From biology to your coffee machine ? How neuromorphic computing may affect our future life

HBP Colloquium

FZ Jülich, October 2018

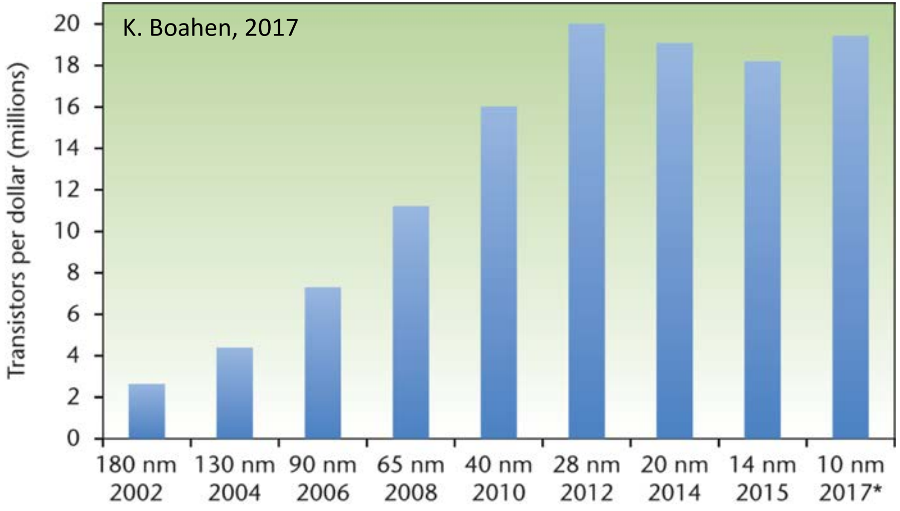
Karlheinz Meier

Ruprecht-Karls-Universität Heidelberg

meierk@kip.uni-heidelberg.de

@brainscales

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386



No more progress from smaller transistors

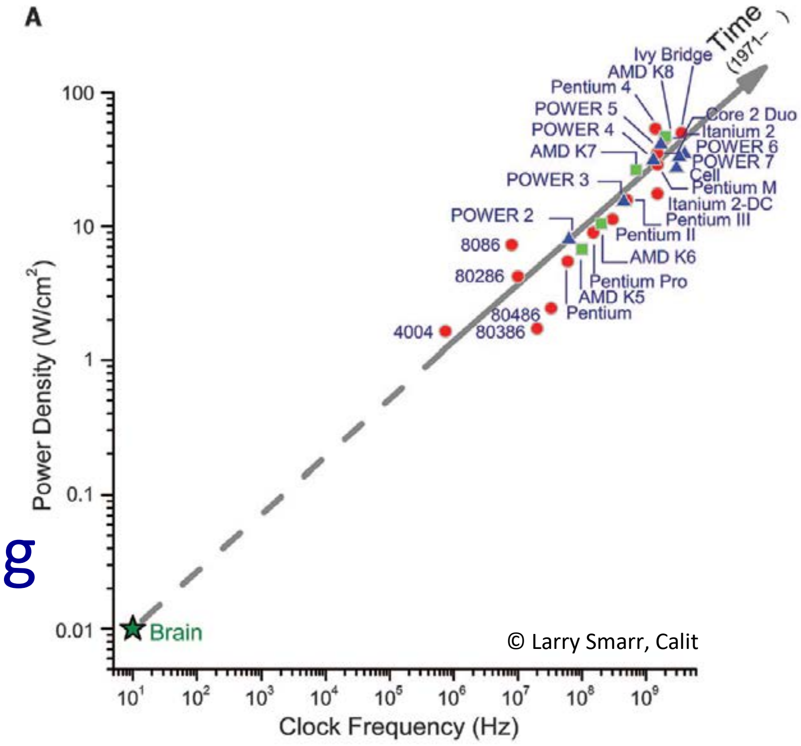
New ARCHITECTURES suddenly interesting !

First : Make use of CMOS devices
Then : Pave the way for non-CMOS

Brain-inspired, brain-derived or neuromorphic computing

Definition Transferring aspects of structure and function from biological substrates to electronic circuits

Structure : Cells – Networks – Connections
Function : Local Processing – Communication – Learning



Assets of brain inspired computing

- Energy efficiency
- Compactness
- Fault tolerance
- Speed
- Configuration and learning replace programming
- Scalability

Conventional computing is moving away from the brain

Neuromorphic : why and how ?

Future computing based on biological information processing

Understanding biological information processing

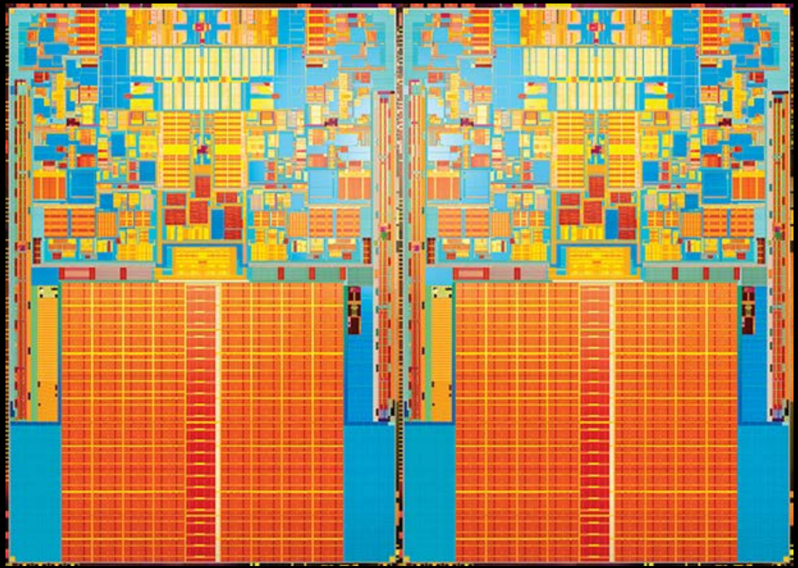
Requires model system to test principles

