Bridging The Gap Between ANNs And SNNs By Calibrating Offset Spikes

Zecheng Hao , Jianhao Ding, Tong Bu, Tiejun Huang, Zhaofei Yu

ICLR2023

Abstract

Spiking Neural Networks (SNNs) have attracted great attention due to their distinctive characteristics of low power consumption and temporal information processing. **ANN-SNN conversion**, as the most commonly used training method for applying SNNs, can ensure that converted SNNs achieve comparable performance to ANNs on large-scale datasets. However, the performance degrades severely under low quantities of time-steps, which hampers the practical applications of SNNs to neuromorphic chips. In this paper, instead of evaluating different conversion errors and then eliminating these errors, we define an **offset spike** to measure the degree of deviation between actual and desired SNN firing rates. We perform a detailed analysis of offset spike and note that the firing of one additional (or one less) spike is the main cause of conversion errors. Based on this, we propose an optimization strategy based on **shifting the initial membrane potential** and we theoretically prove the corresponding **optimal shifting distance for calibrating the spike**. In addition, we also note that our method has a unique **iterative property** that enables further reduction of conversion errors. The experimental results show that our proposed method achieves state-of-the-art performance on CIFAR-10, CIFAR-100, and ImageNet datasets. For example, we reach a top-1 accuracy of 67.12% on ImageNet when using 6 time-steps. To the best of our knowledge, this is the first time an ANN-SNN conversion has been shown to simultaneously achieve high accuracy and ultralow latency on complex datasets.

Offset Spike

According to the experimental result, we have found that the firing of one additional (or one less) spike is the main cause of conversion errors, which implies that we can eliminate errors after adjusting $\sum_{t=1}^{T} \boldsymbol{s}^{l}(t)$ with ± 1.

Definition 1. We define **OFFSET SPIKE** ψ^l of layer l as the difference between the desired total spike count C^l_{designed} and the actual spike count C^l_{actual} during the interval $[0, T]$, that is

$$\psi^l = C^l_{\text{designed}} - C^l_{\text{actual}} = \frac{\boldsymbol{a}^l T}{\theta^l} - \sum_{t=1}^{T} \boldsymbol{s}^l(t), \tag{8}$$

Judge the sign and value of Offset Spike

- The specific value of residual membrane potential $\boldsymbol{v}^l(T)$ can help us judge the sign and value of Offset Spike.

Theorem 3. *Supposing that an ANN with QCFS activation function (equation 7) is converted to an SNN with* $L = T, \lambda^l = \theta^l, \boldsymbol{v}^l(0) = \theta^l/2$*, and the inputs to the* l*-th layer of ANN and SNN are the same, that is,* $\boldsymbol{a}^{l-1} = \boldsymbol{\phi}^{l-1}(T)$*. Then for any* i*-th element of the* l*-th layer, we will have the following conclusions:*

If $\sum_{t=1}^{T} I_i^l(t) \in [-\theta^l/2, \theta^l T + \theta^l/2)$*, when* $v_i^l(T)/\theta^l \in [k, k+1)$*, we will have* $\psi_i^l = a_i^l T/\theta^l - \sum_{t=1}^{T} s_i^l(t) = k$*, where* $k \in \mathbb{Z}$*.*

In fact, even if the input current does not belong to the specific interval, from equation 7, we can derive that when $\sum_{t=1}^{T} \boldsymbol{I}^l(t) < -\theta^l/2, \boldsymbol{a}^l = 0$ and when $\sum_{t=1}^{T} \boldsymbol{I}^l(t) \geqslant \theta^l T + \theta^l/2, \boldsymbol{a}^l = \theta^l$, then we can also directly determine the $\boldsymbol{\psi}^l$ according to the value of $\boldsymbol{\phi}^l(T)$. After we have already acquired the value of $\boldsymbol{\psi}^l$, we will adopt our optimization method for $|\psi_i^l|$ times to eliminate the offset spike on i-th element neuron of the l-th layer.

