Enhancement of Data Reuploading for Photonic Neural Computing without Nonliear Optical Components

Koike-Akino, Toshiaki; Kojima, Keisuke; Taguchi, Mari

TR2025-013 January 28, 2025

Abstract

The data reuploading trick, originally proposed for universal quantum computing, enables the universal approximation property. We recently extended this concept to achieve universal non-quantum photonic computing using practical photonic integrated circuits (PICs) composed solely of 50:50 beam splitters and phase shifters, eliminating the need for nonlinear photonic devices. In this approach, input data are repeatedly embedded as rotation angles. In this presentation, we explore strategies to enhance the performance of this method by increasing the number of layers and optical channels (lanes) in various configurations. For a classical two-dimensional, four- class classification problem of wavy lines, we evaluated a two-mode class-embedding circuit with output, a four- mode configuration, and four stacking methods of two-mode circuits with average pooling, alongside a baseline configuration using projection in a complex domain. The first three configurations demonstrated excellent accuracy. Additionally, we investigated the effect of shifting the order of input data layer by layer, which significantly improved performance in certain applications. These findings highlight a novel architectural approach to photonic neural networks enabled by data reuploading in PICs. This approach offers unique features that set it apart from traditional photonic neural network architectures, providing a promising direction for future advancements in photonic computing

SPIE Photonics West 2025

Enhancement of Data Reuploading for Photonic Neural Computing without Nonliear Optical Components

Toshiaki Koike-Akino[a], Keisuke Kojima[b], and Mari Taguchi[c]

[a]Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA, USA
[b]Boston Quantum Photonics LLC, 588 Boston Post Rd # 315, Weston, MA, USA
[c]Information Technology R&D Center, Mitsubishi Electric Corporation, 5-1-1 Ofuna, Kamakura, Kanagawa 247-8501, Japan

ABSTRACT

The data reuploading trick, originally proposed for universal quantum computing, enables the universal approximation property. We recently extended this concept to achieve universal non-quantum photonic computing using practical photonic integrated circuits (PICs) composed solely of 50:50 beam splitters and phase shifters, eliminating the need for nonlinear photonic devices. In this approach, input data are repeatedly embedded as rotation angles.

In this presentation, we explore strategies to enhance the performance of this method by increasing the number of layers and optical channels (lanes) in various configurations. For a classical two-dimensional, four-class classification problem of wavy lines, we evaluated a two-mode class-embedding circuit with output, a four-mode configuration, and four stacking methods of two-mode circuits with average pooling, alongside a baseline configuration using projection in a complex domain. The first three configurations demonstrated excellent accuracy. Additionally, we investigated the effect of shifting the order of input data layer by layer, which significantly improved performance in certain applications.

These findings highlight a novel architectural approach to photonic neural networks enabled by data reuploading in PICs. This approach offers unique features that set it apart from traditional photonic neural network architectures, providing a promising direction for future advancements in photonic computing.

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

1. INTRODUCTION

Optical neural computing holds the potential to significantly reduce energy consumption compared to conventional digital neural computing. Various proposals and demonstrations of optical neural network (ONN) implementations have emerged, encompassing both quantum and non-quantum approaches.

Most quantum optical neural networks (QONNs)[1,2] require cryogenically controlled photon counters, limiting their scalability. On the other hand, non-quantum ONNs[3] eliminate the need for cryogenics but often rely on nonlinear photonic devices to achieve universality. While linear photonic neural networks[4] have been proposed as an alternative, their inherent unitarity and lack of nonlinear attenuation control often restrict their inference performance.

