CWI
Research institute for mathematics & computer science in the Netherlands

SpikingGamma

Surrogate-Gradient Free and Temporally Precise Online Training of Spiking Neural Networks with Smoothed Delays

June 11, 2026

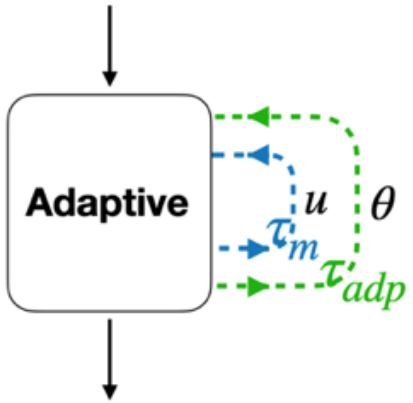
Roel Koopman
CWI

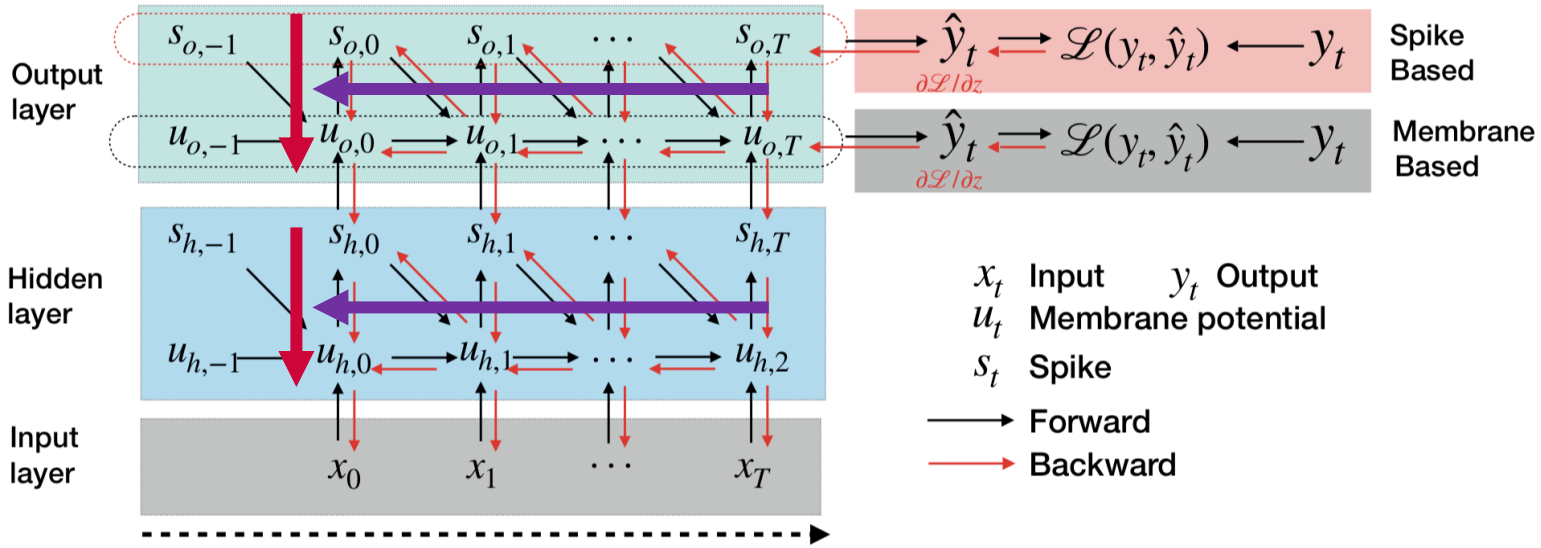
Sebastian Otte
University of Lübeck

Sander Bohté
CWI

The common way of training SNNs

- Treat neurons as (self-)recurrent neural units
- Backpropagate through spikes with surrogate gradients
- Backpropagate through time like in an RNN





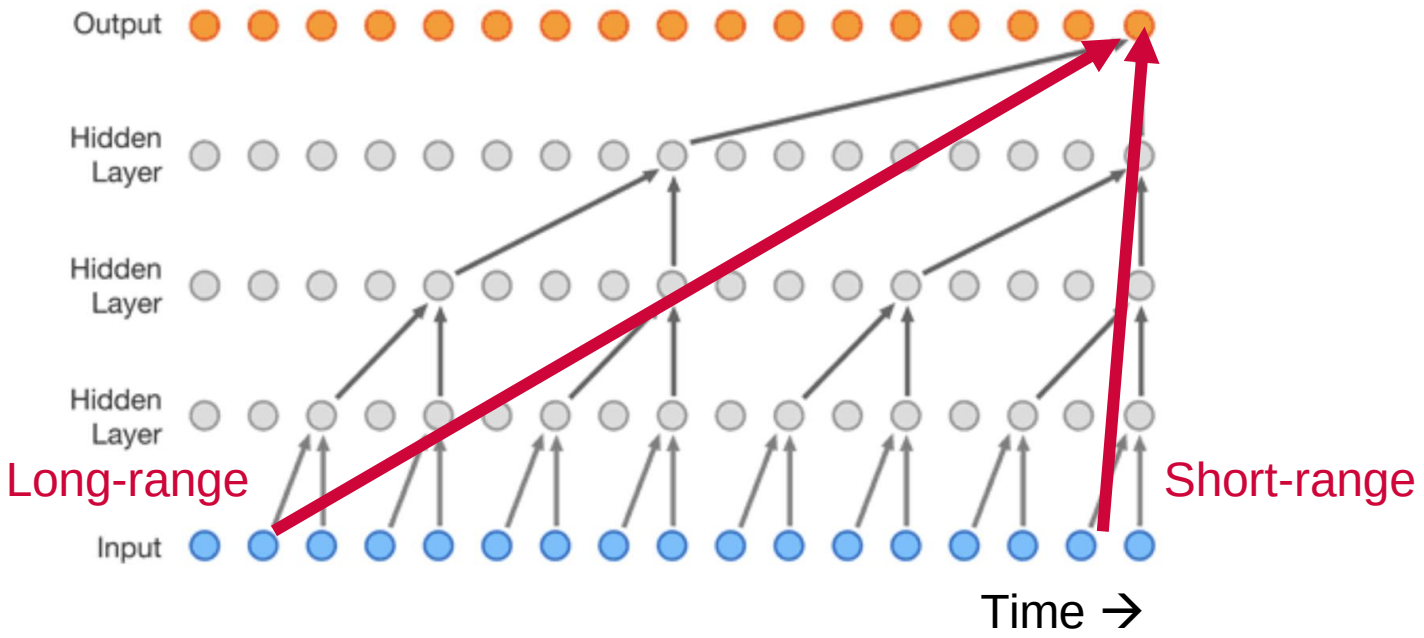
Yin et al. “Accurate and efficient time-domain classification with adaptive spiking recurrent neural networks” 2021 Nature Machine Intelligence

