



Spiking Memristive Architectures for
Learning to Learn

<https://spiking-small.eu>

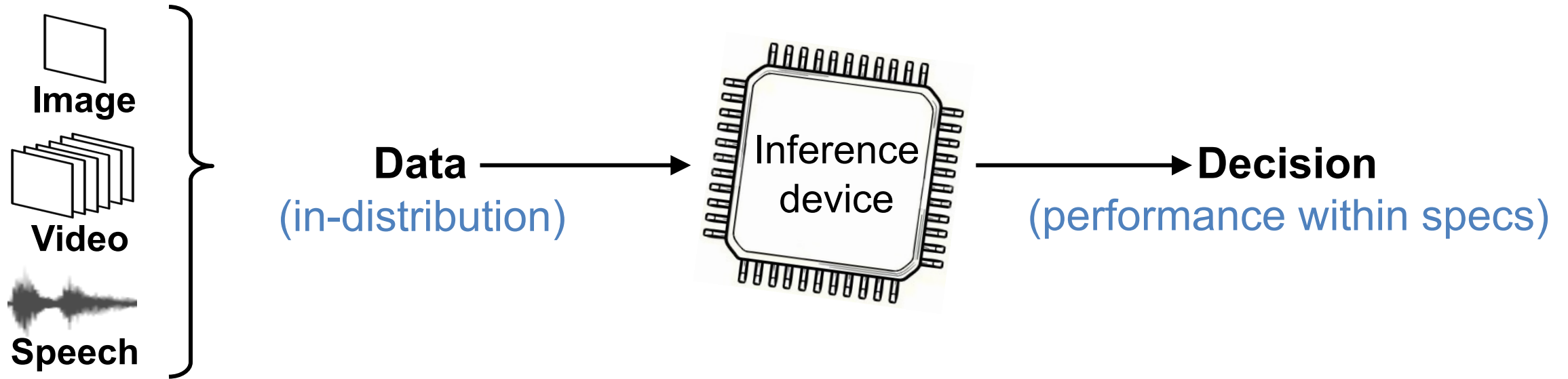
chist-era Projects Seminar
March 2022

Robert Legenstein

<http://www.igi.tugraz.at/legi/>
robert.legenstein@tugraz.at

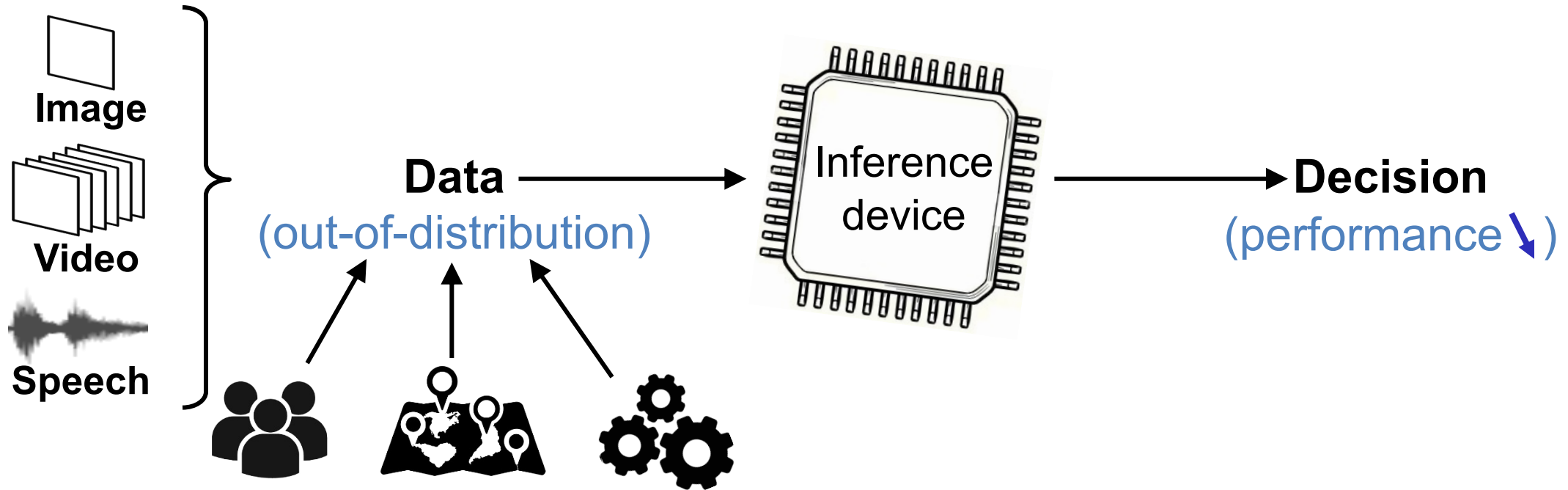
Institute of Theoretical Computer Science
Graz University of Technology, Austria

Scientific Background



Edge-AI: Why we need low-energy and fast on-chip learning solutions

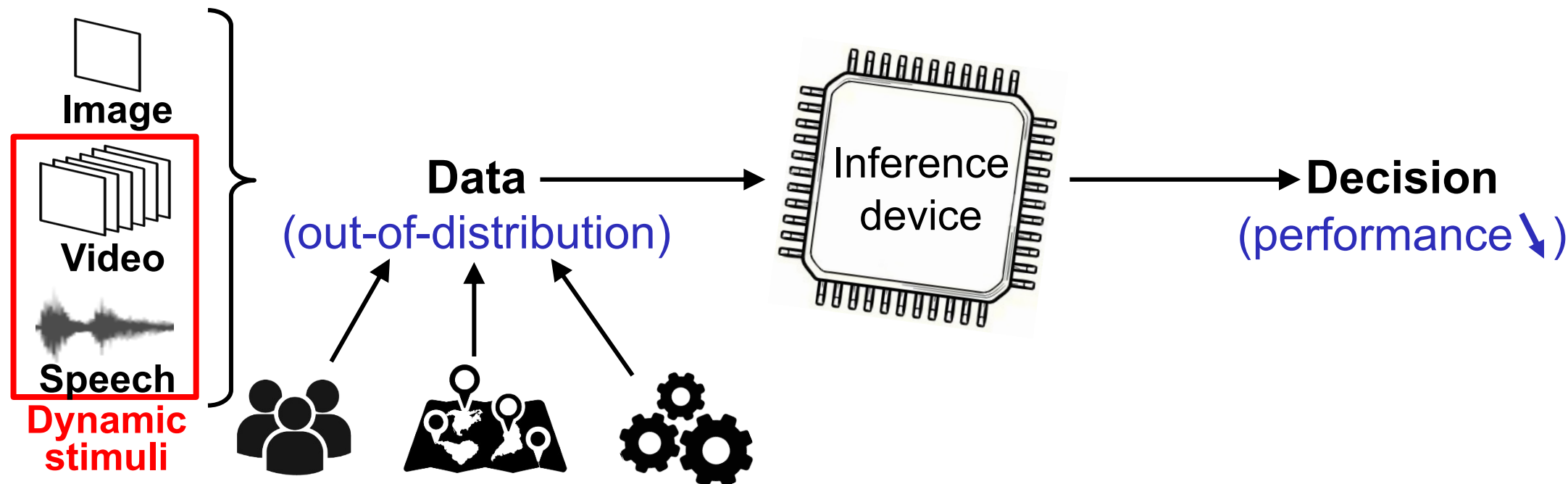
Scientific Background



Different users, environments, task requirements

- More training data before deployment?
 - Issues: cost, robustness, flexibility
- Data exchange with the cloud?
 - Issues: power budget, privacy