Data reuploading,[5] originally proposed for quantum computing, achieves the universal approximation property (UAP) by embedding input data $\vec{x}$ as Pauli rotation angles repeatedly across layers. Nonlinearity is effectively introduced without requiring optical nonlinear components, leveraging repeated embeddings. The validity of this approach has been experimentally demonstrated using quantum photonic devices.[6]

Further author information: (Send correspondence to TK, KK, and MT)
TK : E-mail: koike@merl.com, Telephone: 1 617 621 7543
KK : E-mail: kkojima@bostonqp.com, Telephone: 1 781 803 7636
MT : E-mail: Taguchi.Mari@bp.MitsubishiElectric.co.jp, Telephone: 81 50 3515 8201

We recently extended the concept of data reuploading to non-quantum applications, focusing on linear photonic integrated circuit (PIC) implementations[7] as shown in Fig.1, wherein the continuous wave (CW) light input is modulated by the input signals (x_1, x_2, ...) repeatedly, sandwiched by the trainable weight parameters. Each block consists of phase shifters (PSs) and $50:50$ directional couplers.

In this paper, we further explore its potential for multi-label classification applications, with particular attention to enhancing input data embedding strategies. We propose methods to downsize optical components and enable multi-label classification with a limited number of photonic ports. Our demonstrations through simulations across diverse machine learning tasks showcase the viability of this framework as a practical alternative to existing ONN methods.

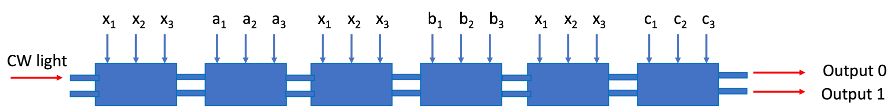


Figure 1: Schematic of a data reuploading block diagram consisting of three layers and three-dimensional input data.

2. NON-BINARY CLASSIFICATION

We first consider a non-binary classification task in a two-dimensional wavy lines problem[5] with $N = 4$ classes. Several strategies can adapt binary classifiers for multi-label classification, including one-vs-one (OvO), one-vs-rest (OvR), and output-coding (OC) approaches.[8] OvO, OvR, and OC require $N(N-1)$, N, and ceil$[\log_2(N)]$ binary classifiers for N-class predictions, respectively.

We first apply the basic data reuploading method with a complex output (Fig. 2). In this photonic implementation, we assume the use of a coherent receiver to detect the complex output. The four target points correspond to distinct classes in the complex plane. The four target points correspond to distinct classes in the complex plane. Next, we consider OvR-based stacking, where N binary-classification data reuploading circuits operate in parallel to generate N-class logits. This approach, however, requires N-times more weight parameters, compared to a single binary classifier, increasing the complexity of the system.

To address the circuit complexity issue, we recently proposed a novel strategy called *implicit classification*.[7] In this method, trainable parameters are shared across all N binary classifiers, reducing the total number of parameters. Additional parameters are introduced for positional embeddings, which encode class information uniquely for each binary classifier. As shown in Fig. 4, class information is added at each layer, enabling weight sharing while extending the circuit length slightly to incorporate the additional class embeddings. This approach strikes a balance between parameter efficiency and classification accuracy, providing a scalable solution for multi-class problems in photonic implementations.

Here, we also introduce another method called multi-mode method. In the case of $N = 4$, we have four input ports and four output ports as shown in Fig. 5. Each block contains multiple Clements-type unitary processors[9] to encode input and weight parameters. We use the basis for the input of $(1, 0, 1, 0)$, which gives better results than $(1, 0, 0, 0)$ or $(1, 0, 0, 1)$ in our configuration.

3. PERFORMANCE EVALUATION

Using four configurations for the PIC building block (summarized in the Appendix), we compared these above four methods, as shown in Fig. 6.

For simulations, we use 4096 training data and 1024 test data, and a batch size of 256. The AdamW optimizer of the PyTorch library is used for training with a learning rate of 0.06. Ten runs with different random number seeds are repeated, and the median and standard deviations for the training and test accuracy are plotted.

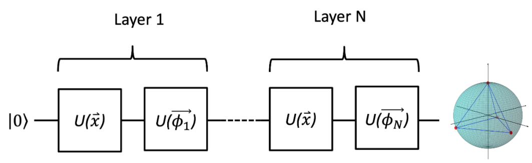


Figure 2: Example of complex classification method where the output is expressed in four target points on the Block sphere.

