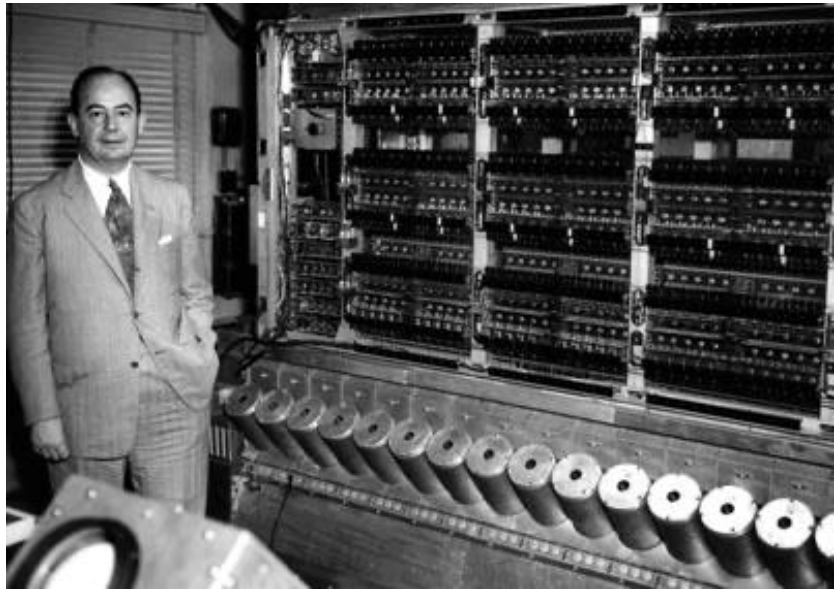
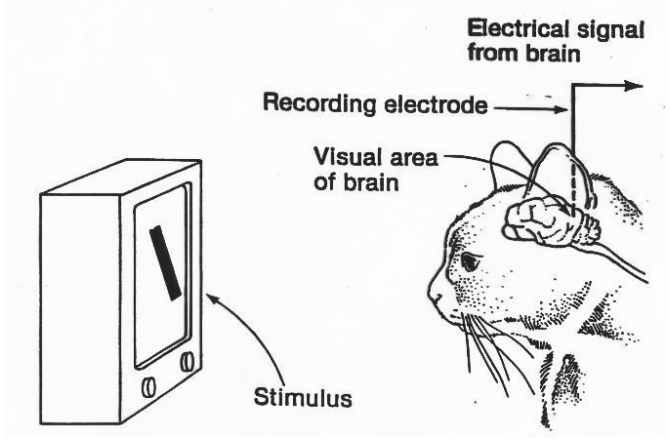
Event-Based Learning of Synaptic Delays in Spiking Neural Networks

