

PAPER • OPEN ACCESS

Fourier fingerprints of ansatzes in quantum machine learning

To cite this article: Melvin Strobl *et al* 2026 *Quantum Sci. Technol.* **11** 045008

View the [article online](#) for updates and enhancements.

You may also like

- [Interpreting variational quantum models with active paths in parameterized quantum circuits](#)
Kyungmin Lee, Hyungjun Jeon, Dongkyu Lee et al.
- [Fourier series guided design of quantum convolutional neural networks for enhanced time series forecasting](#)
Sandra Leticia Juárez-Osorio, Mayra Alejandra Rivera-Ruiz, Andres Mendez-Vazquez et al.
- [Learning to maximize quantum neural network expressivity via effective rank](#)
Juan Yao

Unlock Quantum Computing Potential with Multiphysics Simulation

Design the infrastructure that enables quantum performance.

The future of computing depends on stable, scalable quantum technologies capable of solving problems beyond the reach of classical supercomputers.

COMSOL Multiphysics® empowers engineers to simulate and optimise the supporting systems behind quantum platforms — including RF and microwave components, photonics and optics, cryogenic cooling and thermal management. By modelling the coupled physics that shape the qubit environment, you can reduce noise, improve stability and accelerate quantum hardware development.

» comsol.com/industry/electronics/quantum-computing

Quantum Science and Technology



PAPER

OPEN ACCESS

RECEIVED

25 September 2025

REVISED

29 June 2026

ACCEPTED FOR PUBLICATION

21 July 2026

PUBLISHED

18 August 2026

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Fourier fingerprints of ansatzes in quantum machine learning

Melvin Strobl^{1,*} , M Emre Sahin² , Lucas van der Horst¹ , Eileen Kühn¹ , Achim Streit¹ and Ben Jaderberg³

¹ Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

² The Hartree Centre, STFC, Sci-Tech Daresbury, Warrington WA4 4AD, United Kingdom

³ IBM Quantum, IBM Research Europe, Hursley, Winchester SO21 2JN, United Kingdom

* Author to whom any correspondence should be addressed.

E-mail: melvin.strobl@kit.edu

Keywords: quantum computing, quantum machine learning, quantum Fourier models, ansatz selection, high-energy physics

Abstract

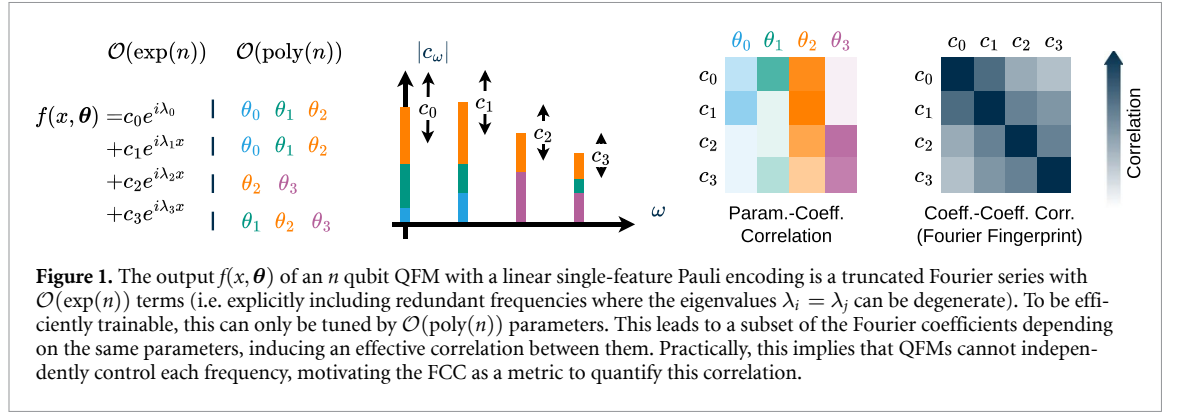
Typical schemes to encode classical data in variational quantum machine learning (QML) lead to quantum Fourier models with $\mathcal{O}(\exp(n))$ Fourier basis functions in the number of qubits. Despite this, in order for the model to be efficiently trainable, the number of parameters must scale as $\mathcal{O}(\text{poly}(n))$. This imbalance implies the existence of correlations between the Fourier modes, which depend on the structure of the circuit. In this work, we demonstrate that this phenomenon exists and show cases where these correlations can be used to predict ansatz performance. For several popular ansatzes, we compute the Fourier coefficient correlations (FCCs) using a linear feature map (FM) and construct the *Fourier fingerprint*, a visual representation of the correlation structure. We subsequently numerically show how, for the problem of learning random Fourier series, the FCC correctly predicts relative performance of ansatzes whilst the widely-used expressibility metric does not. Finally, we demonstrate how our framework applies to the more challenging problem of jet reconstruction in high-energy physics. Overall, our results demonstrate how the Fourier fingerprint is a powerful new tool in the problem of optimal ansatz choice for QML.

1. Introduction

The field of quantum machine learning (QML) has emerged as one of the most rapidly advancing areas of quantum computing in recent years [1]. In particular, QML has seen a significant focus on the use of parameterized quantum circuits (PQCs) as machine learning (ML) models [2, 3]. Here, in a way that is conceptually similar to deep neural networks, the parameters of a quantum circuit are trained to learn the function $f(\mathbf{x}, \boldsymbol{\theta})$ that best approximates a given dataset of input–output pairs $(\mathbf{x}, f(\mathbf{x}, \boldsymbol{\theta}))$. Due to its applicability to a wide range of domains [4–7] and amenability to running on pre-fault-tolerant quantum computers [8, 9], PQCs remain one of the most widely-used paradigms in QML.

