



# Training Deep Normalization-Free Spiking Neural Networks with Lateral Inhibition

Peiyu Liu, Jianhao Ding, Zhaofei Yu

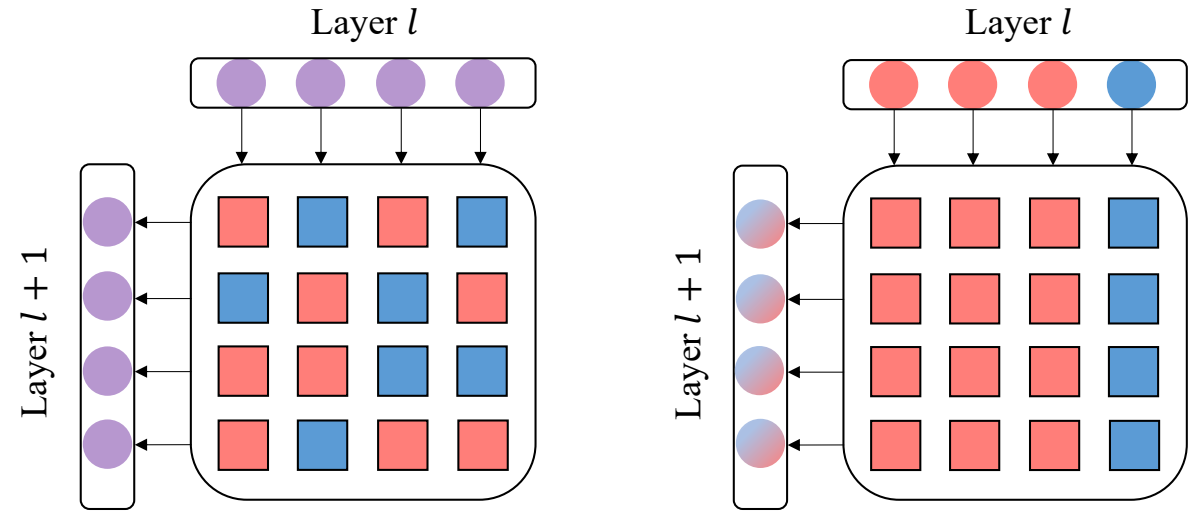
Presenter: Peiyu Liu

[lpy512120@stu.pku.edu.cn](mailto:lpy512120@stu.pku.edu.cn)

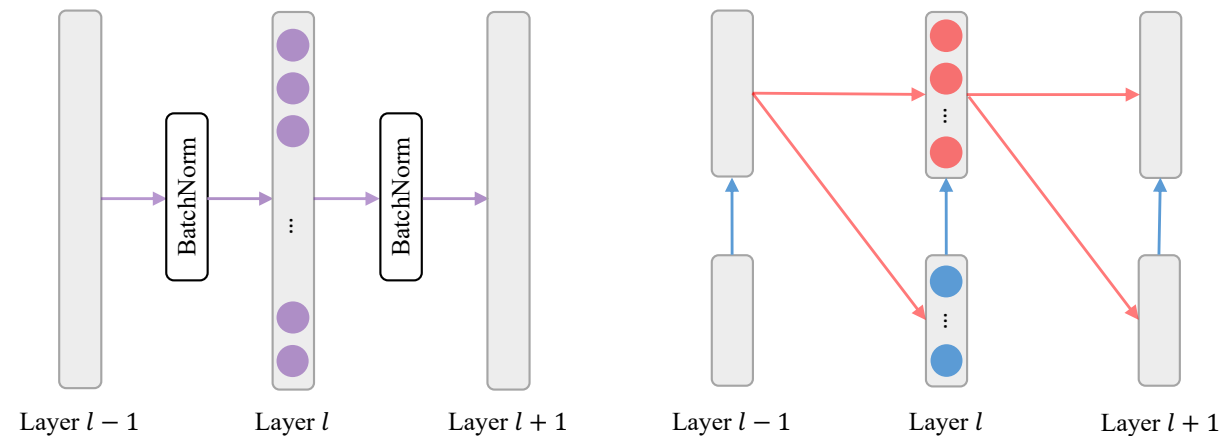
[Project Page](#)

