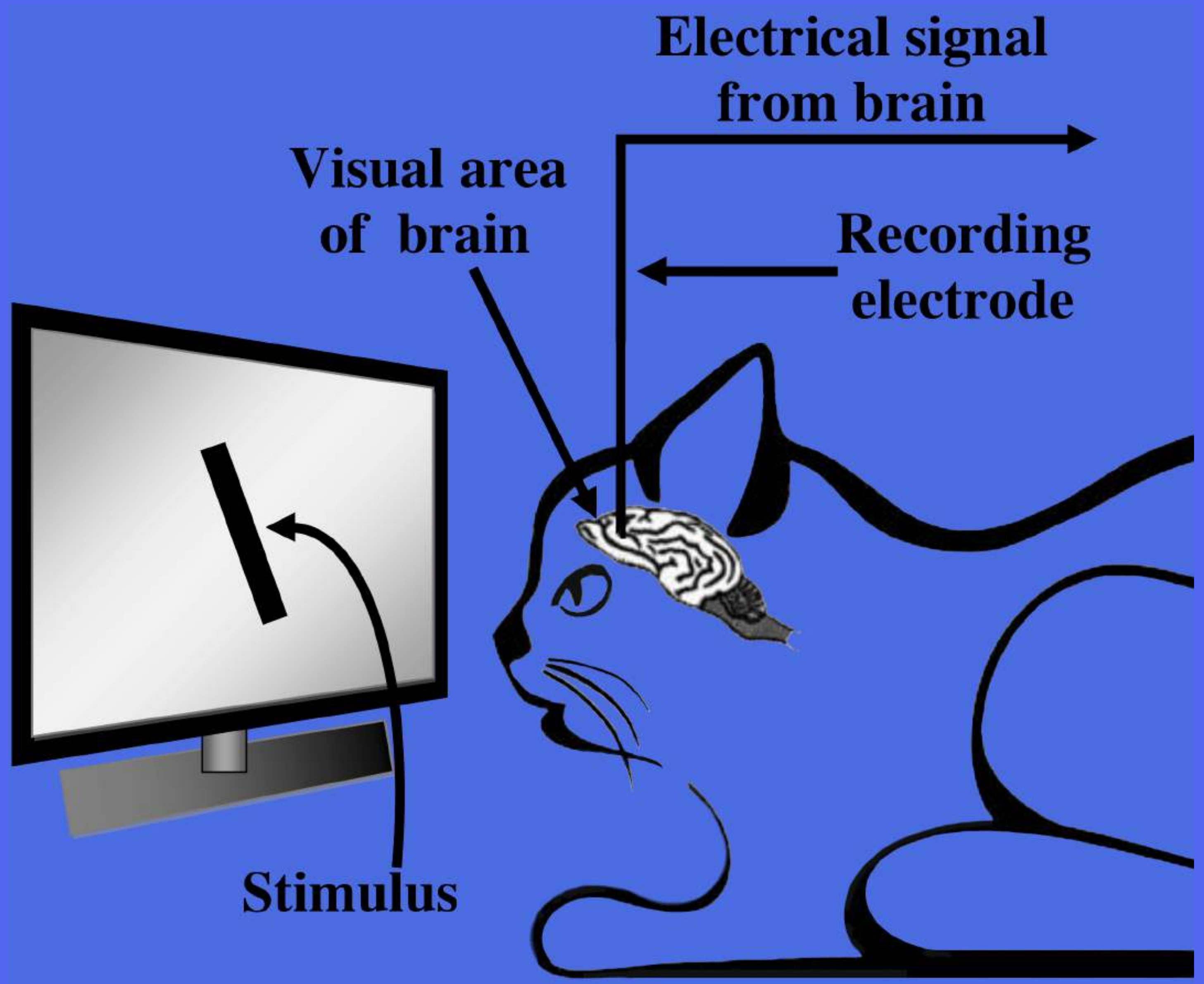
A Biologically Plausible Learning Algorithm for Neural Networks

Dmitry Krotov
MIT-IBM Watson AI Lab
IBM Research

Does brain do deep learning?

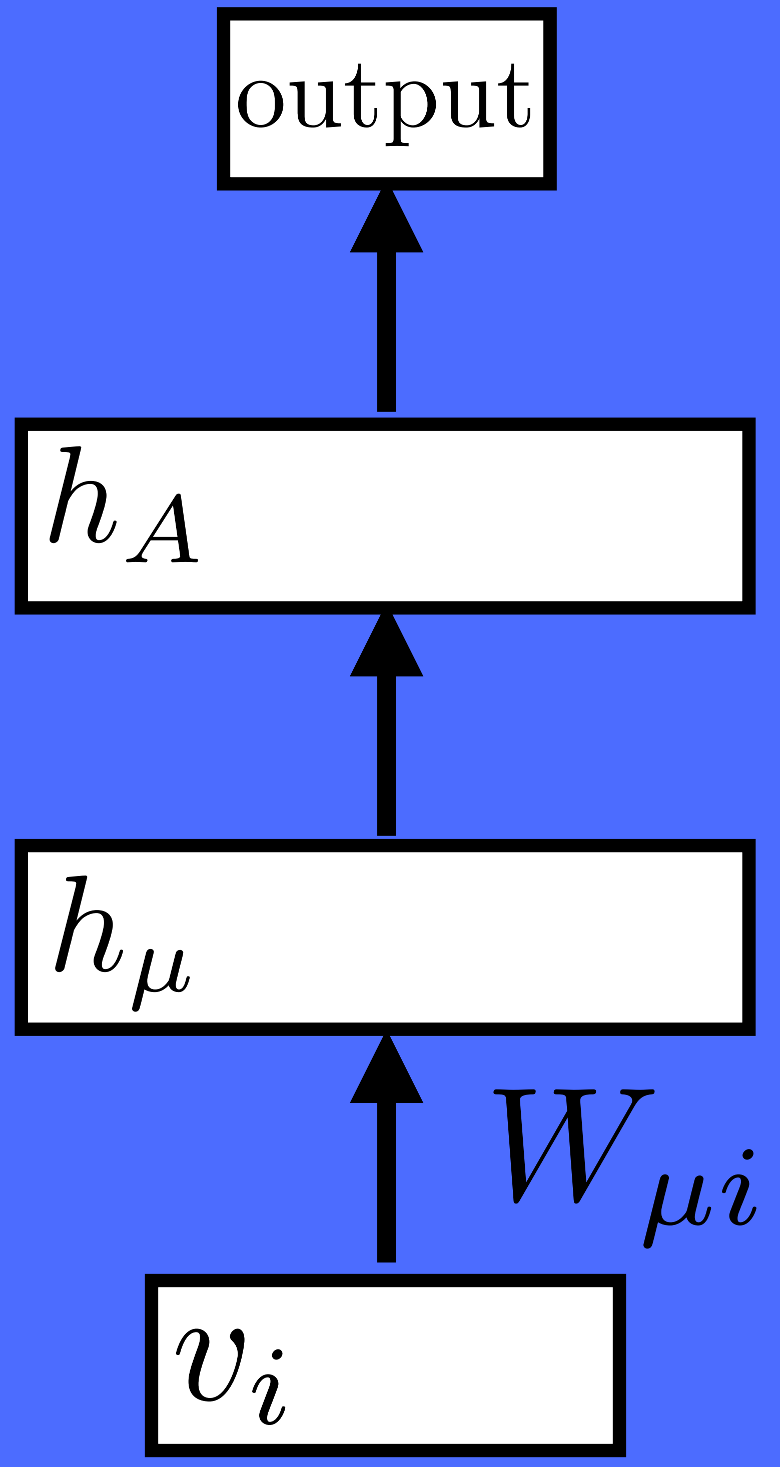
- Architectural similarity between CNNs and visual cortex
- Similarity of the receptive fields of CNNs and direction selective cells

Zeiler and Fergus, 2014



Hubel Wiesel, 1959

Does brain do backpropagation?

