

Layerwise Fourier–Dirac Supervision for Modular Generalisation in Data-Reuploading Quantum Neural Networks

Carrington Kinyanjui^{1,2} Martin Wagah^{2,3}

¹University of Manchester ²Songhai ³University of Cambridge

Abstract

Data-reuploading quantum neural networks (QNNs) [1] provide a natural setting for studying how variational quantum circuits encode structured classical data [2, 3]. Because input variables are repeatedly injected into parameterised rotations, these models can be interpreted as learning trigonometric feature maps over the input space [4]. This connection is consistent with recent Fourier-fingerprint analyses of QML ansatzes, which emphasize that circuit structure shapes the Fourier modes available to a model [7]. In our setting, modular addition provides a concrete finite-group task for observing how a data-reuploading circuit uses such Fourier structure during learning. This makes them especially relevant for finite-group learning problems, where the target function has an explicit spectral structure. We study modular addition as a controlled probe of algorithmic generalisation in data-reuploading QNNs. The task is to learn a y such that:

$$y = (a + b) \bmod p, \quad a, b, y \in \mathbb{Z}_p, \quad (1)$$

from a sparse subset of ordered operand pairs. Although simple to specify, modular addition requires recovery of a cyclic group law rather than interpolation of local numerical patterns [2], and it has become a useful model problem for grokking and Fourier-like algorithm formation [5].

We introduce *layerwise Fourier–Dirac supervision*, a training and diagnostic intervention that attaches finite-modular Fourier readouts after each data-reuploading layer rather than only at the final circuit state. If a QNN gradually builds a cyclic representation of the operands, intermediate states should already contain partial modular evidence. A final readout alone may underuse this distributed information. By supervising and aggregating predictions across layers, we test whether modular arithmetic is more robustly learned as a depth-distributed rule.

The simulated model is a 7-qubit statevector data-reuploading circuit with four reuploading layers, giving a 128-dimensional Hilbert space. Each layer applies operand-dependent trainable single-qubit rotations to periodic encodings of a and b , followed by a fixed ring entangling block. The readout families use computational-basis probabilities, compact expectation features, or finite Fourier–Dirac heads. No direct encoding of $(a + b) \bmod p$ is provided, and the present experiments include neither hardware noise nor shot noise.

The finite-modular readout predicts per-frequency sum features $(\hat{u}_k(a, b), \hat{v}_k(a, b))$ and scores a candidate residue c by

$$D_K(a, b, c) = \sum_{k=1}^K \rho_k \left[\hat{u}_k(a, b) \cos\left(\frac{2\pi kc}{p}\right) + \hat{v}_k(a, b) \sin\left(\frac{2\pi kc}{p}\right) \right]. \quad (2)$$

The full class logit can include a class bias and, in some variants, a residual linear path,

$$\ell_c(a, b) = \beta_c + \gamma D_K(a, b, c) + r_c(a, b). \quad (3)$$

When the predicted Fourier features encode the phase of $(a + b) \bmod p$, this score approximates a modular kernel over $(a + b - c) \bmod p$. In the layerwise setting, the same readout form is attached after each reuploading layer, making intermediate quantum feature states directly probeable for cyclic arithmetic structure.

We compare three layerwise QNN architectures. *Auxiliary layerwise supervision* attaches Fourier–Dirac heads to all intermediate layers during training but deploys only the final prediction. *Learned residual aggregation* learns a weighted mixture over layerwise logits. *Fixed uniform layerwise averaging* removes learned gating and deploys the simple mean of all layerwise logits,

$$\ell_c^{\text{mean}}(a, b) = \frac{1}{L} \sum_{\lambda=1}^L \ell_c^{(\lambda)}(a, b), \quad (4)$$

where L is the number of supervised layers. This fixed-mean model is motivated by a post-hoc diagnostic in which averaging the intermediate heads of the auxiliary architecture produced a stronger held-out rule than the deployed final head.

We evaluate modular addition over prime moduli $p \in \{31, 97, 127\}$ with sparse train/test splits and three split seeds per reported condition. Matched classical Fourier-delta baselines and product-Fourier controls provide interpretive anchors: the controls show what an explicit Fourier representation can do, while the QNN experiments reveal how comparable structure can emerge, fail, or become more robust inside a constrained variational circuit.

