

Experimental data re-uploading with provable enhanced learning capabilities

Martin F. X. Mauser^{1,2}[0000-0001-5391-2837], Solène Four^{1,3}, Lena Marie Predl¹, Riccardo Albiero⁴, Francesco Ceccarelli⁴, Roberto Osellame⁴, Philipp Petersen⁵, Borivoje Dakić^{1,6}, Iris Agresti¹, and Philip Walther^{1,6}

¹ University of Vienna, Faculty of Physics, Vienna Center for Quantum Science and Technology (VCQ), Boltzmannngasse 5, Vienna 1090, Austria

² University of Vienna, Vienna Doctoral School in Physics (VDSP), Boltzmannngasse 5, 1090 Vienna, Austria

³ Master Quantum Engineering, Ecole Normale Supérieure, Université PSL, 75005 Paris, France

⁴ Istituto di Fotonica e Nanotecnologie, Consiglio Nazionale delle Ricerche (IFN-CNR), piazza L. Da Vinci 32, 20133 Milano, Italy

⁵ University of Vienna, Faculty of Mathematics and Research Network Data Science @ Uni Vienna, Kolingasse 14-16, Vienna 1090, Austria

⁶ Institute for Quantum Optics and Quantum Information Sciences (IQOQI), Austrian Academy of Sciences, Boltzmannngasse 3, Vienna 1090, Austria

Abstract. Data re-uploading offers a compelling paradigm for near-term quantum machine learning. In this work, we present the full experimental realization of the originally proposed data re-uploading scheme on a tunable integrated photonic processor. While the deep circuit requirements of multi-layer architectures present severe coherence-time bottlenecks for alternative quantum hardware, the integrated photonic platform naturally accommodates this depth. Crucially, it demonstrates how parameterizing linear optical networks can execute non-linear computations without requiring non-linear optical media. Within the Probably Approximately Correct (PAC) learning framework, we prove that our separated scheme maintains a finite Vapnik-Chervonenkis (VC) dimension scaling linearly with depth. Conversely, we show that for the examined worst-case configuration, the compressed variant popular in other implementations exhibits an infinite VC dimension, breaking formal generalization guarantees. Furthermore, Hessian analysis under these worst-case conditions reveals that the separated architecture ensures a vastly flatter optimization landscape, yielding a maximum Hessian eigenvalue (quantifying the lower bound of sharpness) of ≈ 0.23 compared to $\approx 7.23 \times 10^{12}$ for the compressed model. These results establish our hardware-validated scheme as a highly trainable subroutine with robust worst-case guarantees for scalable optical quantum learning, and simultaneously offers a blueprint for energy-efficient classical optical non-linear computation.

Keywords: Data re-uploading · integrated photonics · PAC learning · VC dimension · worst-case landscape analysis

1 Introduction

One goal of Quantum Machine Learning (QML) algorithms is to minimize the quantum resources required for classical data processing. In this context, the data re-uploading paradigm [5] stands out as an exceptionally resource-efficient framework. By sequentially alternating between data-encoding and trainable processing operations, this approach allows low-width quantum systems, even a single qubit system, to construct highly non-linear transformations [2, 5]. While other implementations often merge encoding and processing into a single joint gate to bypass temporal decoherence bottlenecks, this parameter-merging shortcut alters the hypothesis space, compromising generalization and trainability.

In this work, we demonstrate the original, separate-gate data re-uploading scheme on a tunable integrated photonic platform. We mathematically show that separating the encoding and tunable parameters is essential for optimal convergence. Experimentally, we implemented this learning scheme with single photons, as a building block for complex quantum architectures, and with coherent light, demonstrating how parameterized linear optical networks execute complex non-linear computations without requiring lossy non-linear optical media, suggesting an energy-efficient path towards scalable classical computing models [3].

