Parity-time Symmetric Optical Neural Networks

Haoqin Deng, Mercedes Khajavikhan
1. Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA
2. Department of Physics & Astronomy, Dornsife College of Letters, Arts, & Sciences, University of Southern California, Los Angeles, CA 90089, USA
*khajavik@usc.edu

Abstract: An optical neural architecture is proposed that utilizes parity-time symmetric couplers as its building blocks. Gain–loss contrasts across the array are adjusted as a means to train the network. © 2022 The Author(s)

OCIS codes: (200.0200) Optics in computing

In 2017, a team of researchers from MIT demonstrated a ground-breaking, fully integrated optical neural network (ONN) on a silicon chip [1] by cascading a number of Mach–Zehnder interferometers (MZIs). An arbitrary matrix can be effectively mapped onto this ONN hardware by computing the corresponding phases of each MZI. However, changing phases on chip is undesirable and can significantly overshadow the potential benefits of photonic accelerators [2]. In these arrangements, phase changing is typically accomplished by thermo-optical phase shifters, where a bias current is applied to change the refractive index of an optical waveguide through the thermo-optic effect [1,3]. However, since the thermo-optic coefficient of most optoelectronic materials is relatively small, translating it to a phase change requires a path length that is typically of the order of tens to hundreds of micrometers [3]. Given that for processing \( N \) bits of data, \( O(N^2) \) phase shifters are needed, such schemes can lead to prohibitively large structures as the size of the data increases. Moreover, the time it takes for the phase change to take effect is relatively long, of the order of tens of microseconds [3], which can limit the speed of on-chip training processes, where one needs to frequently vary phases to compute gradients. In this work, we show how parity-time symmetric couplers can be utilized to partially address some of these problems [4].

In a PT coupler, the energy exchange between the two waveguides obeys the following system of equations [4]:

\[
\begin{align*}
  i \frac{da}{dz} - i \frac{\theta}{2} a + \kappa b &= 0 \\
  i \frac{db}{dz} + i \frac{\theta}{2} b + \kappa a &= 0
\end{align*}
\]

where \( a \) and \( b \) represent the electric field in the two waveguides, \( \kappa \) is the coupling strength, \( z \) the propagation length, and \( g \) signifies the gain/loss contrast. The relationship between the input \((a_0 \text{ and } b_0)\) and output \((a \text{ and } b)\) ports (see Fig. 1b) can be derived in two different regimes of operation. Below the PT symmetry breaking point (where \( g/2\kappa \leq 1 \)), the coupling matrix is expressed by:

\[
\begin{pmatrix}
  a \\
  b
\end{pmatrix} = \frac{1}{\cos \theta} \begin{pmatrix}
  \cos(Z \cos \theta - \theta) & i \sin(Z \cos \theta) \\
  i \sin(Z \cos \theta) & \cos(Z \cos \theta + \theta)
\end{pmatrix} \begin{pmatrix}
  a_0 \\
  b_0
\end{pmatrix},
\]

where \( \theta = \sin^{-1}(g/2\kappa) \) and \( Z = \kappa z \). In this work we use PT- couplers exclusively in the PT-unbroken phase. In other words, the gain-loss contrast in the system is only minimally perturbed around zero values (here \( g/2\kappa < 0.2 \)). By adding appropriate constant phases to the input \((−\pi/2, −\pi)\) and output arms \((\pi/2, \pi)\), respectively, the transfer function can be modified to only act in real space:

\[
\begin{pmatrix}
  a \\
  b
\end{pmatrix} = \frac{1}{\cos \theta} \begin{pmatrix}
  \cos(Z \cos \theta - \theta) & -\sin(Z \cos \theta) \\
  -\sin(Z \cos \theta) & \cos(Z \cos \theta + \theta)
\end{pmatrix} \begin{pmatrix}
  a_0 \\
  b_0
\end{pmatrix},
\]