Scientific Background



Different users, environments, task requirements

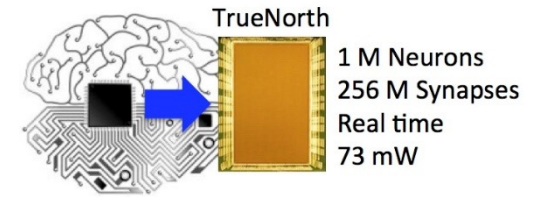
- More training data before deployment?
 - Issues: cost, robustness, flexibility
- Data exchange with the cloud?
 - Issues: power budget, privacy

On-chip training
(**end-to-end**)

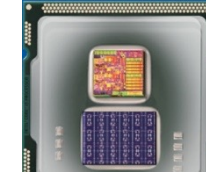
Wanted: Energy-efficient Neuromorphic Hardware



True North (IBM): Energy efficient fully digital neuromorphic chip



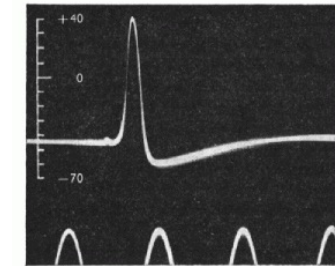
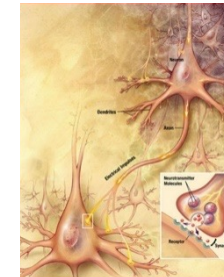
Loihi (intel): Neuromorphic chip with limited learning capabilities



There is a growing demand for low-power autonomously learning AI systems that are employed *in the field*.

We investigate options for learning in low-power unconventional hardware that combines:

1. neuromorphic computation,
2. spike-based communication, and
3. memristive analog memory.



In the field applications need fast training.

To overcome this problem, we investigate the applicability of *learning to learn* (L2L) to spiking memristive neuromorphic hardware, where the hardware is trained to become a good learner for the target application.

We evaluate our approach in a robotics setup.

Key challenges

- Algorithms for fast learning and adaptation in Spiking Neural Networks
- Imperfections of memristive devices
- Coupling of neuromorphic hardware in VLSI with memristive devices
- Application to a real-world robotics scenario

Potential impact

- Solutions for energy efficient inference and learning in edge devices.
- Provide a basic architecture towards low-power adaptive AI systems.

Partners
TU Graz

UZH

IBM

SOTON

USE



Coordinator

PI: Robert Legenstein



Expertise:

- Brain-inspired computing and learning
- Spiking Neural Networks
- Learning algorithms for
 - neuromorphic hardware and memristive devices

Role:

- Learning algorithms for neuromorphic hardware and memristors
- Algorithms for Learning to Learn



Partners
TU Graz

UZH

IBM

SOTON

USE



University of Zurich / ETH Zurich
PI: Giacomo Indiveri



**University of
Zurich^{UZH}**



Expertise:

- Spiking neuromorphic hardware
- Spike-based computation
- Analog and digital chip design

Role:

- Neuromorphic chip design
- Integration with memristive devices

IBM Research Zurich

PI: Angeliki Pantazi



Expertise:

- Memristive Phase-change-memory (PCM) devices
- In-memory computing
- Spike-based computing and learning



Role:

- PCM device development and testing
- Computing and learning paradigms with PCM devices
- Integration of PCM devices with neuromorphic hardware

Partners
TU Graz

UZH

IBM

SOTON

USE



University of Southampton (non-funded)

PI: Themis Prodromakis



Expertise:

- Memristive devices and materials
- Brain-inspired computing



Role:

- Filament-based memristor development and testing
- Integration of memristive devices with neuromorphic hardware

Partners
TU Graz

UZH

IBM

SOTON

USE



University of Seville

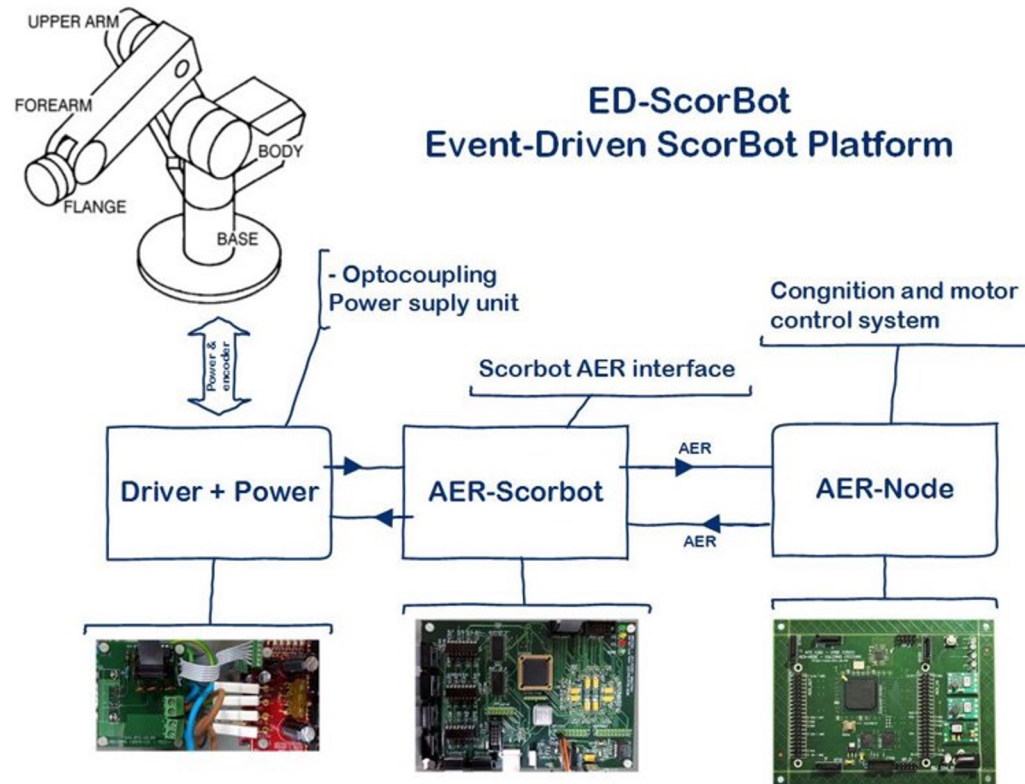
PI: Alejandro Linares-Barranco

Expertise:

- Robotics and Neurorobotics

Role:

- Integration of spiking-memristive system with robotics setup
- Application scenario development and testing



