Human-Centered Computer Systems Lab
Seoul National University

International Conference On Machine Learning

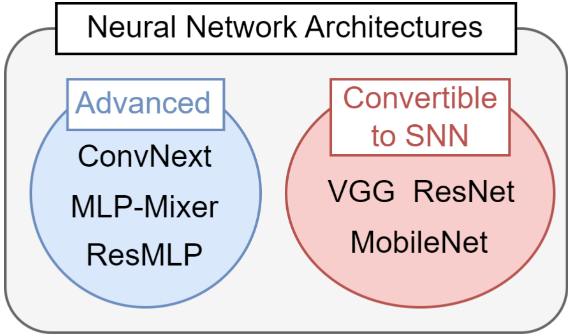
Sign Gradient Descent-based Neuronal Dynamics: ANN-to-SNN Conversion Beyond ReLU Network

Hyunseok Oh
Seoul National University
ohsai@snu.ac.kr

Youngki Lee
Seoul National University
youngkilee@snu.ac.kr

Motivation 1: Performance gap of SNN with SOTA DNNs

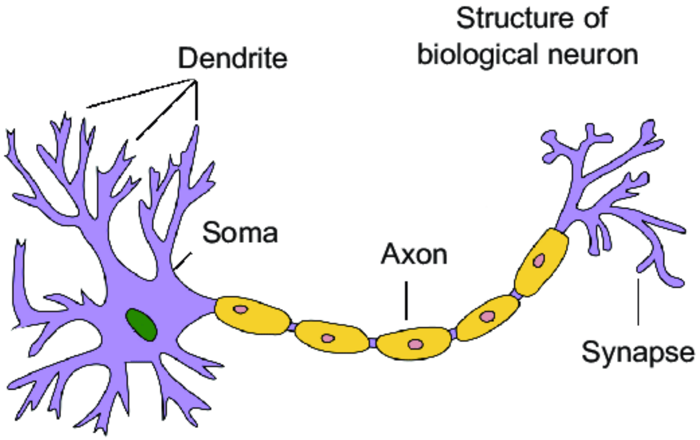
- **SNN** is significantly **energy-efficient** and **biologically plausible**, but it has a **noticeable performance gap** with **state-of-the-art DNNs**
  - Training SNN from scratch
    - Large performance gap with ANNs
  - ANN-to-SNN Conversion
    - Cannot convert nonlinear tensor operator besides ReLU

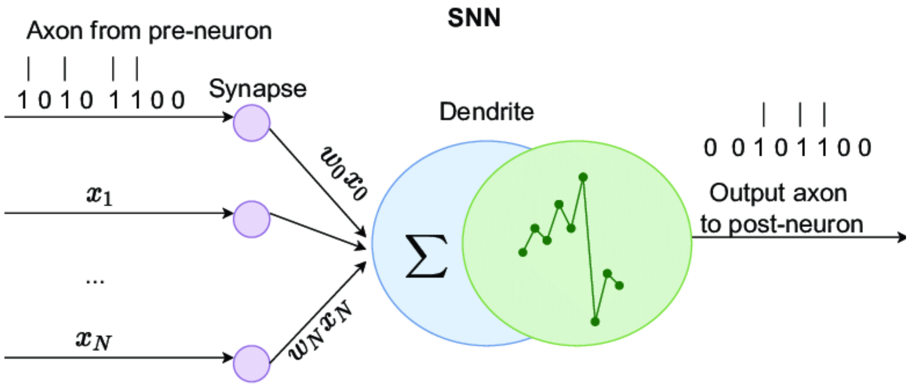


SNN with IF/LIF spiking neuron

Motivation 2: Understanding Behavior of Spiking Neurons

- Which **fundamental mathematical principle** underlies the behavior of spiking neuron, a simplified model of **biological neurons**?

