Gradient Computation With Sparse Observations for Analog Neuromorphic Hardware

Luca Blessing[1] Elias Arnold[1] Eric Müller[1,2] Johannes Schemmel[1] Christian Pehle[1]

[1]Electronic Vision(s) Group, Kirchhoff-Institute for Physics, Heidelberg University, Heidelberg, Germany

[2]European Institute for Neuromorphic Computing, Heidelberg University, Heidelberg, Germany

Introduction

Gradient-based learning has recently been demonstrated to be a suitable learning scheme for analog neuromorphic hardware [1, 2, 3, 7].

While previous approaches either make use of dense observations of neuron dynamics or are limited by assumptions on time constants, spiking behavior, and network topology, here we introduce an approach, based on the EventProp algorithm [8], that does only need spike-time observations from the neuromorphic hardware and is suitable for arbitrary network topologies and loss functions.

1. **Spike Observations are sufficient for estimating gradients.** By training a feed-forward network of 120 Leaky-Integrate and Fire (LIF) neurons on a low dimensional classification task, we demonstrate for the first time that EventProp can be used as an in-the-loop training algorithm for analog neuromorphic hardware.
2. **Numerical Gradient Estimates Converge.** We implement a numerical integration scheme for both the forward- and adjoint dynamics of LIF neurons [8] and show that the computed gradient converges to the analytically known solution in cases where an analytical gradient is known.
3. **Mean Hardware Gradient Estimates agree with Analytical Solution.** We show that the mean gradient estimate obtained using spike time hardware measurements by our algorithm agrees with the analytically known solution in cases where an analytical gradient is known.

Numerical Gradient Estimate Converges

We use a simple numerical integration scheme (explicit Euler integration) for both the forward and adjoint dynamics of a LIF neuron having exponential-shaped, current-based synapses [8]. To quantify the numerical error, we can compare to the analytical gradient in a special case ($\tau_{\text{syn}} = \tau_{\text{mem}}$, $\tau_{\text{refrac}} = 0$) for which an explicit formula is available, cf. [3]. We consider a situation where a LIF neuron receives a single spike with weight w (Fig. 1).

