FOURIER-GEOMETRIC CIRCUIT DESIGN FOR GATE AND ENTANGLEMENT PLACEMENT IN QUANTUM NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

The output of a parameterized quantum circuit (PQC) can be expressed as a finite Fourier series whose accessible frequencies are fixed by the data-encoding gates. While the encoder determines which frequencies can appear, the corresponding Fourier coefficients depend on how the trainable and entangling gates are arranged. Although existing studies provide metrics for characterizing how gate structure affects Fourier coefficients, they do not translate these analyses into an explicit design criterion specifying how gates should be arranged to make a target coefficient reachable. In this paper, we provide such a criterion. Using the adjoint action of the encoding generator, we decompose operator space into two-dimensional invariant planes indexed by frequency and show that each Fourier coefficient is exactly a sum of bilinear projections of the effective state and observable onto the planes at that frequency. Because, for Pauli encodings, high-frequency planes are spanned by mixed multi-qubit Pauli strings, a target coefficient can contribute to the output only when local rotations and entangling layers are arranged so that both the effective state and observable acquire support on one of its planes. For Pauli readouts and commuting two-qubit entanglers, this yields a circuit design rule that specifies, for a given interaction graph, a placement of local rotations and entangling layers that makes a target Fourier coefficient reachable. Using a single encoding layer, we validate the proposed design through a placement ablation, regression tasks from PDEBench and physics-informed Maxwell field modeling, and further demonstrate its robustness to moderate simulated gate noise and damping.

1 INTRODUCTION

Parameterized quantum circuits (PQCs) have emerged as a standard framework for near-term quantum machine learning (QML) due to their compatibility with shallow, low-qubit implementations (Benedetti et al., 2019). However, their design, particularly the arrangement of encoding, local and entangling gates, remains closer to an art than a mathematically grounded discipline because circuits are assembled from a small catalog of templates, chosen with little reference to the function the PQC is meant to represent. This leaves a basic question unanswered. When a target needs particular frequencies, as the oscillatory fields of scientific machine learning do, where should the local and entangling gates be placed?

The closest answer comes from studies of the representational capacity of PQCs. Schuld et al. (2021) showed that the output of a PQC is a finite Fourier series whose accessible frequencies are fixed by the data-encoding gates, and repeating the encoding, either across multiple qubits or through data re-uploading (Pérez-Salinas et al., 2020), enlarges that set in a way that follows directly from the encoder. Therefore, the accessible frequency set is fixed by the encoding gates before training. What is not settled is whether the coefficient at an available frequency is actually nonzero. Coefficients are known to depend on the trainable blocks and the observable, and a growing body of work has since computed them, characterized their statistics, and traced their consequences for generalization and trainability (Casas & Cervera-Lierta, 2023; Barthe & Pérez-Salinas, 2024; Peters & Schuld, 2023; Wiedmann et al., 2025; Campbell et al., 2026; Ragone et al., 2024). However, *these analyses begin from an already chosen circuit.* They tell us which coefficients a particular arrangement of gates

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

produces, but not how that arrangement should be chosen. No existing result specifies how to place local and entangling gates so that a desired Fourier coefficient becomes reachable.

An explicit placement rule is hard to obtain by following the state through the circuit, where a Fourier coefficient appears only when the state meets the observable at the end, so it belongs to the whole circuit rather than to any single gate. Seen through operators, it can be traced. We show that under the adjoint action of the encoding generator, operator space separates into two-dimensional planes, each carrying exactly one frequency, with the input rotating each plane at its own rate. This lets us write the coefficient at a frequency exactly, as the inner product of the effective state and observable with the corresponding planes. We find that high-frequency planes involve multi-qubit operators, so a local measurement cannot reach them unless entangling gates first spread it across qubits. Whether that happens depends on where those gates sit, and the same gates in a different order can leave a Fourier component present in or absent from the output. We exploit this to derive a placement rule for Pauli encodings and readouts with commuting two-qubit entanglers. The rule makes the frequencies a target needs reachable, which is necessary for learning them but does not guarantee it, so we test that link empirically.

We highlight the contributions as follows:
(i) An exact expression for each Fourier coefficient. Using the adjoint action of the encoding generator, we express each coefficient as a sum of bilinear projections of the effective state and observable onto the two-dimensional invariant planes associated with that frequency.
(ii) A placement rule for trainable gates and entanglement. For Pauli encodings and readouts, commuting two-qubit entanglers and a given interaction graph, we give a placement of local rotations and entangling gates that suffices for a target Fourier coefficient to be nonzero at almost every parameter setting, identify which of its conditions are also necessary, and show that each such frequency can reach unit amplitude on its own and that reordering the same gates can suppress it. The rule builds the circuit in linear time.
(iii) Controlled validation. We evaluate the resulting design through parameter-matched comparisons with circuit templates over multiple seeds, a placement ablation that moves the same gates, supervised PDE regression and physics-informed Maxwell modeling, together with damping, gate-noise and finite-shot noise studies.

2 Prior Work

Analyses of a Given Circuit. Schuld et al. (2021) established the finite Fourier representation and its encoder-determined frequency set, Pérez-Salinas et al. (2020) showed that data re-uploading enlarges this set. Casas & Cervera-Lierta (2023) move closer to circuit design by showing that parallel layouts can represent multidimensional series of higher degree than line layouts. Their trainable blocks, however, are treated as generic unitaries. Later works similarly analyze predefined circuits by computing exact spectra and structurally vanishing coefficients (Wiedmann et al., 2025; Nemkov et al., 2023), comparing coefficient statistics across circuit families (Campbell et al., 2026; Mhiri et al., 2025), and studying the frequency profiles and generalization of re-uploading models (Barthe & Pérez-Salinas, 2024; Peters & Schuld, 2023). These works characterize what a given circuit can represent, but do not specify which gate, acting on which qubit and at which position, determines a particular Fourier coefficient. We derive this relationship and use it to construct trainable blocks whose reachable coefficients are determined by design.
Lie-Algebraic and Symmetry-Based Design. The dynamical Lie algebra (DLA) generated by a circuit characterizes its expressivity and trainability, but the circuit is represented through its generating algebra rather than the placement of individual gates. Ragone et al. (2024) explicitly leave the arrangement of gates within a layer to future work (Larocca et al., 2022a; Fontana et al., 2024). Symmetry-aware constructions prescribe circuit structure from algebraic constraints (Larocca et al., 2022b; Meyer et al., 2023; Nguyen et al., 2024), but likewise determine suitable gate sets without specifying their ordering. Consequently, circuits with the same encoding, interaction graph, and gate family may have the same accessible frequencies and generating algebra while differing in whether a spectral component reaches the measured observable. Our work addresses this finer dependence with a placement of local and entangling gates that makes a target Fourier coefficient reachable.
Circuit Templates and Applications. PQCs are commonly built from heuristic circuit templates (Sim et al., 2019; Kandala et al., 2017), including those used for differential equations (Kyriienko et al., 2021; Trahan et al., 2024; Sedykh et al., 2024; Berger et al., 2025; Panichi et al., 2026; Farea

et al., 2025; Xiao et al., 2024; Chen et al., 2026), where oscillatory solutions make spectral representation particularly relevant. Increasing entanglement alone does not necessarily improve performance, and Trahan et al. (2024) found strongly entangled layers suboptimal for their model. This motivates studying not only how much entanglement is used, but where the gates are placed so that the required spectral components can contribute to the output.

3 Fourier-Geometric Circuit Design

Background. A PQC implements a unitary transformation $U(\mathbf{x}, \boldsymbol{\Theta})$ acting on an initial state $|\psi\rangle^{\otimes N}$, where $\mathbf{x} \in \mathbb{R}^d$, $\boldsymbol{\Theta} = (\boldsymbol{\theta}^{(L)}, \boldsymbol{\theta}^{(R)}) \in \mathbb{R}^p$, and N are the input data, trainable parameters, and number of qubits, respectively. We consider an architecture consisting of a data-encoding gate $S(x)$ and trainable blocks W_L, W_R, each composed of parameterized and entangling gates, such that $U(\mathbf{x}, \boldsymbol{\Theta}) = W_L(\boldsymbol{\theta}^{(L)})\, S(x)\, W_R(\boldsymbol{\theta}^{(R)})$. With the density matrix $\rho_0 = (|\psi\rangle\langle\psi|)^{\otimes N}$ and an observable M, the model output is $f(x, \Theta) = \mathrm{Tr}(\rho_0 U^\dagger M U)$. Both ρ_0 and M belong to the Hilbert–Schmidt operator space $\mathcal{B}(\mathcal{H})$ with inner product $\langle A, B\rangle = 2^{-N}\,\mathrm{Tr}(A^\dagger B)$ (Mielnik, 1968), which is spanned by the Pauli strings $P_k \in \{I, X, Y, Z\}^{\otimes N}$. The parameterized gates $R_{P_k}(\theta) = e^{-i\theta P_k/2}$ act on operators through the adjoint map $\mathrm{ad}_{P_k}(A) = [P_k, A]$, since $U^\dagger M U = e^{\mathrm{ad}_{iA}}(M)$ for $U = e^{-iA}$, so gates rotate the Pauli-string components of the state and observable within $\mathcal{B}(\mathcal{H})$. For an encoding gate $S(x) = e^{-ixG/2}$, the adjoint action of the encoding generator G decomposes the operator space into invariant subspaces associated with differences between its eigenvalues. Consequently, the PQC output admits a finite Fourier representation whose accessible frequencies are determined by the spectrum of the encoding generator (Schuld et al., 2021). The trainable blocks W_L and W_R, in contrast, determine how the state and observable project onto these frequency subspaces, and hence the Fourier coefficients. However, *an explicit circuit-level relationship between the placement of trainable and entangling gates and the reachability of a target Fourier coefficient remains unclear.* We establish this relationship and use it to derive a circuit design rule for making target coefficients reachable. The decomposition underlying it holds for any encoding generator, whereas the design rule assumes a specific encoder, readout and entangler, which we state when we derive it.

Closed Invariant Subspace. For any encoding generator G, conjugation by $S(x) = e^{-\frac{ix}{2}G}$ expands into nested commutators with G, which produce the trigonometric terms of the Fourier series, as the following lemma makes precise.

Lemma 1 (Hadamard's Lemma). *Based on the Baker–Campbell–Hausdorff formula (Chevalley, 2018), for a generator G, an operator $\mathcal{O} \in \mathcal{B}(\mathcal{H})$ and a scalar t, $e^{tG}\mathcal{O}e^{-tG} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,\mathrm{ad}_G^k(\mathcal{O})$.*

While Lemma 1 gives an infinite series, the series sums to trigonometric functions whenever two operators $\mathcal{X}_\omega, \mathcal{Y}_\omega \in \mathcal{B}(\mathcal{H})$ span a subspace closed under the adjoint map of G[1], as the following proposition states.

Proposition 1 (Closed Invariant Subspace). *Let G be a generator for encoder $S(x)$ and $\mathcal{X}_\omega, \mathcal{Y}_\omega \in \mathcal{B}(\mathcal{H})$ be operators such that $\mathrm{ad}_G(\cdot) = [G, \cdot]$ closes on $\mathrm{span}\{\mathcal{X}_\omega, \mathcal{Y}_\omega\}$ in the sense that $\exists\, \omega \in \mathbb{R}$ such that $[G, \mathcal{X}_\omega] = -2i\omega\, \mathcal{Y}_\omega$, and $[G, \mathcal{Y}_\omega] = 2i\omega\, \mathcal{X}_\omega$. Then, $\forall x \in \mathbb{R}$,*

$$S(x)^\dagger \mathcal{X}_\omega S(x) = \cos(\omega x)\mathcal{X}_\omega + \sin(\omega x)\mathcal{Y}_\omega, \quad S(x)^\dagger \mathcal{Y}_\omega S(x) = \cos(\omega x)\mathcal{Y}_\omega - \sin(\omega x)\mathcal{X}_\omega. \quad (1)$$

Proposition 1 solves the evolution of a closed pair exactly as a rotation by the angle ωx within its plane,[2] so ω is a *frequency* of the model output in the strict trigonometric sense, fixed by the algebraic structure of the generator. We note that Proposition 1 holds for any G. We now describe *every* closed pair for single-qubit Pauli encodings, which single-qubit Clifford gates map to $G = \sum_{n=1}^{N} \hat{X}_n$, where $\hat{(\cdot)}_n \triangleq I^{\otimes(n-1)} \otimes (\cdot) \otimes I^{\otimes(N-n)}$ acts on qubit n.

[1] Detailed proof of Lemma 1 is provided in Appendix D.1.

[2] Detailed proof of Proposition 1 is provided in the Appendix D.2.

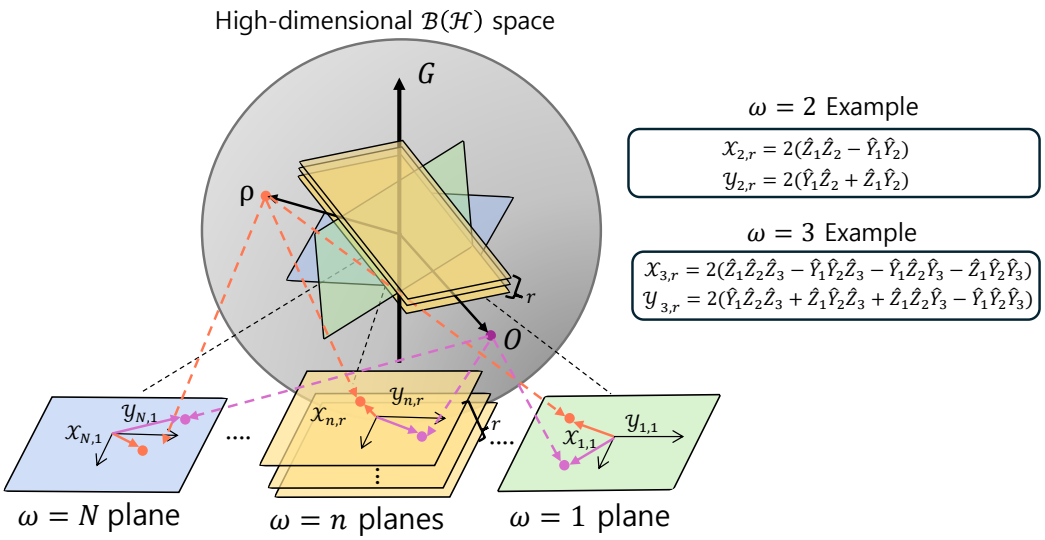


Figure 1: **Invariant subspace projection.** The operator space $\mathcal{B}(\mathcal{H})$ splits into invariant planes, each carrying one frequency, and the coefficient of that frequency is set by the Hilbert–Schmidt inner products of ρ and O with $\mathcal{X}_{\omega,r}$ and $\mathcal{Y}_{\omega,r}$.

Proposition 2 (Complete Harmonic Decomposition)**.** *Let $c = (S_+, S_-, T)$ be pairwise disjoint subsets of $\{1, \ldots, N\}$ and set*

$$E_c = \prod_{j\in S_+} E_j^+ \prod_{j\in S_-} E_j^- \prod_{j\in T} \hat{X}_j, \qquad \omega(c) = |S_+| - |S_-|, \qquad m(c) = |S_+| + |S_-|. \tag{2}$$

Then $\mathrm{ad}_G(E_c) = -2\omega(c)E_c$, and the 4^N operators $\{E_c\}$ form an orthogonal basis of $\mathcal{B}(\mathcal{H})$. For $\omega(c) > 0$ the Hermitian operators $\mathcal{X}_c = E_c + E_c^\dagger$ and $\mathcal{Y}_c = \frac{1}{i}(E_c - E_c^\dagger)$ satisfy $[G, \mathcal{X}_c] = -2i\omega\,\mathcal{Y}_c$ and $[G, \mathcal{Y}_c] = 2i\omega\,\mathcal{X}_c$, with $\|\mathcal{X}_c\|^2 = \|\mathcal{Y}_c\|^2 = 2^{m(c)+1}$.

Hence $\mathcal{B}(\mathcal{H}) = \bigoplus_{\omega=-N}^{N} \mathcal{B}_\omega$, so every observable decomposes over these planes and no frequency component is left unaccounted for. Since $\left(E_j^\pm\right)^2 = 0$, each qubit contributes at most one factor, giving $|\omega| \le m(c) \le N$ and recovering the qubit-count bound on accessible frequencies. We write d_ω for the number of planes at frequency ω and index them by r; the count is given in Appendix E.1. The planes with $S_- = T = \emptyset$ saturate $|\omega| = m(c)$ and are the ones a design must reach to activate the highest frequencies. Higher frequencies therefore require mixed multi-qubit Pauli strings rather than single Pauli operators. Visual intuition is given in Fig. 1[3].

Fourier Series Under Closed Invariant Subspaces. We now show how the Fourier series arises from the closed pairs $\{\mathcal{X}_{\omega,r}, \mathcal{Y}_{\omega,r}\}$.For $U(x) = W_L S(x) W_R$, the identity $\langle\psi|B|\psi\rangle = \mathrm{Tr}(|\psi\rangle\langle\psi|B)$ gives

$$f(x) = \langle\psi|U(x)^\dagger M U(x)|\psi\rangle = \mathrm{Tr}(\underbrace{W_R\rho_0 W_R^\dagger}_{\rho} S(x)^\dagger \underbrace{W_L^\dagger M W_L}_{O} S(x)) = \mathrm{Tr}(\rho S(x)^\dagger O S(x)), \tag{3}$$

where $\rho_0 = |\psi\rangle\langle\psi|$ is the initial state, M the measured observable and O the effective observable. We decompose both ρ and O over all closed pairs $\{(\mathcal{X}_{\omega,r}, \mathcal{Y}_{\omega,r})\}$. By Proposition 1, each pair $(\mathcal{X}_{\omega,r}, \mathcal{Y}_{\omega,r})$ spans a two-dimensional subspace invariant under the adjoint action of G. Conjugation by $S(x)$ therefore rotates each plane by the angle ωx, which yields the Fourier series of the following proposition.