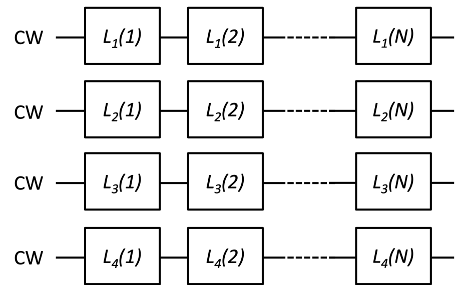


Figure 3: Example of stacking four lanes each with N layers.

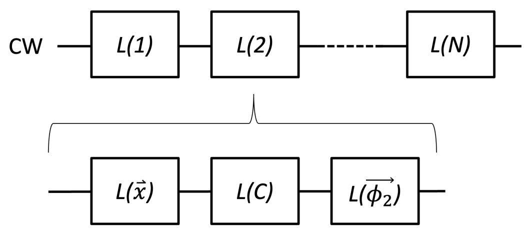


Figure 4: Schematic of the implicit classification method where the class information C is embedded and re-written at each training and inference instances.

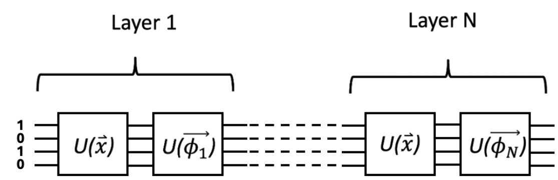


Figure 5: Schematic of four mode method, wherein each output corresponds to each class. Each block contains multiple Clements-type unitary processors.

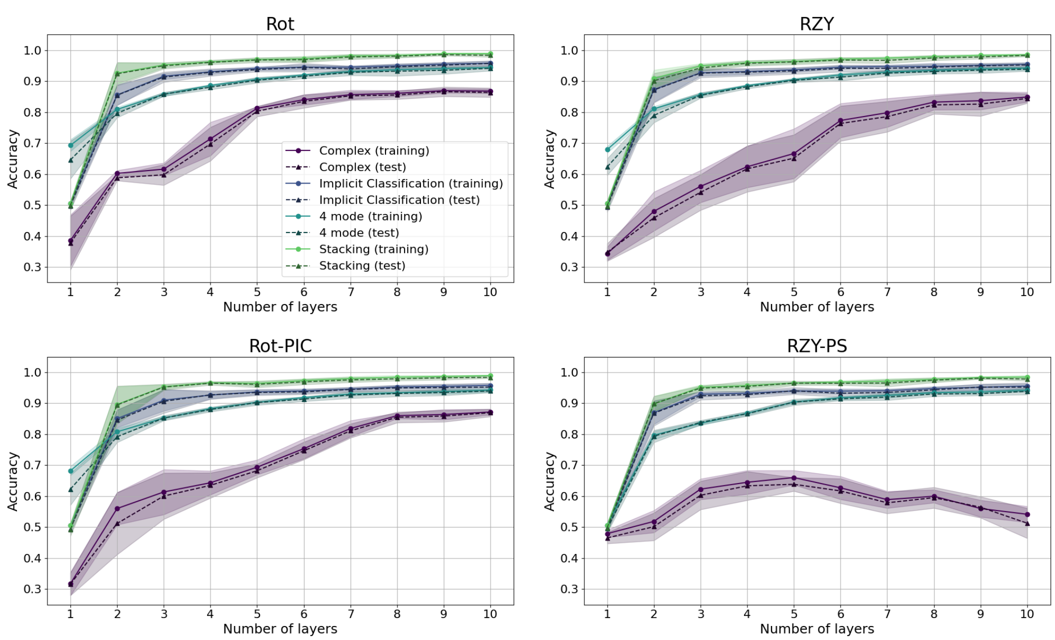


Figure 6: Accuracy of the four methods as a function of the number of layers for the two-dimensional four-class wavy line problem. Rot, Rot-PIC, RZY, and RZY-PS denote the circuit configurations.