Modulus	Train fraction	Fixed-mean QNN	Residual QNN	Matched Fourier-delta
31	30%	0.3948 ± 0.0464	0.3754	0.0228
31	70%	0.8800 ± 0.0322	0.8604	near chance
97	50%	0.9907 ± 0.0019	0.9783	0.9955
97	70%	0.9943 ± 0.0022	0.9811	0.9996
127	50%	0.9963 ± 0.0029	0.9668	0.9997
127	70%	0.9871 ± 0.0024	0.9673	0.9999

Table 1: Held-out exact accuracies reported for the principal modular-addition regimes.

The main result is that fixed uniform layerwise averaging is the strongest QNN architecture in the robust regimes. At $p = 97$, the fixed-mean QNN reaches held-out exact accuracy 0.9907 ± 0.0019 with 50% training coverage and 0.9943 ± 0.0022 with 70% coverage, compared with 0.9783 and 0.9811 for learned residual aggregation. At $p = 127$, fixed averaging reaches 0.9963 ± 0.0029 with 50% coverage and 0.9871 ± 0.0024 with 70% coverage, again outperforming residual aggregation. Matched Fourier-delta baselines approach or exceed QNN performance for $p = 97$ and $p = 127$, while an explicit product-Fourier control solves all evaluated regimes. Thus the contribution does not claim quantum advantage.

The smaller modulus $p = 31$ exposes a distinct regime. At 30% training coverage, fixed-mean QNN performance is weak, reaching only 0.3948 ± 0.0464 , with the residual variant similarly low at 0.3754. Increasing coverage to 50% improves the fixed-mean model to 0.6514 and the residual model to 0.6840. With 70% training coverage, fixed averaging reaches 0.8800 ± 0.0322 , while residual

aggregation reaches 0.8604; the matched Fourier-delta baseline remains near chance. This suggests that, in sparse or small-table conditions, the learner can fit examples while failing to identify a globally consistent modular rule unless coverage is increased or the feature basis makes operand interaction explicit.

We find that naive QNN readouts learn nontrivial cyclic structure but can remain too coarse, producing adjacent-residue errors rather than exact classification. Fourier-Dirac heads are observed to sharpen this cyclic scaffold into a more residue-selective rule. Layerwise aggregation further improves generalisation because useful modular evidence is distributed across intermediate circuit states. Fixed averaging works well because it prevents the model from placing nearly all responsibility on a single final learned readout and instead rewards residue classes supported consistently across depth.

This is supported by further ablations carried out. Feeding intermediate Fourier coefficients back into later circuit inputs through an adapter architecture performs worse than auxiliary supervision and layerwise averaging, suggesting that residue-aligned information is useful as a readout signal but not necessarily as a control signal for subsequent rotations. Same-sum invariance objectives, which train equivalent operand pairs such as $(a + d, b - d)$ to produce consistent predictions, improve seed-0 layerwise variants and support the view that explicitly teaching the modular equivalence class is beneficial.

The contributions are threefold. First, modular addition is used as a finite-group testbed for analysing how data-reuploading QNNs form Fourier-like representations of algebraic rules. Second, layerwise Fourier-Dirac supervision provides a direct way to probe and train intermediate quantum feature states for cyclic arithmetic structure. Third, fixed uniform averaging of layerwise modular evidence consistently outperforms learned residual aggregation in the robust $p = 97$ and $p = 127$ regimes. Although the experiments use exact statevector simulation, layerwise readouts could extend to noisy or hardware-constrained settings as intermediate training signals, early-exit diagnostics, or noise-aware probes of where group-structured information first appears. A hardware implementation would require separate circuit executions, non-destructive measurement strategies, or shallow early-exit variants to avoid destructive mid-circuit measurement collapse.

References

- [1] A. Pérez-Salinas et al., “Data Re-uploading for a Universal Quantum Classifier,” *Quantum*, 2020.
- [2] V. Havlíček et al., “Supervised Learning with Quantum-Enhanced Feature Spaces,” *Nature*, 2019.
- [3] M. Schuld and N. Killoran, “Quantum Machine Learning in Feature Hilbert Spaces,” *Physical Review Letters*, 2019.
- [4] M. Schuld, R. Sweke, and J. J. Meyer, “The Effect of Data Encoding on the Expressive Power of Variational Quantum Machine Learning Models,” *Physical Review A*, 2021.
- [5] A. Power et al., “Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets,” 2022.
- [6] N. Nanda et al., “Progress Measures for Grokking via Mechanistic Interpretability,” 2023.
- [7] M. Strobl et al., “Fourier Fingerprints of Ansatzes in Quantum Machine Learning,” arXiv preprint arXiv:2508.20868, 2025.