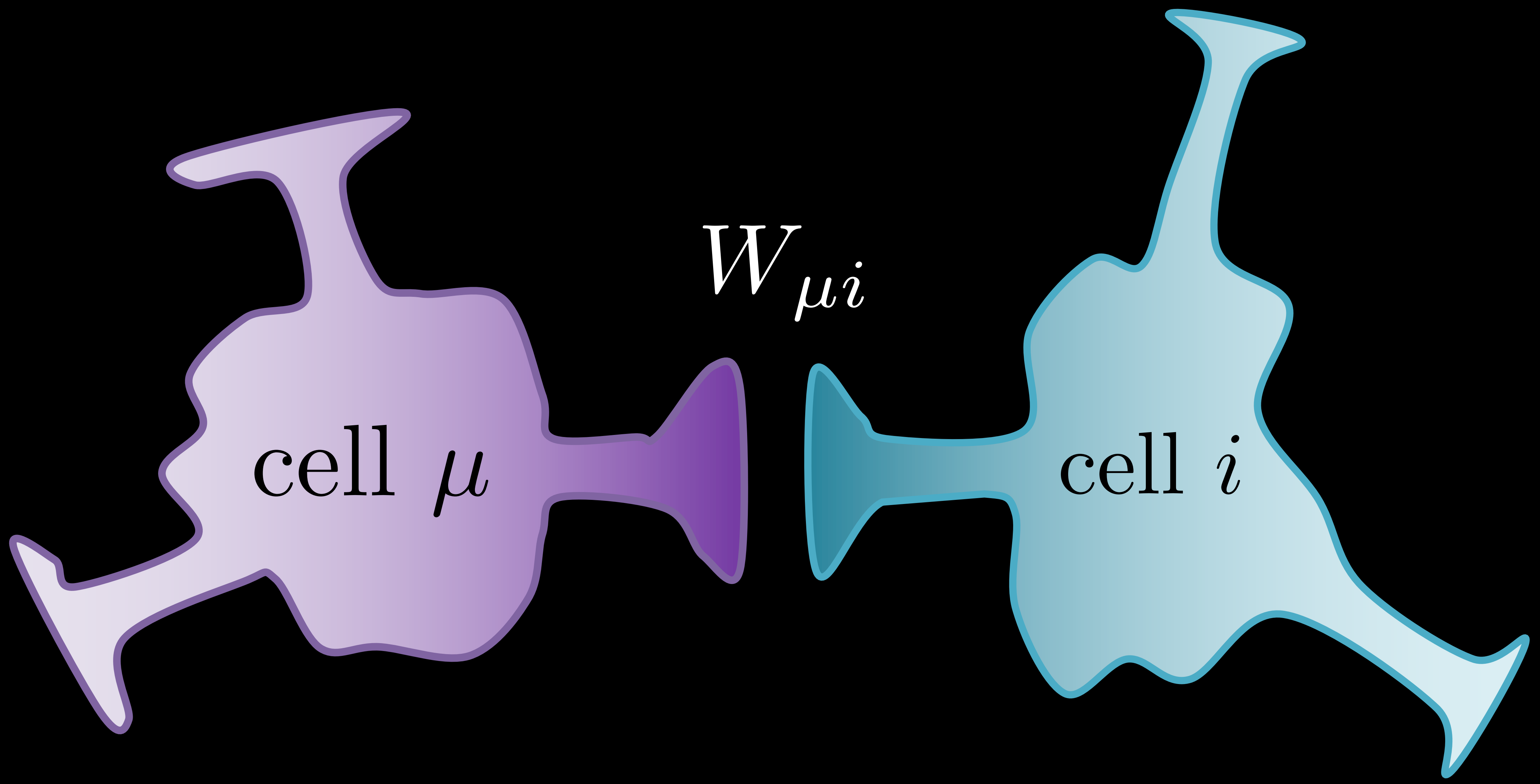


$$L = \sum_{\text{data}} (\text{output} - \text{target})^2$$

$$\Delta W_{\mu i} = -\varepsilon \frac{\partial L}{\partial W_{\mu i}} = \varepsilon \left(\ldots \right)$$

The update rule is non-local

Biological plasticity rules



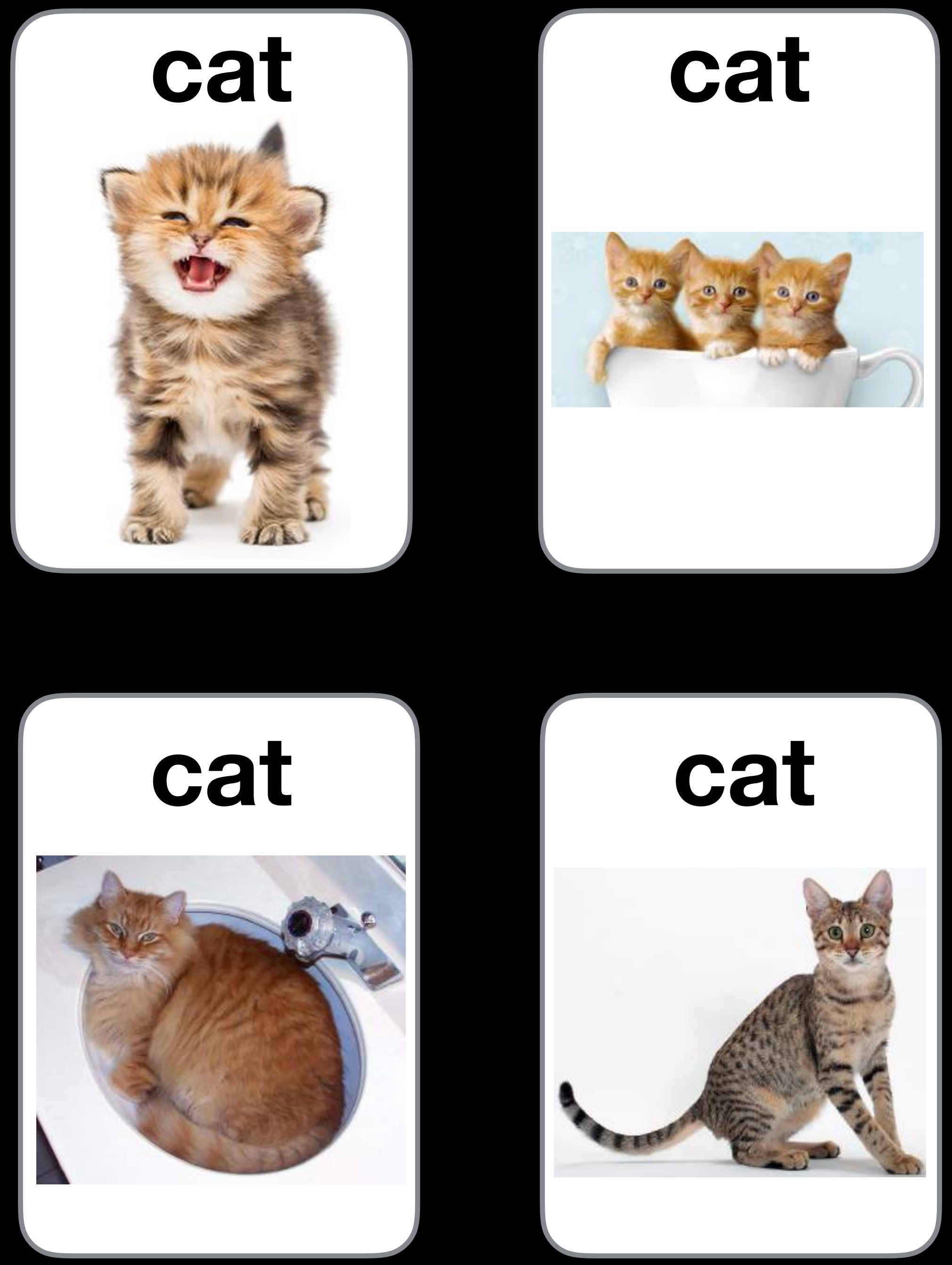
$$\Delta W_{\mu i} = f\left(I_\mu,\ I_i\right)$$

