MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Universal Photonic Neural Networks with Quantum-Free Data Reuploading

Kojima, Keisuke; Koike-Akino, Toshiaki

TR2024-074 June 18, 2024

Abstract

The data reuploading trick was originally proposed for quantum computing to achieve the universal approximation property. In this paper, we introduce data reuploading to realize universal non-quantum photonic computing with practical photonic integrated circuits (PICs). We aim to comprehensively discuss the various advantages and implementation considerations of this approach. Our framework can eliminate the need of quantum squeezed lights, photon counters, and nonlinear photonics, which have been essential for enabling photonic neural networks in conventional configurations. Additionally, we explore ways to minimize the optical components by combining multiple functionalities into a single phase shifter, showing competitive performance when compared to using the same number of phase shifters, all without employing any nonlinear photonic devices. Considering these char- acteristics, our investigation into the use of PICs for data reuploading presents a novel architectural approach to realize photonic neural networks. This approach embodies unique features that distinctly set it apart from traditional photonic neural networks.

SPIE Photonics for Quantum 2024

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Universal Photonic Neural Networks with Quantum-Free Data Reuploading

Keisuke Kojima[a] and Toshiaki Koike-Akino[b]

[a]Boston Quantum Photonics LLC, 588 Boston Post Rd # 315, Weston, MA, USA
[b]Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA, USA

ABSTRACT

The data reuploading trick was originally proposed for quantum computing to achieve the universal approximation property. In this paper, we introduce data reuploading to realize universal non-quantum photonic computing with practical photonic integrated circuits (PICs). We aim to comprehensively discuss the various advantages and implementation considerations of this approach. Our framework can eliminate the need of quantum squeezed lights, photon counters, and nonlinear photonics, which have been essential for enabling photonic neural networks in conventional configurations. Additionally, we explore ways to minimize the optical components by combining multiple functionalities into a single phase shifter, showing competitive performance when compared to using the same number of phase shifters, all without employing any nonlinear photonic devices. Considering these characteristics, our investigation into the use of PICs for data reuploading presents a novel architectural approach to realize photonic neural networks. This approach embodies unique features that distinctly set it apart from traditional photonic neural networks.

Keywords: optical neural networks, optical computing, optical quantum computing, data reuploading, photonic integrated circuits

1. INTRODUCTION

Programmable PICs offer the capability to accelerate neural networks. There are various types of optical neural network (ONN) implementations, including quantum vs. non-quantum counterparts. Quantum optical neural networks (QONNs) have two major methods: discrete-variable (DV-QONN)[1] or continuous-variable (CV-QONN).[2] As a practical challenge, most QONNs require cryogenic-controlled photon counters. Although non-quantum ONN[3] does not need such cryogenics, nonlinear photonic devices are typically required to be universal. While some papers consider linear-photonic neural networks,[4] its unitarity nature without a control of nonlinear attenuation often restricts the inference performance substantially.

Data reuploading[5] shown in Fig. 1 was originally proposed for quantum computing to achieve the universal approximation property (UAP). We discuss its advantages and PIC implementation options. Unlike classical neural networks, as shown in Fig 2, the input data $\vec{x}$ are embedded as Pauli rotation angles repeatedly every layer, and nonlinearity is imposed without using optical nonlinear components due to the repeated embedding. Its validity is experimentally demonstrated using quantum photonic devices.[6] However, its usefulness in non-cryogenic photonics is not well understood in literature. In this paper, we introduce a new non-quantum approach of photonic data reuploading, which has no requirement of nonlinear optical devices, photon counters, and squeezed lights. We investigate several ways to downsize the optical commponents and to realize multi-label classification with limited photonic ports. Our demonstration through different machine learning tasks shows a viable potential of our framework as a practical alternative of ONN methods.