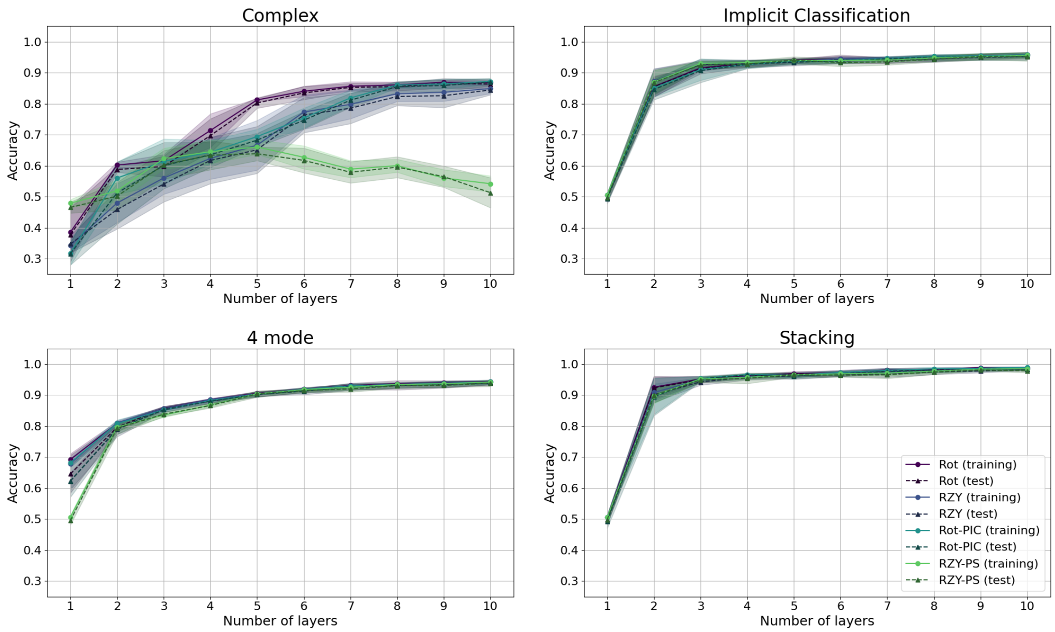


Figure 7: Accuracy of the four methods as a function of the number of layers for the two-dimensional four-class wavy line problem.

This figure demonstrates that overall accuracy tends to improve rapidly with an increase in the number of layers. While the complex method, despite its minimal resource requirements per layer (measured by the number of PSs), does not perform well, the resource-intensive stacking method delivers excellent performance even with just two layers. The resource-efficient implicit classification method also performs remarkably well. In contrast, the four-mode method, despite its higher resource consumption, does not achieve comparable performance.

The same data are now grouped by method and shown in Fig. 7. For the complex method, it is evident that RZY-PS performs poorly, as it lacks the flexibility to fully represent the Bloch sphere. Both Rot-PIC and RZY also exhibit suboptimal performance.

For the other three methods, except in cases with a small number of layers, no significant differences are observed. This suggests that simplifying the circuit using RZY or RZY-PS could provide more efficient circuit designs without compromising performance.

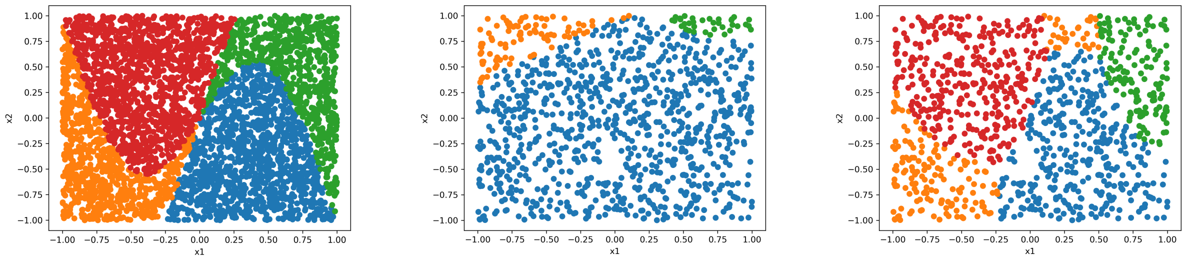
The classification results of the complex classification method are shown in Fig. 8. Fig. 8 (a) displays the training data, while Fig. 8 (d) shows the test data. With one layer, the classification result Fig. 8 (b) is very poor, as indicated by the error plotted in blue in Fig. 8 (e). With 10 layers, the classification results are improved as seen in Fig. 8 (c), and the error performance is also improved as shown in (f).