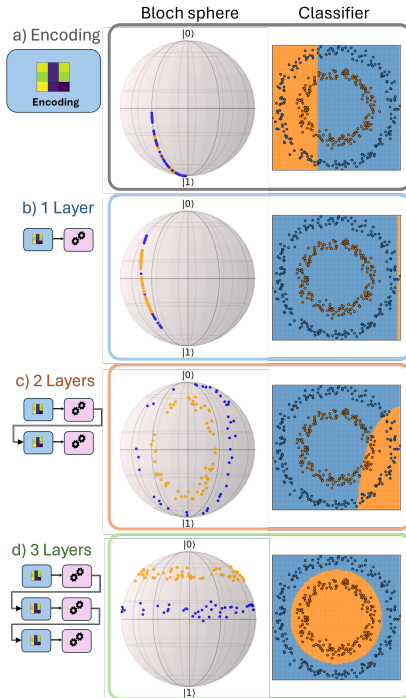


Fig. 1: **Photonic data re-uploading.** The task is to solve the circles dataset, where we want our model to distinguish between inner points and external ones. The first step (a) shows the encoding of the data points and the effect on the initial state on the Bloch sphere. The blue (yellow) color represents the correct label of the points. On the right, we show the decision boundary of the algorithm if we would stop the algorithm here. Hereby, are all the points in the yellow (blue) region are classified as yellow (blue). Then, we add the first rotation (b)) and we show the action of a second (c)) and third (d)) layer. It is visible from the shape of the coloured regions in the background how, increasing the number of layers, the model gets more complex and nonlinear. In this case, already after 3 layers, the algorithm successfully classifies the circles dataset.

2 Methods

We utilize the Vapnik-Chervonenkis (VC) dimension [7], within the Probably Approximately Correct (PAC) learning framework [6] to analyze the correct model com-

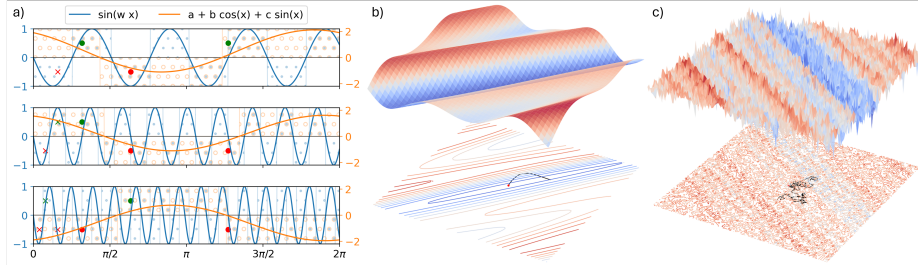


Fig. 2: Generalizability and Trainability of data re-uploading models. **a)** Visualizes how the compressed data re-uploading scheme (indicated with the blue curve) can correctly label arbitrarily many training data points, while the original one (studied in this work) can only do the same for up to 3 points (both models with one layer). This implies that the compressed scheme has an infinite VC dimension, while ours is finite, i.e. $VCdim = 3$ (for one layer). This ensures that our model can generalize to test data points. The marked points indicate the training dataset, while the color and vertical position indicate the true label. The shaded areas indicate the classification result for the two learning models. The shape of the marked points indicate if both models where able to classify all points at that horizontal position correctly (dot) or only the blue sinusoidal classifier with tunable frequency (cross). **b)** and **c)** picture the loss landscape in the two cases (with two layers), projected to 2 dimensions via a principal component analysis for visualization purposes. This demonstrates how the trainability of the model is influenced by its VC dimension. In addition, one possible optimization path using the Adam optimizer is shown on the lower heatmap of the two loss landscapes.

plexity. In our separate-gate scheme we prove that its capacity scales linearly with depth, ensuring a tight generalization error bound:

$$VCdim(\mathcal{H}_{\text{original}}) = 2l + 1 \quad (1)$$

where $\mathcal{H}_{\text{original}}$ describes the set of implementable functions with l layers.

