Mixed precision Quantum Machine Learning on Photonic Quantum and Hybrid Quantum-HPC Systems

Mateusz Slysz[1,2], Krzysztof Kurowski[1], and Grzegorz Waligóra[2]

[1] Poznań Supercomputing and Networking Center, IBCH PAS, Poznań, Poland
{mslysz, krzysztof.kurowski}@man.poznan.pl
[2] Poznań University of Technology, Poznań, Poland
grzegorz.waligora@cs.put.poznan.pl

Abstract. We study mixed precision training in hybrid neural networks combining a photonic quantum processor with classical HPC computation. In hybrid quantum–classical models, gradients are noisy due to finite quantum measurements and classical rounding errors. We analyze gradient variance to examine how these noise sources affect training stability in a simple binary classification task, comparing full precision and mixed precision training. Results show that in low-shot regimes typical of Noisy Intermediate-Scale Quantum (NISQ) devices, quantum noise dominates and mixed precision does not harm training, while at higher shot counts classical numerical errors become increasingly relevant. These findings provide practical guidance for selecting numerical precision in hybrid quantum–HPC workflows.

Keywords: Photonic Quantum Computer · Hybrid Quantum-Classical System · Quantum Machine Learning · Floating-Point Arithmetic · Mixed Precision · Gradient Variance

1 Introduction

Quantum Machine Learning (QML) [10] is an emerging interdisciplinary field that aims to harness the computational capabilities of quantum computers to enhance classical machine learning tasks, including classification, regression, and generative modeling [15, 5]. In recent years, work on hybrid quantum-classical models has grown rapidly, as purely quantum models remain limited by current hardware constraints. Photonic quantum platforms, which encode quantum information in optical modes, have attracted particular interest owing to their potential for high-speed, room-temperature operation and intrinsic robustness to certain noise sources [4]. Photonic QML implementations have demonstrated competitive performance on classification and sampling tasks, indicating practical utility even on small-scale devices [11].

In variational QML, gradients are typically estimated via repeated circuit evaluations, which introduces stochastic noise that can overwhelm small numerical errors from classical computation [3]. In practice, a large number of mea-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

surement shots are required to average out this sampling noise, which increases the computational cost significantly on real quantum processors.

While substantial effort has focused on mitigating quantum noise and reducing shot complexity, comparatively little attention has been given to the role of numerical precision in the classical component of hybrid QML workflows. In classical deep learning, mixed precision arithmetic has emerged as a standard performance optimization technique, enabling substantial reductions in memory footprint and computational time while maintaining convergence behavior [12, 7]. However, its interaction with stochastic gradient estimates arising from quantum sampling remains largely unexplored. Although mixed-precision quantum-classical algorithms have recently been proposed in contexts such as solving linear systems [8], to the best of our knowledge, there is currently no comprehensive investigation of mixed-precision training in hybrid QML workflows, where quantum measurement noise interacts with classical numerical precision.

In this work, we investigate mixed precision quantum machine learning on photonic quantum processors within hybrid High-Performance Computing (HPC) systems. We hypothesize that, in many regimes relevant to NISQ devices [14], quantum sampling noise dominates numerical precision error, enabling aggressive mixed-precision strategies that reduce classical computation time and memory requirements without detriment to training outcomes. We present an empirical study of mixed-precision training in a hybrid photonic quantum-classical machine learning environment deployed on an NVIDIA H100 GPU cluster integrated with an ORCA photonic quantum processor at Poznań Supercomputing and Networking Center.

2 Background and Problem Formulation

2.1 Variational Quantum Machine Learning

Hybrid variational quantum algorithms employ a parameterized quantum circuit $U(\boldsymbol{\theta})$ acting in an initial state $|0\rangle$, where $\boldsymbol{\theta} \in \mathbb{R}^d$ denotes a vector of trainable parameters. For supervised learning tasks, the model output is typically expressed as the expectation value of an observable $\hat{O}$. In real quantum devices, expectation values cannot be evaluated analytically and must be estimated from a finite number of measurement shots N_s. The empirical estimator is given by

$$\hat{f}(\boldsymbol{\theta}, x) = \frac{1}{N_s} \sum_{i=1}^{N_s} o_i, \tag{1}$$

where the values of o_i are stochastic measurement outcomes sampled from the quantum device.