How this results in problems

- BPTT:
  - Extremely memory hungry
  - Exploding/Vanishing gradients
- Surrogate gradients:
  - Approximation errors accumulate
  - Sparseness amplifies gradient vanishing
- These issues add up, limiting scalability, sparsity, and compatibility with real neuromorphic hardware.
- Conclusion: **our current training method needs revisiting**

In contrast: WaveNet

- Timeseries: WaveNet / Causal Convolutional Networks
  - Parallel, no BPTT, no issue with number of timestep (linear memory)

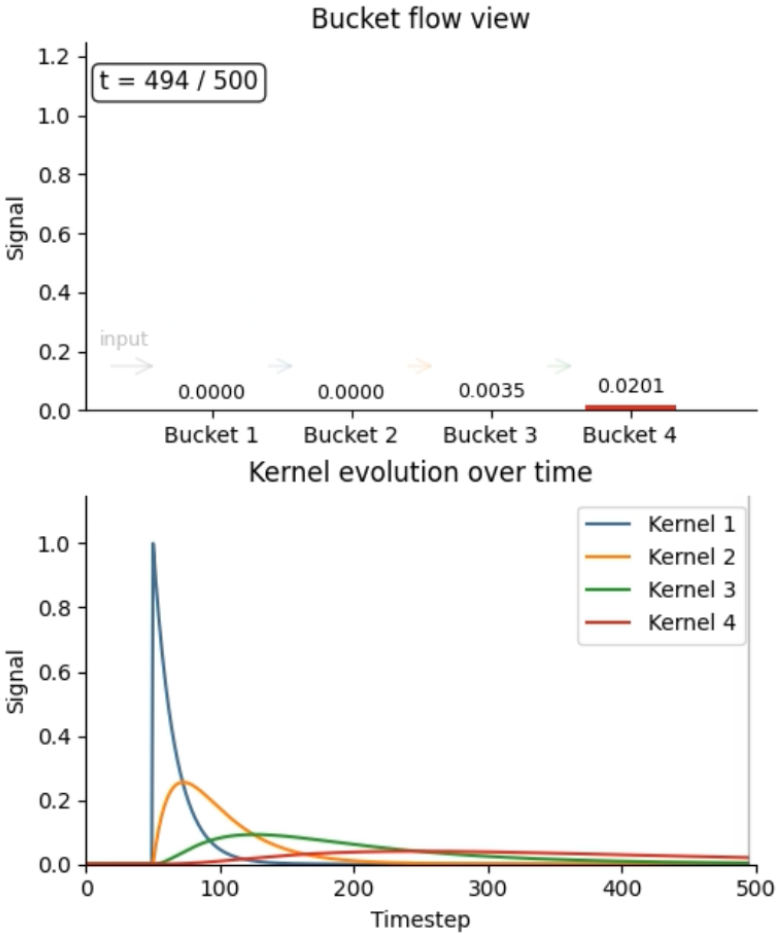


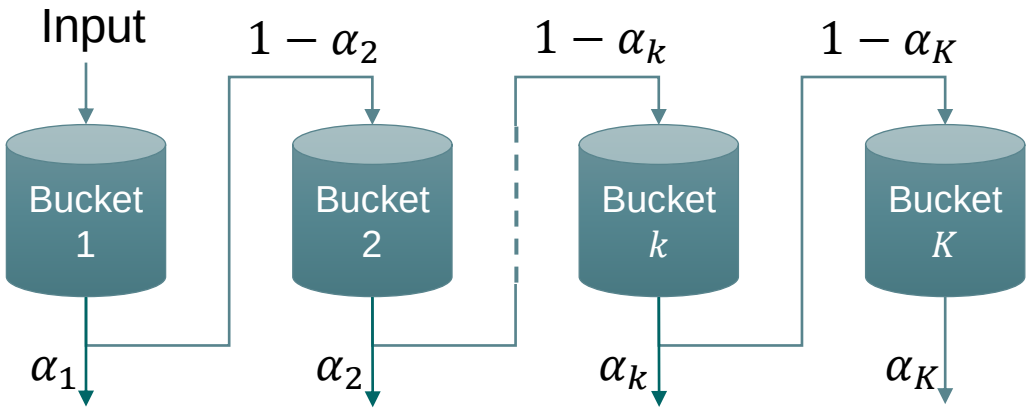
Rethage et al. “A wavenet for speech denoising” 2018 ICASSP
Borovykh et al. “Conditional time series forecasting with convolutional neural networks” 2017 ICANN

Efficient delays: the Gamma Model

- How to efficiently maintain information over different timescales inside a neuron, in a timestep independent manner?
  - Gamma-net (DeVries et al, 1992) / Bucket-model (Drew & Abbott, 2006)

Kernel is a smoothed delay

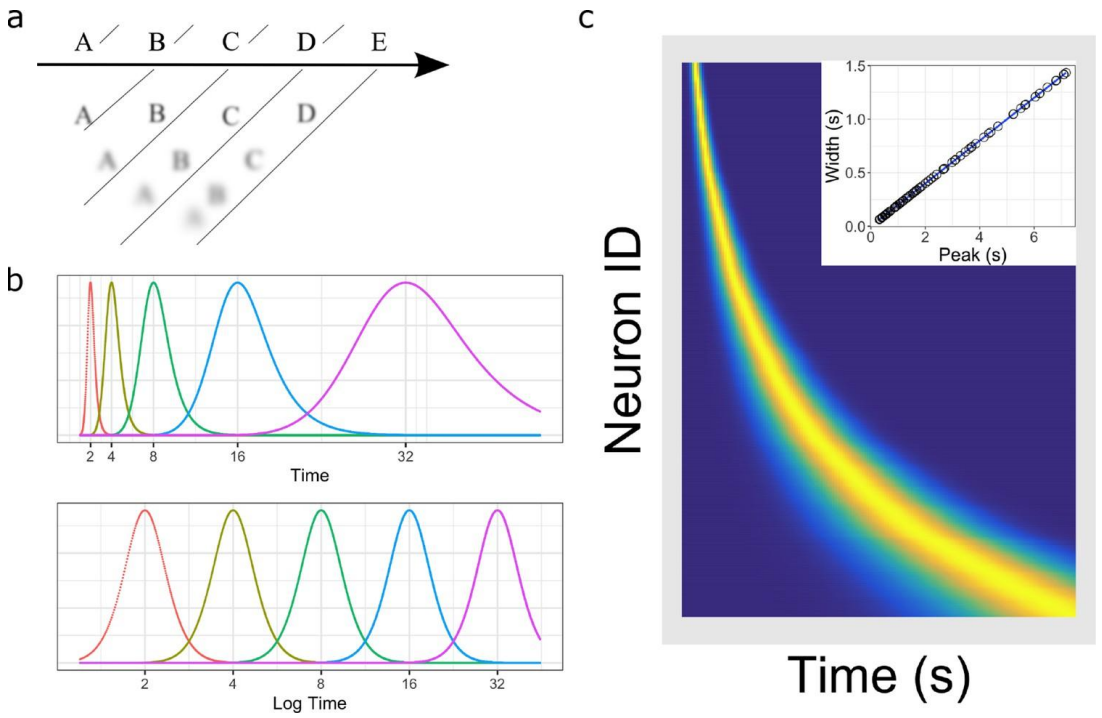




$$k = 1 \quad \text{Bucket}_1[t] = \text{Input}[t] + \text{Bucket}_k[t-1] \cdot \alpha_k$$

$$k > 1 \quad \text{Bucket}_k[t] = \text{Bucket}_{k-1}[t-1] \cdot (1-\alpha_k) + \text{Bucket}_k[t-1] \cdot \alpha_k$$

