A Fourier Lens on Parameterised Quantum Circuits: A Review

Daeyeun Kim, Seungcheol Oh, and Joongheon Kim
Department of Electrical and Computer Engineering, Korea University, Seoul, Republic of Korea
E-mails: {dk725,seungoh,joongheon}@korea.ac.kr

***Abstract*—Variational quantum algorithms (VQAs) are among the most promising near-term approaches for quantum machine learning. For architectures that encode classical data and trainable angles via unitary evolutions generated by Hermitian operators, the expectation values of parameterised quantum circuits (PQCs) can be written as finite Fourier series in input variables. In this spectral view, the set of candidate frequencies is fixed by the eigenvalue structure of the data-encoding Hamiltonians, while the extent to which this spectrum is effectively usable depends on how flexibly the associated Fourier coefficients can be tuned through both the encoding and trainable circuit blocks; we review how these design choices determine spectral richness, degeneracy, coefficient correlations, and their trade-offs with trainability via a Fourier-based account of barren plateaus. We further discuss generalisation and function-approximation performance, including conditions under which overparameterised PQCs can interpolate noisy data yet maintain low test error and when angle-encoding with data re-uploading can realise universal trigonometric approximators for periodic functions. We conclude by highlighting open problems in architecture design, scalable coefficient control, and need for analysis on generalisation capabilities.**

***Index Terms*—variational quantum circuit, quantum machine learning, parameterised quantum circuit, barren plateau, Fourier-series representation.**

I. Introduction

Quantum devices are approaching regimes in which, for certain tasks, they may outperform classical supercomputers [1]. Nevertheless, near-term hardware is expected to remain in the noisy intermediate-scale quantum (NISQ) regime: qubit counts are modest, coherence times short, and gate errors restrict feasible circuit depth [2]. Within these constraints, variational quantum algorithms (VQAs) provide a pragmatic approach. In a VQA, a parameterised quantum circuit (PQC) is executed on quantum hardware multiple times to estimate expectation values, while a classical optimiser updates the circuit parameters to minimise a task-specific loss function. The literature refers to such parameterised models variously as *variational circuits* [3, 4], *quantum neural networks* [5, 6] or *quantum circuit learning* [7]. This hybrid division of labour deploys quantum resources where they are effective and keeps circuits shallow to mitigate limitations of NISQ devices. Recent works highlight that supervised learning with VQAs can exhibit resilience to noise [8] and has potential to achieve quantum advantages [9, 10]. Accordingly, assessing the potential for quantum advantage in machine-learning tasks within this paradigm has become an active research focus in recent years [11-15].

Despite rapid progress, a comprehensive understanding of the capabilities and limitations of VQAs is still being developed. A growing body of work analyses key properties of a PQC architecture and their interdependencies to characterise performance trade-offs and to guide principled, application-specific design. The most prominently discussed properties are expressivity, trainability, generalisation capability, and function-approximation performance. A unifying lens for this interplay, and the focus of this review, is the Fourier-series representation of PQC expectation values. For architectures in which classical data and trainable angles are encoded through gates of the form $\exp(-ixH)$ or $\exp(-i\theta H)$, these expectation values can be expressed as finite trigonometric polynomials in the input variables [16]. The set of frequencies characterises the expressive range of a PQC architecture. The degree of controllability over the Fourier coefficients, both individually and in their correlations, determines how effectively that range can be exploited during training. These connections make the Fourier framework a natural choice for systematic assessment of a PQC architecture.

The remainder of this review is structured as follows. Section II derives the Fourier representation of PQC expectation values and establishes notation. Sections III–V survey prior work that applies the Fourier perspective to expressivity, trainability, and generalisation. Section VI outlines avenues for future research.

II. Preliminaries: Expressing the Output of a Parameterised Quantum Circuit as a Fourier Series

Schuld et al. [16] demonstrated that when classical inputs are encoded into PQCs via time evolutions generated by one or more Hamiltonians, the circuit's output, an expectation value, admits a finite Fourier series representation in those input variables. In what follows, we adopt this framework to derive this Fourier form and define the notations used for the remainder of the review.