[3]Detailed proof of Proposition 2 is provided in Appendix D.3.

Proposition 3 (Projection onto Invariant Subspaces). *Let $\{(\mathcal{X}_{\omega,r}, \mathcal{Y}_{\omega,r})\}_{r=1}^{d_\omega}$ be the invariant pairs, and let $m_{\omega,r} = |S_+^{(r)}| + |S_-^{(r)}|$ denote the number of active qubits of the r-th pair. These operators are mutually orthogonal with $\langle \mathcal{X}_{\omega,r}, \mathcal{X}_{\omega,r}\rangle = \langle \mathcal{Y}_{\omega,r}, \mathcal{Y}_{\omega,r}\rangle = 2^{m_{\omega,r}+1}$, and together with the commutant $\mathcal{B}_0$ they span $\mathcal{B}(\mathcal{H})$. Consequently, for* every *state ρ and* every *observable O, the output decomposes into a constant term and contributions at frequencies $\omega \le N$:*

$$f(x) = \mathrm{Tr}\big(\rho\,\Pi_0 O\big) + \sum_{\omega=1}^{N} \sum_{r=1}^{d_\omega} f_{\omega,r}(x), \tag{4}$$

where Π_0 is the orthogonal projection onto $\mathcal{B}_0$ and

$$f_{\omega,r}(x) = 2^{\,N-m_{\omega,r}-1} \begin{bmatrix} \langle \rho, \mathcal{X}_{\omega,r}\rangle \\ \langle \rho, \mathcal{Y}_{\omega,r}\rangle \end{bmatrix}^{\top} \begin{bmatrix} \cos(\omega x) & -\sin(\omega x) \\ \sin(\omega x) & \cos(\omega x) \end{bmatrix} \begin{bmatrix} \langle O, \mathcal{X}_{\omega,r}\rangle \\ \langle O, \mathcal{Y}_{\omega,r}\rangle \end{bmatrix}. \tag{5}$$

The constant term $\mathrm{Tr}(\rho\Pi_0 O)$ is the $\omega = 0$ component, the part of O that commutes with G. With $\mathbf{a}_{\omega,r}$ and $\mathbf{b}_{\omega,r}$ the projection vectors of ρ and O in equation 5, each component is a single sinusoid

$$f_{\omega,r}(x) = 2^{\,N-m_{\omega,r}-1}\, \|\mathbf{a}_{\omega,r}\|\, \|\mathbf{b}_{\omega,r}\| \cos(\omega x + \varphi_{\omega,r}), \tag{6}$$

where $\varphi_{\omega,r}$ is the angle between the two projections (Fig. 1). We call ω *reachable* if its coefficient c_ω in $f(x) = \sum_{|\omega|\le N} c_\omega e^{i\omega x}$ is not identically zero over the trainable parameters. It is unreachable if $\mathbf{b}_{\omega,r} = 0$ for every r at every parameter setting. Planes at the same ω can cancel, so the converse needs proof, and Corollary 1 below shows that such cancellation is confined to a set of parameters of measure zero.

Multi-Qubit Entanglement Design. Proposition 3 says that a frequency reaches the output only if both ρ and O project onto one of its planes, and Proposition 2 says that every plane at frequency ω involves at least ω active qubits, each carrying a $\hat{Y}$ or $\hat{Z}$ factor. A local readout $M = \hat{Z}_n$ has a single active qubit, and single-qubit rotations keep it on qubit n, so an effective observable built from local gates alone has zero projection onto every plane at $\omega > 1$, however its parameters are trained. Only entangling gates can add active qubits, so the design question is how an entangler transforms the readout and where it should sit relative to the local rotations. We restrict the entangler to the couplings the hardware allows, described by an interaction graph $\mathcal{G} = (V, E)$ with $V = \{1, \dots, N\}$ and neighbour set $\mathcal{N}(n) = \{k : \{n, k\} \in E\}$, and use the Ising entangler (Briegel & Raussendorf, 2001), which represents the entangler class above up to single-qubit Clifford gates, CZ included. Its generator is diagonal and built from commuting $\hat{Z}_a\hat{Z}_b$ couplings, so only the couplings at qubit n act on its Pauli operators, and their action has the closed form of Lemma 2.

Lemma 2 (Ising Entanglement)**.** *For a qubit n, define*

$$U_{\mathrm{ent}} = \exp\Big(-\tfrac{i}{2}\sum_{k\in\mathcal{N}(n)} \gamma_{nk}\hat{Z}_n\hat{Z}_k\Big) = \exp\Big(-\tfrac{i}{2}\hat{Z}_n\Phi_n\Big), \qquad \Phi_n = \sum_{k\in\mathcal{N}(n)} \gamma_{nk}\hat{Z}_k. \tag{7}$$

Then $U_{\mathrm{ent}}^\dagger \hat{Z}_n U_{\mathrm{ent}} = \hat{Z}_n$, $U_{\mathrm{ent}}^\dagger \hat{X}_n U_{\mathrm{ent}} = \hat{X}_n \cos\Phi_n - \hat{Y}_n \sin\Phi_n$, and $U_{\mathrm{ent}}^\dagger \hat{Y}_n U_{\mathrm{ent}} = \hat{Y}_n \cos\Phi_n + \hat{X}_n \sin\Phi_n$.

The three identities answer the design question. The first shows that the entangler leaves $\hat{Z}_n$ unchanged, so an entangler placed directly before the readout has no effect on O. The other two show that it rotates $\hat{X}_n$ and $\hat{Y}_n$ by the operator-valued angle Φ_n, and since $\cos\Phi_n$ and $\sin\Phi_n$ expand into $\hat{Z}$-strings over $\mathcal{N}(n)$, every neighbour in these strings adds an active qubit. The entangler can therefore create support at higher frequencies, but only from the $\hat{X}_n$ and $\hat{Y}_n$ components, which the bare readout lacks. A local rotation must first create these components and hence sit between the readout and the entangler, while a second local rotation on the other side can reshape the strings the entangler produces. This gives the trainable block $W_L = W_L^{\mathrm{post}} U_{\mathrm{ent}} W_L^{\mathrm{pre}}$, and we use the same form $W_R = W_R^{\mathrm{post}} U_{\mathrm{ent}} W_R^{\mathrm{pre}}$ on the state side. Reading $O = W_L^\dagger \hat{Z}_n W_L$ from the inside out gives three stages, illustrated in Fig. 2 for $\omega = 3$ and derived for any N and graph in Appendix E.3.

1. **Expose.** W_L^{post} tilts the readout into $M_1 = m_x\hat{X}_n + m_y\hat{Y}_n + m_z\hat{Z}_n$, creating the $\hat{X}_n$ and $\hat{Y}_n$ components that the entangler acts on. Without this stage the next one is the identity.

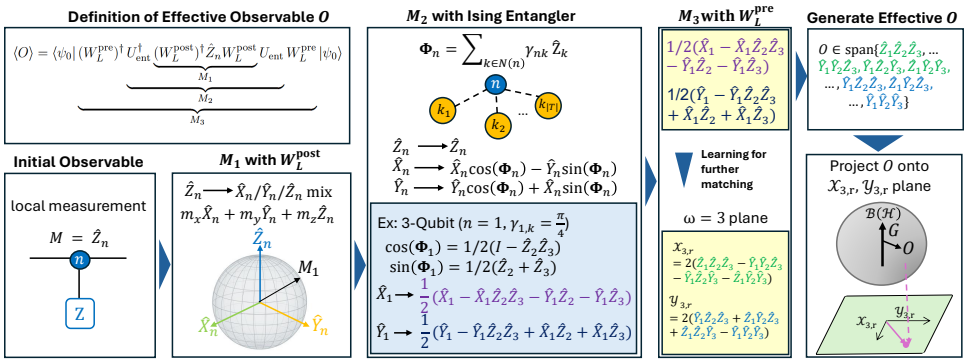


Figure 2: **Design flow for generating an effective observable O.** Local rotations and Ising entanglement transform the initial readout Z_n into mixed Pauli strings, allowing O to project onto the target $\omega = 3$ invariant plane.

2. **Spread.** U_{ent} multiplies these components by $\cos\Phi_n$ and $\sin\Phi_n$, attaching $\hat{Z}$-strings over $\mathcal{N}(n)$ and producing strings on $\{n\}\cup\mathcal{N}(n)$ with up to $\deg(n)+1$ active qubits. The graph $\mathcal{G}$ fixes which qubits can appear, so one round caps the reachable frequency at $\deg(n)+1$ for any N and any graph, which gives $\omega = 3$ for the star graph of Fig. 2. The string that reaches this ceiling has an overlap with its plane proportional to $\prod_k |\sin\gamma_{nk}|$ (Appendix E.3), so by equation 6 the strengths γ_{nk} set the amplitude of the top frequency, which is largest for CZ couplings, $|\gamma_{nk}| = \pi/2$.
3. **Align.** W_L^{pre} rotates the Pauli factors on this support without changing it, reweighting the strings to increase the projection of O onto the target plane, as the matching stage of W_R does for ρ.

Corollary 1 (Sufficient Condition for a Nonzero Coefficient)**.** *Let the local blocks be general single-qubit rotations and ρ_0 a pure product state. If $\mathbf{b}_{\omega,r} \neq 0$ for some plane at frequency $\omega \geq 1$ at one parameter setting, then $c_\omega \neq 0$ at almost every parameter setting. For $W_L = W_L^{\text{post}} U_{\text{ent}} W_L^{\text{pre}}$ with $\sin\gamma_{nk} \neq 0$, this holds for every $\omega \leq \deg(n)+1$.*

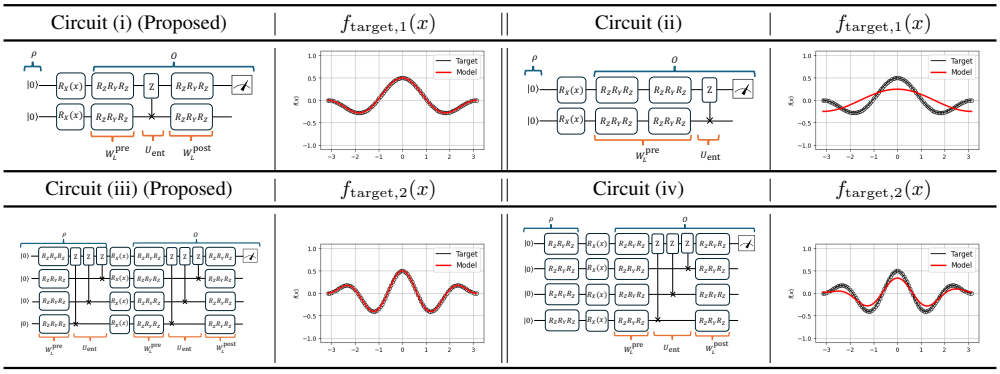
The three stages are therefore also sufficient, and they set the magnitude as well. With CZ couplings and the same entangler in W_R, the circuit can output $f(x) = \cos(\omega x)$ for any single $\omega \leq \deg(n)+1$, whereas any product state ρ limits the amplitude at $\omega = \deg(n)+1$ to $2^{-\deg(n)}$ (Appendix D.6).

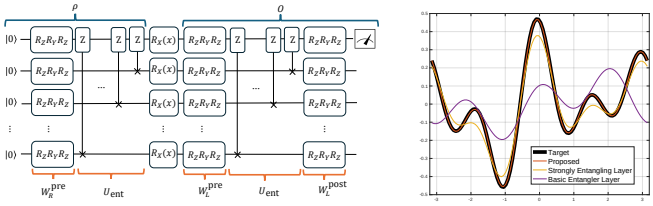
Design procedure. Given the largest frequency $\omega^\star$ a task needs, we measure a qubit with $\deg(n) \geq \omega^\star - 1$ and place Expose, Spread and Align around its couplings on both sides of the encoder (Algorithm 1). The procedure reads only the degrees of $\mathcal{G}$, runs in $O(|E|)$ time and never enumerates the 4^N-dimensional operator space. When the spectrum is unknown, $\omega^\star$ can be estimated from a discrete Fourier transform of the training targets (Boyd, 2001; Xu et al., 2024), and our experiments use $\omega^\star = N$, for which the rule returns the star centered on the measured qubit.

4 Experiments

Experiment Setup. We evaluate the proposed design on sinusoid approximation, supervised regression of the PDEBench advection and Burgers' equations (Takamoto et al., 2022), a placement ablation, a noise study and the physics-informed electromagnetic wave equation. We compare against the Basic Entangler and Strongly Entangling layers of PennyLane (Bergholm et al., 2018), widely used in QNN studies (Suryotrisongko & Musashi, 2022; Dutta et al., 2025; Phalak & Ghosh, 2024; Schuld et al., 2020), Circuits 15–19 of Sim et al. (2019), the circular hardware-efficient ansatz (Kandala et al., 2017; Strobl et al., 2025) and the YZY ansätze of Strobl et al. (2025) (Appendix A). Unless stated otherwise, every circuit encodes its inputs once, which is the setting of our analysis, and a template reaches the parameter budget of the proposed circuit by stacking processing layers. Appendix C.2 repeats the advection comparison and the noise study with data re-uploading, which lies beyond our analysis. Variants marked (γ) also train the entangling strengths γ_{nk} of Lemma 2, both proposed variants use the same two-qubit gates, and 2q in the tables counts them. Regression is scored by the MSE divided by the variance of the target, so a constant predictor scores 1. The physics-informed tasks are scored by the relative L^2 error $\|\hat{u}_\theta - u\|_2/\|u\|_2$ between the prediction $\hat{u}_\theta$ and the analytic solution u on an evaluation grid, which equals 1 for the zero field, and a model

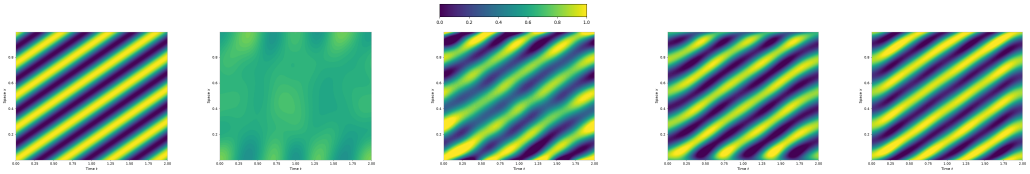
Table 1: Effect of O and ρ on PQC.





(a) **Proposed circuit.** (b) **Performance**

Figure 3: **Proposed circuit architecture and performance comparison.** (a) Structure of the proposed parameterized quantum circuit. (b) Comparison with the considered baseline architectures.



Ground Truth | Basic Ent. | Strongly Ent. ($L = 4$) | **Proposed** | **Proposed** (γ)

Figure 4: **Single-upload reconstructions at 14 qubits**. Advection equation ($\beta = 0.4$).

solves a task when this error falls below 0.1. We report mean $\pm$ s.d. over seeds and compare models with Welch's t-test under Holm correction.

Sinusoid Function Approximation. We test the roles of the effective observable O and effective state $\rho = W_R \rho_0 W_R^\dagger$ with the circuit ablations in Table 1. Circuits (i) and (iii) follow $W_L = W_L^{\text{post}} U_{\text{ent}} W_L^{\text{pre}}$ with $R_Z R_Y R_Z$ local blocks and a CZ entangler, and Circuits (ii) and (iv) remove or misplace these local rotations and entangling layers. For $f_{\text{target},1} = 0.25\cos(x) + 0.25\cos(2x)$, two encoded qubits admit $\omega \leq 2$, and Circuit (i) reconstructs the target while Circuit (ii) underfits the $\omega = 2$ component. The encoder thus fixes which frequencies are admissible, while their amplitudes depend on how O projects onto the invariant planes, and without local dressing O lacks support on the mixed Pauli strings that span the higher-frequency planes. For $f_{\text{target},2} = 0.25\cos(2x) + 0.25\cos(3x)$, Circuit (iii), which dresses both ρ and O, outperforms Circuit (iv), consistent with a coefficient that is bilinear in the two projections and is suppressed when either side lacks support. We then compare the Basic Entangler, the Strongly Entangling layer and the proposed circuit (Fig. 3a) on $f_{\text{target}} = 0.2\cos(2x) + 0.1\sin(x) + 0.11\cos(3x) + 0.15\cos(4x) - 0.1\sin(4x)$. The Basic Entangler fails to capture the target, the Strongly Entangling layer fits it with visible deviation, and the proposed circuit matches it across the domain, with a structure that follows from the invariant-plane analysis rather than a heuristic gate arrangement.

Supervised PDE Regression. We further evaluate the circuits on two PDEBench tasks (Takamoto et al., 2022), the advection equation with $\beta = 0.4$ and Burgers' equation with $\nu = 0.001$ (Table 6, Appendix B.1), whose solutions share their initial condition. We regress each field $u(x, t)$ with x encoded on odd-indexed and t on even-indexed qubits. We first use the setting of our analysis, 14

qubits and a single encoding layer, where a template can reach the budget only by stacking processing layers that add parameters but no frequencies (Table 2). On advection, six of the eight templates, including the Basic Entangler and Circuits 16–19, do not beat the constant predictor, and the best, the circular HEA, reaches 0.155 with four layers and 168 parameters, while the proposed circuit reaches 0.107 with one layer and 126 parameters, or 0.039 when it also trains its entangling strengths, and reproduces the stripes most closely (Fig. 4). Burgers, whose steep fronts make it harder for every circuit, keeps this order, with seven templates near the constant predictor, the circular HEA successful in one of three runs and both proposed variants below 0.21 on every seed. With data re-uploading, a setting beyond our analysis, the proposed circuit also leads every matched template at 7 qubits while using 48 two-qubit gates against 96 to 329 (Appendix C.2).

Table 2: Single-upload regression at 14 qubits over 3 seeds, MSE normalized by the variance of u so that a constant predictor scores 1. L is the number of layers. **Bold** and **blue bold** mark the best and second-best result per task.