Main Scientific Results

Algorithms

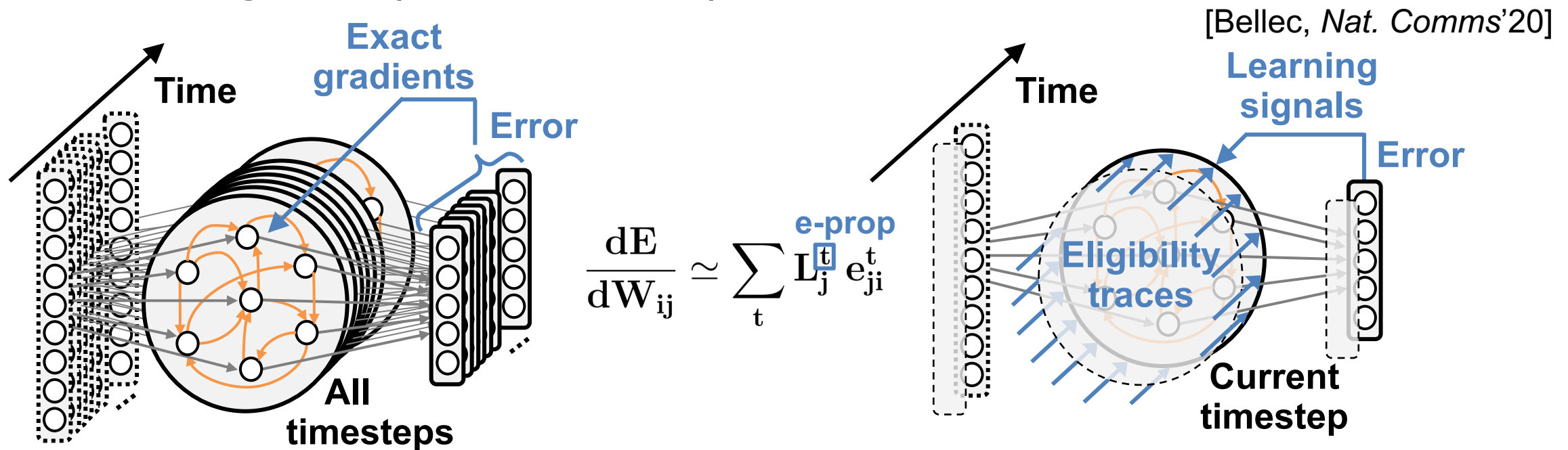
Spiking Hardware

Memristors

Robot Application

Training SNNs with spike-frequency adaptation using hardware-friendly training algorithms

Backprop through time (BPTT, backward)



Biological plausibility ↗

Space and time locality ↗

On-chip memory requirements ↘

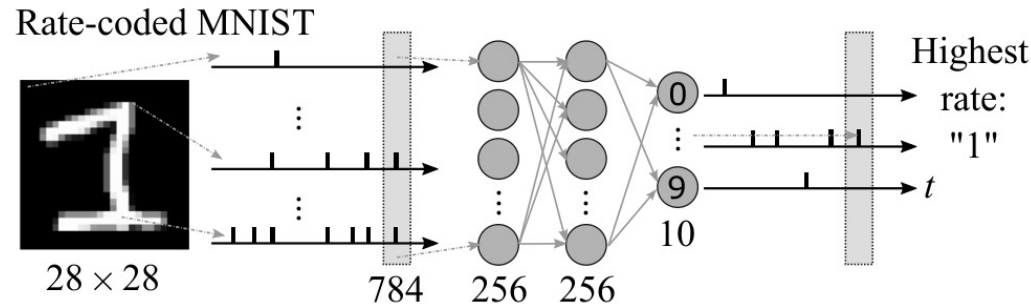
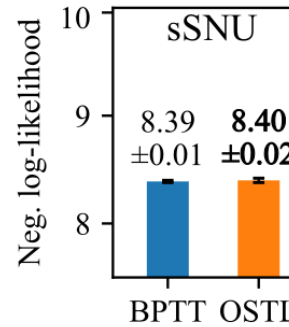
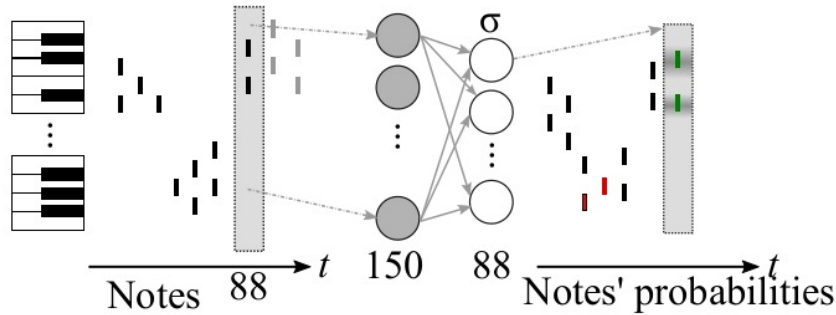
Main Scientific Results

Algorithms

Spiking Hardware

Memristors

Robot Application



Unit	OSTL	BPTT
sSNU	96.33%± 0.10	98.07%± 0.05
SNU	95.34%± 0.24	97.68%± 0.04

- Novel online-learning algorithm for SNNs: Online-spatial-temporal learning (OSTL)
- It separates temporal from spatial (i.e., layer-specific) gradient components, leading to more efficient training.
- Bohnstingl et al. Online spatio-temporal learning in deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems* (2022).

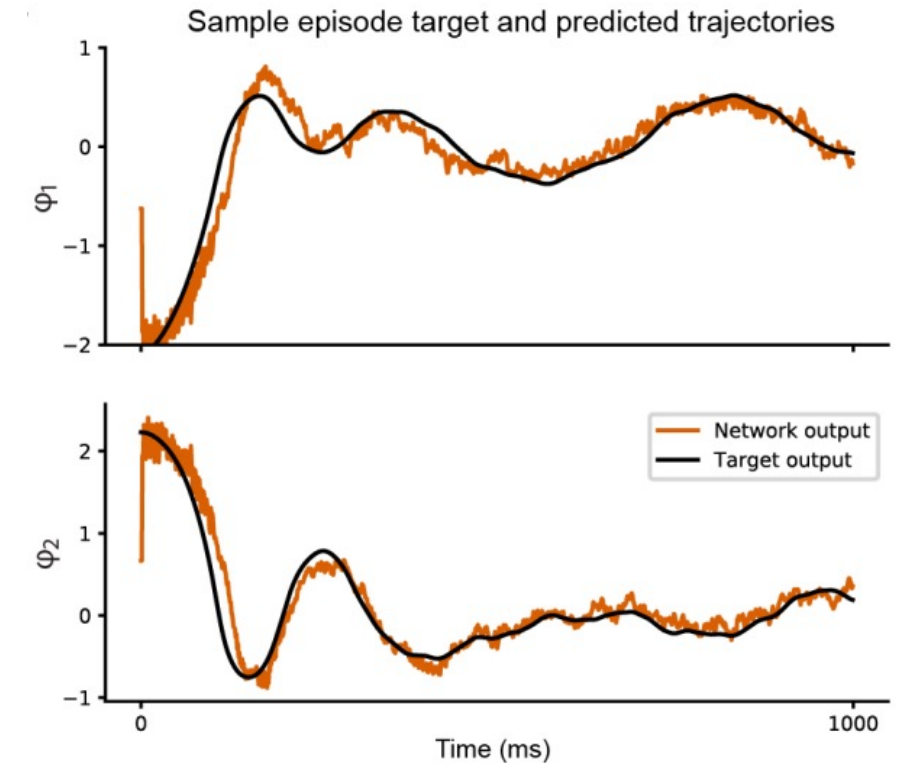
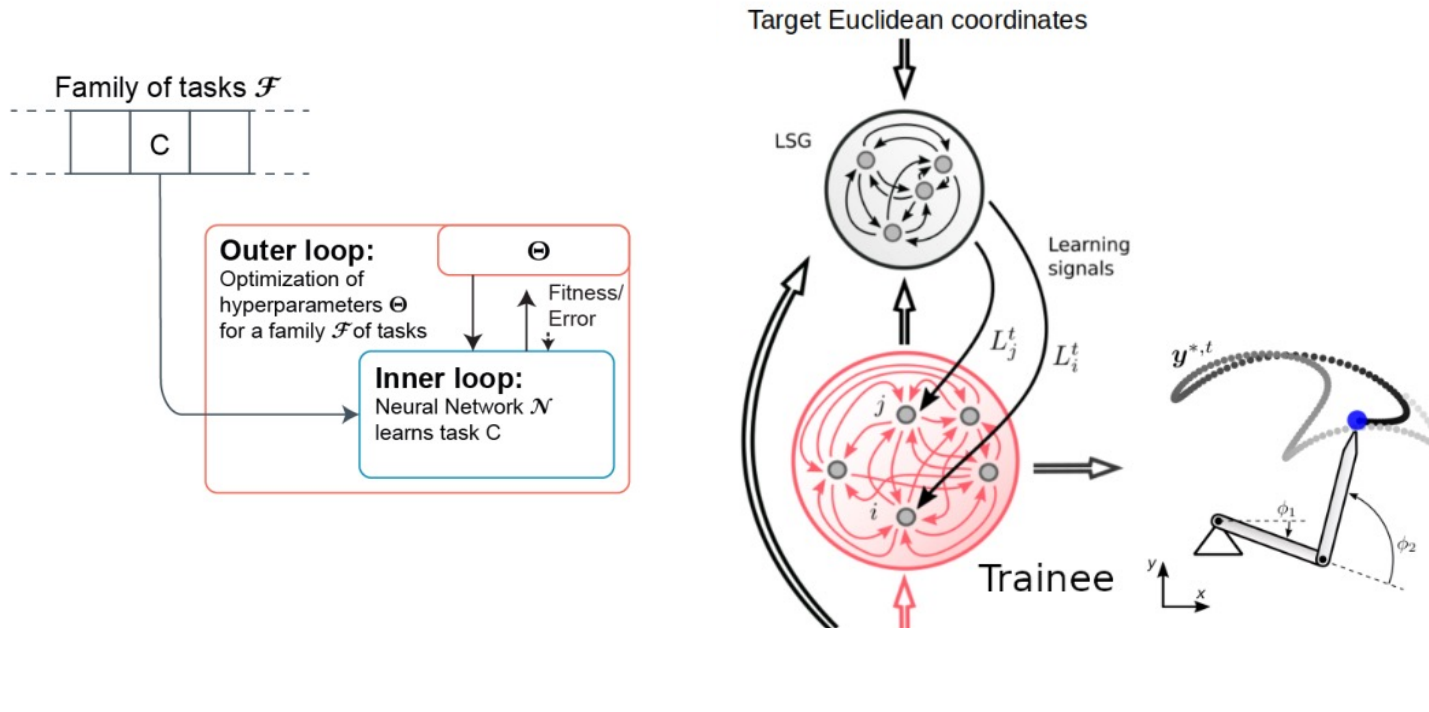
Main Scientific Results

