NICE 2025

Heidelberg (Germany)

A LIF-based Legendre Memory Unit as neuromorphic State Space Model benchmarked on a second-long spatio-temporal task

Benedetto Leto, Gianvito Urgese, Enrico Macii, **Vittorio Fra**

Politecnico di Torino

25 March 2025

The LMU architecture

Legendre Memory Unit - I

Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks

Aaron R. Voelker[1,2] Ivana Kajić[1] Chris Eliasmith[1,2]

[1]Centre for Theoretical Neuroscience, Waterloo, ON [2]Applied Brain Research, Inc.
{arvoelke, i2kajic, celiasmith}@uwaterloo.ca

Voelker et al., Advances in Neural Information Processing Systems 32 (NeurIPS 2019)

The LMU foundation is in the shifted Legendre polynomials

$$P_i(x) = (-1)^i \sum_{j=0}^{i} \binom{i}{j}\binom{i+j}{i}(-x)^i$$

which are used as basis to project a signal $u(t)$ with temporal delay $u(t-\theta')$ within a sliding window of duration θ with $0 \le \theta' \le \theta$:

$$u(t-\theta') \approx \sum_{i=0}^{d-1} P_i\left(\frac{\theta'}{\theta}\right) m_i(t)$$

The highest order $(d-1)$ in the series expansion is related to the dimension d of the memory state-vector $\mathbf{m}(t)$ defined as

$$\theta\dot{\mathbf{m}}(t) = \mathbf{A}\mathbf{m}(t) + \mathbf{B}u(t)$$

with $\mathbf{A}$ and $\mathbf{B}$ representing the ideal state-space matrices determined using the Padé approximants:

$$\mathbf{A} = [a]_{ij} \in \mathbf{R}^{d\times d}; \quad a_{ij} = (2i+1)\begin{cases} -1, & i<j \\ (-1)^{i-j+1}, & i \ge j \end{cases}; \qquad \mathbf{B} = [b]_i \in \mathbf{R}^d; \quad b_i = (2i+1)(-1)^i; \qquad i,j \in [0, d-1]$$

Legendre Memory Unit - II

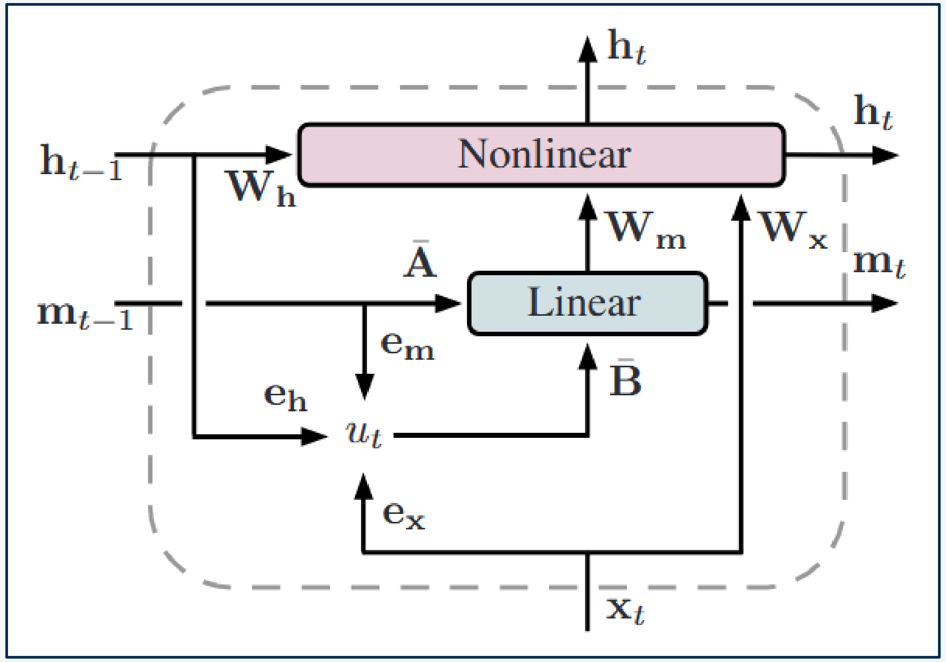
The LMU takes an input $\mathbf{x}$ and produces a hidden state $\mathbf{h}$ and a memory state $\mathbf{m}$ to define the new representation u through learnable encoding vectors $\mathbf{e}$ and learnable kernels $\mathbf{W}$.