In this framework, a typical QML model consists of quantum feature maps (FMs) $\hat{U}_F(\mathbf{x})$, which encode a classical input $\mathbf{x}$ into the Hilbert space, and variational ansatzes $\hat{U}_A(\boldsymbol{\theta})$ which manipulate the input via trainable parameterized gates $\boldsymbol{\theta}$. Importantly, when the FM has the general structure $\hat{U}_F(\mathbf{x}) = \prod_i e^{ig_i(\mathbf{x})H_i}$, the output of these models can be represented as a finite expansion in $g_i(\mathbf{x})$. In the special case of a data-independent Hamiltonian H_i and a linear preprocessing g_i , this expansion reduces to a truncated Fourier series in the input [10], giving rise to the name quantum Fourier models (QFMs). Fully understanding how the choice of FM and ansatz affects the Fourier spectrum of QFMs, both analytically [11, 12] and numerically [13, 14], remains crucial for separating QFMs from classical surrogates [15, 16] and probing dequantization [17, 18].

In this work, we study the pair-wise correlations between Fourier coefficients that arise in efficiently-trainable QFMs, a phenomenon recently highlighted from a redundancy perspective in [12]. Our contribution is complementary: rather than bounding the variance of individual coefficients, we explicitly characterize their pair-wise correlation structure across multiple practical ansatzes. Subsequently, we condense it into a single scalar metric, the Fourier coefficient correlation (FCC), together with the associated



Fourier fingerprint, which reveals how this structure depends on the choice of ansatz. The existence of these correlations means that QFMs cannot in general tune all terms in their Fourier series independently, thereby restricting the class of functions they can practically learn. We theoretically recap this in section 2.1 and emphasize from the outset that this conclusion is asymptotically rigorous only for FMs that produce exponentially many *unique* frequencies. Nevertheless, in section 2.2 we numerically validate the existence of correlations even for widely-used Pauli encodings which have polynomially many unique frequencies, and across several popular ansatzes. An overview of this concept is pictured in figure 1.

Subsequently, we demonstrate in practice how the *FCC*, the average of the fingerprint, can be used as an indicator of ansatzes performance in section 3.1. Here a Pauli encoding with polynomially many unique frequencies is considered again, so the models can be classically simulated using Fourier surrogates [15–18]. This positions our work alongside, but distinct from, the line of research on classical surrogates and dequantization of QFMs: those works study when a classical Fourier model can match or replace a QFM, whereas in this regime we view the FCC not as a separation indicator but as a diagnostic of how effectively a trainable QFM explores independent Fourier modes. We find that, for learning random 1D Fourier series, models with lower FCC perform better, even in cases where expressibility would predict the ansatz to be worse, and we extend our experiments to learning 2D Fourier series, where we observe the same result.

Finally, in section 3.2, we show how the FCC can be used to pick well-performing ansatzes in the more challenging problem of 2D jet reconstruction in high-energy physics (HEP). Given the energy and momentum of a two-particle collision, we train our model to predict the transverse momentum of the largest jet.

Comparing the performance of different ansatzes, we find a similar trend as observed in the experiments concerning random Fourier series, i.e. that models with lower average FCC achieve lower mean squared error (MSE). Overall, our results highlight that the Fourier fingerprint and the resultant FCC metric are practically useful and should be integrated into the process of ansatz choice in QML.

2. Method

2.1. Correlation between coefficients in quantum Fourier models

Let us define an L -layer QFM by $U(\mathbf{x}, \theta) = W^{(L+1)}(\theta)S(\mathbf{x})W^{(L)} \dots W^{(2)}S(\mathbf{x})W^{(1)}$, with trainable unitaries W and n -qubit encoding gates S taking a D dimensional input $\mathbf{x}$. For this class of models, which includes quantum neural networks (QNNs), the output $f(\mathbf{x}, \theta) = \langle 0|^{\otimes n} U^\dagger(\mathbf{x}, \theta) \mathcal{M} U(\mathbf{x}, \theta) |0\rangle^{\otimes n}$ is obtained by measuring an expectation value of an observable $\mathcal{M}$. The seminal work of [10] showed the equivalence of this expression to a truncated Fourier series

$$f(\mathbf{x}, \theta) = \sum_{\omega \in \Omega} c_\omega(\theta) \exp(i\omega \mathbf{x}) \quad (1)$$

where Ω defines the set of frequencies and c_ω the complex valued coefficient belonging to the frequency vector ω .

From equation (1), we first note that the Fourier coefficients of the model are explicitly dependent on the trainable parameters $\theta \in \Theta$ of the ansatz.⁴ The frequency spectrum of this QFM, and specifically models where the output is an expectation value, is defined by the gaps in the eigenspectrum of the FM generator Hamiltonian [10, 12]

$$f(\mathbf{x}, \theta) = \sum_{i=1}^D \sum_{\mathbf{j}, \mathbf{k} \in \prod_{l=1}^L [1, d_l]} c_{\mathbf{j}, \mathbf{k}}(\theta) \exp(i(\Lambda_{i, \mathbf{j}} - \Lambda_{i, \mathbf{k}}) \mathbf{e}_i^T \mathbf{x}_i) \quad | \quad \mathbf{j}, \mathbf{k} \in \prod_{l=1}^L [1, d_l], \quad (2)$$

where $\prod_{l=1}^L$ defines the product of the intervals $j_l, k_l \in [1, d_l]$ in the multi-index $\mathbf{j}, \mathbf{k}$ and $\Lambda_{i, \mathbf{j}} = \lambda_{j_1}^i + \lambda_{j_2}^i + \dots + \lambda_{j_L}^i$ is the sum of all d_l eigenvalues of the l th generator in L layers for input index i mapped to the corresponding scalar feature $x_i \in \mathbf{x}$ via the standard-basis selection vector $\mathbf{e}_i \in \{0, 1\}^D$ (i.e. $\mathbf{e}_i^T \mathbf{x} = x_i$).