Algorithms

Spiking Hardware

Memristors

Robot Application



- **Learning-to-Learn paradigms** for fast learning and adaptation of neuromorphic systems.
- Example: Single-shot robot trajectory learning.
- Subramoney, Bellec, Scherr, Legenstein & Maass (2021). Revisiting the role of synaptic plasticity and network dynamics for fast learning in spiking neural networks. BioRxiv Preprint.

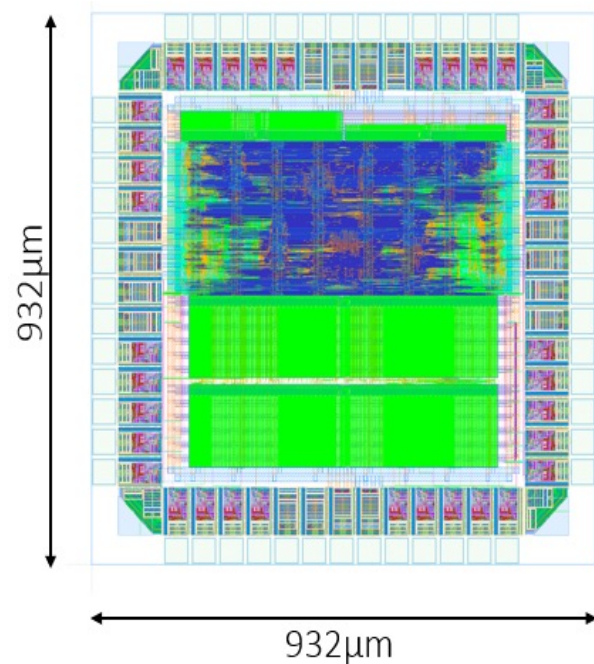
Main Scientific Results

Algorithms

Spiking Hardware

Memristors

Robot Application



Chip codename	ReckOn
Technology	28nm FDSOI CMOS
Implementation	Digital
Area	0.87mm ²
Core area	0.36mm ²
I/O bus	AER
Network type	Long-short-term spiking network
Topology	Max. 256-256-16 (recurrent)
Weight resolution	8-bit
Online learning	Stochastic approx. of BPTT
Learning-to-learn	Support with a dual-chip setup
Timestep	1ms real-time, can be accelerated
Working memory	Up to 32k timesteps
Application	Edge computing (keyword spotting, reinforcement learning)

- Development of a **neuromorphic chip with online learning capabilities (e-prop): ReckOn**
- Supports Learning-to-Learn

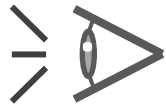
Task-agnostic learning: Three benchmark tasks

Algorithms

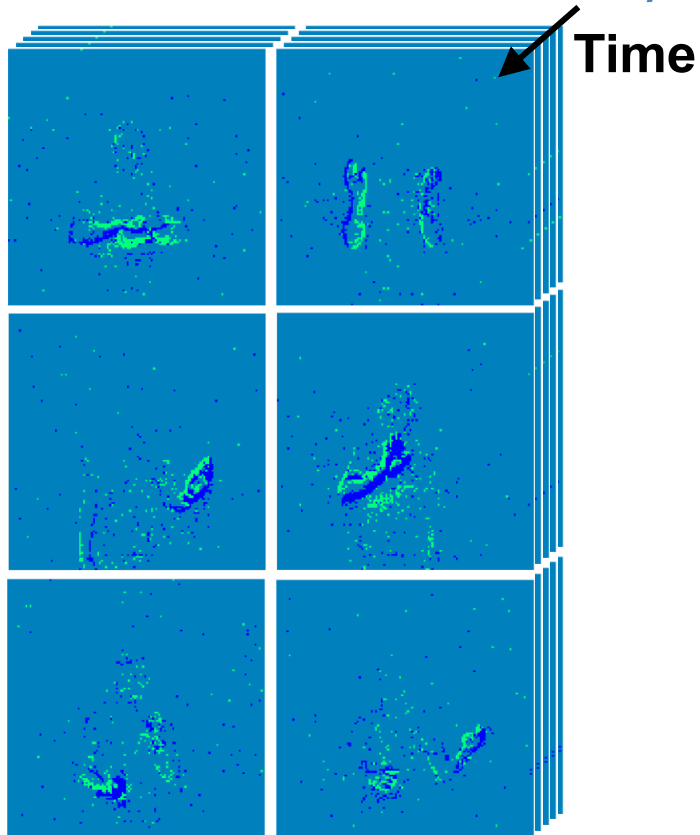
Spiking Hardware

Memristors

Robot Application



IBM DVS Gestures dataset
(10 classes, shrunk to 16x16)