A. Basics of Parameterised Quantum Circuits

For a single real feature $x \in \mathbb{R}$ on n qubits, consider an L-layer PQC that alternates a data-encoding unitary $S_\ell(x)$ and

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-00561377). *(Corresponding author: Joongheon Kim)*

a trainable block $W_\ell(\boldsymbol{\theta}_\ell, \boldsymbol{\phi}_\ell)$. The PQC can be expressed as

$$U(x, \boldsymbol{\theta}, \boldsymbol{\phi}) = W_{L+1}(\boldsymbol{\theta}_{L+1}, \boldsymbol{\phi}_{L+1}) \prod_{\ell=1}^{\overleftarrow{L}} \big[S_\ell(x)\, W_\ell(\boldsymbol{\theta}_\ell, \boldsymbol{\phi}_\ell)\big], \tag{1}$$

where

$$\prod_{\ell=1}^{\overleftarrow{L}} A_\ell := A_L \cdots A_1, \tag{2}$$

and $\boldsymbol{\theta}_\ell$, $\boldsymbol{\phi}_\ell$ collect trainable parameters for layer ℓ.

Data encoding: Each encoding layer is characterised by a Hermitian operator $H_\ell = H_\ell^\dagger$:

$$S_\ell(x) = \exp\big(-i\, x\, H_\ell\big), \qquad \ell = 1, \ldots, L, \tag{3}$$

where H_ℓ may be identical or differ across layers.

Trainable block: The trainable block on n qubits at layer ℓ is

$$W_\ell(\boldsymbol{\theta}_\ell, \boldsymbol{\phi}_\ell) = E_\ell(\boldsymbol{\phi}_\ell) \bigotimes_{j=1}^{n} R^{(j)}(\boldsymbol{\theta}_\ell^{(j)}), \tag{4}$$

where each single-qubit unitary is a fixed-order Euler product

$$R^{(j)}(\boldsymbol{\theta}_\ell^{(j)}) = R_x\big(\theta_{\ell,x}^{(j)}\big)\, R_y\big(\theta_{\ell,y}^{(j)}\big)\, R_z\big(\theta_{\ell,z}^{(j)}\big). \tag{5}$$

Rotation gates about an axis a are expressed in terms of Pauli matrices σ_a as

$$R_a(\theta) = \exp\Big[-\, i\, \frac{\theta}{2}\, \sigma_a\Big], \qquad a \in \{x, y, z\}. \tag{6}$$

In Eq. (4), E_ℓ is a product of entanglers such as controlled-rotation gates about each of the three axes with trainable angular parameters $\boldsymbol{\phi}_\ell$ and controlled-NOT gates.

If the initial state is prepared as $|0\rangle$ and the circuit terminates with an arbitrary observable M, the expectation value of the PQC is

$$f(x, \boldsymbol{\theta}, \boldsymbol{\phi}) = \langle 0|\, U^\dagger(x, \boldsymbol{\theta}, \boldsymbol{\phi})\, M\, U(x, \boldsymbol{\theta}, \boldsymbol{\phi})\, |0\rangle\,. \tag{7}$$

B. Derivation of the Expectation Value as a Fourier Series

In this subsection, to improve readability we suppress the explicit dependence on the trainable parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$.

For clarity, we first treat the case in which all data-encoding layers use a common Hamiltonian H, i.e. $H_\ell = H$ and $S_\ell(x) = S(x) := \exp(-ixH)$ for all ℓ. Architectures with layer-dependent encoding Hamiltonians can be handled by diagonalising each H_ℓ separately; this modifies only the book-keeping of the eigenvalues and leads to the same Fourier form [16].

Write the spectral decomposition $H = V^\dagger \Sigma V$ with $\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, where $d = 2^n$ is the Hilbert-space dimension. Then $S(x) = V^\dagger e^{-ix\Sigma} V$. Insert this into Eq. (1) and absorb the basis change $V, V^\dagger$ into the trainable blocks:

$$\widetilde{W}^{(1)} := V\, W^{(1)}, \tag{8}$$

$$\widetilde{W}^{(\ell)} := V\, W^{(\ell)} V^\dagger, \quad 2 \le \ell \le L, \tag{9}$$

$$\widetilde{W}^{(L+1)} := W^{(L+1)} V^\dagger. \tag{10}$$

The PQC can be rewritten as

$$U(x) = \widetilde{W}^{(L+1)}\, e^{-ix\Sigma}\, \widetilde{W}^{(L)} \cdots e^{-ix\Sigma}\, \widetilde{W}^{(1)}. \tag{11}$$

Thus the x–dependence is fully carried by phases $e^{-ix\lambda_m}$. This change of basis does not restrict generality within the identical-encoder setting and the argument extends straightforwardly to layer-dependent encoders after diagonalising each H_ℓ.