Discretization and layer definitions:

$$\mathbf{m}_t = \bar{\mathbf{A}}\mathbf{m}_{t-1} + \bar{\mathbf{B}}u_t; \quad \bar{\mathbf{A}} = (\Delta t/\theta)\mathbf{A} + \mathbf{I}, \quad \bar{\mathbf{B}} = (\Delta t/\theta)\mathbf{B}$$

$$\mathbf{h}_t = f(\mathbf{W}_\mathrm{x}\mathbf{x}_t + \mathbf{W}_\mathrm{h}\mathbf{h}_{t-1} + \mathbf{W}_\mathrm{m}\mathbf{m}_t)$$

$$u_t = \mathbf{e}_\mathrm{x}{}^\mathrm{T}\mathbf{x}_t + \mathbf{e}_\mathrm{h}{}^\mathrm{T}\mathbf{h}_{t-1} + \mathbf{e}_\mathrm{m}{}^\mathrm{T}\mathbf{m}_{t-1}$$



Voelker et al, Advances in Neural Information Processing Systems 32 (NeurIPS 2019)

Our LIF-based LMU (L^2MU)

Leaky Integrate-and-Fire (LIF) neurons inside LMU

We redesigned the LMU architecture to **fully rely on LIF neurons**[1],[2]:

$$RC\frac{dU(t)}{dt} = -U(t) + I(t)R$$

where $U(t)$ is the **membrane potential** and $I(t)$ is the **input current**, while R and C refer to the electrical equivalent and they define the deacy constant τ_{mem} of the *leaky integrator*.

For our **L^2MU**, we used the `Leaky` neuron model available in snnTorch[3]:

$$U_t = \beta U_{t-1} + W X_t \boxed{- S_{t-1}\Theta}$$

membrane potential decay factor

input

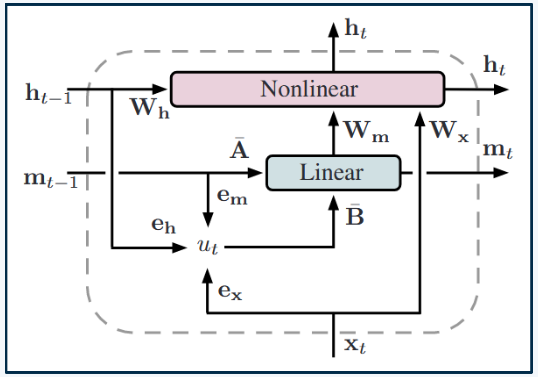
reset by subtraction

Spiking condition: $$S_t = \begin{cases} 1, & if\ U_t > \Theta \\ 0, & otherwise \end{cases}$$