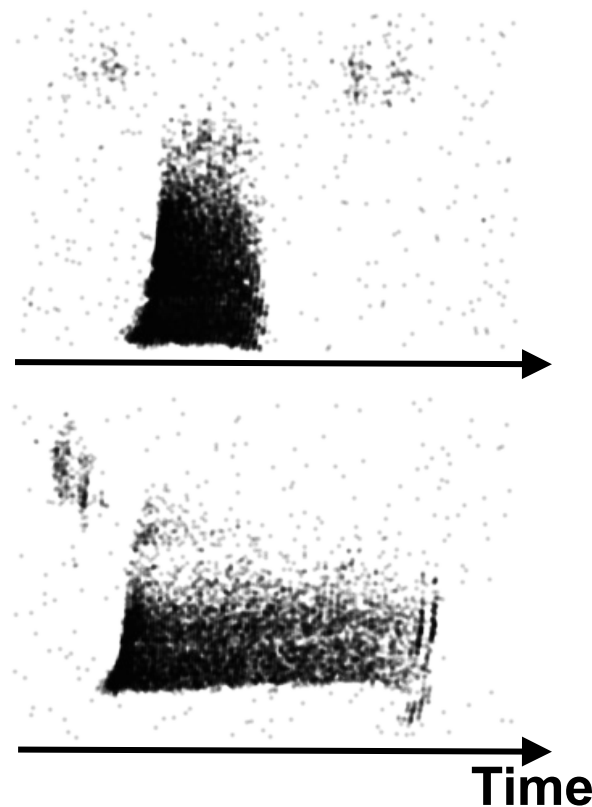


Accuracy: 87.3%



Audition

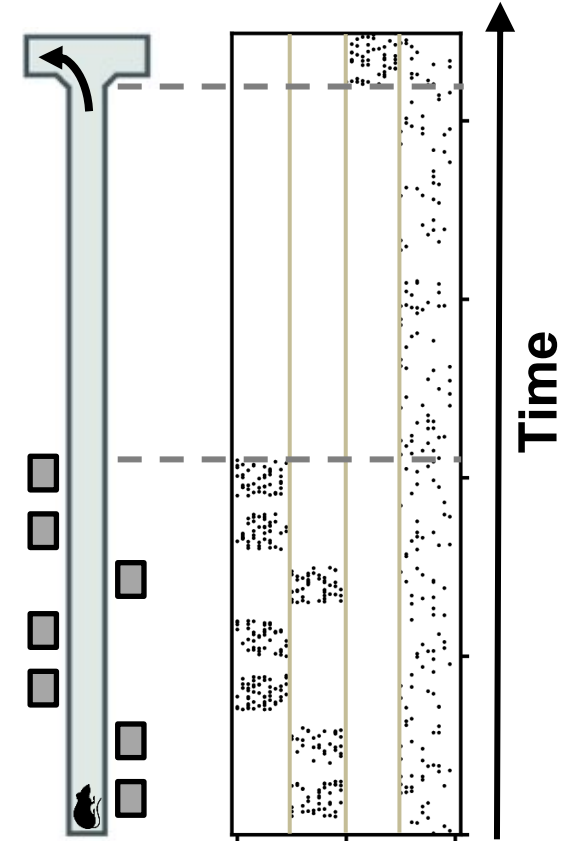
Spiking Heidelberg Digits (EN) dataset
(Keywordspotting)



Accuracy: 90.7%



Delayed-supervision
cue accumulation



Accuracy: 96.4%

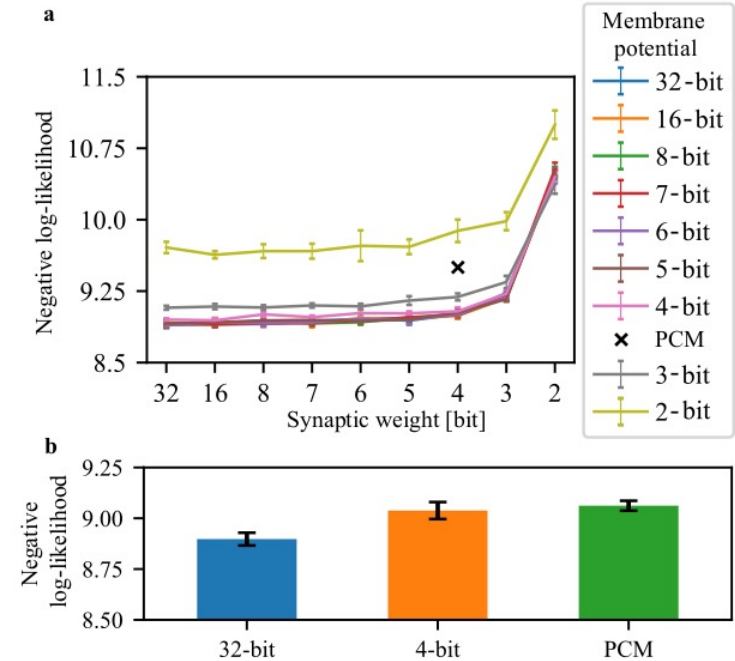
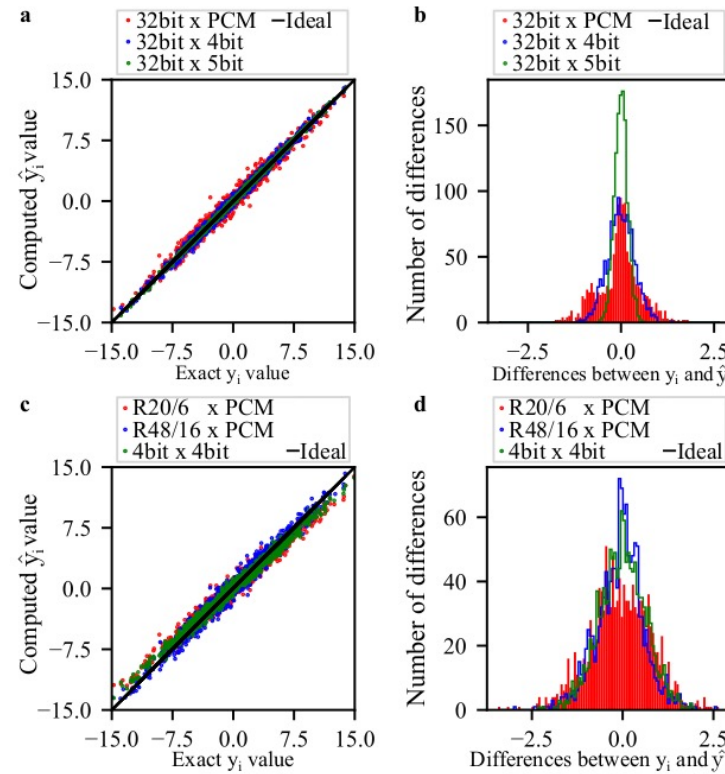
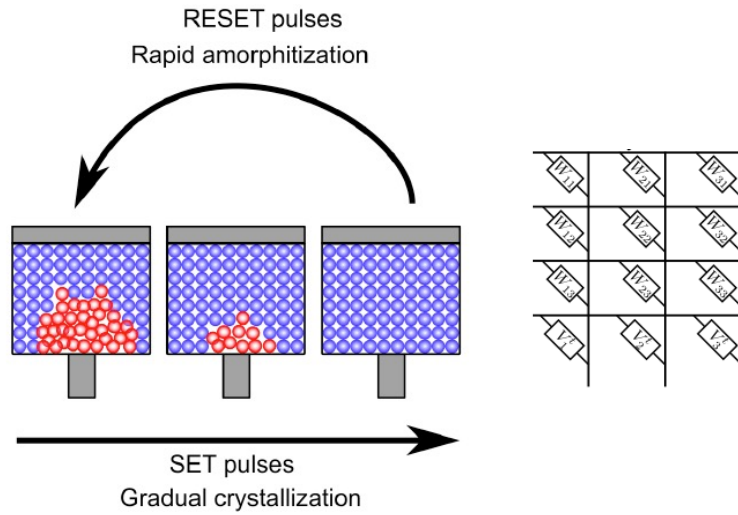
Main Scientific Results

Algorithms

Spiking Hardware

Memristors

Robot Application



- Use PCM memristors in crossbar array for synaptic weights and neural states
- Challenges from intrinsic imperfections of PCM devices
- Bohnstingl et al. (2020). Accelerating Spiking Neural Networks using Memristive Crossbar Arrays. 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 1–4.