Fig. 9 presents the classification results for the four methods using two layers. The results clearly show that the stacking method performs the best, while the complex method performs the worst.

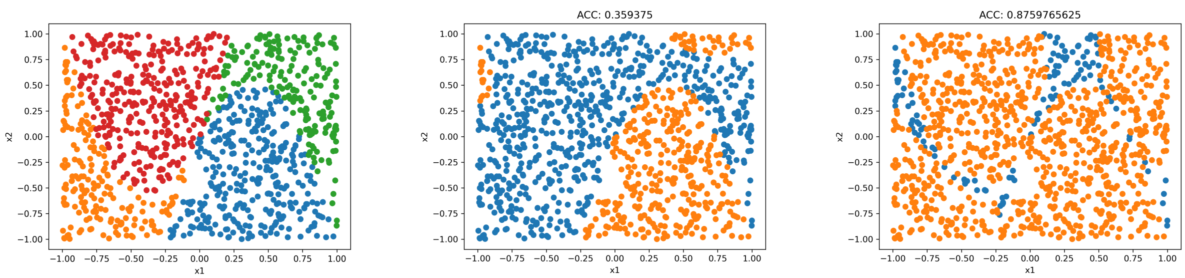
4. EFFECT OF THE DIRECTONAL COUPLER VARIATIONS

It is known that multiport interferometers are sensitive to beamsplitter imperfections,[10] including Clements configurations.[9] The splitting ratio of directional couplers is expressed as $(50 \pm \sigma_{BS})$ %, where σ_{BS} represents the standard deviation of the splitting ratio. Typical silicon photonics fabrication processes can achieve a standard deviation of 2 % or better[10]

Fig. 10 illustrates the classification accuracy for the two-dimensional wavy lines problem using implicit classification, four-mode, and four-stack configurations, each evaluated with five and ten layers under standard