# Motivation: Exploring Cortical Circuits in *Large-Scale* Neural Networks

- E-I interaction: fundamental building blocks of cortical computation.
  - *in-vivo* interaction between excitatory and inhibitory neurons.
  - Gain control (our focus), neural oscillation, selective attention, etc.
- Challenges
  - *Neuroscience*: limited *in-vivo* recording techs.
  - *Deep Learning*: archs and techs not compatible with E-I constraints.
    - Initialization strategy (e.g., Xavier, Kaiming).
    - Normalization (e.g., BatchNorm).

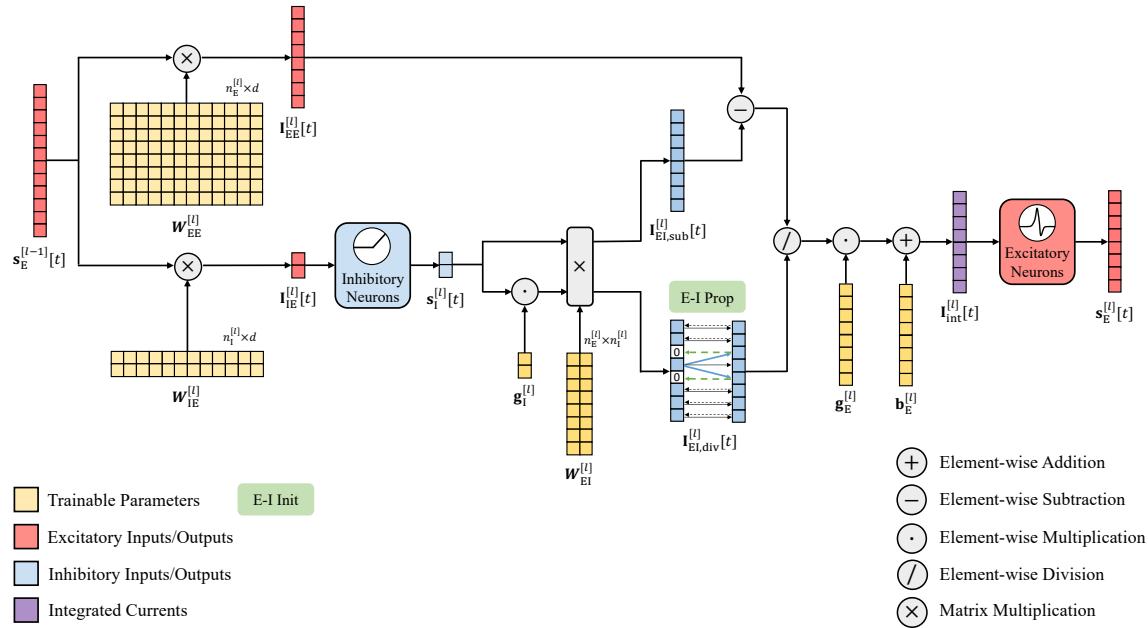


Sign constraint on synaptic weights



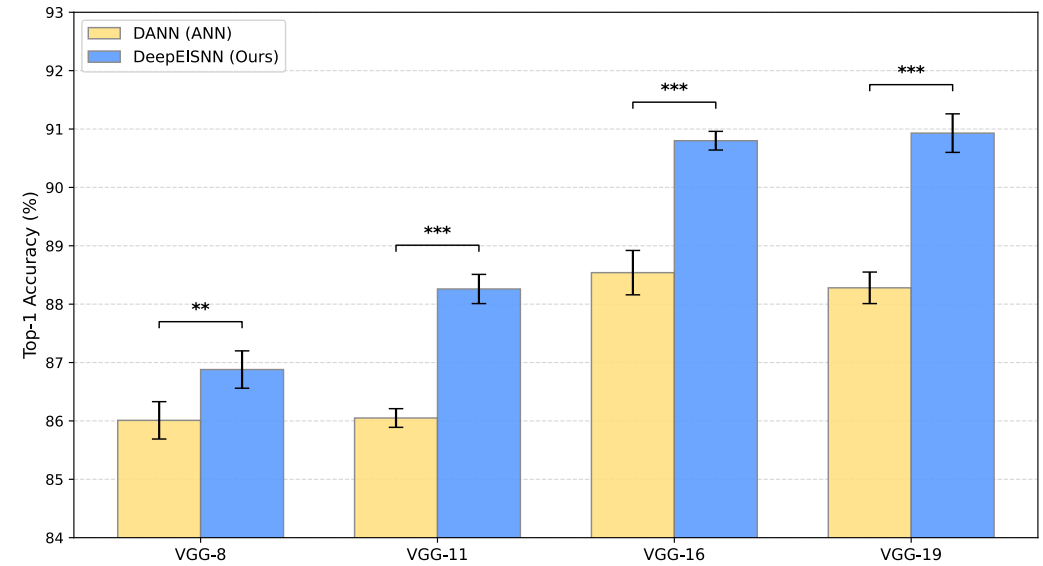
Architectural constraint on connectivity

# Contribution



- Brain-inspired architecture
  - E-I interaction via lateral inhibition.
  - Heterogeneous neuron modelling.
- Normalization-free learning framework
  - *E-I Init*: normalization only at initialization, not training.
  - *E-I Prop*: handling forward and backward instability.