Whilst typically Ω is defined as the *unique* gaps in the spectrum, here it is useful to define a larger set of frequencies that includes degenerate values

$$\tilde{\Omega} = \biguplus_{i=1}^D \{ \Lambda_{i, \mathbf{j}} - \Lambda_{i, \mathbf{k}} \} \quad | \quad \mathbf{j}, \mathbf{k} \in \prod_{l=1}^L [1, d_l], \quad (3)$$

where $\biguplus$ denotes a multiset union, so that $\tilde{\Omega}$ retains every redundant copy of a frequency rather than collapsing them as Ω does. Importantly, whilst the scaling of unique frequencies is determined by the degeneracy of the FM eigenspectrum [19], the absolute number of terms in the Fourier series always grows as $|\tilde{\Omega}| \sim \mathcal{O}(2^{nD})$.

Finally, we note that for QFMs to be efficiently trainable, they must contain a sub-exponential $|\theta| \sim \mathcal{O}(\text{poly}(n))$ number of parameters, since the time complexity of training is linear in $|\theta|$ [2, 20, 21]. Combining these observations about the size of $|\tilde{\Omega}|$ and $|\theta|$, it is immediately apparent that whenever $|\theta| < |\tilde{\Omega}|$, which is the asymptotic regime of interest here, there can be no injective map between the multiset of Fourier *terms* indexed by $(\mathbf{j}, \mathbf{k})$ and θ . Therefore, the coefficients $c_{\omega}(\theta)$ must depend on a *shared* set of ansatz parameters, which implies the existence of correlations between the coefficients of different terms in the Fourier series. Note that the argument is over the redundant set $\tilde{\Omega}$: if instead $|\theta| \geq |\Omega|$ and the ansatz is constructed so that each parameter only appears in coefficients sharing the same unique frequency, the existence of a non-trivial correlation pattern is no longer enforced *a priori* by this counting argument; whether such an ansatz can be realized in practice is an open question discussed in section 4. A complementary, redundancy-explicit derivation following Mhiri *et al* [12] is presented below: writing the trainable unitaries as $W^{(l)}$ and the redundancy set of a unique frequency ω as $R(\omega)$, the Fourier coefficient evaluates to

$$c_{\omega}(\theta) = \sum_{(\mathbf{J}, \mathbf{K}) \in R(\omega)} W_{j_1}^{(1)*} W_{j_2 j_1}^{(2)*} \dots W_{k_2 k_1}^{(2)} W_{k_1}^{(1)} O_{j_L k_L}, \quad (4)$$

which makes manifest that any W -entry appearing in two redundancy sets $R(\omega) \cap R(\omega') \neq \emptyset$ couples the two coefficients with an extended derivation given in appendix E. We provide another perspective on this statement in appendix E.

The practical implication of this result is that, whenever $|\theta| < |\tilde{\Omega}|$, the training of QFMs is in fact an optimization of Fourier series where each term (element of $\tilde{\Omega}$) is not independent; in addition, global constraints such as the Fourier-norm bound of [12] apply independently of this discussion. This is an additional constraint, which is not apparent from the form of equation (1), that reduces the possible solution space. Furthermore, this is in contrast to classical Fourier regression and random-Fourier-feature surrogates [15–18], where the Fourier modes can be tuned independently. We note, following [15, 16], that vanishing FCC corresponds in this sense to behavior close to that of a linear classical regressor; from the inductive-bias perspective the FCC should therefore be interpreted as a structural property of QFMs rather than as a generic figure of merit to be maximized or minimized in absolute terms. In section 3 we numerically show how this constraint for QFMs affects performance, whilst in appendix G we analytically consider how this constraint appears in the (MSE) loss.

⁴ The coefficients also implicitly depend on the FM as the degeneracy of the encoding spectrum upper-bounds the variance of the coefficients [12].

2.2. Computing the Fourier fingerprint

Given a QFM, we numerically compute the correlations between Fourier coefficients as follows. For a randomly initialized set of ansatz parameters θ , we evaluate $f(\mathbf{x}, \theta)$ for many different $\mathbf{x} \in \mathcal{X}$ such that the number of samples satisfies the Shannon–Nyquist [22] criterion, i.e. $|\mathcal{X}| \geq 2\Omega_{\max}$ along each input dimension where $\Omega_{\max}$ is the largest unique frequency in Ω (sufficient for an alias-free recovery of the $|\Omega|$ Fourier coefficients of equation (1)). We then obtain the Fourier coefficients $\mathbf{c}$ using the fast Fourier transform (FFT), specifically as implemented in the QML-Essentials framework [23].

As the FFT requires knowledge of the number of frequencies in advance, in the remainder of this work we consider a single-qubit Pauli encoding, such that $\Omega = \llbracket -nL, nL \rrbracket^D$, based on equation (3) under the assumption $d_l = 2 \forall l \in [1 \dots L]$. We then repeat this procedure M times for randomly sampled ansatz parameters θ from the space of possible ansatz parameters Θ , obtaining a set of M Fourier coefficient vectors. Specifically, each $\theta \in \Theta$ is i.i.d. and drawn from a uniform distribution with the interval $[0, 2\pi]$. Treating each calculated instance of $\mathbf{c}$ as independent samples, we then compute the Pearson correlation coefficient [24]

$$r_{\Theta}(\omega, \omega') = \frac{\sum_{\theta \in \Theta} (c_{\omega}(\theta) - \bar{c}_{\omega})(c_{\omega'}(\theta) - \bar{c}_{\omega'})}{\sqrt{\sum_{\theta \in \Theta} (c_{\omega}(\theta) - \bar{c}_{\omega})^2 \sum_{\theta \in \Theta} (c_{\omega'}(\theta) - \bar{c}_{\omega'})^2}} \quad (5)$$

between each possible pair in the spectrum. Here, $\bar{c}_{\omega}$ and $\bar{c}_{\omega'}$ represent the average value of the coefficient at frequency ω and ω' respectively. Note that the Pearson correlation is normalized by the standard deviation of the coefficients at each frequency. We provide a sanity check of this in figure 8 where we randomly sample coefficients with an increasing number of redundancies towards higher frequencies. Computing the correlation between them results in zeros for all frequency pairs, which confirms that the variance is mitigated, meaning that the decaying variance towards higher frequencies as observed in [12] is taken into account with the FCC metric.

