Quantum Reupload Units: A Scalable and Expressive Approach for Time Series Learning

Léa Cassé
University of Waikato, Hamilton, New Zealand
École Polytechnique, IP Paris, France
lc480@students.waikato.ac.nz

Sabarikirishwaran Ponnambalam
Griffith University
Nathan, Queensland, Australia
sabarikirishwaran@gmail.com

Bernhard Pfahringer
University of Waikato
Hamilton, New Zealand
bernhard.pfahringer@waikato.ac.nz

Albert Bifet
University of Waikato, Hamilton, New Zealand
LTCI, Télécom Paris, IP Paris, France
abifet@waikato.ac.nz

***Abstract*—We propose a single-qubit Quantum Machine Learning (QML) model for time series forecasting, built around the concept of a Quantum Reupload Unit (QRU), a hardware-efficient quantum circuit architecture with shallow depth. The proposed model demonstrates enhanced predictive power compared to variational methods such as quantum circuits (VQC), parameterized quantum circuits (PQC), and quantum residual blocks (QRB). The proposed QRU outperforms classical learning models such as Recurrent Neural Networks (RNNs) and Long-Short Term Memory (LSTM) with the same number of parameters. The novelty of this approach is its ability to model temporal patterns without relying on an extensive memory state, which reduces resource demands while preserving forecast accuracy. The expressivity of the model is evaluated through Fourier spectral decomposition. We analyze the trainability of our model using the absorption witness metric. We benchmarked the proposed model on the Mackey-Glass chaotic time series and the real-world river level dataset from TAIAO. The proposed model consistently exhibits enhanced expressivity over both of the datasets. These results highlight the significance of QRUs as promising candidates for learning models that can be conveniently deployed on noisy intermediate-scale quantum (NISQ) hardware.**

I. Introduction

Quantum Machine Learning (QML) leverages quantum effects like superposition and entanglement, promising enhanced computational capabilities and model expressivity compared to classical approaches [1]. Advances in Noisy Intermediate-Scale Quantum (NISQ) devices have spurred the development of practical quantum learning models, such as Parameterized Quantum Circuits (PQCs) [2] and Variational Quantum Circuits (VQCs) [3]. Despite their successes in classification and regression tasks, these circuits suffer from limited expressivity and challenging optimization due to shallow encoding schemes and barren plateau phenomena [4]. Such limitations are particularly critical in time series forecasting, which demands capturing intricate temporal dependencies and rapid variations in the data [5].

To overcome these issues, Quantum Re-uploading Units (QRUs) have been introduced as single-qubit models that iteratively encode data into quantum circuits, enhancing function approximation capabilities and gradient stability [6], [7]. Complementing this, Quantum Residual Blocks (QRBs) extend PQC expressivity by integrating residual learning principles from classical neural networks [8], [9]. However, QRBs typically require resource-intensive operations involving 2-qubit entanglement gates and ancillary qubits, making them challenging for NISQ devices. To address these hardware constraints, we propose a hybrid architecture—QRU–QRB–Local—that combines QRUs' minimal quantum resource demands with QRBs' enhanced spectral diversity through a shared ancillary strategy.

In this work, we rigorously assess the proposed QRU-based models using spectral decomposition, absorption witness metrics, and quantum state diversity measures. Empirical evaluations on both synthetic benchmark datasets, such as chaotic Mackey–Glass series, and real-world datasets like historical river water levels, demonstrate significant performance improvements over classical recurrent models (RNNs) and standard quantum baselines (PQC, VQC, QRB). Our key contributions include:

- A scalable single-qubit quantum architecture for efficient and expressive time series forecasting.
- Comprehensive validation across diverse datasets confirming strong generalization and trainability.
- A reproducible analytic framework linking quantum model expressivity and spectral properties directly relevant to NISQ applications.

II. Related work

The impact of quantum data feedback loops on the model's expressivity and trainability has got significant attention in recent years. Pérez-Salinas et al. [10] introduced a quantum feedback loop based on iterative data re-uploading, capable of approximating any continuous function without two-qubit gates. Their architecture uses alternating data-encoding and parameterized layers, allowing the model to learn complex decision boundaries has been rigorously shown to satisfy the classical universal approximation theorem.

Building on this, Barthe and Pérez-Salinas [7] analyzed the spectral behavior of data re-uploading models, showing

that iterative encoding smooths out the spectral function and improves trainability. This study introduced the metric -Absorption Witness determine the influence of data re-uploading in the on the learning capacity of the model.

Yet, literature lacks empirical studies on how design choices and encoding schemes affect expressivity and generalization. This work addresses these gaps via a systematic empirical study of QRU-based models for real-world time series forecasting, analyzing their spectra, expressivity, gradients, and feature mappings.

III. Theory

This section outlines the foundational concepts of the Quantum Re-uploading Units (QRU) and Quantum Residual Blocks (QRBs), following increasing model complexity.

A. Quantum Re-uploading Units (QRUs)

QRUs are parameterized quantum circuits (PQCs) leveraging iterative data re-uploading to enhance expressivity. The unitary transformation is given by:

$$U(\theta, x) = \prod_{j=1}^{M} \exp(ig_j x)\, W_j\, \exp(iV_j\theta_j), \tag{1}$$

where g_j and V_j are Hermitian operators [11], W_j are fixed unitaries, and θ_j are trainable parameters. Re-uploading allows the circuit to represent more complex functions than single-layer embeddings [12].

Each layer is defined as:

$$L(j) = U(\theta_j)\, U(x), \tag{2}$$

and the total unitary is:

$$U(\vec{\theta}, \vec{x}) = \prod_{j=1}^{N} \big(U(\vec{\theta}_j)\, U(\vec{x})\big),$$

with $U(\vec{x})$ encoding classical input via rotations R_x, R_y, R_z. This repeated injection enables the circuit to learn complex patterns without duplicating states.

QRUs approximate any continuous function [13]:

$$h_\theta(x) = \sum_{\omega\in\Omega} C_\omega(\theta)\, e^{i\omega x}, \tag{3}$$

where Ω is defined by the eigenvalues of g_j, and spectral richness increases with depth L [7].

QRUs process the same input recursively through trainable and encoding layers, promoting coherence and parameter efficiency while mitigating decoherence [14]. This avoids deep entangling operations, optimizing parameter usage. The output gradient is:

$$\frac{\partial h_\theta(x)}{\partial\theta_j} = \sum_{l=1}^{L} \frac{\partial c_{\omega_l}(\theta)}{\partial\theta_j} e^{i\omega_l x} \tag{4}$$

Repeated encoding disrupts circuit homogeneity, avoiding barren plateaus [15], [16].