The model parameters are optimized by minimizing a loss function $L(\boldsymbol{\theta})$ using classical optimization methods. Gradients are commonly computed using the parameter-shift rule [15], which requires multiple circuit evaluations per parameter. As a consequence, gradient estimates are also reconstructed from finite-shot expectation values and take the stochastic form:

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

$$\widehat{\nabla} L(\boldsymbol{\theta}) = \nabla L(\boldsymbol{\theta}) + \boldsymbol{\epsilon}_{\text{shot}}, \tag{2}$$

where $\boldsymbol{\epsilon}_{\text{shot}}$ denotes sampling noise induced by finite measurement statistics.

2.2 Numerical Precision in Classical Training

Floating-point arithmetic represents real numbers using finite precision formats defined in the IEEE 754 standard [6]. In this work we compare standard FP32 arithmetic with the reduced precision BF16 format.

Reduced precision computation introduces rounding errors, which can be modeled as

$$\widehat{\nabla} L(\boldsymbol{\theta}) = \nabla L(\boldsymbol{\theta}) + \boldsymbol{\epsilon}_{\text{fp}}, \tag{3}$$

where $\boldsymbol{\epsilon}_{\text{fp}}$ denotes numerical rounding error introduced by limited precision.

2.3 Interaction Between Sampling Noise and Precision Error

In hybrid variational quantum machine learning, gradient estimates computed in mixed precision take the combined form:

$$\widehat{\nabla} L(\boldsymbol{\theta}) = \nabla L(\boldsymbol{\theta}) + \boldsymbol{\epsilon}_{\text{shot}} + \boldsymbol{\epsilon}_{\text{fp}}, \tag{4}$$

where $\boldsymbol{\epsilon}_{\text{shot}}$ denotes the stochastic sampling noise of finite quantum measurements, and $\boldsymbol{\epsilon}_{\text{fp}}$ is the numerical rounding error of a low precision classical computation. The relative magnitude of these two error sources determines their impact on optimization. When $\|\boldsymbol{\epsilon}_{\text{shot}}\| \gg \|\boldsymbol{\epsilon}_{\text{fp}}\|$, the effect of reduced numerical precision is expected to be negligible compared to the quantum sampling noise. In contrast, if the shot count is large and $\boldsymbol{\epsilon}_{\text{shot}}$ is small, numerical precision may become a significant factor in convergence.

2.4 Shot Noise Scaling and Gradient Variance

To understand how quantum measurement noise and classical numerical errors affect training, we analyze gradient variance, defined as the variance of parameter gradients across mini-batches.

Consider a quantum observable $\hat{O}$ whose expectation value is estimated using N_s independent measurement shots. Each measurement produces a random outcome o_i with finite variance $\sigma^2 = \text{Var}[o]$. The empirical estimator $\hat{f}$ of the expectation value is given in equation (1). Since the measurement outcomes are independent and identically distributed, the variance of the estimator is

$$\text{Var}[\hat{f}] = \text{Var}\left[\frac{1}{N_s}\sum_{i=1}^{N_s} o_i\right] = \frac{1}{N_s^2}\sum_{i=1}^{N_s}\text{Var}[o_i] = \frac{\sigma^2}{N_s}. \tag{5}$$

Thus, quantum measurement noise decreases inversely with the number of shots. When gradients are computed via the parameter-shift rule, each gradient

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

component is formed from differences of such expectation value estimates. As a result, the variance of gradient estimates inherits the same scaling behavior,

$$\mathrm{Var}[\nabla L(\theta)] \propto \frac{1}{N_s}. \tag{6}$$

This $1/N_s$ dependence provides a theoretical baseline for interpreting experimental gradient variance measurements. Deviations from this scaling indicate the presence of additional noise sources, such as classical numerical precision errors introduced by mixed precision arithmetic.

By comparing FP32 training to mixed precision (BF16/AMP) training across different shot counts, we can determine which source of noise dominates. If variance is similar between FP32 and AMP, quantum shot noise is the main contributor ($\|\boldsymbol{\epsilon}_{\mathrm{shot}}\| \gg \|\boldsymbol{\epsilon}_{\mathrm{fp}}\|$), while larger fluctuations in AMP indicate that classical rounding errors are becoming significant. Systematically varying N_s and recording gradient variance for both precision modes allows us to identify the regimes where shot noise or numerical precision governs the optimization dynamics.