Expanding the prepared state $U(x)\,|0\rangle$ componentwise and collecting the x-dependent phases gives for the ith component

$$\big[U(x)\,|0\rangle\big]_i = \sum_{j_1, \ldots, j_L = 1}^{d} \exp\Big(-ix \sum_{\ell=1}^{L} \lambda_{j_\ell}\Big)\, \widetilde{W}_{i\, j_L}^{(L+1)} \prod_{\ell=2}^{L} \widetilde{W}_{j_\ell\, j_{\ell-1}}^{(\ell)}\, \widetilde{W}_{j_1\, 1}^{(1)}. \tag{12}$$

Introduce the multi-index $\mathbf{j} = (j_1, \ldots, j_L) \in [d]^L$ where

$$[d]^L := \{(j_1, \ldots, j_L) \,:\, j_\ell \in \{1, \ldots, d\}\ \forall \ell\}.$$

Define the eigenvalue sum

$$\Lambda_{\mathbf{j}} := \sum_{\ell=1}^{L} \lambda_{j_\ell}.$$

Using (7) and expanding $f(x)$ in this basis yields

$$f(x) = \sum_{\mathbf{k}, \mathbf{j} \in [d]^L} a_{\mathbf{k},\mathbf{j}}\, e^{\,i(\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{j}})x}, \tag{13}$$

with pair-weights $a_{\mathbf{k},\mathbf{j}}$ depending only on the trainable unitaries and the measurement. More explicitly,

$$a_{\mathbf{k},\mathbf{j}} = \sum_{i,i'} \big(\alpha_{\mathbf{k},i}\big)^*\, M_{ii'}\, \beta_{\mathbf{j},i'}, \tag{14}$$

where

$$\alpha_{\mathbf{k},i} := \big(\widetilde{W}^{(L+1)}\big)_{i\, k_L} \prod_{\ell=2}^{L} \big(\widetilde{W}^{(\ell)}\big)_{k_\ell\, k_{\ell-1}}\, \big(\widetilde{W}^{(1)}\big)_{k_1\, 1}, \tag{15}$$

$$\beta_{\mathbf{j},i'} := \big(\widetilde{W}^{(L+1)}\big)_{i'\, j_L} \prod_{\ell=2}^{L} \big(\widetilde{W}^{(\ell)}\big)_{j_{\ell-1}\, j_\ell}\, \big(\widetilde{W}^{(1)}\big)_{j_1\, 1}. \tag{16}$$

Equations (13)–(16) show that the candidate spectrum of accessible frequencies is fixed entirely by the encoding scheme employed: it is generated by the eigenvalues of the data-encoding Hamiltonians H_ℓ across all layers ℓ. In contrast, the numerical values of the pair-weights, and hence of the Fourier coefficients c_ω, are determined by the trainable blocks and the measurement observable M. More specifically, by collecting like exponents in (13), the frequency set is

$$\Omega := \{\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{j}} :\ \mathbf{j}, \mathbf{k} \in [d]^L\} \subset \mathbb{R}. \tag{17}$$

The Fourier coefficient at $\omega \in \Omega$ is the sum of all pair-weights that induce that frequency:

$$c_\omega = \sum_{\substack{\mathbf{j}, \mathbf{k} \in [d]^L \\ \Lambda_{\mathbf{k}} - \Lambda_{\mathbf{j}} = \omega}} a_{\mathbf{k},\mathbf{j}}. \tag{18}$$

With these definitions, the expectation value of the PQC can be expressed as a finite Fourier series:

$$f(x) = \sum_{\omega \in \Omega} c_\omega \, e^{i\omega x}. \tag{19}$$

Since f is real, $c_{-\omega} = c_\omega^*$ and $0 \in \Omega$; the spectrum is symmetric.

This Fourier-series representation of the expectation value extends straightforwardly to multivariate functions in which several features are encoded via Hamiltonian evolutions; see Appendix A of the original paper by Schuld et al. [16] for details.

III. Expressivity

Expressivity of a PQC through the Fourier lens is governed by two factors: (i) the set of distinct accessible frequencies Ω and (ii) the ability to independently steer the corresponding Fourier coefficients $\{c_\omega\}_{\omega \in \Omega}$. In regimes where the spectrum is sufficiently rich and the coefficients can be controlled with enough flexibility, architectures of this type can act as universal approximators for continuous functions on compact domains under standard assumptions on periodicity and resource scaling. However, near-term hardware constraints make resources scarce, underscoring the importance of analysing the expressivity of a given PQC.