We compute this correlation specifically for an $n = 6$, $L = 1$ QFM with the 1D FM $S = \bigotimes^n R_Y^{(n)}(x)$ and observable $\mathcal{M} = \frac{1}{n} \sum_i \sigma_i^z$, i.e. a model with a single scalar input $x \in \mathbf{x}$ that is encoded redundantly by applying the same rotation angle across all n qubits, which is the standard parallel-encoding setting used throughout the QFM literature for single-feature datasets [10, 12, 13]. In figure 2(a) a visualization of the computed correlations for a range of ansatzes is displayed. It can be observed, that e.g. the hardware-efficient ansatz (HEA) results in a QFM with significantly larger correlations between high-frequency indices whereas *Circuit 15* has a minimal correlation across all coefficients. Generally, the distribution of correlations differs significantly between the ansatzes, prompting us to coin the term *Fourier fingerprint* to describe this unique distribution for each ansatz. Note that the axes are shifted by one coefficient index as we only show the lower triangle of the correlation matrix, i.e. the unique correlation without the trivial diagonal. In appendix B, specifically figure 7(a), we provide similar results for a $S = \bigotimes^n R_X^{(n)}(x)$ FM.

Furthermore, the concept of correlations between Fourier coefficients can be extended to higher input dimensions D as stated in equation (3). Here, ω becomes a vector of frequencies and each coefficient is indexed by $c_{(\omega_0, \omega_1)}$ for the case of $d = 2$. The correlation is then computed, similar to the 1D case, but taking into account each possible combination of frequencies, resulting in figure 2(b) where we visualize the fingerprint of a single ansatz for a 2D model, with FM $S = \bigotimes^n R_X^{(n)}(x_1) \bigotimes^n R_Y^{(n)}(x_2)$. In this work, we use $500 \cdot |\theta| \cdot 2^n \cdot D$ samples to compute the correlations to high accuracy. Notably, the correlation to a fixed error, the number of samples does not need to grow exponentially with n as shown in appendix J.

Based on these fingerprints, we introduce the FCC in equation (6) as a metric that quantifies the average correlation between the coefficients by calculating the mean of the fingerprint,

$$\text{FCC}_{\Theta} := \frac{1}{|\tilde{\Omega}|} \sum_{\omega, \omega' \in \tilde{\Omega}} |r_{\Theta}(\omega, \omega')|. \quad (6)$$

Here, $\tilde{\Omega}$ is the set of all frequency pairs arising from the frequencies $\omega \in \Omega$ of the QFM.

3. Results

3.1. Learning random Fourier series

3.1 To investigate the predictive power of the FCC metric for ansatz performance, we train a model on a truncated Fourier series $\hat{f}$ such that it minimizes the (MSE) loss

$$\mathcal{L}_{\text{MSE}} = \frac{1}{|\mathcal{X}|} \int_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x}, \boldsymbol{\theta}) - \hat{f}(\mathbf{x})|^2 d\mathbf{x}, \quad (7)$$

where the data comes from the family of random Fourier series

$$\hat{f}(\mathbf{x}) = \sum_{\boldsymbol{\omega} \in \Omega} \hat{c}_{\boldsymbol{\omega}} \exp(i\boldsymbol{\omega}\mathbf{x}). \quad (8)$$

Here, $\hat{c}_{\boldsymbol{\omega}}$ are uniformly drawn complex numbers within the unit circle, i.e. $\sqrt{r}e^{-i2\pi p}$ with $r, p \in \mathcal{U}(0, 1]$, satisfying $|\hat{c}_{\boldsymbol{\omega}}| \leq 1$. The frequencies of the dataset Ω are explicitly set to the same frequencies as those generated by the FM of the QFM models Ω . This ensures that the target function is within the learnable domain of the QFMs assuming perfectly independent coefficients. Furthermore, the inputs $\mathbf{x}$ are sampled according to the Nyquist–Shannon sampling theorem [22] ($|\mathbf{x}| \geq 2\Omega_{\text{max}}$), i.e. $|\mathcal{X}| = 2\max(\Omega)$ in the interval $[0, 2\pi)$. Training is then carried out using the Adam optimizer [26] with a learning rate of 0.01 over 2000 and 4000 steps for the 1D and 2D Fourier series respectively; full hyperparameters are summarized in appendix G. We evaluate the (MSE) across multiple randomly seeded models and Fourier coefficients and average over the results to derive a performance index. Specifically, we use 10 different model initialization and 10 different dataset seeds, yielding a total of 100 runs per ansatz.

We compute the expressibility using the Kullback–Leibler (KL)-divergence [27] between the model $\int_{\boldsymbol{\theta}} (|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|)^{\otimes t} d\boldsymbol{\theta}$ and the distribution of the Haar-integral $\int_{\text{Haar}} (|\psi\rangle\langle\psi|)^{\otimes t} d\psi$ of a state t -design with $D_{\text{KL}}(\hat{P}_{\text{Model}}(F; \boldsymbol{\theta}) | P_{\text{Haar}}(F))$ [25]. Here $F = |\langle\psi_{\boldsymbol{\varphi}} | \psi_{\boldsymbol{\phi}}\rangle|^2$ describes the fidelity between two states, and a smaller KL-divergence corresponds to a higher expressibility. We adopt this Haar-distance definition because it is the most widely used in the QML literature [25] and is directly comparable across the ansatzes considered here; alternative formulations (e.g. those based on the dynamical Lie algebra [28]) are complementary but not directly comparable for our small-circuit numerics.