I_μ – information available to cell μ

I_i – information available to cell i

The update rule is local

Supervised backpropagation learning versus biological learning



vs.

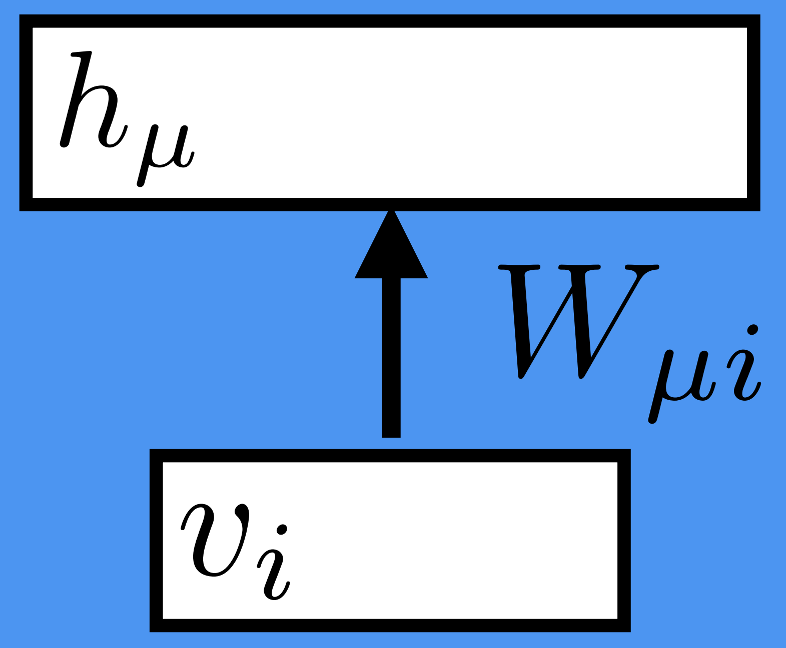
heavily supervised
greedy on labeled data

mostly unsupervised
very few labeled examples

Main question:

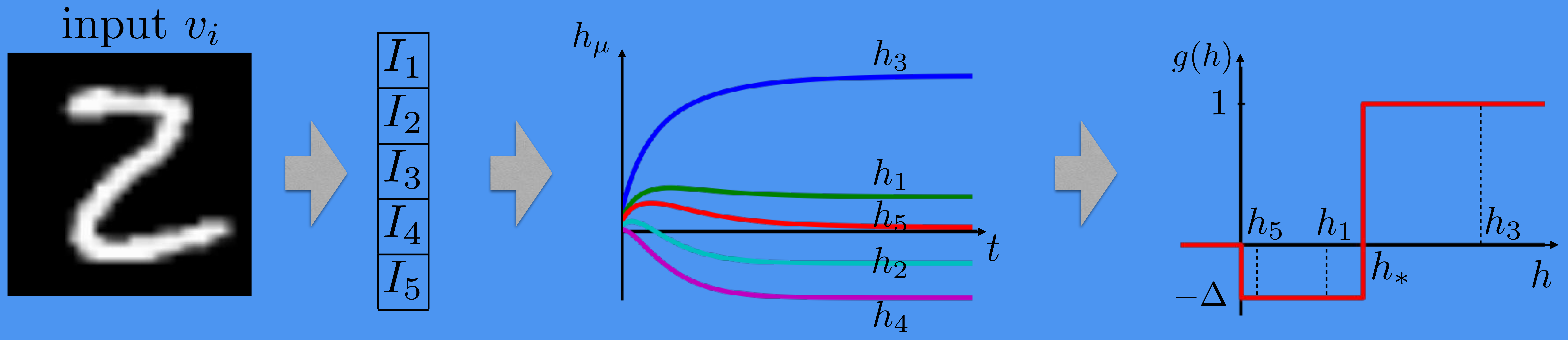
Given unsupervised aspect of learning and locality of synaptic plasticity rules, can we engineer a learning algorithm that leads to a good generalization performance?

Learning algorithm



$$\tau \frac{dh_\mu}{dt} = I_\mu - w_{\text{inh}} \sum_{\nu \neq \mu} r(h_\nu) - h_\mu$$