Eliminate Conversion Error through shifting initial membrane potential

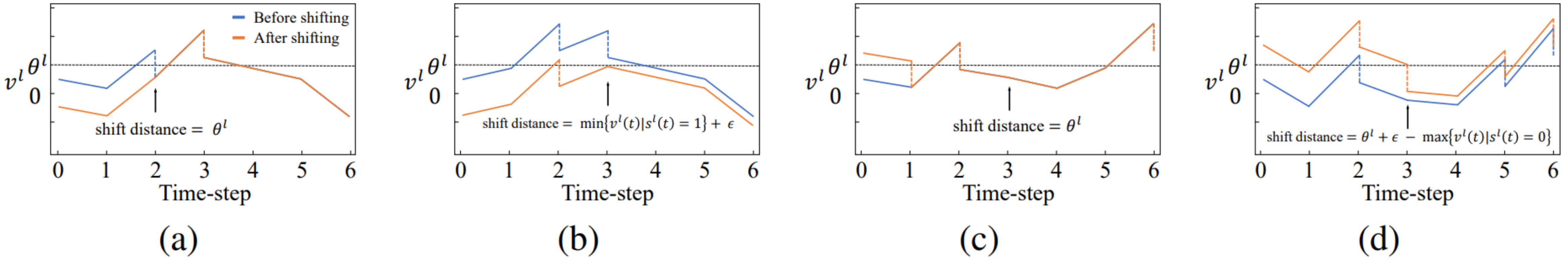


Figure 2: Shifting up (down) the initial membrane potential can increase (decrease) one output spike.

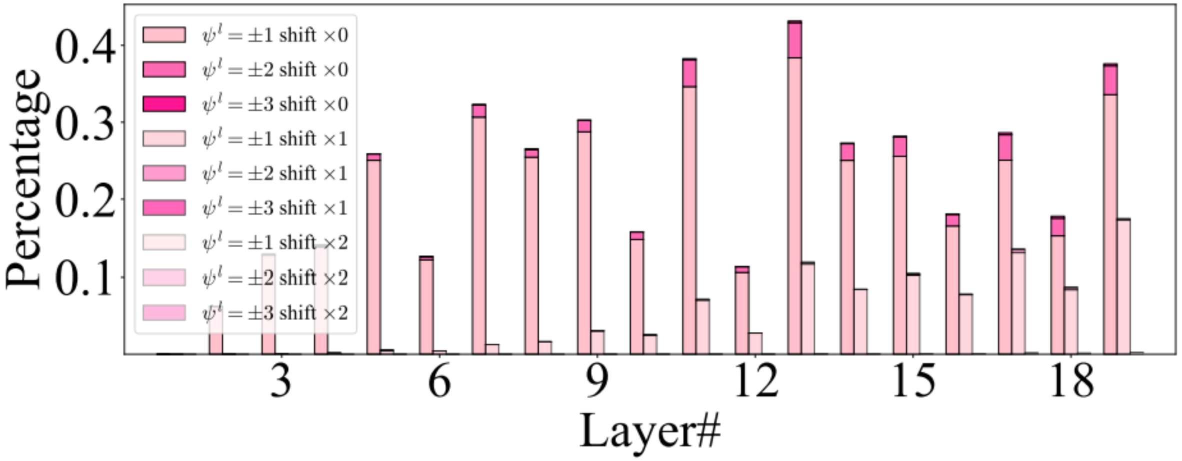
Theorem 2. *If we use $s_i^l(t)$ and $\widetilde{s}_i^l(t)$ to denote the binary spike of the i-th neuron in layer l at time-step t before and after optimization, $v_i^l(0)$ and $\widetilde{v}_i^l(0)$ to represent the initial membrane potential before and after optimization, then $\forall\epsilon\in(0,\theta^l)$, we will have the following conclusions:*

(i) If we set $\widetilde{v}_i^l(0) = v_i^l(0) - \max(\theta^l, \min\{v_i^l(t)|s_i^l(t)=1\}+\epsilon)$, *then* $\sum_{t=1}^{T}\widetilde{s}_i^l(t) = \sum_{t=1}^{T} s_i^l(t) - 1$.