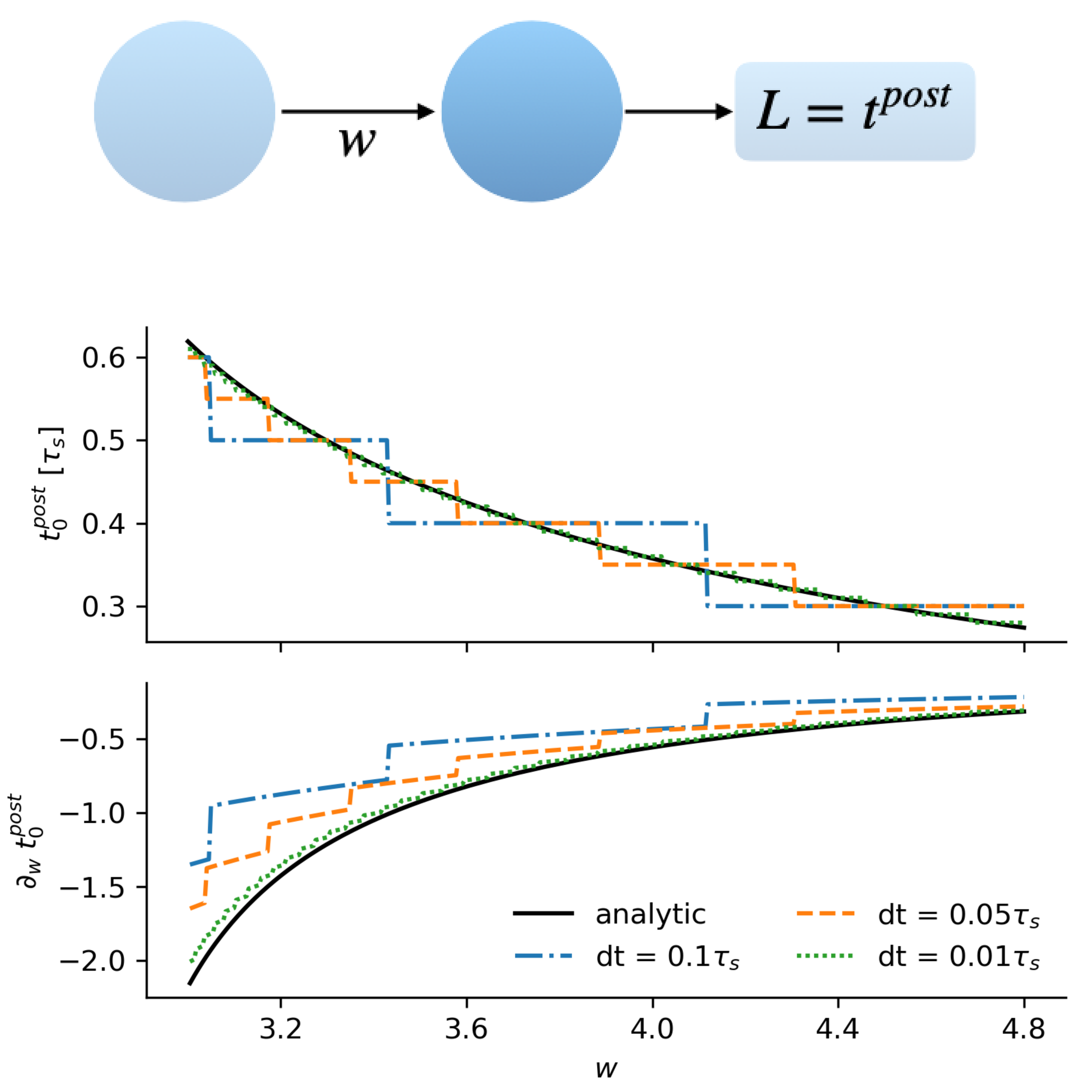

Figure 1. Experiment setup (top), spike time (center) and estimated gradient (bottom) for a loss function $L = t^{\text{post}}$ as a function of the weight w. As the numerical integration timestep $\mathrm{d}t$ decreases the numerically estimated gradient (dashed lines) converges to analytically known gradient (solid-black line), cf. Göltz et al. [3].

Funding & Acknowledgements

This research has received funding from the EC Horizon 2020 Framework Programme under Grant Agreements 785907 (HBP SGA2) and 945539 (HBP SGA3) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Fundation) under Germany's Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster). The authors wish to thank all present and former members of the Electronic Vision(s) research group contributing to the BrainScaleS-2 neuromorphic platform.

Hardware Training Results on Yin-Yang

We choose the Yin-Yang dataset [4], a simple two-dimensional classification task with three classes, for demonstrating the learning algorithm on BrainScaleS-2, an accelerated analog neuromorphic research platform [6, 5].