$$I_\mu = \sum_{i=1}^{N} W_{\mu i} v_i$$



$$\Delta W_{\mu i} \sim g(h_\mu) v_i$$ Hebbian learning

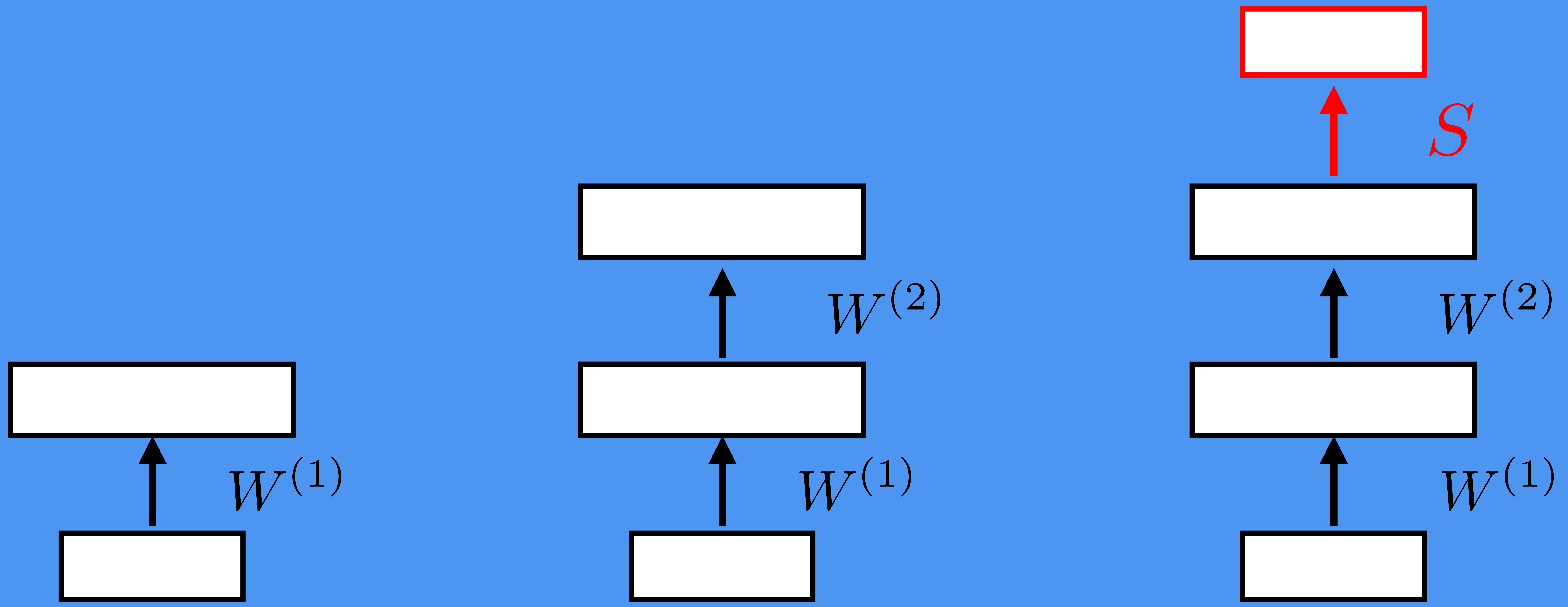
Synaptic plasticity rule

$$\tau_L \frac{dW_{\mu i}}{dt} = g\left(h_\mu\right)\left[v_i - \left(\sum_{k=1}^{N} W_{\mu k} v_k\right) W_{\mu i}\right]$$

$$\sum_{i=1}^{N} W_{\mu i}^2 = 1$$

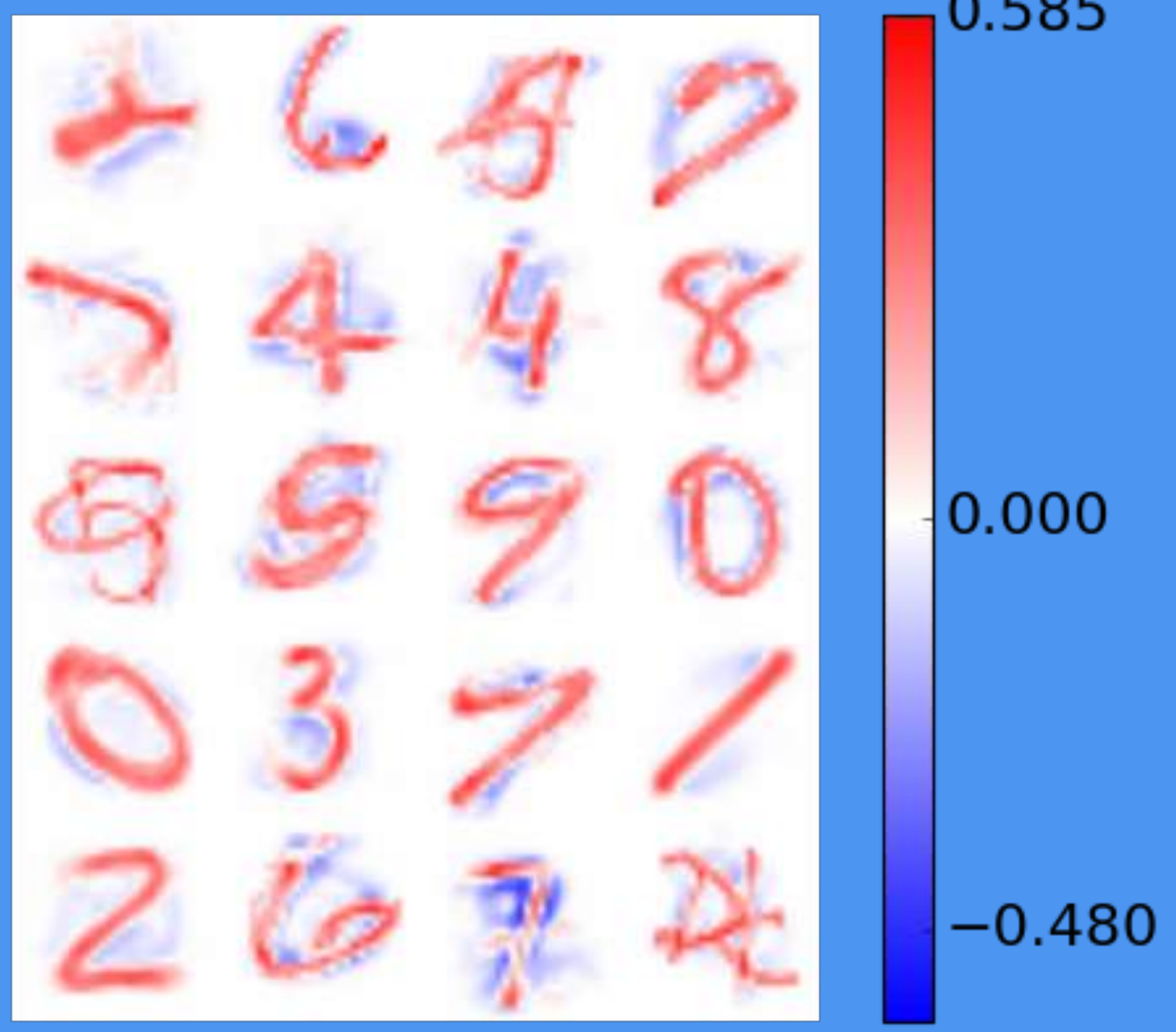
The weights of each hidden unit dynamically converge to the surface of a unit sphere

