中国科学技术大学
University of Science and Technology of China

ICLR

Toward Practical Equilibrium Propagation: Brain-inspired Recurrent Neural Network with Feedback Regulation and Residual Connections

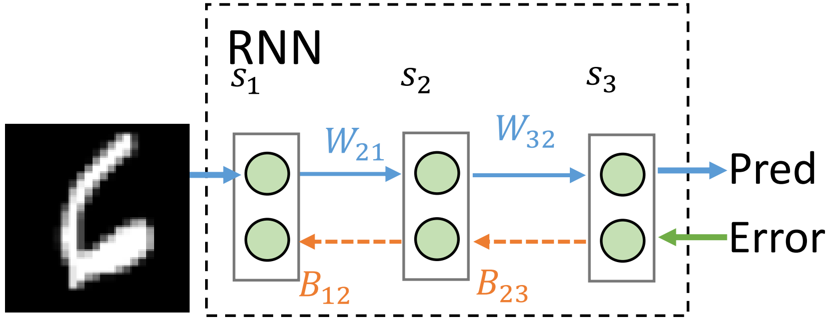
TL;DR: RNNs with feedback regulation and residual connections accelerate Equilibrium Propagation and enhances its scalability.

Zhuo Liu, Tao Chen
School of Microelectronics, University of Science and Technology of China