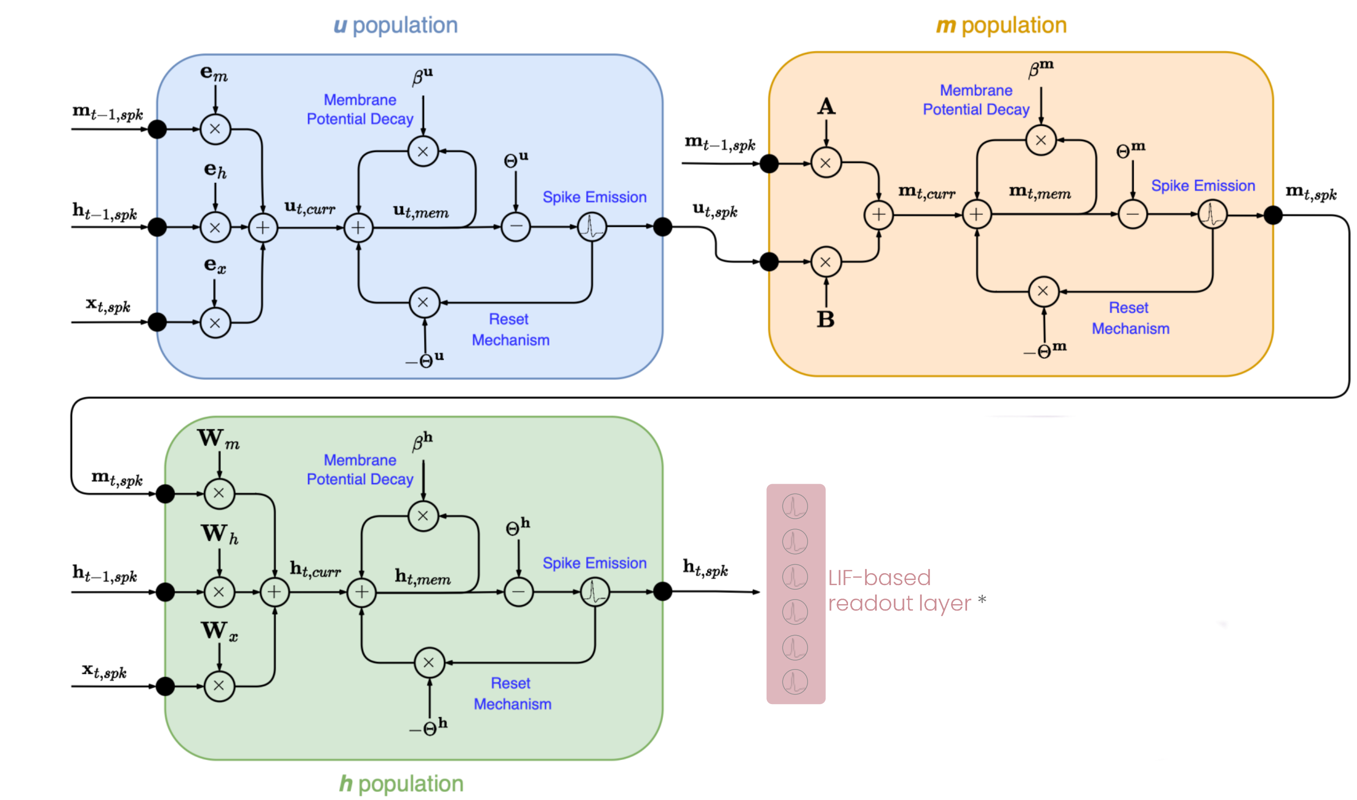
1. *Gerstner and Kistler, Cambridge University Press, 2002*
2. *Eshraghian et al., Proceedings of the IEEE, 2023*
3. *github.com/jeshraghian/snntorch/tree/master*

The L^2MU architecture

Each of the original LMU blocks is converted into a dedicated **population of LIF neurons**:



Voelker et al., Advances in Neural Information Processing Systems 32 (NeurIPS 2019)