A. Set of Distinct Accessible Frequencies

The set of functions a PQC can represent as a function of the input variables and trainable parameters is set by Ω. As seen from Eq. (19), an expectation value can be expressed as a finite trigonometric polynomial; hence the number of unique frequencies $|\Omega|$ fixes the number of trigonometric basis functions accessible by the model and determines the *theoretical* limit of expressivity in this view. From Eq. (17), each $\omega \in \Omega$ is given by a difference between two eigenvalue sums across encoding layers; degeneracies in these sums reduce the number of distinct frequencies and therefore shrink $|\Omega|$.

Concrete constructions that minimise these degeneracies have been proposed in a variety of settings. Shin et al. [17] introduced exponential data-encoding schemes based on scaled non-entangling Pauli generators that realise nearly non-degenerate spectra whose size grows exponentially with the number of encoding gates. Similarly, Kordzanganeh et al. [18] analysed sequential and parallel architectures whose $|\Omega|$ grow exponentially. For a given value of N as a product between the number of qubits Q and data-encoding blocks L, the sequential architecture has N qubits and a single data-encoding block applied on each qubit, whereas the parallel architecture has N data-encoding blocks applied sequentially on a single qubit. Their experimental results showed that the parallel architecture approximates a top-hat function more accurately than the sequential one. Holzer and Turkalj [19] summarised existing data-encoding schemes that help reduce frequency degeneracy.

Beyond increasing the cardinality $|\Omega|$, a quantum model can learn which frequencies to emphasise within Ω. Jaderberg et al. [20] showed that introducing a trainable scalar for each data-encoding Hamiltonian enables rescaling of individual Hamiltonian's eigenvalues independently. By fine-tuning these scalers during training, the model can realise non-uniformly spaced frequency sets and adapt spectral richness to the task at hand. This overcomes a limitation of PQCs without trainable scalars in their data-encoding blocks, where Ω is constrained to integer multiples of a base frequency. They demonstrated the practical benefits of this frequency-pinching approach on a real-world task by solving the Navier–Stokes equations and reporting improved accuracy.

B. Ability to Control Fourier Coefficients

The second factor is the ability to control each Fourier coefficient. This determines the *effective* expressivity. To fully exploit the frequencies accessible to the quantum model, the model must be able to control the amplitudes of each Fourier basis function. The coefficients select a relevant subspace within the full space spanned by the Fourier basis accessible to the PQC.

Independent controllability over individual Fourier coefficients for a PQC whose Fourier space grows exponentially would require resources to grow exponentially. This is not feasible in practice. These resources include the classical memory required to store the parameters and the classical optimiser that must compute gradients of the loss with respect to each parameter. With exponential growth of these resources, the training process quickly becomes intractable because the per-iteration training cost as well as the required memory space increase linearly in $|\boldsymbol{\theta}|$. Moreover, NISQ hardware imposes strict resource constraints. Finite coherence times and the accumulation of gate errors with each operation cap the achievable circuit depth [21]. Consequently, only a limited number of qubits and gates, and thus trainable parameters, can be utilised on current devices.

In practice, if one wants to train variational models efficiently on classical hardware, it is natural to restrict the number of tunable parameters to grow at most polynomially with system size, i.e. $|\boldsymbol{\theta}| = \mathcal{O}(\mathrm{poly}(N))$. On the other hand, for exponential data-encoding schemes one typically has $|\Omega| = \mathcal{O}(\exp(N))$. The discrepancy between these two scalings implies that, in the asymptotic limit, there is no injective map from the set of Fourier coefficients $\{c_\omega\}_{\omega \in \Omega}$ to the parameter vector $\boldsymbol{\theta}$.

Hence the coefficients $c_\omega(\boldsymbol{\theta})$ must be functions of a shared pool of parameters in the trainable blocks, inducing correlations among the coefficients of distinct Fourier terms. Strobl et al. [22] quantified these correlations by sampling many random parameter settings, evaluating the model on a Nyquist grid, extracting the Fourier coefficients via a fast Fourier transform (FFT), and forming the Pearson correlation matrix across all coefficient pairs. The mean absolute correlation over the spectrum defines the Fourier coefficient correlation (FCC), a single-number summary of how strongly different Fourier modes are tied together for a given PQC architecture. Across several common PQC architectures, fingerprints differ markedly, demonstrating circuit-dependent correlation structure. When fitting random 1D and 2D Fourier series whose

frequencies match the model's spectrum, FCC correlates almost linearly with mean-squared error (MSE) between the target and approximated functions: lower FCC corresponds to better approximation of the function by the PQC. In contrast, no clear correlation was observed between theoretical limit of expressivity measured by $|\Omega|$ and FCC.

