Noise-Aware Training of Dynamical Physical Neural Networks of Spintronic Nanodevices

Luca Manneschi[1], Ian T. Vidamour[1], Kilian D. Stenning[2], Charles Swindells[1], Guru Venkat[1],
David Griffin[3], Lai Gui[2], Daanish Sonawala[2], Denis Donskikh[2], Dana Hariga[1], Elisa Donati[4], Susan Stepney[3],
Will R. Branford[2], Jack C. Gartside[2], Thomas J. Hayward[1], Matthew O. A. Ellis[1] and Eleni Vasilaki[1]
[1]The University of Sheffield, Sheffield, S1 4DP, United Kingdom
[2]Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom
[3]University of York, Heslington, York, YO10 5DD, United Kingdom
[4]Institute of Neuroinformatics, University of Zurich and ETHZ, Zurich, Switzerland

Spintronic nano-devices hold significant promise for energy-efficient neuromorphic computing, thanks to their non-volatility, efficient control methods, and complex non-linear dynamics. However, their inherent noise and stochasticity can often hinder performance and necessitate off-chip training. This talk introduces a framework for constructing dynamical physical neural networks (PNNs) from spintronic nano-devices, specifically for temporal tasks, using differentiable digital twins. By incorporating noise directly into the training process via data-driven stochastic models, this approach captures the devices' variability, enabling rapid prototyping, off-device training, and informed network design. Noise-aware training significantly improves transferability to hardware, outperforming deterministic models in tasks such as smart prosthetic gesture prediction, while also requiring fewer experimental measurements than alternative methods. This work highlights how the noisy nature of spintronic systems can be harnessed as a valuable asset for unconventional computing.

Index Terms—**Physical Neural Networks, Neuromorphic Computing, Spintronics.**

I. Introduction

Spintronic devices hold significant promise as physical substrates for neuromorphic computing, thanks to their non-volatile and non-linear behaviour, with potential for low-energy control [1], [2]. When interconnected into physical neural networks (PNNs), these devices can exhibit complex emergent behaviours that go beyond the principles of reservoir computing, offering potential for richer computational capabilities directly within the hardware—an approach especially relevant for edge computing tasks.

In conventional artificial neural networks, the connection weights between neurons are optimised by directly calculating the error gradient via backpropagation. In physical systems, however, measuring gradients directly is often inefficient or impractical due to a lack of accurate models, the presence of noise, variability, and device-specific limitations. As a result, training is commonly performed using alternative methods that estimate gradients indirectly—often by simulating the device's behaviour in software while grounding the training methodologies in experimental data wherever feasible. Examples include Physics-Aware Training (PAT) [3] which uses a digital twin of the physical system to compute approximate gradients, but so far it has primarily been demonstrated on static, feed-forward tasks such as image classification.

Notably, many physical systems naturally exhibit temporal dynamics, which may be particularly advantageous for solving tasks such as signal transformation, time-series forecasting, or classification of long-term dependencies, but these capabilities remain largely under-explored in current techniques. In this work, we present an approach that extends the PAT methodology to address such tasks while crucially accounting for the stochastic variability in device responses, enabling robust learning and deployment.

II. Noise-aware training of physical neural networks

In physical systems their inherent stochasticity and noise can be detrimental, especially when the connection weights are trained off-device. To avoid this 'simulation-reality gap', we present an approach based around the creation of stochastic digital twins that are trained to model the device responses and provide a analogue for training off-device using powerful gradient-based approaches. This approach leverages advanced machine learning models based on stochastic differential equations (SDEs) [4] to create device responses that mimic experimental measurements with similar noise distributions. This allows for off-device training using backpropagation-through-time (BPTT) to learn the interconnection weights between devices to effectively solve a temporal task. Since the off-device training observes realistic noisy device responses, the learning process is able to account for this and create a more robust solution. This means when the interconnection weights are transferred to the real devices for testing, there is minimal drop in performance. An overview of this process if shown in Fig. 1 for a PNN comprised of nanomagnetic ring array.

III. Results

The first stage of this approach involves training models that replicate the responses of target nanomagnetic systems. In this work, we use neural stochastic differential equations (SDEs) [4], which employ a small deep neural network to learn functions for both the deterministic and stochastic time derivatives.

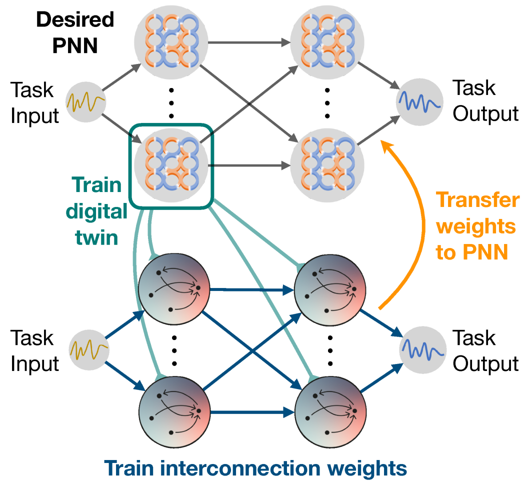

Fig. 1. Offline training loop for physical neural networks. First, a stochastic digital twin based on a neural-SDE is trained on individual spintronic devices to allow offline training using backpropagation-through-time. The interconnection weights are then trained in simulation using realistic responses generated by the model. Finally, the optimised weights are transferred to the experimental system for inference.