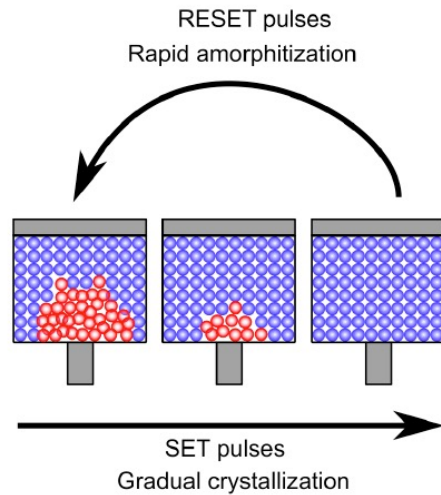
Main Scientific Results

Algorithms

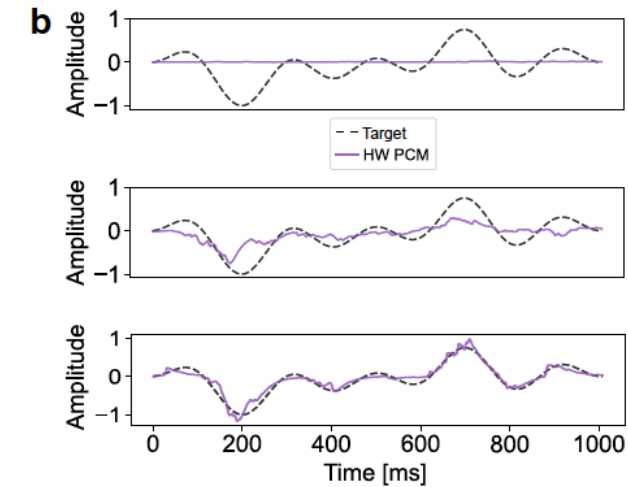
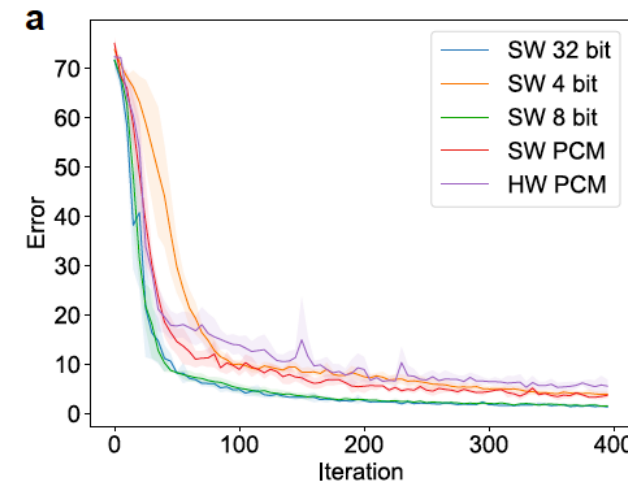
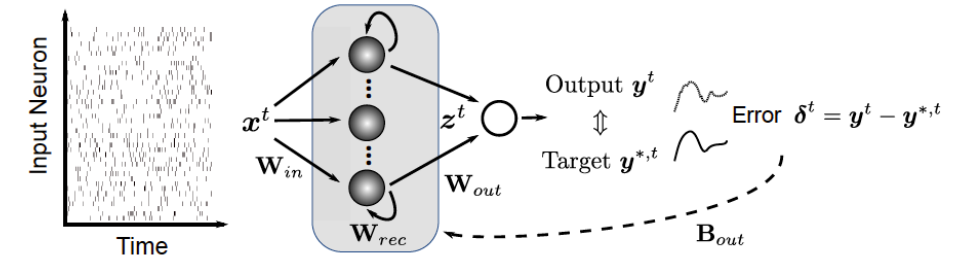
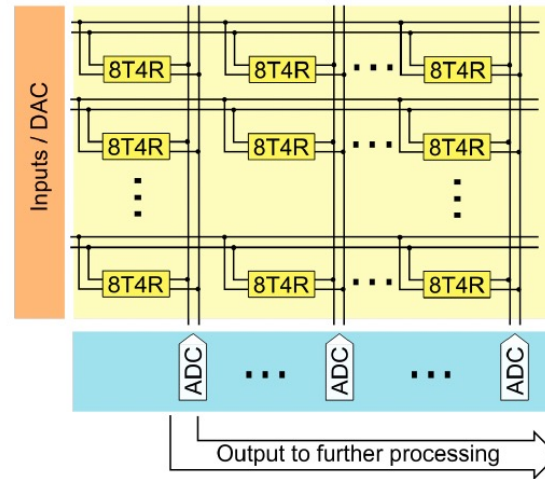
Spiking Hardware

Memristors

Robot Application



HERMES in-memory computing core



- SNN equipped with online learning capabilities and trained with a hardware-in-the-loop approach is robust to limited computational precision and hardware imperfections.
- Learns sensory-motor sequences with e-prop

