北京大學
PEKING UNIVERSITY

International Conference On Machine Learning

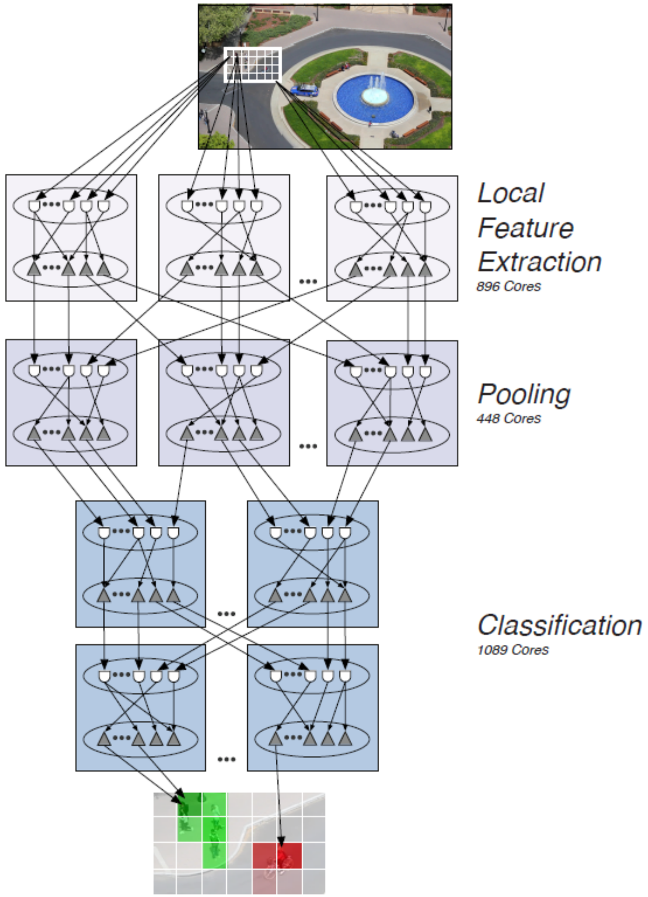
State Transition of Dendritic Spines Improves Learning of Sparse Spiking Neural Networks

Yanqi Chen [1 2] Zhaofei Yu [† 1 2 3] Wei Fang [1 2] Zhengyu Ma [† 2] Tiejun Huang [1 2 3] Yonghong Tian [† 1 2]

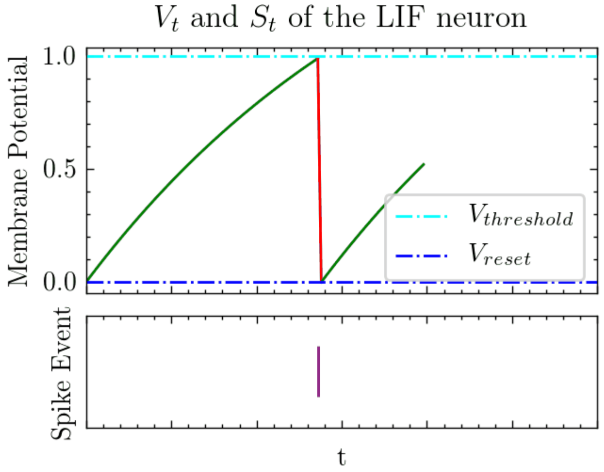
Speaker: Yanqi Chen

Introduction

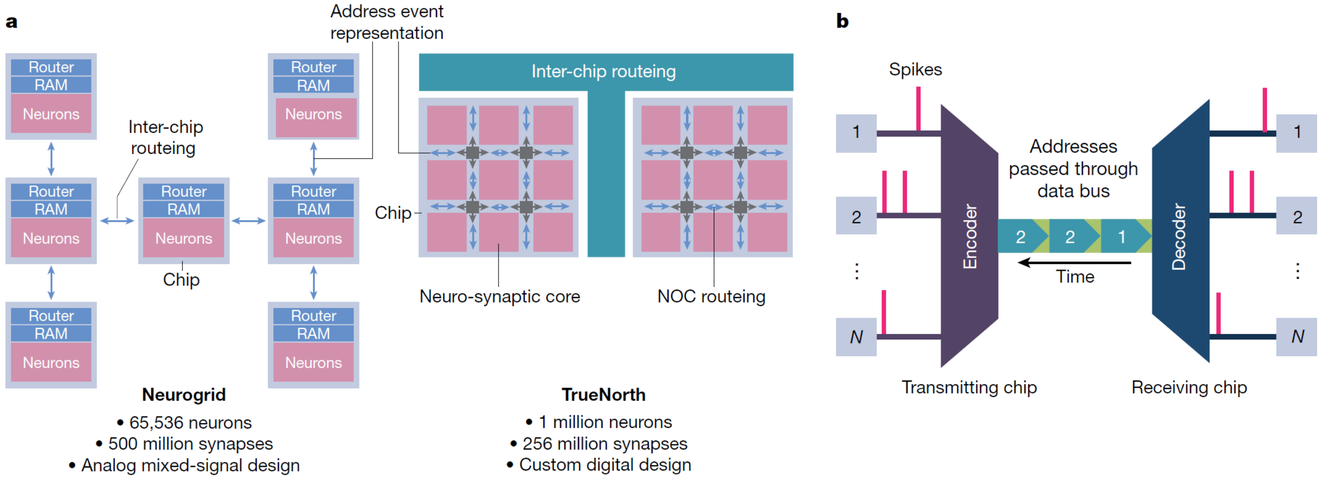
- **Spiking Neural Networks (SNNs)**



Borrowed from Merolla, Paul A., et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface." *Science* 345.6197 (2014): 668-673.