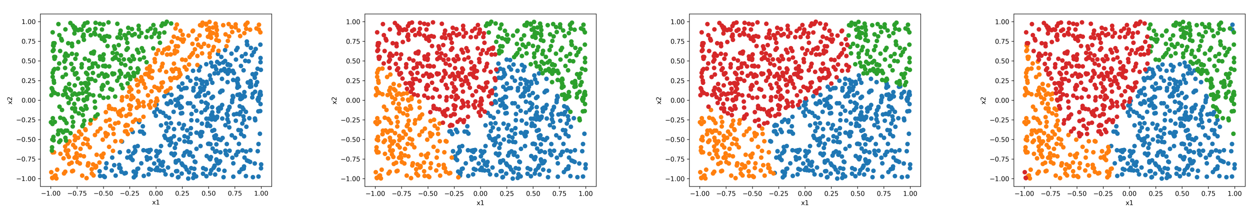


(a) Traing data (b) 1 layer classification (c) 10 layer classification

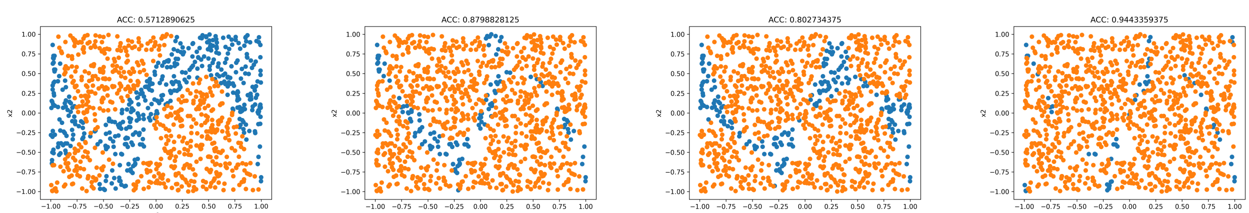


(d) Test data (e) 1 layer classification error (f) 10 layer classification error

Figure 8: Classification results by complex classification method. (b) and (c) show the classification results as indicated by the color, and (e) and (f) show the classification results highlighted by the blue dots.



(a) Complex classification (b) Implicit classification (c) 4 mode classification (d) Stacking classification



(e) Complex error (f) Implicit error (g) 4 mode error (h) Stacking error

Figure 9: Classification results by four methods with 2 layers. (a)-(d) show the classification results as indicated by the color, and (e)-(h) show the classification results highlighted by the blue dots.

deviations of 0, 2 %, and 4 %. It is assumed that all imperfections are compensated for during the online training process. The accuracy tends to decrease gradually as the standard deviation increases. However, in the case of 10 layers, the accuracy remains high when the initial accuracy is already high. Nevertheless, it is desirable to consider the fabriation-tolerant directional coupler designs when implemented in actual PICs.[11]

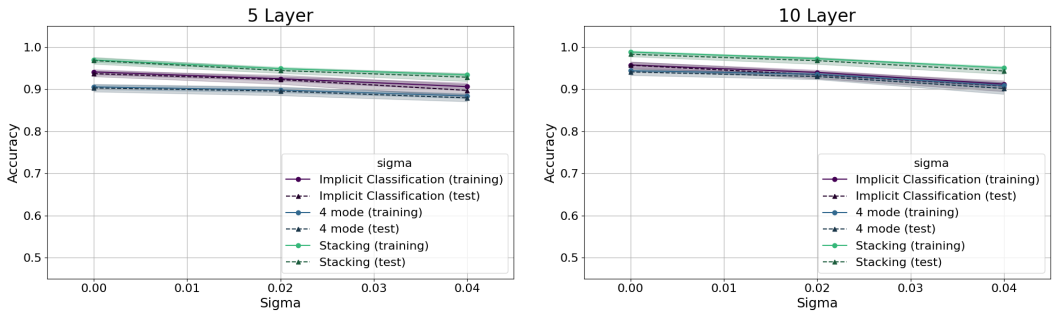


Figure 10: Accuracy of the three methods as a function of the number of layers for the two-dimensional four-class wavy line problem with directional coupler standard deviations of 0, 2 %, and 4 %.

5. PERFORMANCE COMPARISON WITH RESPECT TO THE NUMBER OF PHASE SHIFTERS

Data reuploading demonstrates high accuracy without requiring nonlinear optical components, while its resource demands vary depending on the configuration. A key metric for comparison is the number of phase shifters (PSs). We evaluate the accuracy of the classical neural network (NN) method alongside the four data reuploading methods described earlier. For the data reuploading methods, we use the *RZY-PS* configuration, except for the complex classification method, which employed the *Rot* configuration.

For classical NNs, we consider a single hidden layer with ReLU activation, as outlined by Perez et al.[9] It is important to highlight that classical NNs do not directly utilize the Clements configuration for accuracy calculations and assume a perfect ReLU nonlinear function—an implementation that remains challenging in optical neural networks (ONNs).

Figure 11 presents the simulated accuracy for the four data reuploading methods and the classical NN method. The results indicate that the implicit classification method achieves the highest accuracy when the number of PSs is very small, while the stacking method outperforms others as the number of PSs increases.

Interestingly, the classical NN method performs poorly when evaluated using the mean of 10 data points, though the mean plus sigma results suggest higher accuracy. This disparity underlines the variability in its performance.