Method	$\lvert\theta\rvert$	L	2q	Advection	Burgers	**Method**	$\lvert\theta\rvert$	L	2q	Advection	Burgers
Circuit 18	164	4	65	1.0604 ± 0.0725	1.0222 ± 0.0297	Circuit 19	168	4	65	1.0533 ± 0.0660	1.0220 ± 0.0295
Circuit 17	164	4	28	1.0503 ± 0.0620	1.0222 ± 0.0297	Circuit 16	164	4	28	1.0552 ± 0.0645	1.0223 ± 0.0299
HEA circular	168	4	69	0.1549 ± 0.0469	0.7744 ± 0.2781	YZY (ent.)	168	4	32	1.0989 ± 0.0611	1.0380 ± 0.0501
StronglyEnt.	168	4	56	0.2898 ± 0.0761	0.9267 ± 0.0259	BasicEnt.	154	11	154	1.0335 ± 0.0757	1.0224 ± 0.0302
Proposed	126	1	26	**0.1065 ± 0.0284**	**0.1951 ± 0.0021**	**Proposed** (γ)	152	1	26	$\mathbf{0.0393 \pm 0.0037}$	$\mathbf{0.1890 \pm 0.0143}$

Placement Ablation. We keep the 126 angles and 26 CZ gates of the proposed 14-qubit circuit and change only their order, the graph or the measured qubit (Table 3). Since each Fourier coefficient $c(\omega_x, \omega_t)$ is identically zero or nonzero almost everywhere (Corollary 1), we count the nonzero ones among the 225 exactly at random parameters, without any training (Appendix B.3). Every change removes exactly the pairs our analysis predicts, and No Align shows that reaching the top frequency is not enough, since it keeps $(7, 7)$ but loses 103 of the 225 pairs. To connect these counts to the PDE error, we fit the advection field of Table 2 by least squares on the reachable terms of each variant, which gives a lower bound on the error of every parameter setting. Without Expose, with a leaf readout or with a path, this bound is at least 0.999, near the constant predictor, and without Align it is 0.102, against 0.013 for the proposed circuit. Training every variant the same way, on one eighth of the grid with about one sixth of the optimizer steps of Table 2, gives the proposed circuit 0.63 and every other variant between 0.96 and 1.02.

Table 3: Placement ablation, with gates in time order, R a rotation layer, S the encoder and C or P a star or path of 13 CZ gates. Pairs counts the nonzero $c(\omega_x, \omega_t)$ of 225 and (k_x, k_t) is their largest $(|\omega_x|, |\omega_t|)$, both as predicted. Best possible MSE is the least-squares error on the reachable terms, a lower bound for any parameter setting, and Trained MSE uses a reduced budget over 5 seeds, both normalized by the variance of u.

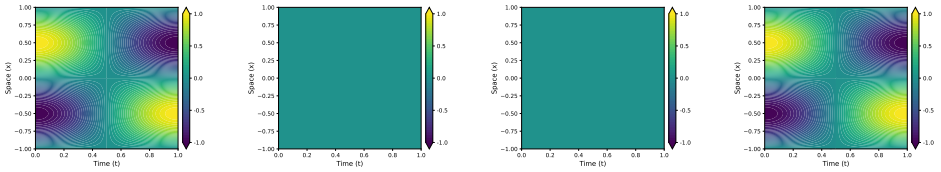
	Proposed	No Expose	No Align	Leaf readout	Path
Gates	$RCSRCR$	$RCSRRC$	$RCSCRR$	$RCSRCR$	$RPSRPR$
(k_x, k_t)	$(7, 7)$	$(1, 0)$	$(7, 7)$	$(1, 1)$	$(2, 1)$
Pairs	225	3	122	9	15
Best possible MSE ↓	**0.013**	1.000	0.102	1.000	0.999
Trained MSE ↓	$\mathbf{0.628 \pm 0.212}$	1.020 ± 0.024	0.957 ± 0.016	1.015 ± 0.012	1.022 ± 0.012

Noise Simulation. We evaluate the trained 14-qubit advection models under depolarizing noise of strength ϵ after every gate, amplitude and phase damping of strength λ after every circuit moment, and finite-shot readout with S shots (Table 4, Appendix C.3). The proposed circuit stays accurate under moderate noise. With trained entangling strengths, its error moves from 0.026 to 0.036 and 0.105 at $\epsilon = 3 \times 10^{-3}$ and 10^{-2}, against 0.250 and 0.473 for HEA, the strongest template, and to 0.067 at $\lambda = 10^{-3}$, while stronger damping at $\lambda = 3 \times 10^{-3}$ raises it to about 0.5. Finite-shot readout is the limiting factor at 14 qubits. Its error grows with the square of the trained readout scale, which is four times smaller for our circuit than for HEA, so our shot error is $16\times$ smaller, but our circuit still needs more than 10^4 shots per point to beat the constant predictor. An exact density-matrix study at 7 qubits with every template, under data re-uploading, is in Appendix C.2.

Physics-Informed Wave Equation. We also solve the electromagnetic wave equation from Maxwell's equations with the physics-informed objective of Table 7 and no solution data. Its solution $\sin(\pi x)\cos(\pi t)$ has one frequency per input, so our rule needs one qubit per input, and we compare circuits at equal depth. With only 2 qubits and one layer of 20 parameters, the proposed

Table 4: Noise at 14 qubits with a single encoding layer, MSE normalized by the variance of u. Gate noise and damping use seed 0 on 128 evaluation points with 512 debiased trajectories, and shots use 3 seeds on 2048 points with 10 draws each. $|w|$ is the trained readout scale.

				Gate noise ϵ		Damping λ		Shots S	
Method	2q	$\lvert w\rvert$	Clean	3×10^{-3}	10^{-2}	10^{-3}	3×10^{-3}	128	512
HEA circular	69	18.48	0.193	0.250	0.473	–	–	1585	397
Proposed	26	5.79	0.064	0.079	0.154	0.082	0.494	156.1	39.3
Proposed (γ)	26	4.64	0.026	0.036	0.105	0.067	0.557	99.3	25.0



Ground Truth | Basic Ent. | Strongly Ent. | **Proposed**

Figure 5: **Physics-informed wave equation** with one layer per circuit. The Strongly Entangling layer on 6 qubits and the proposed circuit on 2 qubits both use 20 parameters.

Table 5: Physics-informed wave equation, mean relative L^2 error. Templates need more layers, since with one layer each outputs the zero field (Table 8).

Circuit	Qubits	L	$\lvert\theta\rvert$	Rel. L^2	Circuit	Qubits	L	$\lvert\theta\rvert$	Rel. L^2
Circuit 16	6	2	36	8.1×10^{-5}	YZY	6	2	38	0.336
Circuit 17	6	2	36	3.1×10^{-5}	YZY (ent.)	6	2	38	0.336
Circuit 18	6	2	38	3.6×10^{-4}	HEA circular	6	2	38	0.455
Circuit 19	6	2	38	3.0×10^{-5}	Circuit 15	6	2	26	1.000
Strongly Ent.	6	3	56	9.9×10^{-3}	Basic Ent.	6	2	14	1.000
Proposed	2	1	20	2.7×10^{-5}					

circuit solves the task with a relative L^2 error of 2.7×10^{-5}, while no template learns anything with one layer even on three times as many qubits, each outputting the zero field (Fig. 5, Table 8). Templates need more layers, two and 36 to 38 parameters for Circuits 16–19 and three and 56 for the Strongly Entangling layer (Table 5), so ours uses a third of their qubits and about half their parameters, although larger templates can exceed its accuracy.

5 Conclusion

In this paper, we presented a geometric framework for analyzing the Fourier expressivity of multi-qubit PQCs. Using the adjoint map, we showed that the data-encoding generator decomposes the operator space into invariant frequency subspaces, where Fourier coefficients are determined by geometric projections between the effective state and observable. This perspective reveals how entanglement activates higher-frequency components by transforming local operators into mixed multi-qubit Pauli strings. For the encoder, readout and entangler we analyze, it yields a placement rule, built from the interaction graph in linear time, that makes every frequency up to the degree of the measured qubit plus one nonzero at almost every parameter setting, and each of these frequencies can reach unit amplitude on its own when the state side is entangled as well. Moving the same gates removes exactly the frequencies that the rule predicts. With trained entangling strengths the proposed circuit leads every parameter-matched template on advection and Burgers while using fewer two-qubit gates, it stays accurate under moderate gate noise and damping, and a single encoding layer suffices to solve the wave equation.

GenAI Usage Disclosure

We occasionally used ChatGPT to refine our wording and grammar. All manuscript and contents of this paper were checked and reviewed by authors.

Ethics statement

All authors have read and adhere to the ICLR Code of Ethics. This work develops theory and a design rule for parameterized quantum circuits and evaluates them in classical simulation. It involves no human subjects, no personal or sensitive data and no deployed system. The supervised experiments use the public PDEBench datasets (Takamoto et al., 2022) under their license, and every other target is synthetic and given in closed form. We see no direct path to harmful use, and the computational cost of the work is limited to classical simulation on a single workstation (Table 9).

Reproducibility statement

The proofs of all lemmas and propositions are in Appendix D, the proof of Corollary 1 is in Appendix D.6, the scope of the results is discussed in Appendix E.5, and the design procedure is stated as Algorithm 1. Appendix A defines every baseline circuit, and Appendix B gives the data, encoding, objective and training settings of each experiment, with the number of seeds and the hardware and software we used (Table 9). The comparison tables list the parameter and two-qubit gate counts of every circuit, and we report the mean and standard deviation over seeds with Holm-corrected Welch tests. The supervised data are public PDEBench samples, the other targets are given in closed form, and the reachable-pair counts of Table 3 require no training and follow from the procedure of Appendix B.3. Our code, built on PennyLane (Bergholm et al., 2018), JAX and Optax, is included in the supplementary material and will be made public upon publication.

References

Alice Barthe and Adrián Pérez-Salinas. Gradients and frequency profiles of quantum re-uploading models. *Quantum*, 8:1523, 2024.

Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.

Simon Berger, Norbert Hosters, and Matthias Möller. Trainable embedding quantum physics informed neural networks for solving nonlinear pdes. *Scientific Reports*, 15(1):18823, 2025. doi: 10.1038/s41598-025-02959-z.

Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M Sohaib Alam, Guillermo Alonso-Linaje, B AkashNarayanan, Ali Asadi, et al. Pennylane: automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018.

John P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications, 2 edition, 2001.

Hans J. Briegel and Robert Raussendorf. Persistent entanglement in arrays of interacting particles. *Physical Review Letters*, 86(5):910–913, 2001.

Kyle J. S. Campbell, Luigi Del Debbio, and Petros Wallden. Circuit harmonic matrices: A spectral framework for quantum machine learning. *arXiv preprint arXiv:2604.04292*, 2026.

Berta Casas and Alba Cervera-Lierta. Multidimensional Fourier series with quantum circuits. *Physical Review A*, 107(6):062612, 2023.

Ziv Chen, Gal G. Shaviner, Hemanth Chandravamsi, Shimon Pisnoy, Steven H. Frankel, and Uzi Pereg. Quantum physics-informed neural networks for Maxwell's equations: Circuit design, "black hole" barren plateaus mitigation, and GPU acceleration. *Quantum Machine Intelligence*, 8:21, 2026.

Claude Chevalley. *Theory of Lie Groups*. Courier Dover Publications, 2018.

Shivanya Shomir Dutta, Sahil Sandeep, Nandhini D, and Amutha S. Hybrid quantum neural networks: harnessing dressed quantum circuits for enhanced tsunami prediction via earthquake data fusion. *EPJ Quantum Technology*, 12:4, 2025.

Afrah Farea, Saiful Khan, and Mustafa Serdar Celebi. QCPINN: Quantum-classical physics-informed neural networks for solving PDEs. *arXiv preprint arXiv:2503.16678*, 2025.

Enrico Fontana, Dylan Herman, Shouvanik Chakrabarti, Niraj Kumar, Romina Yalovetzky, Jamie Heredge, Shree Hari Sureshbabu, and Marco Pistoia. Characterizing barren plateaus in quantum ansätze with the adjoint representation. *Nature Communications*, 15(1):7171, 2024.

Crispin Gardiner and Peter Zoller. *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*. Springer Science & Business Media, 2004.

Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017. doi: 10.1038/nature23879.

Oleksandr Kyriienko, Annie E. Paine, and Vincent E. Elfving. Solving nonlinear differential equations with differentiable quantum circuits. *Physical Review A*, 103(5):052416, 2021. doi: 10.1103/PhysRevA.103.052416.

Martin Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J Coles, and Marco Cerezo. Diagnosing barren plateaus with tools from quantum optimal control. *Quantum*, 6:824, 2022a.

Martín Larocca, Frédéric Sauvage, Faris M. Sbahi, Guillaume Verdon, Patrick J. Coles, and M. Cerezo. Group-invariant quantum machine learning. *PRX Quantum*, 3(3):030341, 2022b. doi: 10.1103/PRXQuantum.3.030341.

Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert. Exploiting symmetry in variational quantum machine learning. *PRX Quantum*, 4(1):010328, 2023. doi: 10.1103/PRXQuantum.4.010328.

Hela Mhiri, Léo Monbroussou, Mario Herrero-Gonzalez, Slimane Thabet, Elham Kashefi, and Jonas Landman. Constrained and vanishing expressivity of quantum Fourier models. *Quantum*, 9:1847, 2025.

Bogdan Mielnik. Geometry of quantum states. *Communications in Mathematical Physics*, 9(1): 55–80, 1968.

Nikita A Nemkov, Evgeniy O Kiktenko, and Aleksey K Fedorov. Fourier expansion in variational quantum algorithms. *Physical Review A*, 108(3):032406, 2023.

Quynh T Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J Coles, Frederic Sauvage, Martin Larocca, and Marco Cerezo. Theory for equivariant quantum neural networks. *PRX Quantum*, 5(2):020328, 2024.

Giorgio Panichi, Sebastiano Corli, and Enrico Prati. Quantum physics-informed neural networks for multi-variable partial differential equations. *Physical Review Applied*, 25(1):014001, 2026.

Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020.

Evan Peters and Maria Schuld. Generalization despite overfitting in quantum machine learning models. *Quantum*, 7:1210, 2023.

Koustubh Phalak and Swaroop Ghosh. Non-parametric greedy optimization of parametric quantum circuits. In *International Symposium on Quality Electronic Design (ISQED)*, pp. 1–7. IEEE, 2024.

Michael Ragone, Bojko N Bakalov, Frédéric Sauvage, Alexander F Kemper, Carlos Ortiz Marrero, Martín Larocca, and M Cerezo. A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nature Communications*, 15:7172, 2024.

Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103:032430, 2021.

Alexandr Sedykh, Maninadh Podapaka, Asel Sagingalieva, Karan Pinto, Markus Pflitsch, and Alexey Melnikov. Hybrid quantum physics-informed neural networks for simulating computational fluid dynamics in complex shapes. *Machine Learning: Science and Technology*, 5(2): 025045, 2024.

Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, 2019.

Melvin Strobl, M. Emre Sahin, Lucas van der Horst, Eileen Kuehn, Achim Streit, and Ben Jaderberg. Fourier fingerprints of ansatzes in quantum machine learning. *arXiv preprint arXiv:2508.20868*, 2025.

Hatma Suryotrisongko and Yasuo Musashi. Evaluating hybrid quantum-classical deep learning for cybersecurity botnet DGA detection. *Procedia Computer Science*, 197:223–229, 2022.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. *Advances in Neural Information Processing Systems*, 35:1596–1611, 2022.

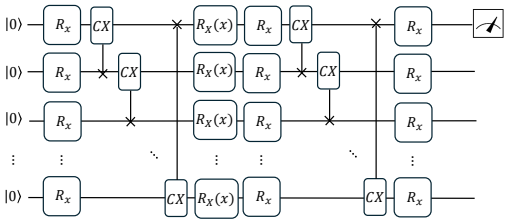
Corey Trahan, Mark Loveland, and Samuel Dent. Quantum physics-informed neural networks. *Entropy*, 26(8):649, 2024. doi: 10.3390/e26080649.

Marco Wiedmann, Maniraman Periyasamy, and Daniel D. Scherer. Fourier analysis of variational quantum circuits for supervised learning. In *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 01, pp. 1785–1795, 2025.

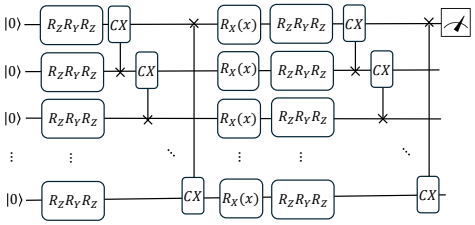
Y. Xiao, L. M. Yang, C. Shu, S. C. Chew, B. C. Khoo, Y. D. Cui, and Y. Y. Liu. Physics-informed quantum neural network for solving forward and inverse problems of partial differential equations. *Physics of Fluids*, 36(9):097145, 2024. doi: 10.1063/5.0226232.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with $10k$ parameters. In *International Conference on Learning Representations*, 2024.