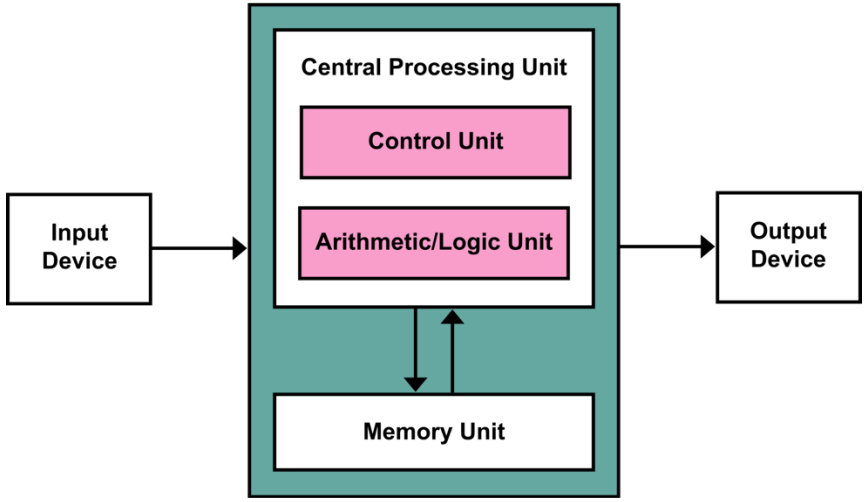
Balázs Mészáros

The Alan Turing Institute
LEVERHULME TRUST
UNIVERSITY OF SUSSEX

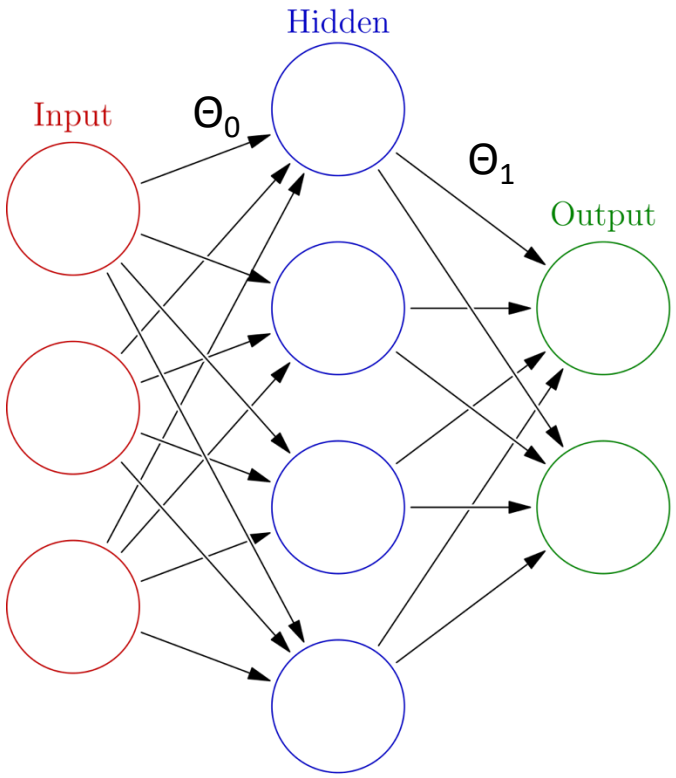
Computer Science 🤝 Neuroscience



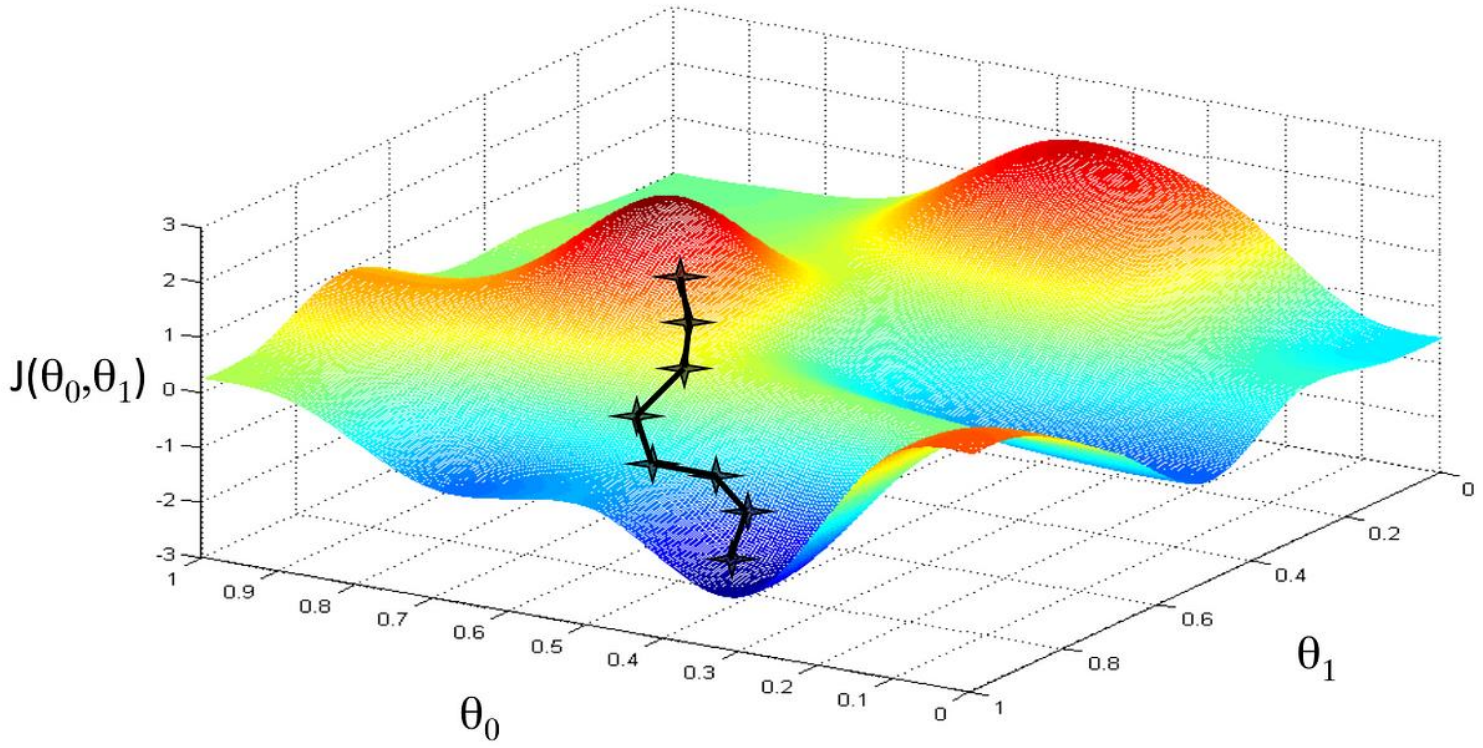




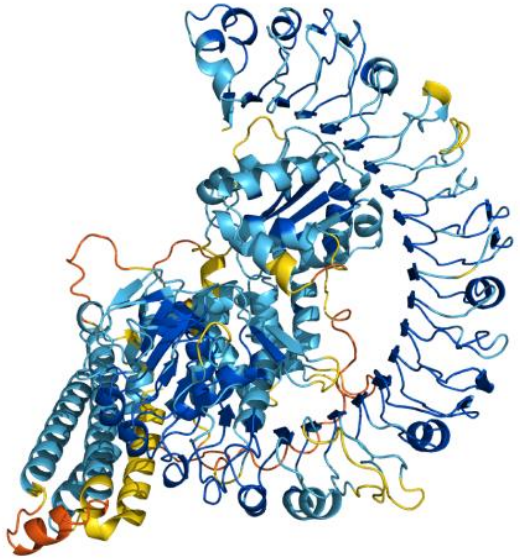
Artificial Neural Networks



Loss function = f(Expect output,Predicted output)