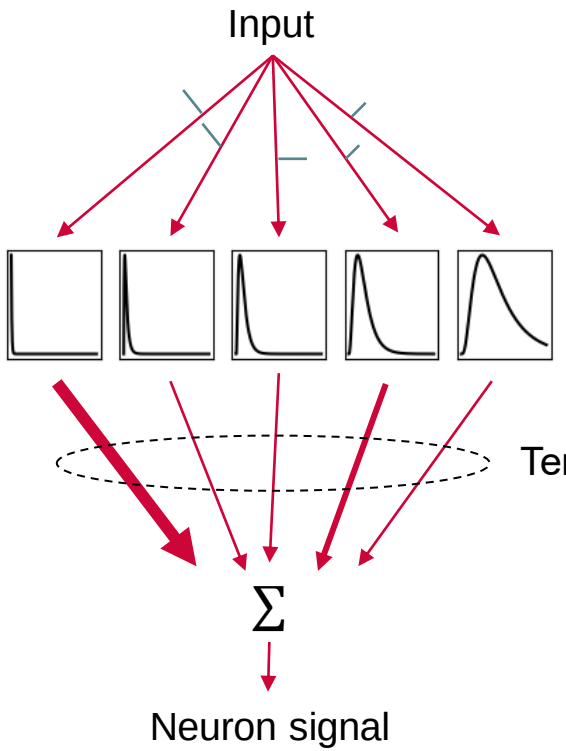
Ad lib: time in our brain

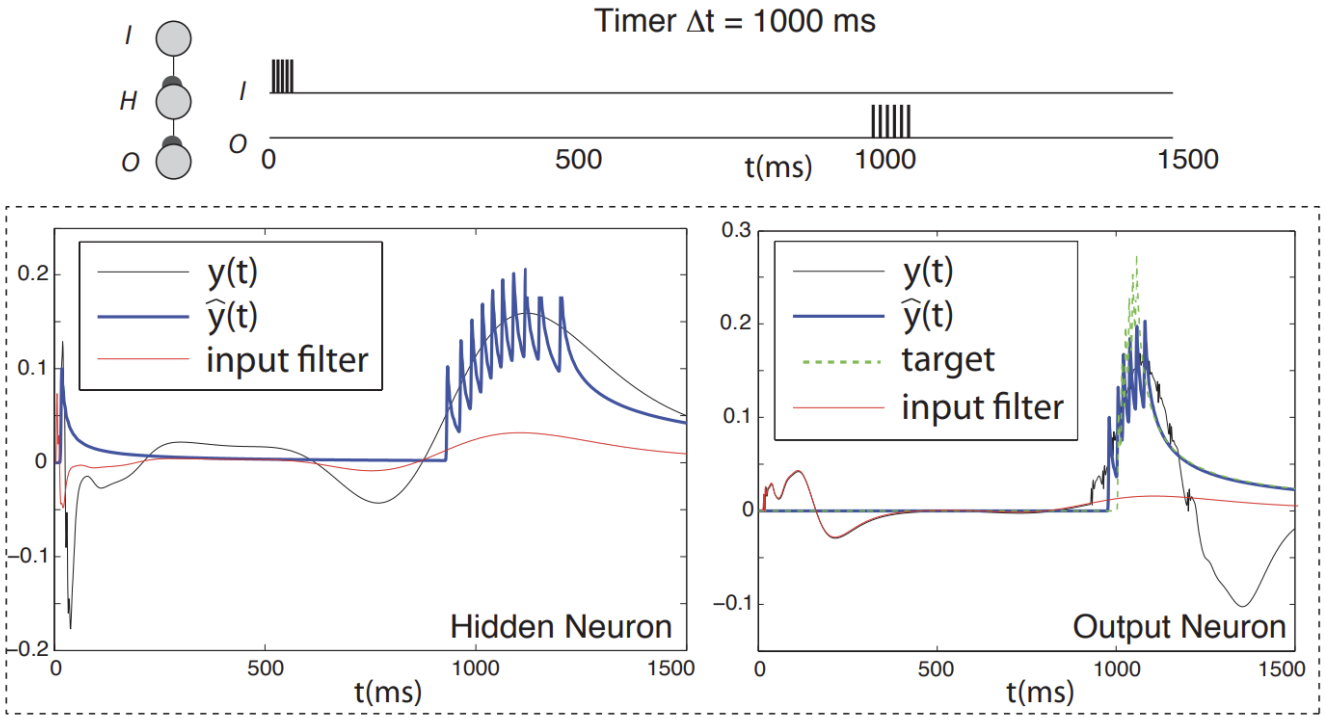


- Hippocampus research
- Information decays rapidly at first, then more gradually over time **(a)**
- Evidence for neurons / “time cells” that respond at different timescales **(b,c)**

Cao et al. "Internally generated time in the rodent hippocampus is logarithmically compressed." 2022 eLife

Related work





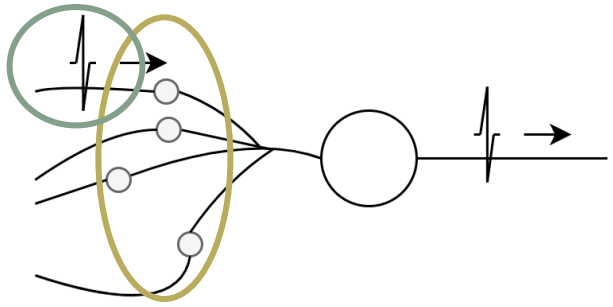
Bohte, S. M. (ICANN, 2011)