Further author information: (Send correspondence to KK and TK)
KK : E-mail: kkojima@bostonqp.com, Telephone: 1 505 123 1234
TK : E-mail: koike@merl.com, Telephone: 1 617 621 7543

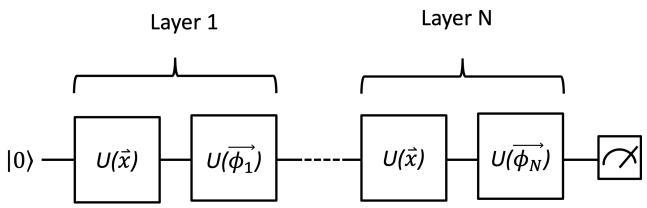

Figure 1: The original data reuploading method using a single qubit, where the input $\vec{x}$ are repeatedly embedded every layer.[5]

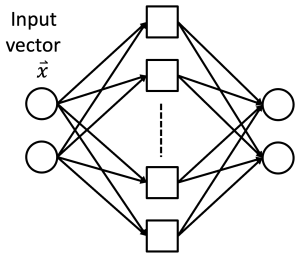

Figure 2: Schematic of a classical neural network.

2. PIC IMPLEMENTATION

An arbitrary single qubit operation can be decomposed into Pauli Z/Y/Z rotations:[7]

$$Rot(\phi,\theta,\omega) = RZ(\omega)RY(\theta)RZ(\phi) = \begin{bmatrix} e^{-i(\phi+\omega)/2}\cos(\theta/2) & -e^{i(\phi-\omega)/2}\sin(\theta/2) \\ e^{-i(\phi-\omega)/2}\sin(\theta/2) & e^{i(\phi+\omega)/2}\cos(\theta/2) \end{bmatrix}. \quad (1)$$

Fig. 3 shows a case of three-dimensional input data $\vec{x} = [x_1, x_2, x_3]$ and a three-layer network to illustrate how the operations are conducted. The network has trainable weight parameters $\vec{a} = [a_1, a_2, a_3]$, $\vec{b} = [b_1, b_2, b_3]$ and $\vec{c} = [c_1, c_2, c_3]$ for the first, second, and third layers respectively. Similar to single-qubit operations in the quantum case, we have two modes and continuous wave (CW) light fed into one mode. The weight variables are trained such that the measurement of the output lights at two modes provides a task prediction, e.g., as a binary class probability.

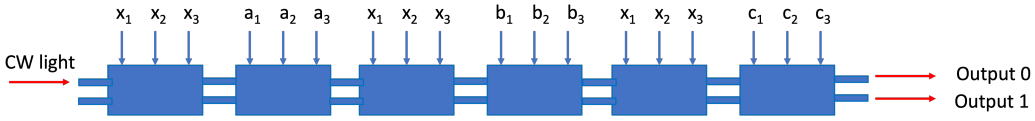

Figure 3: Example of three-dimensional input data and three-layer data reuploading network.

Each of the block in Fig. 3 is a rotation block expressed as in Eq. (1), and its most direct PIC realization is shown in Fig. 4 and we call it *Rot*. We use three pairs of differential phase shifters (PSs) to represent the rotation in Pauli Z, Y, and Z axis.

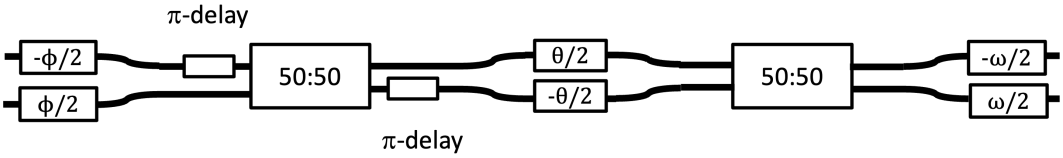

Figure 4: Schematic of a close translation to a PIC representing a rotation (*Rot*).

3. BINARY CLASSIFICATION

We first apply the PIC circuit for binary classification problems: three-dimensional sphere and six-dimensional hyper-spheres benchmark.[5] For the measurement at the last layer, we use logit: $\log(p_0) - \log(p_1)$ for the measured power p_0 and p_1 at output mode 0 and 1, respectively. For the three-dimensional case, we use one block of *Rot* for the input data and another block for the three weight parameters to form a layer. In the case of six-dimensional

case, we use two blocks of *Rot* to express six input data, and two blocks of *Rot* for six weight parameters to form a layer. We use 4096 training data and 1024 test data. The AdamW optimizer of the PyTorch library is used for training with a learning rate of 0.03. Ten runs with different random number seeds are repeated, and the maximum test accuracy is reported. Fig. 5 shows the classification accuracy as a function of the number of layers, showing good performance.