B. Quantum Residual Blocks (QRBs)

QRBs, inspired by classical residual networks [9], enhance expressivity and trainability via residual connections [8]. A QRB applies:

$$R(x, \theta) = \frac{1}{\sqrt{2}} \left[I + L(x, \theta)\right], \tag{5}$$

acting on $|\psi\rangle$ as:

$$R(x, \theta)|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\psi\rangle + L(x, \theta)|\psi\rangle\right), \tag{6}$$

with $L(x, \theta) = U(x)W(\theta)$:

$$U(x) = e^{-ixG}, \quad G|g_k\rangle = \omega_k|g_k\rangle, \tag{7}$$

where $W(\theta)$ is a trainable unitary. The residual link adds $|\psi\rangle$ to $L(x, \theta)|\psi\rangle$, allowing information to bypass transformations and improving gradient flow.

Without residuals, frequencies from $U(x)|\psi\rangle$ come from $\{\pm\omega_k\}$. With residuals, cross-terms arise:

$$\big(|\psi\rangle + L|\psi\rangle\big)\big(\langle\psi| + \langle\psi|L^\dagger\big), \tag{8}$$

producing a richer frequency set:

$$\Omega_R = \{\omega_k - \omega_j,\ \pm\omega_k \mid k, j \in [d]\}, \tag{9}$$

This extends expressivity, allowing QRBs to model a broader function class. By leveraging interference and superposition, QRBs preserve gradient variance and counteract barren plateaus.

IV. Methodology

A. Training Configuration

All quantum circuits are implemented using the `PennyLane` framework and simulated on a qubit-based statevector backend. Each circuit processes an input vector $x \in \mathbb{R}^n$, where n denotes the windowed sequence length derived from the time-series data. The dataset is segmented into overlapping subsequences of fixed length n, which are encoded into quantum states via repeated data re-uploading. Each circuit is parameterized by a trainable parameter set $\boldsymbol{\theta} \in \mathbb{R}^P$, where the total number of parameters P is a function of the model-specific structure and the circuit depth d. The parameter count is adjusted to ensure architectural parity across models, enabling fair comparisons of expressibility and learning capacity.

[1] A re-uploading unit consists of a sequence of parametrized rotations R_X, R_Y, R_Z generally denoted as $U(x)$, applied to each input feature(s). Fig. 1 illustrates the re-uploading scheme at depth d, where the same pattern is repeated across all $k \in \{1, \ldots, n\}$ input sequence.

The combined unitary transformation of QRU can be denoted as $U(\theta, x)$ with observable O:

$$h_\theta(x) \;=\; \langle 0\,|\,U^\dagger(\theta, x)\, O\, U(\theta, x)\,|0\rangle, \tag{10}$$

[1]Full Implementation details are available in the codebase at https://github.com/Sabarikirishwaran/QreuploadUnit

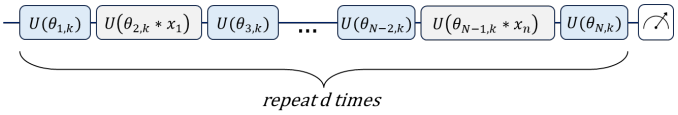


Fig. 1: Schematic of the re-uploading architecture consisting of alternating layers of trainable unitary gates and data-dependent parameterized gates. Each layer applies a sequence of unitaries $U(\theta_{i,k})$, interleaved with input-modulated operations of the form $U(\theta_{j,k} \cdot x_j)$, where $x_j \in \mathbb{R}$ is the classical input feature. The subscript $k \in \{1, 2, \ldots, d\}$ denotes the layer index (depth), and $i \in \{1, 2, \ldots, n\}$ indexes the position within the input sequence of length n.

This work has compared following architectures with baseline QRU model:

Model	Description	Params	Qubits
QRU(baseline)	Single-qubit with input re-uploading using R_X, R_Y, R_Z per feature multiplied by a trainable parameter.	$3\ n\ d$	1
PQC	Single-qubit with one input encoding followed by parameterized layer.	$3\ n\ d$	1
VQC	Multi-qubits - one qubit per input and entanglement layer via pennylane's basic entanglement layer template.	$3\ n\ d$	n
QRU–QRB-Local	Two-qubit with residual CRX connection and one ancilla qubit.	$3\ n\ d + d$	2
QRU–QRB-Global	d-ancilla qubits with residual CRX connection.	$3\ n\ d + d$	d+1

TABLE I: Quantum circuit models with descriptions, parameter count, and qubit scaling with input size n and depth d.

B. Dataset

Each model was trained using a sliding-window strategy, where a fixed-length input sequence of n time steps was used to predict the subsequent value in the series. The input values were encoded into quantum states via parameterized single-qubit rotational gates with each gate angle modulated by both the input data and trainable parameters. All quantum circuits operated on a single data qubit with variable depth, where the depth corresponds to the number of data re-uploading layers. Training was performed using the Adam optimizer with a learning rate of 0.01 for 300 epochs. The loss function employed was the Huber loss, defined as:

$$L(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & \text{if } |y - \hat{y}| \leq \delta \\ \delta(|y - \hat{y}| - \frac{1}{2}\delta), & \text{otherwise} \end{cases} \tag{11}$$

where $\delta = 0.1$. The accuracy is measured as the fraction of predictions within a threshold of 0.1 from the true value.

1) Synthetic Dataset: Sinusoidal Wave: consists of a univariate sinusoidal function defined as:

$$x(t) = A \sin(2\pi f t), \tag{12}$$

where A represents the amplitude and f the frequency.

2) Real-World Dataset: River Water Levels: Tested the models on realistic time-series prediction tasks, The Coromandel River and Rain Gauge Time Series dataset contains ten-year historical river and rain gauge measurements from the Coromandel region, provided by the TAIAO project[2]. This dataset captures real-world complexity due to measurement noise, inherent variability, and uncertainties. In contrast to synthetically generated data, it exhibits non-stationary and irregular patterns, therefore suitable for assessing model robustness and generalization.

3) Chaotic Dataset: Mackey-Glass Time Series: it is a well-known chaotic time-series governed by the following delayed differential equation:

$$\frac{dx}{dt} = \beta \frac{x(t - \tau)}{1 + x(t - \tau)^n} - \gamma\, x(t). \tag{13}$$

parameters used for this generation:

$$\tau = 17, \quad \beta = 0.2, \quad \gamma = 0.1, \quad n = 10, \quad x(0) = 1.2.$$

Therefore, a discrete-time approximation based on the Euler method is used to generate the series. Owing to its chaotic and nonlinear dynamics, the Mackey–Glass series poses a significant modeling challenge, requiring models to accurately capture short-term dependencies to approximate the underlying target function effectively.

V. Experiments

This section presents a series of controlled experiments carried out to evaluate and compare the trainability and generalization characteristics of multiple quantum circuit architectures. These experiments are designed to test how distinct architectural design choices—such as reuploading (QRU), variational entanglement (VQC), and static parameterization (PQC)—influence its performance on chaotic Mackey–Glass series forecasting task.