- Tiny MLP networks, not efficiently computable

CWI
Research institute for mathematics & computer science in the Netherlands

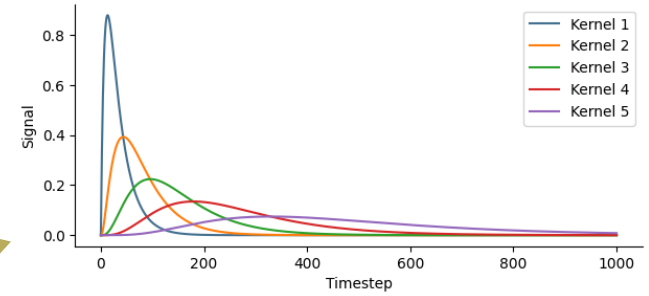
Introducing our model

Starting at the synapse

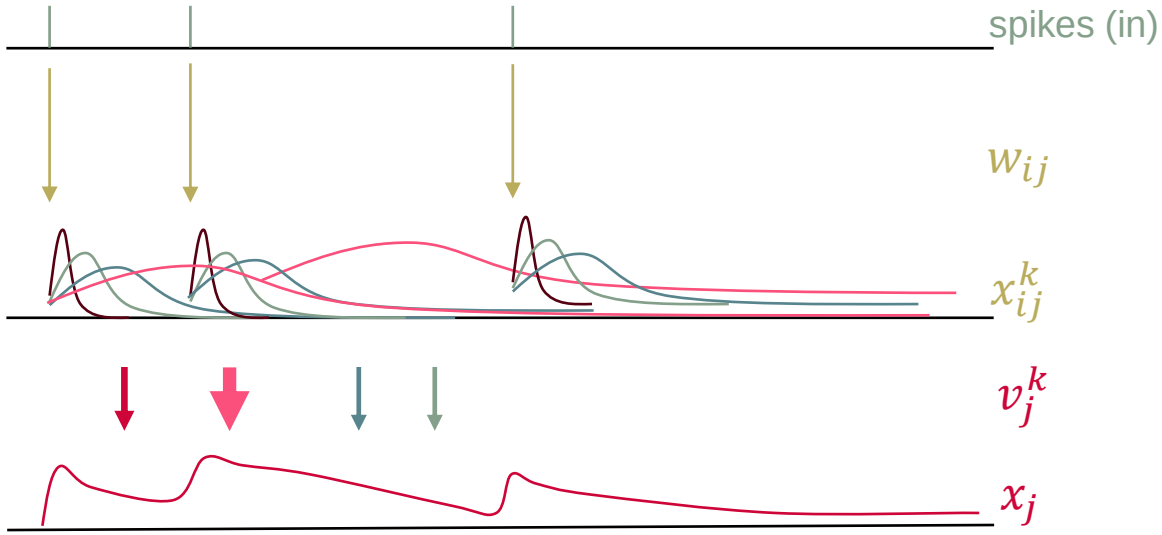


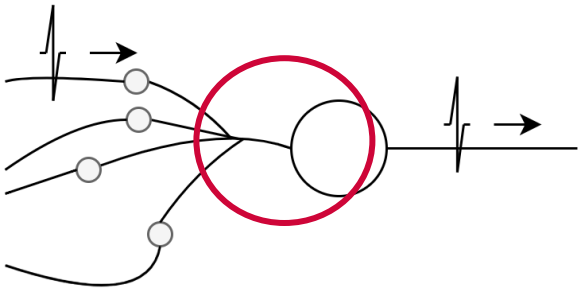
From neuron i to j,
for kernel $1 \leq k \leq K$:

$$x_{ij}^{k} = \sum_{t<t_i} \kappa^{k}(t - t_i) \cdot w_{ij}$$



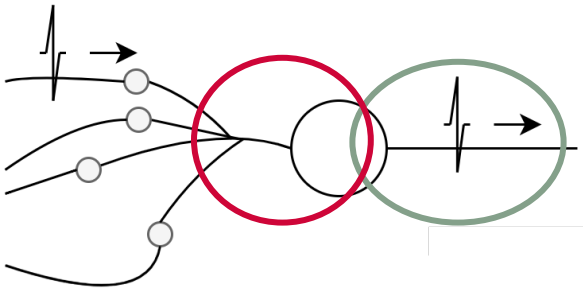
"A representation of the past in the spatial domain" akin to a temporal convolution





$$x_j = \sum_{k} \sum_{i} x_{ij}^{k} \cdot v_j^{k}$$

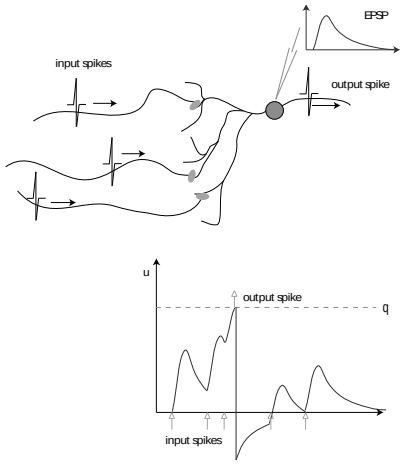
"What segment of the past (i.e., which kernel), is most important for this neuron?"

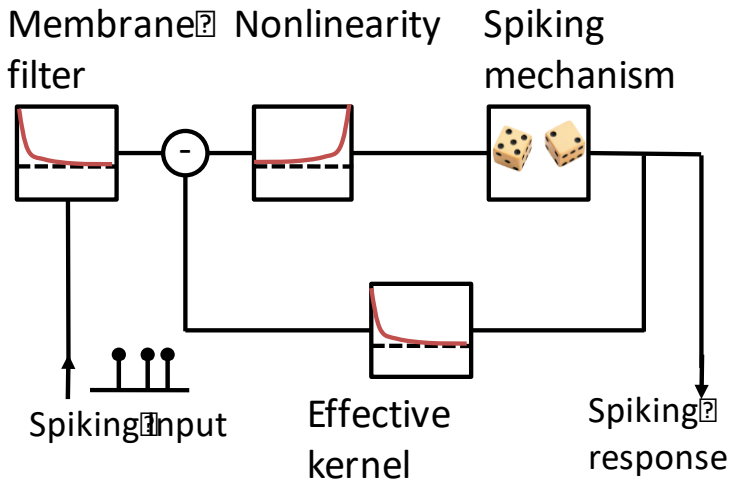


$$y_j = ReLU(x_j)$$

Negative part of x_j is rectified since binary spikes can only signal one direction