3 Experimental Setup

We evaluate mixed precision training in a hybrid quantum-classical scenario using a simplified binary classification version of the MNIST dataset [9], where the task is to distinguish between digits 0 and 1. Input images are normalized and flattened into 28×28 vectors before being fed into the network. This task provides a controlled environment to study the interaction between quantum measurement noise and classical numerical precision, while remaining computationally tractable on current hardware. All experiments were performed using the ORCA PT-1 [1, 2] photonic quantum processors installed at the Poznań Supercomputing and Networking Center (PCSS). The ORCA PT-1 systems provide 8 optical modes (qumodes) and implement the boson sampling paradigm using a time-bin encoding of the optical modes. The devices operate in a single-loop architecture with 7 programmable beam-splitter parameters, as each adjacent pair of qumodes is connected by a single beam splitter. The system also supports computations in a double-loop configuration, which provides 14 programmable beam-splitter parameters. All experiments were performed on a high-performance computing node equipped with an NVIDIA H100 GPU and connected to the ORCA photonic quantum processor.

The hybrid model combines classical neural network layers with a photonic quantum processing layer implemented via the ORCA photonic platform. The classical pre-quantum block consists of fully connected layers designed to extract features from the high-dimensional input: we use two layers mapping $784 \rightarrow 64$ neurons, each followed by a Rectified Linear Unit (ReLU) activation. This block outputs the parameters required by the quantum layer. The quantum layer itself is implemented as a programmable interferometer (PTLayer) with $m = 6$ optical modes, using a time-multiplexed boson sampling architecture and photon-number detection to encode and process the classical features. Following

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

the quantum layer, a classical post-quantum block maps the quantum outputs to a single logit for binary classification; the post-quantum block consists of one hidden layer with 64 neurons and a ReLU activation. Automatic Mixed Precision (AMP) [12] is applied on the classical blocks, while the PTLayer computations remain in full FP32 to preserve quantum numerical accuracy.

For each experiment, we vary both the number of shots $N_s \in \{10, 20, 50, 100\}$ and the numerical precision mode (FP32 or AMP BF16 for classical layers), while keeping all other hyperparameters fixed, including a learning rate of 1×10^{-3}, batch size of 128, and 10 training epochs. To evaluate the effect of numerical precision on training stability, we measure the final loss and the gradient variance across mini-batches for each combination of shot number and precision mode. This allows us to quantify the relative contributions of stochastic quantum noise and classical rounding errors in the hybrid training loop. The entire experimental pipeline is implemented in PyTorch [13], using a PTLayer wrapper to isolate FP32 quantum computations from AMP classical operations. The results for these runs are gathered in Table 1 and visualised on a plot in Figure 1. Each configuration was evaluated over three independent training runs, and the reported gradient variances represent averages across these runs. The observed trends were qualitatively consistent, suggesting that the reported behavior reflects systematic effects rather than random fluctuations.

Shots (N_s)	Precision	Final Loss	Total Time [s]	Gradient Variance
10	FP32	0.00332	25.29	0.211
10	AMP BF16	0.00101	25.15	0.056
20	FP32	0.00039	31.25	0.0071
20	AMP BF16	0.00231	31.89	0.321
50	FP32	0.00034	50.55	$1.32{\cdot}10^{-5}$
50	AMP BF16	0.00065	51.02	0.0117
100	FP32	0.00029	80.70	$1.14{\cdot}10^{-4}$
100	AMP BF16	0.00177	81.78	0.0952

Table 1. Final loss, training time, and gradient variance for different shot counts and precision modes collected in experiments.

The hybrid model successfully learns the binary MNIST classification task across all tested shot counts and precision modes, with final losses below 0.01. Training under both FP32 and mixed precision (AMP/BF16) remains stable, although gradient fluctuations vary across shot regimes.