Dynamic of the **L**eaky **I**ntegrate-and-**F**ire Neurons (Constant input)



Borrowed from Roy, Kaushik, et al. "Towards spike-based machine intelligence with neuromorphic computing." *Nature* 575.7784 (2019): 607-617.

PEKING UNIVERSITY

ICML
International Conference On Machine Learning

Motivation

- Directly trained SNNs are going deeper

- Restricted number of synapses in a single neuromorphic chip

Motivation

- **Directly trained SNNs are going deeper**

2016	2018	2019	2020	2021
2 FC	2 Conv, 2 FC	5 Conv, 2 FC	ResNet-11	SEW ResNet-152
Lee, Jun Haeng, et al. *Frontiers in neuroscience* 10 (2016): 508.	Wu, Yujie, et al. *Frontiers in neuroscience* 12 (2018): 331.	Wu, Yujie, et al. *AAAI*. Vol. 33. No. 01. 2019.	Lee, Chankyu, et al. *Frontiers in neuroscience* 14 (2020): 119.	Fang, Wei, et al. *NeurIPS* 34 (2021): 21056-21069.

- **Restricted number of synapses in a single neuromorphic chip**

Motivation

- **Directly trained SNNs are going deeper**

2016 — 2 FC
Lee, Jun Haeng, et al. *Frontiers in neuroscience* 10 (2016): 508.

2018 — 2 Conv, 2 FC
Wu, Yujie, et al. *Frontiers in neuroscience* 12 (2018): 331.

2019 — 5 Conv, 2 FC
Wu, Yujie, et al. *AAAI*. Vol. 33. No. 01. 2019.

2020 — ResNet-11
Lee, Chankyu, et al. *Frontiers in neuroscience* 14 (2020): 119.

2021 — SEW ResNet-152
Fang, Wei, et al. *NeurIPS* 34 (2021): 21056-21069.

- **Restricted number of synapses in a single neuromorphic chip**

	Max # Neuron	Max # Synapse
SpiNNaker	1.6×10^4	1.6×10^7
DYNAPs	10^3	6.4×10^4
TrueNorth	10^6	2.56×10^8
Loihi	1.3×10^5	1.3×10^8
BrainScaleS	512	14080
Tianjic	4×10^4	10^7
Darwin 2	1.5×10^4	$>10^7$

ICML International Conference On Machine Learning

Contributions

- **Bio-inspired pruning algorithm for deep SNNs**

I. Inspired by **filopodial model** of **spinogenesis**

II. Theoretical proof of convergence

III. SOTA low performance loss on ImageNet and deep SNN

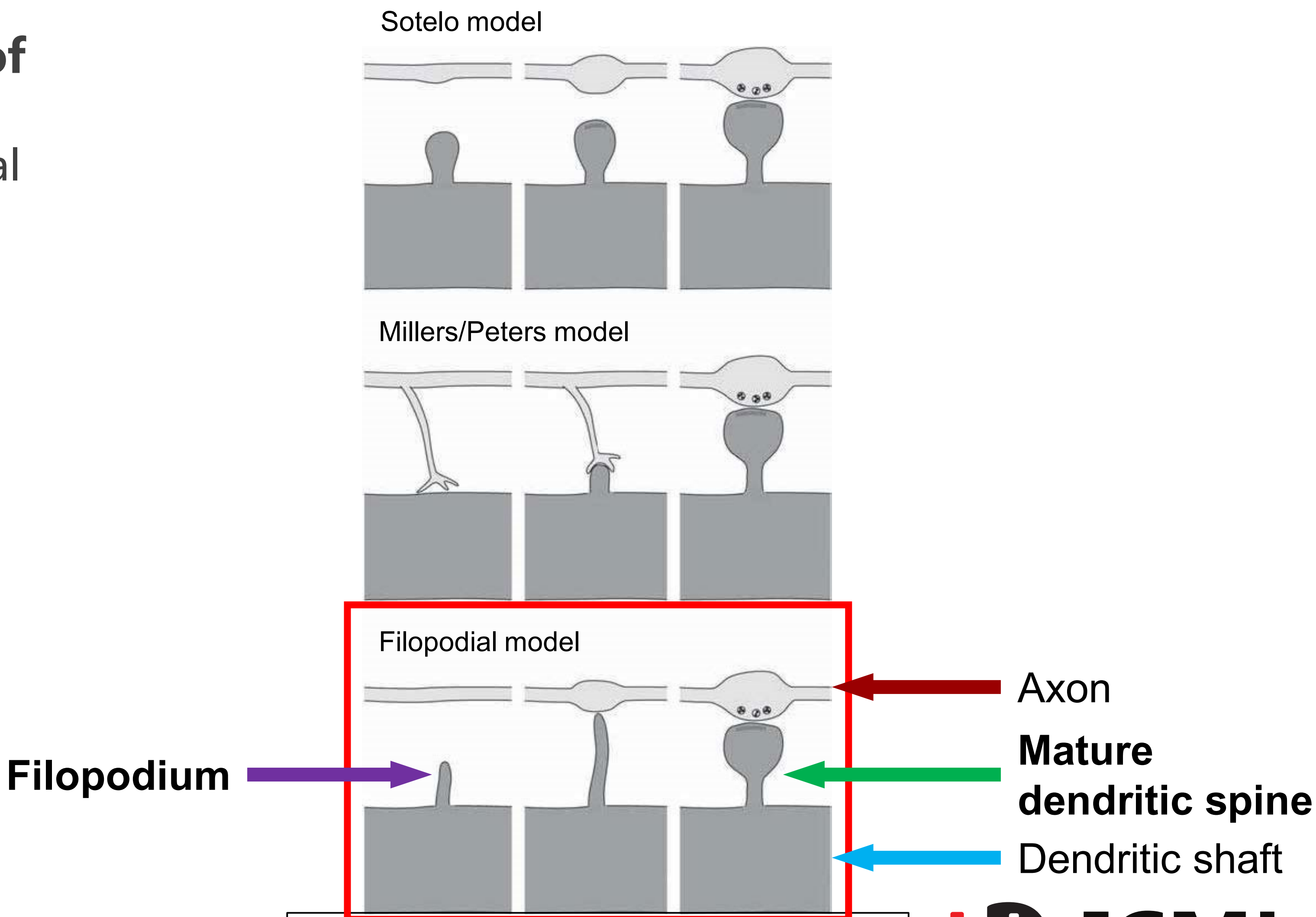
IV. In-depth discussion of pruning settings