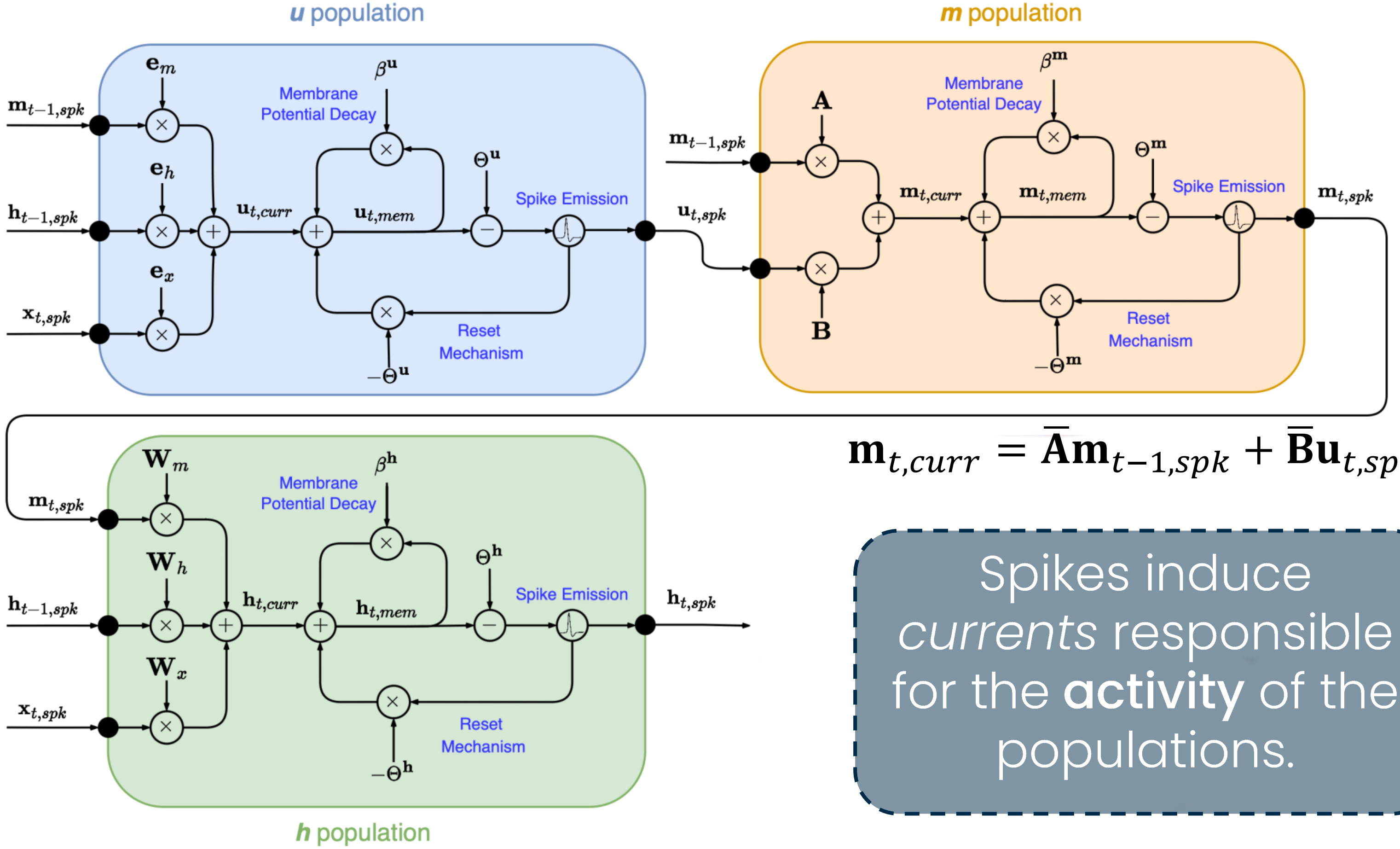
** details in the next slides*

The L²MU equations

We rewrote the original LMU equations in such a way that each block, i.e. population:

a. transmits **spikes**
b. receives **spikes**

$$\mathbf{u}_{t,curr} = \mathbf{e_x}^{\mathrm{T}}\mathbf{x}_{t,spk} + \mathbf{e_h}^{\mathrm{T}}\mathbf{h}_{t-1,spk} + \mathbf{e_m}^{\mathrm{T}}\mathbf{m}_{t-1,spk}$$