We additionally measure total runtime as a function of the number of measurement shots N_s and numerical precision. These measurements allow us to analyze how quantum sampling noise and classical numerical precision interact during training.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

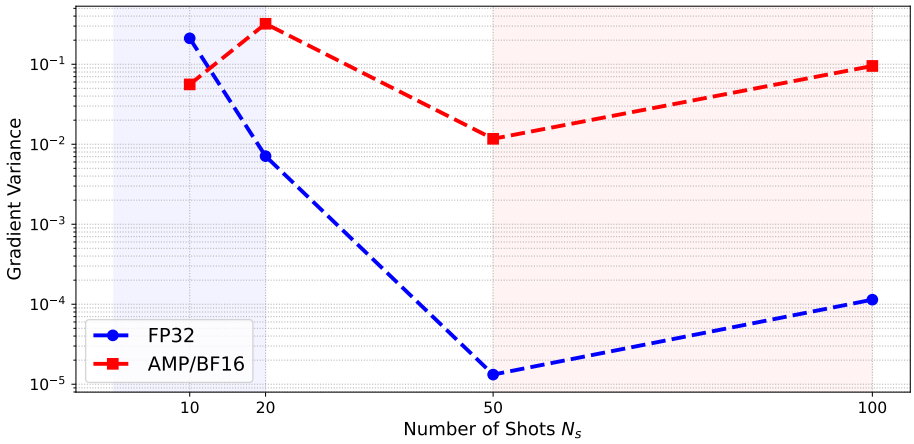

Fig. 1. Gradient variance of the hybrid photonic quantum–classical neural network as a function of the number of measurement shots N_s. Shaded regions highlight operational regimes: low-shot (blue) where quantum sampling noise dominates, and high-shot (red) where classical numerical precision becomes significant.

4 Discussion

The experimental results confirm the expected inverse scaling of gradient variance with respect to the number of measurement shots N_s under full FP32 precision. As predicted by the theoretical analysis, increasing N_s systematically reduces stochastic fluctuations in the gradient estimates, closely following the ideal $1/N_s$ trend. This validates that quantum measurement noise behaves as statistically independent sampling noise in the studied regime.

In contrast, mixed precision (BF16/AMP) training exhibits deviations from ideal scaling. While gradient variance also decreases with increasing N_s, its magnitude remains consistently higher than in FP32 for moderate and large shot counts. This behavior indicates that when quantum shot noise becomes sufficiently small, classical numerical precision errors begin to contribute noticeably to gradient fluctuations. Interestingly, at very low shot counts, the variance under mixed precision does not significantly differ from FP32 values. In this regime, quantum sampling noise dominates the optimization dynamics, effectively masking rounding effects introduced by reduced precision arithmetic. As N_s increases and shot noise diminishes, the relative contribution of numerical precision errors becomes more visible, leading to a divergence between FP32 and AMP variance curves.

All these observations suggest the existence of distinct operational regimes. In the low-shot regime, optimization is governed primarily by quantum measurement noise, and mixed precision does not significantly degrade training stability. In higher-shot regimes, however, where quantum noise is suppressed, classical

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

rounding errors can become comparable in magnitude, influencing gradient stability and potentially affecting convergence dynamics.

Finally, the observed total runtime increases approximately linearly with N_s, reflecting the direct proportionality between the number of quantum circuit evaluations and the number of measurement samples required for expectation estimation. These results suggest that, in hybrid photonic–GPU workflows, optimizing shot allocation has a significantly larger impact on performance than reducing classical numerical precision. Mixed precision may still provide memory or throughput benefits in larger-scale models, however, in the present setting, quantum sampling remains the primary bottleneck.

5 Conclusions and Future Work

In this work, we investigated mixed precision training in a hybrid neural network combining a photonic quantum processor with classical computation on an HPC cluster. Our experiments on a binary MNIST classification task demonstrate that the interplay between quantum measurement noise and classical numerical precision governs training stability. In low-shot regimes typical of NISQ devices, quantum sampling noise dominates, enabling aggressive mixed-precision strategies (BF16/AMP) without degrading model performance. As the number of shots increases, classical rounding errors become more significant, highlighting the need for precision-aware algorithm design in hybrid quantum–HPC workflows. From a practical perspective, this interplay provides guidance for precision selection in hybrid quantum–HPC workflows.