Two **fundamentally different** modeling approaches

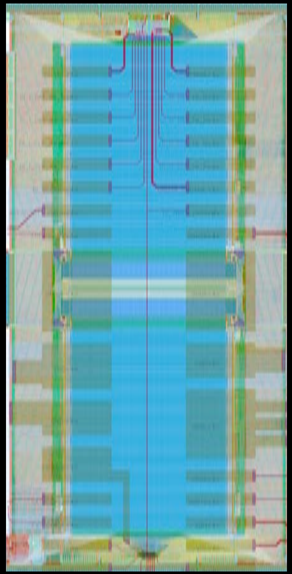
- **SYMBOLIC MODEL (Turing)**
  represents model parameters as symbols (binary numbers)
  represents function as instructions (algorithms)
- **PHYSICAL MODEL (non Turing)**
  represents model parameters as physical quantities (like the brain)
  represents function as circuit dynamics (physical laws)

Combined into hybrid system

Same transistors - where‘s the neuron ?

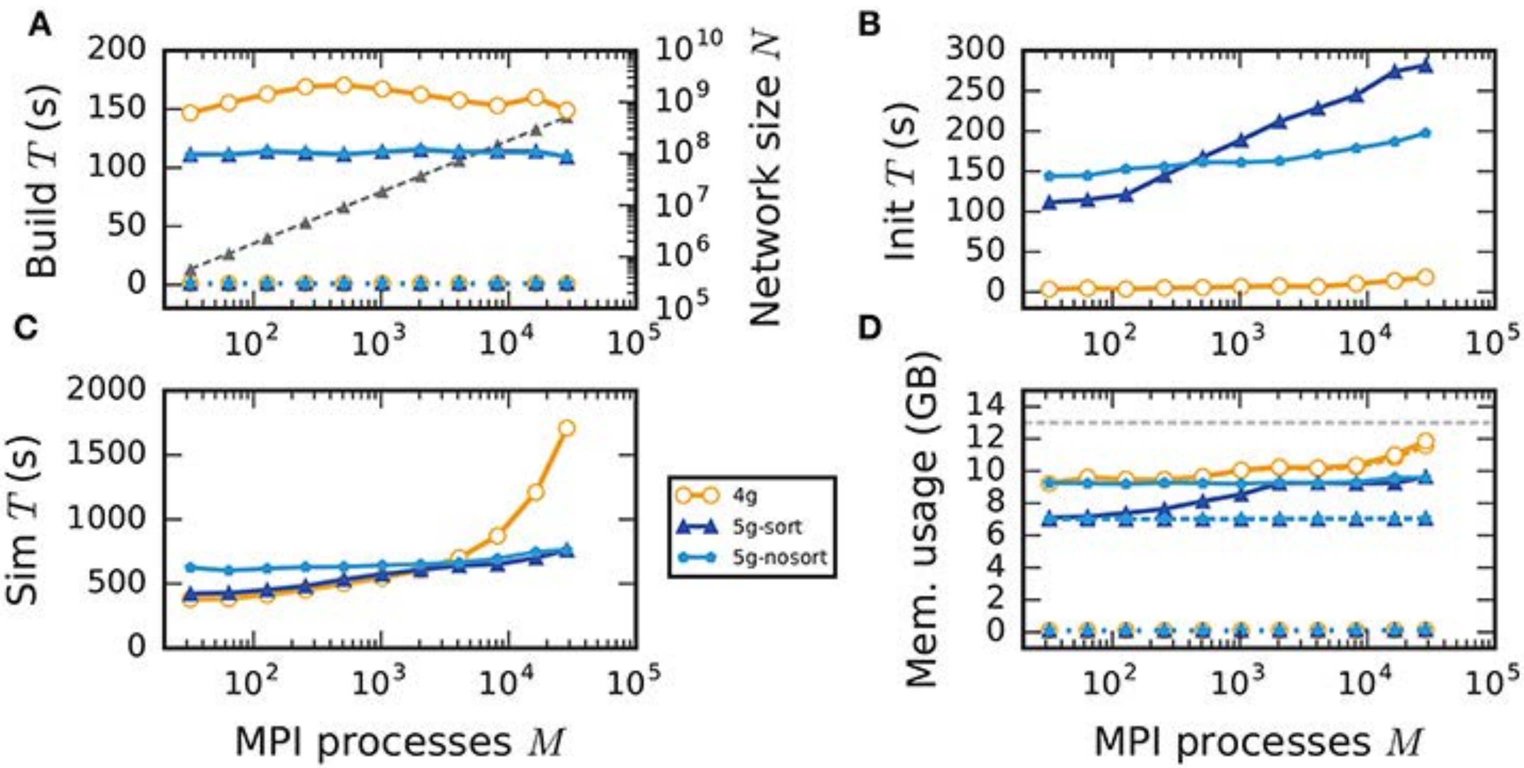


© Intel Corporation



The BrainScaleS Project

Not to scale



The simulation approach

Weak scaling (*sizeup*) on modern petascale many-processor machines

Jordan, Jakob, et al. Extremely scalable spiking neural network simulation code: from laptops to exascale computers. *Frontiers in Neuroinformatics*, 2018, 12. Jg., S. 2.