Filopodial Model

- **Spinogenesis—The emergence of dendritic spines**
  - Connect dendrite to axon terminal
  - Form synapse between neurons
  - Changes of spine size and shape
- **Synaptic weight & Spine size**
  - Positive correlation

Dendritic Filopodia

1. **Precursors of dendritic spines**
2. **Lack clear heads (w.r.t. mature spines)**
3. **No synaptic contact**



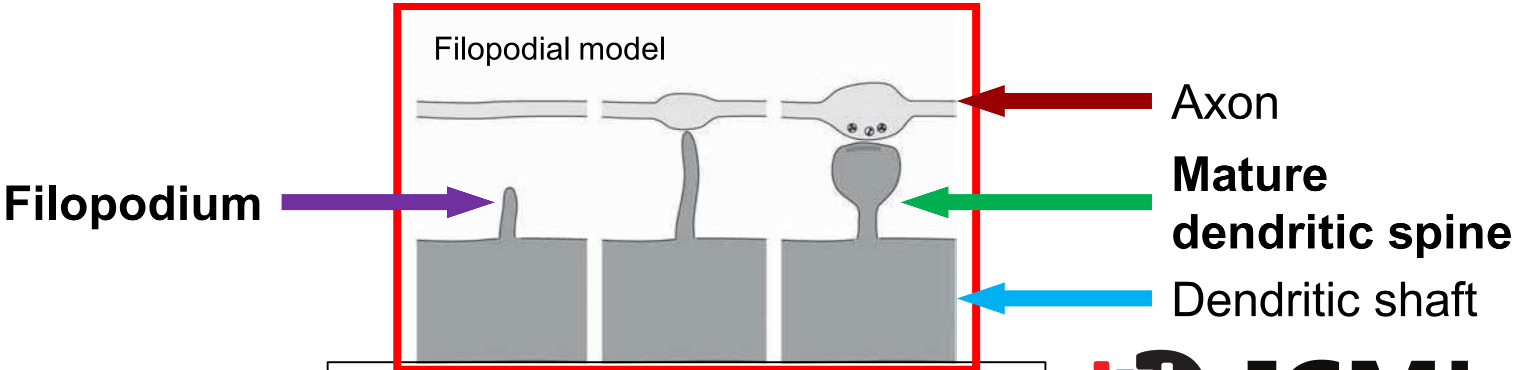
Three models for spinogenesis
Borrowed from Yuste, Rafael. Dendritic spines. MIT press, 2010.

Filopodial Model

- **Spinogenesis—The emergence of dendritic spines**
  - Connect dendrite to axon terminal
  - Form synapse between neurons
  - Changes of spine size and shape
- **Synaptic weight & Spine size**
  - Positive correlation

Dendritic Filopodia

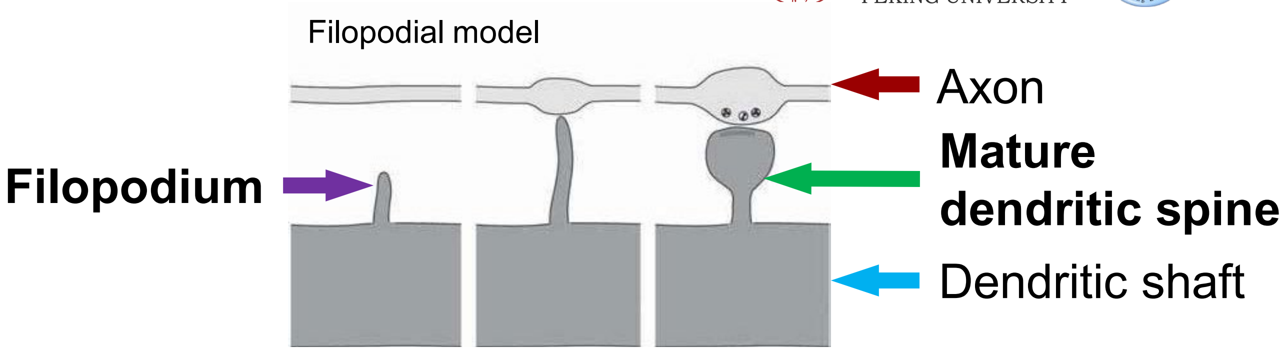
1. **Precursors of dendritic spines**
2. **Lack clear heads (w.r.t. mature spines)**
3. **No synaptic contact**