Comparison with DANN with VGG-8 on CIFAR-10

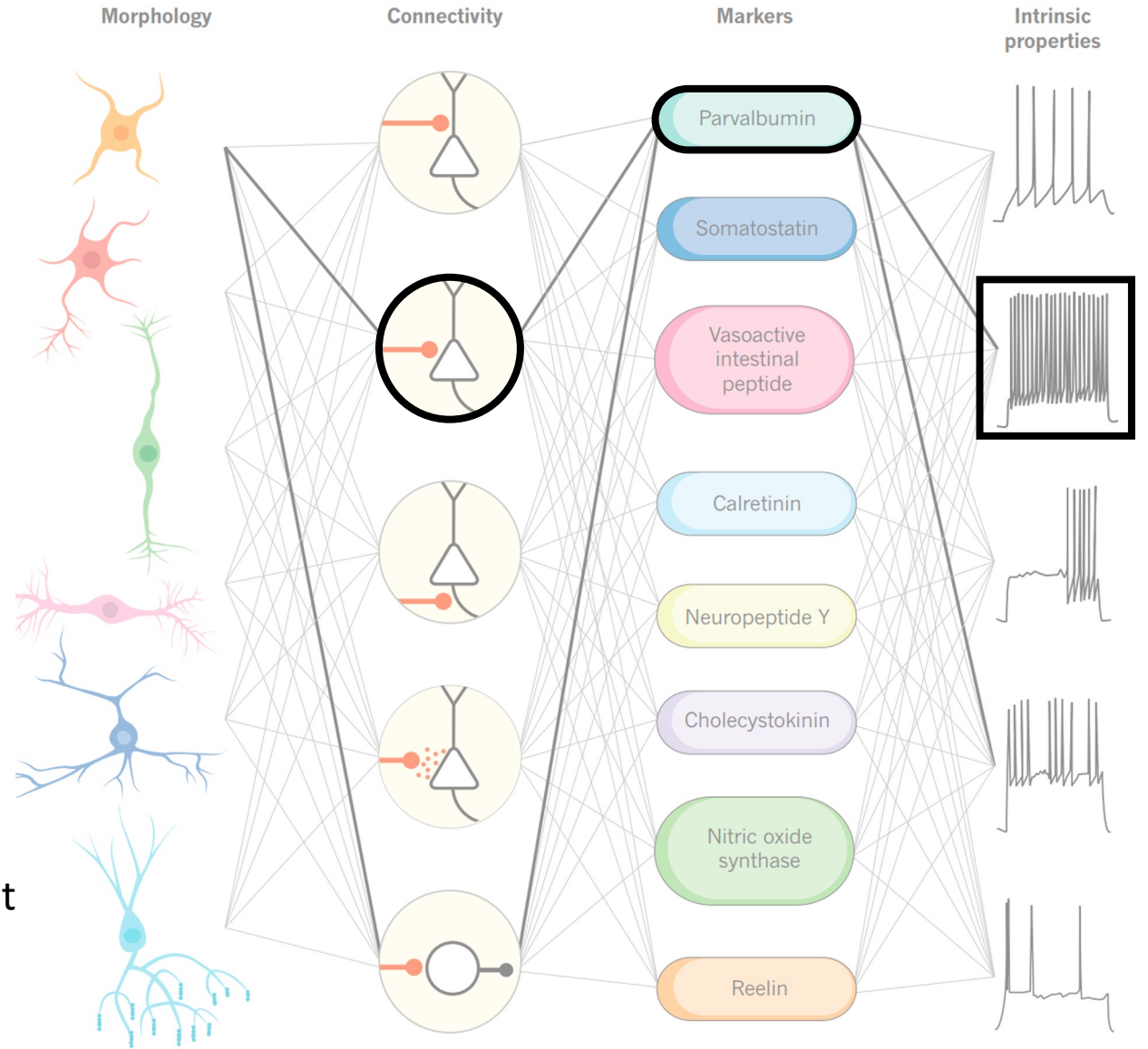


- Scalability & competitive performance
  - SOTA performance among models with E-I constraints.
  - Robust performance across multiple datasets.

# Key Design Choices

## *Heterogeneous neuron modelling*

- Excitatory neurons
  - Regular spiking
  - Feedforward activation
  - Modelled by LIF models
- Parvalbumin<sup>+</sup> inhibitory neurons
  - Prominent in cortex, and fast spiking
  - Lateral inhibition
    - **Synaptic inhibition** for *E-I balance*
    - **Somatic inhibition** for *gain control*
    - Putting together helps achieve norm. effect at initialization (*E-I Init*).