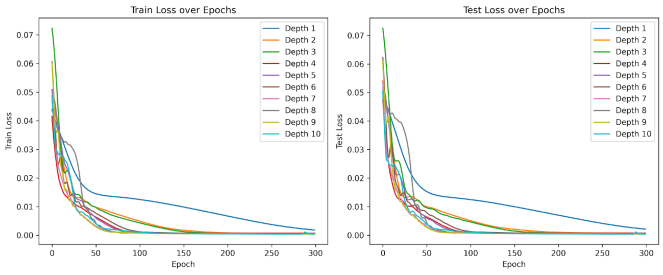
A. Influence of Circuit Depth

Circuit depth is a key hyperparameter in QRU models, directly affecting training behavior. We systematically assess how the number of re-uploading layers impacts performance. As shown in Figure 2a, loss convergence is optimal at depth 3; deeper circuits offer little improvement. Figure 2b shows that

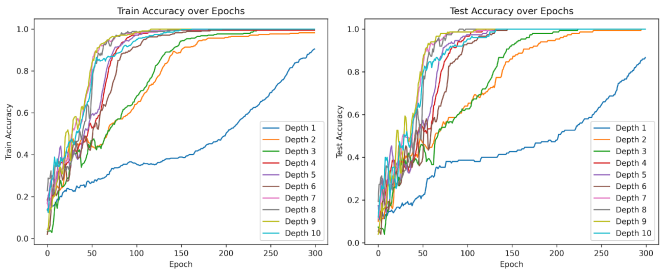
[2]The dataset can be accessed at: https://taiao.ai/datasets/coromandel-river-and-rain-gauge-time-series.en/

accuracy also saturates beyond depth 3–6, indicating sufficient expressivity at moderate depths for the chosen sequence length (3).

A depth-1 model, using single-pass encoding, underperforms due to limited expressivity, mirroring VQC/PQC strategies. These results question the assumption that deeper circuits always improve performance. With the same number of parameters, QRUs match or exceed VQC expressivity without excessive depth, avoiding issues like overparameterization or instability.



(a) Losses over epochs



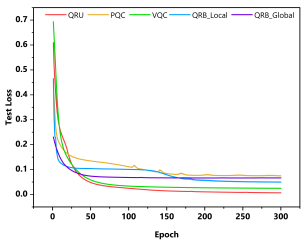
(b) Accuracies over epochs

Fig. 2: Performances comparison of a QRU for different depth: (a) Losses over epochs, (b) Accuracies over epochs.

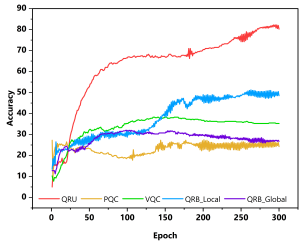
B. Experimental Evaluation of the models

The primary objective of this experiment is to evaluate the performance of three quantum circuits: QRU, PQC, and VQC on the chaotic Mackey–Glass time series forecasting task. This experiment aims to verify the hypothesis that the iterative reuploading in QRU would lead to enhanced expressivity and loss minimization, enabling it to capture complex temporal dynamics more effectively than PQC and VQC. By comparing these architectures, it is possible to assess the effects of data encoding strategies on loss convergence and gradient flow during training.

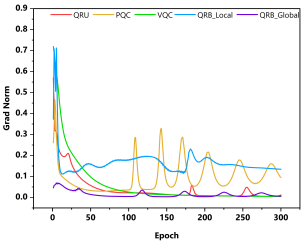
Figure 3 shows that QRU outperforms others in convergence speed, gradient stability, and accuracy—reflecting a smoother loss landscape due to its iterative reuploading and trainable encoding. PQC, lacking reuploading and entanglement, shows fast but suboptimal convergence and vanishing gradients, indicating a barren plateau. VQC improves over PQC via entanglement, enabling richer mappings, but suffers from unstable gradients and poorer generalization.



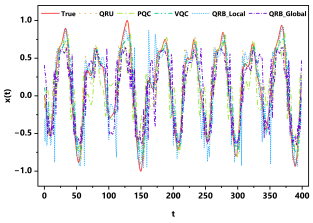
(a) Loss over epochs



(b) Accuracy over epochs



(c) Norm of gradients

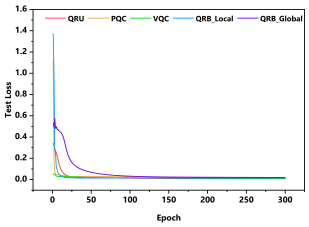


(d) Predictions on test set

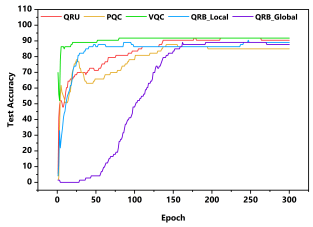
Fig. 3: Performances comparison between QRU, PQC, VQC, QRU-QRB-Local and QRU-QRB-Global: (a) Loss over epochs, (b) Accuracy over epochs, (c) Gradient norm evolution, (d) Prediction vs True values on test data.

These results align with prior studies on the expressivity–trainability trade-off in entangled circuits [17], reinforcing QRU's advantage. To assess architectural effects on spectral diversity and trainability, we compare QRU, VQC, and QRU–QRB–Local (entangled with a shared ancilla) on the Mackey–Glass forecasting task with chaotic short-term dynamics.

Further, VQC struggles to approximate the target function mainly due to the gradient vanishing as early as the 80th epoch, and hence the model struck at one of the local minima and couldn't get past it. This is attributed to low parameter count and the barren plateau effect.



(a) Loss over epochs



(b) Accuracy over epochs

Fig. 4: Performances comparison between a QRU and a QRU with QRB: (a) Loss over epochs, (b) Accuracy over epochs.

This experiment explores the trade-off between spectral expressivity and optimization stability for QRU and QRU–QRB–Global in a river water level prediction task. QRU–QRB–Global uses one ancilla qubit per re-uploading layer,

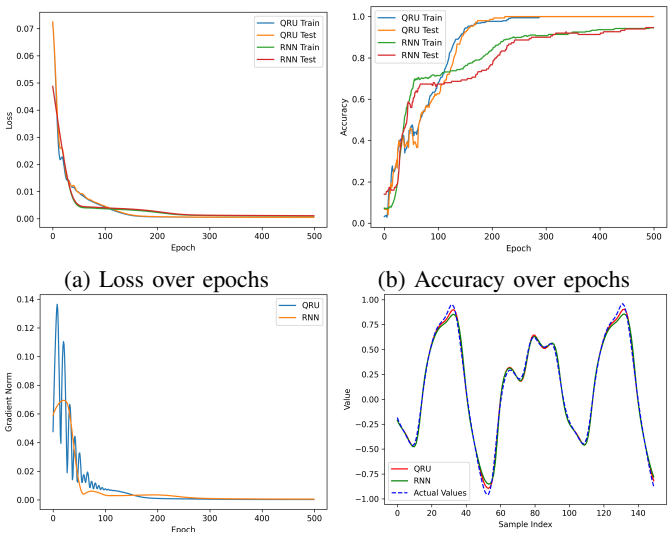
enabling independent residual transformations and combinatorial frequency mixing. While this increases expressivity, it also leads to greater circuit depth.