Another useful metric to analyse coefficient controllability is the variance of the coefficients. For a fixed spectrum Ω and observable O, the variance $\mathrm{Var}_\theta[c_\omega(\theta)]$ over random PQC parameters quantifies how much a single coefficient can be changed by tuning $\boldsymbol{\theta}$; small variance indicates poor coefficient-level controllability at that frequency. Schuld et al. [16] observed that with a fixed data-encoding Hamiltonian across the circuit, $\mathrm{Var}_\theta[c_\omega(\theta)]$ tends to decrease with increasing frequency. This is because the frequency of a mode is produced by differences of eigenvalue sums $\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{j}}$, where distinct paths are characterised by $\mathbf{j}$ and $\mathbf{k}$. The number of paths that realise a given ω typically shrinks toward the spectral edges. As a result, the number of pair-weights $a_{\mathbf{k},\mathbf{j}}$ that contribute to a single Fourier coefficient via Eq. (18) decreases as the frequency increases.

This logic leads to an important observation: controllability over Fourier coefficients depends on the data-encoding blocks as well as on the trainable blocks. The trainable parameters are shared across modes, but a smaller number of distinct paths to generate a given ω means fewer distinct parameter pathways feed into the coefficient of that particular ω.

Mhiri et al. [23] formalised this both mathematically and empirically. They defined the frequency generator $R(\omega)$ as the set of all possible paths generating the frequency ω and the cardinality $|R(\omega)|$ as the *frequency redundancy*.

Under the common 2-design assumption for the trainable blocks, they show that

$$\mathrm{Var}_\theta[c_\omega(\theta)] \in \mathcal{O}\Big(\alpha \, \tfrac{|R_e(\omega)|}{d}\Big), \tag{20}$$

where $\alpha := \big(d\|O\|_2^2 - \mathrm{Tr}(O)^2\big)/d^2$, $d = 2^n$, and $|R_e(\omega)| := |R(\omega)|/d^2$ is the normalised redundancy. Thus, on average over random parameters, modes with larger redundancy have larger accessible variance, whereas low-redundancy modes have strongly constrained controllability over their coefficients; moreover, because $d = 2^n$, all variances decay exponentially in n.

Relaxing to ε-approximate 2-designs, Mhiri et al. also proved upper bounds of the form

$$\mathrm{Var}_\theta[c_\omega(\theta)] \in \mathcal{O}\big(Q_\varepsilon(|R_e(\omega)|)\big), \tag{21}$$

where Q_ε is at most quadratic in $|R_e(\omega)|$. Hence, for any fixed ε, the same inductive bias remains: lower $|R_e(\omega)|$ implies tighter variance constraints, and for many encodings a broad subset of modes is likely to exhibit exponentially small variance, which they termed *vanishing coefficients*.

They further distinguished coefficient-level concentration from model-level concentration, giving regimes where $\mathrm{Var}_\theta[f(x,\theta)]$ need not vanish exponentially even though many $c_\omega(\theta)$ do and vice versa. This clarifies how redundancy-driven constraints shape effective expressivity.

Lastly, the authors introduced the idea of a global ℓ_2 budget on coefficients. Independently of parameterisation and encoding schemes employed, if the data-encoding Hamiltonians are fixed, one has the Fourier norm bounds

$$|f(x,\theta)|^2 \le \|O\|_\infty^2, \tag{22}$$

$$\sum_{\omega\in\Omega} |c_\omega(\theta)|^2 \;\le\; \|O\|_\infty^2, \tag{23}$$

where $\|O\|_\infty$ is the largest singular value (or absolute eigenvalue) of the observable O. The square of this quantity acts as a global amplitude budget shared across all Fourier modes. Combined with the redundancy-controlled variances, this implies competition between modes: allocating significant weight to some coefficients necessarily limits what remains available to others. This budget constraint is orthogonal to spectrum size and persists for any L and any choice of trainable unitaries.

In summary, there are structural trade-offs between the two factors that affect the expressivity of a PQC: the spectrum of accessible frequencies and the controllability of coefficients. Encoding schemes that maximise $|\Omega|$ by removing degeneracies improve basis richness but simultaneously push most $|R(\omega)|$ to $\mathcal{O}(1)$, amplifying vanishing-coefficient effects unless the trainable layers deviate substantially from 2-designs. Encoding schemes with more degeneracy have smaller $|\Omega|$ but can yield higher redundancy, which helps maintain variance in some coefficients.