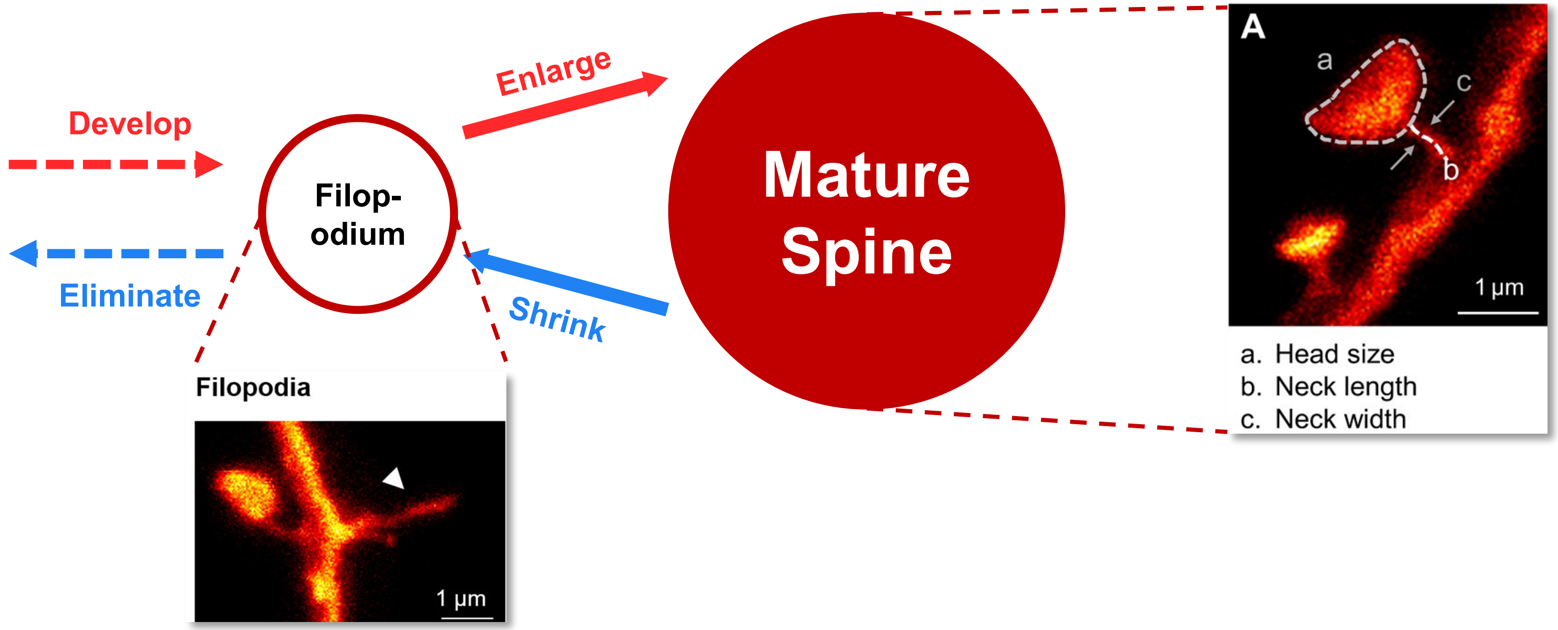


Three models for spinogenesis
Borrowed from Yuste, Rafael. Dendritic spines. MIT press, 2010.

Filopodial Model

- **Pruning model**
  - Transition from mature spine → filopodium
- **Weight learning & Structure learning**





Spine and filopodia images from Steffens, Heinz, et al. "Stable but not rigid: Chronic in vivo STED nanoscopy reveals extensive remodeling of spines, indicating multiple drivers of plasticity." *Science Advances* 7.24 (2021): eabf2806.

Reparameterization of Weights

- **Model transitions of spines in two folds**

Filopodium ⟷ Mature Spine

$$\boxed{w} = \mathrm{sign}\boxed{(\theta)} \cdot (|\theta| - d)_+$$

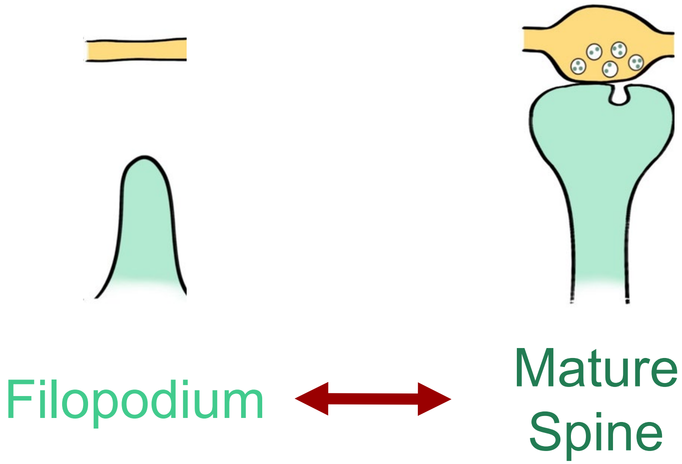
Synaptic Weight

Spine Size

Inhibitory ⟷ Excitatory

Reparameterization of Weights

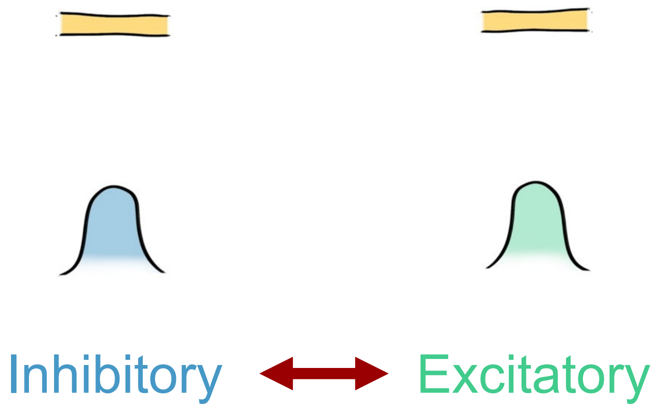
- **Model transitions of spines in two folds**