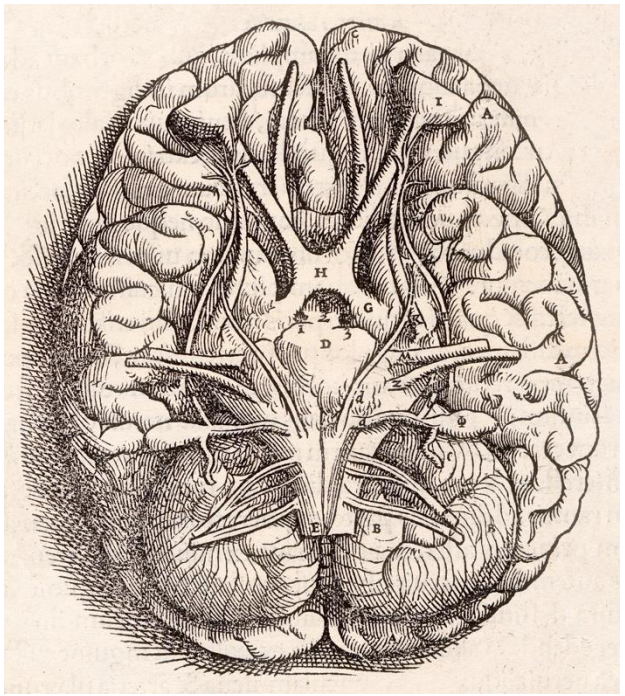


ANNs are powerful



ANNs are inefficient

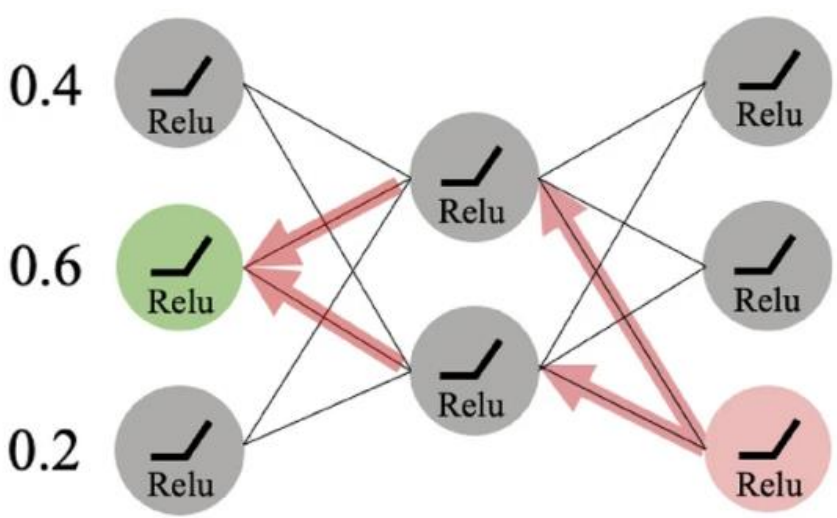
1000 megawatt hours

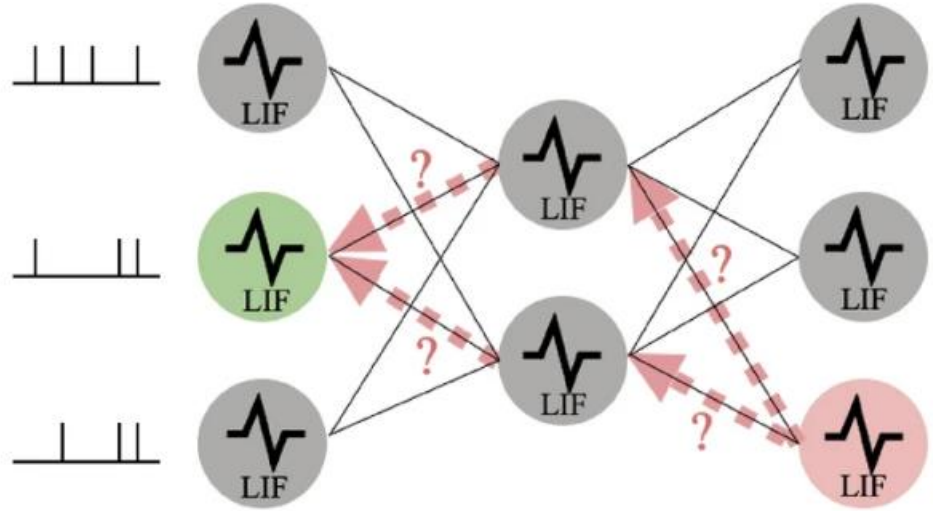


20 watts

Using the brain for 5700 years!!

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*.


0.4
0.6
0.2
Relu
Relu
Relu
Relu
Relu
Relu
Relu
Relu
Relu

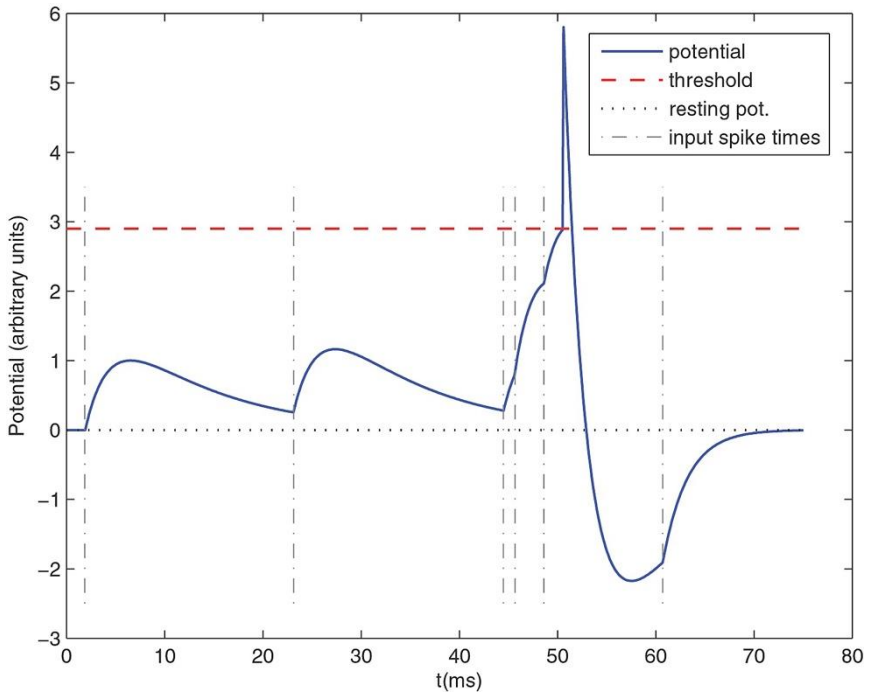

LIF
LIF
LIF
LIF
LIF
LIF
LIF
LIF
?
?
?
?


TESLA
NVIDIA

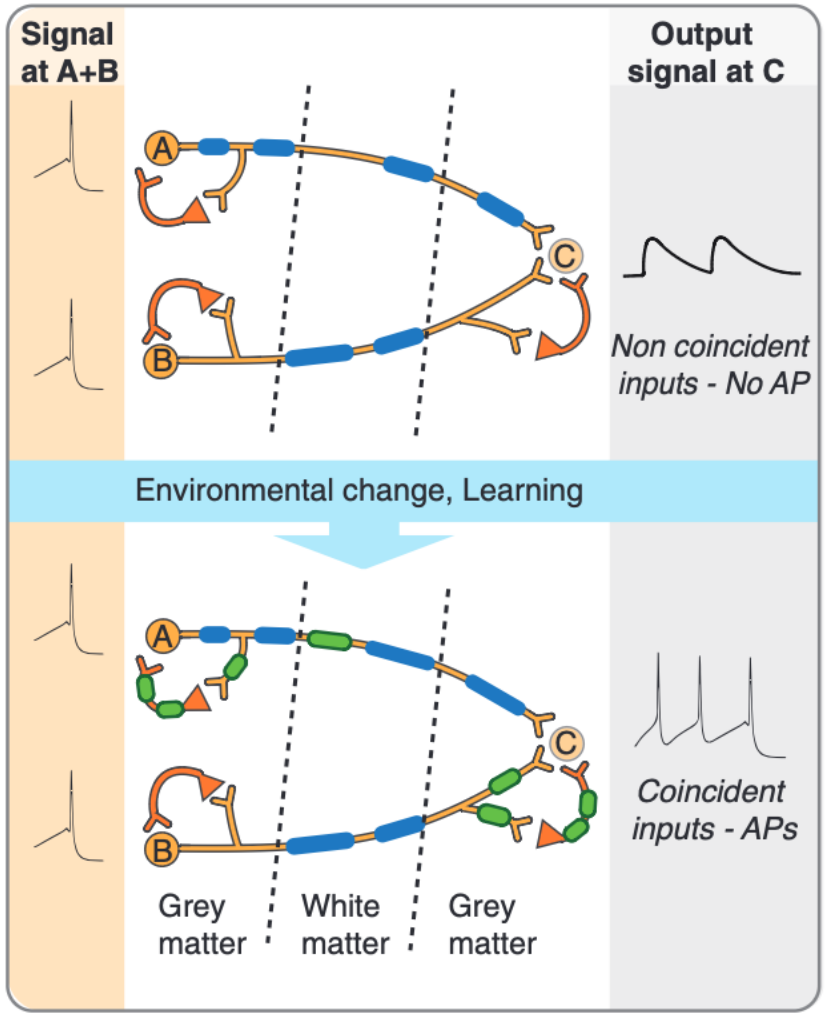

intel labs
Loihi 2

Leaky Integrate-and-Fire neuron

Free dynamics	Transition condition	Jumps at transition
$\tau_{\text{mem}} \frac{\mathrm{d}}{\mathrm{d}t} V = -V + I$ $\tau_{\text{syn}} \frac{\mathrm{d}}{\mathrm{d}t} I = -I$	$(V)_n - \vartheta = 0$ $(\dot{V})_n \neq 0$ for any n	$(V^+)_n = 0$ $I^+ = I^- + We_n$