IV. Trainability

Trainability refers to how the statistics of the loss gradient, typically $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_{\theta_j} C]$, scale with the number of qubits n and with circuit depth. McClean et al. [6] showed that, for a broad class of randomly initialised parameterised circuits whose parameter blocks approximate a unitary 2-design, the variance of gradients of global cost functions decays exponentially in n, i.e. $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_{\theta_j} C] = \mathcal{O}(2^{-n})$. This phenomenon is known as a barren plateau (BP). Subsequent works identified several mechanisms that can cause or aggravate BPs, including high expressivity [24], extensive entanglement [25], properties of the learning task [26], the choice of cost function [27], and noise [28].

A. A Spectral Picture of Barren Plateaus

In Sec. II we expressed the PQC output as a Fourier series in the input variables, $f(x,\boldsymbol{\theta},\boldsymbol{\phi}) = \sum_{\omega\in\Omega} c_\omega(\boldsymbol{\theta},\boldsymbol{\phi})\, e^{i\omega x}$. To understand trainability, it is convenient to examine its dependence on a *single* parameter while holding all other quantities fixed. Define

$$g_i(\vartheta) \;:=\; f\big(x,\, \theta_1, \ldots, \theta_{i-1}, \vartheta, \theta_{i+1}, \ldots, \theta_L\big), \tag{24}$$

and expand g_i into a Fourier series in ϑ,

$$g_i(\vartheta) \;=\; \sum_{k\in\Omega} c_k^{(i)}(\vartheta)\, e^{ikx}, \qquad P_i \;:=\; \sum_{k\in\Omega} |c_k^{(i)}(\vartheta)|^2, \tag{25}$$

where P_i is the total spectral power associated with the ith parameter.

Consider the ith trainable block $U_i(\theta_i)$ and the state $\rho_{i-1}(\boldsymbol{\theta}_{<i})$ entering that block. Following an analysis based on second moments of the parameter ensemble, Okumura and Ohzeki [29] bounded the deviation of P_i from its Haar-typical value in terms of the block's closeness to a unitary 2-design. Specifically,

$$\left| P_i - \frac{1}{2^n+1} \right| \;\le\; \epsilon^{(2)}_{U_i}\big(\rho_{i-1}\big), \tag{26}$$

where the 2-design deficit is

$$\epsilon^{(2)}_{U_i}(\rho) \;:=\; \left\| \Phi^{(2)}_{\mathcal{U}_i}\big(\rho^{\otimes 2}\big) - \Phi^{(2)}_{\mathrm{Haar}}\big(\rho^{\otimes 2}\big) \right\|_1, \tag{27}$$

with $\Phi^{(2)}_{\mathcal{U}_i}(X)$ denoting the ensemble-averaged channel

$$\Phi^{(2)}_{\mathcal{U}_i}(X) := \mathbb{E}_{\theta_i}\left[U_i^{\otimes 2} X (U_i^\dagger)^{\otimes 2} \right], \tag{28}$$

and $\Phi^{(2)}_{\mathrm{Haar}}$ the corresponding Haar-twirl on two copies. Thus, $\epsilon^{(2)}_{U_i} \ge 0$ and equals 0 iff the ensemble $\{U_i(\theta_i)\}$ forms an exact unitary 2-design on the support of ρ. In the ideal 2-design limit $\epsilon^{(2)}_{U_i} \to 0$, Eq. (26) yields

$$\sum_{k\in\mathbb{Z}} |c_k^{(i)}|^2 \;=\; \frac{1}{2^n+1}. \tag{29}$$

Equation (29) shows that the total spectral power of g_i is exponentially suppressed in the number of qubits. Since the set of accessible parameter frequencies typically grows with depth, the average magnitude of nonconstant coefficients $|c_k^{(i)}|$ becomes exponentially small. As gradients are derivatives of the Fourier series,

$$\partial_{\theta_i} f \;=\; \sum_{k\in\mathbb{Z}} ik\, c_k^{(i)}\, e^{ik\theta_i},$$

this collapse of spectral power translates into exponentially vanishing gradient statistics, giving a spectral view of barren plateaus. In practice, this implies that resolving gradients to a fixed relative precision requires a number of measurements that grows exponentially in n when the trainable blocks approach a 2-design.

B. Problem-Specific Ansätze and Empirical Correlations

The spectral viewpoint unifies the trade-offs between expressivity and trainability: as $U_i(\theta_i)$ approaches a 2-design, the model explores larger swathes of state space but concentrates spectral power near constants, making gradients small. Because BPs depend on the joint choice of architecture, measurement, cost function, and noise, there is no circuit family that is uniformly optimal across all aspects of a PQC.