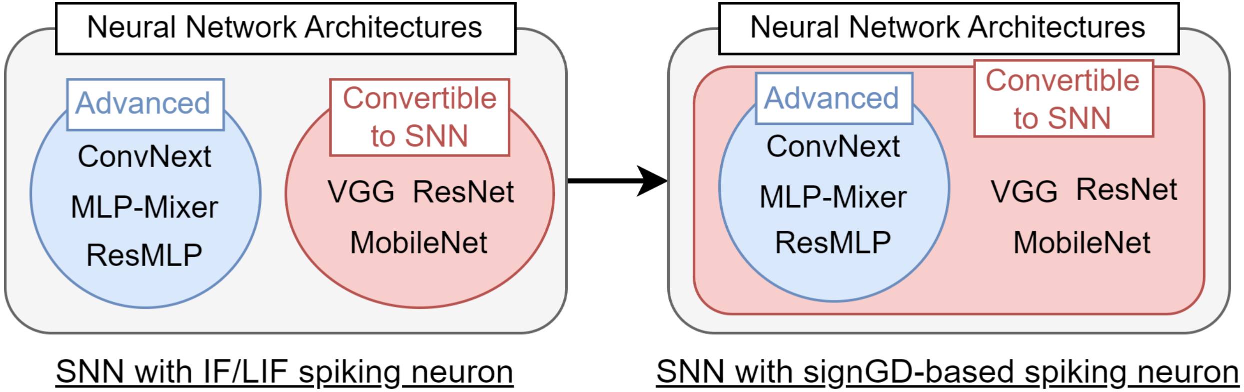




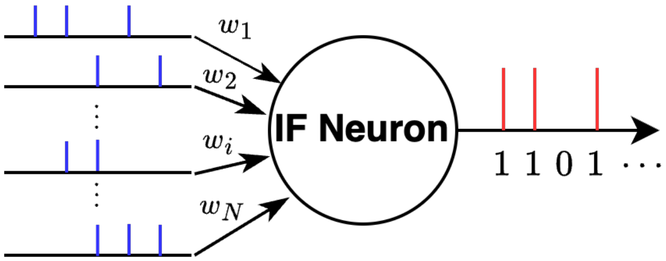
Yamazaki, Kashu & Vo, Khoa & Bulsara, Darshan & Le, Ngan. (2022). Spiking Neural Networks and Their Applications: A Review. Brain sciences. 12. 10.3390/brainsci12070863.

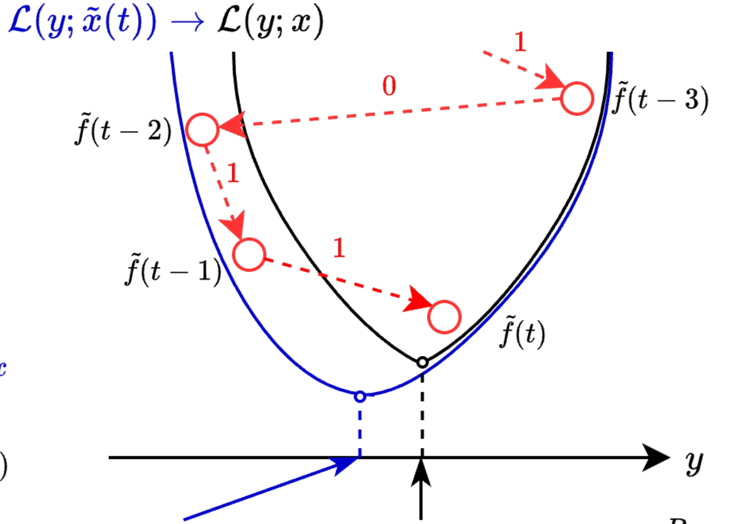
Key Contributions

- Mathematical Findings
  - Discrete Dynamics of Spiking Neurons (IF/LIF) are Sub-gradient Method
- Application
  - ANN-to-SNN Conversion Beyond ReLU Network



Theory: Discrete Dynamics of Simple Spiking Neuron Models are Sub-gradient Method





Input: $\tilde{x}(t) = \frac{t}{t+1}\tilde{x}(t-1) + \frac{1}{t}\sum_{i=1}^{N} I_i(t) \xrightarrow[t\to\infty]{} x$