In appendix H we provide details on the comparison between the expressibility and the FCC in terms of complexity. For the following experiments, we utilize the implementation from the QML-Essentials package [23] where a lower metric value also corresponds to a higher expressibility, with $500 \cdot |\boldsymbol{\theta}| \cdot 2^n$ parameter samples and 75 histogram bins. Here we find that computing the FCC incurs a lower overhead both on classical and quantum hardware, making it a more scalable metric for comparing ansatzes.

The results for the 1D case are visualized in figure 3(a). Here, the FCC and expressibility are represented by different colors, whereas each symbol indicates a different QML ansatz. Note, that for visualization purposes, we display the $1 - \text{Expressibility}$ such that a low value in the figure corresponds with a high value for the expressibility metric. In line with the observations from figure 2(a), *Circuit 19* shows a similar FCC value to *Circuit 15*, but the resulting (MSE) is slightly higher. For *Circuit YZY* and *Circuit 18*, the MSEs are almost identical, while the FCCs deviate slightly. Following this, we observe that, overall, the FCC correlates almost linearly with the (MSE), with the lowest and highest values of the FCC being 4.57×10^{-3} (*Circuit 15*) and 0.323 (*Circuit 18*), respectively. The expressibility, in contrast, exhibits a noticeably weaker linear correlation with the final (MSE) on these benchmarks, and does not consistently order the ansatzes. Note that this observation is restricted to the chosen Haar-distance estimator and the Fourier-regression task; expressibility remains a meaningful and physically motivated quantity in its own right, capturing controllability of the variational unitary, and is in general expected to track the controllability of the trainable layers [28]. The takeaway here is therefore not that expressibility ‘fails’, but that for this concrete predictive question FCC is the more informative summary statistic.

The polynomial unique-spectrum size of the Pauli-style QFMs considered here means that the present learning task can be solved efficiently classically. Either exactly via least-squares Fourier regression on Ω , or approximately via random-Fourier-feature surrogates [17, 18]. We do not claim a quantum advantage in this regime; rather, the controlled benchmark serves to compare *ansatz-selection metrics* on QFMs that are widely used in practice. The relevance of the FCC to dequantization (i.e. its potential as a probe for when an RFF-style surrogate suffices) is discussed in section 4.

Employing a second Pauli encoding gate, orthogonal to the first, we replicate this training on a two-dimensional Fourier series, subsequently calculating the FCC. The Fourier series dataset is then naturally extended to a two-dimensional dataset, yielding the results shown in figure 3(b).

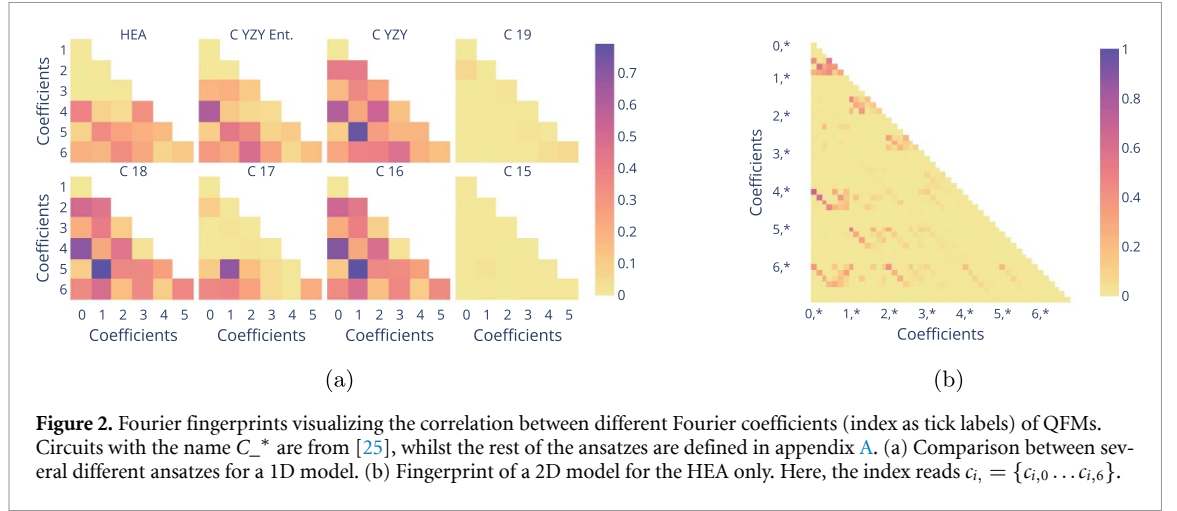


Figure 2. Fourier fingerprints visualizing the correlation between different Fourier coefficients (index as tick labels) of QFMs. Circuits with the name C_* are from [25], whilst the rest of the ansatzes are defined in appendix A. (a) Comparison between several different ansatzes for a 1D model. (b) Fingerprint of a 2D model for the HEA only. Here, the index reads $c_i, = \{c_{i,0} \dots c_{i,6}\}$.

