

## Information Sciences

Special Topic: Novel Optoelectronic Devices

**Strategies for training optical neural networks**Qipeng Yang<sup>1</sup>, Bowen Bai<sup>1,\*</sup>, Weiwei Hu<sup>1</sup> & Xingjun Wang<sup>1,2,3,4,\*</sup><sup>1</sup>State Key Laboratory of Advanced Optical Communications System and Networks, School of Electronics, Peking University, Beijing 100871, China;<sup>2</sup>Yangtze Delta Institute of Optoelectronics, Peking University, Nantong 226010, China;<sup>3</sup>Frontiers Science Center for Nano-optoelectronics, Peking University, Beijing 100871, China;<sup>4</sup>Peng Cheng Laboratory, Shenzhen 518055, China\*Corresponding authors (emails: [bowenbai@pku.edu.cn](mailto:bowenbai@pku.edu.cn) (Bowen Bai); [xjwang@pku.edu.cn](mailto:xjwang@pku.edu.cn) (Xingjun Wang))

Received 30 June 2022; Revised 11 October 2022; Accepted 28 October 2022; Published online 8 November 2022

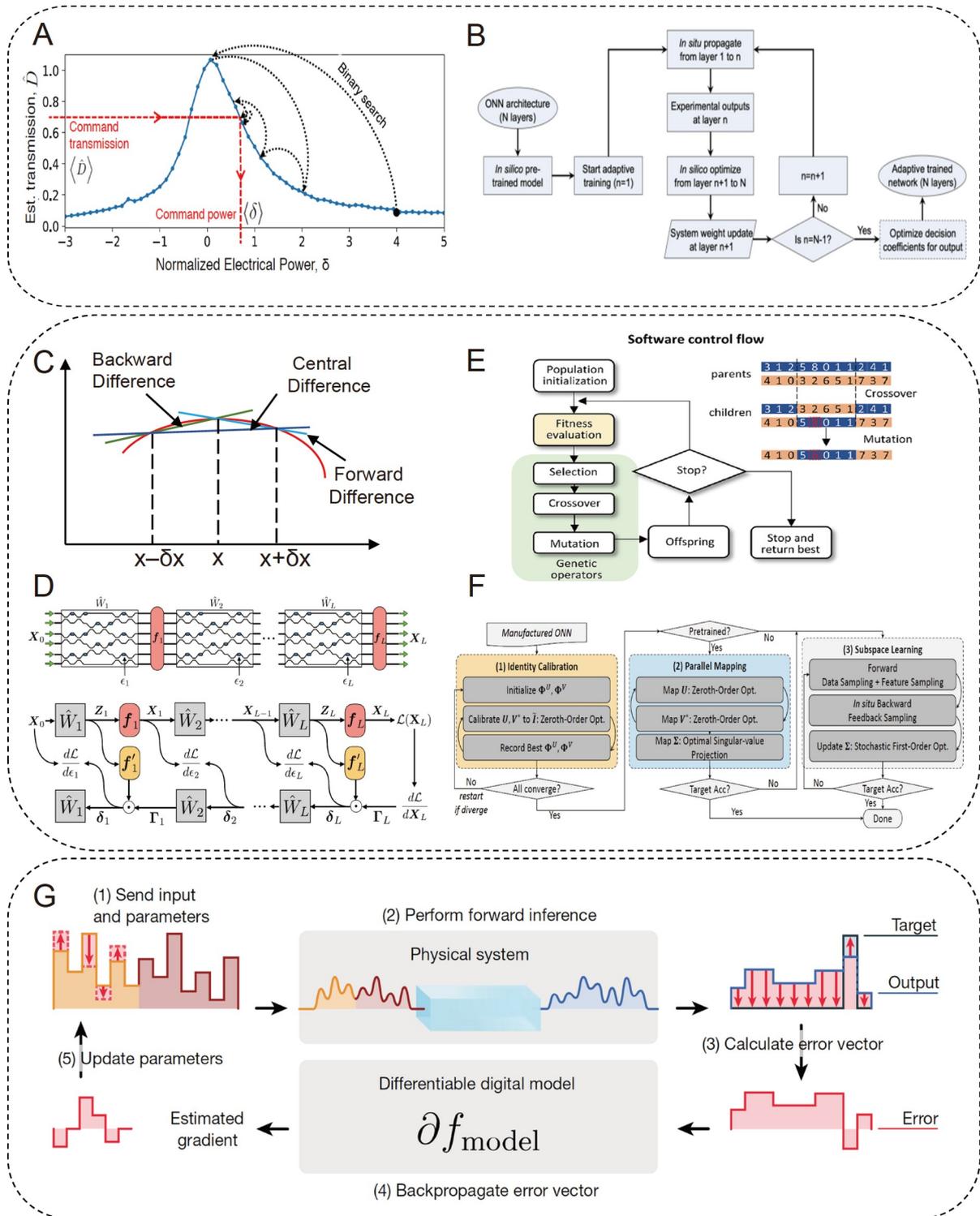
Deep learning has flourished in different areas in recent years, such as computer vision and natural language processing. However, with the end of Moore's law, these applications that rely heavily on computing power are facing bottlenecks. Optical neural networks (ONNs) [1] use light to perform calculations [2,3] featuring high speed and low energy consumption, and they are widely regarded as the next-generation application-specific integrated circuit (ASIC) [4,5] for artificial intelligence.