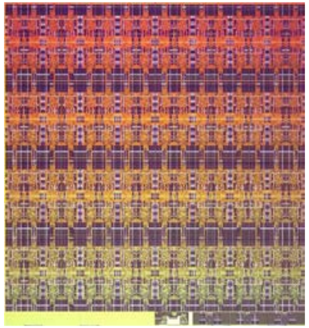
Neuromorphic implementations : Towards biological realism

SpiNNaker
Biologically Inspired Massively Parallel Architectures

IBM
TrueNorth

intel
Loihi

BrainScaleS



Biological realism →

← Ease of use by traditional programming tools

Many-core (ARM) architecture Optimized spike communication network Programmable local learning x0.01 real-time to real-time	Full-custom-digital neural circuits No local learning (TrueNorth) Programmable local learning (Loihi) Exploit economy of scale x0.01 real-time to x100 real-time	Analog neural cores Digital spike communication Biological local learning Programmable local learning x10.000 real-time

The HBP Neuromorphic Computing Strategy

2nd generation emerged from co-design process in HBP

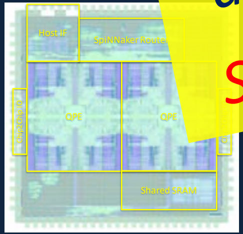
1st generation SpiNNaker-1 Machine

Many-core system
1 Million ARM cores
Real-time simulator

Many-core architecture

SIMULATION

Towards the 2nd generation SpiNNaker-2



144 Cortex M4F per chip
36 GIPS/Watt per chip
x10 performance with constant power

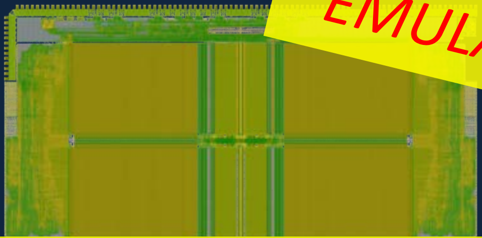
1st generation BrainScaleS-1 Machine

Physical model system
4M neurons, 1B plastic synapes
Accelerated emulator

Physical model

EMULATION

Towards the 2nd generation BrainScaleS-2



On-chip plasticity processors
Flexible hybrid plasticity
Active dendrites

Common software ecosystem, remote access, open user facility
Close cooperation with (theoretical) neuroscience, application domain

These are neuromorphic processors designed and built from the transistor up !

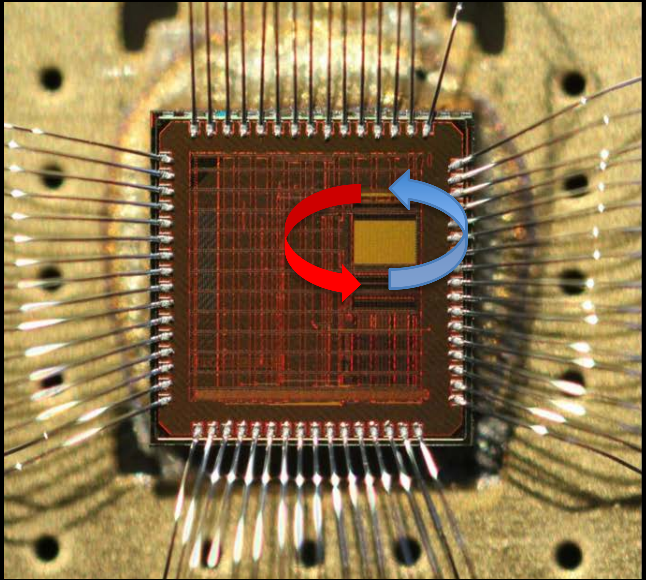
Assets of brain inspired computing ?

- Energy efficiency
- Compactness
- Fault tolerance
- Speed
- Configuration and learning replace programming
- Scalability

2nd generation – HBP made

BrainScaleS-2

65 nm prototype chip in the lab



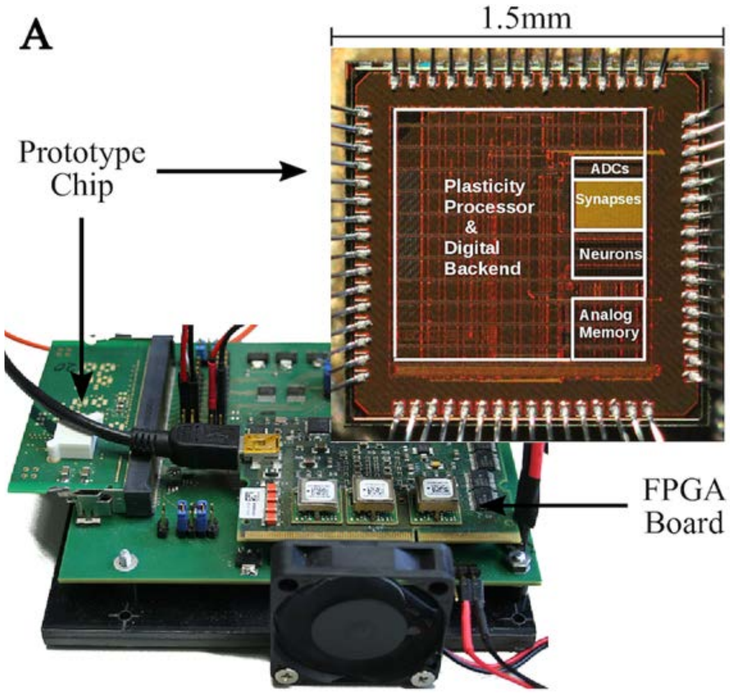
- Hybrid plasticity with on-chip processor (PPU): on-chip loops, time-scales from ms to years
  - Input : timing correlations, rates, membrane potentials, external signals
  - Change : synaptic weights, neuromodulation, network structure