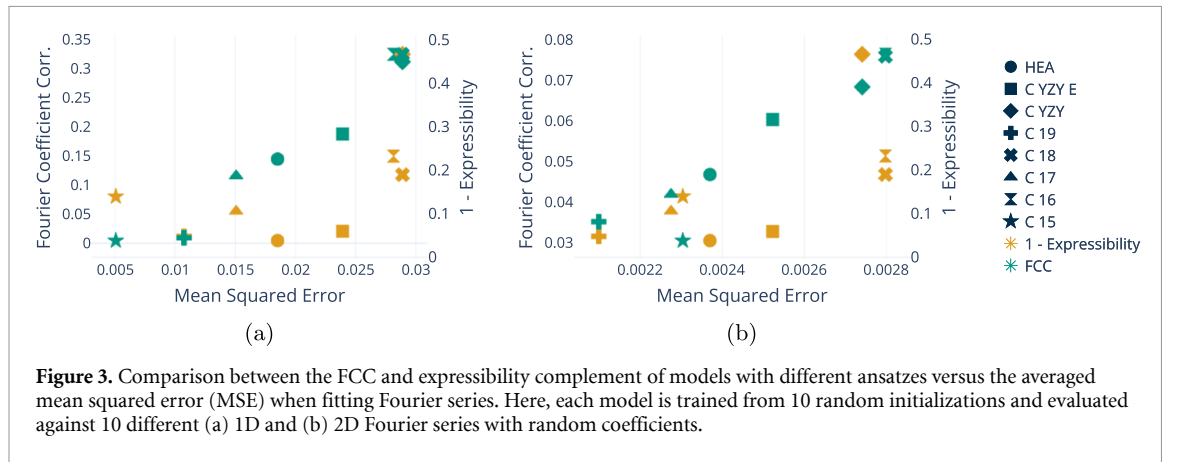


Figure 3. Comparison between the FCC and expressibility complement of models with different ansatzes versus the averaged mean squared error (MSE) when fitting Fourier series. Here, each model is trained from 10 random initializations and evaluated against 10 different (a) 1D and (b) 2D Fourier series with random coefficients.

Here, one can see that *Circuit 15* is now an outlier in the otherwise almost linear relationship between (MSE) and FCC. This outlier can be explained by the fact that *Circuit 15* lacks the imaginary part of the coefficients when being joint by a Pauli-RX FM, which we used to encode the second input dimension (see appendix B for results specifically concerning this FM). The absence of an imaginary part depending on the rotational axis was also shown in [29] for this specific circuit and is included in appendix C.

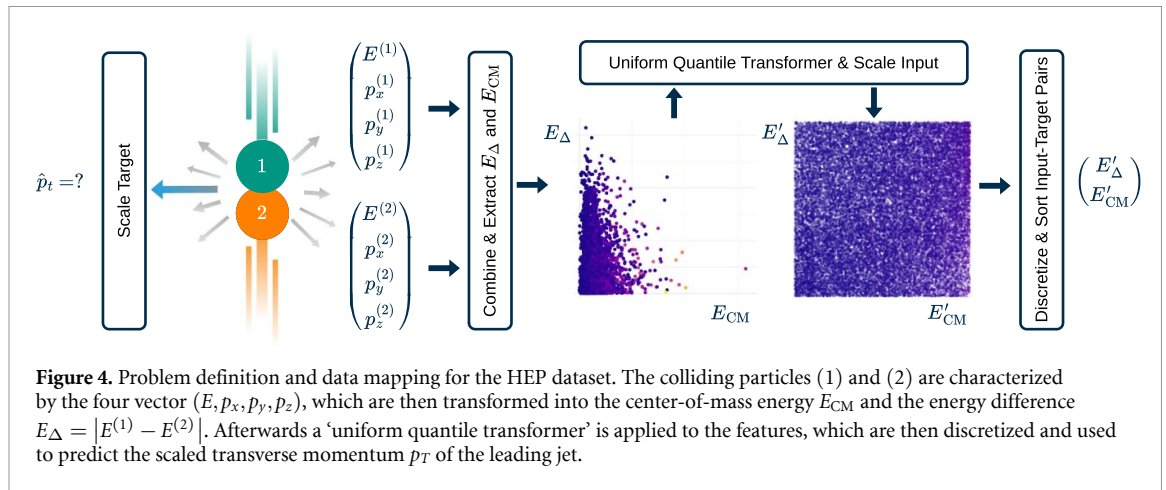
Overall, the results on the 1D and 2D Fourier series reveal a clear relationship between the (MSE) and the FCC, indicating that this metric provides a more powerful tool for predicting model performance in these cases. Importantly, in both sets of experiments, the widely-used expressibility metric incorrectly identifies which ansatzes would perform the best. We note that the standard deviations are significantly smaller than the plotted values for all measures and are mostly consistent across all ansatzes (see appendix F).

3.2. Application to HEP

HEP is a widely studied application and benchmark for QML [6, 30–36]. In this section, we apply the same methodology from section 3.1 to the more challenging HEP problem of 2D jet reconstruction.

Specifically, we train a QFM to reconstruct the highest transverse momentum p_T (i.e. the transverse momentum of the leading jet) using the four-vector (E, p_x, p_y, p_z) of two colliding particles. In each of the synthetically generated collision events, two protons collide along the z -axis, where their constituent partons interact and produce, through a series of further interactions, decays and final hadronization, sprays of particles observed by the detector as a cluster of hadron jets. Since the partons travel exclusively along z and the large hadron collider detector is symmetric about this axis, only the transverse momenta components in the x and y directions are relevant. In the following numerical experiments, we utilize the dataset published in [37].