Stacking the layers

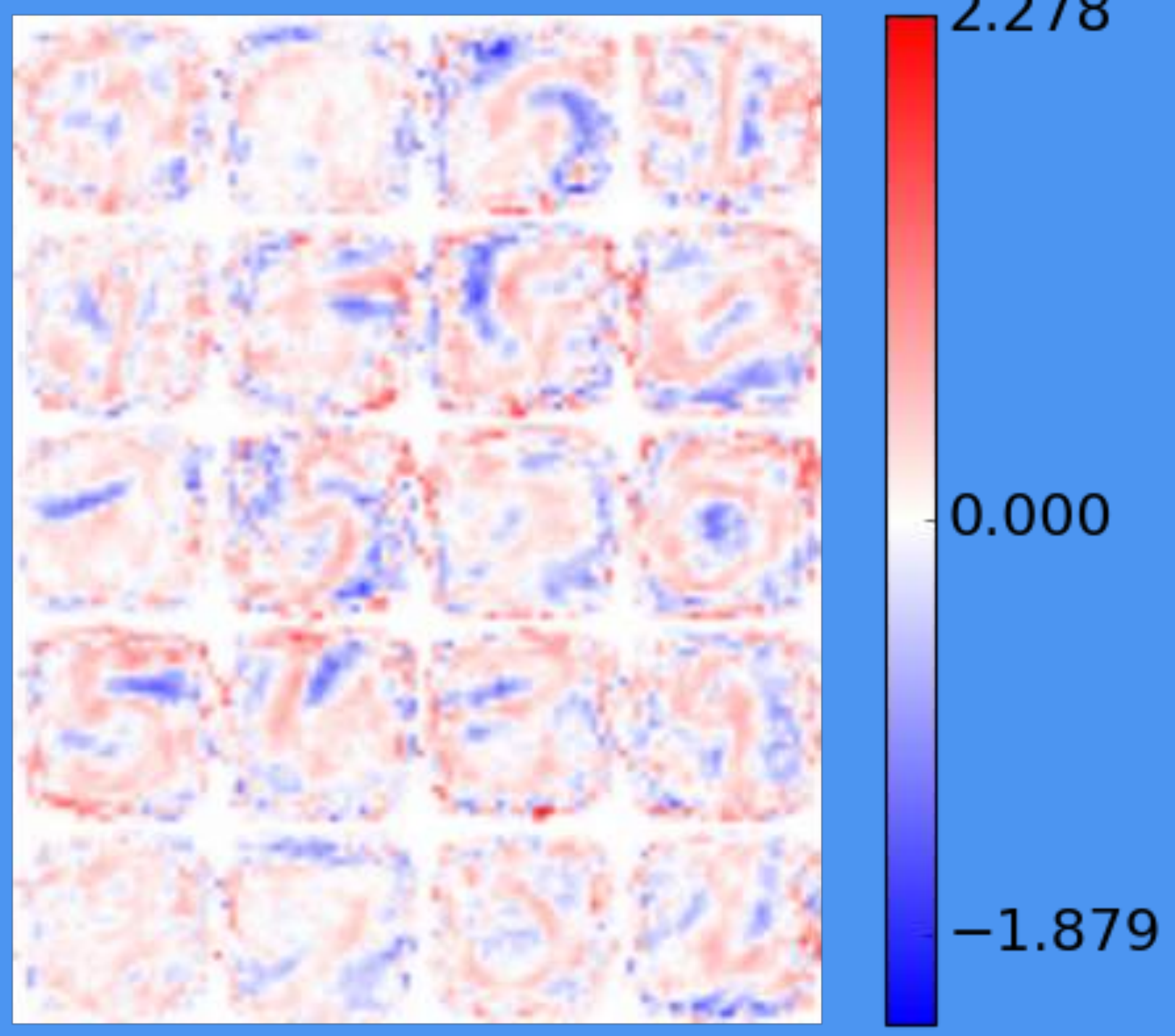


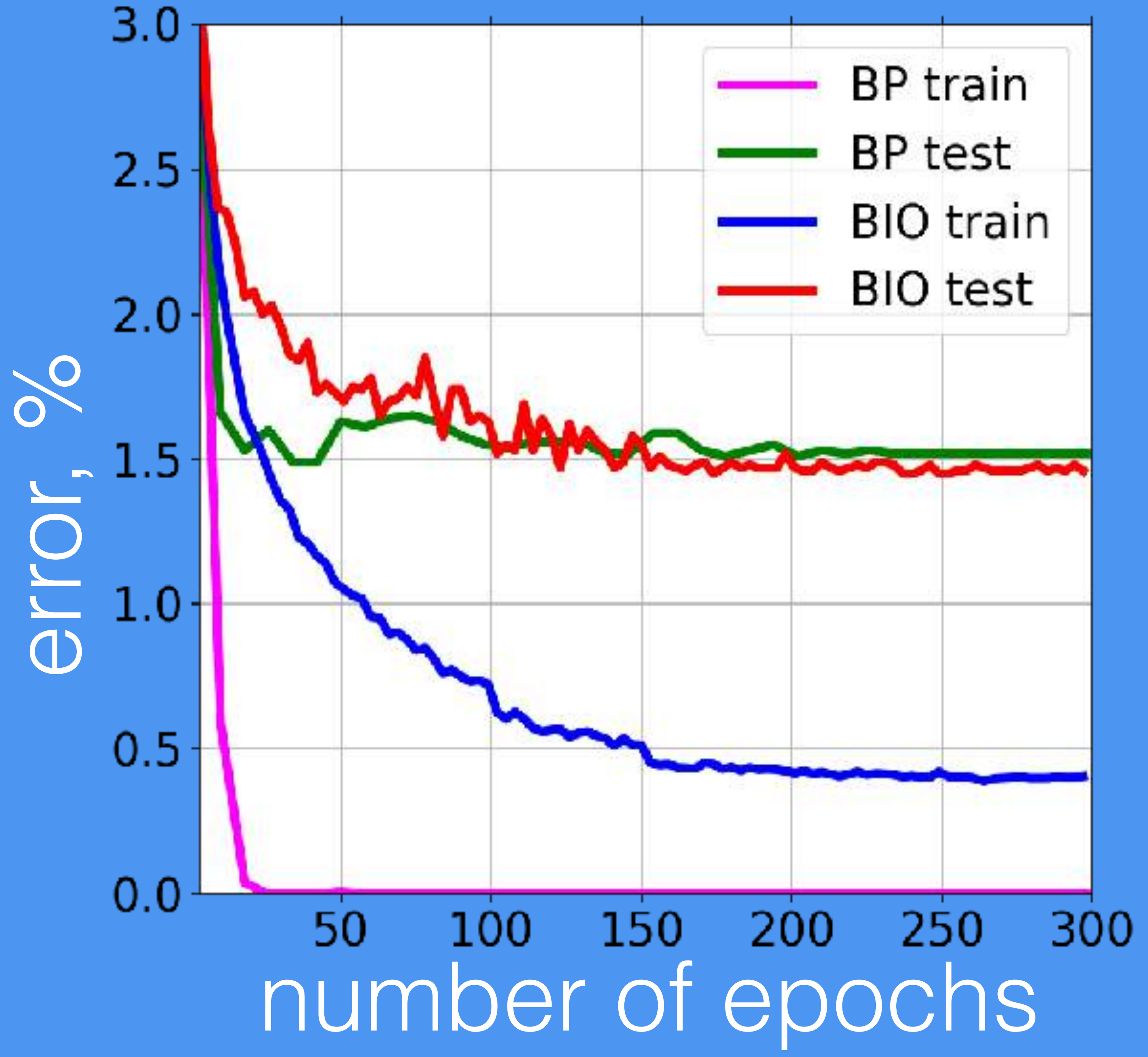
Generalization performance

“biological training”



end-to-end training





Alternative ideas on biologically plausible learning

frontiers in Computational Neuroscience
ORIGINAL RESEARCH
published: 04 May 2017
doi: 10.3389/fncom.2017.00024