At present, ONNs have demonstrated many applications, such as vowel recognition and handwritten digit recognition. Compared with current mainstream deep learning applications, for example, end-to-end object detection [6], the tasks implemented on ONNs are very naive. One reason is that the number of neurons in current ONNs is much less than their electronic competitors. However, such limitation can be removed by novel architecture design and improvement of optoelectronic integration technology eventually. Another reason is that the strategies for training ONNs are still in a very early stage without a unified framework. Different from digital neural networks, the backpropagation method is not suitable for all ONNs because the gradient information is not easily available in some architectures [7]. To overcome this dilemma, three classes of training strategies are specifically designed. Compared to their conventional electrical counterparts, ONNs using these strategies have shown the comparable performance. The first class relies on fine tuning to map parameters from pre-trained models *in-silico*. Although these strategies can improve the inference performance, the ONNs fail to exploit the speed advantage due to the long training time caused by fabrication deviation and environmental fluctuation. Therefore, another class based on *in-situ* training is proposed. These strategies, directly training neural networks on the physical implementation, can adaptively find optimal parameters under system distortion and are more suitable for realistic application. However, restricted by approaches to obtain internal information from ONNs, *in-situ* training strategies (gradient-based and gradient-free) still suffer from a slow convergence rate for training large-scale ONNs. The third class proposes a hybrid *in silico-in situ* algorithm that utilizes backpropagation to the physical system. Although it shows potential to train physical neural networks (PNN), PNN can only be considered as an extra supplement

to existing systems that computing resources are mostly consumed in inference, since digital training models are used.

One of the representative studies for the first class is shown in Figure 1A, which utilizes lookup tables to map weights with different voltages [8]. With both lookup table and binary search algorithms for fine tuning, parameter deviations can be minimized. Because of the optical neurons' inconsistency, recharacterization is always required in different systems. Such a method is only suitable for simple demonstrations, and it is difficult to be utilized on large-scale ONNs. In ref. [9], an adaptive training method corrects pre-trained models layer by layer sequentially to compensate imperfections of the physical system (Figure 1B). The ONN achieves satisfied accuracies for high-speed image and video recognition and a computing performance superior to advanced electronic computing platforms. Although this training strategy makes the ONNs robust, it takes much time to train the neural networks with multiple iterations *in-silico*. Instead of mapping pre-trained parameters to ONNs, the second class directly train ONNs *in-situ*. As shown in Figure 1C, a gradient-based training method is proposed in ref. [7]. This method obtains gradient information by finite difference, but it is not suitable for large-scale ONNs because the training time grows exponentially with the number of parameters. Another gradient-based training method shown in Figure 1D directly uses the backpropagation algorithms (widely applied in neural networks) combined with the adjoint method in ref. [10]. The ONN can successfully learn the XOR gate. It can optimize parameters in parallel, however, internal information, such as field intensity and phase in each neuron, is still needed to calculate the gradient. Therefore, a portion of optical power is used for monitoring neurons, resulting in optical signal-to-noise ratio deterioration and reduction in calculation accuracy. Moreover, lossless assumption greatly limits the training accuracy. One similar method without such assumption [11] requires optical fields to be reversely injected into ONNs for implementing backpropagation and the internal complex optical field should also be measured at the same time. The ONNs achieve accuracy comparable to *in-silico* training with an electronic computer on the tasks of object classification and matrix-vector multiplication. To obtain optimal parameters without extra loss, gradient-free strategies, such as the genetic algorithm and zeroth order optimization algorithm, are applied. As shown in Figure 1E, a genetic algorithm, inspired by the evolution theory, is proposed to train ONNs in ref. [12]. Using the three genetic operators (crossover, selection, and mutation), optimal voltages applied to the phase shifters are selected. The viability of the ONN has been confirmed using the crossbar switch and the iris classification task. Zeroth-order optimization (Figure 1F), as another gradient-free strategy widely used in black-box optimization, is also proposed in ref. [13]. Instead of direct measurement, it estimates the gradient direction from sampling. Although gradient-free methods can optimize parameters with no extra loss, a high convergence rate is not always guaranteed. To combine the *in-situ* and *in-silico* training, one hybrid algorithm called physical aware training (PAT) is proposed in ref. [14] shown in Figure 1G. Since physical systems are hard to be analytically differentiated, the usual backpropagation cannot be directly applied. Instead, digital models such as deep neural networks are used to estimate gradients. Updated parameters along with data inputs are sent to physical systems to implement correct classification. Although the PAT provides potential to train physical neural networks, using digital models makes it only benefit inference. As we point out previously, current strategies have not fulfilled the requirement for massive and efficient ONNs training yet, and brand-new training strategies are still expected.

