Model-Free Optimization Algorithms for Physical Neural Networks

A Skalli[1], S Sunada[2], M Goldmann[3], M Gebski[4], N Haghighi[5], S Reitzenstein[5], J A Lott[5], T Czyszanowski[4], and D Brunner[1]

[1]*Optics, Université Marie et Louis Pasteur, CNRS, institut FEMTO-ST, Besançon, France*
[2]*Mechanical Engineering, Faculty of Mechanical Engineering, Institute of Science and Engineering, Kanazawa University, Kanazawa, Japan*
[3]*Akhetonics GmbH, Berlin, Germany*
[4]*Institute of Physics, Lodz University of Technology, Łódź, Poland*
[5]*Institut für Festkörperphysik, Technische Universität Berlin, Berlin, Germany*
Contact Email: anas.skalli@femto-st.fr

Artificial neural networks (ANNs), have become ubiquitous and revolutionized many applications ranging from computer vision to medical diagnoses. However, they offer a fundamentally connectionist and distributed approach to computing, in stark contrast to classical computers that use the von Neumann architecture. This distinction has sparked renewed interest in developing unconventional hardware to support more efficient implementations of ANNs, rather than merely emulating them on traditional systems. Photonics stands out as a particularly promising platform, providing scalability, high speed, energy efficiency, and the ability for parallel information processing.

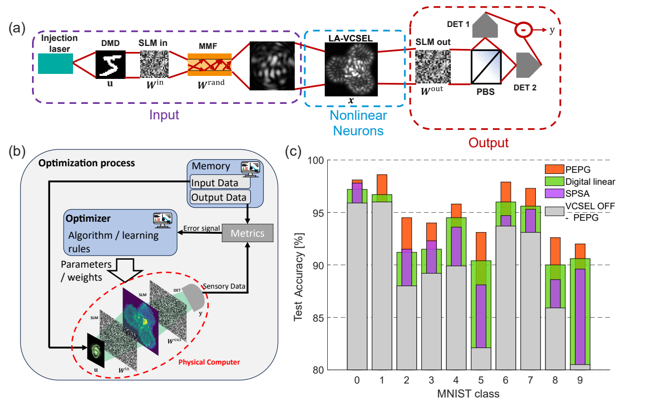

Figure 1: (a) Working principle of our experimental photonic neural network. (b) Schematic outlining the model free optimization process of a physical computer. (c) Test accuracy on the MNIST dataset for different model free algorithms

However, fully realized autonomous optical neural networks (ONNs) with *in situ* learning capabilities are still rare. In this work, we demonstrate a fully autonomous and parallel ONN using a multimode vertical cavity surface emitting laser (VCSEL) using off-the-shelf components. Our ONN is highly efficient and is scalable both in network size and inference bandwidth towards the GHz range. Moreover, high performance hardware-compatible optimization algorithms are necessary in order to minimize reliance on external von Neumann computers to fully exploit the potential of ONNs [4,5]. As such we present and extensively study several algorithms which are broadly compatible with a wide range of systems. We then apply these algorithms to optimize our ONN, and benchmark them using the MNIST dataset. We show that our ONN can achieve high accuracy and convergence efficiency, even under limited hardware resources. Crucially, we compare these different algorithms in terms of scaling and optimization efficiency in term of convergence time which is crucial when working with limited external resources. Our work provides some guidance for the design of future ONNs as well as a simple and flexible way to train them.

The ONN presented here was first implemented in [2,3]. Whereas previously it followed the reservoir computing (RC) concept, where input and internal weights are fixed and only the output weights are trained, we now present an improved version of the setup where all connections in the ONN can be trained yielding therefore a highly tunable network [4]. The experimental scheme is shown in Fig. 1(a). Input images u displayed on a digital micromirror device (DMD) are passed through a phase mask displayed on a spatial light modulator (SLM) which encodes the input weights $\mathbf{W}^{\text{in}}$. The phase modulated input is injected onto the LA-VCSEL through a multimode fiber (MMF) which passively implements a random linear mixing $\mathbf{W}^{\text{rand}}$. The VCSEL then transforms the injected information non-linearly yielding the perturbed mode profile x. The final part of the PNN is its output layer. The VCSEL's surface is imaged onto a second SLM, operating in intensity modulation giving us analog output weights Wout, positive and negative weights are obtained via balanced detection through polarization filtering via subtracting signals from DET_1 and DET_2 in real time. Thus 5000+ nodes are implemented fully in parallel. $\mathbf{W}^{\mathrm{out}}$ and $\mathbf{W}^{\mathrm{in}}$ are trained via iterative optimization based on evolutionary search algorithms or gradient descent using gradient estimation methods from reinforcement learning as shown in Fig. 1(b).

In addition, we implemented and extensively studied the hyperparameter behavior of different hardware compatible optimization algorithms, namely SPSA, PEPG, CMA-ES and FD, to train our ONN. We also compared the convergence efficiency of these strategies in the context of hardware optimization under limited digital resources for a wide range of LA-VCSELs, and found that PEPG is the most efficient and high-performance algorithm. In contrast, due to its computational overhead, CMA-ES is not well suited for high dimensional optimization problems in an autonomous hardware context. Finally, SPSA offers somewhat of a happy medium between to two. We then studied the performance of PEPG and SPSA for all classes of the MNIST dataset and compared them to two baselines. The first being a digital linear classifier, while the second was our ONN with the LA-VCSEL switched off, which corresponds to a hardware linear system. We found that PEPG is able to overcome the digital linear classifier limit of 93.37 %, while SPSA approaches it yet falls short as shown in Fig. 1(c).

Crucially, both outperform the linear hardware system accuracy of 89.43% by a significant margin achieving 95.39 and 92.71% for PEPG and SPSA respectively, highlighting the high dimensional nature of the transformation produced by the LA-VCSEL. Notably, a significant portion of physical NNs in the literature still rely on offline, *i.e.* software weights on a digital computer. This approach, while understandable since it offers great flexibility, is not in the spirit of building autonomous computing systems. Crucially, it artificially inflates the performance of physical NNs while tying them down, in most cases, to a slow digital computer handling the states of the system and applying software weights, which undermines the potential gains in terms of speed and energy efficiency that physical systems might otherwise offer.

In conclusion, our study provides a robust framework for training large-scale physical neural networks based on model-free optimization algorithms that are broadly applicable to a wide range of systems, we also study the impact of these algorithms in terms of energy and discuss the cost of optimization which is often overlooked. We also introduce experimental innovations that overcome challenges in the field. The combination of model-free optimization, high-resolution tunable weight control, and an advanced VCSEL-based platform culminates in an ONN that is both high-performing and energy-efficient.

References

[1] A Skalli, J Robertson, D Owen-Newns, M Hejda, X Porte, S Reitzenstein, A Hurtado and D Brunner, Opt. Mater. Express **12**, 2395 (2022)

[2] X Porte, A Skalli, N Haghighi, S Reitzenstein, J A Lott and D Brunner, J. Phys. Photonics **3**, 024017 (2021)

[3] A Skalli, X Porte, N Haghighi, S Reitzenstein, J A Lott and D Brunner, Opt. Mater. Express **12**, 2793 (2022)

[4] A Skalli, S Sunada, M Goldmann, M Gebski, S Reitzenstein, J A Lott, T Czyszanowski and D Brunner, arXiv:2503.16943 (2025)

[5] A Momeni, B Rahmani, B Scellier, *et al.*, arXiv:2406.03372 (2024)