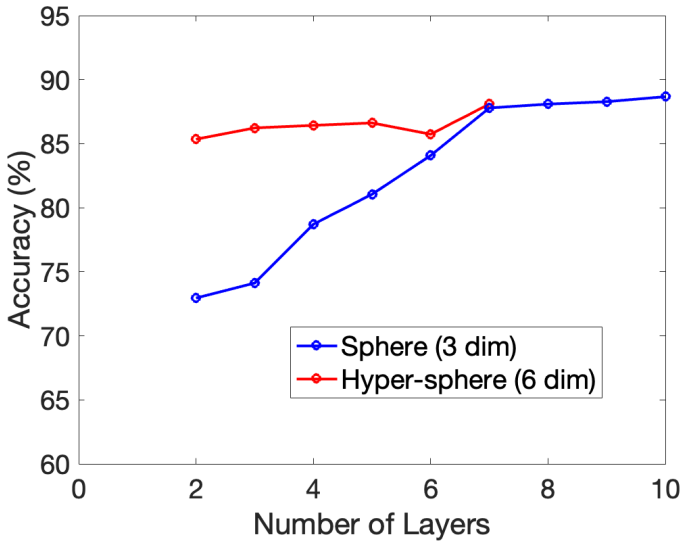

Figure 5: Accuracy of the photonic data reuploading method as a function of the number of layers for the three-dimensional sphere and six-dimensional hyper-sphere problems.

4. NON-BINARY CLASSIFICATION

We also consider a non-binary classification task in a two-dimensional wavy lines problem[5] with $N = 4$ classes. With 2-mode optical data reuploading circuits, it is not straightforward to classify multiple classes. There are multiple options including one-vs-one (OvO), one-vs-rest (OvR), and output-coding (OC) strategies[8] to use binary classifiers for multi-label classifications. OvO, OvR, and OC require $N(N-1)$, N, and ceil$[\log_2(N)]$ binary classifiers for N-class predictions, respectively. We consider OvR-based stacking, where N binary-classifing data reuploading circuits are used in parallel to obtain N-class logits. It requires N-times more weight parameters. To reduce the total number of parameters, we introduce a new strategy called *implicit classification*, where the trainable parameters are shared across all N binary classifiers but with extra parameters for positional embedding used in each binary classifier. For implicit classification method, the class information is added at each layer as shown in Fig. 7. In this case, we can share the weights, but the circuit length of the stack is longer with an extra class information.

Fig. 8 shows the accuracy of the stacking method and the implicit classification method as a function of the number of layers for the wavy lines problem. Fig. 9 shows the accuracy of these methods as a function of the number of the rotation blocks (*Rot*). These figures indicate that the stacking method achieves higher accuracy for the same number of layers. However, both methods perform similarly when considering the number of blocks. This is because the implicit classification method uses photonic circuit resources efficiently. Fig. 10 shows the training data of the two-dimensional wavy line problem with four classes of different colors, and Fig. 11 shows the best test data with four stacks and seven layers demonstrating 98.05 % accuracy.

5. PIC VARIANTS

To reduce the number of PSs, we consider three variants of circuit simplifications below. We first remove the two π-delays from the first 50:50 directional coupler as shown in Fig. 12, denoted as *Rot-PIC*. This is equivalent to a

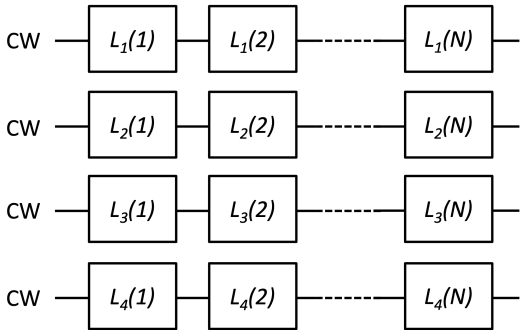

Figure 6: Example of stacking four lanes each with N layers.