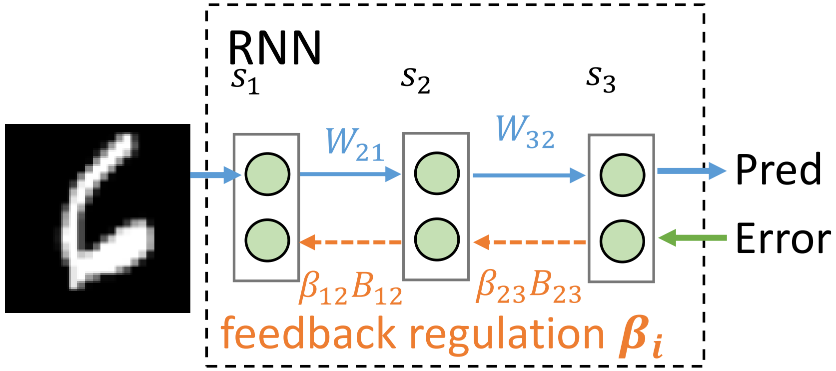
Presentation: Zhuo Liu

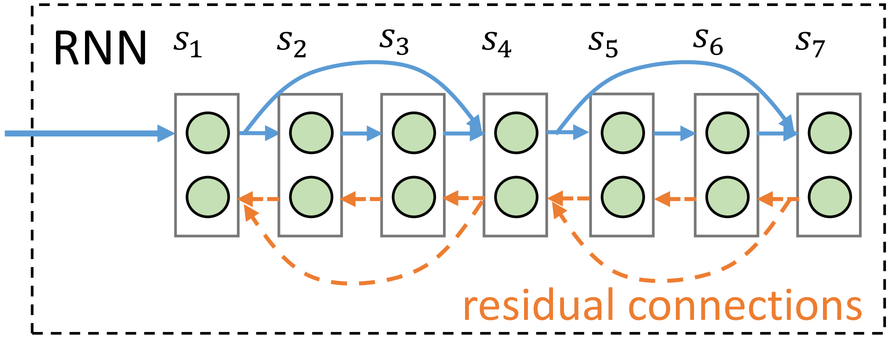
2026.3.28

Background and contribution



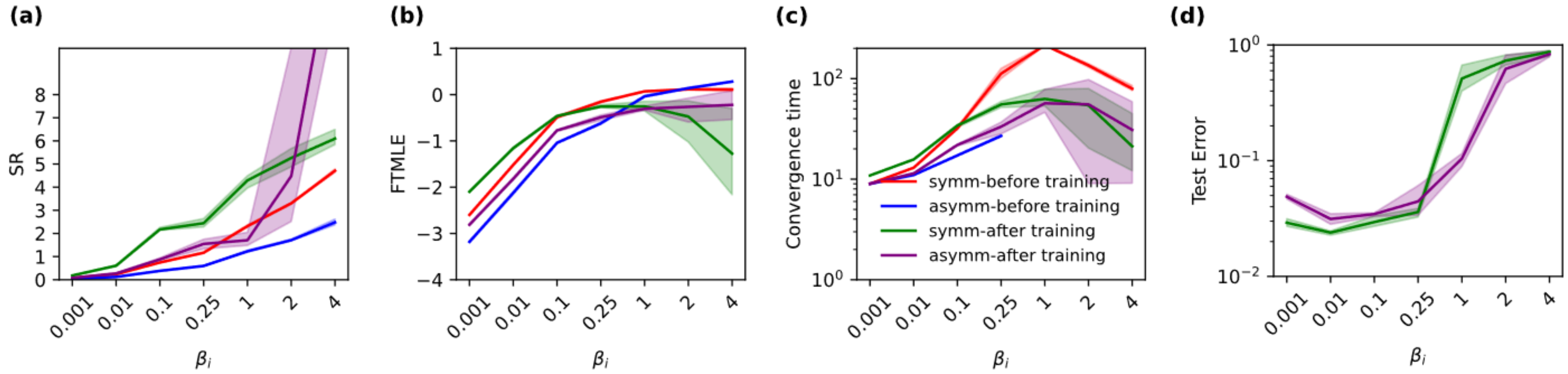
The RNN in original Equilibrium Propagation is hard to converge especially in deeper architectures (harmful to the scalability).



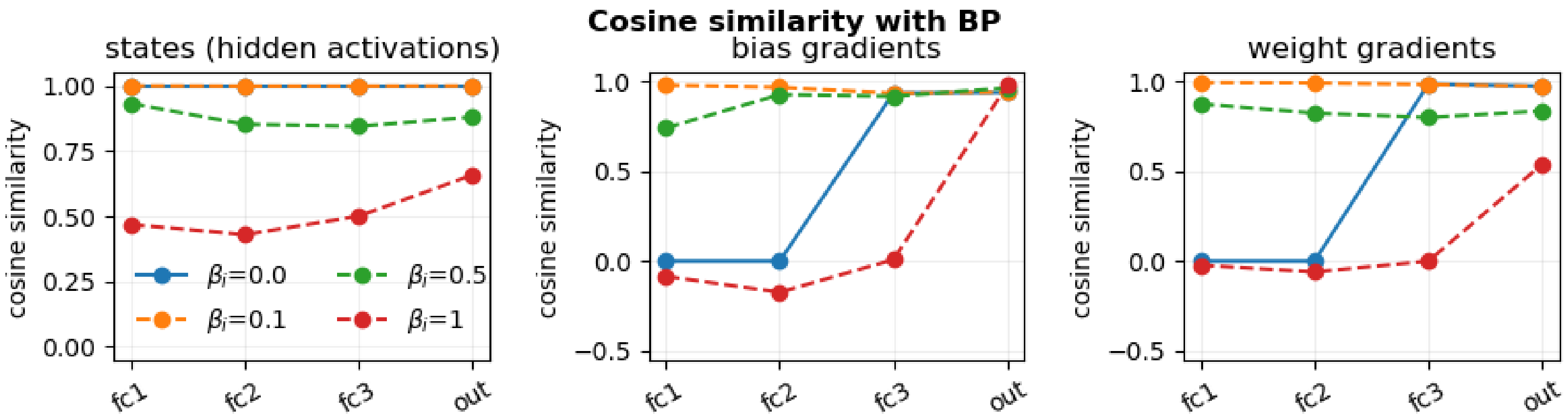


In this work, we use weak feedback regulation and residual connections to make it more practical for less training cost and better scalability.

Our method: 1/ Weak feedback accelerates convergence of RNNs



Weak feedback always leads to smaller spectral radii and FTMLE, enabling RNNs to converge faster.