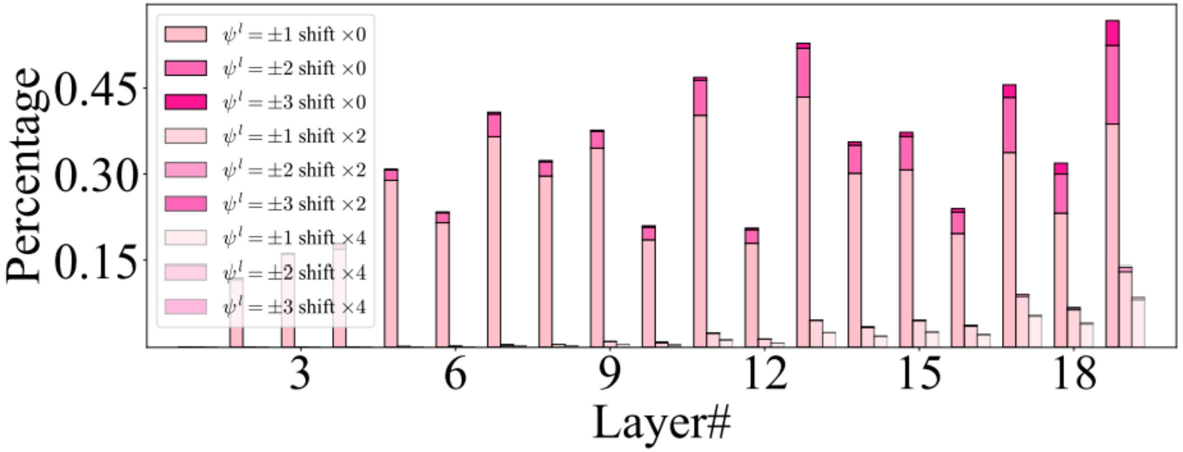
(ii) If we set $\widetilde{v}_i^l(0) = v_i^l(0) + \max(\theta^l, \theta^l+\epsilon-\max\{v_i^l(t)|s_i^l(t)=0\})$, *then* $\sum_{t=1}^{T}\widetilde{s}_i^l(t) = \sum_{t=1}^{T} s_i^l(t) + 1$.

Eliminate Conversion Error through shifting initial membrane potential

- Judge the sign and value of Offset Spike : according to $\boldsymbol{v}^l(T)$
- One can reuse the optimal shifting distance mentioned in Theorem 2 to increase (or decrease) one output spike each time.
- The performance of the converted SNN increases with the iteration.



(a) ResNet-20 on CIFAR-10



(b) ResNet-20 on CIFAR-100

NELVT

Experiments

Table 3: The ratio and MSE after multiple iterations

Dataset	Architecture	Baseline		**Ours**		**Ours** $\times 2$		**Ours** $\times 4$	
		Ratio	MSE	Ratio	MSE	Ratio	MSE	Ratio	MSE
CIFAR-10	VGG-16	88.33%	0.120	97.65%	0.024	99.83%	0.002	99.84%	0.002
	ResNet-20	62.38%	0.512	82.56%	0.179	99.73%	0.003	99.76%	0.002
CIFAR-100	VGG-16	82.90%	0.192	98.42%	0.016	99.86%	0.001	99.87%	0.001
	ResNet-20	41.59%	1.641	67.08%	0.453	86.03%	0.165	91.29%	0.101

Experiments

Method	ANN	Architecture	T=1	T=2	T=4	T=8	T=16	T=32
CIFAR-100 Dataset								
SNM	74.13%	VGG-16	-	-	-	-	-	71.80%
SNNC-AP	77.89%		-	-	-	-	-	73.55%
OPI	76.31%		-	-	-	60.49%	70.72%	74.82%
QCFS	76.28%		-	63.79%	69.62%	73.96%	76.24%	77.01%
Ours	76.28%		74.24%	76.03%	76.26%	76.52%	76.77%	76.96%
RMP	68.72%	ResNet-20	-	-	-	-	-	27.64%
OPI	70.43%		-	-	-	23.09%	52.34%	67.18%
QCFS	69.94%		-	19.96%	34.14%	55.37%	67.33%	69.82%
Ours	69.97%		59.22%	64.21%	65.18%	67.17%	69.44%	70.29%
ImageNet Dataset								
SNNC-AP	75.36%	VGG-16	-	-	-	-	-	63.64%
SNM	73.18%		-	-	-	-	-	64.78%
OPI	74.85%		-	-	-	6.25%	36.02%	64.70%
QCFS	74.29%		-	-	-	19.12%	50.97%	68.47%
Ours	74.19%		63.84%	70.59%	72.94%	73.82%	74.09%	74.33%
SNNC-AP	75.66%	ResNet-34	-	-	-	-	-	64.54%
QCFS	74.32%		-	-	-	35.06%	59.35%	69.37%
Ours	74.22%		69.11%	72.66%	73.81%	74.17%	74.14%	73.93%

Conclusion

In this paper, we first define offset spike to measure the degree of deviation between the actual and desired SNN firing rates. Then we analyze the distribution of offset spike and demonstrate that we can infer the specific value of the deviation according to the corresponding residual membrane potential. Furthermore, we propose an optimization method to eliminate offset spike by shifting the initial membrane potential up and down. Finally, we demonstrate the superiority of our method on CIFAR-10/100 and ImageNet datasets. Our results will further facilitate the relevant research and application of SNNs to neuromorphic chips.

Thanks for Listening!