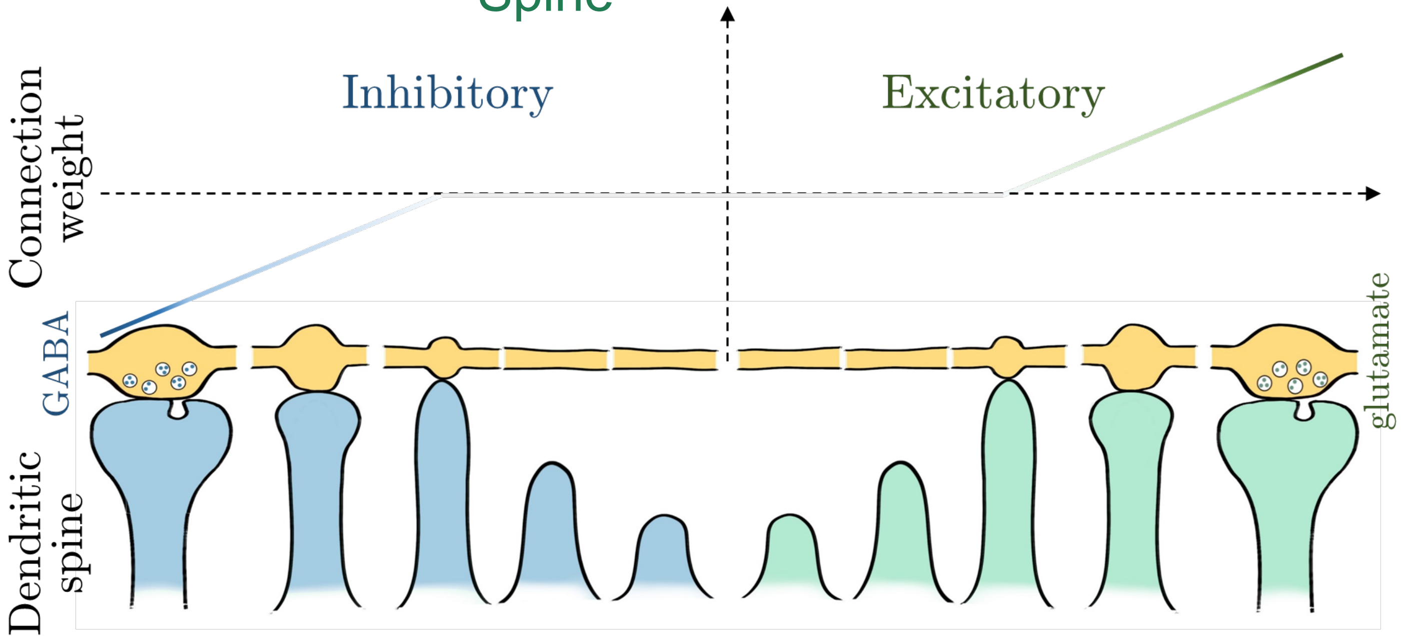


$$\boxed{w} = \text{sign}\boxed{(\theta)} \cdot (|\theta| - d)_+$$

Synaptic Weight

Spine Size





Reparameterization of Weights

北京大学 PEKING UNIVERSITY

- **Model transitions of spines in two folds**

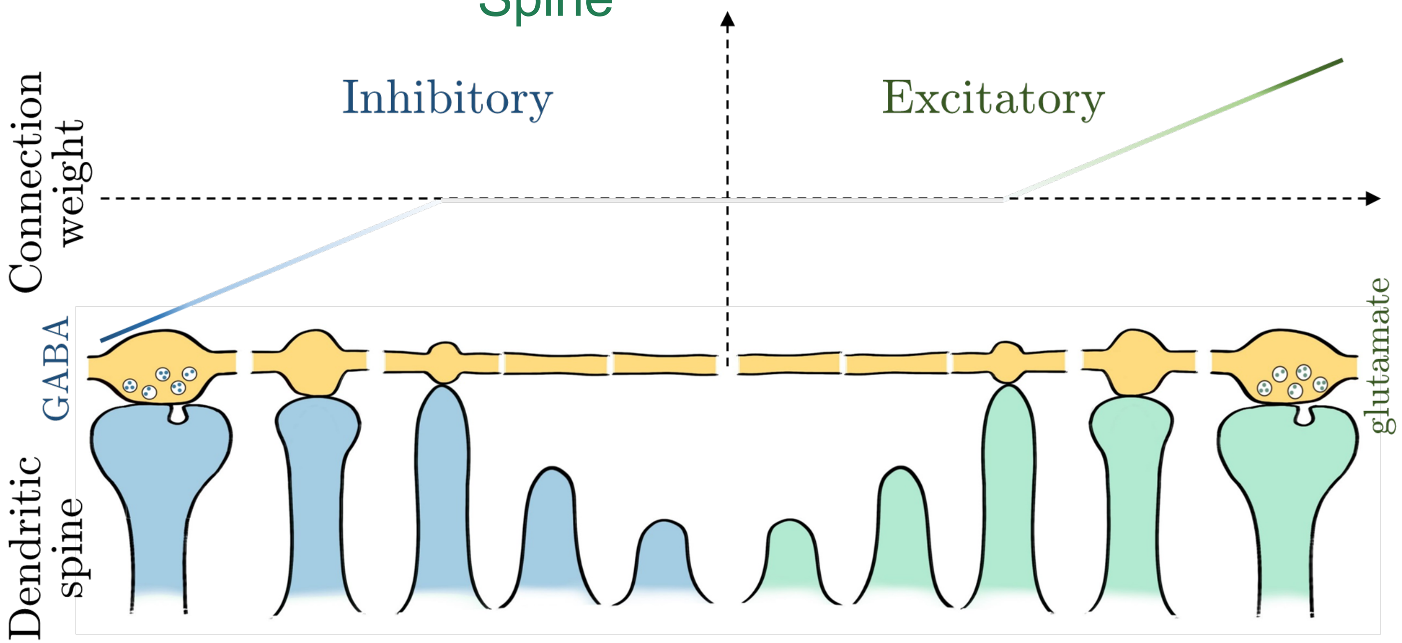
$$\boxed{w} = \mathrm{sign}\boxed{(\theta)} \cdot (|\theta| - d)_+$$

