	R^2↑	.817	.618	.440	.279	.879	.744	.687	.674	.961	.876	.795	.738	.977	.974	.972	.963	.775
		RSE↓	.475	.668	.752	.905	.376	.536	.569	.580	.204	.333	.465	.521	.263	.284	.338	.348	.476

Ablation Study

Method	Encoder	Metric	Metr-la 6	24	48	96	Electricity 6	24	48	96
Spike-TCN	Convolutional	R^2↑	.783	.603	.468	.326	.970	.963	.958	.953
		RSE↓	.491	.664	.769	.935	.333	.342	.368	.389
	Delta	R^2↑	.751	.582	.458	.317	.963	.956	.948	.942
		RSE↓	.525	.676	.768	.871	.344	.371	.432	.460
	Repetition	R^2↑	.024*	.024*	.022*	.020*	.878	.710*	.710*	.710*
		RSE↓	1.05*	1.04*	1.04*	1.05*	.662	1.03*	1.03*	1.03*
Spike-RNN	Convolutional	R^2↑	.846	.622	.433	.283	.977	.972	.962	.960
		RSE↓	.412	.648	.794	.935	.267	.296	.346	.481
	Delta	R^2↑	.839	.616	.430	.277	.969	.966	.962	.876
		RSE↓	.420	.652	.799	.938	.301	.318	.344	.685
	Repetition	R^2↑	.817	.578	.021*	.021*	.901	.816	.710*	.710*
		RSE↓	.481	.684	1.04*	1.04*	.592	.766	1.03*	1.04*
iSpikformer	Convolutional	R^2↑	.817	.618	.440	.279	.977	.974	.972	.963
		RSE↓	.475	.668	.752	.905	.263	.284	.338	.348
	Delta	R^2↑	.804	.601	.434	.272	.972	.969	.960	.944
		RSE↓	.496	.666	.759	.910	.274	.302	.391	.455
	Repetition	R^2↑	.692	.548	.238	.021*	.962	.953	.849	.710*
		RSE↓	.573	.708	.847	1.04*	.289	.557	.705	1.03*

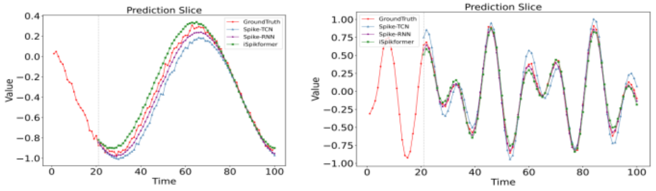


Energy Consumption and Analysis

- Estimation of Energy Consumption

Model	Param(M)	OPs (G)	Energy (mJ)	Energy Reduction	R^2
TCN	0.460	0.14	0.64	63.60% ↓	.973
Spike-TCN	0.461	0.15	0.23		.963
GRU	1.288	1.32	6.07	75.05% ↓	.972
Spike-GRU	1.289	1.63	1.51		.964
iTransformer	1.634	2.05	9.47	66.30% ↓	.977
iSpikformer	1.634	3.55	3.19		.974

- Low-frequency and High-frequency Temporal Analysis



THANKS!

Contact us:
czlv22@m.fudan.edu.cn
yansenwang@microsoft.com
zhengxq@fudan.edu.cn