  - Modelled by ReLU with gain factors



Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature*, 2014.

# Key Design Choices

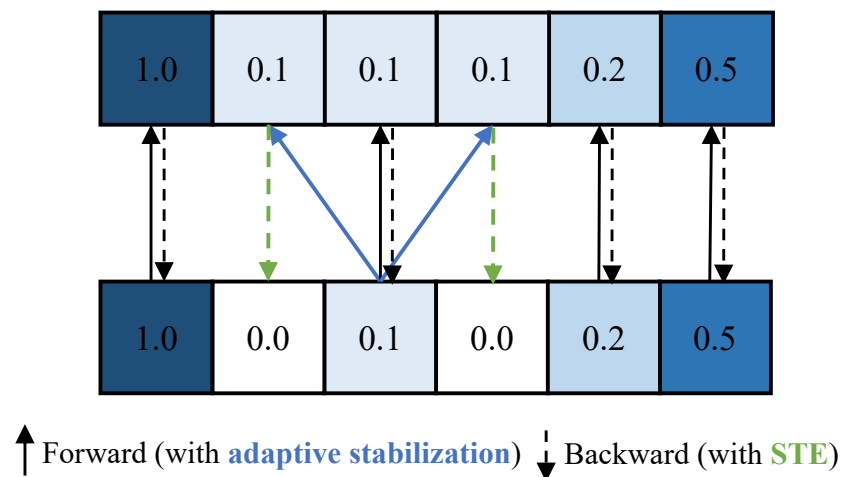
*E-I Prop: Overcoming forward and backward instability via decoupling*

- Forward instability from divide-by-zero error.
  - Convention is adding a small constant  $\epsilon$  to the denominator.
  - However, it turns out non-trivial to find appropriate  $\epsilon$ .
  - This leads to *adaptive stabilization*.
- Backward instability from *adaptive stabilization*.
  - Gradient sink.
  - Use STE to redirect gradient flow.