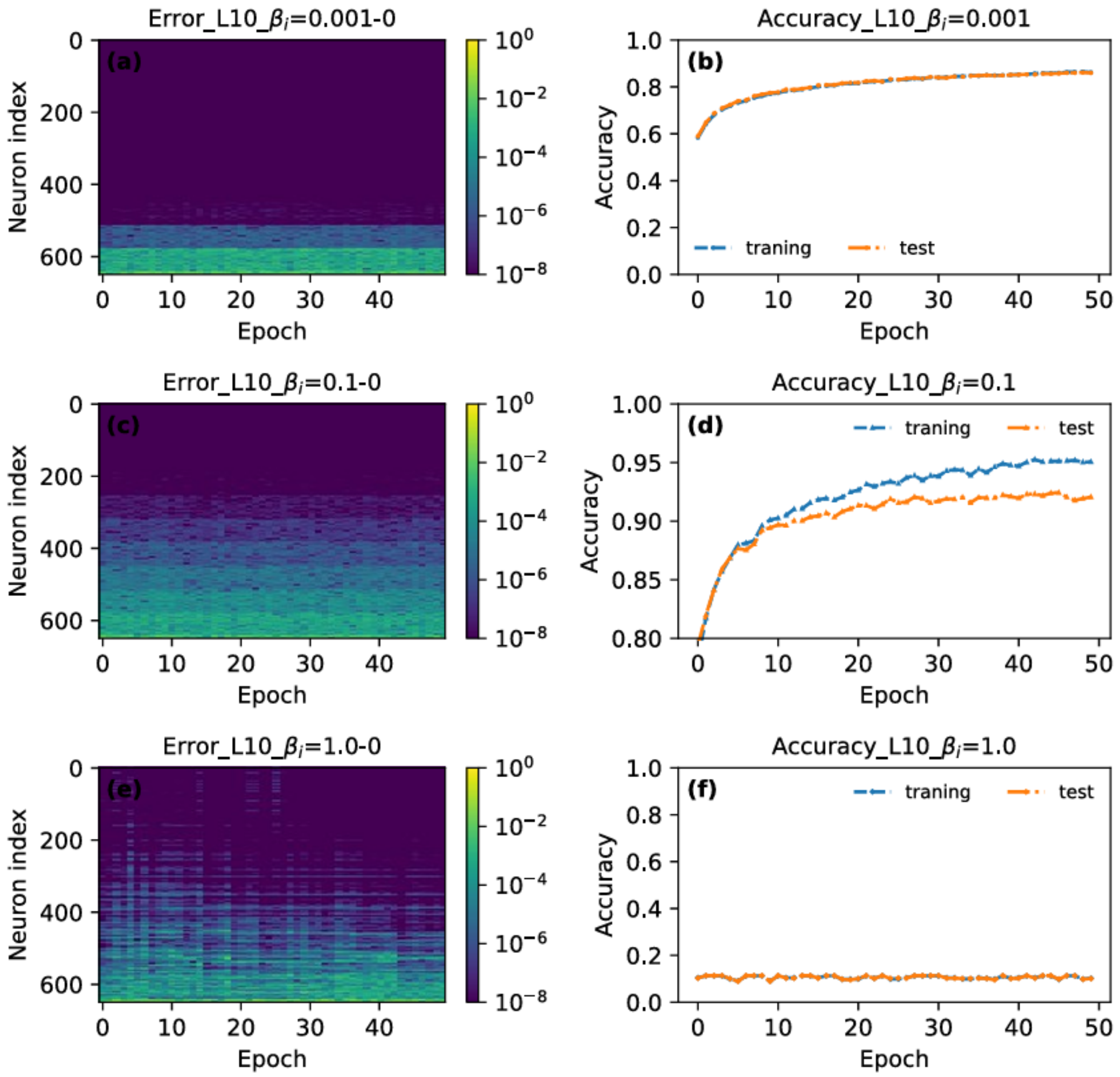
Weak feedback also causes the RNN to behave more like FNNs, with its gradient direction more consistent with backpropagation (BP).

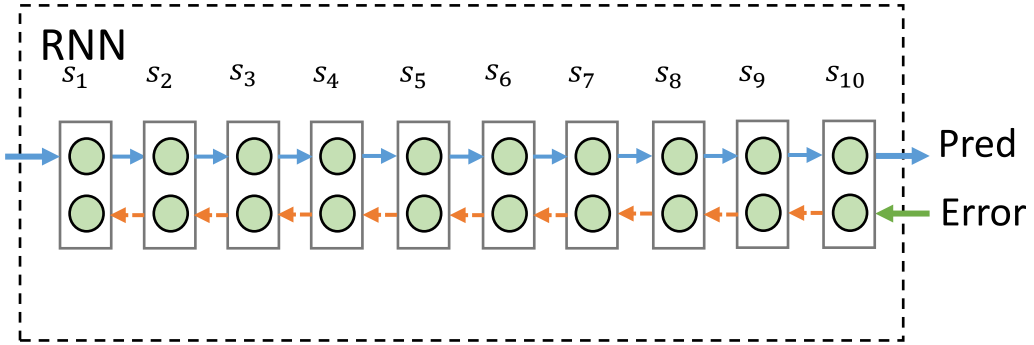
Our method: 1/ Weak feedback accelerates convergence of RNNs