Output: $y(t) = \frac{t}{t+1}y(t-1) + \frac{1}{t}s(t)$ (Rate coding)

$$\tilde{f}(t) = \frac{t}{t+1}y(t) - \frac{u(0)-\theta_{th}}{\theta_{th}(t+1)}$$

$$\arg\min_y \mathcal{L}(y; \tilde{x}(t)) \to \arg\min_y \mathcal{L}(y; x) = \mathrm{ReLU1}\left(\frac{R}{\theta_{th}}x\right)$$

Discrete Neuronal Dynamics of IF Neuron $\equiv$ Optimization Trajectory of Subgradient Method

Theory: Discrete Dynamics of Simple Spiking Neuron Models are Sub-gradient Method

Dynamics equivalence of IF neuron with sub-gradient method (Informal).

Dynamical system of **IF neuron** with rate-coded input $\tilde{x}(t)$ and output $y(t)$ is equivalent to the **sub-gradient method** with **diminishing step size** $\frac{1}{t+1}$ over an optimization problem $\min_{y\in\mathbb{R}} \mathcal{L}(y; x)$, approximated with $x \leftarrow \tilde{x}(t+1)$ as,

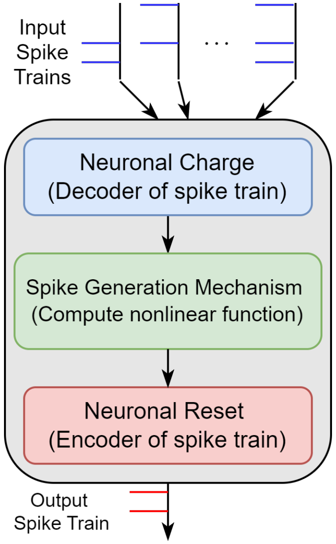
$$\tilde{f}(t) = \tilde{f}(t-1) - \frac{1}{t+1} \cdot \tilde{g}\left(\tilde{f}(t-1); \tilde{x}(t)\right) \qquad \text{(sub-gradient method)}$$

$$\mathcal{L}(y; x) = \mathrm{ReLU}\left(\frac{R}{\theta_{th}} x - y\right) + \frac{1}{2} y^2 \qquad \text{(objective function)}$$

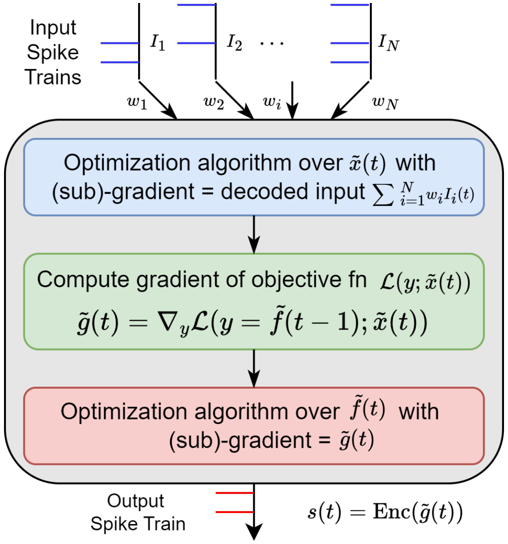
$$\tilde{f}(t) = \frac{t}{t+1} y(t) - \frac{u(0) - \theta_{th}}{\theta_{th}(t+1)} \qquad (t\text{-th approximation from y(t)})$$

where $\tilde{g}(y; x)$ is a sub-gradient of $\mathcal{L}(y; x)$. Minimizer of the problem is $\mathrm{ReLU1}(\frac{R}{\theta_{th}} x)$. (**LIF neuron** version is also on the paper)