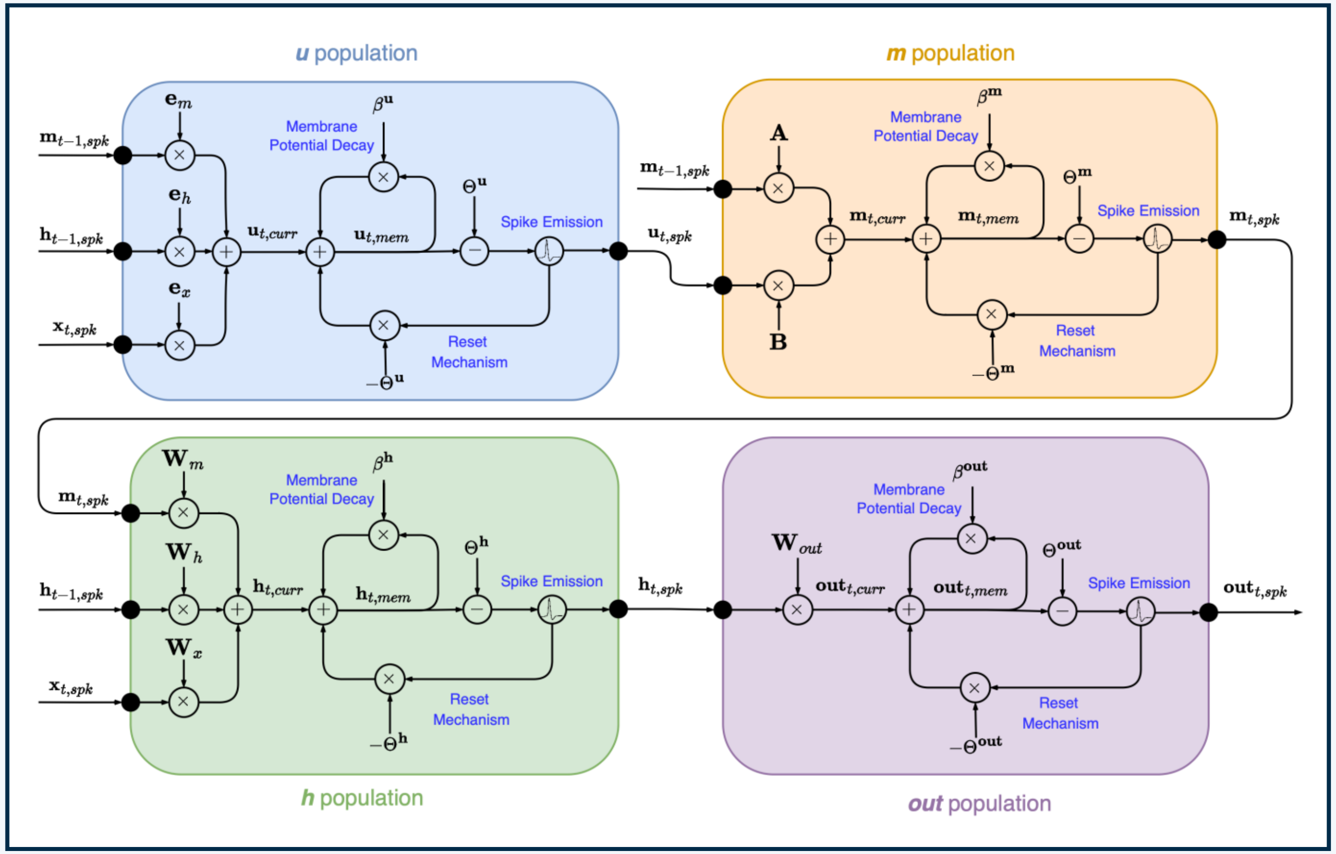
$$\mathbf{m}_{t,curr} = \bar{\mathbf{A}}\mathbf{m}_{t-1,spk} + \bar{\mathbf{B}}\mathbf{u}_{t,spk}$$

Spikes induce *currents* responsible for the **activity** of the populations.

$$\mathbf{h}_{t,curr} = \mathbf{W_x}\mathbf{x}_{t,spk} + \mathbf{W_h}\mathbf{h}_{t-1,spk} + \mathbf{W_m}\mathbf{m}_{t,spk}$$

The complete model

The readout layer is made of an additional population of LIF neurons, the ***out population***, which receives spikes from the ***h population***:



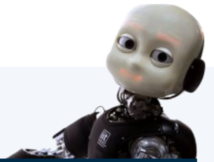
The **spiking activity** of the ***out population*** is used to make predictions

Every layer and all the populations underwent **hyperparameter optimization** through the Neural Network Intelligence toolkit[1].