In our network, we also assume a constant \( \kappa \) and \( Z = \kappa z = 1 \).
Fig. 1 shows a schematic of a two-layer PT-symmetric optical neural network that is used in our simulations. In layer 1, $N_1$ pixels of the incoming data are encoded in light amplitude (provided by a series of laser sources/beams). After modulating the data on the carrier frequency, it travels in a triangular-shaped array containing $N_1(N_1 - 1)/2$ PT-symmetric couplers, followed by $N_2$ amplifiers/attenuators, another triangular-shaped array of PT-couplers containing $N_2(N_2 - 1)/2$ components, and finally $N_2$ nonlinear elements. Layer 1 is followed by layer 2, which is similar to the first layer in architecture but with different number of elements ($N_3$, and $N_3$ instead of $N_2$ and $N_2$) and ends in $N_3$ optical detectors. The output of the detectors is then sent to an electronic circuit to calculate the PT-coupler gain/loss parameters ($\theta$'s) in order to implement the gradient descent algorithm in the training cycles. In this example, $N_1, N_2, N_3$ are the sizes of input, hidden, and output layers, respectively.

The simulations are performed for digit recognition task on MNIST dataset. To accomplish this, the $28 \times 28$-pixel images are subsampled by a factor of 16 to be $7 \times 7$-pixel images for computing efficiency improvement. In our studies, we use the input layer of size $7 \times 7$ ($N_1 = 49$), the hidden layer of size $N_2 = 20$, and an output layer of $N_3 = 10$ dimensionality (corresponding to 10 digits). We use a sigmoid activation function for the hidden layer (this choice is regardless of the hardware used for the implementation of the nonlinear function), SoftMax activation function for the output layer, and cross-entropy as the loss function. The simulations are run with Python programs on an Intel i9-9900k CPU. We also assume that all parameters are randomly initialized. For on-chip training, we compute the numerical gradients of the designed parameters using finite difference method.

First, we model an MZI-based optical neural network in which phases of the MZIs serve as the parameters. The MZI mesh is arranged in the triangular fashion inspired by [1], which uses singular-value decomposition (SVD). Fig. 2a shows that this model achieves a peak training accuracy of 69% and a peak testing accuracy of 70.2%. Second, we replace MZIs with PT couplers. In this case the training parameters are gain/loss factors. We use the same topology of the mesh in the second and third simulations in order to allow a direct comparison to be made. Fig. 2b shows that a peak testing accuracy of 66.5% and a peak training accuracy of 67.2%.

![Fig. 2 (a) Training and testing accuracies of the MZI-based optical neural network. (b) Training and testing accuracies of the parity–time-symmetric optical neural network](image)

In conclusion, we demonstrated that PT-ONN architecture using gain–loss contrast as the training parameter can achieve on-chip training and testing accuracies comparable to those reported in ONNs composed of MZI devices with phase shifters. Since varying gain/loss coefficients can be more efficient than changing phases in terms of space, power consumption, and speed, our PT-ONN architecture can potentially require a smaller footprint and accelerate on-chip training at lower powers.

Acknowledgments. This work was supported by Office of Naval Research (N00014-19-1-2052, N00014-20-1-2522, N00014-20-1-2789); Air Force Office of Scientific Research (FA9550-20-1-0322, FA9550-21-1-0202); Army Research Office (W911NF-17-1-0481); United States-Israel Binational Science Foundation (2016381); National Science Foundation (CBET 1805200, ECCS 2000538, ECCS 2011171); Defense Advanced Research Projects Agency (D18AP000058).

References
1. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, “Deep learning with coherent nanophotonic circuits,” Nat. Photonics 11, 441–446 (2017).
2. J. Moughames, X. Porte, M. Thiel, G. Ulliac, L. Larger, M. Jacquot, M. Kadic, and D. Brunner, “Three-dimensional waveguide interconnects for scalable integration of photonic neural networks,” Optica 7, 640–646 (2020).
3. N. C. Harris, Y. Ma, J. Mower, T. Baehr-Jones, D. Englund, M. Hochberg, and C. Galland, “Efficient, compact and low loss thermo-optic phase shifter in silicon,” Opt. Express 22, 10487–10493 (2014).
4. H. Deng and M. Khajavikhan, “Parity–time symmetric optical neural networks,” Optica 8, 1328-1333 (2021)