Architecture	Approach	Testing Accuracy	Epoch / Batch size -T/K	Wall Clock Time (HH:MM:SS)
2HL	P-EP	98.05%±0.10%	50/20-100/20	1:56:-
	Ours	98.39%±0.04%	50/500-10/10	0:01:16
	BP	98.26%±0.06%	50/500-1/1	0:00:18
3HL	P-EP	97.99%±0.18%	100/20-180/20	8:27:-
	Ours	97.83%±0.13%	100/20-18/10	1:01:54
	Ours	98.36%±0.06%	50/500-18/10	0:02:11
	BP	98.36%±0.08%	50/500-1/1	0:00:24
Conv 32c-64c	P-EP	98.98%±0.04%	40/20-200/10	8:58:-
	Ours	99.14%±0.02%	40/128-20/10	0:12:28
	BP	98.93%±0.18%	40/128-1/1	0:01:01

Compared to P-EP, the time and computational costs are significantly reduced owing to fewer iterations (less computation).

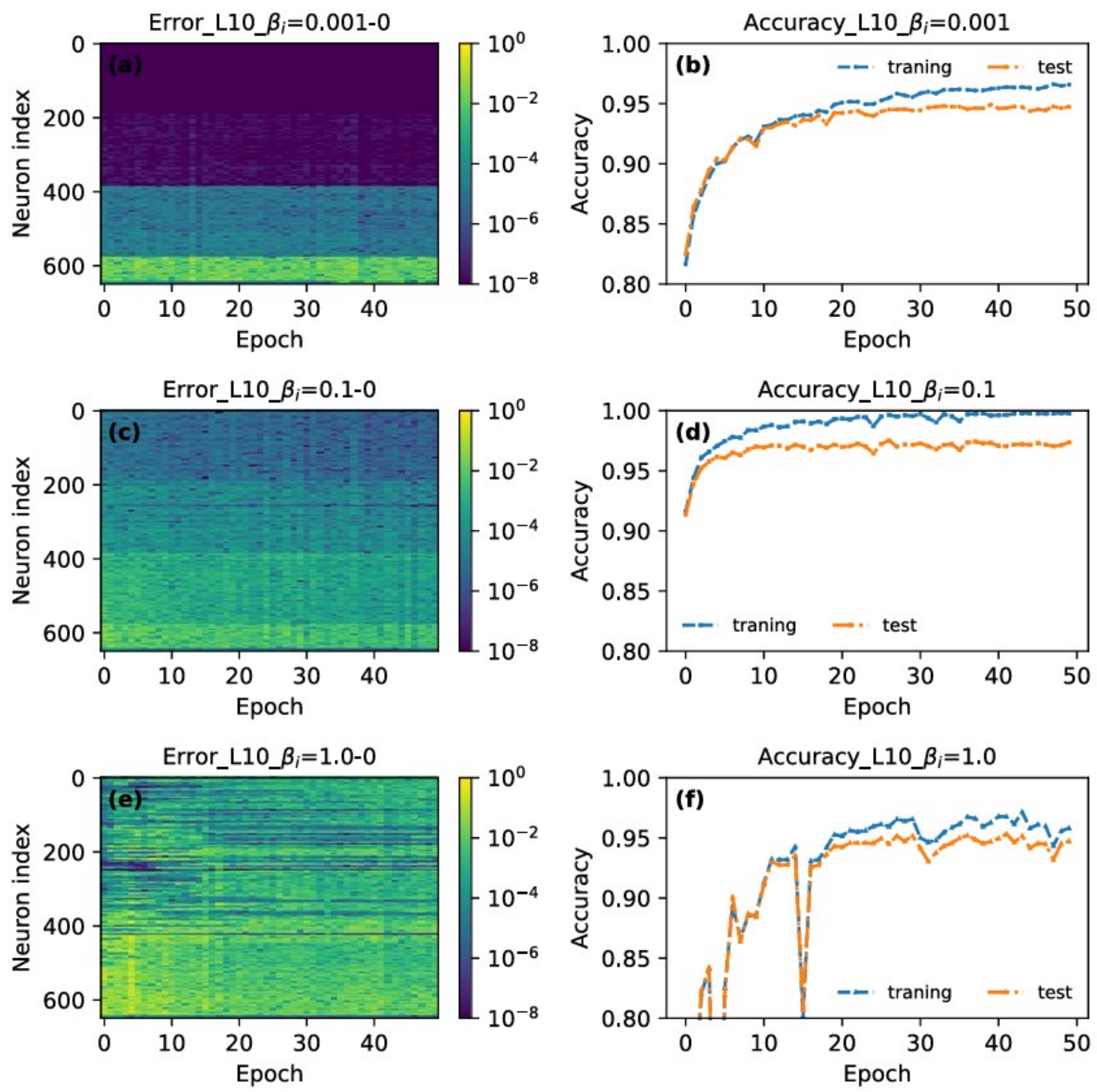
2/ but it leads to obvious gradient vanishing in deeper networks