Main Scientific Results

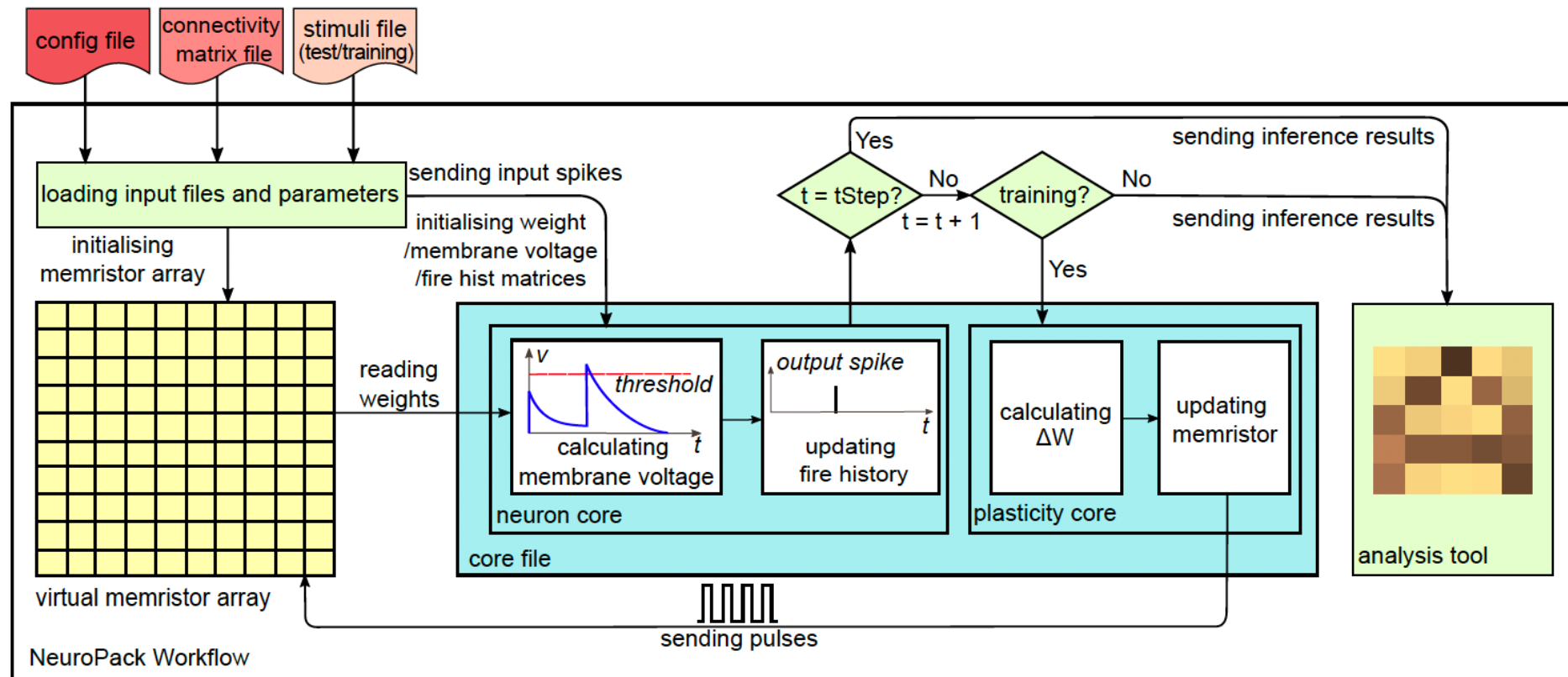
Algorithms

Spiking Hardware

Memristors

Robot Application

NeuroPack platform (SOTON):



Main Scientific Results

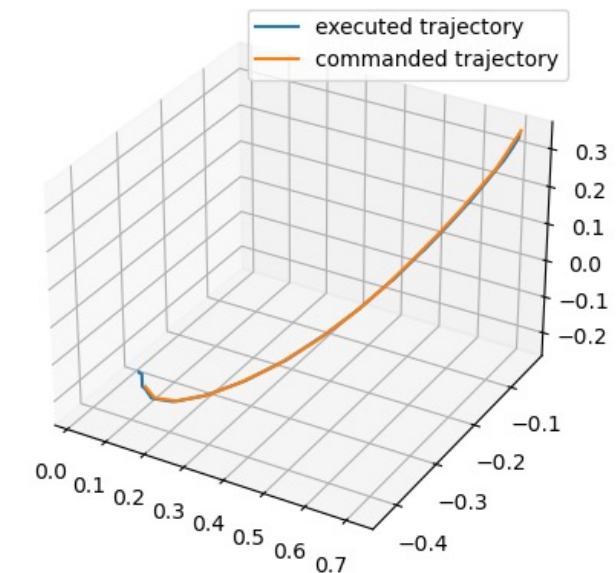
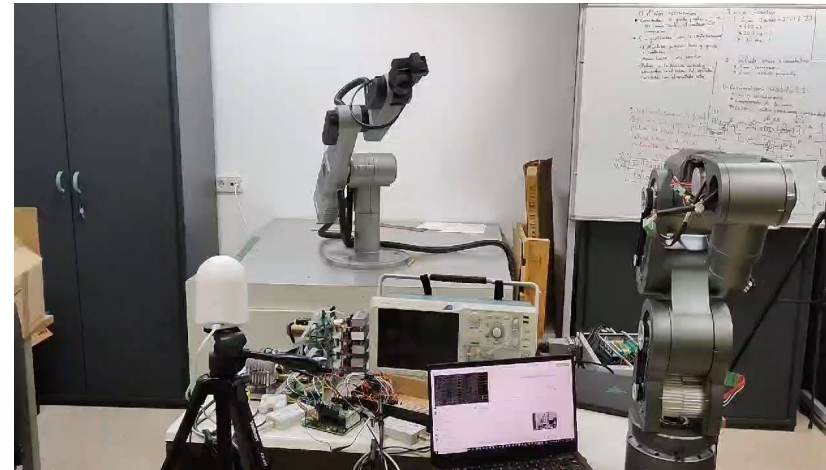
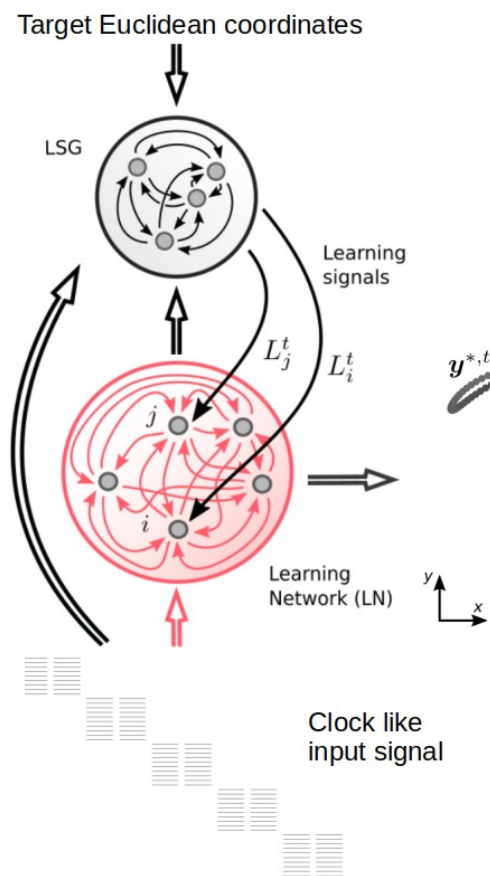
Algorithms

Spiking Hardware

Memristors

Robot Application

- We moved from simulation to the real robot



2020

- T Bohnstingl, S Woźniak, W Maass, A Pantazi & E Eleftheriou (2020a). Online spatio-temporal learning in deep neural networks. ArXiv Preprint ArXiv:2007.12723.
- T Limbacher & R Legenstein (2020). H-Mem: Harnessing synaptic plasticity with Hebbian Memory Networks. **NeurIPS 2020**.
- T Bohnstingl, A Pantazi & E Eleftheriou (2020b). Accelerating Spiking Neural Networks using Memristive Crossbar Arrays. 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 1–4.
- A Linares-Barranco, F Perez-Peña, A Jimenez-Fernandez & E Chicca (2020). ED-BioRob: A Neuromorphic Robotic Arm With FPGA-Based Infrastructure for Bio-Inspired Spiking Motor Controllers. Frontiers on Neurorobotics.
- C Giotis, A Serb, S Stathopoulos, L Michalas, A Khiat & T Prodromakis (2020). Bidirectional Volatile Signatures of Metal–Oxide Memristors–Part I: Characterization. IEEE Transactions on Electron Devices, vol. 67, no. 11.
- C Giotis, A Serb, S Stathopoulos, L Michalas, A Khiat & T Prodromakis(2020). Bidirectional Volatile Signatures of Metal–Oxide Memristors–Part II: Modeling. IEEE Transactions on Electron Devices, vol. 67, no. 11.