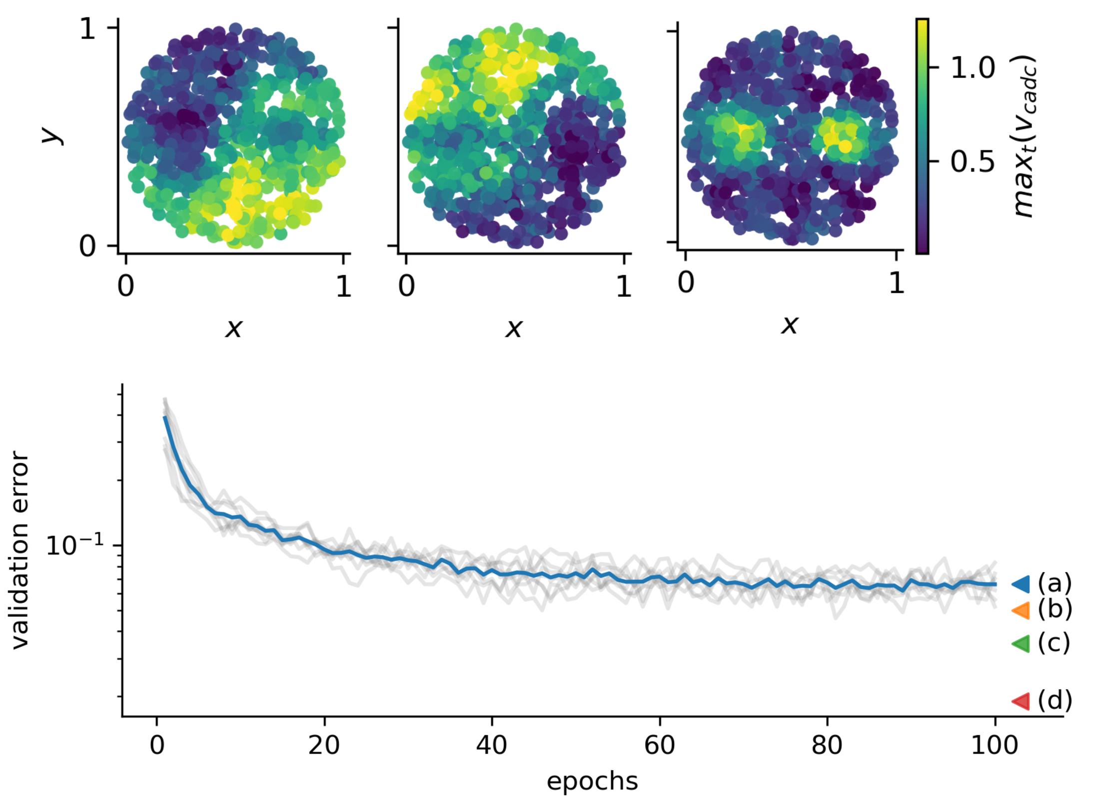

Figure 2. Maximum membrane values over time on validation set after training (top) and validation accuracy (bottom) for hardware-in-the-loop training with EventProp on the Yin-Yang dataset.

Gradient estimator			loss	test acc.
analytical [3]	sim.		TTFS	95.9 ±0.7
analytical [3]	hw	(b)	TTFS	95.0 ±0.9
EventProp [8]	sim.	(d)	TTFS	98.1 ±0.2
Surr. Gradient	sim.		Max	96.0 ±0.5
Surr. Gradient	hw		Max	93.8 ±0.8
EventProp	sim.	(c)	Max	96.5 ±0.7
EventProp	hw	(a)	Max	93.4 ±0.9

Table 1. Test accuracy on the Yin-Yang task for three gradient estimators, two loss definitions using numerical integration and hardware emulation. The results marked by 'TTFS' are using a loss based on time-to-first-spike decoding, the 'Max' results are based on 'maximum membrane value over time'.