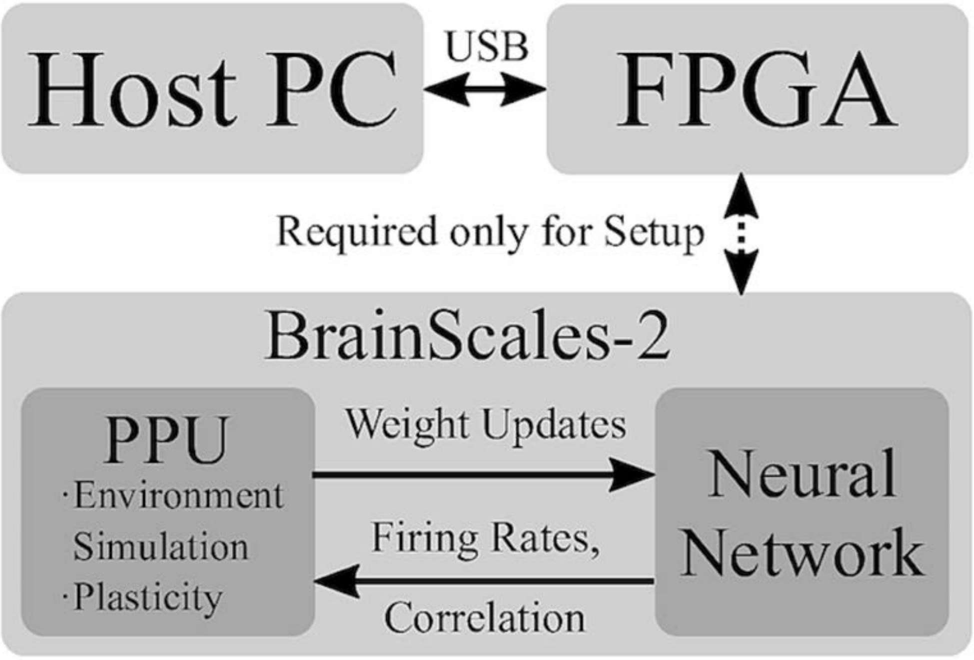
- Structured neurons
  - Multicompartment neurons
  - Active, non-linear dendrites, backpropagating APs
  - NMDA, Ca plateau potentials

- Public evaluation system by end-2018
- Full-size prototypes by mid-2020
- Full size system by 2023
- Funding pending