Figure 4 shows that QRU achieves stable convergence with consistent gradient flow, whereas QRU–QRB–Global shows erratic gradient fluctuations, likely due to non-convex interactions from multi-ancilla entanglement. Thus, QRU proves more stable and trainable despite lower spectral capacity.

To benchmark quantum vs. classical representation power, we compare QRU and an RNN on the Mackey–Glass forecasting task, with both models constrained to 27 trainable parameters. Total parameters $=$ depth $\times\, r_g \,\times$ n_input where the circuit depth was set to 3 and the sliding window size n was fixed at 3. Each QRU layer utilized parameterized single-qubit rotational gates R_X, R_Y, R_Z - hence $r_g = 3$ with input-dependent angles modulated by trainable weights.

The RNN architecture was implemented with a scalar hidden state updated through a recurrence relation $h = \tanh(a \cdot h + b \cdot x_j + c)$ where h is the hidden state (initialized to 0), x_j is the input at time step j, a, b, c are trainable parameters.

The performance of the models was analyzed based on four key metrics: loss, accuracy, gradient norm evolution, and prediction fidelity on the test set. Figure 5 presents the comparative results.



(a) Loss over epochs (b) Accuracy over epochs

(c) Norm of gradients (d) Predictions on test set

Fig. 5: Performances comparison between a QRU and a RNN: (a) Loss over epochs, (b) Accuracy over epochs, (c) Gradient norm evolution, (d) Prediction vs True values on test data.

The results show that QRU reaches near-perfect accuracy faster than the classical RNN. As seen in Figure 5b, QRU approaches 100% accuracy in 200 epochs, while RNN plateaus around 90%. Similarly, Figure 5a shows faster loss convergence for QRU. Both models start with high gradient activity (Figure 5c), then stabilize; QRU exhibits more oscillations, suggesting higher sensitivity to updates and potentially aiding faster learning.

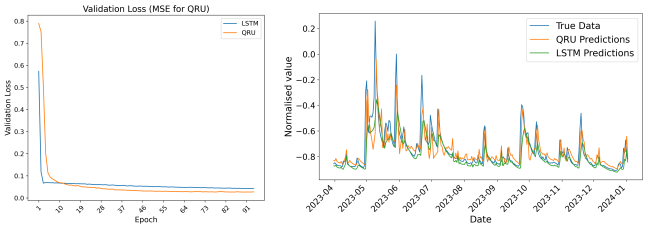


Fig. 6: Left: Validation loss for QRU vs. LSTM. Right: Test set predictions: QRU, LSTM, and actual river levels.

Figures 5d and 6 confirm good generalization in both models, with QRU predictions slightly closer to the ground truth. These results support the idea that quantum-enhanced models like QRU can outperform classical like RNNs or LSTM in time-series tasks with equal parameter counts.

Model Name	Dataset	Epoch	Training Loss	Accuracy
QRU	Mackey's Glass	300	0.00675	80%
PQC	Mackey's Glass	300	0.07446	24%
VQC (baseline)	Mackey's Glass	300	0.02483	35%
QRU–QRB–Local	Mackey's Glass	300	0.04924	49%
QRU–QRB–Global	Mackey's Glass	300	0.06649	27%
QRU	River Level	300	0.001781	93%
PQC	River Level	300	0.003226	87%
VQC (baseline)	River Level	300	0.002416	91%
QRU–QRB–Local	River Level	300	0.01532	89%
QRU–QRB–Global	River Level	300	0.01927	88%

TABLE II: Performance comparison of quantum circuit models on chaotic and real-world time series datasets. Styled to match IEEE tabular conventions.

We observe that all models achieve very similar results on the river water level dataset. We hypothesize that this is because the global trends of the signal are relatively easy to capture, while the fine-grained local oscillations are significantly more challenging to predict. This difficulty in modeling the small-scale variations seems to affect all models in a comparable way.

C. Quantum Processing Unit

We trained the QRU on river water level data, then performed predictions on the test set using both a simulation and a Quantum Processing Unit (QPU) from the IBM Quantum platform: `ibm_fez`.

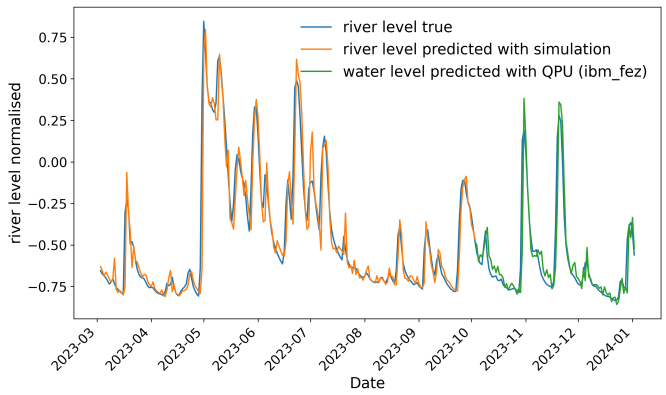
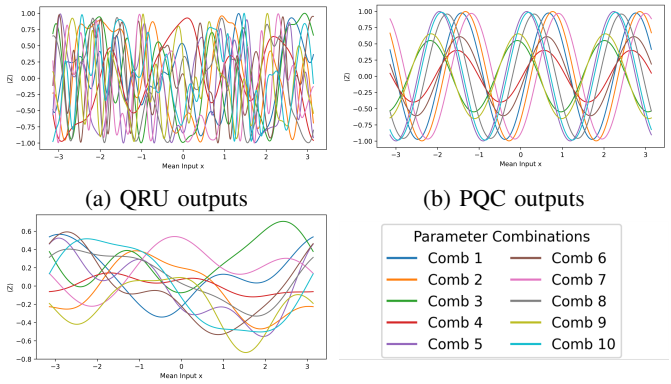


Fig. 7: River level prediction over one month using a QRU model trained in simulation. Blue: prediction run on simulator. Orange: same circuit executed on QPU.

While both capture general trends, hardware noise introduces slight amplitude distortions. We can see that the QRU, due to its low qubit count and shallow depth, is well suited for the NISQ era while still providing good performance.

VI. Discussion



(a) QRU outputs (b) PQC outputs

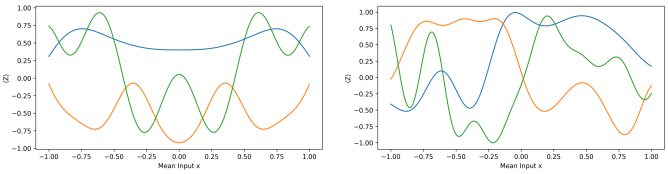
(c) VQC outputs (d) Legend

Fig. 8: Circuit outputs $h_\theta(x)$ for various combinations of trainable parameters across three quantum architectures: QRU, PQC and VQC.

Figure 8 shows the Fourier-decomposed circuit output $h_\theta(x)$, measured on the data qubit ($\langle Z \rangle$) for multiple random parameter sets θ, across PQC, VQC, and QRU circuits. PQC displays regular, periodic oscillations with simple patterns, indicating a limited set of accessible frequencies and phases. VQC yields more irregular outputs, with richer yet still constrained frequency combinations due to architectural limits.