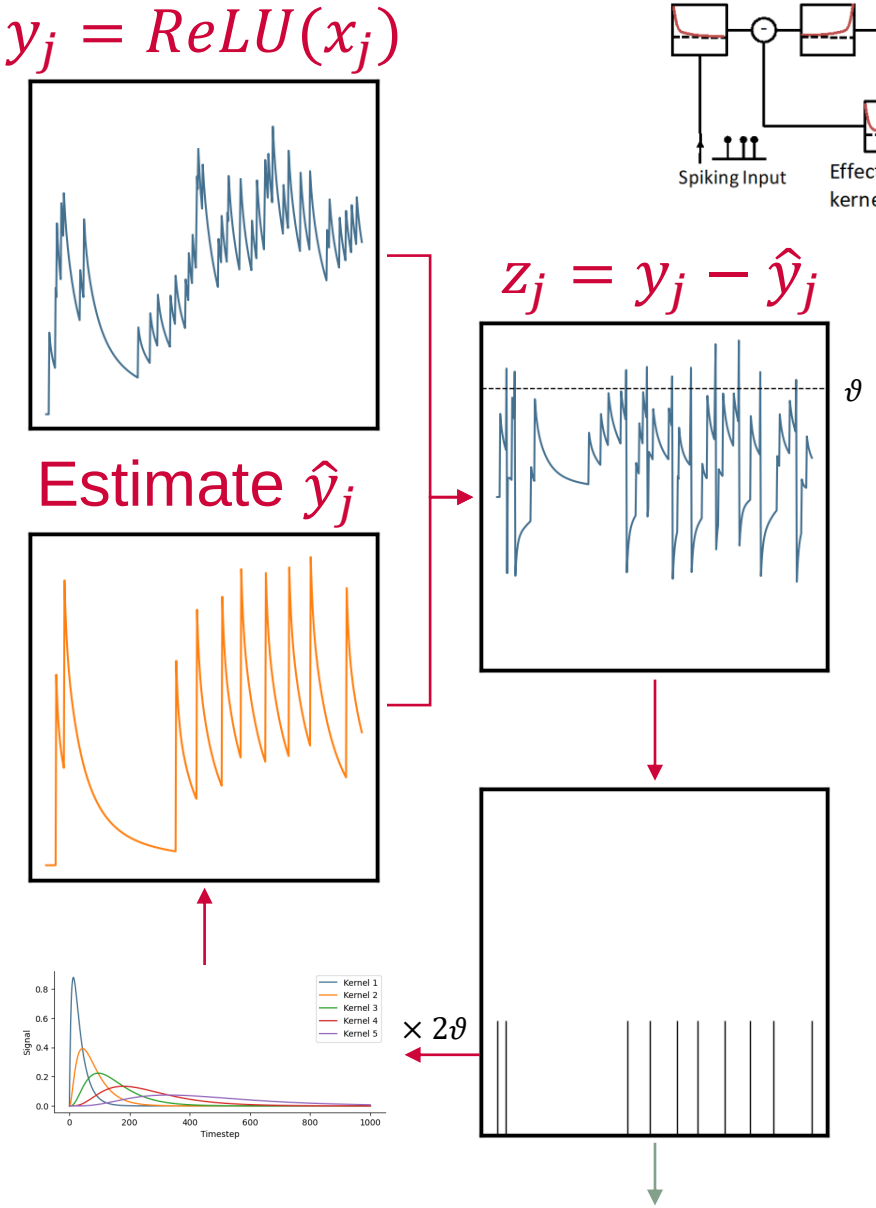
A Spiking Neuron

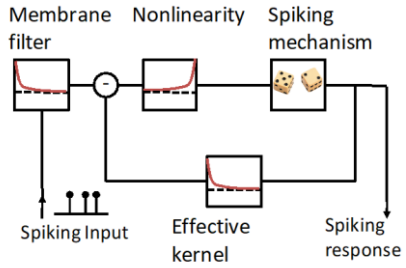


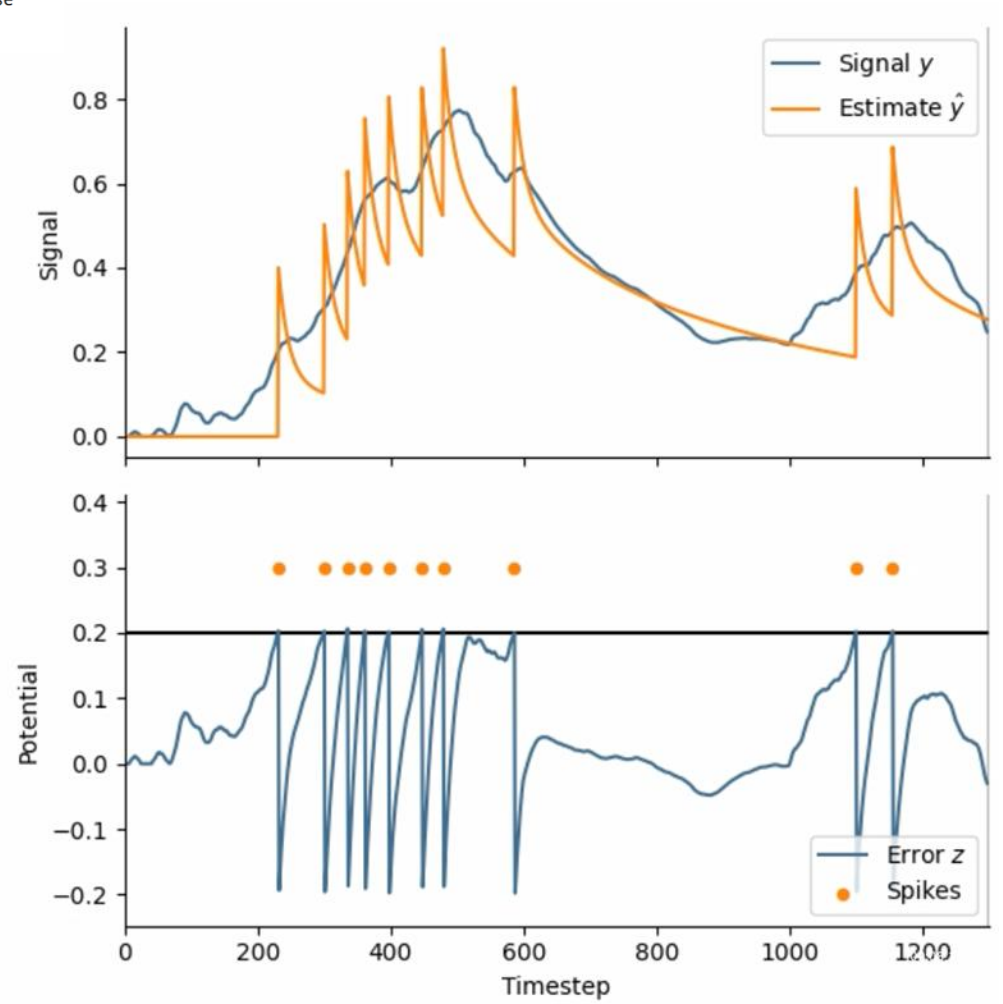


AD/DA = Analog Digital / Digital Analog converter

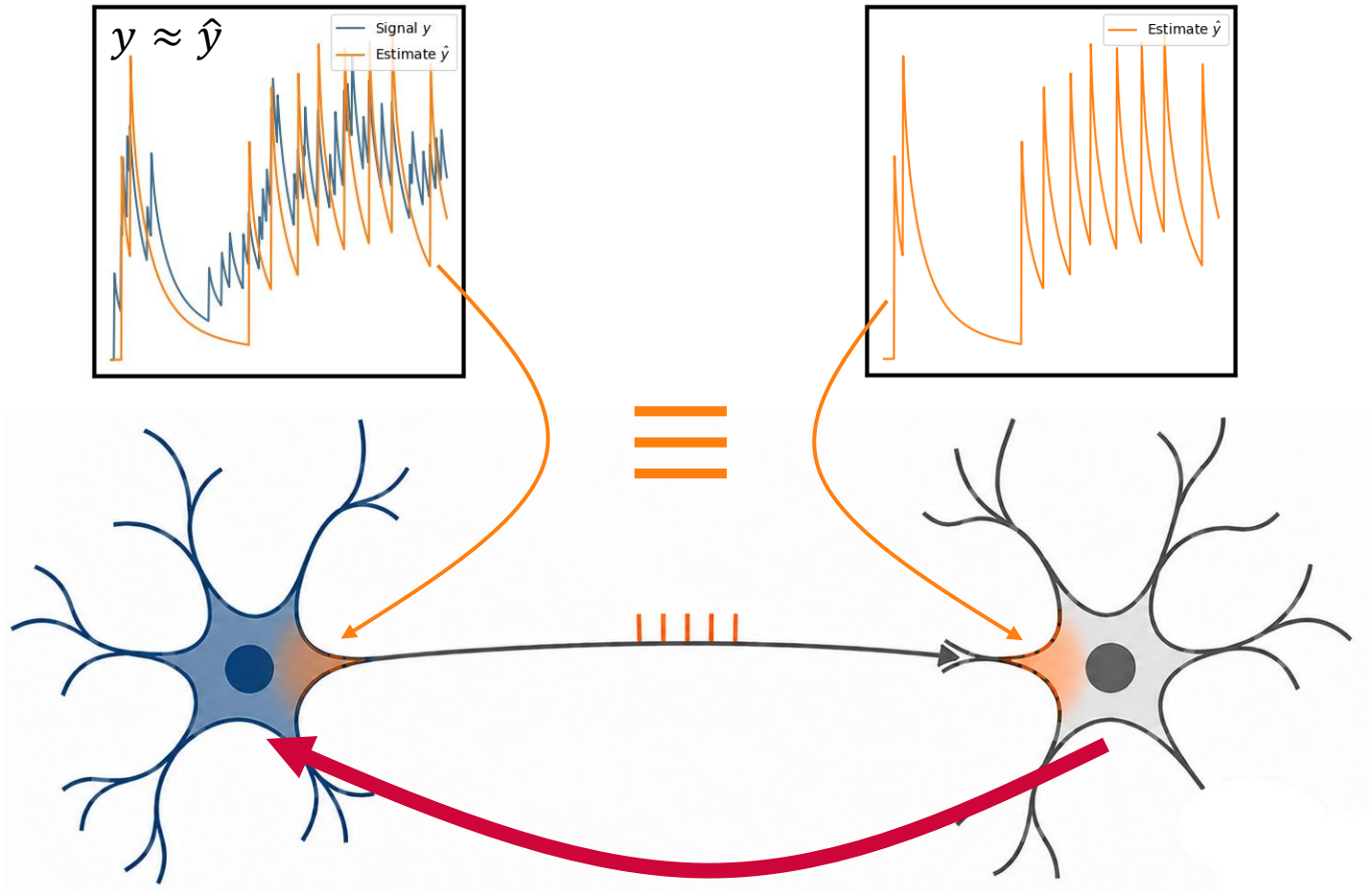
From analog back to spikes using Σ−Δ coding