In-the-the-loop experiments on silicon – External computers only for set-up

Compactness

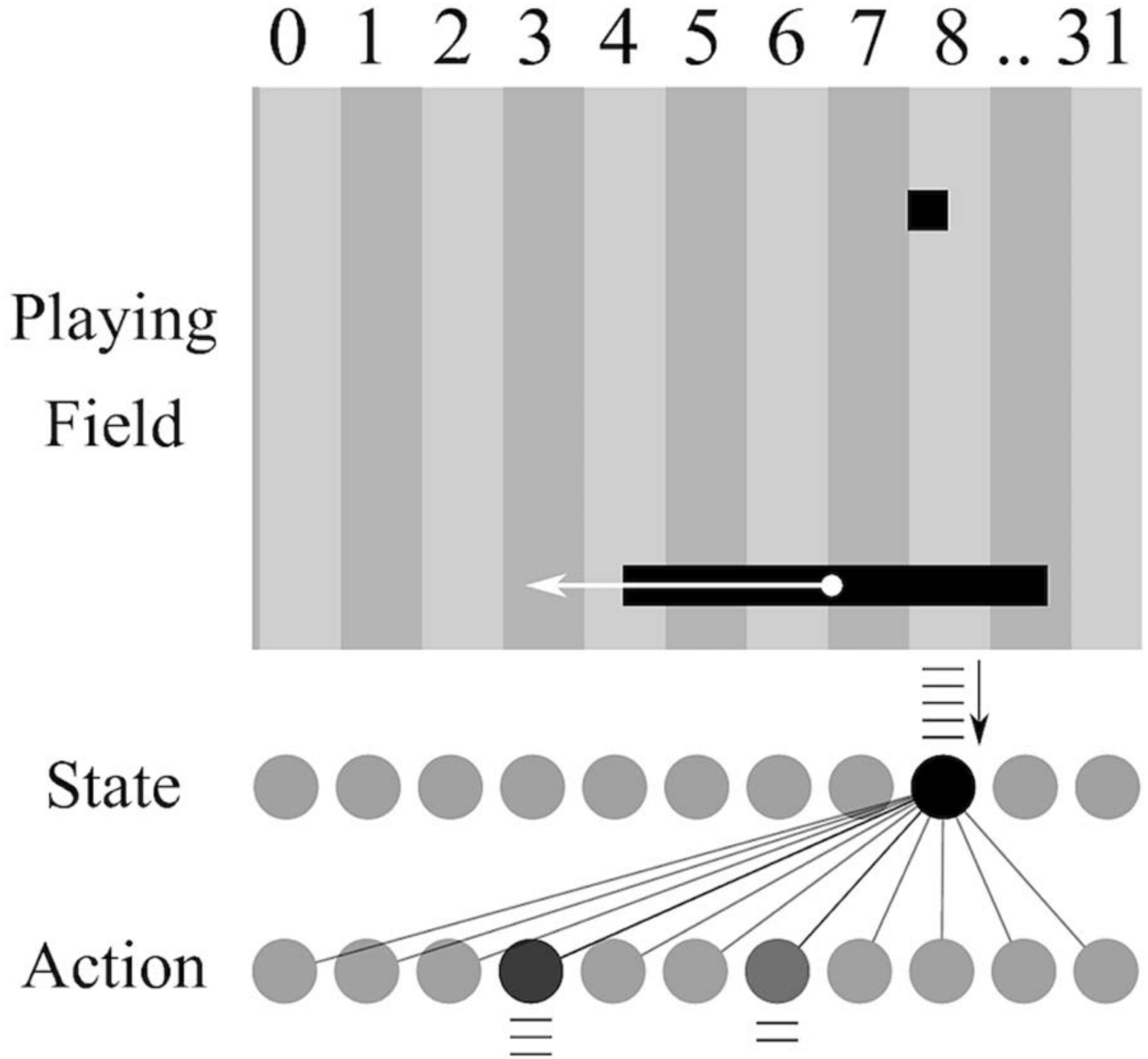




Aamir, S. A. et al. (2018). An Accelerated LIF Neuronal Network Array for a Large Scale Mixed-Signal Neuromorphic Architecture. IEEE Transactions on Circuits and Systems I: Regular Papers , 1 14

Wunderlich et al., 2018, to be submitted to Frontiers of Neuromorphic Engineering