In contrast, the QRU exhibits complex, non-linear oscillations, revealing access to a broader function space. This aligns with the Fourier representation of single-qubit outputs, where the accessible frequency set Ω depends on the rotation generators and circuit design. QRUs, via data reuploading, significantly expand Ω, improving functional expressivity—especially with greater depth (see Appendix A).

Figure 9 illustrates $h_\theta(x)$ for QRU at various $x \in [-\pi, \pi]$, across nine parameter sets and two encoding gate orders: $(R_x R_y R_x)$ and $(R_x R_y R_z)$, followed by measurement.



(a) $R_x - R_y - R_x$ (b) $R_x - R_y - R_z$

Fig. 9: Hypothesis functions $h_\theta(x)$ for 3 different parameter configurations across two circuit architectures. Each colored curve represents a randomly sampled trainable parameter set, showing the variability in expressivity.

Replacing the second R_x with R_z alters the generated frequencies and phase factors, enhancing expressivity, whereas R_x-R_y-R_x is limited to fit the even functions only. These plots exemplify how specific gate sequencing yield distinct hypothesis landscapes. The enhanced frequency variability also indicates the importance of designing optimal training configuration to avoid barren plateaus while ensuring robust generalization [18].

A. Fourier Analysis

Fourier analysis decomposes a given periodic function into an infinite sum of complex exponentials. In continuous form, a function $f(x)$ can be represented as

$$f(x) = \sum_{k=-\infty}^{\infty} C_k \, e^{ikx}, \tag{14}$$

where $C_k \in \mathbb{C}$ are the complex Fourier coefficients that capture the amplitude and phase of each frequency mode k. Each quantum circuit under study (PQC, VQC, QRU, etc.) produces, depending on the classical input $\vec{x}$ (or sometimes a scalar x), an output function $h_{\boldsymbol{\theta}}$ such as eq 3. Using a Fourier analysis of this, we write:

$$h_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{k \in \Omega} C_k(\boldsymbol{\theta}) \, e^{ik \cdot \mathbf{x}},$$

Where Ω is the set of frequencies accessible to the architecture, $C_k(\boldsymbol{\theta}) = |C_k(\boldsymbol{\theta})| e^{i\phi_k}$ is a complex coefficient, depends on trainable parameters set $\boldsymbol{\theta}$. When working with discrete samples $f_n = f(x_n)$ at points $x_n = n\,\Delta x$ for $n = 0, 1, \ldots, N-1$, we typically use the Discrete Fourier Transform (DFT), or in practice its efficient implementation known as the Fast Fourier Transform (FFT). For N samples:

$$\hat{F}_k = \sum_{n=0}^{N-1} f_n \, e^{-2\pi i \, \frac{kn}{N}}, \quad k = 0, 1, \ldots, N-1. \tag{15}$$

These discrete-frequency coefficients $\hat{F}_k$ are related to the continuous Fourier coefficients C_k through sampling relations

and the Nyquist criterion (to avoid aliasing) [19]. We used zero-padding [20] to enhance the resolution of the discrete spectrum.

1) Amplitude spectrum: Our goal is to investigate the expressivity of various quantum circuit architectures by analyzing the Fourier decomposition of their outputs with respect to input data.
We form sliding windows of size `n_input` over the input:

$$\{x_0, x_1, x_2\}, \quad \{x_1, x_2, x_3\}, \quad \dots$$

Each window is passed to the circuit with randomly initialized parameters in multiple runs. For each circuit and each random parameter set, we collect the outputs $\langle Z \rangle$ as a function of the *mean* of the group (to plot in one dimension). A function `compute_fft` takes the output signal $\{y_n\}$ (the circuit outputs) of size N_{groups}:

$$y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N_{\text{groups}}-1} \end{bmatrix},$$

and zero-pads, then, the FFT of this padded signal is computed, and the corresponding frequency grid is derived using `np.fft.fftfreq`. The amplitudes

$$\text{Amplitude}(k) = \frac{2}{N_{\text{groups}}} \left| \hat{F}_k \right|$$

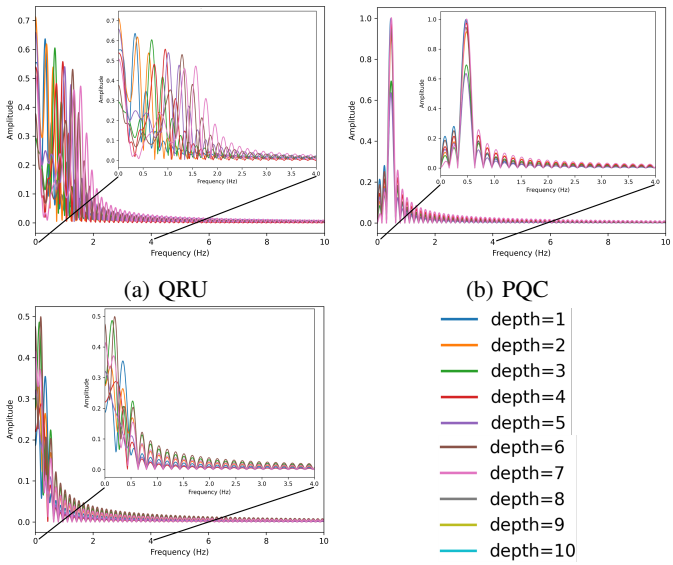
are then visualized.
The script generate figures for each circuit across different conditions:

a) Varying depth: Figure 10 shows the amplitudes spectrums for different depth for each quantum circuits.

Increasing the circuit depth can provide more opportunities for frequency mixing. `PQC` and `VQC` often exhibit fewer non-zero frequency coefficients: some modes remain completely absent. `QRU` circuits present a richer spectrum, including both low and moderate (sometimes high) frequencies. As depth grows, more frequency components become visible, although many remain very small in amplitude.