Hardware Gradient Estimation

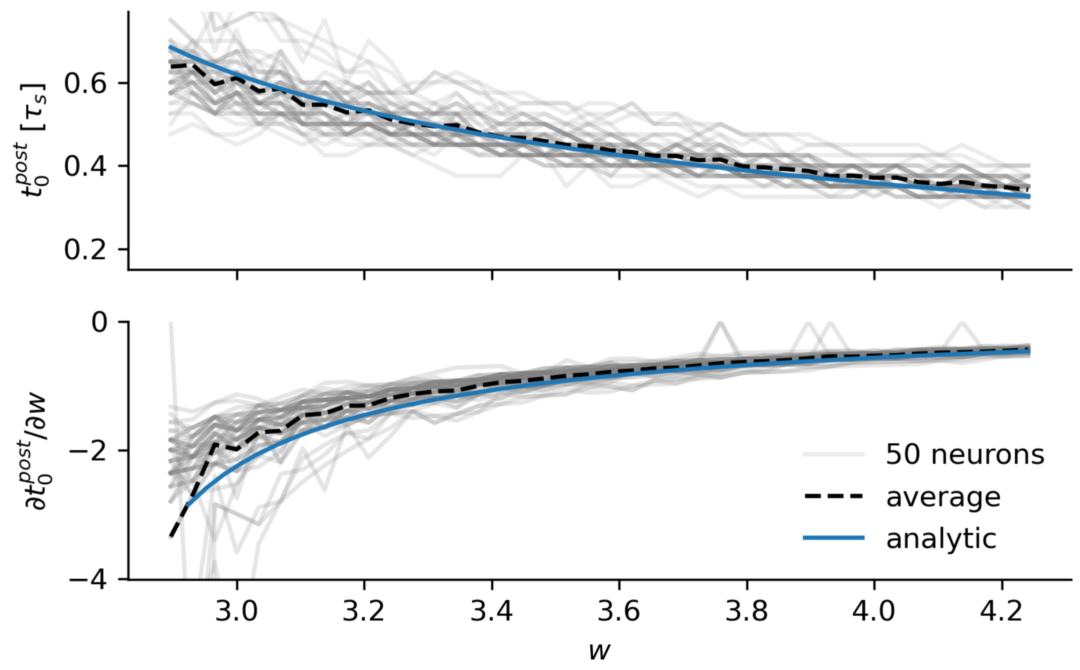

Figure 3. Spike time (top) acquired from runs on BrainScaleS-2 and estimated gradient (bottom) for a loss function $L = t^{\text{post}}$ as a function of the weight w. Results for $n = 50$ hardware neuron circuits are shown in light grey, the average in dashed black. The analytical result (blue line, see Fig. 1) agrees well with the hardware average.

Comparison to Surrogate Gradient

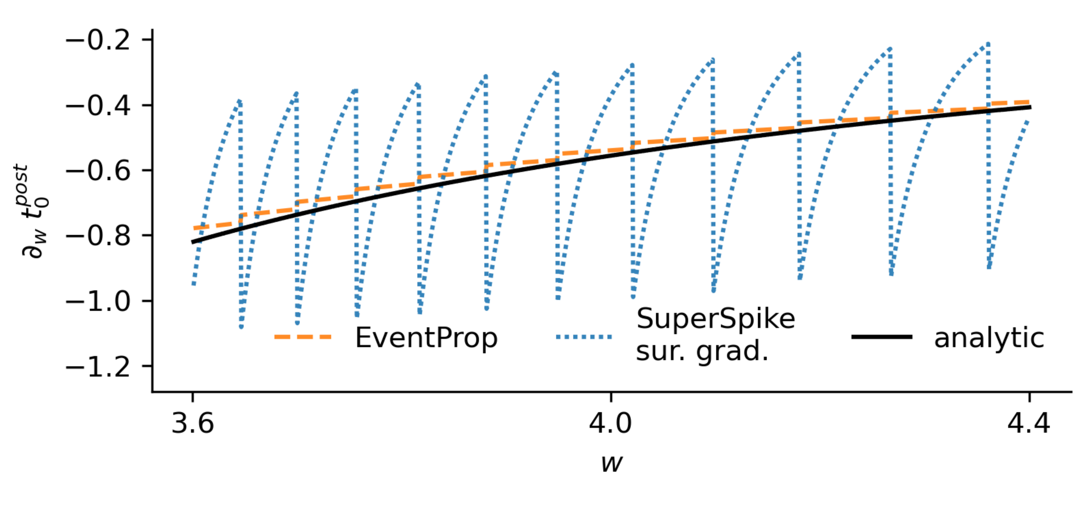

Figure 4. Numerical comparison of surrogate and EventProp gradient estimate evaluated with the same numerical precision to the analytical result. Same experiment as in Fig. 1.