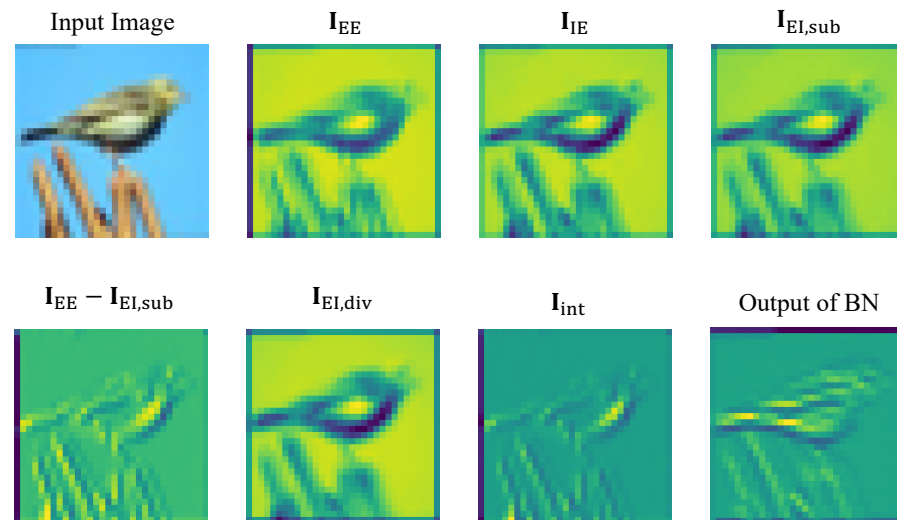
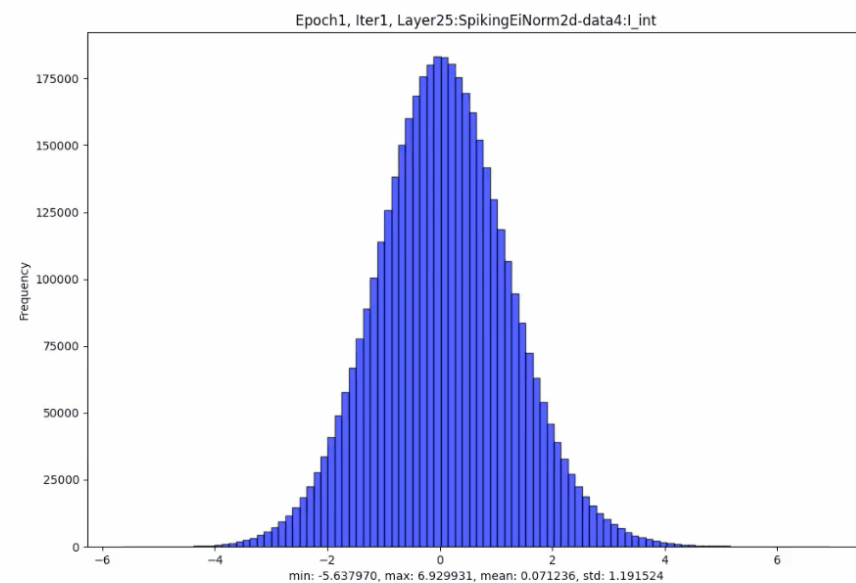
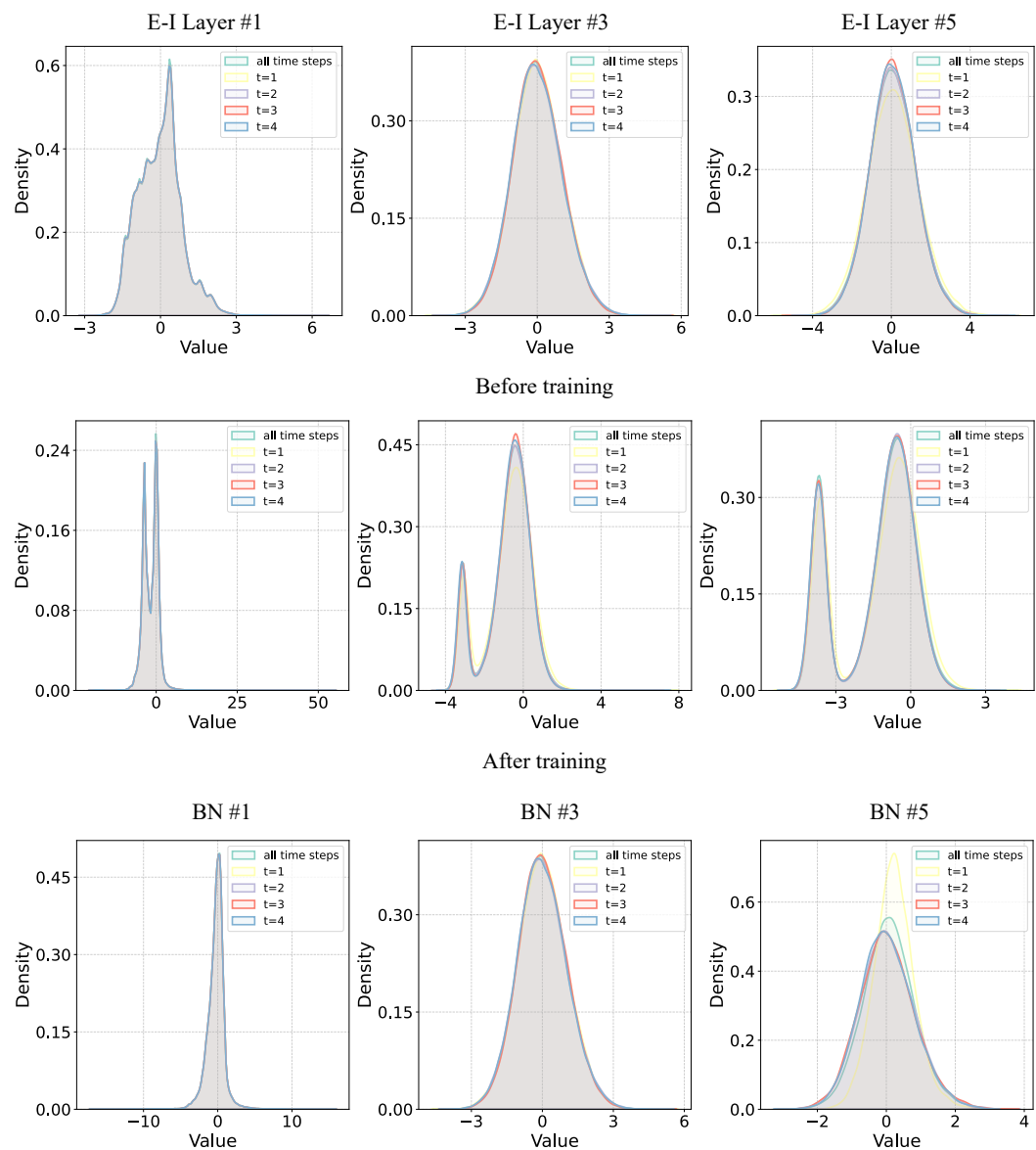
---

Fixed $\epsilon = 10^{-8}$	Failed to converge
Fixed $\epsilon = 10^{-7}$	Failed to converge
Fixed $\epsilon = 10^{-6}$	Collapsed (Epoch 22)
Fixed $\epsilon = 10^{-5}$	Collapsed (Epoch 5)

---



# Visualization of Learned Representation



# Conclusion

## For neuroscience

- Healthy network activity under E-I interaction emerges from synergy of
  - structural constraints (lateral inhibition), and
  - intrinsic neuronal dynamics (heterogeneous E/I neurons).

## For deep learning

- Biological constraints are not merely limitations, but also solutions.
  - E-I constraints + E-I interaction -> normalization-free training
- E-I interaction yields novel representations.
- Biological E-I interaction may inspire more developments in deep learning.
  - Beyond gain control, E-I dynamics like *neural oscillation* may help *feature binding* in the temporal dimension.

Thank you!

THANK  
YOU