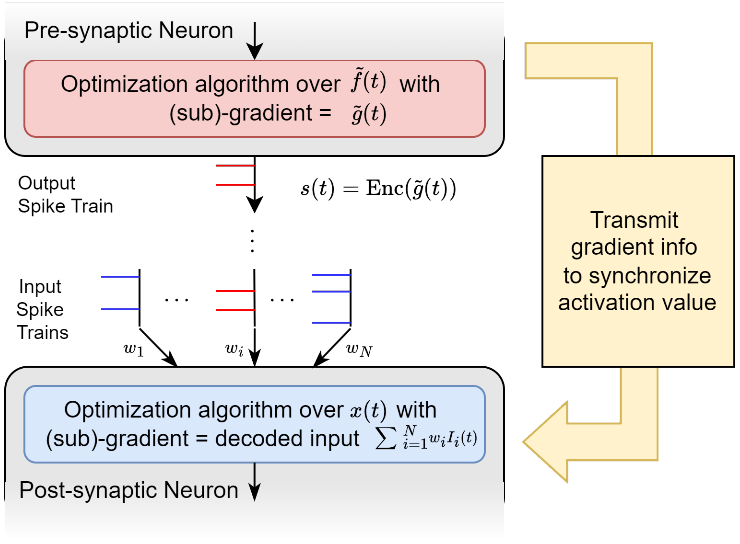
Theory: Discrete Dynamics of Simple Spiking Neuron Models are Sub-gradient Method



Sequential stages of neuronal dynamics



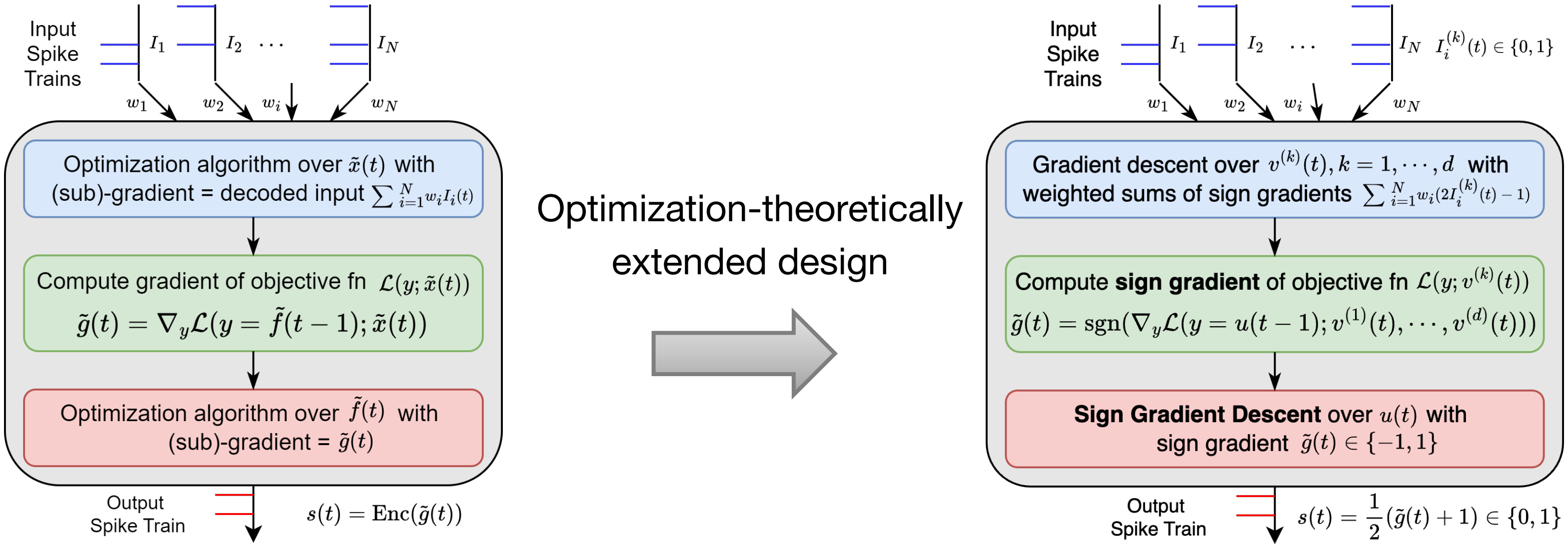
Optimization-theoretic interpretation of dynamics stages