Empirically, studies comparing circuit templates report that pushing expressivity too far can hurt downstream performance: classification accuracy often saturates or even degrades at high expressivity [30-32]. From the Fourier perspective, one interpretation is that many learning targets occupy only a modest spectral subset; architectures that drive spectra towards Haar-like coverage simultaneously shrink $|c_\omega|$ for $\omega \neq 0$, hindering optimisation. This motivates problem-specific architecture design that tailors the accessible spectra through both data encoders and trainable layers to the task at hand rather than maximising expressivity indiscriminately. Such tailored design is regarded as one of the leading strategies against BPs [24, 33].

V. Generalisation Capability

From the Fourier viewpoint, models generalise best when most of their spectral weight lies on task-relevant, typically low frequencies that capture the coarse structure of the target function, while avoiding over-emphasis on high frequencies that encode small-scale fluctuations and noise. At the same time, a controlled tail of higher frequencies can absorb label noise without harming test error. This mirrors observations in modern deep neural networks: highly overparameterised models can still generalise, a phenomenon known as benign overfitting (BO) [34].

Peters and Schuld [35] provided a characterisation of BO for quantum models using the Fourier-series representation of PQC expectation values. In their framework, the test error of an interpolating model splits into (i) a variance term, where high-frequency modes spike near training points to soak up noise, and (ii) a bias term, governed by how much true low-frequency signal bleeds into alias modes. Two practical messages follow: (i) allowing many high-frequency alias modes can help by concentrating noise locally while remaining smooth elsewhere; (ii) if most spectral weight sits on the task's low-frequency band with a thin tail across aliases, and the circuit dimension grows faster than the sample size, an interpolating PQC can still achieve vanishing generalisation error.

The original Fourier-series view of PQCs already argued that expected generalisation should guide PQC design along with other circuit characteristics such as expressibility and trainability [16]. Schuld et al. highlighted as a promising research direction the development of modern generalisation measures for PQCs that exploit their Fourier structure. To the best of our knowledge, this task remains largely open and is a promising direction for future work.

VI. Conclusion and Future Directions

The Fourier–series viewpoint offers a compact account of how PQC architectures behave across expressivity, trainability, and generalisation. Spectral spread determines which functions are representable, and coefficient controllability determines whether that expressive range is usable. Spectral concentration, whether induced by approximate 2-design behaviour or by degeneracy, helps explain the emergence of barren plateaus. Recent generalisation analyses likewise reduce to how a PQC allocates effective spectral power and budget across frequencies, with benign overfitting arising when low-frequency concentration is paired with a sufficiently diffuse high-frequency tail. Certain PQC architectures can act as universal function approximators, subject to resource constraints inherited from hardware limitations and architecture-specific scaling laws.

Interesting areas of research for future work include:

1) **Spectrum-aware encoder design:** Methods for engineering Ω, e.g. tunable spectral scaling, controlled degeneracy breaking, and systematic constructions that balance the richness of the frequency spectrum with coefficient accessibility.
2) **Scalable coefficient control:** A deeper understanding of how parameter sharing, circuit depth, and entangling structure influence variance and covariance of $c_\omega(\boldsymbol{\theta})$, and how to design architectures with favourable controllability under NISQ constraints.
3) **Generalisation beyond simple models:** Fourier-based analyses for deeper, multi-feature, and re-uploading architectures, including PQC-specific generalisation metrics that incorporate aliasing structure, degeneracy, and coefficient constraints.

Overall, the Fourier framework condenses several seemingly disparate behaviours of PQCs into a single analytic language. As methods for spectral engineering and spectral diagnostics mature, this perspective is well positioned to guide principled PQC architecture selection and to clarify the practical capabilities of near-term quantum machine learning.

VII. Acknowledgement

This work was supported by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT) (IITP-2026-RS-2024-00436887). (Corresponding author: Joongheon Kim).

References

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[2] J. Preskill, "Quantum computing in the nisq era and beyond," *Quantum*, vol. 2, p. 79, 2018.

[3] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New Journal of Physics*, vol. 18, no. 2, p. 023023, 2016.

[4] J. Romero and A. Aspuru-Guzik, "Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions," *Advanced Quantum Technologies*, vol. 4, no. 1, p. 2000003, 2021.

[5] E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," *arXiv preprint arXiv:1802.06002*, 2018.

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

[7] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Physical Review A*, vol. 98, no. 3, p. 032309, 2018.