PONG : Closed action-perception loop

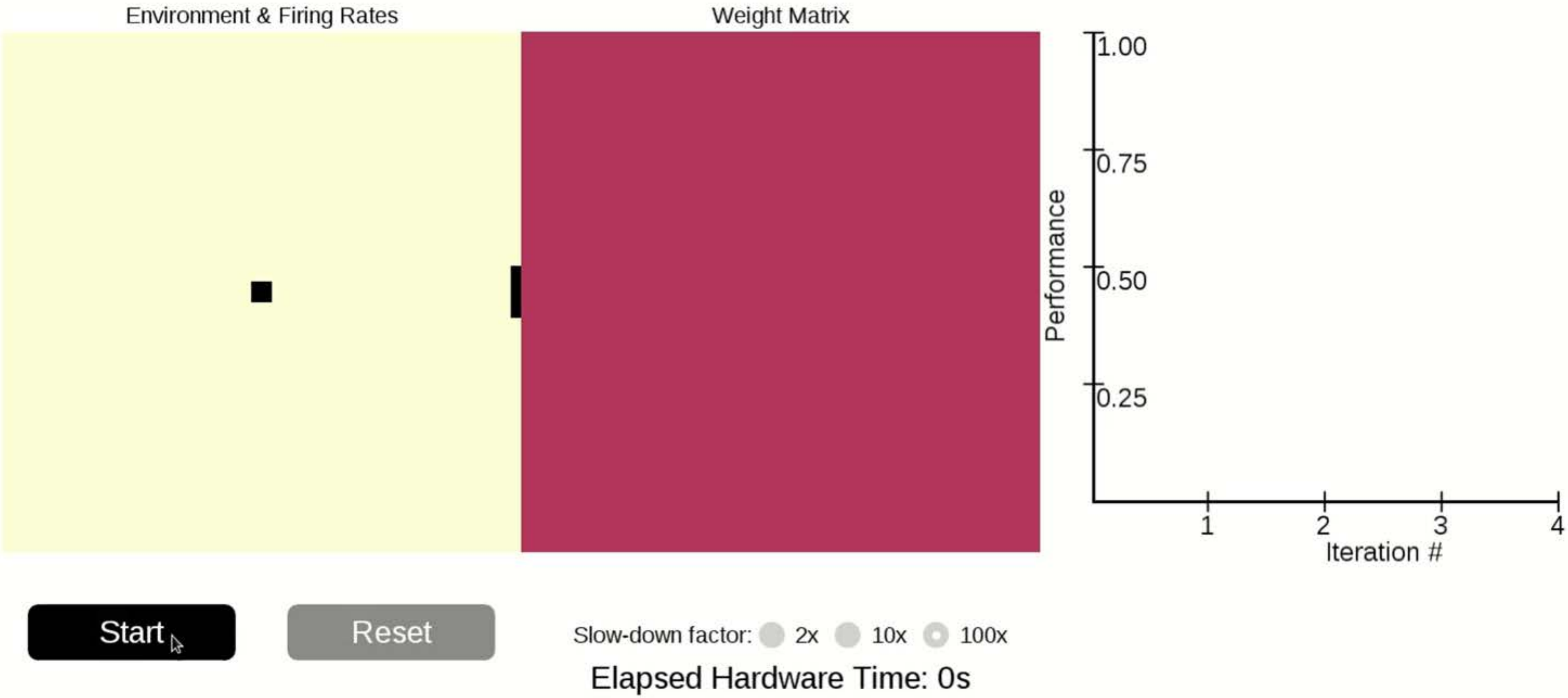


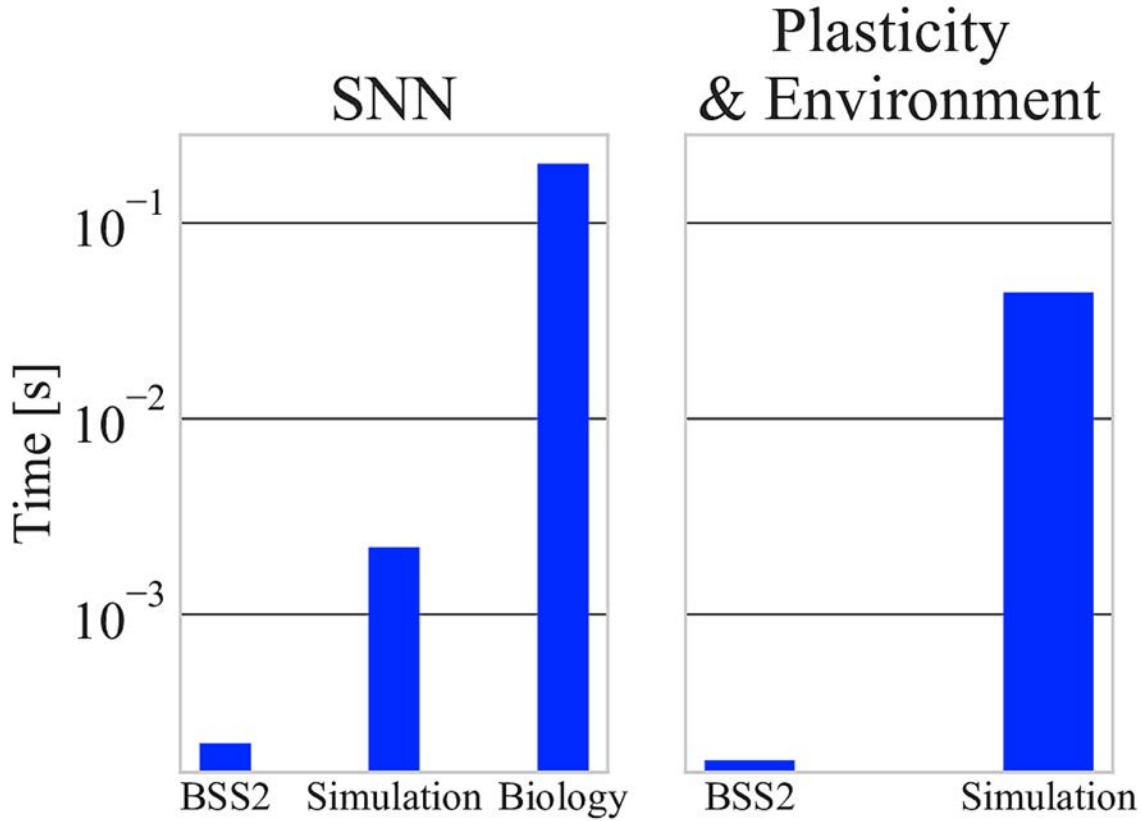
- Reinforcement learning by reward-modulated spike-timing-dependent plasticity
- R-STDP : three-factor learning rule that modulates the effect of unsupervised STDP using a reward signal
- Inspired by the activity of dopamine in the brain found to encode expected reward (e.g. Hollerman and Schultz, 1998)

$$\Delta w_{ij} = \beta \cdot (R - b) \cdot e_{ij}$$

Wunderlich et al., 2018, to be submitted to Frontiers of Neuromorphic Engineering

Local Learning Demo on BrainScaleS 2





220 µs	2.3 ms	200 ms	180 µs	50 ms
13 µJ	55 mJ	O(pJ)	10 µJ	1.2 J

BSS2 : Chip   Simulation : NEST   Biology : Real time

Speed and energy

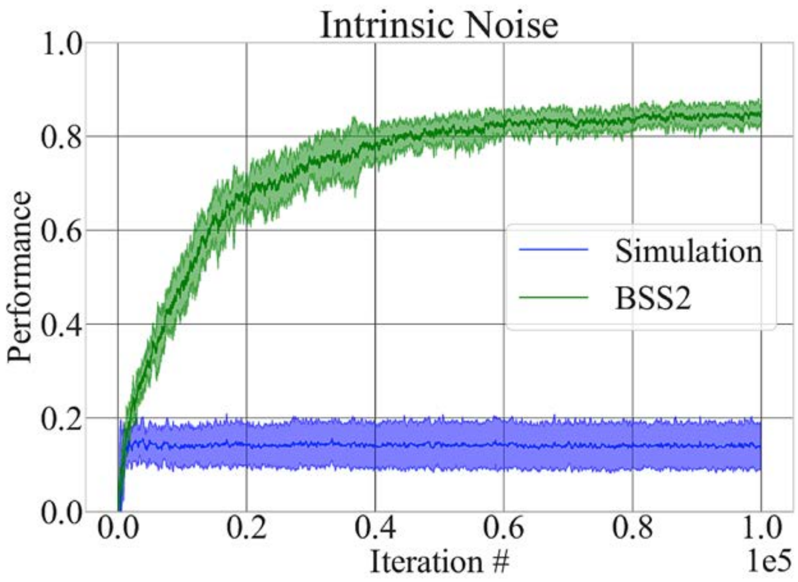
Times and energies for single experiment iterations