In contrast to section 3.1, our assumption that the frequencies of the dataset match those of the QFM cannot be guaranteed in this case, making this problem potentially harder for the model to learn. We therefore expect, on theoretical grounds, that the FCC–(MSE) correlation is weaker on real-world



data than on the controlled Fourier-series benchmark of section 3.1: the FCC is constructed under the assumption that the target spectrum aligns with the model spectrum, which holds by construction in figure 3 but not for arbitrary physical data. Analyzing the dataset to identify a subset of the most relevant input features for prediction using simple feature importance analysis, we find that the type of particle and charge have negligible influence on the leading p_T , which aligns with physical expectations. Meanwhile, we identify two input features that can be computed from the data that have a large effect on the leading p_T , the center-of-mass energy $E_{CM} = \sqrt{(E^{(1)} + E^{(2)})^2 - (p_z^{(1)} + p_z^{(2)})^2}$ and the absolute energy difference $E_{\Delta} = |E^{(1)} - E^{(2)}|$ where $(\cdot)^{(1)}$ and $(\cdot)^{(2)}$ depicts the property of the first and second particle respectively. Note that due to the aforementioned symmetry, $p_x = p_y = 0$. Considering the distribution for the resulting features, we found that they are centered around low values (see mid of figure 4). Therefore, we apply a uniform quantile transformer to E_{CM} and E_{Δ} resulting in a uniform distribution of the input features. To make them fit the natural periodicity of the input encoding gates, we subsequently apply a ‘MinMax’ scaling to fit them in the interval $[0, 2\pi)$. We then discretize this distribution to ensure evenly spaced input features as conceptually visualized in figure 4. Similarly, the p_t target is scaled to match the range of the measured expectation value within $[-1, 1]$.

We first benchmark the complexity of this problem by applying a classical ML model. We train a classical multi-layer perceptron (MLP) consisting of 2 hidden layers and 8 neurons per layer, resulting in 105 trainable parameters, on the supervised learning task of predicting p_T from the input features. While being small in size, the number of parameters of this classical model still exceeds those of the QFM tested in this work.

We train the classical model with a batch size of 256 on a combination of the KL divergence (KLDiv) and the Huber loss as $\mathcal{L} = \text{KLDiv} + 0.001 \cdot \text{HuberLoss}$ using the Adam optimizer [26] with a learning rate of 0.05, adjusted dynamically via a learning rate scheduler. We justify the choice of this loss function as we aim to learn the momenta distribution of the dataset, due to the probabilistic nature of the task and no one-to-one input to output mapping, which KLDiv naturally models. KL-divergence is used exclusively for validation and testing, while we add a small Huber loss during training, as it empirically improved convergence stability.

Training is performed on 200 000 events with a ratio of 80% for training, 10% for validation and 10% for testing. On average, across 10 random models, the classical MLP obtained a KLDiv score of 0.163 ± 0.049 . However, selecting the 10 best-performing models from a pool of 30 improved the result significantly, yielding a KLDiv of 0.035 ± 0.005 . Despite trying multiple engineering tactics—including varying model architecture, model size, and learning strategies—performance did not improve. Consequently, we selected the simplest solution that still yielded competitive results.

Following an evaluation of the classical model, we shift our attention to the QFM. As in previous experiments described in section 3.1, we use 10 different model initializations and 10 different dataset initializations. In contrast to the preliminary experiments on MLP, we reduce the dataset size to 40 000 with a ratio of 80% for training, 10% for validation and 10% for testing, both for any subsequent experiments on the MLP and QFM. We switch the previous ‘MinMax’ scaler of the classical MLP to a ‘uniform quantile transformer’ as in the QFM; the resulting two-stage pipeline (uniform-quantile $\rightarrow$ ‘MinMax’ to $[0, 2\pi)$, then discretization) is identical to the one used by the QFM so that both models see the same effective inputs. Similarly, the target values using ‘MinMax’ scaling and compared with the

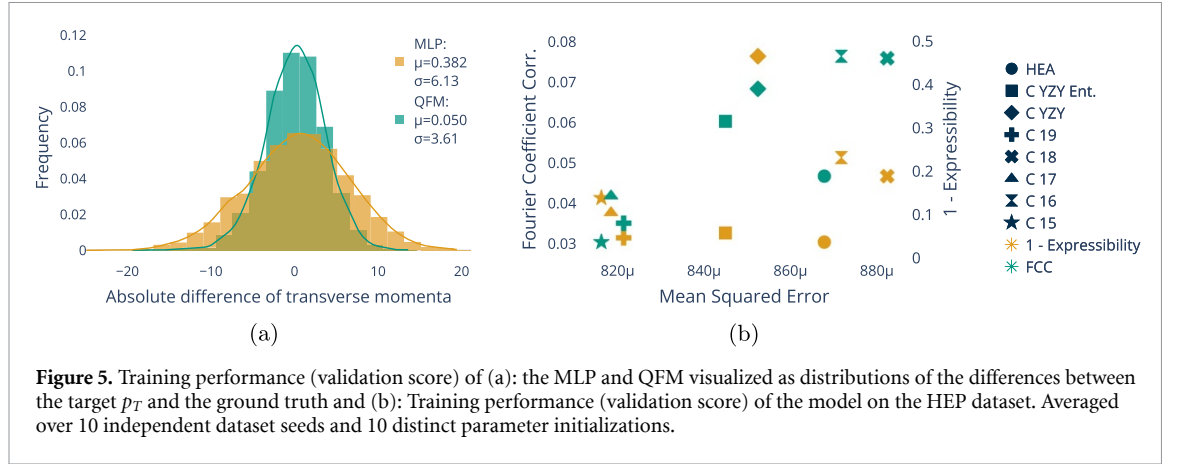


Figure 5. Training performance (validation score) of (a): the MLP and QFM visualized as distributions of the differences between the target p_T and the ground truth and (b): Training performance (validation score) of the model on the HEP dataset. Averaged over 10 independent dataset seeds and 10 distinct parameter initializations.

predicted values combining the (MSE) and KL-divergence loss [27] $\mathcal{L} = \text{MSE} + 0.001 \cdot \text{KLDiv}$ yields the best results in conjunction with an Adam optimizer [26] with a learning rate of 0.005.