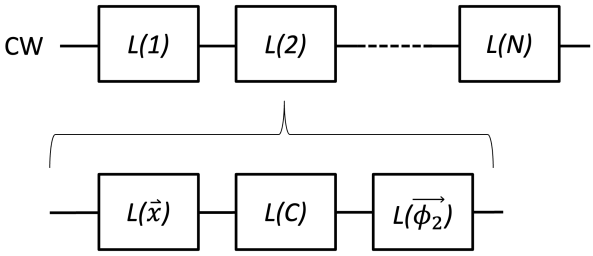

Figure 7: Schematic of the implicit classification method where the class information C is embedded and re-written at each training and inference instances.

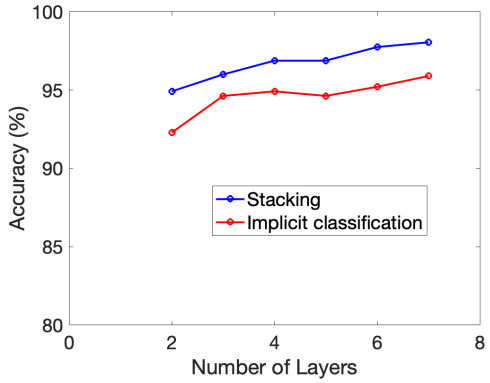

Figure 8: Accuracy of the stacking and the implicit classification methods as a function of the number of layers for the two-dimensional four-class wavy line problem.

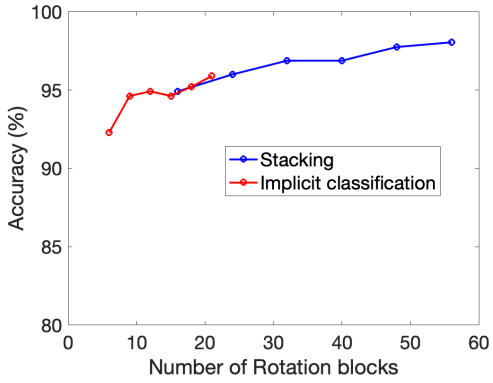

Figure 9: Accuracy of the stacking and the implicit classification methods as a function of the number of *Rot* blocks for the two-dimensional four-class wavy line problem.

2×2 universal unitary optical processor used by Macho-Ortez et al.,[9] and also equivalent to $Rot(\phi-\pi,\theta+\pi,\omega)$. Therefore, this fixed phase difference should be absorbed during the training process.

When multiple blocks are cascaded, the last differential PS pair of each block is directly connected to the first differential PS pair of the next block as shown in Fig. 13. Therefore, the PS pair can be combined, as long as electrical summation can be made. We call it RZY.

Further, as shown in Fig. 14, the differential PS can be simplified as a single-ended PS which has twice the amount of phase shift, and this can be called PZY-PS.

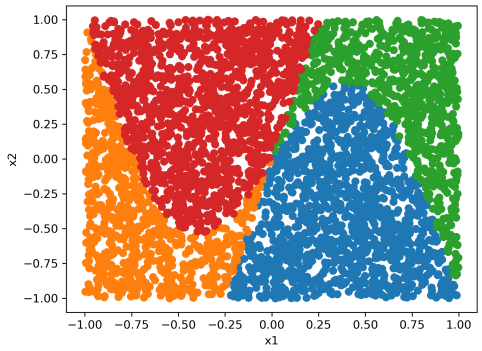

Figure 10: Plot of the training data for the two-dimensional four-class wavy line problem. Each color represents different classes.

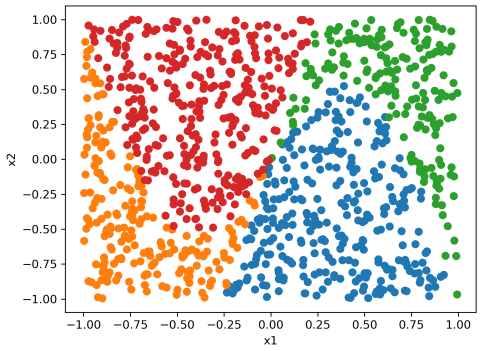

Figure 11: Plot of the test data for the two-dimensional four-class wavy line problem, when the accuracy of 98.05% was achieved.