Hardware-in-the-loop Training

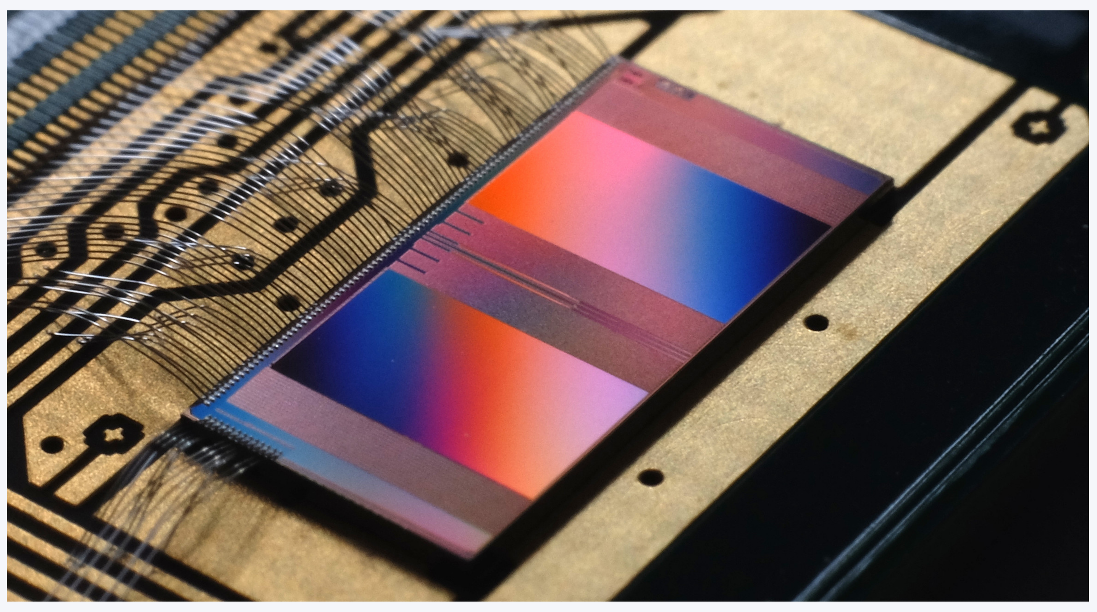

Figure 5. A BrainScaleS-2 neuromorphic chip (4 mm × 8 mm) bonded to a carrier board, see [6] for detailed information.

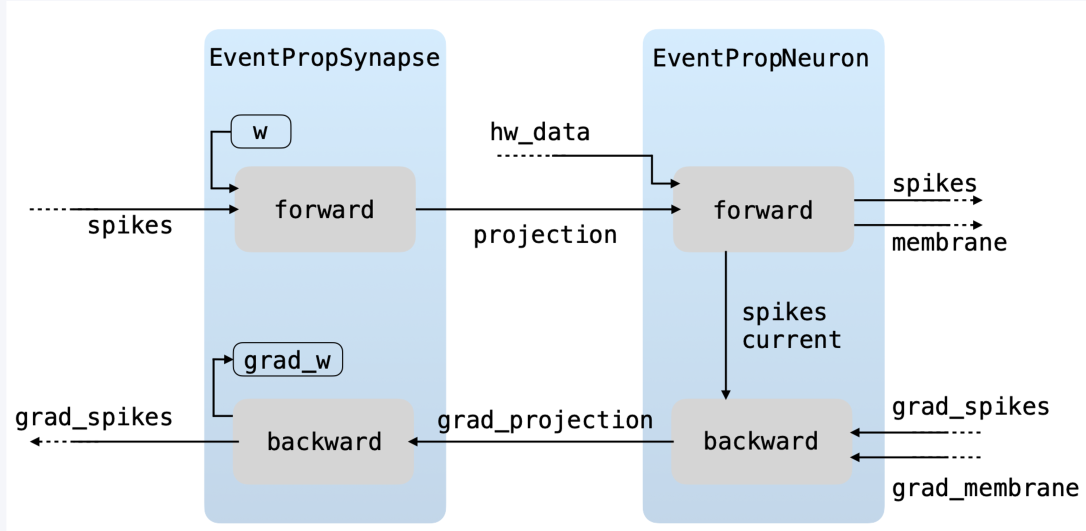

Figure 6. Dataflow occuring during hardware-in-the-loop training.