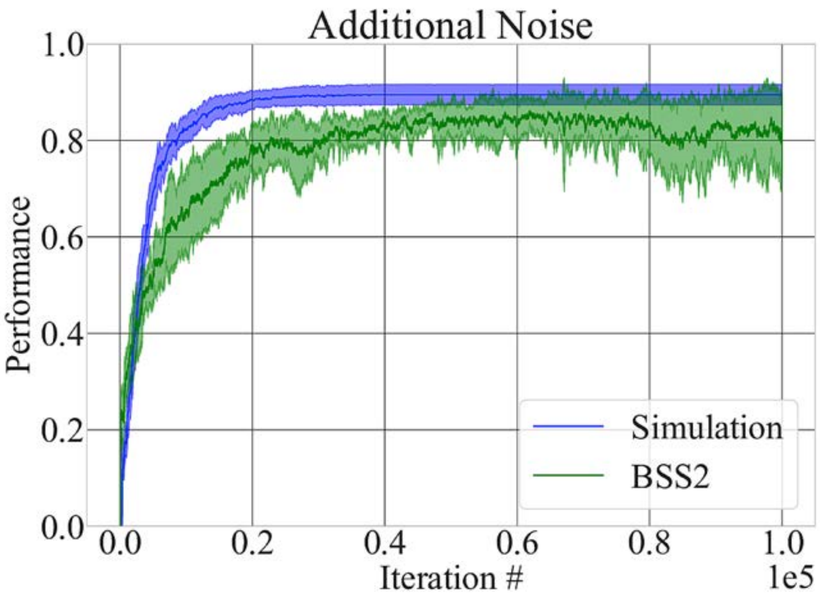
Energy measured directly through procesdsor curent

Simulation running on one single core of an Intel i7-4771 CPU

Wunderlich et al., 2018, to be submitted to Frontiers of Neuromorphic Engineering

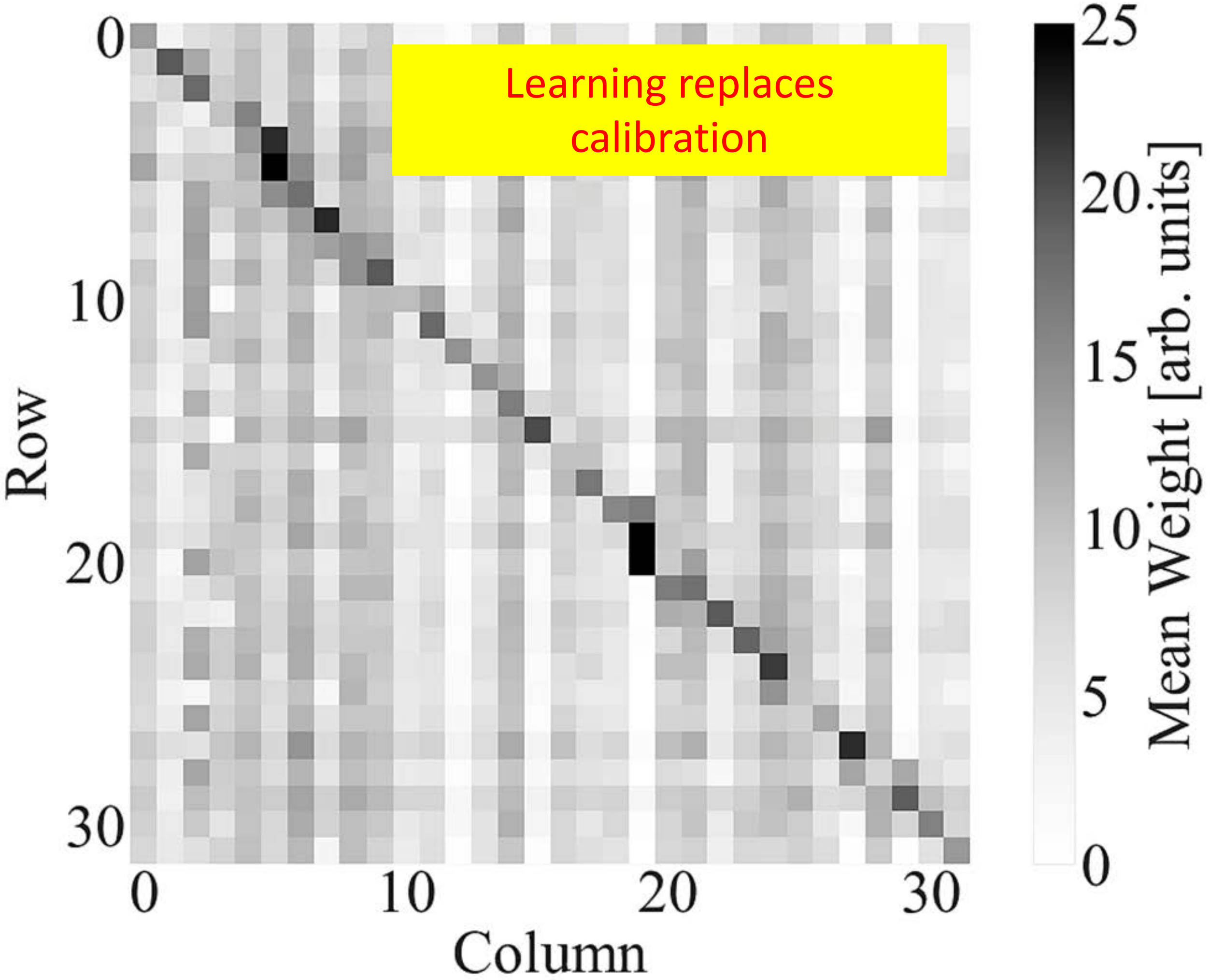
Learning success in software (NEST) with and without artificial noise and hardware with intrinsic noise