For future work, we plan to scale to larger datasets and more complex architectures, including multi-class classification and deeper hybrid networks. In addition, utilising multi-GPU and multi-QPU setups could significantly accelerate hybrid computations, enabling the simultaneous parallelization of classical and photonic operations as initially demonstrated in [16]. Finally, extending these methods to other quantum machine learning tasks, such as generative modeling or quantum kernel methods, will help generalize our findings. Furthermore, investigating adaptive precision strategies that dynamically balance quantum noise and classical numerical accuracy during training represents a promising direction for improving efficiency and robustness.

Acknowledgments. We acknowledge the Poznań Supercomputing and Networking Center for providing access to the ORCA PT-1 photonic quantum systems and GPU cluster. This research has been funded by the Program of the Polish Ministry of Science and Higher Education "Applied Doctorate" realized in years 2022–2026 (agreement no. DWD/6/0142/2022) and by the Poznań University of Technology, Poland (project no. 0311/SBAD/ 0764).

References

1. AbuGhanem, M.: Toward scalable fault-tolerant photonic quantum computers. The Journal of Supercomputing **82**(2), 51 (2026). https://doi.org/10.1007/s11227-025-08132-7, https://doi.org/10.1007/s11227-025-08132-7

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50

2. Bradler, K., Wallner, H.: Certain properties and applications of shallow bosonic circuits. arXiv preprint arXiv:2112.09766 (2021)
3. Facelli, G., Roberts, D.D., Wallner, H., Makarovskiy, A., Holmes, Z., Clements, W.R.: Exact gradients for linear optics with single photons. arXiv preprint arXiv:2409.16369 (2024)
4. Flamini, F., Spagnolo, N., Sciarrino, F.: Photonic quantum information processing: a review. Reports on Progress in Physics **82**(1), 016001 (2019)
5. Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M.: Supervised learning with quantum-enhanced feature spaces. Nature **567**(7747), 209–212 (2019)
6. IEEE: IEEE Standard for Binary Floating-Point Arithmetic. ANSI/IEEE Std 754-1985 pp. 1–20 (1985). https://doi.org/10.1109/IEEESTD.1985.82928
7. Kalamkar, D., Mudigere, D., Mellempudi, N., Das, D., Banerjee, K., Avancha, S., Vooturi, D.T., Jammalamadaka, N., Huang, J., Yuen, H., Yang, J., Park, J., Heinecke, A., Georganas, E., Srinivasan, S., Kundu, A., Smelyanskiy, M., Kaul, B., Dubey, P.: A Study of BFLOAT16 for Deep Learning Training (2019), https://arxiv.org/abs/1905.12322
8. Koska, O., Baboulin, M., Gazda, A.: A mixed-precision quantum-classical algorithm for solving linear systems (2025), https://arxiv.org/abs/2502.02212
9. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
10. Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum algorithms for supervised and unsupervised machine learning (2013), https://arxiv.org/abs/1307.0411
11. Madsen, L.S., Laudenbach, F., Askarani, M.F., Rortais, F., Vincent, T., Bulmer, J.F., Miatto, F.M., Neuhaus, L., Helt, L.G., Collins, M.J., et al.: Quantum computational advantage with a programmable photonic processor. Nature **606**(7912), 75–81 (2022)
12. Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., Wu, H.: Mixed Precision Training. In: International Conference on Learning Representations (ICLR) (2018), https://openreview.net/forum?id=r1gs9JgRZ
13. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
14. Preskill, J.: Quantum Computing in the NISQ era and beyond. Quantum **2**, 79 (2018). https://doi.org/10.22331/q-2018-08-06-79, https://doi.org/10.22331/q-2018-08-06-79
15. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J., Killoran, N.: Evaluating analytic gradients on quantum hardware. Physical Review A **99**(3), 032331 (2019). https://doi.org/10.1103/PhysRevA.99.032331
16. Slysz, M., Łukasz Grodzki, Rydlichowski, P., Siera, D., Kurowski, K., Waligóra, G., Węglarz, J.: Solving combinatorial optimization and machine learning problems on hybrid near-term quantum photonic computers. Future Generation Computer Systems **174**, 107934 (2026). https://doi.org/https://doi.org/10.1016/j.future.2025.107934, https://www.sciencedirect.com/science/article/pii/S0167739X25002298

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_50