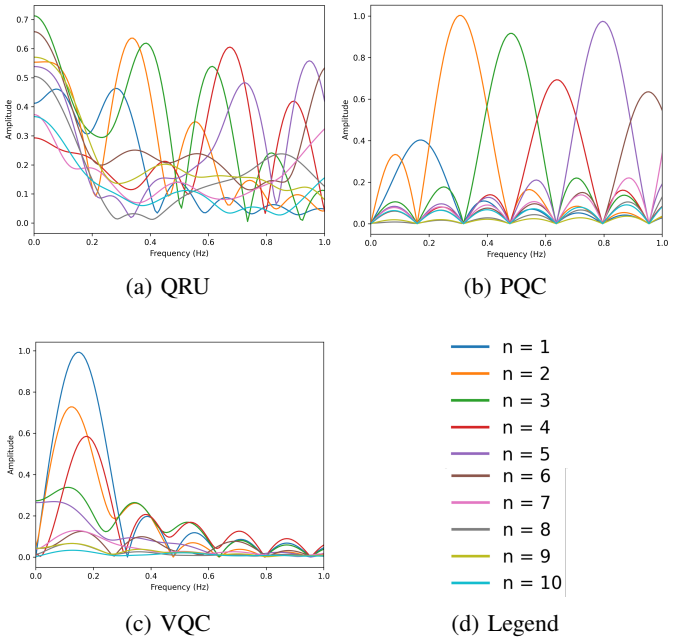
b) Varying the number of input data points: Figure 11 shows spectrums for different number os data taken in input by the three quantum circuits. When the input window `n_input` or the total sampling points `N` is small, several frequency modes simply cannot be excited because of insufficient or overly constrained input encoding. As `n_input` or `N` grows:

- The accessible frequencies spread out more densely.
- The amplitude of certain modes decreases (particularly high-frequency modes). This is due to fact that the more your increase the number of data, the more your increase the depth of the circuit, a phenomenon reminiscent of results discussed in literature. [7]



(a) QRU (b) PQC (c) VQC (d) Legend

Fig. 10: Amplitude spectra of Fourier coefficients for PQC, VQC, and QRU architectures as a function of reuploading depth.



(a) QRU (b) PQC (c) VQC (d) Legend

Fig. 11: Amplitude spectra of Fourier coefficients for PQC, VQC, and QRU architectures as a function of the number of inputs in the sliding window.

c) Exploring random parameter sets: Even with a fixed architecture, varying parameters yield different frequency spectra, as shown in Figure 12. This figure compares six quantum models, each tested with 10 random parameter

sets—identical across all circuits. The `QRU + Ent.` architecture adds parametric CRX gates from the data qubit to an ancillary one, but only the data qubit is measured, unlike `QRU-QRB-Local`.

Some models, like `PQC` and `VQC`, consistently show zero amplitude at certain frequencies, indicating intrinsic spectral limitations. In contrast, `QRU` variants (including `QRU-QRB-Local`, `QRU-QRB-Global`, and `QRU + Ent.`) activate all low-frequency modes. While some `QRU-QRB` configurations may miss certain frequencies, these appear accessible with better parameters—i.e., not inherently forbidden. Residual connections introduce interference that selectively filters frequency ranges, yielding periodic hypotheses. Overall, `QRU`-based architectures access richer frequency spectra, confirming that architecture defines the accessible modes, while parameters determine their activation.

B. Feature Mapping

This analysis evaluates the feature mapping and separation capabilities of five quantum architectures—VQC, PQC, QRU, QRU–QRB–Global, and QRU–QRB–Local—by projecting their outputs into the complex plane of Fourier coefficients. Using 1000 forward passes with random parameter initializations on the MG time series test data, we extract complex Fourier coefficients via FFT and visualize their real and imaginary parts.

VQC and QRU–QRB–Local yield a uniform spread, indicating good input separation. PQC produces tightly clustered features, suggesting limited separation despite internal correlation. QRU significantly improves upon PQC by mapping inputs more distinctly within clusters. QRU–QRB–Global shows the most restricted feature spread due to limited mapping capacity. While deeper reuploading layers can improve separation, they also increase circuit complexity.

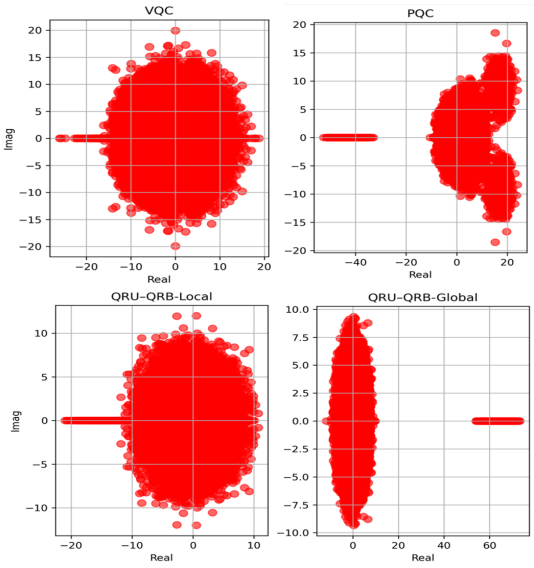
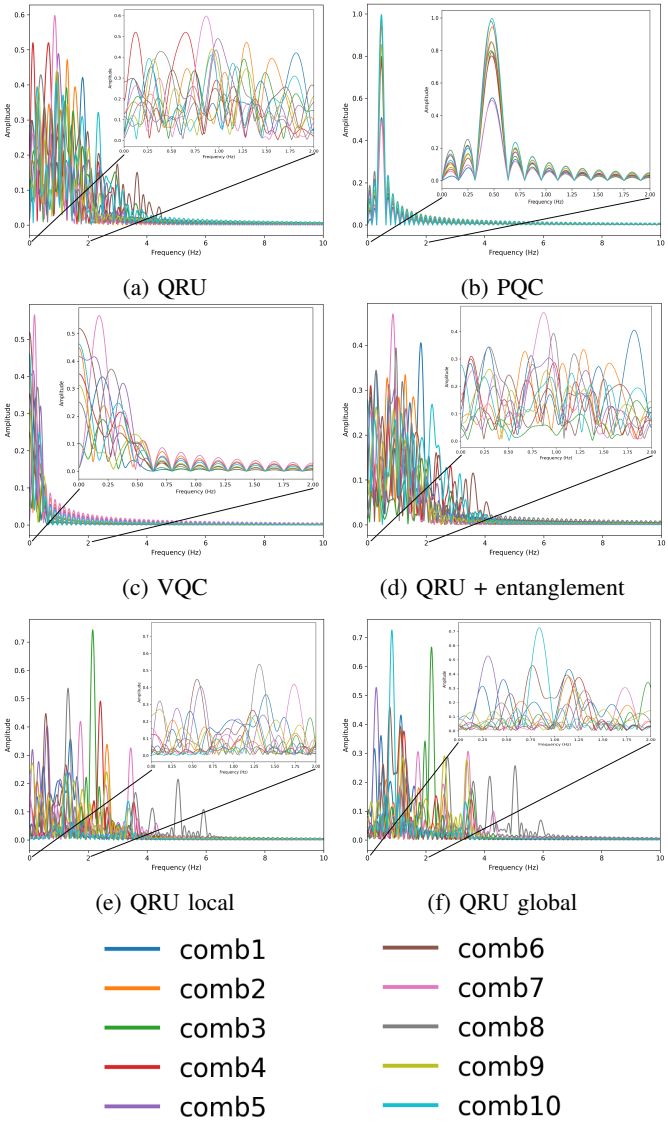


Fig. 13: Complex plane representation of FFT coefficients for VQC, PQC, QRU, QRU–QRB–Global, and QRU–QRB–Local architectures.



(a) QRU (b) PQC (c) VQC (d) QRU + entanglement (e) QRU local (f) QRU global (g) Legend

Fig. 12: Amplitude spectra of Fourier coefficients for various quantum circuit architectures, with randomly sampled parameters.