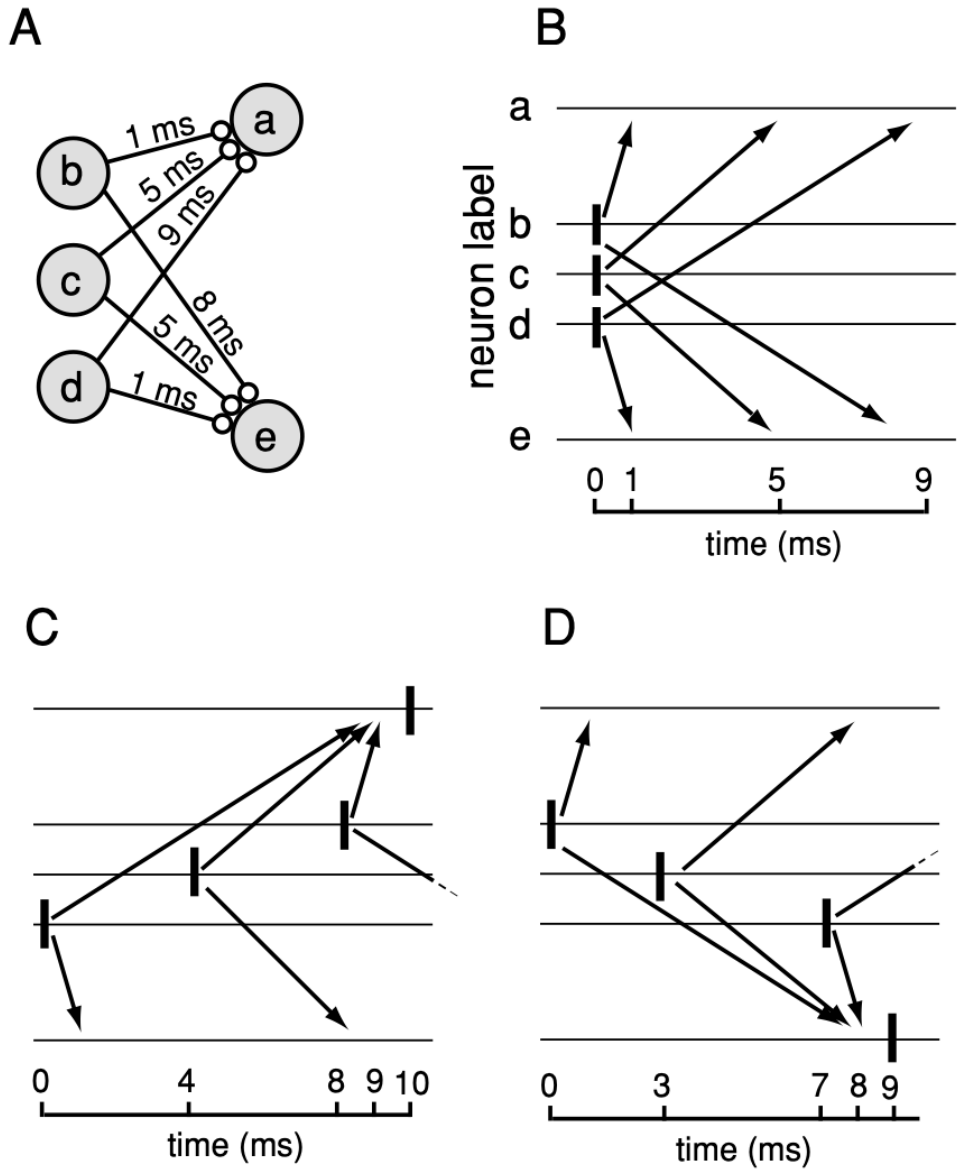
Myelination



Bonetto, G., Belin, D., & Káradóttir, R. T. (2021). Myelin: A gatekeeper of activity-dependent circuit plasticity?. *Science*, *374*(6569), eaba6905.

Delays

- *we can take our point of view to the extreme and say that the network "has memories of all past and future events"*



Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural computation*, *18*(2), 245-282.

Backward propagation

Free dynamics	Transition condition	Jump at transition
$\tau_{\text{mem}}\lambda_V' = -\lambda_V - \frac{\partial l_V}{\partial V}$ $\tau_{\text{syn}}\lambda_I' = -\lambda_I + \lambda_V$	$t - t_k^{\text{post}} = 0$ for any k	$(\lambda_V^-)_{n(k)} = (\lambda_V^+)_{n(k)} + \frac{1}{\tau_{\text{mem}}(\dot{V}^-)_{n(k)}}\Big[\vartheta(\lambda_V^+)_{n(k)} + \left(W^\top(\lambda_V^+ - \lambda_I)\right)_{n(k)} + \frac{\partial l_{\text{p}}}{\partial t_k^{\text{post}}} + l_V^- - l_V^+\Big]$

Gradient updates: $$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}w_{ji}} = -\tau_{\text{syn}} \sum_{\text{spikes from } i} (\lambda_I)_j$$

Wunderlich, T. C., & Pehle, C. (2021). Event-based backpropagation can compute exact gradients for spiking neural networks. *Scientific Reports*, *11*(1), 12829.

Synaptic delays

Free dynamics	Transition condition	Jumps at transition
$\tau_{\text{mem}} \frac{\mathrm{d}}{\mathrm{d}t} V = -V + I$ $\tau_{\text{syn}} \frac{\mathrm{d}}{\mathrm{d}t} I = -I$	$(V)_n - \vartheta = 0$ $(\dot{V})_n \neq 0$ for any n	$(V^+)_n = 0$ $I^+ = I^- + We_n$ ← Delay here

Fixed and learnable synaptic delays

- EventProp updates weights only at spike time:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}w_{ji}} = -\tau_{\mathrm{syn}} \sum_{\text{spikes from } i} (\lambda_I)_j$$

- That doesn't change when delays are included:

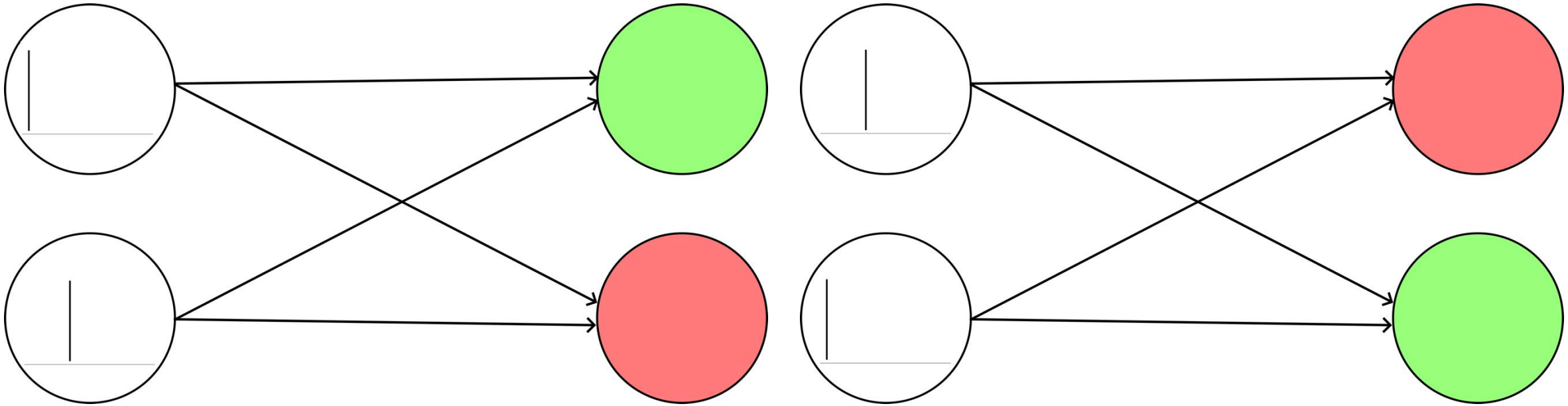
$$\frac{d\mathcal{L}}{dw_{ji}} = -\tau_{syn} \sum_{t_s \in \substack{\text{spikes} \\ \text{from } i}} (\lambda_I)_j \Big|_{t_s + d_{ji}}$$

- And delay updates also only happen at spike times

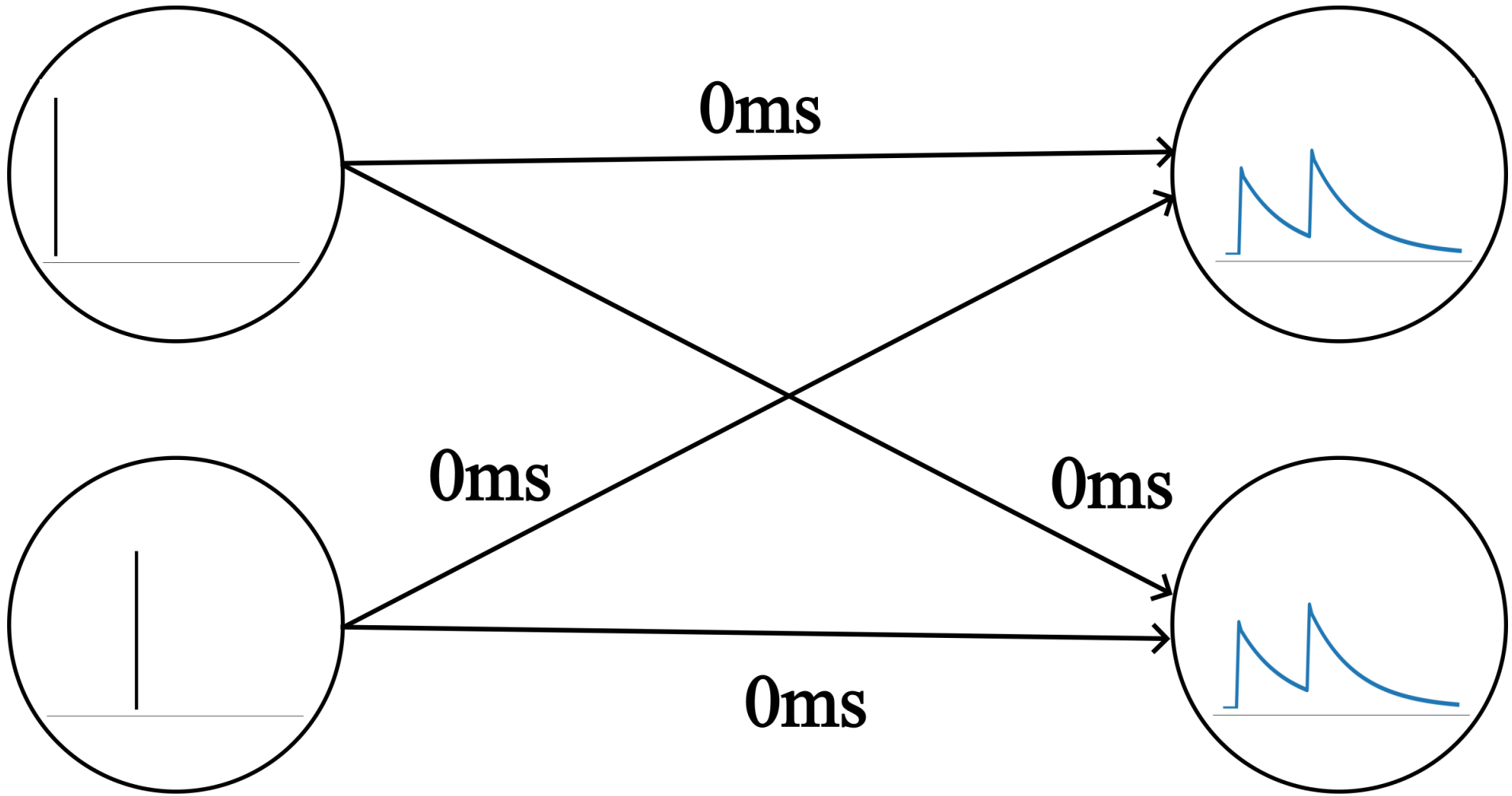
$$\frac{d\mathcal{L}}{dd_{ji}} = w_{ji} \sum_{t_s \in \substack{\text{spikes} \\ \text{from } i}} (\lambda_I - \lambda_V)_j \Big|_{t_s + d_{ji}}$$