[8] R. LaRose and B. Coyle, "Robust data encodings for quantum classifiers," *Physical Review A*, vol. 102, no. 3, p. 032420, 2020.

[9] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.

[10] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill *et al.*, "Quantum advantage in learning from experiments," *Science*, vol. 376, no. 6598, pp. 1182–1186, 2022.

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

[12] M. T. West, S.-L. Tsang, J. S. Low, C. D. Hill, C. Leckie, L. C. Hollenberg, S. M. Erfani, and M. Usman, "Towards quantum enhanced adversarial robustness in machine learning," *Nature Machine Intelligence*, vol. 5, no. 6, pp. 581–589, 2023.

[13] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, "Training deep quantum neural networks," *Nature communications*, vol. 11, no. 1, p. 808, 2020.

[14] Y. Liu, S. Arunachalam, and K. Temme, "A rigorous and robust quantum speed-up in supervised machine learning," *Nature Physics*, vol. 17, no. 9, pp. 1013–1017, 2021.

[15] M. Schuld and N. Killoran, "Quantum machine learning in feature hilbert spaces," *Physical review letters*, vol. 122, no. 4, p. 040504, 2019.

[16] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Physical Review A*, vol. 103, no. 3, p. 032430, 2021.

[17] S. Shin, Y. S. Teo, and H. Jeong, "Exponential data encoding for quantum supervised learning," *Physical Review A*, vol. 107, no. 1, p. 012422, 2023.

[18] M. Kordzanganeh, P. Sekatski, L. Fedichkin, and A. Melnikov, "An exponentially-growing family of universal quantum circuits," *Machine Learning: Science and Technology*, vol. 4, no. 3, 2023.

[19] P. Holzer and I. Turkalj, "Spectral invariance and maximality properties of the frequency spectrum of quantum neural networks," *arXiv preprint arXiv:2402.14515*, 2024.

[20] B. Jaderberg, A. A. Gentile, Y. A. Berrada, E. Shishenina, and V. E. Elfving, "Let quantum neural networks choose their own frequencies," *Physical Review A*, vol. 109, no. 4, p. 042421, 2024.

[21] J. W. Z. Lau, K. H. Lim, H. Shrotriya, and L. C. Kwek, "Nisq computing: where are we and where do we go?" *AAPPS bulletin*, vol. 32, no. 1, p. 27, 2022.

[22] M. Strobl, M. E. Sahin, L. van der Horst, E. Kuehn, A. Streit, and B. Jaderberg, "Fourier fingerprints of ansatzes in quantum machine learning," *arXiv preprint arXiv:2508.20868*, 2025.

[23] H. Mhiri, L. Monbroussou, M. Herrero-Gonzalez, S. Thabet, E. Kashefi, and J. Landman, "Constrained and vanishing expressivity of quantum fourier models," *Quantum*, vol. 9, p. 1847, 2025.

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

[25] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, "Entanglement-induced barren plateaus," *PRX quantum*, vol. 2, no. 4, p. 040316, 2021.

[26] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, "Barren plateaus preclude learning scramblers," *Physical Review Letters*, vol. 126, no. 19, p. 190501, 2021.

[27] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nature communications*, vol. 12, no. 1, p. 1791, 2021.

[28] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," *Nature communications*, vol. 12, no. 1, p. 6961, 2021.

[29] S. Okumura and M. Ohzeki, "Fourier coefficient of parameterized quantum circuits and barren plateau problem," *arXiv preprint arXiv:2309.06740*, 2023.

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

[31] T. Hubregtsen, J. Pichlmeier, P. Stecher, and K. Bertels, "Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility and entangling capability," *Quantum Machine Intelligence*, vol. 3, no. 1, p. 1, 2021.

[32] J. Tangpanitanon, S. Thanasilp, N. Dangniam, M.-A. Lemonde, and D. G. Angelakis, "Expressibility and trainability of parametrized analog quantum systems for machine learning applications," *Physical Review Research*, vol. 2, p. 043364, 2020.

[33] M. T. West, J. Heredge, M. Sevior, and M. Usman, "Provably trainable rotationally equivariant quantum machine learning," *PRX Quantum*, vol. 5, no. 3, p. 030320, 2024.

[34] P. L. Bartlett, P. M. Long, G. Lugosi, and A. Tsigler, "Benign overfitting in linear regression," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 063–30 070, 2020.

[35] E. Peters and M. Schuld, "Generalization despite overfitting in quantum machine learning models," *Quantum*, 2023, accepted 2023-12-11.