C. Absorption Witness

The QRU does not entirely prevent the vanishing gradient phenomenon; this will depend both on its architecture and on the data. In order to identify the architecture of the QRU that minimizes this barren plateau effect as much as possible, we rely on the Absorption Witness (AW) metric introduced by Barthe et al. [7].

AW concretely serves to quantify the ability of a reuploading model to absorb or efficiently integrate input data into its parameters, such that it does not overly rely on external variations like the data itself. More precisely, it measures how insensitive the circuit is to changes induced by the data during

training, by comparing the variance of gradients with and without data.

In practice, a low value of the absorption witness indicates that the circuit can easily "absorb" the data into its parameters without causing significant variation in the output, which facilitates training by avoiding the so-called vanishing gradients phenomenon. This guides circuit design to improve trainability, notably by avoiding structures that might cause gradient disappearance or lead to difficult optimization behaviors.

To empirically evaluate the practical relevance of the absorption witness, the following numerical procedure is implemented. To study the trainability of the QRU, we analyze the norm of the gradient $\nabla_\theta h_\theta(x) \in \mathbb{R}^m$, where $h_\theta(x)$ denotes the output of the circuit and m the total number of parameters. For each random sample $\theta \sim \mathcal{U}([-\pi, \pi]^m)$, we compute:

$$\|\nabla_\theta h_\theta(x)\| = \sqrt{\sum_{j=1}^{m} \left(\frac{\partial h_\theta(x)}{\partial \theta_j} \right)^2} \tag{16}$$

The procedure is repeated over N independent random draws (in our case $N = 100$), and then the variance of the gradient norm is calculated over the parameter space $\mathrm{Var}_\theta \left(\|\nabla_\theta h_\theta(x)\| \right)$. This variance reflects how the sensitivity to the input x varies depending on the initialization of the parameters. Following the definition of the absorption witness $B^{(2)}_{R,j}$ introduced in [7], we expect that if the data x can be fully absorbed through a simple parameter shift $\theta \mapsto \theta + \delta$, then adding more data re-uploading steps should not significantly affect the gradient variance. In such cases, the gradient distribution remains unchanged, and we observe $AW_{\text{num}} \approx 0$.

The empirical absorption witness is then defined as:

$$AW_{\text{num}} = \left| \mathrm{Var}_\theta \left(\|\nabla^{(\text{QRU})}_\theta h(x)\| \right) - \mathrm{Var}_\theta \left(\|\nabla^{(\text{PQC})}_\theta h(x)\| \right) \right|$$

since we are here analyzing the difference in behavior between the layer with input data x and its version without data, when averaging over several possible parameter combinations.

We then test the QRU model on our three data types and evaluate the AW metric for different depths of the QRU in order to understand which one appears to have the best trainability, thus best mitigating the vanishing gradient.

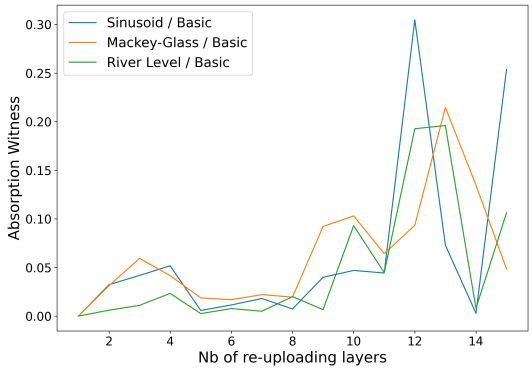


Fig. 14: Numerical Absorption Witness values for the QRU evaluated on three different datasets as a function of the number of re-uploading layers.

As shown in Figure 14, the absorption witness provides a clear picture of the trainability of QRUs with increasing depth. In the first region, from depth 1 up to around depth 11, the values of the witness remain small (but not strictly zero). This regime is favorable, as it indicates that re-uploading the data is not overwhelming the circuit: the input is effectively absorbed and integrated with the parameters, and the gradients are still exploitable for learning.

However, beyond depth 11, we observe that the witness increases significantly. This reveals that the circuit becomes too sensitive to repeated data injection, which in turn causes the gradients to vanish. This explains why performance saturates and why trainability deteriorates after a certain number of re-uploadings: the circuit enters a barren plateau regime where the gradients are progressively wiped out.

Interestingly, at depth 14, the witness value drops again to nearly zero. This does not indicate improved trainability, but rather that the circuit states have become effectively random due to excessive re-uploading, consistent with the findings of [21]. In this regime, the circuit loses meaningful structure, and training becomes ineffective.

Overall, these results confirm the existence of a sweet spot in QRU depth: shallow to moderately deep architectures retain good trainability, while excessive depth leads to barren plateaus or randomization effects that undermine learning.

Quantum State Diversity

The quantum state diversity analysis quantitatively evaluates the expressivity of different quantum circuit architectures by computing the Kullback-Leibler (KL) divergence between the state fidelity distribution and that of the ideal Haar-distributed state ensemble. In this analysis, pairs of quantum states are generated by randomly sampling pairs of trainable parameters for each architecture (for 1000 times), including QRU, PQC, VQC, and QRU-QRB-Local and QRU-QRB-Global. The KL divergence is defined as

$$D_{\text{KL}}(P_F \parallel P_{\text{Haar}}) = \sum_j P_F(j) \log \left(\frac{P_F(j)}{P_{\text{Haar}}(j)} \right), \tag{17}$$

where the analytical Haar fidelity distribution is given by

$$p_{\text{Haar}}(F) = (N-1)(1-F)^{N-2}, \tag{18}$$

with N denoting the dimension of the Hilbert space. A lower KL divergence value indicates higher expressivity by signifying a broader exploration of the quantum state space. The computed KL divergence values reveal that the QRU architecture exhibits enhanced expressivity with a divergence of 0.03421, outperforming both PQC (0.04421) and the QRB-Local (0.20022). Notably, the QRB-Global achieves exceptionally low KL divergence of 0.00001, thereby demonstrating that its entangling strategy significantly broadens the distribution of accessible quantum states. These findings underscore that the design choices in quantum circuits critically determine the circuit's capability to generate diverse quantum state representations for different initial parameterizations, which is directly connected to the generalization of the quantum circuit [22], [23], [24].

The failure case of QRU-QRB architectures, the underperformance of QRU-QRB compared to the baseline QRU is attributed to a fundamental trade-off between increased expressivity and decreased trainability. While QRBs are designed to enrich the model's frequency spectrum and enhance expressivity through residual connections, this added architectural complexity introduces significant optimization instability. Experiments demonstrate that the QRU-QRB-Global variant, which uses an ancilla qubit for each layer, suffers from violent gradient fluctuations during training. This behavior is attributed to non-convex interactions from multi-ancilla entanglement, creating a difficult optimization landscape.