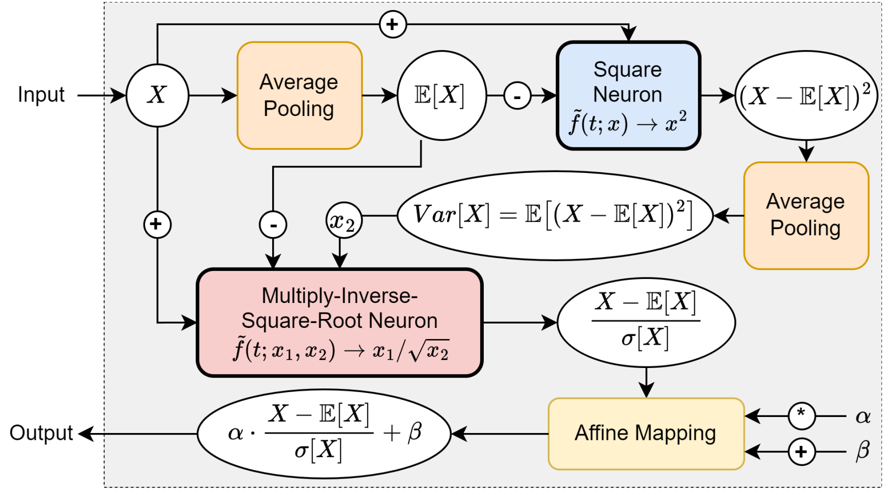
A spike train synchronizes an activation value between pre- and post-synaptic neuron

Application: ANN-to-SNN Conversion Beyond ReLU Network

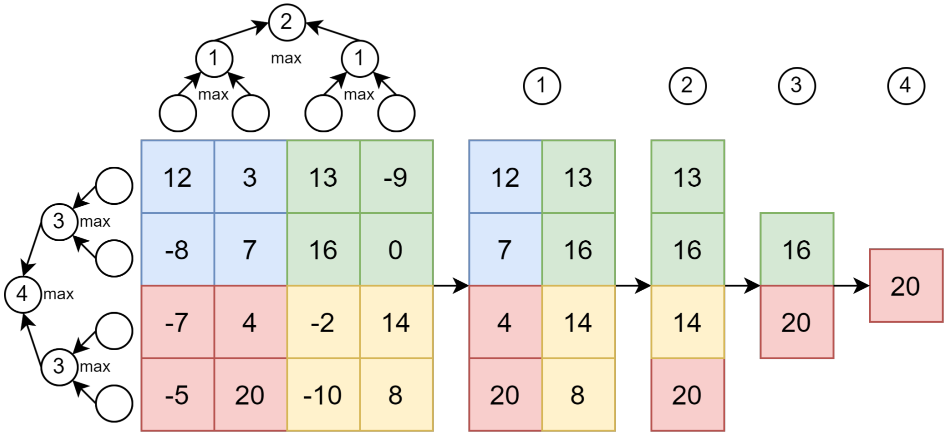
- **Sign gradient descent**(signGD)-based **neuronal dynamics**
  1. **signGD** instead of **sub-gradient** method to design neuronal dynamics.
  2. **Generalize LR schedule** of optimizer form of neuronal dynamics