It is particularly noteworthy that data reuploading achieves accuracy comparable to or better than classical NNs, all without relying on nonlinear optical components. This highlights the potential of data reuploading as a resource-efficient alternative in scenarios where nonlinear optical devices are impractical or undesirable.

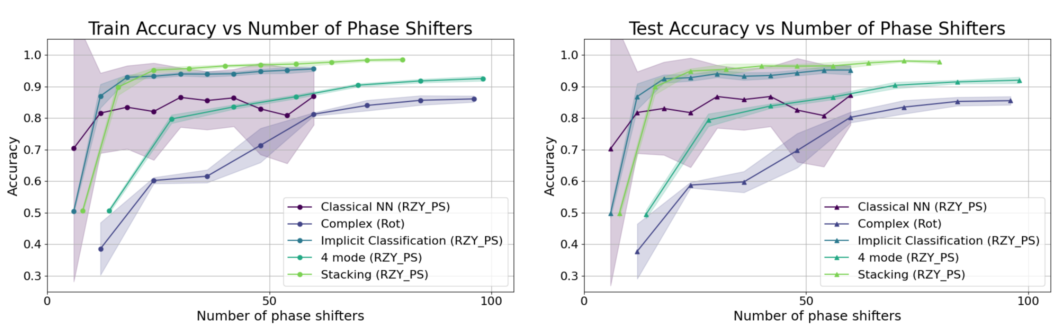


Figure 11: Accuracy of the four methods and the classical NN as a function of the number of PSs for the two-dimensional four-class wavy line problem.

6. INPUT DATA ORDERING

So far, we have used the same ordering for input data across all layers, specifically $(x_1,\, x_2)$. However, there is no inherent reason to maintain the same order. For instance, in the two-dimensional four-class problem, we evaluated $(x_1,\, x_2)$ for odd layers and $(x_2,\, x_1)$ for even layers. The difference in performance was marginal and not statistically significant. However, in certain problems, this approach led to significant improvements.

For the three-dimensional hypersphere problem, which involves classifying whether a point lies inside or outside a unit sphere, we compared two scenarios. In one, no input data shift was applied; in the other, input data alternated as $(x_1,\, x_2,\, x_3)$, $(x_2,\, x_3,\, x_1)$, and $(x_3,\, x_1,\, x_2)$. As shown in Fig. 12, the method with input data shifts demonstrated a clear advantage.

This suggests that for data reuploading, reordering input data can provide significant benefits in certain applications.

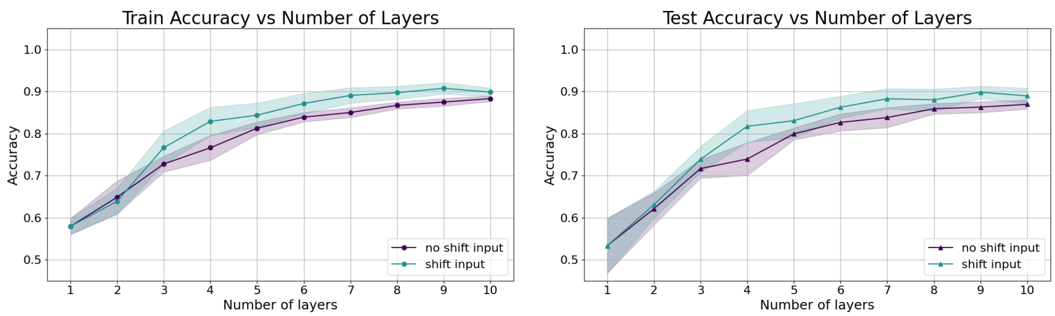


Figure 12: Training and test accuracy for the three-dimensional two-class sphere problem, comparing cases with and without input data order shifting.

7. CONCLUSION

Data reuploading was evaluated for multi-class classification problems using four distinct methods and various circuit configurations. Configurations such as two-mode class-embedding circuits, four-mode configurations, and stacked two-mode circuits with average pooling demonstrated excellent accuracy for tasks like the two-dimensional, four-class wavy lines problem. Additionally, layer-by-layer input data reordering was shown to significantly enhance performance in certain applications.