This method was applied to two candidate spintronic nanodevices: a nanomagnetic ring array (NRA) [5] and artificial spin vortex ice (ASVI) [6], both of which have previously been investigated for neuromorphic computing. Figure 2(a) shows the model's predicted device response for the NRA device under a given driving magnetic field. Repetitions of the same field sequence produce different trajectories due to device stochasticity, and our trained model successfully captures this behaviour. This approach yields a noisy output that reflects the inherent variability of the physical system and may assist in identifying parameter regimes during the optimisation process that avoid large noise fluctuations.

An example of this is shown in Fig. 2(b). The task uses the chaotic Mackey-Glass time series as a benchmark for regression performance, requiring the system to forecast five steps ahead—a standard challenge in neuromorphic spintronic systems [7]. In this result, we train a three-layer neural network using the device models in simulation. The optimised connection weights are then transferred to the experimental system, where a single device is time-multiplexed to emulate multiple devices. In this case, we observe a mean squared error approximately an order of magnitude lower than existing results in the literature [7]. Furthermore, we apply this methodology to a temporal classification task based on handwritten digits, where a two-layer PNN achieves comparable improvements. In both cases, incorporating stochastic effects is crucial for the successful transfer of weights to the physical network.

In summary, we have proposed a framework for training physical neural networks composed of spintronic devices using stochastic models. These models provide noisy predictions of device dynamics under input signals, enabling simulated optimisation of the connectivity that can then be transferred to the physical system with minimal loss of accuracy. This approach is general and can be applied across a range of neuromorphic systems.

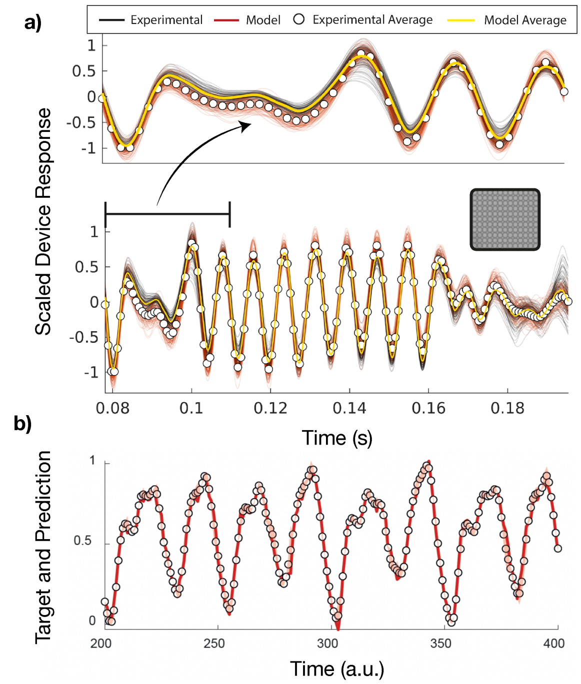

Fig. 2. Results of a) training a neural-SDE to predict the dynamics of a nanomagnetic ring array response, and b) testing of the physical neural network to forward predict the Mackey-Glass task.

Acknowledgements

The authors gratefully acknowledge funding from the EPSRC (Projects No. EP/V006339/1, EP/V006029/1, EP/W524335/1, EP/Y003276/1 and EP/S009647/1), The Eric and Wendy Schmidt Fellowship Program, Horizon 2020 FET-Open (Agreement No 861618), and the Royal Academy of Engineering.

REFERENCES

[1] J. Grollier, D. Querlioz, K. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, "Neuromorphic spintronics," *Nature electronics*, vol. 3, no. 7, pp. 360–370, 2020.

[2] D. A. Allwood, M. O. Ellis, D. Griffin, T. J. Hayward, L. Manneschi, M. F. Musameh, S. O'Keefe, S. Stepney, C. Swindells, M. A. Trefzer *et al.*, "A perspective on physical reservoir computing with nanomagnetic devices," *Applied Physics Letters*, vol. 122, no. 4, 2023.

[3] L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, and P. L. McMahon, "Deep physical neural networks trained with backpropagation," *Nature*, vol. 601, no. 7894, pp. 549–555, 2022.

[4] P. Kidger, J. Foster, X. Li, and T. J. Lyons, "Neural sdes as infinite-dimensional gans," in *International conference on machine learning*. PMLR, 2021, pp. 5453–5463.

[5] R. W. Dawidek, T. J. Hayward, I. T. Vidamour, T. Broomhall, G. Venkat, M. Al Mamoori, A. Mullen, S. J. Kyle, P. Fry, N.-J. Steinke, J. Cooper, F. Maccherozzi, S. S. Dhesi, L. Aballe, M. Foerster, J. Prat, E. Vasilaki, M. O. A. Ellis, and D. A. Allwood, "Dynamically-Driven Emergence in a Nanomagnetic System," *Advanced Functional Materials*, vol. 31, no. 8, 2021.

[6] J. C. Gartside, K. D. Stenning, A. Vanstone, H. H. Holder, D. M. Arroo, T. Dion, F. Caravelli, H. Kurebayashi, and W. R. Branford, "Reconfigurable training and reservoir computing in an artificial spin-vortex ice via spin-wave fingerprinting," *Nature Nanotechnology*, vol. 17, no. 5, pp. 460–469, 2022.

[7] K. D. Stenning, J. C. Gartside, L. Manneschi, C. T. Cheung, T. Chen, A. Vanstone, J. Love, H. Holder, F. Caravelli, H. Kurebayashi *et al.*, "Neuromorphic overparameterisation and few-shot learning in multilayer physical neural networks," *Nature Communications*, vol. 15, no. 1, p. 7377, 2024.