Synaptic Weight

Spine Size

Filopodium ↔ Mature Spine

Inhibitory ↔ Excitatory



Theorem 4.1 (Convergence). *For a spiking neural network, where each synaptic weight w is dominated by corresponding spine size θ through a soft threshold mapping*

$$w = \mathrm{sign}(\theta) \cdot (|\theta| - d)_+, d \geq 0, \tag{18}$$

if we apply a smooth approximation

$$w = f(\theta) := \frac{1}{\alpha} \log\left(\frac{1 + e^{\alpha(\theta - d)}}{1 + e^{-\alpha(\theta + d)}}\right), \alpha \gg 1, \tag{19}$$

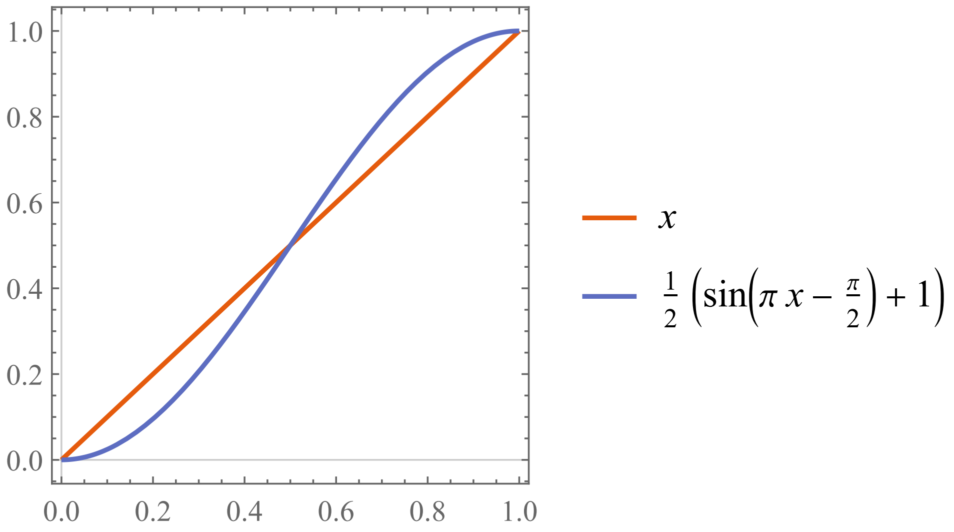
and define the pseudo partial derivative during computing gradients as $\frac{\partial w}{\partial \theta}_p \equiv 1$, the loss function $\mathcal{L}$ is L-smooth and lower bounded, the sequence $\{\mathcal{L}(\theta^t)\}_{t \in \mathbb{N}}$ must converge if learning rate $\eta < \frac{4}{L(1+e^{\alpha d})}$.