A key advantage of this approach is its ability to achieve universal photonic computing using only 50:50 beam splitters and phase shifters, eliminating the need for nonlinear photonic devices, photon counters, or squeezed light sources. This simplicity reduces resource requirements while maintaining strong performance, offering a practical and efficient alternative to conventional optical neural network architectures. These findings highlight the potential of data reuploading as a promising direction for advancing photonic computing.

Appendix

For the building block of the PIC in Fig. 1, we summarize the four distinct configurations for rotations as described in previous work.[7] Any arbitrary single-qubit operation can be decomposed into a sequence of Pauli Z, Y, and Z rotations:[12]

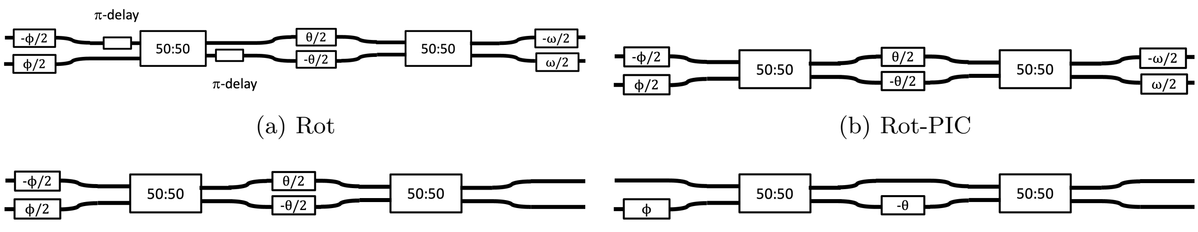
$$Rot(\phi,\theta,\omega) = RZ(\omega)RY(\theta)RZ(\phi) = \begin{bmatrix} e^{-i(\phi+\omega)/2}\cos(\theta/2) & -e^{i(\phi-\omega)/2}\sin(\theta/2) \\ e^{-i(\phi-\omega)/2}\sin(\theta/2) & e^{i(\phi+\omega)/2}\cos(\theta/2) \end{bmatrix}. \tag{1}$$

Rot: The configuration for mathematical direct implementation to Eq. 1 is shown in Fig. 13 (a). It employs three pairs of differential phase shifters (PSs) to perform rotations along the Pauli Z, Y, and Z axes.

Rot-PIC: By removing two π-delays from the first 50:50 coupler in *Rot*, this becomes equivalent to a 2×2 universal unitary optical processor used by Macho-Ortez et al.,[13] and also equivalent to $Rot(\phi - \pi, \theta + \pi, \omega)$, and is shown in Fig. 13 (b). The fixed phase difference is absorbed during training.

RZY: In cascaded blocks, the last PS pair of one block merges with the first PS pair of the next (Fig. 13 (c)), provided electrical summation is feasible.

PZY-PS: As shown in Fig. 13 (d), differential PS pairs are replaced by single-ended PSs with twice the phase shift.



(a) Rot (b) Rot-PIC (c) RZY (d) RZY-PS

Figure 13: Schematics of PIC building blocks.

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] Kojima, K. and Koike-Akino, T., "Universal photonic neural networks with quantum-free data reuploading," in [*SPIE Photonics for Quantum*], SPIE (2024).

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

[9] 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).

[10] Hamerly, R., Bandyopadhyay, S., and Englund, D., "Stability of self-configuring large multiport interferometers," *Phys. Rev. Appl.* **18**, 024018 (Aug 2022).

[11] Cabanillas, J. M. F., Zhang, B., and Popović, M. A., "Demonstration of 3±0.12 db power splitting over 145 nm optical bandwidth in a 31-μm long 3-db rapid adiabatic coupler," in [*Optical Fiber Communication Conference*], Th1A–2, Optica Publishing Group (2020).

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

[13] 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).