Circumventing spikes in backward path

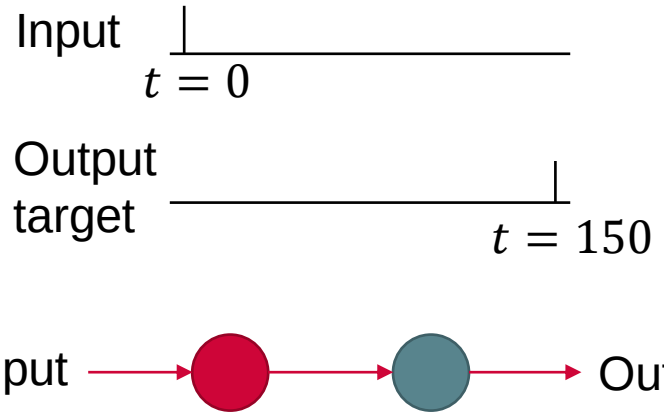


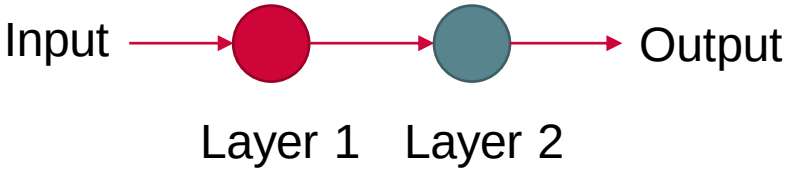
Straight Through Estimator $\hat{y} \to y$

Since $\hat{y}_j(t)$ estimates $y_j(t)$, we have $\frac{\partial \hat{y}_j(t)}{\partial y_j(t)} \approx 1$, and thus no need to go backward through spikes.