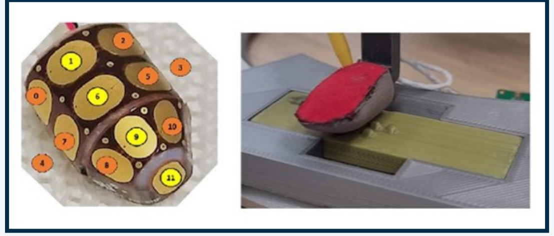
1. github.com/microsoft/nni

Braille letter reading

Second-long spatio-temporal task



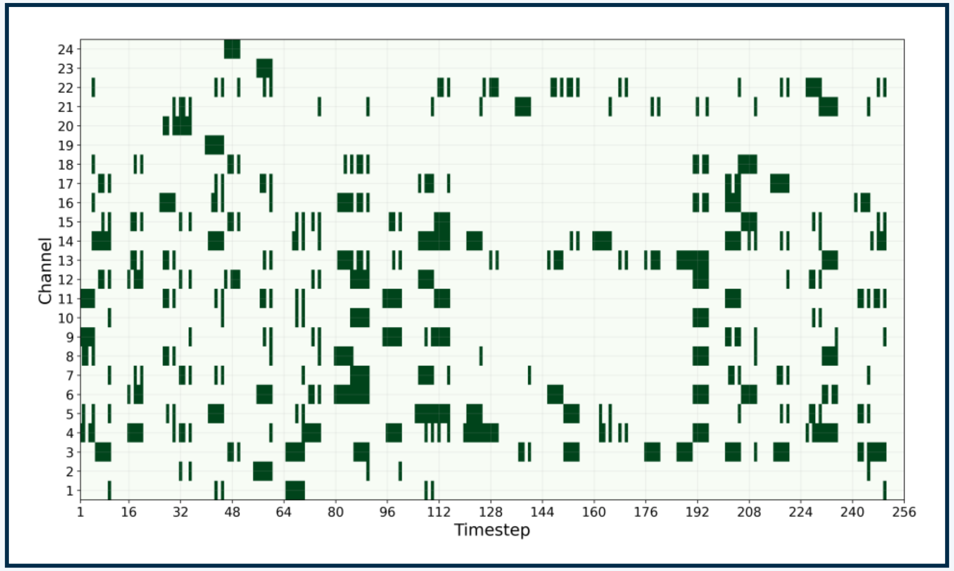
The Braille dataset[1] contains pressure readings from the iCub artificial fingertip sliding over individual Braile characters and collecting samples of 1.35 s.



Müller-Cleve et al., Frontiers in Neuroscience 16:951164, 2022

Complete dataset:

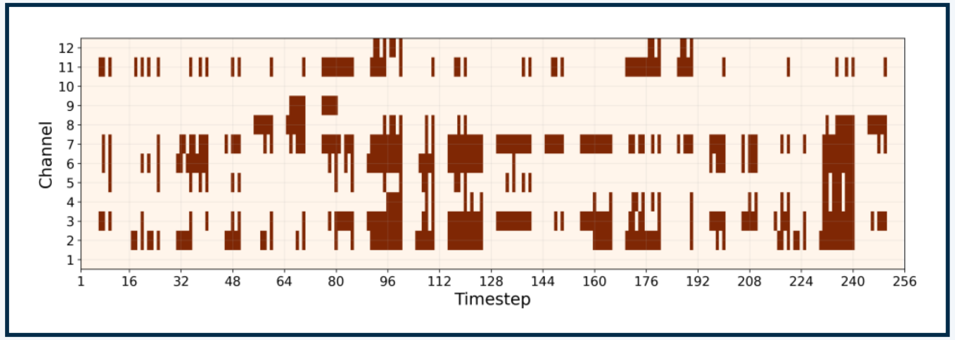
26 characters and 'Space'



Müller-Cleve et al., Frontiers in Neuroscience 16:951164, 2022

7-class subset:

{'A', 'E', 'I', 'O', 'U', 'Y', 'Space'}



Pedersen et al., Nature Communications 15:8122, 2024

$$Ch_{i,subset} = Ch_{2i,whole} \vee Ch_{2i-1,whole}$$

Events are created from the pressure readings through sigma-delta modulators.

1. 10.5281/zenodo.7050094

Classification accuracy and NeuroBench[1] results

Müller-Cleve et al., Frontiers in Neuroscience 16:951164, 2022

Pedersen et al., Nature Communications 15:8122, 2024

Braille dataset	Signal type	Architecture	Accuracy (%)
Complete	Encoded*	Spiking RNN	81.8
		LSTM	82.3
Complete	Encoded	L^2MU	85.6
7-class subset	Encoded	Spiking RNN	95.0
7-class subset	Encoded	L^2MU	97.1

Complete dataset

Metrics	Value
Test accuracy	85.55%
Parameters	4.90×10^5
Footprint	1.96 MB
Connection sparsity	0.13
Activation sparsity	0.97
Membrane updates	2.39×10^5
Effective MACs	0
Effective ACs	2.04×10^6
Dense operations	12.67×10^7