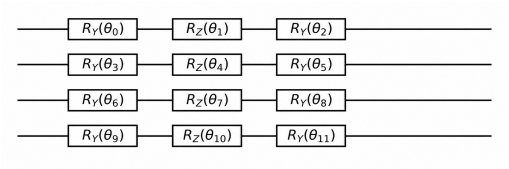
A BASELINE CIRCUITS



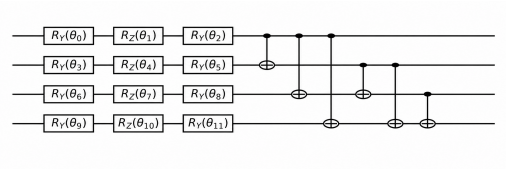
(a) Basic Entangler



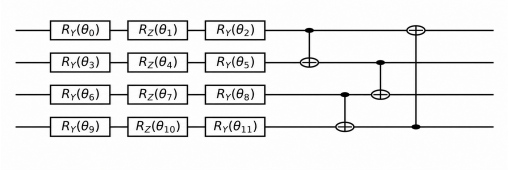
(b) Strongly Entangling



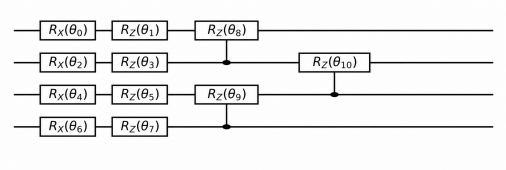
(c) YZY



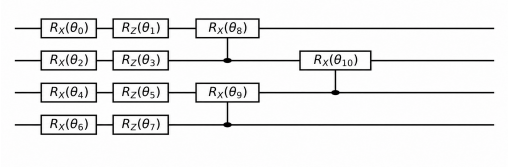
(d) YZY Entangling



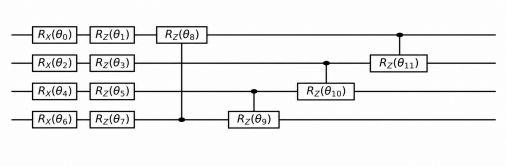
(e) Hardware Efficient Ansatz (HEA)



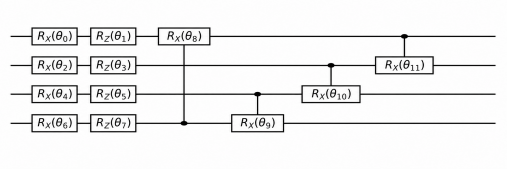
(f) Circuit 16



(g) Circuit 17



(h) Circuit 18



(i) Circuit 19

Figure 6: PQC baseline architectures considered in the experiments.

B EXPERIMENTAL SETUP

In this section, we describe the details of datasets that are utilized in the experiments.

B.1 DETAILED EXPERIMENT SCENARIOS

Sinusoidal Function Approximation. In this experiment, we use a synthetic dataset generated from a sinusoidal target function. Input points are uniformly sampled over a fixed interval, and target values are obtained directly from the analytic function. This dataset provides a controlled benchmark for evaluating the model's ability to learn oscillatory behavior and frequency-dependent patterns. The proposed model was trained using the Adam optimizer with the default PyTorch settings. The training was performed for 1000 epochs, where each epoch iterates over all mini-batches in the training dataloader. The loss function used for optimization was the smooth $L1$ loss.

Supervised PDE Regression. Both settings regress PDEBench solutions (Takamoto et al., 2022) and scale the coordinates by π before encoding (Table 6). The single-upload setting uses the first sample of the advection data with $\beta = 0.4$ and of the Burgers data with $\nu = 0.001$, which share their initial condition, each on its full 201×1024 grid with the raw field u. Every circuit uses 14 qubits and encodes the coordinates only in its first layer, so additional layers add parameters but no frequencies, and every template is depth-scaled to at least the 152 parameters of the larger proposed variant. Each circuit is trained with Adam at learning rate 0.015, without clipping or decay, for 6 epochs at batch size 64 on a random 80% of the grid (15,432 steps), and we report the MSE over the full grid divided by the variance of u, 1.70×10^{-3} for advection and 1.01×10^{-3} for Burgers, so a constant predictor scores 1. The variant with trainable entangling strengths replaces each CZ by a controlled-phase gate initialized at π, where it equals CZ, and all circuits share the same 3 seeds. The re-uploading setting uses advection only. There, every circuit uses 7 qubits and re-encodes the input between blocks, on a 48×24 space-time grid with an 80/20 train-test split and targets standardized to zero mean and unit variance, uses the readout $a\langle Z_0 \rangle + b$, and is trained with Adam for 30 epochs at batch size 64 over 8 seeds. Each model uses its best learning rate from $\{0.005, 0.02, 0.05, 0.1, 0.2\}$. The noise study retrains 5 of the 8 seeds with the same procedure, requires each to reproduce its noiseless test error, and averages finite-shot readout over 10 draws per seed.

Table 6: PDEBench regression setup.

Component	Notation	Equation
Advection equation	β	$\partial_t u(x,t) + \beta\, \partial_x u(x,t) = 0$
Burgers equation	ν	$\partial_t u(x,t) + \partial_x \left(u^2(x,t)/2\right) = (\nu/\pi)\, \partial_{xx} u(x,t)$
Input encoding	(x,t)	$q_1 \leftarrow x, \quad q_2 \leftarrow t, \quad q_3 \leftarrow x, \quad q_4 \leftarrow t, \quad \cdots$
Model mapping	f_θ	$f_\theta : (x,t) \mapsto \hat{u}_\theta(x,t)$
Training objective	$\mathcal{L}$	$\min_\theta \frac{1}{N} \sum_{i=1}^{N} \lvert \hat{u}_\theta(x_i,t_i) - u(x_i,t_i) \rvert^2$

Physics-Informed Wave Equation. We solve $\partial_{tt} E = \partial_{xx} E$ on $[-1,1] \times [0,1]$ with $E(\pm 1, t) = 0$, $E(x,0) = \sin(\pi x)$ and $\partial_t E(x,0) = 0$, whose exact solution is $E = \sin(\pi x)\cos(\pi t)$, with the objective of Table 7. Every circuit encodes x and t once with $R_X(\pi\,\cdot)$, one input per qubit, so the solution has frequency 1 in each input, and the proposed circuit is the $R\,C\,S\,R\,C\,R$ block of Table 3 on 2 qubits, with a single CZ gate as C. Every circuit is trained for 300 steps with Adam at a constant learning rate of 0.015 after gradient clipping at norm 1, resampling 1000 collocation, 200 boundary and 200 initial points at every step, over 3 shared seeds. We score the relative L^2 error against the exact solution on a 64×32 grid. Rather than matching one parameter budget, we report each template with one and two layers, and the Strongly Entangling layer with one and three, and list every parameter count in Table 8. Every template with one layer outputs the zero field, which satisfies the equation and the boundary conditions but leaves the initial-condition loss at 0.51. With more parameters, templates can exceed our accuracy, and in earlier runs with 56 parameters and 500 steps, Circuit 18 reached 1.5×10^{-6}.

Table 7: Wave-equation training objective.

Component	Notation	Equation
Governing equation	$E_\theta(x,t)$	$\partial_{tt} E_\theta - \partial_{xx} E_\theta = 0, \quad x \in [-1,1],\ t \in [0,1]$
PDE residual	$\mathcal{L}_{\mathrm{PDE}}$	$\frac{1}{N_c} \sum_{i=1}^{N_c} \lvert \partial_{tt} E_\theta(x_i,t_i) - \partial_{xx} E_\theta(x_i,t_i) \rvert^2$
Boundary condition	$\mathcal{L}_{\mathrm{BC}}$	$E_\theta(-1,t) = 0, \quad E_\theta(1,t) = 0$
Initial condition	$\mathcal{L}_{\mathrm{IC}}$	$E_\theta(x,0) = \sin(\pi x), \quad \partial_t E_\theta(x,0) = 0$
Total objective	$\mathcal{L}$	$\mathcal{L}_{\mathrm{PDE}} + 10\,\mathcal{L}_{\mathrm{BC}} + 10\,\mathcal{L}_{\mathrm{IC}}$

B.2 Computational Resources

Table 9 lists the hardware and software we used.

Table 8: Physics-informed wave equation with one encoding layer, mean relative L^2 error over 3 seeds. Every template uses 6 qubits.

One layer				More layers, 6 qubits			
Circuit	Qubits	$\|\theta\|$	Rel. L^2	Circuit	L	$\|\theta\|$	Rel. L^2
HEA circular	6	20	1.0004	Circuit 16	2	36	8.1×10^{-5}
Strongly Ent.	6	20	1.0004	Circuit 17	2	36	3.1×10^{-5}
Circuit 15	6	14	1.0003	Circuit 18	2	38	3.6×10^{-4}
Circuit 16	6	19	1.0006	Circuit 19	2	38	3.0×10^{-5}
Circuit 17	6	19	1.0005	YZY	2	38	0.336
Circuit 18	6	20	1.0006	YZY (ent.)	2	38	0.336
Circuit 19	6	20	1.0005	HEA circular	2	38	0.455
YZY, YZY (ent.)	6	20	1.0006	Circuit 15	2	26	1.000
Basic Ent.	6	8	1.0004	Basic Ent.	2	14	1.000
Proposed	2	20	2.7×10^{-5}	Strongly Ent.	3	56	9.9×10^{-3}

Table 9: Computational resources and software libraries used in the experiments.

Resource / Library	Specification
CPU	AMD Ryzen 7 9700X 8-Core Processor, 3.80 GHz
RAM	64 GB
GPU	NVIDIA GeForce RTX 5080, 16 GB
Operating System	64-bit Windows, x64-based processor
Storage	1.82 TB SSD
Quantum ML Library	PennyLane
ML / Auto-differentiation Backend	JAX
Optimizer Library	Optax
Numerical Libraries	NumPy, Pandas
Visualization Library	Matplotlib

B.3 Placement Ablation

Circuits. All circuits act on 14 qubits, encode x on odd-indexed and t on even-indexed qubits with $R_X(\pi\,\cdot)$, and measure $\hat{Z}$ on qubit 1, which encodes x, except Leaf readout, which measures qubit 2, which encodes t. Each R applies $R_Z R_Y R_Z$ to every qubit (42 angles), C applies CZ between qubit 1 and every other qubit, and P applies CZ along the path 2–1–3–4–$\cdots$–14, on which qubit 1 has one x-neighbour and one t-neighbour. The proposed circuit $R\,C\,S\,R\,C\,R$ is $W_R = CR$ and $W_L = RCR$ in operator order, so the rotation between the encoder and the star of W_L is Align and the rotation after that star is Expose. No Expose, $R\,C\,S\,R\,R\,C$, therefore places the star directly before the readout, and No Align, $R\,C\,S\,C\,R\,R$, places it directly after the encoder.

Counting reachable pairs. With one encoding layer, $f(x,t) = \sum c(\omega_x,\omega_t)\, e^{i\pi(\omega_x x+\omega_t t)}$ with $|\omega_x|, |\omega_t| \le 7$, which gives 225 coefficients. By Corollary 1, whose proof holds verbatim for two inputs (Appendix D.6), each coefficient is, as a function of the angles, either identically zero or nonzero at almost every setting. We therefore draw all 126 angles uniformly at random, evaluate f on a 16×16 grid of encoding angles over one period and obtain every coefficient exactly by a two-dimensional discrete Fourier transform, which does not alias because $16 > 2 \cdot 7 + 1$. A pair is reachable if its coefficient is nonzero, and (k_x, k_t) is the largest $|\omega_x|$ and $|\omega_t|$ over the reachable pairs. Over 20 draws, reachable coefficients exceed 10^{-10} in magnitude at every draw and the others stay below 10^{-15}, so a single draw gives the exact count, and no data or training enters it.

Predictions. Without Expose, U_{ent} commutes with $\hat{Z}_n$, so O acts on n alone and reaches only $\omega_t = 0$ and $|\omega_x| \le 1$. The leaf readout and the path have $\deg(n) = 1$ and $\deg(n) = 2$, and the bound $|\omega_x| \le |L \cap V_x|$, $|\omega_t| \le |L \cap V_t|$ of Appendix D.6 gives $(1,1)$ and $(2,1)$, with every pair inside reachable at almost every parameter setting. Without Align, the merged rotations tilt the readout into $M_1 = m_x\hat{X}_n + m_y\hat{Y}_n + m_z\hat{Z}_n$ and the star gives $O = m_z\hat{Z}_n + (m_x\hat{X}_n + m_y\hat{Y}_n)\hat{Z}_{\mathcal{N}(n)}$. Each $\hat{Z}_k = (E_k^+ + E_k^-)/2$ adds ± 1 to the frequency of its input and all seven t-qubits are leaves, so the strings through the leaves reach every ω_x but only odd ω_t, and $\hat{Z}_n$ adds $(\pm 1, 0)$, which leaves $15 \times 8 + 2 = 122$ pairs.

Best possible MSE. The output of a variant is $w\, f(x,t) + b$, where f contains only its reachable pairs, so its error can never fall below the least-squares fit of u on the constant and on $\cos(\pi(\omega_x x + \omega_t t))$

and $\sin(\pi(\omega_x x + \omega_t t))$ over those pairs. We compute this fit on the full grid of Table 2 and divide the mean squared residual by the variance of u. It needs no training and bounds from below the error of every parameter setting, whatever the training budget.

Training. We train each variant with the settings of Table 2 on the grid subsampled with stride 8 ($25{,}728$ points), which gives $2{,}412$ optimizer steps against $15{,}432$, with seeds 0 to 4 shared across variants. At this budget the proposed circuit reaches 0.628 ± 0.212 and the four variants stay between 0.957 and 1.022, with Holm-corrected Welch $p = 0.082$ for every variant, limited by the spread of the proposed circuit across seeds. The variant without Expose stays at the constant predictor on every seed, between 1.000 and 1.058, as its output has a single frequency in x and none in t. Every trained error lies above the bound of its variant.

C Additional Experiment Results

C.1 Case Study of O and ρ

We provide additional ablation results to further support the projection-based interpretation of Proposition 3. In the main text, we argued that the encoder determines the accessible frequency set, while the actual Fourier amplitudes depend on the projections of the effective observable O and effective state ρ onto the corresponding invariant planes. Tables 10 and 11 provide detailed circuit-level evidence for this mechanism.

Table 10 examines the effect of entanglement placement on the effective observable O. Circuit (i) follows the proposed structure, where local rotations are placed before and after the entangling layer. This allows the initially local measurement operator to be first rotated into a noncommuting Pauli direction, then dressed by entanglement into mixed multi-qubit Pauli strings, and finally adjusted by the remaining local rotations. As a result, the effective observable can acquire nonzero overlap with higher-frequency invariant planes. This explains why Circuit (i) accurately reconstructs $f_{\text{target},1}(x)$.

In contrast, Circuits (ii)–(iv) modify the placement of the local rotations and entangling blocks. Although these circuits may still contain entanglement, their ordering does not effectively promote the measured observable into the mixed Pauli-string components required for higher-frequency activation. Consequently, the projection of O onto the relevant invariant subspaces is weakened, and the learned function underfits the higher-frequency oscillations. This confirms that entanglement alone is not sufficient; its placement relative to local rotations is essential for shaping the effective observable.

Table 11 further investigates the role of the effective state $\rho = W_R \rho_0 W_R^\dagger$. Since each Fourier coefficient depends on the joint projection of both ρ and O, a target frequency can be suppressed if either side has insufficient overlap with the corresponding invariant plane. Circuit (i) jointly dresses both the state and observable, enabling both operators to support the mixed Pauli strings associated with $f_{\text{target},2}(x)$. This leads to the best reconstruction performance.

The remaining circuits break this mechanism in different ways. Circuit (ii) mainly shapes the observable side but leaves the state insufficiently prepared. Circuit (iii) introduces entanglement on the state side without first generating the necessary local noncommuting components, limiting its ability to populate the target invariant subspaces. Circuit (iv) partially restores observable-side dressing, but the state remains unentangled, which limits the size of its projection onto the target planes. These results show that higher-frequency learning requires coordinated design of both O and ρ, rather than arbitrary entanglement placement.

Overall, the appendix results reinforce the main conclusion: the success of the proposed architecture comes from its ability to align both the effective observable and effective state with the invariant frequency subspaces induced by the data-encoding generator. This provides a circuit-level explanation for why the proposed entanglement placement activates higher Fourier components more reliably than heuristic alternatives.

Table 10: Effect of entanglement placement and O on PQC.

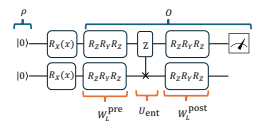
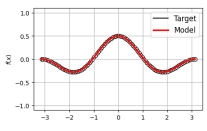
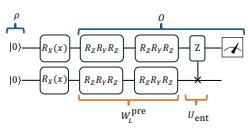
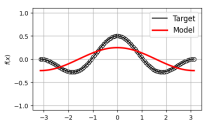
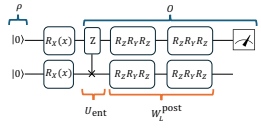
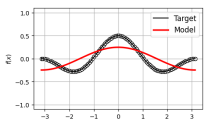
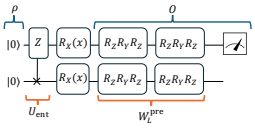
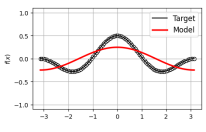
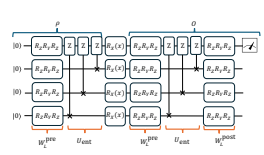
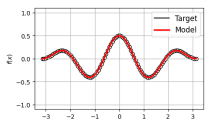
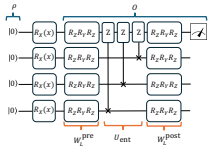
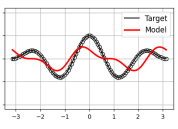
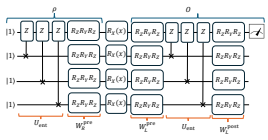
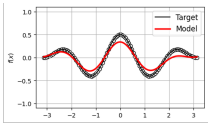
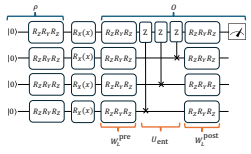
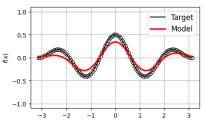
Circuit (i)	$f_{\text{target},1}(x)$	Circuit (ii)	$f_{\text{target},1}(x)$
			
Circuit (iii)	$f_{\text{target},1}(x)$	Circuit (iv)	$f_{\text{target},1}(x)$
			

Table 11: Effect of changing initial quantum state on PQC.

Circuit (i) (Proposed)	$f_{\text{target},2}(x)$	Circuit (ii)	$f_{\text{target},2}(x)$
			
Circuit (iii)	$f_{\text{target},2}(x)$	**Circuit (iv)**	$f_{\text{target},2}(x)$
			

C.2 Results with Data Re-uploading

Our analysis covers one encoding layer, so we treat data re-uploading as a test beyond it. The setting favors the templates, since every added layer re-encodes the input (Appendix F). We run it on advection at 7 qubits, which keeps an exact density-matrix noise study tractable, with every circuit depth-scaled to about 330 parameters (Appendix B.1). Both proposed variants lead every matched template sharing our $R_X(\pi x)$ encoding (Holm $p < 0.001$), the better reaching 0.104 against 0.169 for the nearest, Circuit 17, while using fewer parameters and 48 two-qubit gates against the 96–329 of the entangling templates (Table 12).