Check for updates

test error = 2-3%

Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation

*Benjamin Scellier * and Yoshua Bengio †*

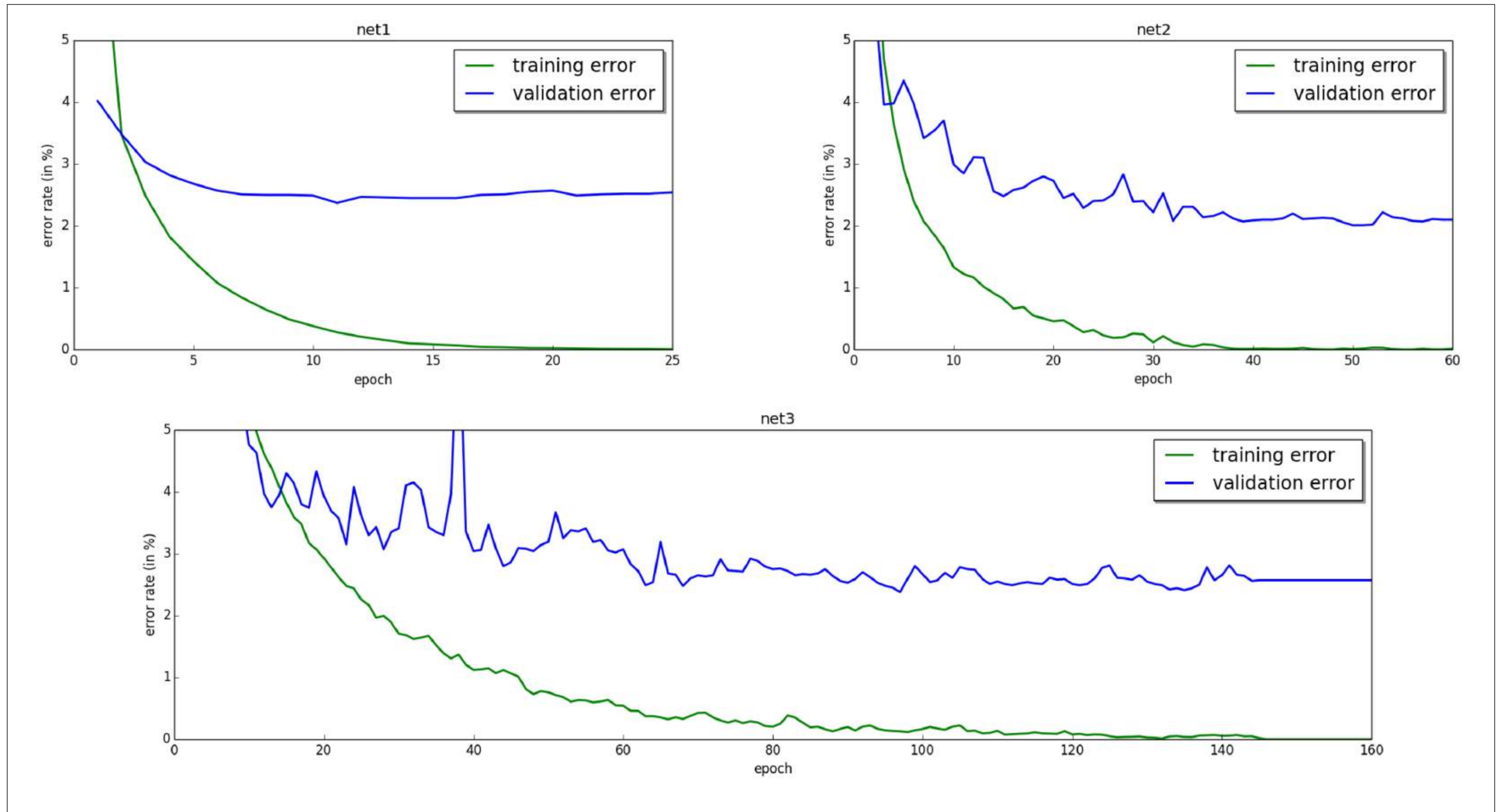


FIGURE 3 | Training and validation error for neural networks with one hidden layer of 500 units (top left), two hidden layers of 500 units **(top right)**, and three hidden layers of 500 units **(bottom)**. The training error eventually decreases to 0.00% in all three cases.

Dendritic error backpropagation in deep cortical microcircuits

João Sacramento[1*], Rui Ponte Costa[1], Yoshua Bengio[2], Walter Senn[1*]

[1]Department of Physiology
University of Bern, Switzerland

[2]Montreal Institute for Learning Algorithms
Université de Montréal, Quebec, Canada

30 Dec 2017

test error = 1.96%

Abstract

Animal behaviour depends on learning to associate sensory stimuli with the desired motor command. Understanding how the brain orchestrates the necessary synaptic modifications across different brain areas has remained a longstanding puzzle. Here, we introduce a multi-area neuronal network model in which synaptic plasticity continuously adapts the network towards a global desired output. In this model synaptic learning is driven by a local dendritic prediction error that