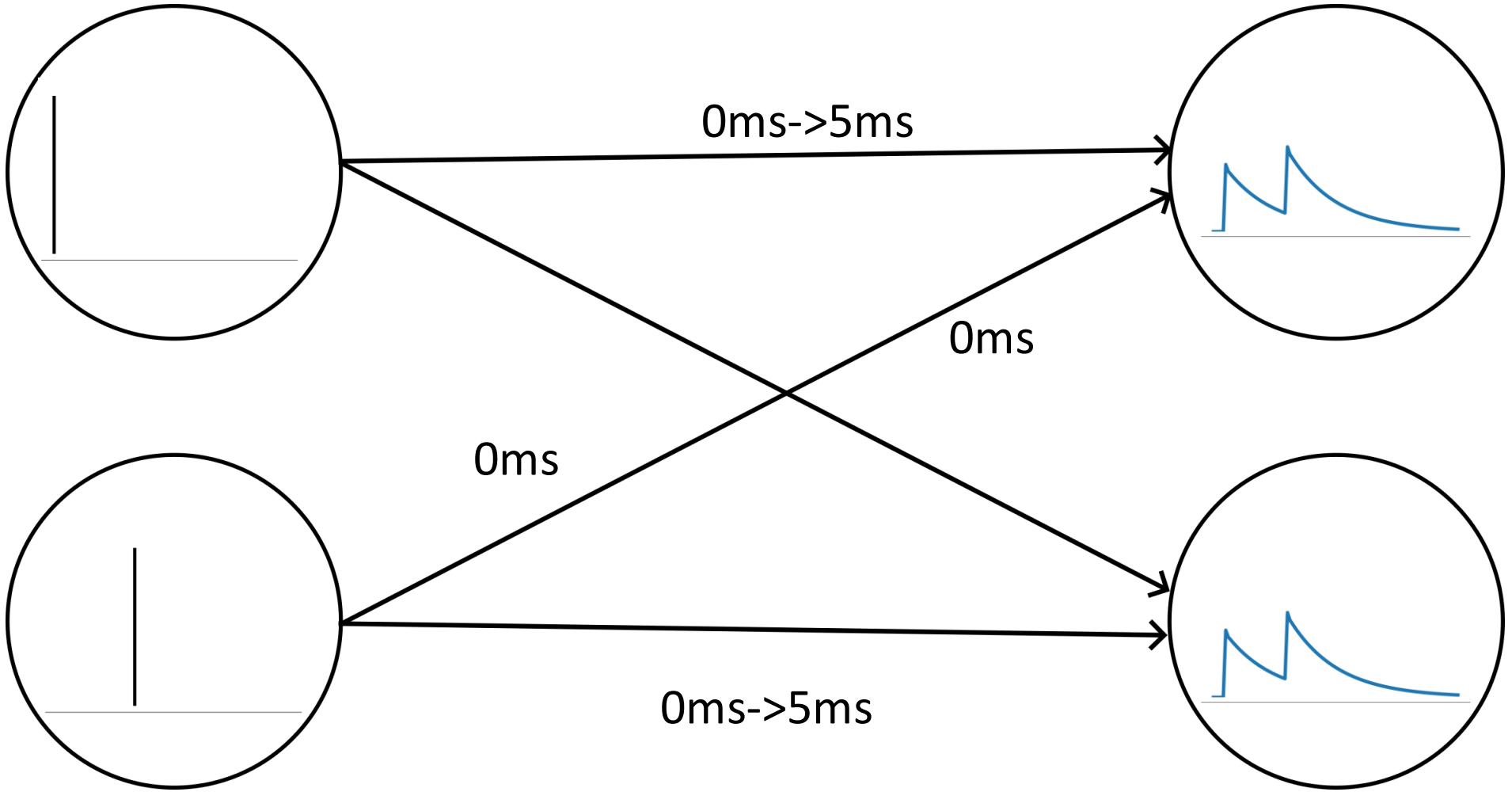
Class 1

Class 2

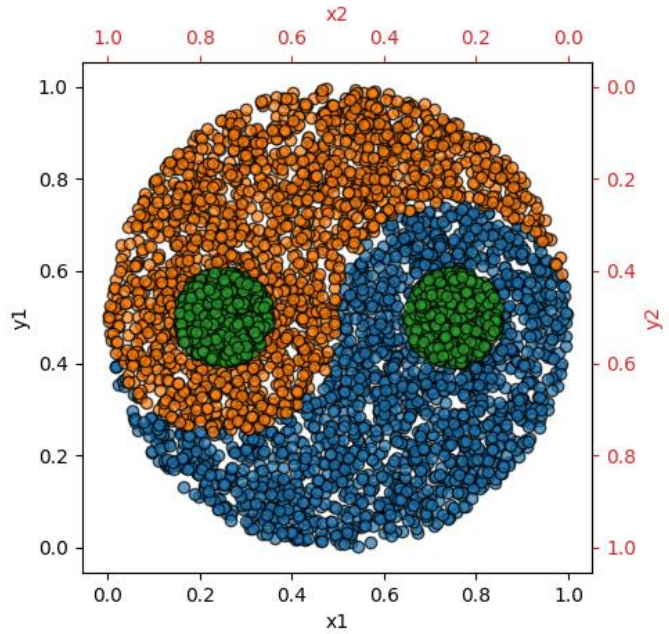


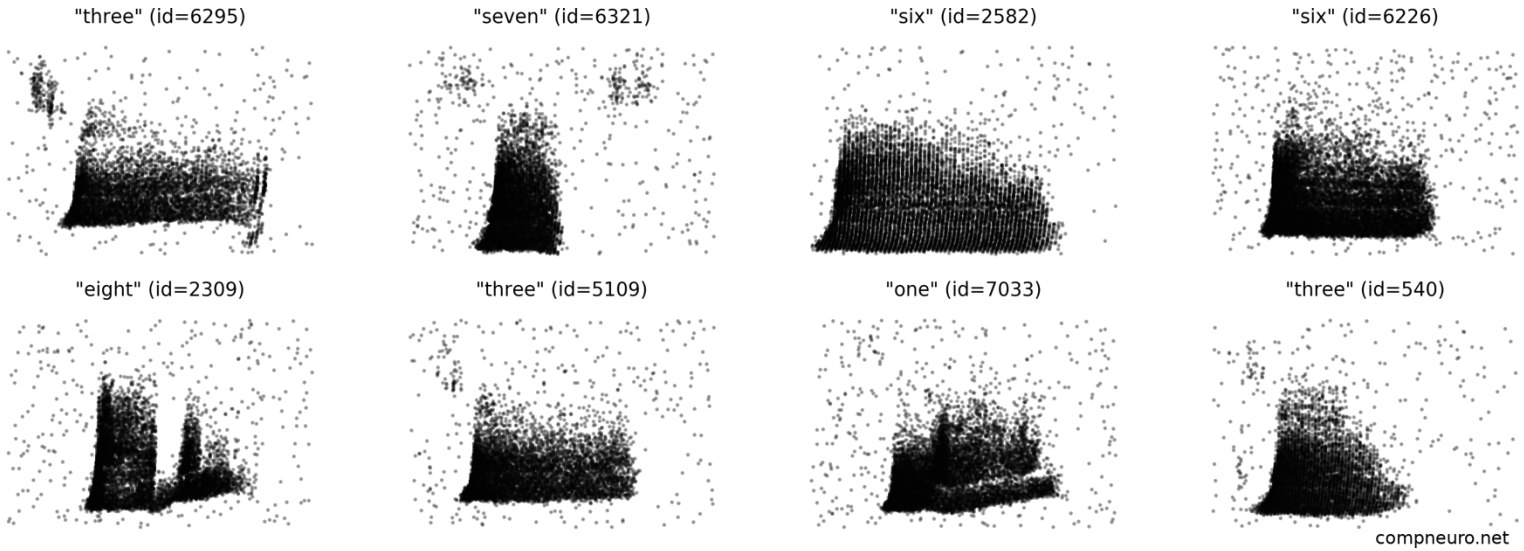
Class 1




0ms->5ms
0ms
0ms
0ms->5ms

Improved results on YY, SHD and SSC

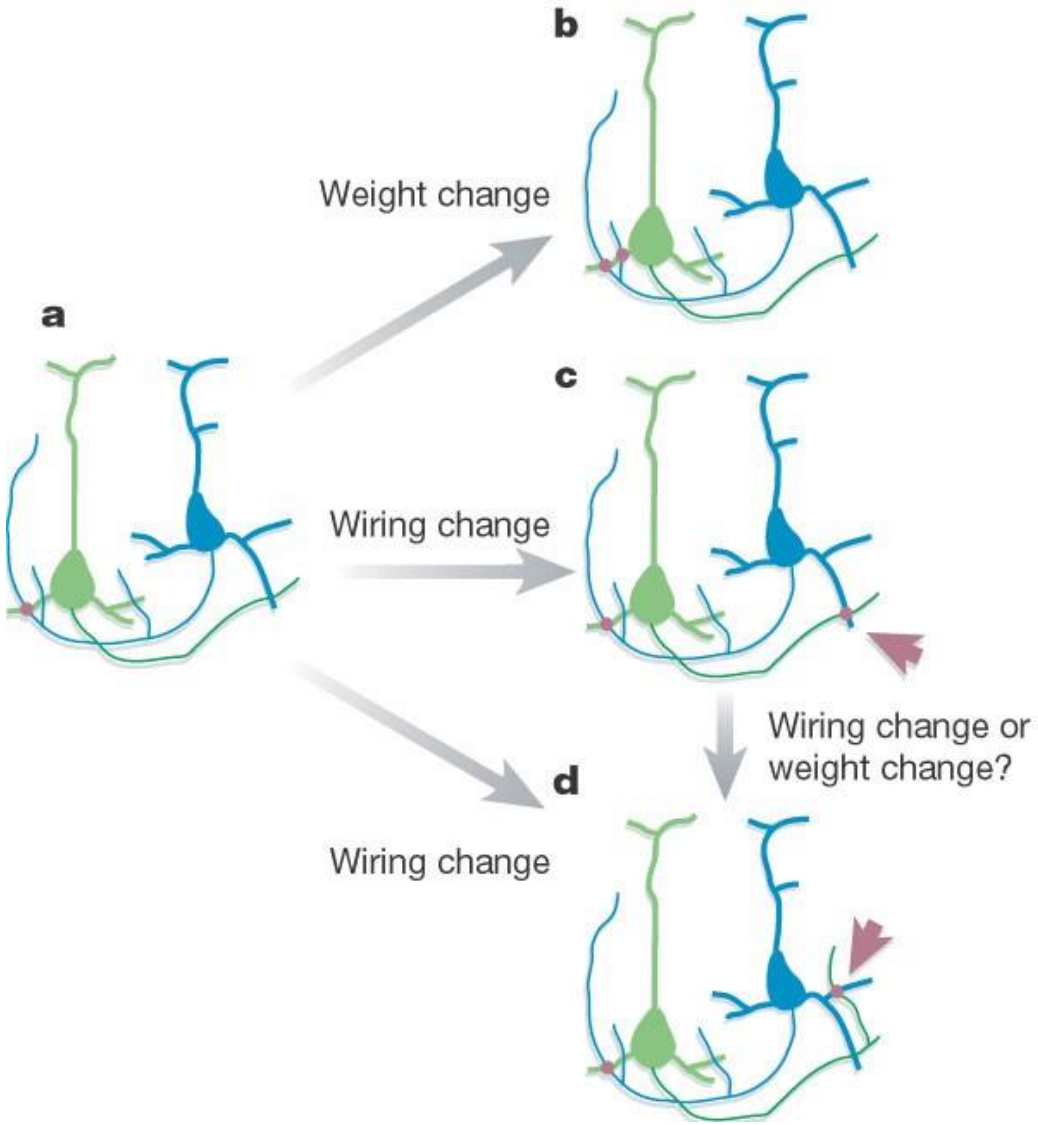




Kriener, L., Göltz, J., & Petrovici, M. A. (2022, March). The yin-yang dataset. In *Proceedings of the 2022 Annual Neuro-Inspired Computational Elements Conference* (pp. 107-111).

Cramer, B., Stradmann, Y., Schemmel, J., & Zenke, F. (2020). The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(7), 2744-2757.