2021

- Subramoney, A., Bellec, G., Scherr, F., Legenstein, R., & Maass, W. (2021). Revisiting the role of synaptic plasticity and network dynamics for fast learning in spiking neural networks. *BioRxiv*.
- Y Demirag, F Moro, T Dalgaty, G Navarro, C Frenkel, G Indiveri, E Vianello & M Payvand (2021). PCM-trace: Scalable Synaptic Eligibility Traces with Resistivity Drift of Phase-Change Materials. IEEE International Symposium on Circuits and Systems (ISCAS).
- Kraisnikovic, C., Maass, W., & Legenstein, R. (2021). Spike-based symbolic computations on bit strings and numbers. *Neuro-Symbolic Artificial Intelligence: The State of the Art*, 214–234.
- Traub, M., Legenstein, R., & Otte, S. (2021). Many-Joint Robot Arm Control with Recurrent Spiking Neural Networks. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Demirag, Y., Frenkel, C., Payvand, M., & Indiveri, G. (2021). Online Training of Spiking Recurrent Neural Networks with Phase-Change Memory Synapses. *ArXiv Preprint ArXiv:2108.01804v2*.
- Salaj, D., Subramoney, A., Kraisnikovic, C., Bellec, G., Legenstein, R., & Maass, W. (2021). Spike frequency adaptation supports network computations on temporally dispersed information. **eLife**.

2022

- Frenkel, C., & Indiveri, G. (2022). ReckOn: A 28nm Sub-mm² Task-Agnostic Spiking Recurrent Neural Network Processor Enabling On-Chip Learning over Second-Long Timescales. In *2022 IEEE International Solid-State Circuits Conference (ISSCC)* (Vol. 65, pp. 1-3). IEEE.
- Bohnstingl et al. Online spatio-temporal learning in deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- Piñero-Fuentes E., Canas-Moreno S., Rios-Navarro A., Cascado-Caballero D., Jimenez-Fernandez A. and Linares-Barranco, A., An MPSoC-based on-line Edge Infrastructure for Embedded Neuromorphic Robotic Controllers. Accepted for publication at ISCAS (2022).
- Linares-Barranco, A., Piñero-Fuentes E., Canas-Moreno S., Rios-Navarro A., Maryada, Jingyue Zhao, Zendrikov D., Indiveri G., Towards hardware Implementation of WTA for CPG-based control of a Spiking Robotic Arm. Accepted for publication at ISCAS (2022).
- J. Huang et al., NeuroPack: An Algorithm-level Python-based Simulator for Memristor-empowered Neuro-inspired Computing, arXiv:2201.03339 (2022)

Dissemination

Publications

Webpage, Talks, Popular Science

Webpage: <https://spiking-small.eu>

- February 24, 2020: Plenary lecture: “In-memory Computing for AI Applications”, SPIE 2020.
- April 27, 2020: Invited talk at the IBM Research & Brain-inspired computing, virtual seminar, MLP2020, University of Warsaw, Warsaw, PL.
- September 29, 2020: Invited talk at the Bernstein conference on Computational Neuroscience - Workshop on Dynamic Probabilistic Inference in the Brain. Talk title: Inference from Facts with Hebbian Plasticity.
- October 2020: Keynote lecture at the "Brain-inspired computing", Ghost Day, Applied Machine Learning Conference.
- November 2020: Invited panellist at "Inspiring stories of women's success in STEM", WiCAS Event at the IEEE International Conference on Electronics, Circuits & Systems.
- December 8, 2020: Poster presentation and oral spotlight talk at the Conference on Advances in Neural Information Processing Systems (**NeurIPS 2020**). Talk title H-Mem: Harnessing synaptic plasticity with Hebbian memory networks.
- December 8, 2020: Popular science article published in Planet Research: “Future of Computing: Learning-Based, Energy-Efficient, and Brain-Inspired” by Robert Legenstein.
- January 18, 2021: Keynote lecture: “In-memory Computing for AI Applications”, HiPEAC 2021.
- Apr-Oct 2021: **Art exhibition** “SHOWING STYRIA: what will be – Towards a Plurality of Futures”. Video showing our trained SNN playing a video game. Robert Legenstein
- July 2021: Invited Talk “Brain Computation: Integrating perspectives from Neuroscience, Computer Science, and Deep Learning”, Workshop on Beyond von Neumann Architectures, Lagrange Mathematics and Computation Research Center, online, Speaker: Robert Legenstein.
- Sept 2021: Invited Talk “Brain-Inspired Algorithms for Energy-Efficient Machine Learning”, BeyondAI International Summer School, online, Graz University of Technology, Speaker: Robert Legenstein.

Webpage: <https://spiking-small.eu>

- Sept 2021: Poster presentation “Spike-based symbolic computations on symbol sequences”, Bernstein Conference 2021, online, Speaker: Ceca Krasinikovic.
- G. Indiveri, "Neuromorphic intelligence: Using populations of inhomogeneous spiking neurons to carry out robust computation in neuromorphic processors", ISSCC Tutorial, 2022.
- ReckOn also appeared on **Tweakers** (large magazine dedicated to electronics and technology news, in Dutch): <https://tweakers.net/reviews/9916/losjes-gebaseerd-op-onze-hersenen-uitdagingen-voor-energiezuinige-ai.html>
- March 2022: Design, Automation and Test in Europe Conference (DATE22 , panelist lecture)
- February 2022: IEEE International Solid-State Circuits Conference. G. Indiveri (ISSCC, forum lecture)
- December 2021: IEEE International Electron Devices Meeting. G. Indiveri (IEDM, short-course lecture)
- May 2021: IEEE International Symposium on Circuits and Systems. G. Indiveri (**ISCAS, keynote lecture**)
- April 2021: A. Pantazi, “Neural dynamics and learning in Spiking Neural Networks”, lecture at FORTH
- April 2021: A. Pantazi, “Brain-inspired computing”, lecture at IEEE Switzerland General Assembly
- April 2021: S. Woźniak, T. Bohnstingl and A. Pantazi, “Biologically Inspired Dynamics Enhance Deep Learning ” in ERCIM News Special Theme: Brain-Inspired Computing

Dissemination

Publications

Webpage, Talks, Popular Science

Webpage: <https://spiking-small.eu>

- May 23, 2021: A. Pantazi, “Neural dynamics and learning in SNNs”, ISCAS 2021 Tutorial Spiking Neural Networks - Algorithms, Circuits, and Systems based on CMOS and emerging devices, IEEE International Symposium on Circuits and Systems
- October 6th-7th, 2021: Woźniak, S., Bohnstingl, T., Pantazi, A. Neuromorphic Computing & Spiking Neural Networks, mPP tutorial at CERN, <https://indico.cern.ch/event/1070931/>
- November 17th, 2021: Pantazi, A., Bohnstingl, T., Overview talk on Neuromorphic Computing, Virtual event at CERN (CMS)
- Nov 23-24, 2021: Pantazi, A. Human-like AI – Neuromorphic computing & Human-like AI, Press Event at IBM Research Europe, Zurich, <https://lnkd.in/gcgZFft9>
- September 1, 2021: Intel Neuromorphic Research Center Forum invited talk: “Spike-based motor controllers for Neuromorphic Robotics”
Speaker: Alejandro Linares-Barranco
- September 14, 2021: Neurotech Talks: Educational Webinar Day 9: “WHY INTEGRATING NEUROMORPHIC TECHNOLOGIES IN ROBOTS?”
Speaker: Alejandro Linares-Barranco

- Verilog code of ReckOn as well as its documentation have been released in open source
<https://github.com/ChFrenkel/ReckOn>
under a SolderPad Hardware License v2.1.
- The ArC 2 system has been developed with input from 100s of customers of ArC Instruments (SOTON).
associated software is provided as open-source
www.arc-instruments.co.uk
- Partner SOTON has released NeuroPack, a software package for network simulations with memrisive devices in open source.
- IBM has released the source code of OSTL to the scientific community,
<https://github.com/IBM/ostl>
- TUG releases all published code on public repositories
<https://github.com/IGITUGraz>



SMALL

Spiking Memristive Architectures for
Learning to Learn

<https://spiking-small.eu>