Conversely, the compressed model maps to highly oscillating spaces $\mathcal{H}_{\text{compressed}} = \{x \mapsto \sin(\omega x)\}$. Under a worst-case dataset configuration where data points scale as $X_i = 2^i$ with random labels $y_i \in \{0, 1\}$, any arbitrary labeling can be shattered by setting the frequency $\omega = 2\pi \sum_{i=0}^N \frac{y_i}{2^i}$. Consequently, even a single compressed layer features an infinite capacity:

$$VCdim(\mathcal{H}_{\text{compressed}}) = \infty \quad (2)$$

This strips the compressed architecture of formal PAC generalizability guarantees [4]. Hessian matrix analysis quantitatively confirms how this complexity explosion damages trainability under worst-case conditions. The optimization sharpness, evaluated via the largest Hessian eigenvalue (λ_{max}), reveals an extreme divergence from a smooth, stable landscape in our architecture ($\lambda_{\text{max}} \approx 0.23$) to an untrainable, highly rugged profile in the compressed variant ($\lambda_{\text{max}} \approx 7.23 \times 10^{12}$).

Experimental Setup The scheme was realized on a tunable integrated optical circuit fabricated via femtosecond laser waveguide writing, operating in the dual-rail qubit convention. Each layer consists of Mach-Zehnder Interferometers (MZIs) mapping a 2D feature vector $x = \{x_1, x_2\}$ via internal phase shifts $\theta = \frac{\pi}{2}x_2$ and external phase shifts $\phi = \frac{\pi}{2}x_1$. Stacking $\frac{n}{2}$ encoding unitaries with

$\frac{n}{2}$ trainable unitaries per base block handles higher-dimensional inputs ($x \in R^n$). Inputs were either single photons from a Spontaneous Parametric Down Conversion (SPDC) source or coherent light. Measured computational basis probabilities were post-processed using Linear Discriminant Analysis (LDA) [1] to determine the classification threshold τ .

3 Results and Discussion

The photonic hardware implementation of data re-uploading was evaluated across four increasingly complex tasks, i.e. circles ($d = 2$), moons ($d = 2$), a 9-dimensional custom Tetromino layout and Overhead MNIST ($d = 20$, with PCA preprocessing) classification. The model achieved high classification accuracies with minimal resources. Our results prove that our hardware-validated implementation acts as a universal approximator and an effective learner, offering a highly trainable optimization landscape, high resilience to finite photon shot noise, and provable generalizability both in a quantum and energy-efficient classical domain.

References

1. Dorfer, M., Kelz, R., Widmer, G.: Deep Linear Discriminant Analysis (Feb 2016). <https://doi.org/10.48550/arXiv.1511.04707>
2. Goto, T., Tran, Q.H., Nakajima, K.: Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces. *Physical Review Letters* **127**(9), 090506 (2021)
3. Hamerly, R., Bernstein, L., Sludds, A., Soljačić, M., Englund, D.: Large-scale optical neural networks based on photoelectric multiplication. *Physical Review X* **9**(2), 021032 (2019)
4. Mauser, M.F.X., Four, S., Predl, L.M., Albiero, R., Ceccarelli, F., Osellame, R., Petersen, P., Dakić, B., Agresti, I., Walther, P.: Experimental data reuploading with provable enhanced learning capabilities. *Science Advances* **12**(15), eaeb1397 (Apr 2026). <https://doi.org/10.1126/sciadv.aeb1397>
5. Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J.I.: Data re-uploading for a universal quantum classifier. *Quantum* **4**, 226 (2020)
6. Valiant, L.G.: A theory of the learnable. *Commun. ACM* **27**(11), 1134–1142 (Nov 1984). <https://doi.org/10.1145/1968.1972>
7. Vapnik, V.N., Chervonenkis, A.Y.: On the uniform convergence of relative frequencies of events to their probabilities. In: *Measures of complexity: festschrift for alexey chervonenkis*, pp. 11–30. Springer (2015)