VII. Conclusion and Future Work

This work systematically evaluated the Quantum Reupload Unit (QRU)—a hardware-efficient, single-qubit architecture tailored for time series forecasting. A rigorous theoretical and empirical study demonstrated that iterative data reuploading paired with trainable parameters in shallow circuits enables rich function approximation, competitive expressivity, and superior gradient behavior compared to multi-qubit variational quantum circuits (VQCs) and classical recurrent models. Comprehensive Fourier analysis confirmed that QRUs inherently activate a denser and more diverse set of frequency components. The absorption witness metric validated that repeated data encoding restructures the loss landscape and enhances gradient flow not reproducible by mathematical reparameterization techniques. Despite the constraints of NISQ hardware, QRUs exhibit strong predictive performance on both chaotic synthetic and real-world datasets. The trade-off between expressivity and qubit overhead in QRB-augmented variants provides valuable design insights for designing future quantum circuit architectures. This study further establishes QRUs as a scalable model for learning time-dependent sequences and presents a reproducible, interpretable framework for evaluating quantum circuit expressivity. Future extensions of this work will focus on developing frameworks for multivariate sequential data forecasting, adaptive reuploading schemes, and hybrid QLSTM-based forecasting approaches.

Appendix A
Spectral Resolution of Circuits

The expressivity of a quantum circuit can be characterized by the set of frequency components appearing in the Fourier expansion of its output.

In a QRU, each layer applies a unitary $U(x) = e^{iHx}$, with $H = \sum_k w_k |h_k\rangle\langle h_k|$. Given an initial state $|\psi\rangle = \sum_j \psi_j |h_j\rangle$ and observable O, the output is

$$f(x;\theta) = \sum_{j,l} \psi_j^* \psi_l \, \langle h_j|O|h_l\rangle \, e^{i(w_l - w_j)x}.$$

Defining $c_{jl}(\theta) = \psi_j^* \psi_l \, \langle h_j|O|h_l\rangle$, the accessible frequency set is

$$\Omega_{\text{QRU}} = \{w_l - w_j \mid j, l \in [1, d]\}.$$

Concatenating L layers yields frequencies of the form $\sum_i (w_{l_i} - w_{j_i})$, scaling linearly as $\mathcal{O}(L)$.

In the QRU–QRB–Local model, a shared ancilla is entangled via a CRX gate. Let $V_0 = U_{\text{post}} \, R_Y(\theta_1 x)$ and $V_1 = U_{\text{post}} \, R_X(\alpha\theta_4) R_Y(\theta_1 x)$. After applying a Hadamard and projecting the ancilla onto $|0\rangle$, the effective state becomes

$$|\psi_{\text{eff}}\rangle = \tfrac{1}{2}(V_0|0\rangle + V_1|0\rangle),$$

and the output is

$$h(x) = \tfrac{1}{4}\left[\langle V_0|Z|V_0\rangle + \langle V_1|Z|V_1\rangle + 2\,\Re\langle V_0|Z|V_1\rangle\right].$$

The cross terms introduce mixed frequencies $e^{i(\omega_{RY} \pm \omega_{RX})x}$, and with L layers, the spectrum scales as $\mathcal{O}(L^2)$.

In the QRU–QRB–Global model, each layer has its own ancilla, allowing independent entanglement. A single layer accesses

$$\Omega_{\text{QRB–Global}} = \{0\} \cup \{\pm w_k\} \cup \{w_k - w_j\}.$$

Over L layers, the number of combinations grows exponentially as $\mathcal{O}(2^L)$, maximizing frequency diversity at the cost of hardware.

RU uses a single qubit and requires $L \sim N_{\text{freq}}$ to match complex spectra. QRU–QRB–Local introduces interference via shared ancilla, suitable for correlated signals but less diverse. QRU–QRB–Global reaches maximal spectral resolution with 2^L terms, requiring one ancilla per layer and allowing $L \sim \log_2(N_{\text{freq}})$ for rich targets.

References

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[2] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, "Parameterized quantum circuits as machine learning models," *Quantum science and technology*, vol. 4, no. 4, p. 043001, 2019.

[3] S. Y.-C. Chen, C.-H. H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, "Variational quantum circuits for deep reinforcement learning," *IEEE access*, vol. 8, pp. 141 007–141 024, 2020.

[4] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature communications*, vol. 9, no. 1, p. 4812, 2018.

[5] C. Chatfield, *Time-series forecasting*. Chapman and Hall/CRC, 2000.

[6] N. L. Wach, M. S. Rudolph, F. Jendrzejewski, and S. Schmitt, "Data re-uploading with a single qudit," *Quantum Machine Intelligence*, vol. 5, no. 2, p. 36, 2023.

[7] A. Barthe and A. Pérez-Salinas, "Gradients and frequency profiles of quantum re-uploading models," *Quantum*, vol. 8, p. 1523, 2024.

[8] J. Wen, Z. Huang, D. Cai, and L. Qian, "Enhancing the expressivity of quantum neural networks with residual connections," *Communications Physics*, vol. 7, no. 1, p. 220, 2024.

[9] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

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

[11] V. Derkach and M. Malamud, "The extension theory of hermitian operators and the moment problem," *Journal of mathematical sciences*, vol. 73, no. 2, pp. 141–242, 1995.

[12] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, "Quantum embeddings for machine learning," *arXiv preprint arXiv:2001.03622*, 2020.

[13] M. T. Augustine, "A survey on universal approximation theorems," *arXiv preprint arXiv:2407.12895*, 2024.

[14] M. Schuld, "Supervised quantum machine learning models are kernel methods," 2021. [Online]. Available: https://arxiv.org/abs/2101.11020

[15] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[16] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, "A review of barren plateaus in variational quantum computing," *arXiv preprint arXiv:2405.00781*, 2024.

[17] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, "Connecting ansatz expressibility to gradient magnitudes and barren plateaus," *PRX Quantum*, vol. 3, no. 1, p. 010313, 2022.

[18] L. Cassé, B. Pfahringer, A. Bifet, and F. Magniette, "Optimizing hyperparameters for quantum data re-uploaders in calorimetric particle identification," *arXiv preprint arXiv:2412.12397*, 2024.

[19] C. Desoer and Y.-T. Wang, "On the generalized nyquist stability criterion," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 187–196, 1980.

[20] Y. Du, G. Feng, H. Li, and S. Zhou, "Accurate carrier-removal technique based on zero padding in fourier transform method for carrier interferogram analysis," *Optik*, vol. 125, no. 3, pp. 1056–1061, 2014.

[21] X. Wang, H.-X. Tao, and R.-B. Wu, "Predictive performance of deep quantum data re-uploading models," 2025. [Online]. Available: https://arxiv.org/abs/2505.20337

[22] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019.

[23] G. R. Schleder, A. C. Padilha, and C. M. Acosta, "On the use of the haar measure for computing circuit expressibility," *arXiv preprint arXiv:2007.14421*, 2020.

[24] Z. Wang, Y. Du, M.-H. Hsieh, and D. Tao, "Enhancing the expressivity of quantum neural networks with residual connections," *npj Quantum Information*, vol. 7, no. 1, pp. 1–10, 2021.