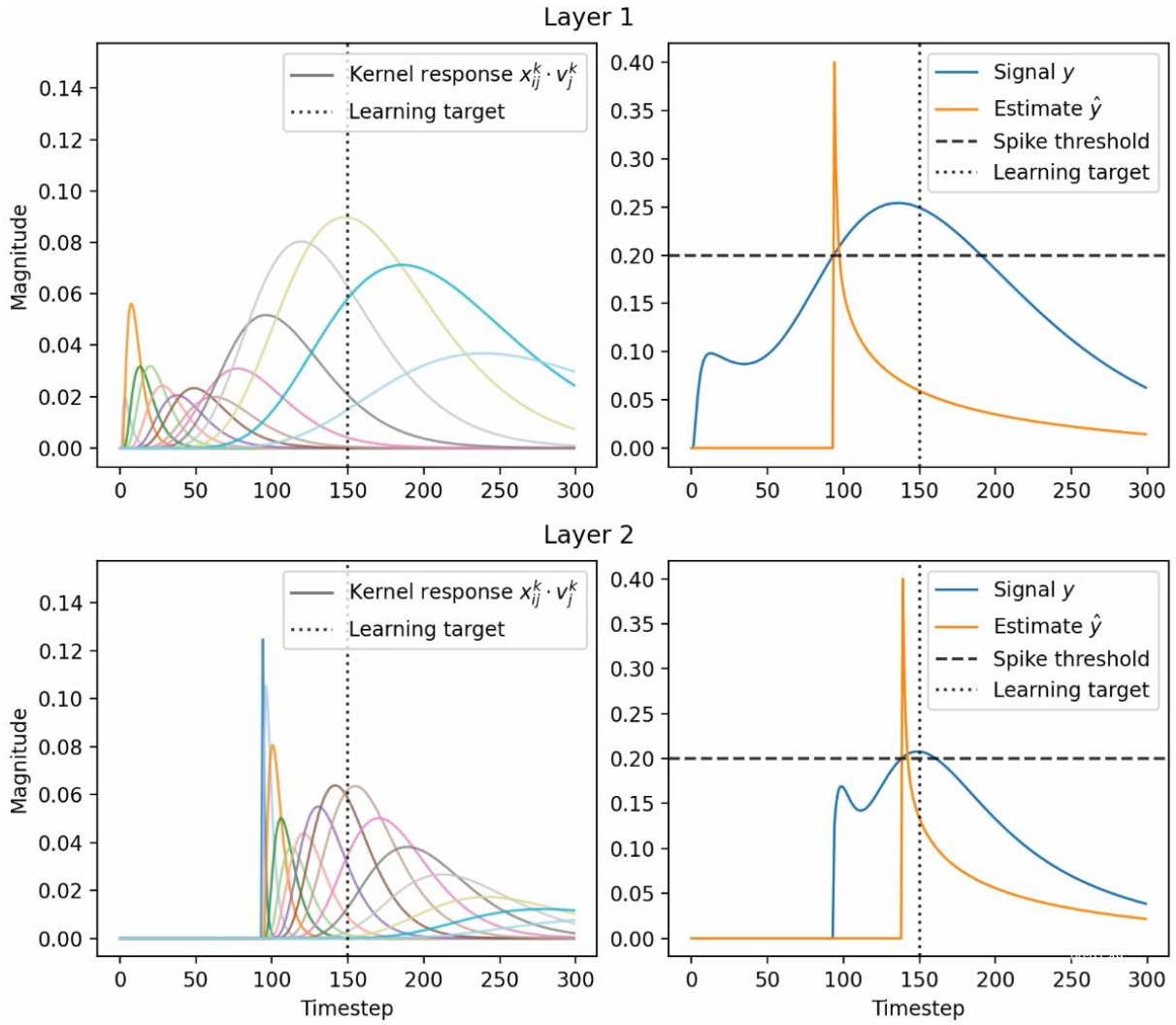
Insensitive to timestep size

Temporal credit assignment in sparse networks

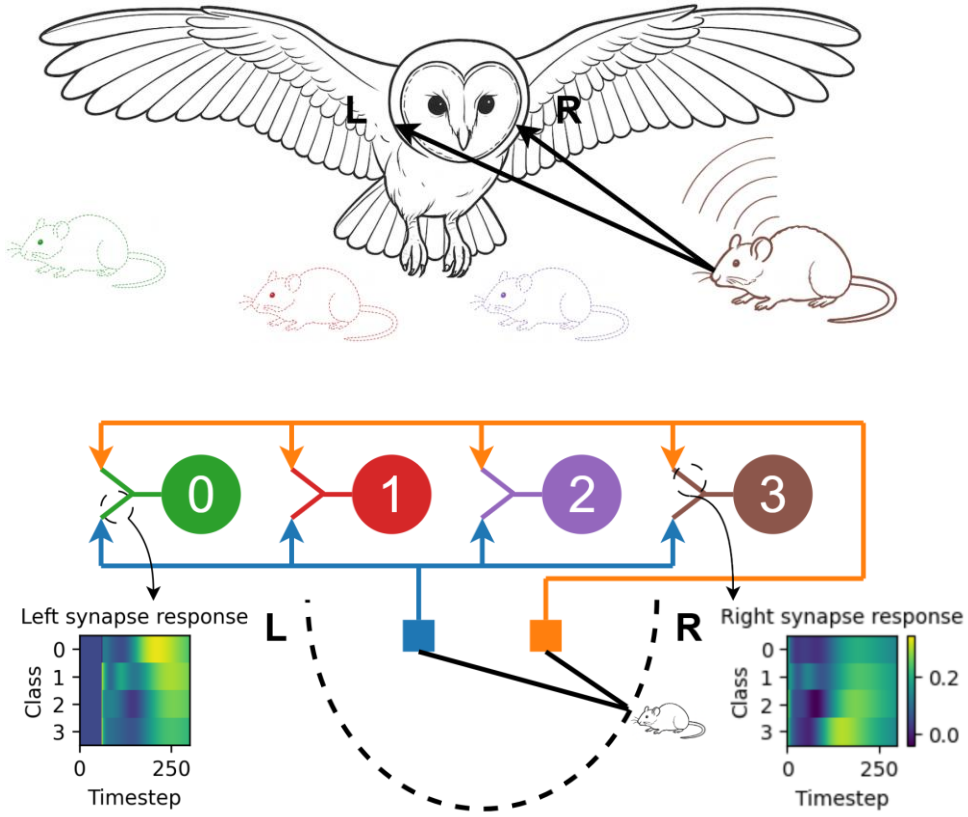




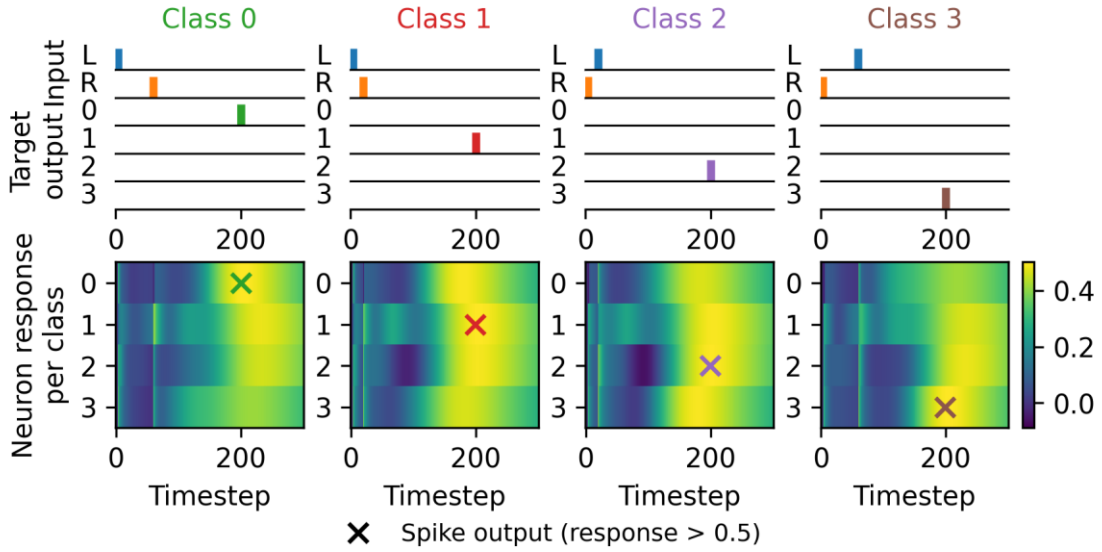
Each layer: 1 neuron, 15 buckets, and 15 trainable bucket weights