Threshold Scheduler

- **Increasing threshold of soft threshold function over training process**

$$d^t = D \cdot f(t/T)$$

  - Choice of scheduler function f

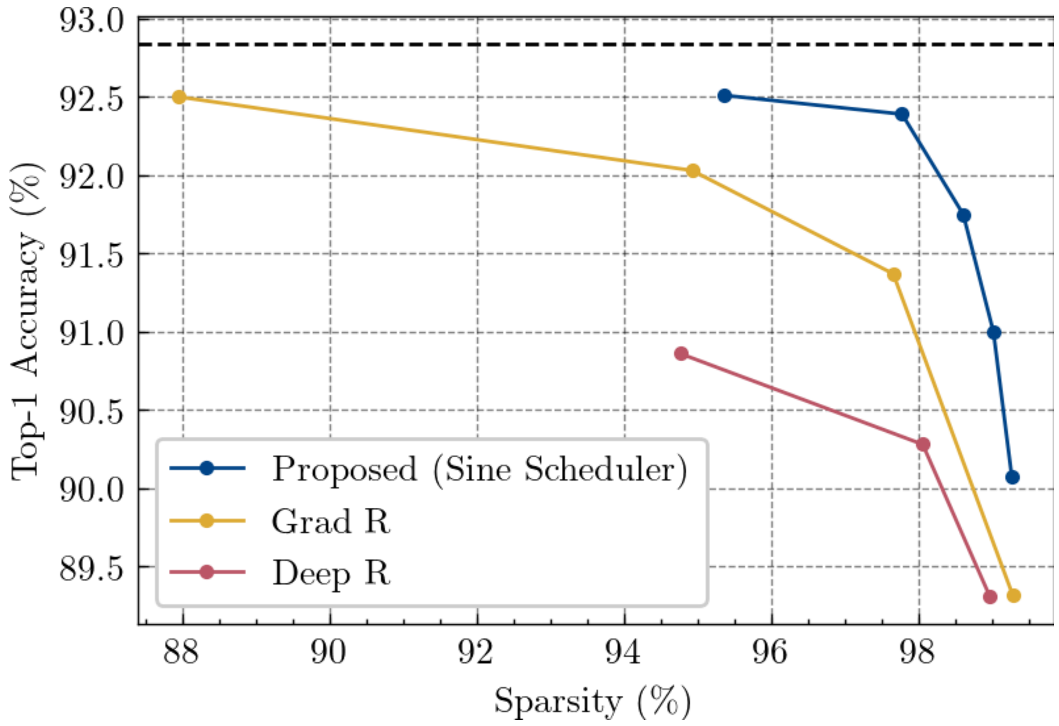


  - Act as proximal gradient descent under L1 regularized loss

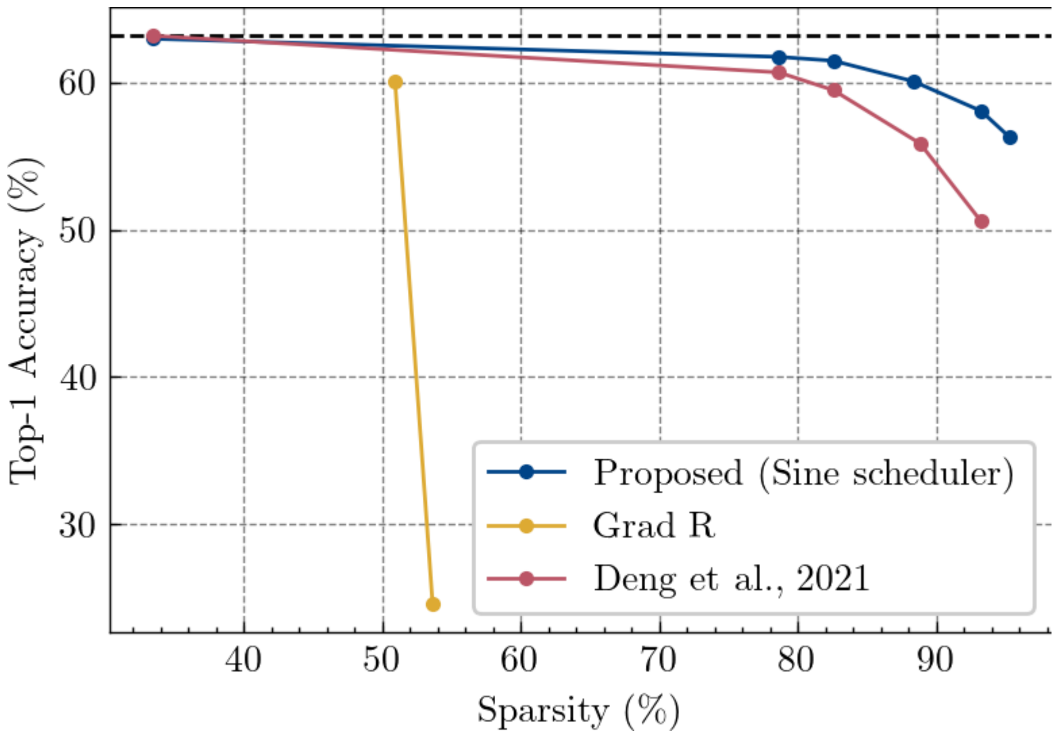
$$\boldsymbol{w}^t = \underset{\boldsymbol{w}}{\operatorname{argmin}} \left\{ \frac{1}{2\eta} \left\| \boldsymbol{w} - (\boldsymbol{w}^{t-1} \eta \nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}^{t-1})) \right\|_2^2 + \Delta d^t \|\boldsymbol{w}\|_1 \right\}$$