Structural Plasticity

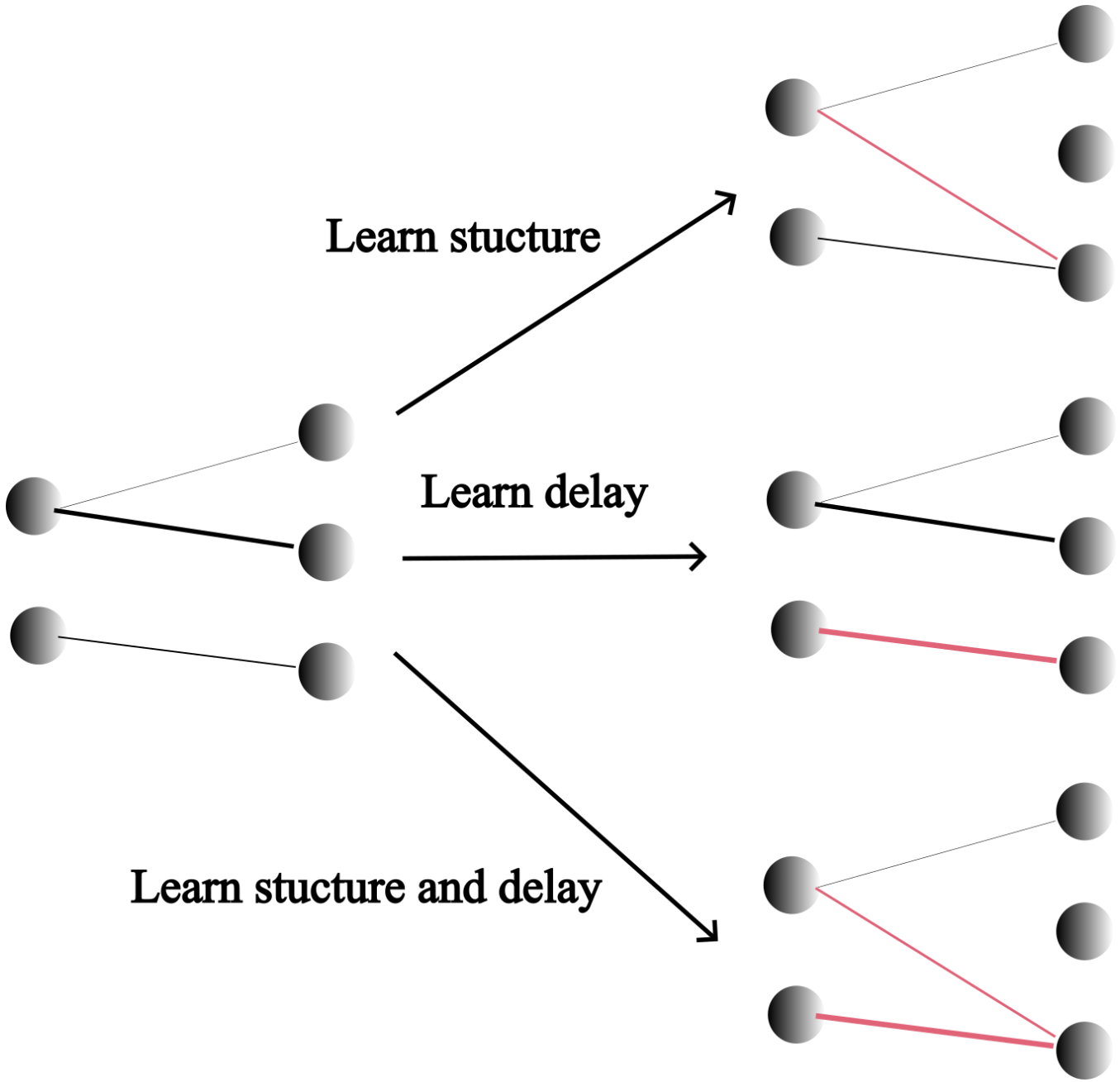


Chklovskii, D. B., Mel, B. W., & Svoboda, K. (2004). Cortical rewiring and information storage. *Nature*, *431*(7010), 782-788.

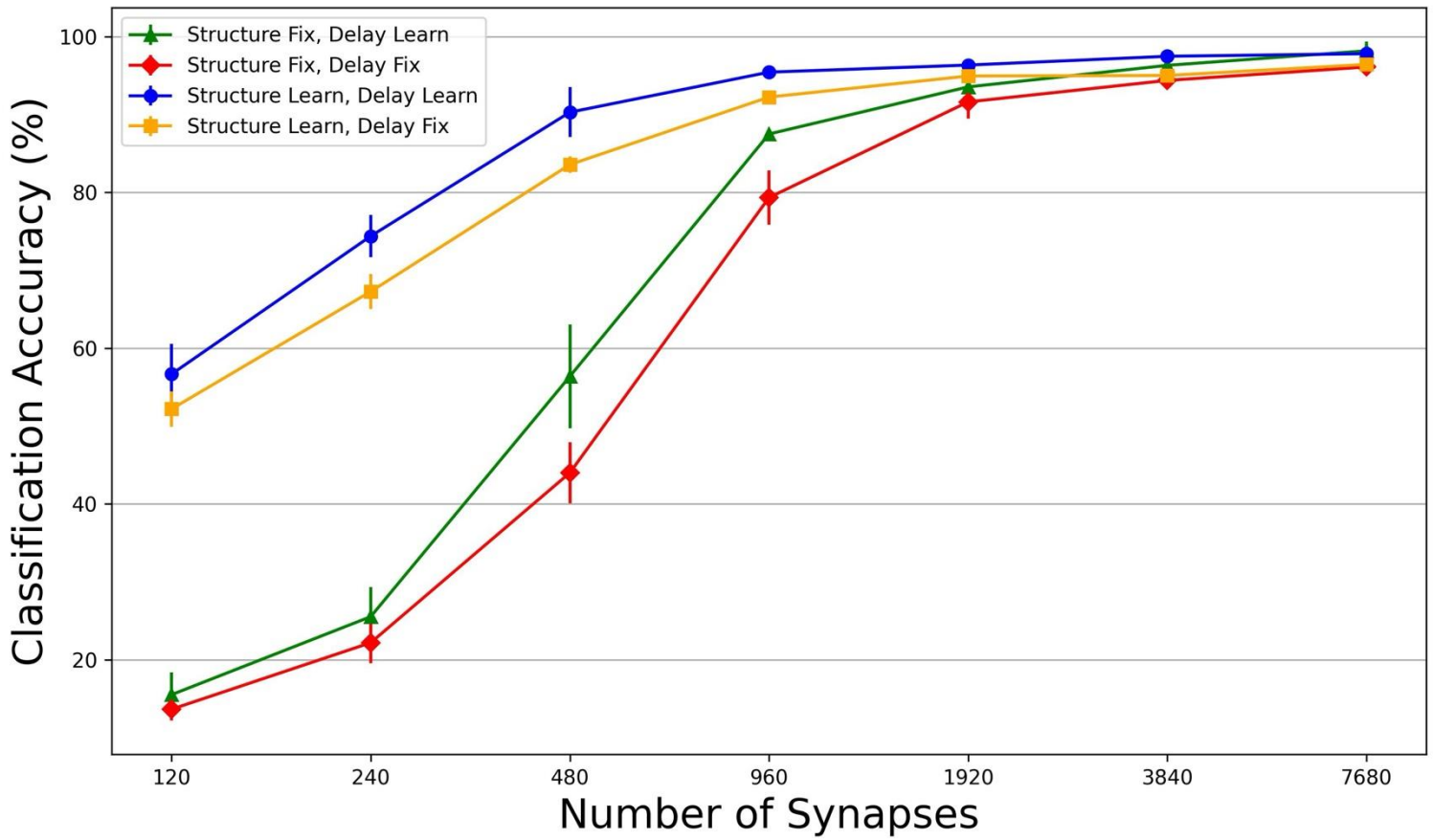
DEEP R

for i *in* $[1, N_{iterations}]$ **do**
  for *all active connections* k $(\theta_k \geq 0)$ **do**
    $\theta_k \leftarrow \theta_k - \eta \frac{\partial}{\partial \theta_k} E_{\mathbf{X}, \mathbf{Y}^*}(\boldsymbol{\theta}) - \eta\alpha + \sqrt{2\eta T}\, \nu_k$;
    if $\theta_k < 0$ **then** set connection k dormant ;
  end
  while *number of active connections lower than* K **do**
    select a dormant connection k' with uniform probability and activate it;
    $\theta_{k'} \leftarrow 0$
  end
end

Bellec, G., Kappel, D., Maass, W., & Legenstein, R. (2017). Deep rewiring: Training very sparse deep networks. *arXiv preprint arXiv:1711.05136*.



Mészáros, B., Knight, J., & Nowotny, T. (2024). Learning Delays Through Gradients and Structure: Emergence of Spatiotemporal Patterns in Spiking Neural Networks. *arXiv preprint arXiv:2407.18917*.



Mészáros, B., Knight, J., & Nowotny, T. (2024). Learning Delays Through Gradients and Structure: Emergence of Spatiotemporal Patterns in Spiking Neural Networks. *arXiv preprint arXiv:2407.18917*.

Future plans

- Combine ”long term” and ”short term” delay learning
  - Based on a heuristic we could decide if a synapse should be dropped and a new one should be introduced
    - This could be interpreted in a densely connected network as well, we can randomly sample a new delay value for the same connection
    - Or we could even introduce multiple connections between pairs of neurons with different delay values
  - The delays could still be tuned with high precision through gradient updates
- Explore recurrent delays with a bias: delays defined by the spatial distance of cells