Table 1 shows a comparison of the performance for the three cases, i.e., two-dimensional circle, two-dimensional wavy lines using implicit classification, and six-dimensional hypersphere problems. The performance among the four different configurations show comparable results despite the circuit complexity. There might be small performance degradation with *RZY-PS*, but further detailed study will be needed to conclude.

Note that reducing the number of PSs is typically associated with PSs requiring larger amount of phase shift. Therefore, the choice of circuitry needs to be carefully examined by the overall chip footprint and power consumption.

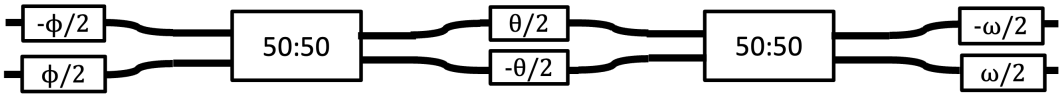

Figure 12: Schematic of a PIC representing a rotation without π-delays: *Rot-PIC*.

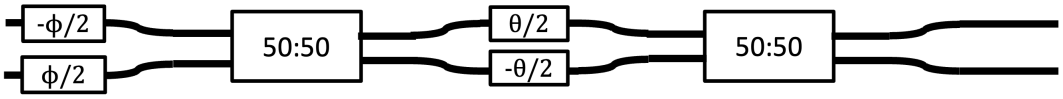

Figure 13: Schematic of a simplified circuit where the PSs in the neighboring sections are combined: *RZY*.

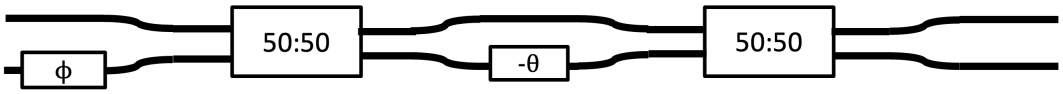

Figure 14: Schematic of a PIC representing a rotation by combining two differential PSs into one: *RZY-PS*.

6. PERFORMANCE COMPARISON WITH CLASSICAL NEURAL NETWORK

Data reuploading demonstrates comparable accuracy without nonlinear optical components, to classical NNs when compared with the same number of phase shifters as shown in Fig. 15 for the case of six-dimensional hypersphere problem. For the data reuploading, we assume the *RZY-PS* configuration. As we discussed earlier, if more conservative circuit configuration is desirable, then the number of PSs needs to be adjusted. For the

Table 1: Performance of different methods across various dimensions and configurations.

	Circle	Wavy lines	Hypersphere	# of PS/layer
dimension	2	2	6	
# of layers	10	5	5	
# of stacks	1	1	1	
Rot	99.31	94.63	85.94	12
Rot-PIC	98.93	95.70	85.94	12
RZY	97.85	95.80	84.28	8
RZY-PS	97.56	94.14	84.77	4

classical NNs, we assume one hidden layer and rectified linear unit (ReLU) activation as in Perez et al.,[5] simulated using the Scikit-Learn package and Adam optimizer. For the number of PSs, we assume Clements photonic unitary operation configuration.[10] Note that the classical NN does not directly use the Clements configuration for the accuracy calculation, and that this assumes perfect ReLU nonlinear function which is very difficult to achieve in a reliable manner for ONN realization. Despite these caveats, it is notable that data reuploading achieves comparable accuracy with classical NNs, without any optical nonlinear device.

In fact, it is theoretically proven that data reuploading achieves UAP.[5] Hence, it is expected that the accuracy can be further improved with more layers, and further study will follow.

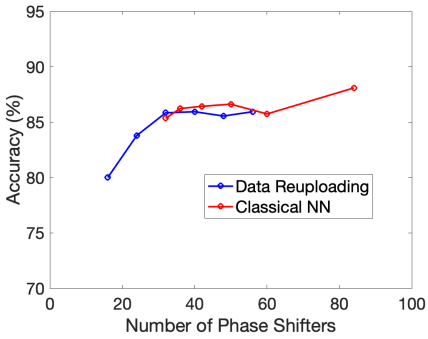

Figure 15: Comparison of accuracy vs. number of phase shifters for the hypersphere (6 dimensions) problem.