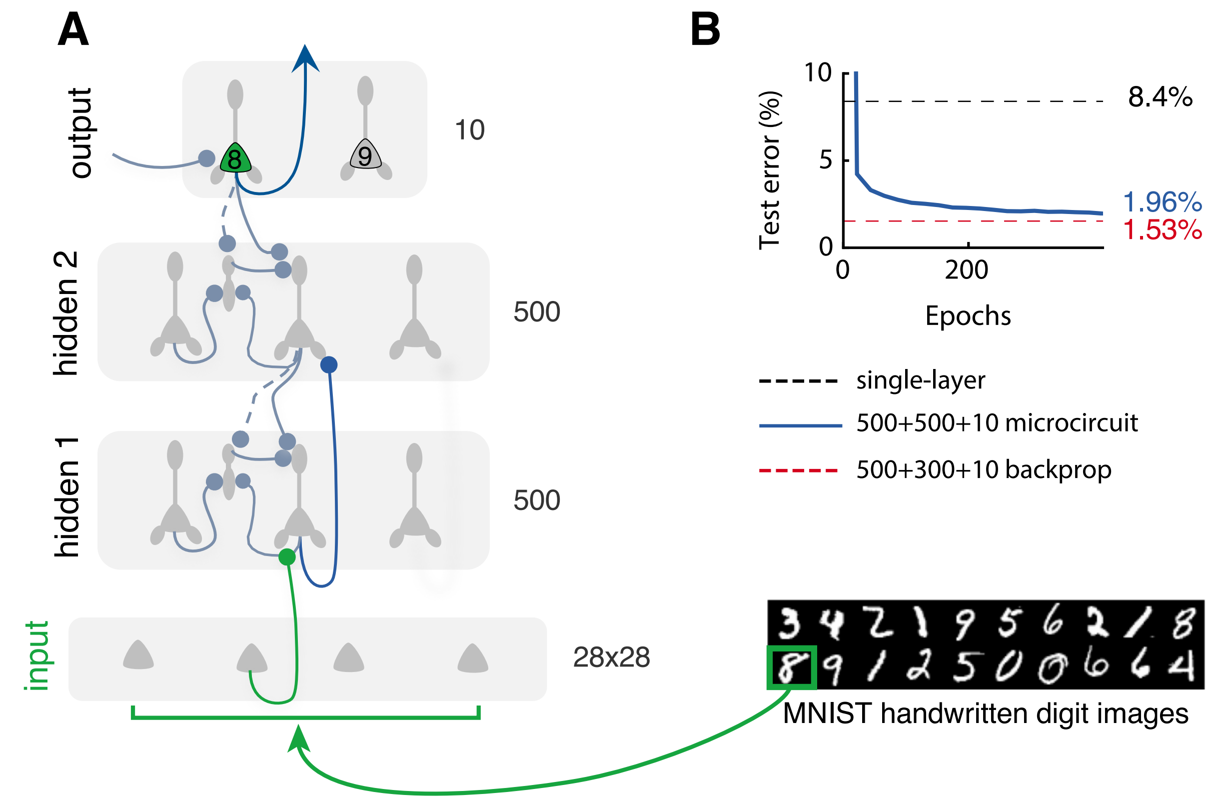
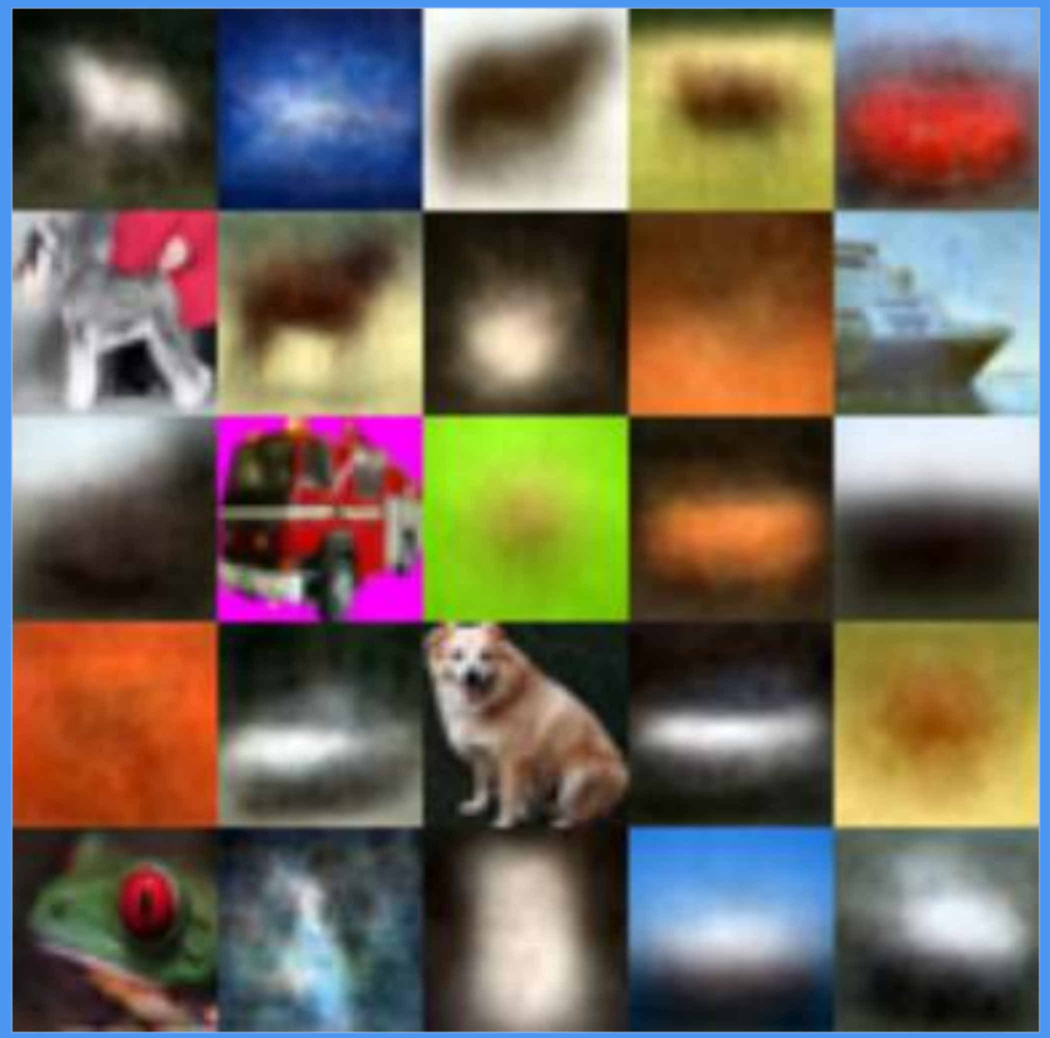