Application: ANN-to-SNN Conversion Beyond ReLU Network



Layer Normalization computation
with signGD-based neurons



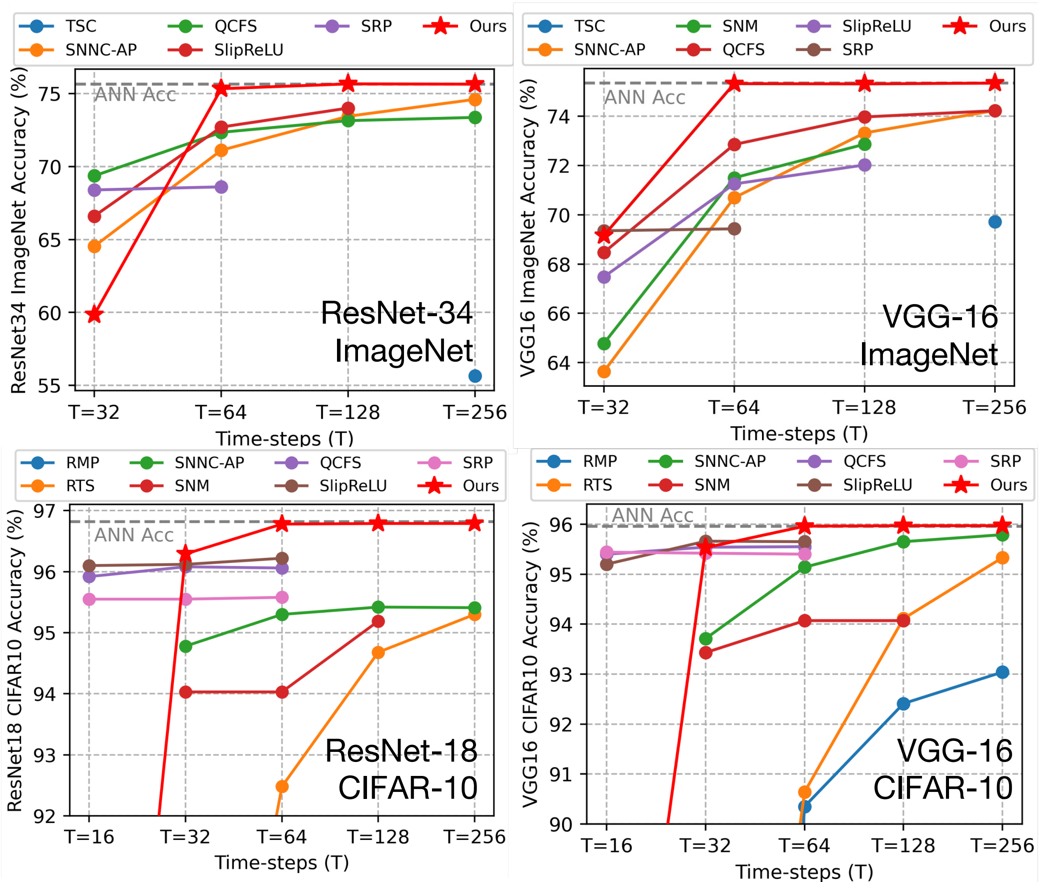
4x4 **Max pooling** with signGD-based neuron
of binary-input maximum operator

Evaluation: Converting Novel DNN Architectures

- ConvNext, MLP-Mixer (**LayerNorm + GELU** support)
- ResMLP (**GELU** support)
- VGG16, ResNet34 (**Exact Max Pooling** Support)

Converted DNN Models	ANN Acc.	SNN Simulation time-steps T = 32	T = 64	T = 128	T = 256
ConvNext-B	84.06	0.11	5.07	72.60	**81.07**
MLP-Mixer-B32	76.59	0.11	0.35	50.06	72.97
ResMLP-S24	80.76	**72.94**	**76.91**	**77.99**	78.04
RegNetX-3.2GF	81.19	26.85	**77.74**	**80.93**	**80.99**
VGG16 (MaxPool2D)	73.36	38.08	67.04	71.33	71.50
ResNet34 (MaxPool2D)	73.30	58.09	72.38	73.31	73.29

Evaluation: Performance Comparison with Prior Works



Summary

- We mathematically prove that **discrete neuronal dynamics of IF/LIF spiking neuron models** are equivalent to optimization dynamics of **sub-gradient method**
- We extend the theory to design **a new spiking neuron model** that can approximate **arbitrary nonlinear ops**.
- With our neuron, we expand **ANN-to-SNN conversion beyond ReLU networks**, e.g., ConvNext, MLP-Mixer.