Learning coincidence detection



Each **synapse** has 25 buckets and 25 trainable bucket weights

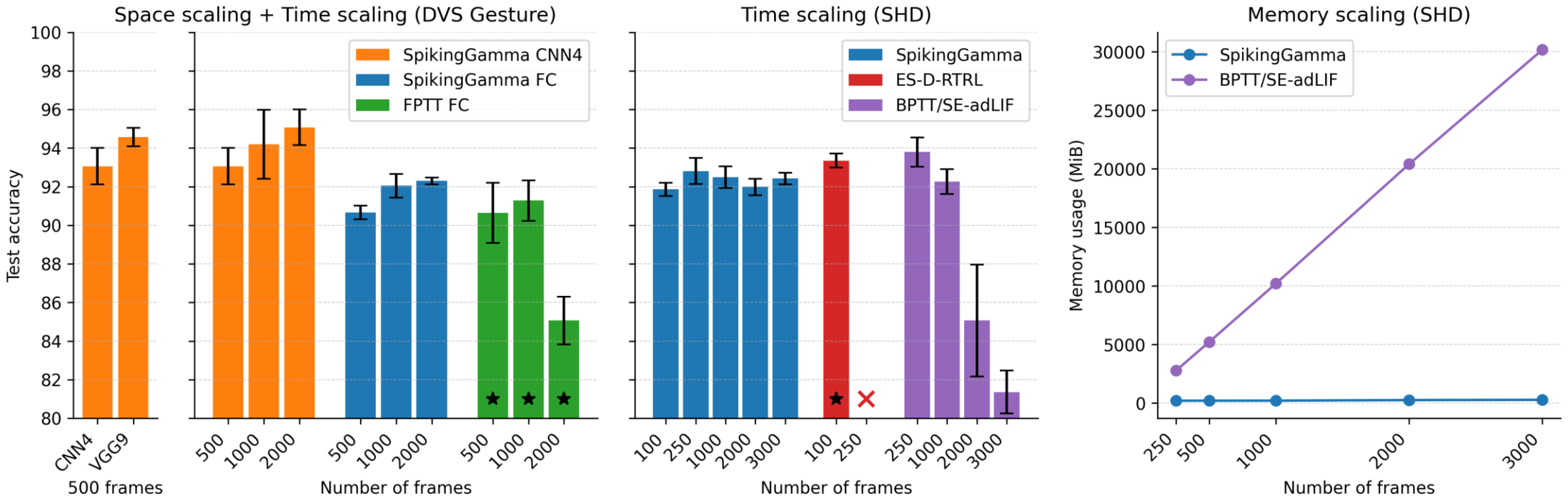


CWI
Research institute for mathematics & computer science in the Netherlands

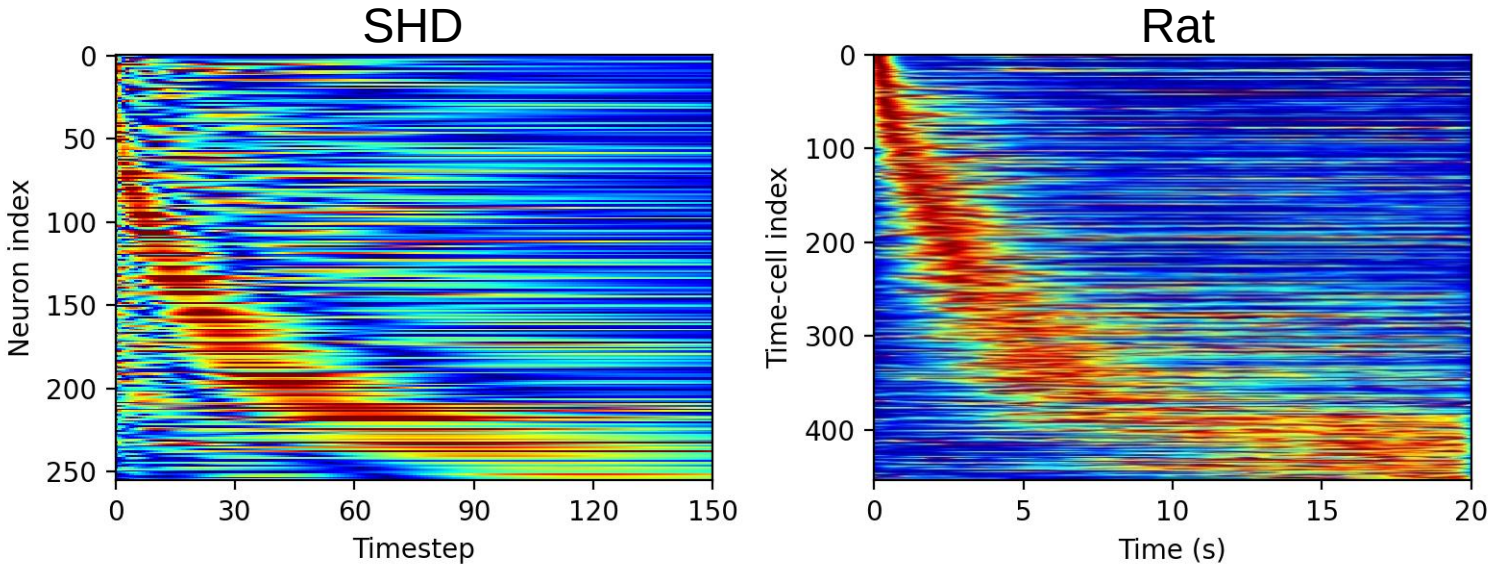
Scaling through space and time

Task	Frames	Method	Architecture	Accuracy ± std (peak)
DVS Gesture	1000	FPTT	2-layer FC, LTC	- (91.28 ± 1.05)
	500	DECOLLE	4-layer CNN, CUBA	- (95.54 ± 0.16)
	2000	SpikingGamma	2-layer FC	**92.30 ± 0.18 (93.81 ± 0.18)**
	2000	SpikingGamma	4-layer CNN	**95.08 ± 0.93 (96.08 ± 0.18)**
SHD	50	OTTT	[512 × 3], LIF	71.2 ± 0.8 (-)
	50	OSTL	[512 × 3], LIF	70.6 ± 0.7 (-)
	50	OTPE	[512 × 3], LIF	75.4 ± 0.5 (-)
	100	DECOLLE	[450], ALIF	62.01 ± 0.61 (-)
	100	ES-D-RTRL	[1024 × 3], RadLIF	- (93.35 ± 0.36)
	100	e-prop	[450], ALIF	80.79 ± 0.39 (-)
	250	SpikingGamma	[256 × 3]	**92.81 ± 0.68 (93.55 ± 0.48)**
SSC	100	ES-D-RTRL	[1024 × 3], RadLIF	- (66.63 ± 0.53)
	250	SpikingGamma	[512 × 3]	**75.63 ± 0.44 (75.91 ± 0.54)**

Further scaling through time



Bio-plausible neuron response dynamics



- Left: The neuron responses to a stimulus at $t = 0$ after training on SHD. Sorted by peak response time.
- Right: Firing patterns of rats' time-cells during a temporal bisection task.

Shimbo et al. "Scalable representation of time in the hippocampus" 2021 Science Advances

CWI
Research institute for mathematics & computer science in the Netherlands

Conclusion

- We introduced a new neuron model and learning method that:
  - Is less susceptible to gradient vanishing caused by sparsity
  - Has no memory problems
  - Is fully online
  - Can train delays
  - Can scale

- Preprint is available on arxiv
- Source code will be available soon

https://arxiv.org/abs/2602.01978