Table 12: Supervised advection regression at 7 qubits with re-uploading, test MSE over 8 seeds. **Bold** and **blue bold** mark the best and second-best result.

Method	$\lvert\theta\rvert$	2q	MSE ↓	**Method**	$\lvert\theta\rvert$	2q	MSE ↓
Circuit 18	338	112	0.1699 ± 0.0181	Circuit 19	338	112	0.1762 ± 0.0216
Circuit 17	322	96	0.1694 ± 0.0263	Circuit 16	322	96	0.1833 ± 0.0168
HEA circular	338	112	0.4378 ± 0.0525	YZY (ent.)	338	96	0.3336 ± 0.0292
StronglyEnt.	338	112	0.1835 ± 0.0106	BasicEnt.	331	329	0.5170 ± 0.0146
Proposed	254	48	$\mathbf{0.1158 \pm 0.0189}$	**Proposed** (γ)	302	48	$\mathbf{0.1041 \pm 0.0073}$

Noise at 7 qubits. We evaluate the noiselessly trained advection models of Table 12 under three noise models, after retraining each from its original seed to reproduce its noiseless error exactly. Amplitude and phase damping of strength λ acts on every qubit after every circuit moment, so exposure grows with depth. Depolarizing noise of strength ϵ acts after every two-qubit gate and at $\epsilon/10$ after every single-qubit gate, so exposure grows with gate count. Finite-shot readout estimates each expectation from S shots. In Table 13, the proposed circuit, which has the fewest two-qubit gates, has the lowest mean error at $\lambda = 10^{-3}$, at both gate-noise levels and at $S = 512$, while the chain templates, Circuits 16 and 17, lead under stronger damping and at $S = 128$.

Table 13: Test MSE under noise of the advection models of Table 12, mean over 5 seeds. **Bold** and **blue bold** mark the best and second-best per column.

					Damping λ		Gate noise ϵ		Shots S	
Method	$\lvert\theta\rvert$	2q	Dep.	Clean	10^{-3}	3×10^{-3}	3×10^{-3}	10^{-2}	128	512
Circuit 18	338	112	174	0.1757	0.2551	0.5216	0.2122	0.3829	0.2523	0.1966
Circuit 19	338	112	174	0.1857	0.2809	0.5880	0.2246	0.4187	0.3020	0.2130
Circuit 17	322	96	98	0.1743	0.2028	**0.3280**	0.1935	0.3031	**0.2422**	0.1919
Circuit 16	322	96	98	0.1837	0.2134	**0.3468**	0.2024	0.3129	**0.2435**	0.1990
StronglyEnt.	338	112	92	0.1845	0.3076	0.6529	0.2785	0.6202	0.4749	0.2595
YZY (ent.)	338	96	114	0.3349	0.4391	0.7393	0.3995	0.6468	0.7325	0.4291
HEA circular	338	112	190	0.4154	0.6470	0.9626	0.5014	0.8092	0.8090	0.5181
BasicEnt.	331	329	423	0.5179	0.9198	1.0191	0.7691	1.0022	0.6810	0.5537
Proposed	254	48	64	**0.1203**	**0.1811**	0.4182	**0.1477**	**0.2986**	0.3619	**0.1804**
Proposed (γ)	302	48	64	**0.1077**	**0.1705**	0.4186	**0.1358**	**0.2918**	0.3449	**0.1692**

C.3 Noise at 14 Qubits

An exact density matrix at 14 qubits has 4^{14} entries per state, so we simulate gate noise and damping by quantum trajectories, with the noise models of the 7-qubit study in Appendix C.2. Gate noise is a Pauli channel, which the trajectories unravel exactly, and damping uses the Monte-Carlo wavefunction method. Against an exact simulator at 5 qubits with 800 trajectories, every tested setting agrees within 1.3 standard errors. The MSE of a trajectory-averaged prediction carries a positive bias equal to the variance of that average, and we subtract its sample estimate. We compare against HEA, the strongest template in Table 2, because trajectory simulation of the Strongly Entangling layer exceeded our compute budget, and HEA's damping cells stay unresolved at 512 trajectories, since its readout scale of 18.5 amplifies the trajectory variance. Circuits 16–19 and YZY (ent.) ignore the readout, with $|w| \leq 0.09$, so noise cannot change their error. Finite-shot readout adds a variance of about $w^2(1 - \langle Z_0 \rangle^2)/S$ to each prediction, and the shot errors of different circuits scale with w^2 to within 1%.

C.4 Advection Equation

We further evaluate the proposed circuit on the one-dimensional advection equation, which describes transport-dominated dynamics where an initial profile propagates through space at a constant speed without changing its shape. The governing equation is

$$\partial_t u(x,t) + \beta \partial_x u(x,t) = 0, \tag{8}$$

where β denotes the advection speed. This task is useful for testing whether the model can preserve coherent spatiotemporal wave patterns, since the solution appears as diagonal stripe structures in the space-time domain. We form supervised samples $\{((x_i,t_i), u(x_i,t_i))\}_{i=1}^N$ from the PDEBench space-time grid and train the quantum model $f_\theta : (x,t) \mapsto \hat{u}_\theta(x,t)$ by minimizing the mean-squared error

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} \left| \hat{u}_\theta(x_i,t_i) - u(x_i,t_i) \right|^2 . \tag{9}$$

Figures 7, 8 and 4 show the reconstruction results for $\beta = 0.1$, 0.2 and 0.4. Across all advection speeds, the Basic Entangler and Strongly Entangling baselines fail to accurately reproduce the diagonal transport patterns, producing distorted or over-smoothed fields. In contrast, the proposed circuit closely matches the ground-truth stripe patterns for all tested values of β, indicating that the proposed entanglement design can effectively represent transport-dominated spatiotemporal fields.

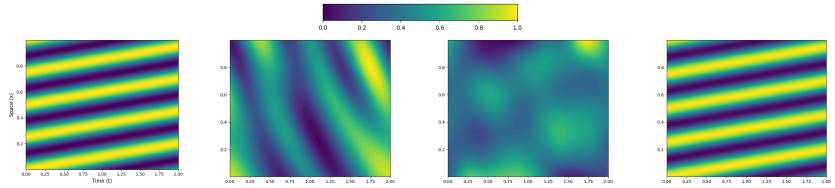


Figure 7: **Prediction results of the algorithms**. Advection equation ($\beta = 0.1$).

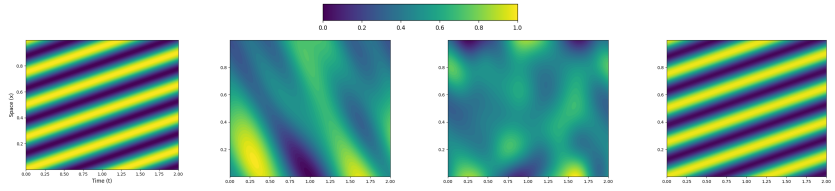


Figure 8: **Prediction results of the algorithms**. Advection equation ($\beta = 0.2$).

C.5 Adaptation to Environment Changes

We further test whether a model can learn a family of fields across propagation environments. An antenna phase vector $\boldsymbol{\phi} \in [0, 2\pi]^M$ with $M = 4$ is concatenated with the space-time coordinates, so that each input is $[x, t, \phi_1, \ldots, \phi_4]$, and the model learns the parametric field $\hat{E}_\theta(x, t; \boldsymbol{\phi})$ over $x, t \in [0, 1]$. The target is $E(x, t; \boldsymbol{\phi}) = \cos(\pi(t - x)) + A(\boldsymbol{\phi}) \cos(\pi(t + x) + \psi(\boldsymbol{\phi}))$, where $Ae^{j\psi} = \frac{1}{4}\sum_m e^{j\phi_m}$ is the effective array response, and every model is trained with the wave-equation objective of Table 14, with boundary and initial values taken from E. Each epoch resamples 2000 collocation points over $(x, t, \boldsymbol{\phi})$, 400 boundary points at $x \in \{0, 1\}$ and 400 initial points at $t = 0$, and every model is trained for 500 epochs with Adam under an exponential learning-rate decay of 0.99 every 50 steps after gradient clipping at global norm 1.0.

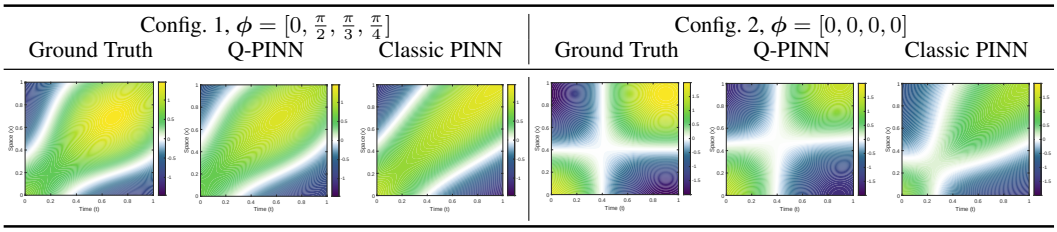
Table 14: Antenna-configuration field-learning setup. $\mathcal{L}_{\mathrm{PDE}}$ is the residual of $\partial_{tt}\hat{E}_\theta - \partial_{xx}\hat{E}_\theta$, and $\mathcal{L}_{\mathrm{BC}}$, $\mathcal{L}_{\mathrm{IC}}$ match E at $x \in \{0, 1\}$ and $E, \partial_t E$ at $t = 0$.

Component	Notation	Equation
Parametric field	E_θ	$E_\theta : (x, t, \boldsymbol{\phi}) \mapsto \hat{E}_\theta(x, t; \boldsymbol{\phi}), \quad x, t \in [0, 1],\ \boldsymbol{\phi} \in [0, 2\pi]^M$
Effective array response	$\Gamma(\boldsymbol{\phi})$	$\Gamma(\boldsymbol{\phi}) = \frac{1}{M}\sum_{m=1}^{M} e^{j\phi_m} = A(\boldsymbol{\phi})e^{j\psi(\boldsymbol{\phi})}$
Amplitude / phase	A, ψ	$A(\boldsymbol{\phi}) = \lvert\Gamma(\boldsymbol{\phi})\rvert, \quad \psi(\boldsymbol{\phi}) = \arg(\Gamma(\boldsymbol{\phi}))$
Analytical field	$E(x, t; \boldsymbol{\phi})$	$E(x, t; \boldsymbol{\phi}) = \cos\big(\pi(t - x)\big) + A(\boldsymbol{\phi}) \cos\big(\pi(t + x) + \psi(\boldsymbol{\phi})\big)$
Training objective	$\mathcal{L}$	$\mathcal{L}_{\mathrm{PDE}} + 10\,\mathcal{L}_{\mathrm{BC}} + 10\,\mathcal{L}_{\mathrm{IC}}$

The Q-PINN uses 6 qubits, encoding one input per qubit, and four re-uploading layers, each applying general single-qubit rotations, a star-graph CZ entangler, the encoding $R_X(fz)$ of the qubit's input z with a trainable frequency f, further rotations and a second star-graph entangler, followed by a scaled and shifted $\langle Z_0 \rangle$ readout (170 parameters). It is trained at learning rate 0.015 and retrained over three seeds, one of which reproduces the reported model within seed-level noise. The tanh PINN is a fully connected network with three hidden layers of width 128 (34,049 parameters) that receives $(\pi x, \pi t)$ and the phases in radians, as the Q-PINN does, and is trained at learning rate 10^{-3}. With inputs normalized to $[-1, 1]$ and learning rate 5×10^{-3}, the same network reaches 0.11 and 0.16 over three seeds. The tanh network of comparable size, with two hidden layers of 10 units (191 parameters), takes normalized inputs and the learning rate with the lowest training loss among $\{10^{-3}, 5 \times 10^{-3}, 1.5 \times 10^{-2}\}$ over five seeds.

Evaluation uses a 100×100 grid in (x, t) for each of six configurations, Config. 1 $[0, \pi/2, \pi/3, \pi/4]$ and Config. 2 $[0, 0, 0, 0]$ of Table 15 together with $[\pi/2, \pi/2, \pi/5, \pi/5]$, $[0, \pi/2, \pi, 3\pi/4]$, $[\pi/5, \pi/5, \pi/5, \pi/5]$ and $[0, \pi/7, 0, \pi/7]$, and reports the relative L^2 error $\|\hat{E}_\theta - E\|_F / \|E\|_F$. On Config. 1 and Config. 2, the Q-PINN reconstructs the field to relative L^2 errors of 0.35 ± 0.04 and 0.25 ± 0.01 over three seeds, while the tanh network of comparable size reaches 0.50 ± 0.08 and 0.61 ± 0.11 over five seeds (Holm $p \leq 0.022$). Over the six configurations the reported Q-PINN averages 0.33 and is more accurate than the tanh PINN on three of them (average 0.44).

Table 15: Field reconstruction under two antenna configurations.



D PROOFS OF LEMMAS AND PROPOSITIONS

Definition 1 (Baker-Campbell-Hausdorff). *Let A, B be elements of a Lie algebra for which the BCH series converges. Then there exists C such that*

$$e^A e^B = e^C, \tag{10}$$

where C is given by the BCH expansion

$$\begin{aligned} C =& A + B + \frac{1}{2}[A, B] \\ &+ \frac{1}{12}[A, [A, B]] + \frac{1}{12}[B, [B, A]] \\ &- \frac{1}{24}[B, [A, [A, B]]] + \cdots . \end{aligned} \tag{11}$$

D.1 PROOF OF HADAMARD'S LEMMA (LEMMA 1)

Proof. Define $F(t) = e^{tA} B e^{-tA}$. Then $F(0) = B$. Differentiate using the product rule:

$$F'(t) = A e^{tA} B e^{-tA} - e^{tA} B A e^{-tA} = [A, F(t)]. \tag{12}$$

Evaluating at $t = 0$ gives

$$F'(0) = [A, B]. \tag{13}$$

Differentiate again:

$$F''(t) = [A, F'(t)] = [A, [A, F(t)]] \quad \Rightarrow \quad F''(0) = [A, [A, B]]. \tag{14}$$

By induction,

$$F^{(k)}(t) = \mathrm{ad}_A^k(F(t)) \quad \Rightarrow \quad F^{(k)}(0) = \mathrm{ad}_A^k(B), \tag{15}$$

where $\mathrm{ad}_A^0(B) := B$ and $\mathrm{ad}_A^{k+1}(B) := [A, \mathrm{ad}_A^k(B)]$.

Apply the Taylor expansion of $F(t)$ about $t = 0$:

$$F(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} F^{(k)}(0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{ad}_A^k(B). \tag{16}$$

Setting $t = 1$ yields $e^A B e^{-A} = \sum_{k \geq 0} \frac{1}{k!} \mathrm{ad}_A^k(B)$, and the general-$t$ form follows. This completes the proof. □

D.2 PROOF OF PROPOSITION 1 (CLOSED INVARIANT SUBSPACE.)

Proof. Let $S(t) = e^{-\frac{it}{2}A}$, by the Hadamard's lemma,

$$e^{it/2A} B e^{-it/2A} = \sum_{k \geq 0} \frac{1}{k!} \left(\frac{it}{2}\right)^k \mathrm{ad}_A^k(B), \tag{17}$$

where $[A, B] = \mathrm{ad}_A(B) = -2i\omega C, \qquad [A, C] = \mathrm{ad}_A(C) = 2i\omega B.$

Using this, compute the first few nested commutators:

$$\mathrm{ad}_A(B) = -2i\omega C, \qquad \mathrm{ad}_A^2(B) = \mathrm{ad}_A(-2i\omega C) = -2i\omega(2i\omega B) = 4\omega^2 B,$$

$$\mathrm{ad}_A^3(B) = \mathrm{ad}_A(4\omega^2 B) = 4\omega^2(-2i\omega C) = -8i\omega^3 C, \qquad \mathrm{ad}_A^4(B) = 16\omega^4 B,$$

Every $k \geq 0$ is either even or odd, so setting $k = 2m$ and $k = 2m + 1$ with $m \geq 0$ splits the Hadamard series into two sums:

$$e^{\frac{it}{2}A} B e^{-\frac{it}{2}A} = \underbrace{\sum_{m \geq 0} \frac{1}{(2m)!} \left(\frac{it}{2}\right)^{2m} \mathrm{ad}_A^{2m}(B)}_{\text{Even}} + \underbrace{\sum_{m \geq 0} \frac{1}{(2m+1)!} \left(\frac{it}{2}\right)^{2m+1} \mathrm{ad}_A^{2m+1}(B)}_{\text{Odd}}.$$

Even terms. $i^{2m} = (-1)^m$ and $\mathrm{ad}_A^{2m}(B) = 2^{2m}\omega^{2m}B$:

$$\frac{1}{(2m)!} \cdot \frac{(-1)^m t^{2m}}{2^{2m}} \cdot 2^{2m}\omega^{2m} B = \frac{(-1)^m (\omega t)^{2m}}{(2m)!}\, B.$$

Odd terms. $\left(\frac{it}{2}\right)^{2m+1} = \frac{i\,(-1)^m t^{2m+1}}{2^{2m+1}}$ and $\mathrm{ad}_A^{2m+1}(B) = -i\,2^{2m+1}\omega^{2m+1}C$:

$$\frac{1}{(2m+1)!} \cdot \frac{i\,(-1)^m t^{2m+1}}{2^{2m+1}} \cdot \left(-i\,2^{2m+1}\omega^{2m+1}\right) C = \frac{(-1)^m (\omega t)^{2m+1}}{(2m+1)!}\, C.$$

Resumming the two series gives the first identity:

$$e^{\frac{it}{2}A} B e^{-\frac{it}{2}A} = B \sum_{m\geq 0} \frac{(-1)^m(\omega t)^{2m}}{(2m)!} + C \sum_{m\geq 0} \frac{(-1)^m(\omega t)^{2m+1}}{(2m+1)!} \tag{18}$$

$$= B\cos(\omega t) + C\sin(\omega t). \tag{19}$$

We can make similar case for

$$e^{\frac{it}{2}A} C e^{-\frac{it}{2}A} = C\cos(\omega t) - B\sin(\omega t). \tag{20}$$

This completes the proof. □

D.3 Proof of Proposition 2

Proof. Each qubit j is assigned one of $\hat{I}_j$, $\hat{X}_j$, $E_j^+ = \hat{Z}_j + i\hat{Y}_j$, $E_j^- = \hat{Z}_j - i\hat{Y}_j$, and $\langle A, B\rangle = 2^{-N}\,\mathrm{Tr}(A^\dagger B)$.