Discussion

We have demonstrated a gradient estimation algorithm for analog neuromorphic hardware, which only requires spike observations and makes no assumptions on network topology or loss function. This has the potential to enable scalable gradient estimation in large-scale neuromorphic hardware, as the continuous measurement of observables would be prohibitive in this case. While the results reported here are encouraging further work is needed in several areas:

- We would like to demonstrate the algorithm on further tasks, particularly ones that are not feasible using surrogate-gradient-based in-the-loop training due to hardware trace-memory limitations.
- We would like to use hardware observables to learn the dynamics, instead of assuming a particular model.

References

[1] E. Arnold, G. Böcherer, E. Müller, P. Spilger, J. Schemmel, S. Calabrò, and M. Kuschnerov. Spiking neural network equalization on neuromorphic hardware for IM/DD optical communication. In *Proc. Eur. Conf. Optical Commun. (ECOC)*, Basel, Switzerland, September 2022. Paper Th1C.5. arXiv:2206.00401.

[2] B. Cramer, S. Billaudelle, S. Kanya, A. Leibfried, A. Grübl, V. Karasenko, C. Pehle, K. Schreiber, Y. Stradmann, J. Weis, J. Schemmel, and F. Zenke. Surrogate gradients for analog neuromorphic computing. *Proceedings of the National Academy of Sciences*, 119(4):e2109194119, 2022. doi:10.1073/pnas.2109194119.

[3] J. Göltz, L. Kriener, A. Baumbach, S. Billaudelle, O. Breitwieser, B. Cramer, D. Dold, A. F. Kungl, W. Senn, J. Schemmel, K. Meier, and M. A. Petrovici. Fast and energy-efficient neuromorphic deep learning with first-spike times. *Nature machine intelligence*, 3(9):823–835, 2021. doi:10.1038/s42256-021-00388-x.

[4] L. Kriener, J. Göltz, and M. A. Petrovici. The yin-yang dataset. In *Neuro-Inspired Computational Elements Conference*, pages 107–111, 2022. doi:10.1145/3517343.3517380.

[5] E. Müller, E. Arnold, O. Breitwieser, M. Czierlinski, A. Emmel, J. Kaiser, C. Mauch, S. Schmitt, P. Spilger, R. Stock, Y. Stradmann, J. Weis, A. Baumbach, S. Billaudelle, B. Cramer, F. Ebert, J. Göltz, J. Ilmberger, V. Karasenko, M. Kleider, A. Leibfried, C. Pehle, and J. Schemmel. A scalable approach to modeling on accelerated neuromorphic hardware. *Front. Neurosci.*, 16, 2022. doi:10.3389/fnins.2022.884128.

[6] C. Pehle, S. Billaudelle, B. Cramer, J. Kaiser, K. Schreiber, Y. Stradmann, J. Weis, A. Leibfried, E. Müller, and J. Schemmel. The BrainScaleS-2 accelerated neuromorphic system with hybrid plasticity. *Front. Neurosci.*, 16, 2022. doi:10.3389/fnins.2022.795876.

[7] S. Schmitt, J. Klähn, G. Bellec, A. Grübl, M. Guettler, A. Hartel, S. Hartmann, D. Husmann, K. Husmann, S. Jeltsch, V. Karasenko, M. Kleider, C. Koke, A. Kononov, C. Mauch, E. Müller, P. Müller, J. Partzsch, M. A. Petrovici, S. Schiefer, S. Scholze, V. Thanasoulis, B. Vogginger, R. Legenstein, W. Maass, C. Mayr, R. Schüffny, J. Schemmel, and K. Meier. Neuromorphic hardware in the loop: Training a deep spiking network on the brainscales wafer-scale system. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2227–2234. IEEE, 2017. doi:10.1109/IJCNN.2017.7966125.

[8] T. C. Wunderlich and C. Pehle. Event-based backpropagation can compute exact gradients for spiking neural networks. *Scientific Reports*, 11(1):1–17, 2021. doi:10.1038/s41598-021-91786-z.