Results

- **Acc. vs Sparsity**



CIFAR-10



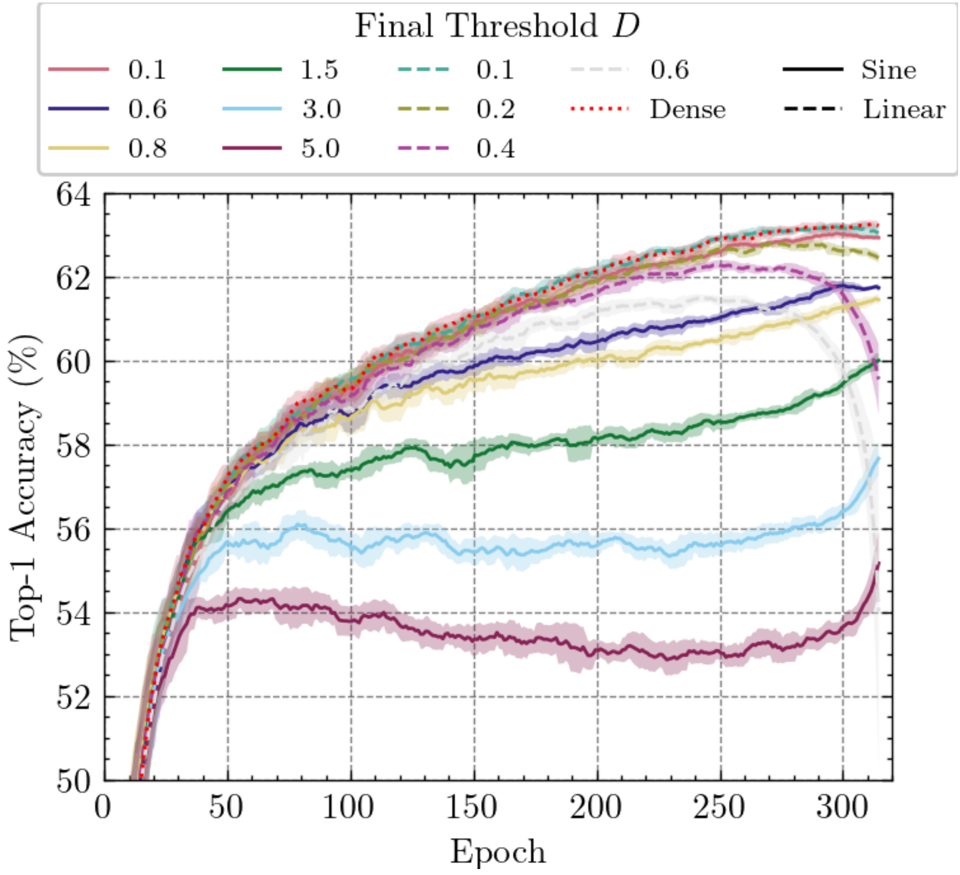
ImageNet

PEKING UNIVERSITY

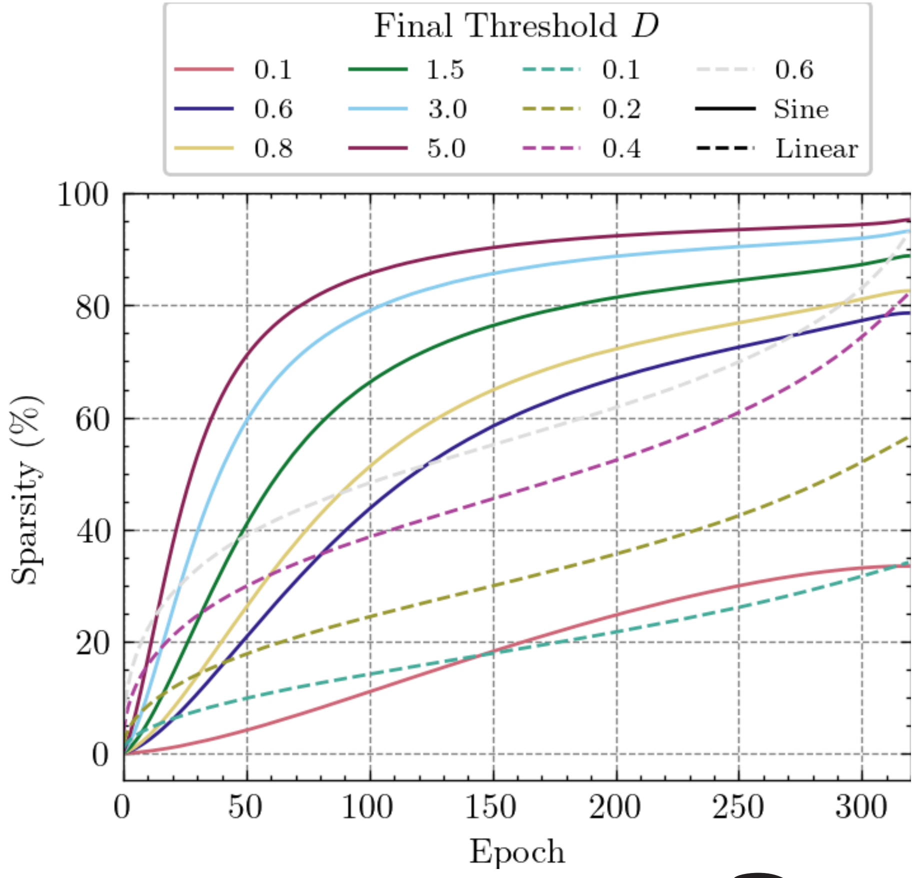
ICML
International Conference On Machine Learning

Results

- **Final threshold D control the sparsity**



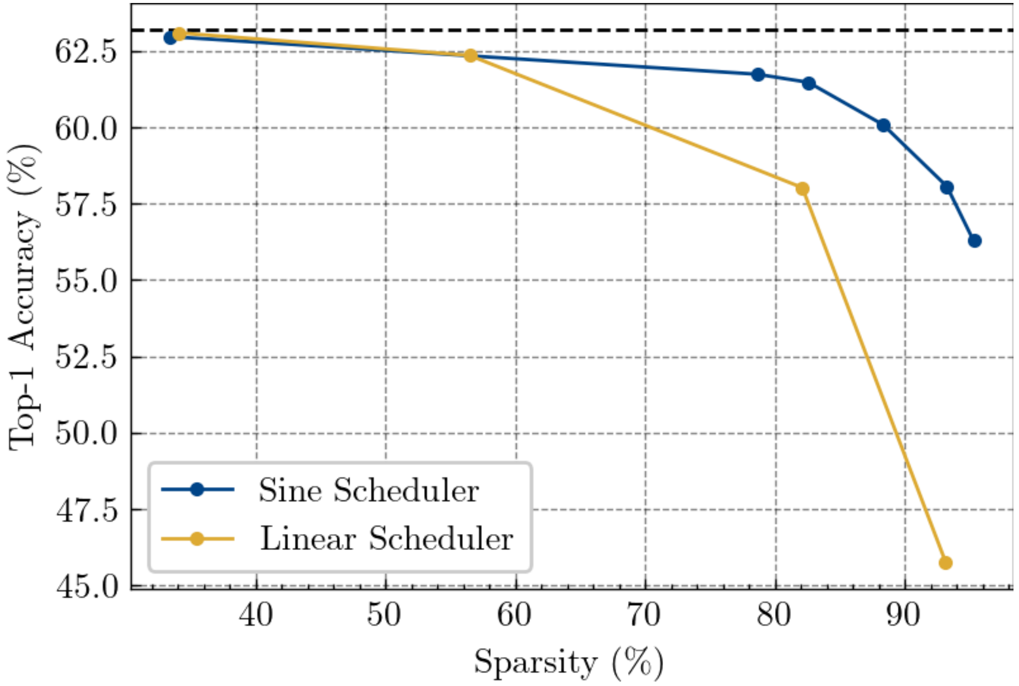
- **Different schedulers has different behaviors**



ICML
International Conference On Machine Learning

Results

- **Choice of schedulers matters**
  - Sine > Linear



  - Sine has milder penalty at the end of training stage

Summary & Discussion

- **A pruning method for really deep SNNs**
  - Gradient-based method — coupled with any directly trained SNNs
  - Effcient — Low accuracy loss under high sparsity

- **Based on elaborate model of synapses**
  - Bio-plausibility
  - May shed light on more meticulously designed models

- **Future work**
  - Pursuing the optimal threshold schedulers
  - Adding more biological ingredients (More Yummy!)