Step 1 (single-qubit eigen-operators). Only the j-th term of $G = \sum_k \hat{X}_k$ fails to commute with an operator supported on qubit j. From $[X, Z] = -2iY$ and $[X, Y] = 2iZ$,

$$\mathrm{ad}_G\big(E_j^\pm\big) = \big[\hat{X}_j,\ \hat{Z}_j \pm i\hat{Y}_j\big] = -2i\hat{Y}_j \pm i\big(2i\hat{Z}_j\big) = \mp 2\big(\hat{Z}_j \pm i\hat{Y}_j\big) = \mp 2\,E_j^\pm, \tag{21}$$

while $\mathrm{ad}_G(\hat{I}_j) = \mathrm{ad}_G(\hat{X}_j) = 0$.

Step 2 (eigenvalue of a product). By the Jacobi identity ad_G is a derivation, $\mathrm{ad}_G(AB) = \mathrm{ad}_G(A)B + A\,\mathrm{ad}_G(B)$. The factors of E_c act on pairwise disjoint qubits, so applying the derivation rule across them and using Step 1,

$$\mathrm{ad}_G(E_c) = \Big(\sum_{j\in S_+}(-2) \;+\; \sum_{j\in S_-}(+2) \;+\; \sum_{j\in T} 0\Big) E_c = -2\big(|S_+| - |S_-|\big)E_c = -2\,\omega(c)\,E_c. \tag{22}$$

Step 3 (orthogonality, norms, and completeness). On a single qubit, with $\langle A, B\rangle_1 = \frac{1}{2}\,\mathrm{Tr}(A^\dagger B)$,

$$\langle \hat{I}, \hat{I}\rangle_1 = \langle \hat{X}, \hat{X}\rangle_1 = 1, \qquad \langle E^+, E^+\rangle_1 = \langle E^-, E^-\rangle_1 = 2, \tag{23}$$

where the last equality uses $(E^+)^\dagger E^+ = (Z - iY)(Z + iY) = 2I + 2X$. All cross terms vanish: $\mathrm{Tr}(X) = \mathrm{Tr}(Y) = \mathrm{Tr}(Z) = 0$ kills $\langle \hat{I}, \cdot\rangle_1$; $XZ = -iY$ and $XY = iZ$ are traceless, so $\langle \hat{X}, E^\pm\rangle_1 = 0$; and $\{Z, Y\} = 0$ gives $(E^+)^\dagger E^- = (Z - iY)^2 = -i\{Z, Y\} = 0$, so $\langle E^+, E^-\rangle_1 = 0$. Hence $\{\hat{I}, \hat{X}, E^+, E^-\}$ is an orthogonal basis of $\mathcal{B}(\mathcal{H}_1)$.

The inner product factorises over tensor products, $\big\langle \bigotimes_j A_j,\ \bigotimes_j B_j\big\rangle = \prod_j \langle A_j, B_j\rangle_1$, so the 4^N operators $\{E_c\}$ are mutually orthogonal with

$$\|E_c\|^2 = 2^{|S_+|+|S_-|} = 2^{m(c)} \neq 0. \tag{24}$$

Being $4^N = \dim \mathcal{B}(\mathcal{H})$ in number, they form an orthogonal basis.

Step 4 (Hermitian pairs and closure). The factors commute and $(E_j^\pm)^\dagger = E_j^\mp$, so $E_c^\dagger = E_{\bar{c}}$ with $\bar{c} = (S_-, S_+, T)$; thus $\omega(\bar{c}) = -\omega(c)$, $m(\bar{c}) = m(c)$, $E_c^\dagger$ is again a basis element, and $\mathrm{ad}_G(E_c^\dagger) = +2\omega(c)E_c^\dagger$. Therefore $\mathcal{X}_c^\dagger = \mathcal{X}_c$, $\mathcal{Y}_c^\dagger = \mathcal{Y}_c$, and

$$[G, \mathcal{X}_c] = -2\omega\big(E_c - E_c^\dagger\big) = -2i\omega\,\mathcal{Y}_c, \qquad [G, \mathcal{Y}_c] = \tfrac{1}{i}(-2\omega)\big(E_c + E_c^\dagger\big) = 2i\omega\,\mathcal{X}_c, \tag{25}$$

which is the hypothesis of Proposition 1. For $\omega(c) > 0$ we have $S_+ \neq S_-$, so E_c and $E_c^\dagger$ are *distinct* basis elements and hence orthogonal, giving

$$\|\mathcal{X}_c\|^2 = \|E_c\|^2 + \|E_c^\dagger\|^2 = 2^{m(c)+1} = \|\mathcal{Y}_c\|^2, \qquad \langle \mathcal{X}_c, \mathcal{Y}_c \rangle = 0. \tag{26}$$

The same argument gives $\mathcal{X}_c, \mathcal{Y}_c \perp \mathcal{X}_{c'}, \mathcal{Y}_{c'}$ whenever $c \neq c'$ with $\omega(c) = \omega(c') > 0$.

Step 5 (decomposition). Since $\operatorname{span}_{\mathbb{C}}\{E_c, E_c^\dagger\} = \operatorname{span}_{\mathbb{C}}\{\mathcal{X}_c, \mathcal{Y}_c\}$, grouping the basis $\{E_c\}$ by $\omega(c)$ yields the orthogonal decomposition

$$\mathcal{B}(\mathcal{H}) = \mathcal{B}_0 \oplus \bigoplus_{\omega=1}^{N} \mathcal{B}_\omega, \qquad \mathcal{B}_0 = \operatorname{span}\{E_c : \omega(c) = 0\} = \ker \operatorname{ad}_G, \tag{27}$$

with $\mathcal{B}_\omega = \operatorname{span}\{\mathcal{X}_c, \mathcal{Y}_c : \omega(c) = \omega\}$. Finally $(E_j^\pm)^2 = \hat{Z}_j^2 - \hat{Y}_j^2 \pm i\{\hat{Z}_j, \hat{Y}_j\} = 0$, so each qubit carries at most one $E^\pm$ factor and $|\omega(c)| \leq m(c) \leq N$. This completes the proof. □

D.4 Proof of Proposition 3

Proof. Let $U(x) = W_L S(x) W_R$ with $S(x) = e^{-ixG/2}$ and $G = \sum_j \hat{X}_j$. Using $\langle\psi|B|\psi\rangle = \operatorname{Tr}(|\psi\rangle\langle\psi|B)$ and cyclicity,

$$f(x) = \langle\psi|U(x)^\dagger M U(x)|\psi\rangle = \operatorname{Tr}\Big(\underbrace{W_R \rho_0 W_R^\dagger}_{\rho}\, S(x)^\dagger \underbrace{W_L^\dagger M W_L}_{O}\, S(x)\Big) = \operatorname{Tr}\big(\rho\, S(x)^\dagger O S(x)\big). \tag{28}$$

Step 1 (trace to inner product). With $\langle A, B\rangle = 2^{-N}\operatorname{Tr}(A^\dagger B)$ and $\rho = \rho^\dagger$,

$$\operatorname{Tr}(\rho B) = \operatorname{Tr}(\rho^\dagger B) = 2^N \langle \rho, B\rangle, \quad f(x) = 2^N \big\langle \rho,\ S(x)^\dagger O S(x)\big\rangle. \tag{29}$$

Because ρ, O, $\mathcal{X}_{\omega,r}$ and $\mathcal{Y}_{\omega,r}$ are all Hermitian, every inner product below is real.

Step 2 (orthogonal projection onto one plane). By Proposition 2 the pairs $\{(\mathcal{X}_{\omega,r}, \mathcal{Y}_{\omega,r})\}$ together with $\mathcal{B}_0$ are mutually orthogonal and span $\mathcal{B}(\mathcal{H})$, with $\|\mathcal{X}_{\omega,r}\|^2 = \|\mathcal{Y}_{\omega,r}\|^2 = 2^{m_{\omega,r}+1}$. The basis is orthogonal but *not* normalised, so the projection carries the norm factor:

$$\Pi_{\omega,r}(O) = \frac{\langle O, \mathcal{X}_{\omega,r}\rangle\, \mathcal{X}_{\omega,r} + \langle O, \mathcal{Y}_{\omega,r}\rangle\, \mathcal{Y}_{\omega,r}}{2^{\,m_{\omega,r}+1}}, \qquad O = \Pi_0 O + \sum_{\omega=1}^{N}\sum_{r=1}^{d_\omega} \Pi_{\omega,r}(O), \tag{30}$$

where Π_0 is the orthogonal projection onto $\mathcal{B}_0$.

Step 3 (the $\omega = 0$ term). $\Pi_0 O \in \mathcal{B}_0 = \ker \operatorname{ad}_G$ commutes with G and therefore with $S(x)$, so $S(x)^\dagger(\Pi_0 O)S(x) = \Pi_0 O$ and this term contributes $2^N\langle \rho, \Pi_0 O\rangle = \operatorname{Tr}(\rho\,\Pi_0 O)$, independent of x.

Step 4 (rotation within a plane). Proposition 1 gives $S^\dagger \mathcal{X} S = \cos(\omega x)\mathcal{X} + \sin(\omega x)\mathcal{Y}$ and $S^\dagger \mathcal{Y} S = \cos(\omega x)\mathcal{Y} - \sin(\omega x)\mathcal{X}$. Substituting into $\Pi_{\omega,r}(O)$ and collecting the $\mathcal{X}$ and $\mathcal{Y}$ components,

$$\begin{aligned} S^\dagger \Pi_{\omega,r}(O) S = 2^{-(m_{\omega,r}+1)} \Big[&\big(\langle O, \mathcal{X}\rangle \cos\omega x - \langle O, \mathcal{Y}\rangle \sin\omega x\big)\mathcal{X} \\ &+ \big(\langle O, \mathcal{X}\rangle \sin\omega x + \langle O, \mathcal{Y}\rangle \cos\omega x\big)\mathcal{Y}\Big]. \end{aligned} \tag{31}$$

Step 5 (contraction with ρ). Applying $2^N\langle \rho, \cdot\rangle$ to the previous line,

$$\begin{aligned} f_{\omega,r}(x) = 2^{\,N-m_{\omega,r}-1} \Big[&\langle \rho, \mathcal{X}\rangle \big(\langle O, \mathcal{X}\rangle \cos\omega x - \langle O, \mathcal{Y}\rangle \sin\omega x\big) \\ &+ \langle \rho, \mathcal{Y}\rangle \big(\langle O, \mathcal{X}\rangle \sin\omega x + \langle O, \mathcal{Y}\rangle \cos\omega x\big)\Big], \end{aligned} \tag{32}$$

which is precisely the bilinear form

$$f_{\omega,r}(x) = 2^{\,N-m_{\omega,r}-1} \begin{bmatrix} \langle \rho, \mathcal{X}_{\omega,r}\rangle \\ \langle \rho, \mathcal{Y}_{\omega,r}\rangle \end{bmatrix}^\top \begin{bmatrix} \cos(\omega x) & -\sin(\omega x) \\ \sin(\omega x) & \cos(\omega x) \end{bmatrix} \begin{bmatrix} \langle O, \mathcal{X}_{\omega,r}\rangle \\ \langle O, \mathcal{Y}_{\omega,r}\rangle \end{bmatrix}. \tag{33}$$

Summing over all (ω, r) and adding the constant term of Step 3 gives equation 5. □

D.5 Proof of Lemma 2

Proof. Define the local Ising generator on qubit n as

$$H_n := \sum_{k \in \mathcal{N}(n)} \gamma_{nk} Z_n Z_k. \tag{34}$$

Since Z_n acts only on qubit n and each Z_k acts on $k \neq n$, all factors commute and

$$H_n = \sum_{k \in \mathcal{N}(n)} \gamma_{nk}\, Z_n Z_k = Z_n \sum_{k \in \mathcal{N}(n)} \gamma_{nk} Z_k =: Z_n \Phi_n. \tag{35}$$

Therefore,

$$U^{\text{ent}} = \exp\left(-\frac{i}{2} H_n\right) = \exp\left(-\frac{i}{2} Z_n \Phi_n\right). \tag{36}$$

Throughout, Φ_n commutes with X_n, Y_n, Z_n because it is supported only on $\mathcal{N}(n)$.

Conjugation of Z_n. Because $[Z_n \Phi_n, Z_n] = 0$, conjugation leaves Z_n invariant:

$$(U^{\text{ent}})^\dagger Z_n U^{\text{ent}} = Z_n. \tag{37}$$

Conjugation of X_n. Let $A := \frac{i}{2} Z_n \Phi_n$, so that $(U^{\text{ent}})^\dagger X_n U^{\text{ent}} = e^A X_n e^{-A}$. Using the Baker–Campbell–Hausdorff (BCH) expansion in adjoint form,

$$e^A X_n e^{-A} = \sum_{m=0}^{\infty} \frac{1}{m!} \operatorname{ad}_A^m(X_n), \qquad \operatorname{ad}_A(B) := [A, B], \tag{38}$$

and the Pauli commutators $[Z_n, X_n] = 2iY_n$ and $[Z_n, Y_n] = -2iX_n$, together with $[\Phi_n, X_n] = [\Phi_n, Y_n] = 0$, yield

$$\operatorname{ad}_{Z_n\Phi_n}(X_n) = [Z_n \Phi_n, X_n] = [Z_n, X_n]\Phi_n = 2iY_n \Phi_n, \tag{39}$$

$$\operatorname{ad}^2_{Z_n\Phi_n}(X_n) = [Z_n \Phi_n, 2iY_n \Phi_n] = 2i[Z_n, Y_n]\Phi_n^2 = 4X_n \Phi_n^2. \tag{40}$$

By induction, for all $m \geq 0$,

$$\operatorname{ad}^{2m}_{Z_n\Phi_n}(X_n) = 2^{2m} X_n \Phi_n^{2m}, \qquad \operatorname{ad}^{2m+1}_{Z_n\Phi_n}(X_n) = 2^{2m+1} i Y_n \Phi_n^{2m+1}. \tag{41}$$

Since $A = \frac{i}{2} Z_n \Phi_n$, it follows that

$$\begin{aligned}
(U^{\text{ent}})^\dagger X_n U^{\text{ent}} &= \sum_{m=0}^{\infty} \frac{1}{(2m)!} \left(\frac{i}{2}\right)^{2m} 2^{2m} X_n \Phi_n^{2m} + \sum_{m=0}^{\infty} \frac{1}{(2m+1)!} \left(\frac{i}{2}\right)^{2m+1} 2^{2m+1} i Y_n \Phi_n^{2m+1} \\
&= X_n \sum_{m=0}^{\infty} \frac{(-1)^m \Phi_n^{2m}}{(2m)!} - Y_n \sum_{m=0}^{\infty} \frac{(-1)^m \Phi_n^{2m+1}}{(2m+1)!} \\
&= X_n \cos \Phi_n - Y_n \sin \Phi_n,
\end{aligned} \tag{42}$$

where $\cos \Phi_n$ and $\sin \Phi_n$ are defined by their convergent power series.

Conjugation of Y_n. The same argument gives

$$(U^{\text{ent}})^\dagger Y_n U^{\text{ent}} = Y_n \cos \Phi_n + X_n \sin \Phi_n. \tag{43}$$

This completes the proof. □

D.6 Proof of Corollary 1

We prove a sharper form, where *almost every* means outside a closed set of measure zero. Let the local blocks be general single-qubit rotations, ρ_0 a pure product state and $\omega \geq 1$. **(i)** If $\mathbf{b}_{\omega,r} \neq 0$ for some r at some $\boldsymbol{\theta}^{(L)}$, then $c_\omega(\boldsymbol{\Theta}) \neq 0$ for almost every $\boldsymbol{\Theta}$, and otherwise $c_\omega \equiv 0$. **(ii)** For $W_L = W_L^{\text{post}} U_{\text{ent}} W_L^{\text{pre}}$ with readout $\hat{Z}_n$ and every $\sin \gamma_{nk} \neq 0$, (i) applies to every $\omega \leq \deg(n)+1$,

and $c_\omega \equiv 0$ for $\omega > \deg(n) + 1$. **(iii)** At $\omega = \deg(n) + 1$ a single plane carries c_ω, and any product state ρ limits the amplitude there to $2^{-\deg(n)}$. With $|\gamma_{nk}| = \pi/2$ and $W_R = W_R^{\text{post}} U_{\text{ent}} W_R^{\text{pre}}$, the circuit can output $f(x) = \cos(\omega x)$ for any single $\omega \le \deg(n) + 1$. Throughout, $L = \{n\} \cup \mathcal{N}(n)$ and Q_ω is the orthogonal projection onto $\operatorname{span}\{E_c : \omega(c) = \omega\}$.