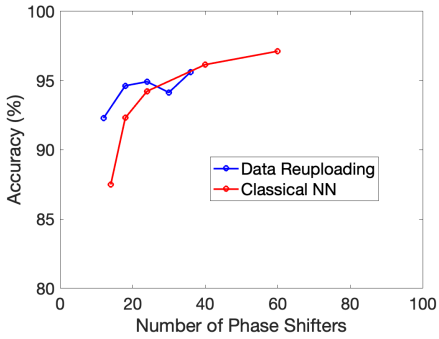

Figure 16: Comparison of accuracy vs. number of phase shifters for the wavy lines problem (2 dimensions, 4 classes) problem.

7. PHASE SHIFTER DISCUSSION

One potential drawback in implementing photonic data reuploading is the speed of the phase shifter. For many photonic neural networks, phase shifters are implemented using thermal effects to adjust the weights, which is acceptable during the inference stage as fast reconfiguration is not needed. However, data reuploading requires faster phase shifters (PSs) since data needs to be applied at every layer in the PIC. While parity-time symmetry can provide decent acceleration,[11] it is still not easy to achieve modulation speed in the order of MHz. Alternatively, using ring modulators for phase modulation can be effective. Despite the reported 0.97 dB intensity modulation experimentally demonstrated using silicon ring modulators,[12] this might be acceptable in differential PSs as both arms will experience the same intensity change if biased properly. Other material systems, like InP, GaAs, and thin-film lithium niobate, where electro-optic phase shifters are available, could also be solutions.

8. CONCLUSION

Data reuploading is evaluated in the context of non-quantum photonic neural networks from various aspects. Our method does not require nonlinear photonic devices, photon counters, or squeezed light sources. We have demonstrated its advantages in different PIC configurations, binary/non-binary classification strategies We plan to experimentally validate the performance of the proposed approach.

REFERENCES

[1] Steinbrecher, G. R., Olson, J. P., Englund, D., and Carolan, J., "Quantum optical neural networks," *npj Quantum Information* **5**(1), 60 (2019).

[2] Killoran, N., Bromley, T. R., Arrazola, J. M., Schuld, M., Quesada, N., and Lloyd, S., "Continuous-variable quantum neural networks," *Physical Review Research* **1**(3), 033063 (2019).

[3] Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D., et al., "Deep learning with coherent nanophotonic circuits," *Nature photonics* **11**(7), 441–446 (2017).

[4] Lu, K. and Guo, X., "Efficient training of unitary optical neural networks," *Optics Express* **31**(24), 39616–39623 (2023).

[5] Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., and Latorre, J. I., "Data re-uploading for a universal quantum classifier," *Quantum* **4**, 226 (Feb. 2020).

[6] Ono, T., Roga, W., Wakui, K., Fujiwara, M., Miki, S., Terai, H., and Takeoka, M., "Demonstration of a bosonic quantum classifier with data reuploading," *Physical review letters* **131**(1), 013601 (2023).

[7] https://docs.pennylane.ai/en/stable/code/api/pennylane.Rot.html (2023).

[8] Bishop, C. M., "Pattern recognition and machine learning," *Springer google schola* **2**, 1122–1128 (2006).

[9] Macho-Ortiz, A., Pérez-López, D., and Capmany, J., "Optical implementation of 2 × 2 universal unitary matrix transformations," *Laser & Photonics Reviews* **15**(7), 2000473 (2021).

[10] Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S., and Walmsley, I. A., "Optimal design for universal multiport interferometers," *Optica* **3**, 1460–1465 (Dec 2016).

[11] Chang, C., Li, T., Wu, Y., Zhou, P., and Zou, Y., "Fast-response, energy-efficient thermo-optic silicon phase shifter based on non-hermitian engineering," in [*Optical Fiber Communication Conference (OFC) 2022*], *Optical Fiber Communication Conference (OFC) 2022* , M3E.5, Optica Publishing Group (2022).

[12] Alam, A. S., Mikkelsen, J. C., Poon, J. K. S., and Aitchison, J. S., "Silicon ring modulator based optical phase shifter," in [*CLEO 2023*], *CLEO 2023* , SF2K.7, Optica Publishing Group (2023).