7-class subset

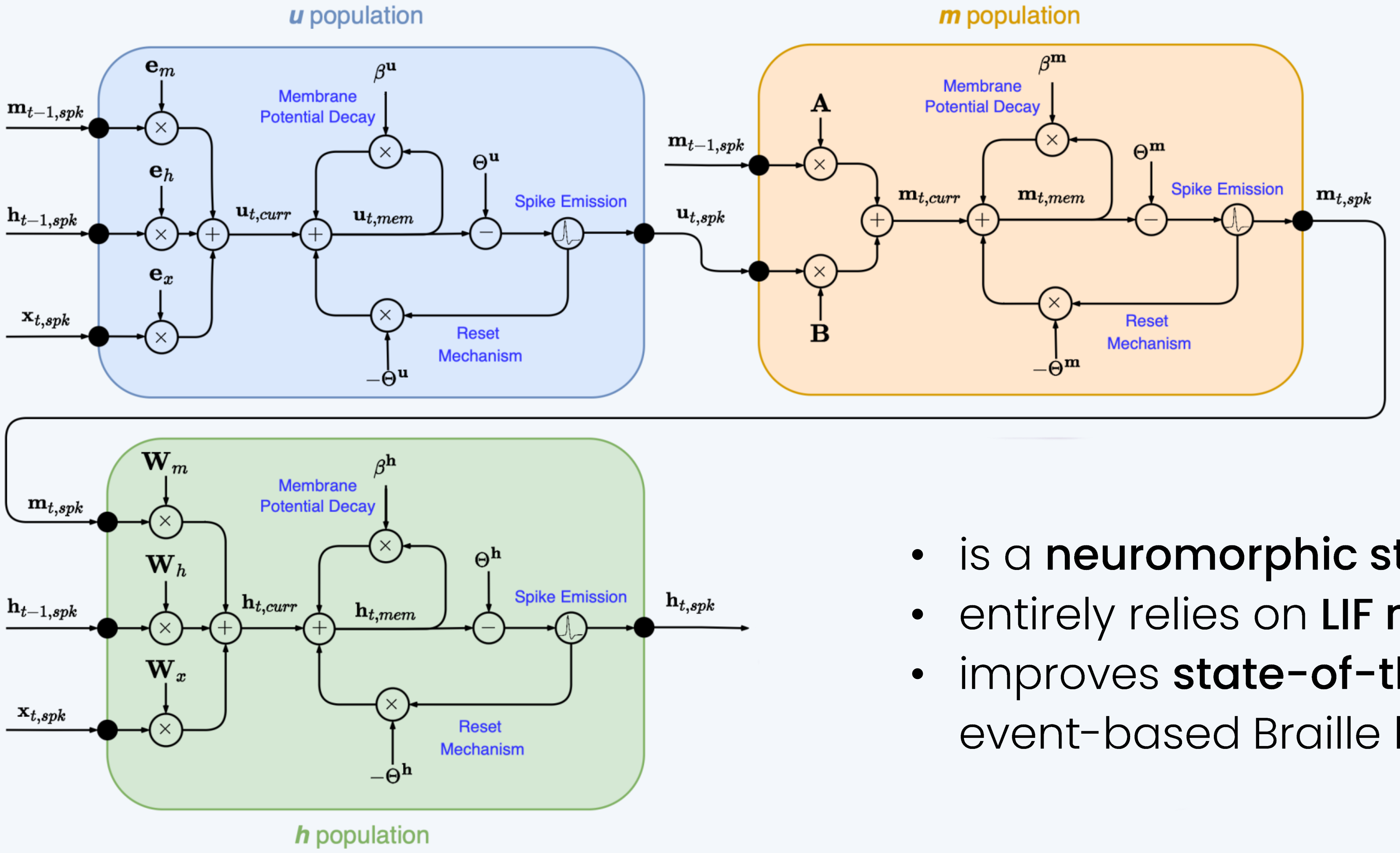
Metrics	Value
Test accuracy	97.14%
Parameters	1.19×10^5
Footprint	0.48 MB
Connection sparsity	0.21
Activation sparsity	0.96
Membrane updates	1.23×10^5
Effective MACs	0
Effective ACs	0.90×10^6
Dense operations	3.15×10^7

- <u>Activation sparsity</u>: relative number of inactive neurons within the network
- <u>Membrane updates</u>: number of updates of the neurons' membrane potential
- <u>Effective MACs</u>: number of operations with non-binary activations
- <u>Effective ACs</u>: number of operations with binary activations

1. *Yik et al., Nature Communications 16:1545, 2025*

Conclusion

Our **LIF-based LMU (L²MU)**:



- is a **neuromorphic state space model**
- entirely relies on **LIF neuron populations**
- improves **state-of-the-art** results on event-based Braille letter reading

Thank you!

Politecnico
di Torino

1859