Proof of (i). By Proposition 2, $S(x)^\dagger E_c S(x) = e^{-i\omega(c)x} E_c$, so expanding O in $\{E_c\}$ and using $\mathcal{X}_c = E_c + E_c^\dagger$, $\mathcal{Y}_c = -i(E_c - E_c^\dagger)$ gives

$$f(x) = \sum\nolimits_{|\omega| \le N} e^{-i\omega x} \operatorname{Tr}\big(\rho\, Q_\omega(O)\big), \qquad \|\mathbf{b}_{\omega,r}\| = 2\,|\langle E_c, O\rangle|, \tag{44}$$

so $|c_\omega| = |\operatorname{Tr}(\rho\, Q_\omega(O))|$ because f is real, likewise $\|\mathbf{a}_{\omega,r}\| = 2|\langle E_c, \rho\rangle|$, and $\mathbf{b}_{\omega,r} \neq 0$ for some r exactly when $Q_\omega(O) \neq 0$. If this never happens, $c_\omega \equiv 0$. Otherwise let $Q_\star = Q_\omega(O) \neq 0$ at some $\boldsymbol{\theta}_\star^{(L)}$ and write $W_R = V W_R^{\text{pre}}$. The reachable states include $V\sigma V^\dagger$ for every pure product state σ, and these span $\mathcal{B}(\mathcal{H})$, while $\rho \mapsto \operatorname{Tr}(\rho\, Q_\star)$ is a nonzero functional, equal to $2^N \|Q_\star\|^2$ at $\rho = Q_\star^\dagger$, so some $\boldsymbol{\theta}^{(R)}$ gives $c_\omega \neq 0$. Each gate $e^{-i\theta P/2} = \cos(\theta/2) I - i \sin(\theta/2) P$, including every factor of U_{ent}, is a trigonometric polynomial in its half-angle, so $|c_\omega(\boldsymbol{\Theta})|^2$ is a real trigonometric polynomial that is not identically zero. Such a polynomial has finitely many zeros per period in one variable, and by induction on the number of variables with Fubini's theorem it vanishes only on a closed set of measure zero.

Proof of (ii). By Appendix E.3, O is supported on L, which gives equation 84 and $c_\omega \equiv 0$ for $\omega > \deg(n) + 1$. For $1 \le \omega \le \deg(n) + 1$ and $\deg(n) \ge 1$, the case $\deg(n) = 0$ being immediate, choose $K \subseteq \mathcal{N}(n)$ with $|K| = \deg(n) + 1 - \omega$, let W_L^{post} give $\mu \neq 0$, and let W_L^{pre} map $\hat{Z}_k$ to $\hat{X}_k$ on K and act trivially elsewhere. Since every string of equation 83 carries $\hat{I}$ or $\hat{Z}$ on each neighbour, the only string of O that overlaps E_c for $c = (L \setminus K, \emptyset, K)$, of frequency ω, is $\hat{Y}_n \hat{Z}_{\mathcal{N}(n)\setminus K} \hat{X}_K$, the image of the top string. Hence $\|\mathbf{b}_{\omega,r}\| = 2|\mu| \prod_k |\sin \gamma_{nk}| \neq 0$ and (i) applies.

Proof of (iii). Every configuration with $\langle E_c, O\rangle \neq 0$ lies inside L, so $\omega(c) = |L|$ forces $S_+ = L$, only this plane contributes, and $|c_{\deg(n)+1}| = 2^{-|L|} |\langle E_c, O\rangle|\, |\operatorname{Tr}(\rho E_c)|$. Since $\hat{Z}_j + i\hat{Y}_j = 2\,|-\rangle\langle +|$, E_c has trace norm 2^N, so $|\langle E_c, O\rangle| \le \|O\| = 1$, and a product state gives $|\operatorname{Tr}(\rho E_c)| = \prod_{j \in L} |\operatorname{Tr}(\rho_j(\hat{Z}_j + i\hat{Y}_j))| \le 1$, which bounds the amplitude $2|c_{\deg(n)+1}|$ by $2^{1-|L|} = 2^{-\deg(n)}$. For attainment, let $|\gamma_{nk}| = \pi/2$, so $\cos\Phi_n$ and $\sin\Phi_n$ reduce to 0 or $\pm\hat{Z}_{\mathcal{N}(n)}$, and fix $L' \subseteq L$ with $n \in L'$ and $|L'| = \omega$. Choosing $M_1 = \hat{X}_n$ and a W_L^{pre} that rotates the factor on n to $\hat{Z}_n$ and $\hat{Z}_k$ to $\hat{X}_k$ for $k \notin L'$ gives $O = \pm\hat{Z}_{L'} \hat{X}_{\mathcal{N}(n)\setminus L'}$. On the state side, W_R^{pre} prepares $|+\rangle$ on L' and $|0\rangle$ elsewhere, U_{ent} turns L' into a star graph state up to $\hat{Z}$ rotations, and W_R^{post} undoes these rotations, applies a Hadamard gate to qubit n and rotates the qubits outside L' to $|+\rangle$, which yields $(|+\rangle^{\otimes\omega} + |-\rangle^{\otimes\omega})/\sqrt{2}$ on L'. The encoder gives $|\pm\rangle^{\otimes\omega}$ the phases $e^{\mp i\omega x/2}$ and $\hat{Z}_{L'}$ exchanges them, so $f(x) = \pm\cos(\omega x)$, with the sign set by a π rotation in W_L^{post}. At $\omega = \deg(n) + 1$ no output of this block does better, since averaging f over the shifts $x \mapsto x + 2\pi j/|L|$ keeps only the frequencies 0 and $|L|$ and leaves an amplitude of at most 1.

Remarks. CZ differs from the Ising form at $|\gamma_{nk}| = \pi/2$ only by single-qubit $\hat{Z}$ rotations, which the adjacent local blocks absorb. When x and t are encoded on disjoint qubit sets V_x and V_t, each E_c carries a frequency vector and the proof of (i) holds verbatim. The construction of (ii), which also uses the string $\hat{X}_n \hat{Z}_{\mathcal{N}(n)}$ when $n \in T$, then reaches every vector with $|\omega_x| \le |L \cap V_x|$ and $|\omega_t| \le |L \cap V_t|$, which for the 14-qubit star of our advection experiment is $|\omega_x|, |\omega_t| \le 7$.

E Detailed Explanation

E.1 Defining Closed-Invariant Subspaces

Equipped with Proposition 1, our objective is to represent the quantum model's output as Fourier series representation in terms of the operators. The question we can ask is how we can define such set of operator pairs. For this, we employ the method of *eigen-operators of the adjoint map* (Gardiner & Zoller, 2004). This technique builds operator pairs from eigen-operators that close under Lie commutation with the generator. To construct the target operators $\mathcal{X}$ and $\mathcal{Y}$ from such eigen-operators, we begin by establishing the fundamental building blocks on the single-qubit. Additionally, the key observation is that the encoding unitary $S(x)$ is often constructed from a single Pauli generator type; in our case, it is generated by Pauli generator X. In this light, we first define on each qubit n the local eigen-operators as $E_n^+ = Z_n + iY_n$, $E_n^- = Z_n - iY_n$, which satisfy $[X_n, E_n^+] = -2E_n^+$, $[X_n, E_n^-] = 2E_n^-$.

For the global multi-qubit generator $G = \sum_{n=1}^N \hat{X}_n$, we apply tensor product to these local eigen-operators over pairwise disjoint index sets $S_+, S_-, T \subseteq \{1, \dots, N\}$. Because $\hat{X}_k$ commutes with G just as I_k does, a site outside $S_+ \cup S_-$ may carry either factor, and both must be retained for the resulting family to span $\mathcal{B}(\mathcal{H})$. Specifically, define

$$E_{S_+,S_-,T} \triangleq \bigotimes_{n\in S_+} E_n^+ \otimes \bigotimes_{m\in S_-} E_m^- \otimes \bigotimes_{t\in T} X_t \otimes \bigotimes_{k\notin S_+\cup S_-\cup T} I_k. \tag{45}$$

Then, if $n \in S_+$, the commutation is described as

$$[\hat{X}_n, E_{S_+,S_-,T}] = I^{\otimes(n-1)} \otimes [X_n, E_n^+] \otimes I^{\otimes(N-n)} = (-2)E_{S_+,S_-,T}, \tag{46}$$

and likewise $[\hat{X}_m, E_{S_+,S_-,T}] = (+2)E_{S_+,S_-,T}$ for $m \in S_-$, while a site $j \in T$ or a site carrying the identity contributes nothing, since $[\hat{X}_j, \hat{X}_j] = 0$ and $[\hat{X}_j, \hat{I}_j] = 0$. Thus, Lie commutation of G with $E_{S_+,S_-,T}$ can be as

$$[G, E_{S_+,S_-,T}] = \sum\nolimits_{j=1}^N [\hat{X}_j, E_{S_+,S_-,T}] = \sum_{j\in S_+} -2E_{S_+,S_-,T} + \sum_{j\in S_-} 2E_{S_+,S_-,T} \tag{47}$$

$$= -2(|S_+| - |S_-|)E_{S_+,S_-,T} = -2\omega E_{S_+,S_-,T}. \tag{48}$$

The result shows that the difference of the cardinality of the disjoint sets determines $\omega = |S_+| - |S_-|$.

Furthermore, note that there could be many combinations of disjoint sets (S_+, S_-, T) that yield the same ω. We express such set as

$$\mathcal{S}_\omega \triangleq \Big\{(S_+, S_-, T) : S_+, S_-, T \text{ pairwise disjoint},\ S_+, S_-, T \subseteq \{1, \dots, N\},\ |S_+| - |S_-| = \omega\Big\}. \tag{49}$$

Denote $d_\omega = |\mathcal{S}_\omega|$, the cardinality of equation 49. This cardinality has a closed form. Each qubit independently carries one of I_k, X_k, E_k^+, E_k^-, contributing z^0, z^0, z^{+1}, z^{-1} respectively when the frequency is tracked by a formal variable z, so

$$\sum_{(S_+,S_-,T)} z^{\omega(S_+,S_-,T)} = \left(2 + z + z^{-1}\right)^N = z^{-N}(1+z)^{2N}, \tag{50}$$

and reading off the coefficient of z^ω gives

$$d_\omega = \binom{2N}{N+\omega}, \qquad \omega = -N, \dots, N. \tag{51}$$

In particular $d_0 = \binom{2N}{N} = \dim \mathcal{B}_0$, and $d_0 + 2\sum_{\omega=1}^N d_\omega = 4^N$, confirming that the family exhausts $\mathcal{B}(\mathcal{H})$. The extreme planes $|\omega| = N$ are attained only by $S_- = T = \varnothing$, so $d_{\pm N} = 1$; for $N = 2$ one has $d_1 = \binom{4}{3} = 4$ and $d_2 = \binom{4}{4} = 1$. We choose an enumeration $\{(S_+^{(r)}, S_-^{(r)}, T^{(r)})\}_{r=1}^{d_\omega}$ of $\mathcal{S}_\omega$ and define $E_{\omega,r} \triangleq E_{S_+^{(r)},S_-^{(r)},T^{(r)}}$. Accordingly, the exact operators can be formulated as

$$\mathcal{X}_{\omega,r} = E_{\omega,r} + E_{\omega,r}^\dagger, \quad \mathcal{Y}_{\omega,r} = \frac{1}{i}(E_{\omega,r} - E_{\omega,r}^\dagger). \tag{52}$$

Appendix D.3 proves that the operators in equation 52 satisfy equation 1, that pairs belonging to distinct (ω, r) are mutually orthogonal, and that $\|\mathcal{X}_{\omega,r}\|^2 = \|\mathcal{Y}_{\omega,r}\|^2 = 2^{m_{\omega,r}+1}$, as used in Proposition 3.

Furthermore, we note that high-frequency invariant-plane operators are not made purely of single-qubit Pauli string. For instance, for a configuration with $S_+ = S \subseteq \{1, \dots, N\}$ and $S_- = T = \varnothing$, the operators $\mathcal{X}_{\omega,r}$ and $\mathcal{Y}_{\omega,r}$ contain mixed multi-qubit Pauli strings supported on S for j-th qubit, including terms of the form $\hat{Y}_j \otimes \hat{Z}_{S\setminus\{j\}}$, $j \in S$, as well as higher-order mixtures with multiple Y factors. This reflects that higher ω components are associated with multi-qubit operator support, rather than purely single-qubit Pauli terms.

E.2 Complete 2-Qubit System

Consider an encoder with generator $G = \hat{X}_1 + \hat{X}_2$ and a $\hat{Z}_1$ measurement on qubit 1. Throughout, for any single-qubit operator $A \in \mathbb{C}^{2\times 2}$ we denote its N-qubit embedding by

$$\hat{A}_n \triangleq I^{\otimes(n-1)} \otimes A \otimes I^{\otimes(N-n)}. \tag{53}$$

In particular, for $N = 2$ we have $\hat{A}_1 = A \otimes I$ and $\hat{A}_2 = I \otimes A$.

We adopt the proposed decomposition

$$W_L = W_L^{\text{post}}\, U_{\text{ent}}\, W_L^{\text{pre}}, \tag{54}$$

where $W_L^{\text{post}} = W_{L,1}^{\text{post}} \otimes W_{L,2}^{\text{post}}$ and $W_L^{\text{pre}} = W_{L,1}^{\text{pre}} \otimes W_{L,2}^{\text{pre}}$. The Ising entangler is

$$U_{\text{ent}} = \exp\left(-\frac{i}{2}\gamma\, \hat{Z}_1\hat{Z}_2\right). \tag{55}$$

For $N = 2$, the frequency index is determined by $\omega = |S_+| - |S_-|$ for pairwise disjoint sets $S_+, S_-, T \subseteq \{1, 2\}$. The highest frequency is $\omega = 2$, which occurs only for $(S_+, S_-, T) = (\{1, 2\}, \varnothing, \varnothing)$, so $d_2 = 1$ and $r = 1$ in this case.

Using ad_G, an explicit eigen-operator is

$$E_{2,1} = (\hat{Z}_1 + i\hat{Y}_1)(\hat{Z}_2 + i\hat{Y}_2), \tag{56}$$

which yields the invariant operators

$$\mathcal{X}_{2,1} = E_{2,1} + E_{2,1}^\dagger = 2\,(\hat{Z}_1\hat{Z}_2 - \hat{Y}_1\hat{Y}_2), \tag{57}$$

$$\mathcal{Y}_{2,1} = \frac{1}{i}(E_{2,1} - E_{2,1}^\dagger) = 2\,(\hat{Z}_1\hat{Y}_2 + \hat{Y}_1\hat{Z}_2). \tag{58}$$

In particular, the $\omega = 2$ plane contains the mixed Pauli strings $\hat{Y}_1\hat{Z}_2$ and $\hat{Z}_1\hat{Y}_2$.

Let the full circuit be

$$U(x) = (W_{L,1}^{\text{post}} \otimes W_{L,2}^{\text{post}})\, U_{\text{ent}}\, (W_{L,1}^{\text{pre}} \otimes W_{L,2}^{\text{pre}})\, (R_{X_1}(x) \otimes R_{X_2}(x)), \tag{59}$$

and the model output is

$$f(x) = \langle 00|U(x)^\dagger \hat{Z}_1\, U(x)|00\rangle. \tag{60}$$

Write the data-encoding as

$$S(x) = e^{-ixG/2}, \tag{61}$$

and write the x-independent trainable part as

$$W_L = (W_{L,1}^{\text{post}} \otimes W_{L,2}^{\text{post}})\, U_{\text{ent}}\, (W_{L,1}^{\text{pre}} \otimes W_{L,2}^{\text{pre}}), \tag{62}$$

so that $U(x) = W_L S(x)$.

Consider the conjugated observable

$$M_1 := (W_L^{\text{post}})^\dagger \hat{Z}_1 W_L^{\text{post}}. \tag{63}$$

Since $W_L^{\text{post}} = W_{L,1}^{\text{post}} \otimes W_{L,2}^{\text{post}}$ and $\hat{Z}_1 = Z \otimes I$, only $W_{L,1}^{\text{post}}$ affects the conjugation. Parameterize

$$W_{L,1}^{\text{post}} = \exp\left(-\frac{i}{2}\beta\, \mathbf{n}\cdot\boldsymbol{\sigma}\right), \qquad \boldsymbol{\sigma} = (X, Y, Z), \quad \|\mathbf{n}\| = 1, \tag{64}$$

and define the single-qubit generator $A := i\frac{\beta}{2}\mathbf{n}\cdot\boldsymbol{\sigma}$. Embed it to the two-qubit space as $\hat{A}_1 := A \otimes I$ so that

$$M_1 = e^{\hat{A}_1}\, \hat{Z}_1\, e^{-\hat{A}_1}. \tag{65}$$

Applying Hadamard's lemma and evaluating commutators on qubit 1 yields

$$M_1 = m_x \hat{X}_1 + m_y \hat{Y}_1 + m_z \hat{Z}_1, \qquad m_x^2 + m_y^2 + m_z^2 = 1. \tag{66}$$

Now conjugate through the entangler:

$$M_2 := U_{\text{ent}}^\dagger M_1 U_{\text{ent}}. \tag{67}$$

$$\begin{aligned} M_1 &= (W_L^{\text{post}})^\dagger \hat{Z}_1 W_L^{\text{post}} \\ &= \big((W_{L,1}^{\text{post}})^\dagger \otimes (W_{L,2}^{\text{post}})^\dagger\big)(Z \otimes I)\big(W_{L,1}^{\text{post}} \otimes W_{L,2}^{\text{post}}\big) \\ &= \big((W_{L,1}^{\text{post}})^\dagger Z\, W_{L,1}^{\text{post}}\big) \otimes \big((W_{L,2}^{\text{post}})^\dagger I\, W_{L,2}^{\text{post}}\big) \\ &= \big((W_{L,1}^{\text{post}})^\dagger Z\, W_{L,1}^{\text{post}}\big) \otimes I. \end{aligned} \tag{68}$$

Applying Hadamard's lemma gives

$$e^A Z e^{-A} = \sum_{k=0}^{\infty} \frac{1}{k!}\, \mathrm{ad}_A^k(Z), \qquad \mathrm{ad}_A(B) = [A, B]. \tag{69}$$