Exploiting noise

Wunderlich et al., 2018, to be submitted to Frontiers of Neuromorphic Engineering



Weight matrix on chip after 10^5 iterations

Variabiluity after learning represents self-calibration

Learning replaces calibration

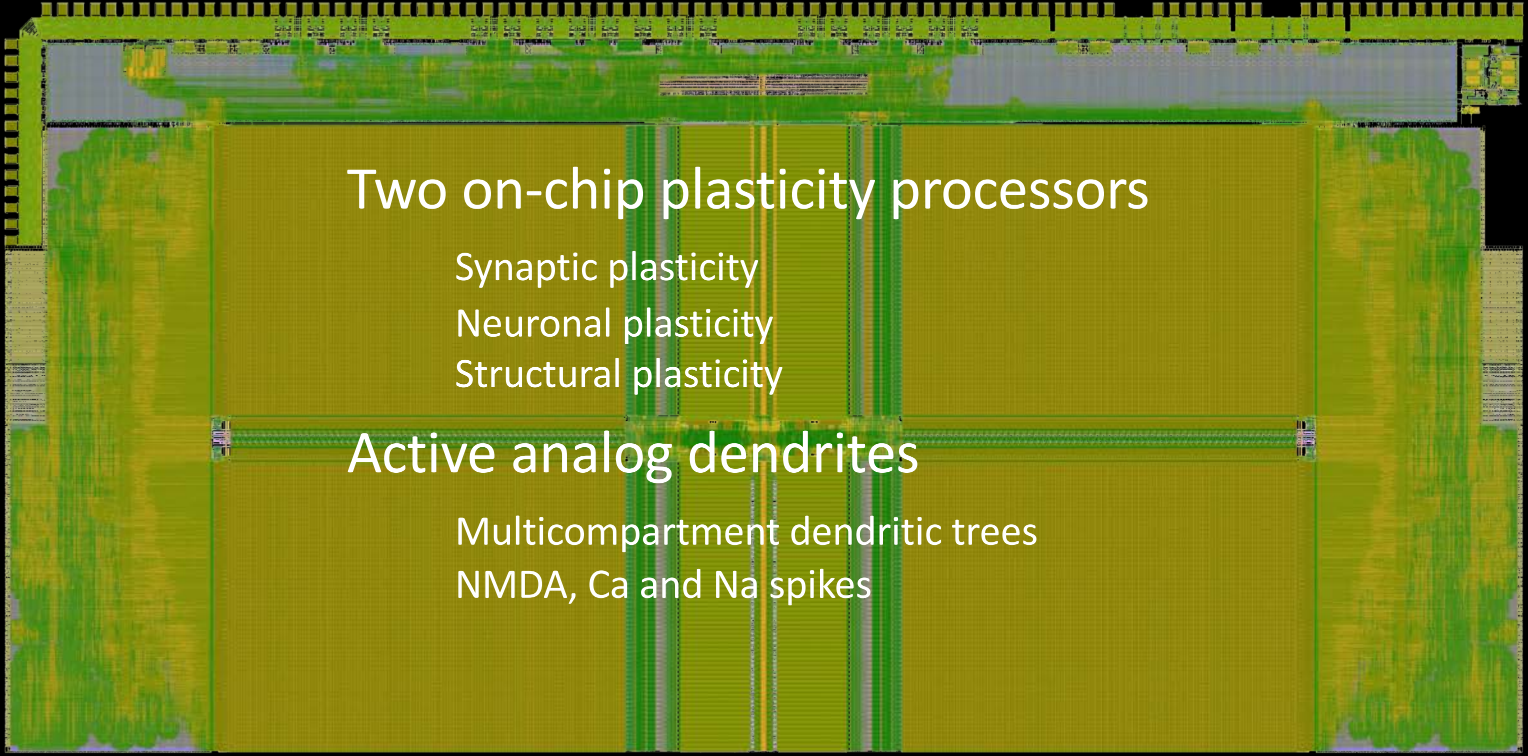
Wunderlich et al., 2018, to be submitted to Frontiers of Neuromorphic Engineering

Scalability

BrainScaleS-2

Next generation neuromorphic computing

Hybrid architecture analog-custom digital-processor nased



Two on-chip plasticity processors

- Synaptic plasticity
- Neuronal plasticity
- Structural plasticity

Active analog dendrites

- Multicompartment dendritic trees
- NMDA, Ca and Na spikes

Coffee machines ?

More general : Distributed sensing / IoT

IoT Challenges

- Energy consumption
- Size
- Data transfer
- Local preporcessing
- Adaptation

NMC technology

- Energy efficiency
- Compactness
- Local intelligence (learning)
- Adaptability

All enabled by biological priciples

Concrete examples

- Personal medical devices (EEG, ECG, drug delivery, nrve stimulation, brain implants, sensory substituation)
- Intelligent adaptive control (engines, manufacturing plants)

Project evaluation under way with BASF, Huawei, Airbus, Daimler

Why do we need big HBP systems, then ?

- Rapid prototyping (like LTL, continuous learning, one-shot learning, dendritic computation)
- Neural FPGA concept