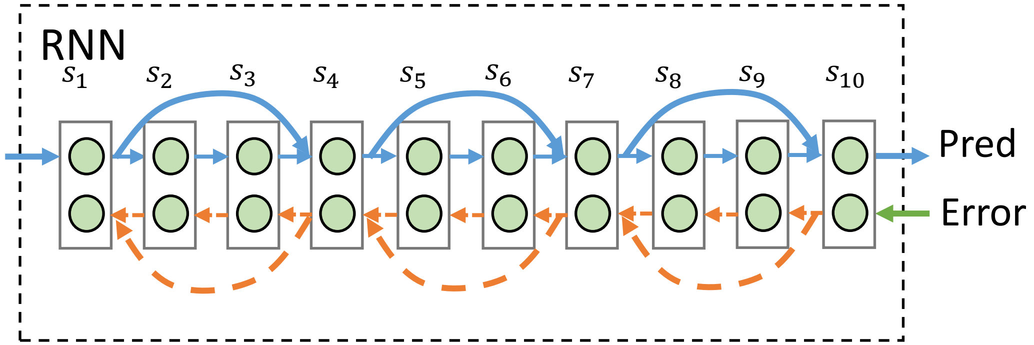




Neurons near the input (with lower indices) receive negligible gradients even if without feedback regulation in EP. It makes EP hard to scale to deep networks.

Our method: 2/ residual connections preserve gradient flow





With residual connections, gradient flow is restored, and networks with 10 (or even 20) hidden layers can be trained, enabling EP to scale to deeper architectures.

Our method: 2/ residual connections preserve gradient flow

Architecture	Approach	MNIST-Testing	CIFAR-10-Testing
5HL	BP	97.69%±0.10%	49.23%±0.81%
	Ours	97.64%±0.10%	50.72%±0.17%
10HL	BP	97.61%±0.04%	48.23%±1.26%
	Ours	92.49%±0.32%	34.90%±0.38%
	Ours-Residual	97.49%±0.05%	44.46%±0.51%
20HL	BP	97.48%±0.07%	47.35%±1.49%
	Ours-Residual	95.95%±0.18%	43.61%±1.17%
Conv 32c-64c-128c	BP	99.34%±0.04%	75.45%±0.46%
	Ours	99.27%±0.07%	75.04%±0.51%

With residual connections, gradient flow is restored, and networks with 10 (or even 20) hidden layers can be trained, enabling EP to scale to deeper architectures.

Takeaways

1. Weak feedback accelerates convergence.
   - It reduces the computation and time cost by changing dynamics of RNN.
   - It makes RNNs behave more like FNNs, with the gradients more consistent with BP-trained FNNs.

2. Residual connections preserve gradient flow.
   -It makes EP feasible for deeper networks, e.g. 20-hidden-layer FNN.

Therefore,
Our work accelerates Equilibrium Propagation and enhances its scalability, taking an important step towards the practical application of EP.

Thank you!