Editorial: Focus issue on machine learning for neuromorphic engineering

Other Journal Item

Author(s):
Payvand, Melika; Neftci, Emre; Zenke, Friedemann

Publication date:
2023-09-01

Permanent link:
https://doi.org/10.3929/ethz-b-000631833

Originally published in:
Neuromorphic Computing and Engineering 3(3), https://doi.org/10.1088/2634-4386/acee1a

NEUROMORPHIC
Computing and Engineering

OPEN ACCESS

RECEIVED
24 July 2023

ACCEPTED FOR PUBLICATION
8 August 2023

PUBLISHED
17 August 2023

EDITORIAL

Editorial: Focus issue on machine learning for neuromorphic engineering

Melika Payvand[1], **Emre Neftci**[2,3] **and Friedemann Zenke**[4,5]

1 Institute of Neuroinformatics, University of Zürich and ETH Zürich, Zürich, Switzerland
2 Peter Grünberg Institute, Forschungszentrum Jülich, Jülich, Germany
3 Electrical Engineering and Information Technology, RWTH Aachen, Aachen, Germany
4 Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
5 Faculty of Natural Sciences, University of Basel, Basel, Switzerland

Spiking neural network (SNN) hardware is poised to reduce the energy for training and inference of neural networks by leveraging sparsity and locality in computation and learning. Despite tremendous progress in recent years, key challenges remain in translating scalable ideas from machine learning (ML) to neuromorphic engineering. These challenges span SNN algorithms, hardware, and applications. On the one hand, training algorithms SNNs need improvements. On the other hand, it is vital to exploit sparsity more effectively on the available hardware and enhance efficiency on real-world tasks well-suited for SNN processing.

This NCE Focus Issue is a collection of solutions for the above challenges. On the algorithmic side, Rossbroich *et al* (2022) propose a new initialization method through fluctuating the neurons for improved direct training of SNNs with backpropagation. Further, Moraitis *et al* (2022) introduce an unsupervised local training algorithm based on a combination of Hebbian plasticity with a soft winner take all mechanism. On the hardware side, two studies introduce new methods to exploit sparsity on current hardware (graphics processing unit (GPU) and field programmable gate array (FPGA)) to improve inference efficiency through Complementary Sparsity (Turner *et al* 2022) and Procedural connectivity (Hunter *et al* 2022). Finally, on the application side, DeWolf *et al* (2023) introduce a welcome closed-loop benchmark control task based on a robotic arm simulated in the popular Mojoco platform to showcase the inherent power efficiency and low latency of event-based computation.

SNNs and neuromorphic hardware are particularly well-suited to closed-loop control tasks, where latency and energy are paramount. In DeWolf *et al* (2023), the authors demonstrate an SNN-based controller that can control a robot arm simulated in the popular Mujoco platform. Their approach utilizes a mature set of tools to convert an adaptive controller into SNNs, leveraging deep neural networks and the neural engineering framework (Nengo). The former is used to obtain a reduced representation of the high-dimensional state space, and the latter is used to implement the algebraic equations of the adaptive controller. The resulting problem is then converted into a network that can run on the Intel Loihi research chip. This remarkable work showcases the inherent power efficiency and low latency of event-based computation in close-loop tasks and establishes an important benchmark for neuromorphic computing.

Optimal weight initialization is critical for deep neural network performance. However, efficient weight initialization strategies for deep spiking networks are largely missing. To fill this knowledge gap, Rossbroich *et al* (2022) propose a simple initialization strategy inspired by the balanced state in neurobiology whereby individual neurons receive fluctuation-driven input, which ensures unimpeded activity propagation in spiking convolutional neural networks (CNNs) up to seven layers deep. They show empirically that fluctuation-driven initialization results in better task performance on spatiotemporal problems and faster training. For training deeper SNNs, the authors propose a set of homeostatic loss functions that restore and maintain activity propagation and improve task performance.

Unlike traditional image classification, which relies on deep CNNs and requires millions of multiply-accumulate operations, SNNs communicate through binary spikes, reducing the need for multiplicative operations. Despite the popular solutions for directly training SNNs using, for example, the Surrogate Gradient approach, training of SNNs for deeper networks remains challenging. Existing ML libraries for SNNs based on PyTorch, JAX, and TensorFlow struggle with handling the sparsity of SNNs. To address this, Turner *et al* (2022) introduce the mlGenn platform, which employs procedural connectivity for 2D convolutions. In this technique,neurons' outgoing sparse random connectivity is regenerated on the fly when they spike rather than being stored in memory. This modification results in a 2.5× increase in simulation speed compared to the state-of-the-art SNN simulation tools.

Although the sparse activity and connectivity of SNNs should result in a proportional reduction in computing requirements, the irregular patterns of neuron interconnectivity and activity reduce the expected gains on current hardware. Hunter *et al* (2022) address this problem by structuring the sparsity to match the requirements of the target hardware for implementing sparse activation-sparse connectivity networks on FPGA. This restructuring is achieved by overlaying multiple sparse matrices to form a single dense structure if no two sparse matrices contain non-zero elements at the same location. This approach, which they call Complementary Sparsity, yields 100× improvement in throughput and energy efficiency on FPGA compared to optimized dense implementations.

Local unsupervised learning rules, which only require locally available information at the synapse, are highly desirable for neuromorphic engineering thanks to their efficiency and online capability. Yet, most local learning rules had limited success in achieving competitive task performance on real-world data. Moraitis *et al* (2022) revisit the problem of local learning. They derive a SoftHebb, a local learning rule for soft winner-take-all networks from a Bayesian generative framework and show empirically that for deep neural networks with a suitable positive expansion ratio, the learning rule minimizes cross-entropy in downstream supervised lazy learning settings in which only a linear readout head is trained.

The future for translating scalable ideas from machine to neuromorphic engineering looks bright and vital advances follow one another at the algorithms, hardware, and application level. This collection of articles highlights recent progress at all levels and showcases exciting advances in a dynamically expanding field.

Data availability statement

No new data were created or analysed in this study.

ORCID iD

Friedemann Zenke https://orcid.org/0000-0003-1883-644X

References

DeWolf T, Patel K, Jaworski P, Leontie R, Hays J and Eliasmith C 2023 *Neuromorph. Comput. Eng.* **3** 014007
Hunter K, Spracklen L and Ahmad S 2022 *Neuromorph. Comput. Eng.* **2** 034004
Moraitis T, Toichkin D, Journé A, Chua Y and Guo Q 2022 *Neuromorph. Comput. Eng.* **2** 044017
Rossbroich J, Gygax J and Zenke F 2022 *Neuromorph. Comput. Eng.* **2** 044016
Turner J P, Knight J C, Subramanian A and Nowotny T 2022 *Neuromorph. Comput. Eng.* **2** 024002