To bring ONNs into reality, several challenges must be overcome. The first one is a mechanism to quickly obtain internal information from ONNs without extra loss during training. Meanwhile, small intensity or



**Figure 1** Training strategies for ONNs. (A) and (B) Two different training strategies to map parameters from pre-trained models. (C)–(F) Four *in-situ* training strategies using finite difference, backpropagation with the adjoint method, genetic algorithm, and zeroth-order optimization, respectively. (G) Physical aware training strategies for physical neural networks.

phase variations should be accurately perceived. Gradient information should be easily acquired and then parallel gradient descent that accelerates training can be implemented, which makes rapid training possible. The second challenge is software-hardware co-design. Unlike digital neural networks, the training strategies for ONNs are not once for all, and they should be dedicatedly designed for different architectures. Since training ONNs cannot be accomplished without digital electronics, the communication time between electronic and optical devices is another limitation to restricting the training speed. Monolithic integration technology could be one possible approach to overcome this obstacle. Besides, reprogrammable photonic nonlinear activation functions [15] are highly required since different types of functions can greatly impact training efficiency in different tasks. It would be a critical step for blooming of ONNs and artificial intelligence once the massive and efficient training strategy was fulfilled.

### Data availability

The original data are available from corresponding authors upon reasonable request.

### Acknowledgements

The authors thank Bitao Shen, Ming Jin, and Yuansheng Tao for helpful suggestions on the manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (62105008 and 62235003) and the China Postdoctoral Science Foundation (2021T140004).

### Conflict of interest

The authors declare that they have no conflict of interest.

### References

- 1 Ashtiani F, Geers AJ, Aflatouni F. An on-chip photonic deep neural network for image classification. *Nature* 2022; **606**: 501–506.
- 2 Bai BW, Shu HW, Wang XJ, *et al.* Towards silicon photonic neural networks for artificial intelligence. *Sci China Inf Sci* 2020; **63**: 160403.
- 3 Zou WW, Ma BW, Xu SF, *et al.* Towards an intelligent photonic system. *Sci China Inf Sci* 2020; **63**: 160401.
- 4 Shu H, Chang L, Tao Y, *et al.* Microcomb-driven silicon photonic systems. *Nature* 2022; **605**: 457–463.
- 5 Li Y, An N, Lu Z, *et al.* Nonlinear co-generation of graphene plasmons for optoelectronic logic operations. *Nat Commun* 2022; **13**: 3138.
- 6 Carion N, Massa F, Synnaeve G, *et al.* End-to-end object detection with transformers. In: *Lecture Notes in Computer Science*. Cham: Springer, 2020. **12346**: 213–229.
- 7 Shen Y, Harris NC, Skirlo S, *et al.* Deep learning with coherent nanophotonic circuits. *Nat Photonics* 2017; **11**: 441–446.
- 8 Tait AN, Jayatilaka H, De Lima TF, *et al.* Feedback control for microring weight banks. *Opt Express* 2018; **26**: 26422.
- 9 Zhou T, Lin X, Wu J, *et al.* Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat Photonics* 2021; **15**: 367–373.
- 10 Hughes TW, Minkov M, Shi Y, *et al.* Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* 2018; **5**: 864.
- 11 Zhou T, Fang L, Yan T, *et al.* In situ optical backpropagation training of diffractive optical neural networks. *Photon Res*

- 2020; **8**: 940.
- 12 Zhang H, Thompson J, Gu M, *et al.* Efficient on-chip training of optical neural networks using genetic algorithm. *ACS Photonics* 2021; **8**: 1662–1672.
  - 13 Gu J, Zhu H, Feng C, *et al.* L2ight: Enabling on-chip learning for optical neural networks via efficient in-situ subspace optimization. In: *Proceedings of Advances in Neural Information Processing Systems*, 2021. **11**: 8649–8661.
  - 14 Wright LG, Onodera T, Stein MM, *et al.* Deep physical neural networks trained with backpropagation. *Nature* 2022; **601**: 549–555.
  - 15 Bandyopadhyay S, Sludds A, Krastanov S, *et al.* Single chip photonic deep neural network with accelerated training. 2022; arXiv: [2208.01623](https://arxiv.org/abs/2208.01623).