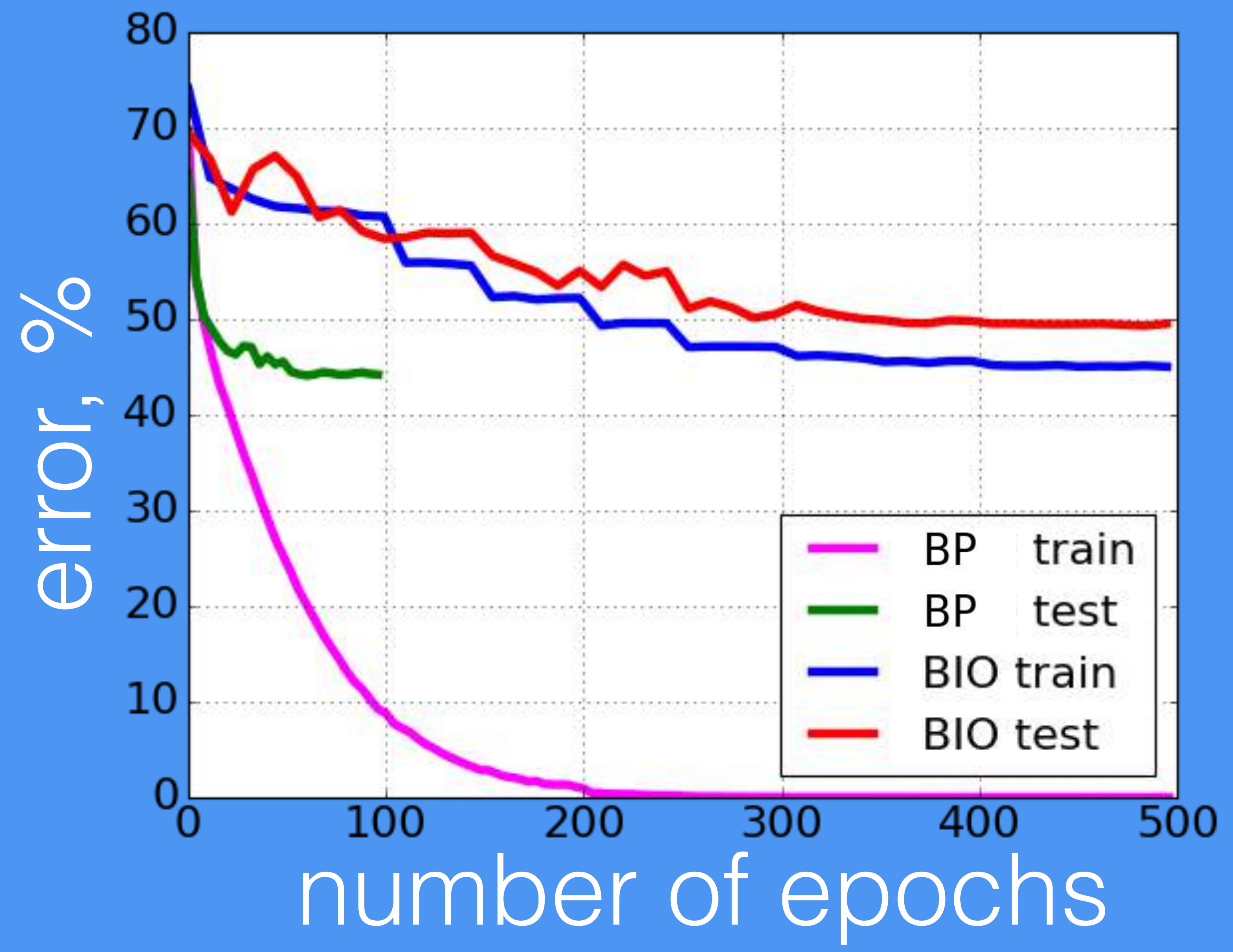
Figure 4: **Learning to classify real-world, structured stimuli with a multi-area network.** (**A**) A 784-500-500-10 (i.e. with two hidden areas) network of pyramidal neurons learns to recognize and classify handwritten digits from the MNIST data set. Only a subset of connections is shown to enhance clarity. (**B**) Competitive accuracy ($< 2\%$, an empirical signature of backprop-like learning) is achieved on the standard MNIST testing dataset by our network (solid blue). For comparison the performance of a shallow learner (i.e. a network in which only output weights are adapted, dashed black) and of a standard artificial neural network trained with backprop (dashed red, see Methods) are also shown.

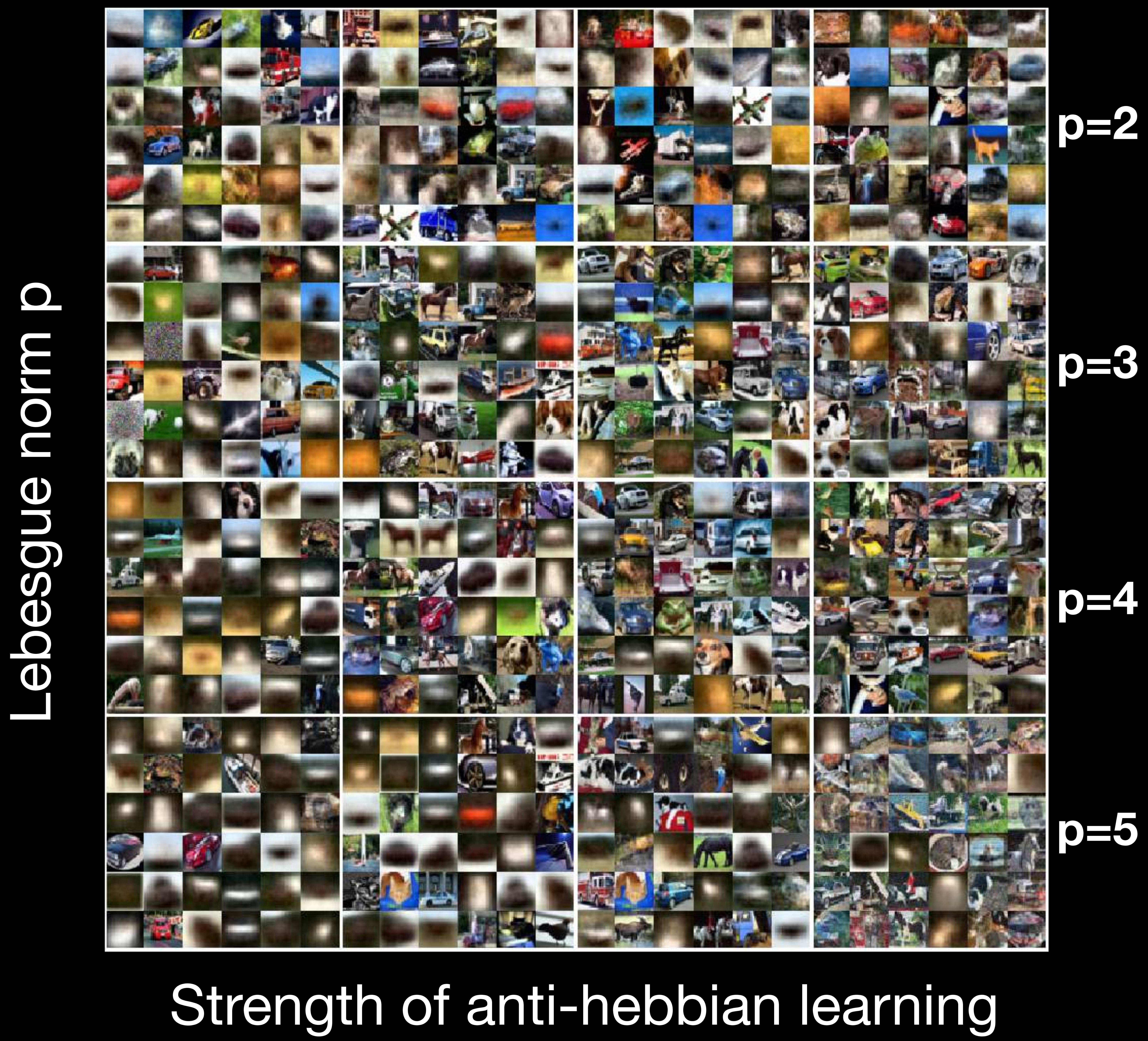
Generalization performance on CIFAR-10

“biological training”



end-to-end training





Work with L.Grinberg

Conclusions

1. Generalization performance of the “biologically plausible” neural network is close to that of the neural network trained end-to-end.

2. The weights of the intermediate layers $W^{(1)}$, $W^{(2)}$, etc., do not have information about the task the network will have to solve eventually. Thus, they produce a “general” representation of the data.

D.Krotov, J.Hopfield: “Unsupervised learning by competing hidden units”, arXiv:1806.10181

GitHub

github.com/DimaKrotov/Biological_Learning

Twitter:

@dimakrotov