Using the Pauli commutator identity

$$[\mathbf{u}\cdot\boldsymbol{\sigma},\, \mathbf{v}\cdot\boldsymbol{\sigma}] = 2i(\mathbf{u}\times\mathbf{v})\cdot\boldsymbol{\sigma}, \tag{70}$$

with $\mathbf{e}_z = (0, 0, 1)$ (so $Z = \mathbf{e}_z \cdot \boldsymbol{\sigma}$), we obtain

$$[A, Z] = i\frac{\beta}{2}[\mathbf{n}\cdot\boldsymbol{\sigma},\, \mathbf{e}_z\cdot\boldsymbol{\sigma}] = -\beta(\mathbf{n}\times\mathbf{e}_z)\cdot\boldsymbol{\sigma} = \beta(n_x Y - n_y X). \tag{71}$$

Here, for any real vectors $\mathbf{u} = (u_x, u_y, u_z)^\top$ and $\mathbf{v} = (v_x, v_y, v_z)^\top$ in $\mathbb{R}^3$, we use the shorthand

$$\mathbf{u}\cdot\boldsymbol{\sigma} := u_x X + u_y Y + u_z Z, \qquad \mathbf{v}\cdot\boldsymbol{\sigma} := v_x X + v_y Y + v_z Z.$$

We also write the rotation axis as a unit vector $\mathbf{n} = (n_x, n_y, n_z)^\top \in \mathbb{R}^3$ with $\|\mathbf{n}\| = 1$, and define $\mathbf{e}_z := (0, 0, 1)^\top$ so that $Z = \mathbf{e}_z \cdot \boldsymbol{\sigma}$. The series sums to the closed-form Rodrigues rotation, yielding

$$(W_{L,1}^{\text{post}})^\dagger Z\, W_{L,1}^{\text{post}} = m_x X + m_y Y + m_z Z, \qquad m_x^2 + m_y^2 + m_z^2 = 1, \tag{72}$$

for real coefficients (m_x, m_y, m_z) determined by $(\beta, \mathbf{n})$. Combining equation 68 and equation 72 gives

$$M_1 = (m_x X + m_y Y + m_z Z) \otimes I = m_x \hat{X}_1 + m_y \hat{Y}_1 + m_z \hat{Z}_1. \tag{73}$$

Conjugating through $U_{\text{ent}} = \exp(-\frac{i}{2}\gamma \hat{Z}_1 \hat{Z}_2)$ gives

$$U_{\text{ent}}^\dagger \hat{Z}_1 U_{\text{ent}} = \hat{Z}_1, \tag{74}$$

$$U_{\text{ent}}^\dagger \hat{X}_1 U_{\text{ent}} = \hat{X}_1 \cos\gamma - \hat{Y}_1 \hat{Z}_2 \sin\gamma, \tag{75}$$

$$U_{\text{ent}}^\dagger \hat{Y}_1 U_{\text{ent}} = \hat{Y}_1 \cos\gamma + \hat{X}_1 \hat{Z}_2 \sin\gamma, \tag{76}$$

and therefore

$$M_2 = m_z \hat{Z}_1 + m_x(\hat{X}_1 \cos\gamma - \hat{Y}_1 \hat{Z}_2 \sin\gamma) + m_y(\hat{Y}_1 \cos\gamma + \hat{X}_1 \hat{Z}_2 \sin\gamma). \tag{77}$$

Finally,

$$O := (W_L^{\text{pre}})^\dagger M_2 W_L^{\text{pre}}, \qquad W_L^{\text{pre}} = W_{L,1}^{\text{pre}} \otimes W_{L,2}^{\text{pre}}. \tag{78}$$

For example, on the mixed term $\hat{Y}_1 \hat{Z}_2$,

$$(W_L^{\text{pre}})^\dagger (\hat{Y}_1 \hat{Z}_2) W_L^{\text{pre}} = \big((W_{L,1}^{\text{pre}})^\dagger Y\, W_{L,1}^{\text{pre}}\big) \otimes \big((W_{L,2}^{\text{pre}})^\dagger Z\, W_{L,2}^{\text{pre}}\big), \tag{79}$$

so W_L^{pre} preserves the support $\{1, 2\}$ while reweighting and mixing the local Pauli factors.

E.3 Generalization to N-Qubit Systems

We now carry the same three stages through for arbitrary N and arbitrary interaction graph $\mathcal{G} = (V, E)$, for a single encoding layer and a single entangling round, and show that the graph then determines a hard ceiling on the frequencies a given readout can reach.

Fix the measured qubit n, write $\mathcal{N}(n) = \{k : \{n, k\} \in E\}$ for its neighbour set and $d_n = |\mathcal{N}(n)|$ for its degree. The measurement is $M = \hat{Z}_n$ and the trainable block is $W_L = W_L^{\text{post}} U_{\text{ent}} W_L^{\text{pre}}$ with $W_L^{\text{post}} = \bigotimes_j W_{L,j}^{\text{post}}$ and $W_L^{\text{pre}} = \bigotimes_j W_{L,j}^{\text{pre}}$.

Stage 1 (Expose). Since W_L^{post} is a tensor product and $\hat{Z}_n$ acts only on qubit n, only $W_{L,n}^{\text{post}}$ affects the conjugation. The Rodrigues argument of Appendix E.2 applies verbatim on that factor, giving

$$M_1 = (W_L^{\text{post}})^\dagger \hat{Z}_n W_L^{\text{post}} = m_x \hat{X}_n + m_y \hat{Y}_n + m_z \hat{Z}_n, \qquad m_x^2 + m_y^2 + m_z^2 = 1, \tag{80}$$

independently of N and of the graph.

Stage 2 (Spread). The operators $\hat{Z}_k$, $k \in \mathcal{N}(n)$, commute, so $e^{-i\Phi_n} = \prod_{k\in\mathcal{N}(n)} (\cos\gamma_{nk}\, \hat{I} - i \sin\gamma_{nk} \hat{Z}_k)$. Expanding the product over subsets $T \subseteq \mathcal{N}(n)$ and writing $\hat{Z}_T \triangleq \bigotimes_{t\in T} \hat{Z}_t$,

$$\kappa_T \triangleq \prod_{k\in\mathcal{N}(n)\setminus T} \cos\gamma_{nk} \prod_{k\in T} \sin\gamma_{nk}, \tag{81}$$

$$\cos\Phi_n = \sum_{|T|\,\text{even}} (-1)^{|T|/2} \kappa_T \hat{Z}_T, \qquad \sin\Phi_n = \sum_{|T|\,\text{odd}} (-1)^{\frac{|T|-1}{2}} \kappa_T \hat{Z}_T. \tag{82}$$

Substituting into Lemma 2,

$$\begin{aligned} M_2 = U_{\text{ent}}^\dagger M_1 U_{\text{ent}} = m_z \hat{Z}_n &+ \sum_{|T|\,\text{even}} (-1)^{\frac{|T|}{2}} \kappa_T \big(m_x \hat{X}_n + m_y \hat{Y}_n\big) \hat{Z}_T \\ &+ \sum_{|T|\,\text{odd}} (-1)^{\frac{|T|-1}{2}} \kappa_T \big(m_y \hat{X}_n - m_x \hat{Y}_n\big) \hat{Z}_T. \end{aligned} \tag{83}$$

Every string in M_2 is supported on $\{n\} \cup T$ for some $T \subseteq \mathcal{N}(n)$; the graph fixes which qubits can appear, and the largest support is $\{n\} \cup \mathcal{N}(n)$.

Stage 3 (Align). W_L^{pre} is a tensor product, so conjugation acts qubit-wise and maps each single-qubit factor into a combination of $\hat{X}$, $\hat{Y}$, $\hat{Z}$ on the same qubit. It reweights and mixes the Pauli factors but cannot enlarge the support set. The effective observable $O = W_L^\dagger \hat{Z}_n W_L$ is therefore supported on $\{n\} \cup \mathcal{N}(n)$, of size $d_n + 1$.

Graph-limited reachability. This support constraint caps the reachable spectrum. By Proposition 2, a Pauli string belongs to a plane with configuration (S_+, S_-, T) only if it carries $\hat{Y}$ or $\hat{Z}$ on $S_+ \cup S_-$ and $\hat{X}$ on T, so its frequency obeys $|\omega| \le |S_+| + |S_-|$, which is at most the size of its support. Hence

$$\langle \mathcal{X}_{\omega,r}, O\rangle = \langle \mathcal{Y}_{\omega,r}, O\rangle = 0 \qquad \text{for all } \omega > d_n + 1. \tag{84}$$

The bound is attained. The string $\hat{Y}_n \hat{Z}_{\mathcal{N}(n)}$ appears in equation 83 with coefficient $\pm\mu\kappa_{\mathcal{N}(n)}$, where $\mu = m_y$ for d_n even and $\mu = m_x$ for d_n odd, and carries exactly one $\hat{Y}$ factor, so it lies in $\mathcal{Y}_{d_n+1,r}$ for the plane with $S_+ = \{n\} \cup \mathcal{N}(n)$ and $S_- = T = \emptyset$. Since all other strings of M_2 have smaller support and are orthogonal to it,

$$\big|\langle \mathcal{Y}_{d_n+1,r}, M_2\rangle\big| = 2\,|\mu| \prod_{k\in\mathcal{N}(n)} |\sin\gamma_{nk}|. \tag{85}$$

Two consequences follow. First, the reachable spectrum is set by the local degree rather than by N: measuring a qubit with d_n neighbours caps the accessible frequency at $d_n + 1$ no matter how many qubits the circuit has or how the parameters are trained. Activating frequency ω therefore

requires $\deg(n) \geq \omega - 1$ in the interaction graph, a condition a designer can check before training, and Corollary 1 shows that it also suffices. Second, by equation 85 the top-plane overlap scales as $\prod_k |\sin \gamma_{nk}|$, so the top frequency becomes unreachable if any coupling angle is 0 or π. Since $\sum_{T \subseteq \mathcal{N}(n)} \kappa_T^2 = \prod_k (\cos^2 \gamma_{nk} + \sin^2 \gamma_{nk}) = 1$, the coupling angles distribute a fixed budget of the effective observable across its Pauli supports $\{n\} \cup T$, so no single choice of γ maximizes its overlap with every invariant plane at once, which is why we treat the entangling strengths as trainable parameters.

Setting $d_n = 1$ recovers Appendix E.2: $\kappa_\emptyset = \cos\gamma$, $\kappa_{\{2\}} = \sin\gamma$, the reachable ceiling is $\omega = 2$, and the top-plane overlap is $2|m_x \sin\gamma|$, matching. The star graph of Fig. 2 has $d_1 = 2$ and ceiling $\omega = 3$, which is the frequency shown there. This analysis assumes a single entangling round; with L rounds the support spreads to the radius-L neighbourhood of n and the ceiling grows accordingly, which we do not analyze here.

E.4 Design Procedure

Algorithm 1 Placement rule for a target frequency

Require: interaction graph $\mathcal{G} = (V, E)$ with $|V| = N$, target frequency $\omega^\star \leq N$, with $\omega^\star = N$ if the spectrum is unknown

1: choose the readout qubit $n \in \arg\max_{v \in V} \deg(v)$
2: **if** $\deg(n) < \omega^\star - 1$ **then**
3: **return** add entangling rounds or data re-uploading ▷ one round reaches at most $\deg(n) + 1$
4: **end if**
5: $U_{\text{ent}} \leftarrow \prod_{k \in \mathcal{N}(n)} e^{-i\gamma_{nk} \hat{Z}_n \hat{Z}_k / 2}$ ▷ Spread, CZ at $|\gamma_{nk}| = \pi/2$
6: $W_L \leftarrow W_L^{\text{post}}\, U_{\text{ent}}\, W_L^{\text{pre}}$ ▷ Expose, Spread, Align
7: $W_R \leftarrow W_R^{\text{post}}\, U_{\text{ent}}\, W_R^{\text{pre}}$ ▷ state side, sets the amplitude
8: **return** $U(x) = W_L\, S(x)\, W_R$ with readout $\hat{Z}_n$

The procedure reads each vertex degree once and places $2\deg(n)$ two-qubit gates and four layers of single-qubit rotations, so it runs in $O(|V| + |E|)$ time. The invariant planes enter only its correctness proof, never its execution, so the 4^N-dimensional operator space is never formed. By Corollary 1 and Appendix D.6, every $\omega \leq \deg(n) + 1$ of the returned circuit has a nonzero coefficient at almost every parameter setting, and each can reach amplitude 1 on its own. For $\omega^\star = N$, a vertex of degree $N - 1$ needs $N - 1$ edges, so the star centered on the measured qubit is the smallest graph that reaches every frequency of one encoding layer in one round, and it is the circuit of our experiments. The star is native to platforms with all-to-all connectivity, such as trapped ions with Mølmer–Sørensen gates, while on bounded-degree devices, such as heavy-hex processors with degree at most three, one round reaches $\omega \leq 4$ and larger targets need further rounds or re-uploading.

E.5 Scope of the Results

Any encoding generator. Let $G = \sum_k \lambda_k \Pi_k$ be the spectral decomposition of a Hermitian generator. Every operator splits as $A = \sum_{j,k} \Pi_j A \Pi_k$ with $\mathrm{ad}_G(\Pi_j A \Pi_k) = (\lambda_j - \lambda_k)\Pi_j A \Pi_k$, and ad_G is self-adjoint for the Hilbert–Schmidt inner product, so $\mathcal{B}(\mathcal{H})$ is the orthogonal sum of the eigenspaces $\mathcal{B}_\omega = \{A : \mathrm{ad}_G(A) = -2\omega A\}$, on which $S(x)^\dagger A S(x) = e^{-i\omega x} A$. For an orthogonal basis $\{E_{\omega,r}\}_r$ of $\mathcal{B}_\omega$ with $\omega > 0$, the operators $\mathcal{X}_{\omega,r} = E_{\omega,r} + E_{\omega,r}^\dagger$ and $\mathcal{Y}_{\omega,r} = -i(E_{\omega,r} - E_{\omega,r}^\dagger)$ are closed pairs in the sense of Proposition 1, mutually orthogonal and of equal norm, and the proof of Proposition 3 applies verbatim with $2^{N - m_{\omega,r} - 1}$ replaced by $2^N / \|\mathcal{X}_{\omega,r}\|^2$. The frequencies are the half-differences of the eigenvalues of G (Schuld et al., 2021), and Proposition 2 is the explicit basis for $G = \sum_n \hat{X}_n$.

Pauli encodings and readouts. If qubit j is encoded along a Pauli axis σ_j and the readout is a Pauli operator on qubit n, single-qubit Clifford gates map $\sum_j \sigma_j$ to $\sum_j \hat{X}_j$ and the readout to $\hat{Z}_n$. These gates merge into the adjacent rotation layers, which are general, so every result holds with the Pauli labels exchanged.

Entanglers. Let P_n be a Pauli operator on the measured qubit and Q_k a Pauli operator on $k \in \mathcal{N}(n)$. Single-qubit Clifford gates that map P_n to $\hat{Z}_n$ and each Q_k to $\hat{Z}_k$ turn $\exp\big(-\frac{i}{2}\sum_k \gamma_{nk} P_n Q_k\big)$ into U_{ent} of Lemma 2, with signs absorbed into γ_{nk}, and they merge into the adjacent rotation layers. Lemma 2, Corollary 1, the amplitude results of Appendix D.6 and Algorithm 1 therefore hold unchanged for every entangler of commuting two-qubit couplings that share their factor P_n. This class contains CZ and controlled-phase gates, $CNOT$ gates with the measured qubit as common control or common target, whose single-qubit factors commute with the couplings, echoed cross-resonance gates with a common control, and Mølmer–Sørensen $\hat{X}\hat{X}$ gates. For any other layer of two-qubit gates on the edges at n, O remains supported on $\{n\} \cup \mathcal{N}(n)$, so the ceiling $\deg(n) + 1$ still holds, and part (i) of the corollary applies whenever O reaches the top plane at one parameter setting. This can fail. In our numerical checks, a star of iSWAP gates, whose $\hat{X}\hat{X}$ and $\hat{Y}\hat{Y}$ terms do not share a factor on n, leaves $c_{\deg(n)+1} \equiv 0$, whereas a star of $\sqrt{\mathrm{iSWAP}}$ gates activates it, so the class above is sufficient but not necessary.

F Data Reuploading

Data Reuploading. The previous analysis considered a single data-encoding block $S(x)$. A direct extension is to encode the input multiple times throughout the circuit:

$$U_{\mathrm{DR}}(x, \Theta) = W_L S(x) W_{L-1} S(x) \cdots W_1 S(x) W_0. \tag{86}$$

This is known as data reuploading (Pérez-Salinas et al., 2020). From the Fourier viewpoint, each encoding block introduces another opportunity to generate input-dependent rotations in the operator space. Therefore, repeating the encoding increases the range of accessible Fourier components. In particular, if one encoding layer with N qubits can generate frequencies up to $\omega \leq N$, then L reuploading layers can, in principle, generate higher frequencies up to $\omega \leq LN$. Thus, data reuploading provides a practical way to increase spectral expressivity without increasing the number of qubits. Access to higher frequencies does not mean that the model uses them. Our results are proved for a single encoding layer, and we do not extend Proposition 3, Corollary 1 or Algorithm 1 to $L > 1$. Our re-uploading experiments are therefore empirical and test whether the placement remains useful beyond the proved setting.

G Limitations and Potentials

Our placement analysis covers one encoding layer and one entangling round on each side of the encoder, with single-qubit Pauli encodings and readouts and with entanglers whose two-qubit couplings commute and share a factor on the measured qubit (Appendix E.5), and we evaluate data re-uploading only empirically. One round reaches frequencies up to $\deg(n) + 1$, so on devices of bounded connectivity, such as heavy-hex processors of degree three, larger targets need further entangling rounds, which our analysis does not cover, or re-uploading. High-frequency planes are spanned by Pauli strings that act on many qubits, and local noise such as depolarization damps a string exponentially in the number of qubits it acts on, so the frequencies that our rule activates are also the most exposed to noise. At 14 qubits, finite-shot readout needs more than 10^4 shots per point before any circuit we tested beats the constant predictor (Table 4). We evaluate all circuits in classical simulation and leave experiments on quantum hardware to future work. In broader impact, an interpretable design rule can help practitioners build PQCs that are both expressive and controllable, and it can reduce unnecessary circuit depth and two-qubit gate counts.