A comparison between the results obtained by the MLP and QFM is shown in figure 5(a). Specifically, we plot the distribution of absolute differences between the predicted and true p_T values in each case. Here, we observe that both models achieve a low mean deviation of 0.382 and 0.050 for the MLP and QFM, respectively. However, the standard deviation of the absolute differences is lower for the QFM with a standard deviation of 3.61 compared to 6.13 for the MLP. Overall, our results validate that the problem is non-trivial to solve with a standard MLP, and that our trained QFM produces a competitive benchmark.

In figure 5(b) we show the comparison of the predicted performance of the FCC and expressibility for the HEP dataset. When observing the FCC, *Circuit 17* and the ‘Hardware Efficient Ansatz’ can be identified as outliers. The ansatz with the lowest and highest FCC, *Circuit 15* and *Circuit 18*, correlate to the lowest and highest (MSE), respectively. Although the order of MSEs is not perfectly predicted by the FCCs here, which is consistent with the expectation, set out earlier in this section, that real-world data need not satisfy the spectrum-alignment assumption underlying the FCC as cleanly as the controlled task of figure 3, we observe a clear overall trend linking correlation and performance. Compared to expressibility, both metrics successfully predict the same clusters of ansatzes while the FCC is computationally less expensive as discussed in appendix H. In addition to the results considering the (MSE), we provide the results for the KLDiv and Huber loss in appendix I.

4. Conclusion and discussion

In this work, we have analyzed the existence of correlations between the Fourier coefficients of QFMs and its impact on performance for several popular ansatzes. In section 2, we introduced the theoretical motivation as to why these correlations exist and how they can be numerically calculated by repeatedly sampling random instances of the model. In section 3.1, we showed how, for the task of learning random Fourier series, models with lower average FCC consistently perform better for both 1D and 2D inputs. Finally, in section 3.2 we numerically evidenced how this relationship extends to problems of practical interest such as the more challenging setting of 2D jet reconstruction in HEP. Compared to the analytical redundancy treatment of Mhiri *et al* [12], our contribution is twofold: first, we make the parameter-coefficient correlation structure explicit and visualizable through the Fourier fingerprint; second, we condense it into a single scalar metric, the FCC, which we numerically demonstrate to be predictive of QFM performance across multiple ansatzes, input dimensionalities, and a real-world HEP task.

Despite our numerical results, we note that correlations between the Fourier coefficients of efficiently-trainable QFMs are only asymptotically guaranteed for FMs that produce exponentially many *unique* frequencies. This includes the Golomb encoding [38] as well as those in [39, 40]. Whilst the FMs used in this work produce QFMs with exponentially many basis functions, in contrast, there are only a linear number of unique modes due to degeneracy in the eigenspectrum. It therefore remains an interesting open question whether one can design an ansatz that, in combination with such a degenerate FM, would produce Fourier coefficients that are independent in a stronger, non-asymptotic sense, for example by ensuring that each parameter only appears in the coefficients of terms with the same nonzero frequency.

In the case of QNNs this would be restricted to nonzero-frequency modes, since c_0 generally depends on the spectral gap between every generator and itself. This restriction does not, however, apply to QFMs built from quantum kernels, whose frequencies are indexed by generator eigenvalues rather than gaps. Nevertheless, our results highlight that this phenomenon does not occur in popular ansatzes. Understanding what such an ansatz would look like, and if it is efficiently implementable, remains an interesting open question. Notably, classical preprocessing would turn cause the Fourier series representation of the QFM to become truncated. The effect of preprocessing as in [13] on the FCC is unknown and subject for further research.

For QFMs whose unique-frequency set Ω grows only polynomially in n , which includes all Pauli-style encodings considered in this work, the function class is in principle dequantizable: a least-squares fit on the basis $\{e^{i\omega^\top x}\}_{\omega \in \Omega}$, or a random-Fourier-feature approximation thereof [17, 18], recovers the same expressivity classically. Recent work has further sharpened the conditions under which classical and quantum models meaningfully disagree [15, 16]. Within this picture, the FCC acquires an additional interpretation: a model whose Fourier coefficients are strongly correlated effectively occupies a much smaller subspace of the spanning Fourier basis, suggesting that an even smaller classical surrogate could reproduce its outputs. Conversely, exponential-frequency encodings (e.g. [38–40]) are precisely the regime in which classical RFF surrogates lose their efficiency guarantees, and where the FCC is most stringent. Quantitatively connecting the FCC to RFF sample complexity is a promising direction for future work.

The aspect of trainability should also be considered an important area of future research. Ragone *et al* [28] showed that ansatzes with high expressibility are very unlikely to be trainable, raising the interesting question of whether a similar pattern translates to the FCC. The numerical results presented in this paper do not strictly imply that a high expressibility correlates with a low FCC (see figure 3), suggesting a non-trivial relationship between these metrics. Therefore, it cannot be ruled out that a lower FCC, i.e. ansatzes with highly-independent coefficients, are generally not trainable. A thorough numerical study of trainability and Barren-plateau behaviour as a function of the FCC, including its dependence on n , L , and feature-map type, is beyond the scope of this manuscript and is left as an explicit avenue for follow-up work.

Finally, in this work, we studied only one instance of a real-world dataset. As a future direction of research, it remains to be seen whether the same trends hold for other problems, whereby ansatzes with the lowest average FCC perform the best. In particular, this heuristic has the underlying assumption that the frequencies of the dataset are themselves independent. If this is not true, one may find that the best QML model is one which purposefully produces Fourier coefficients with large correlations specifically tied to the data. In this way, the FCC may be better thought of as an inductive bias to encode into QFMs, instead of as a metric to always minimize.

Acknowledgments

We thank Gabriel Mejía Ruiz for helpful discussions on mathematical formulations and Maja Franz for valuable discussions concerning QFMs in a broader view.

We want to express our gratitude to Ryan Sweke and Christa Zoufal for their useful discussions and insights.

MS, EK and AS acknowledge support by the state of Baden-Württemberg through bwHPC.

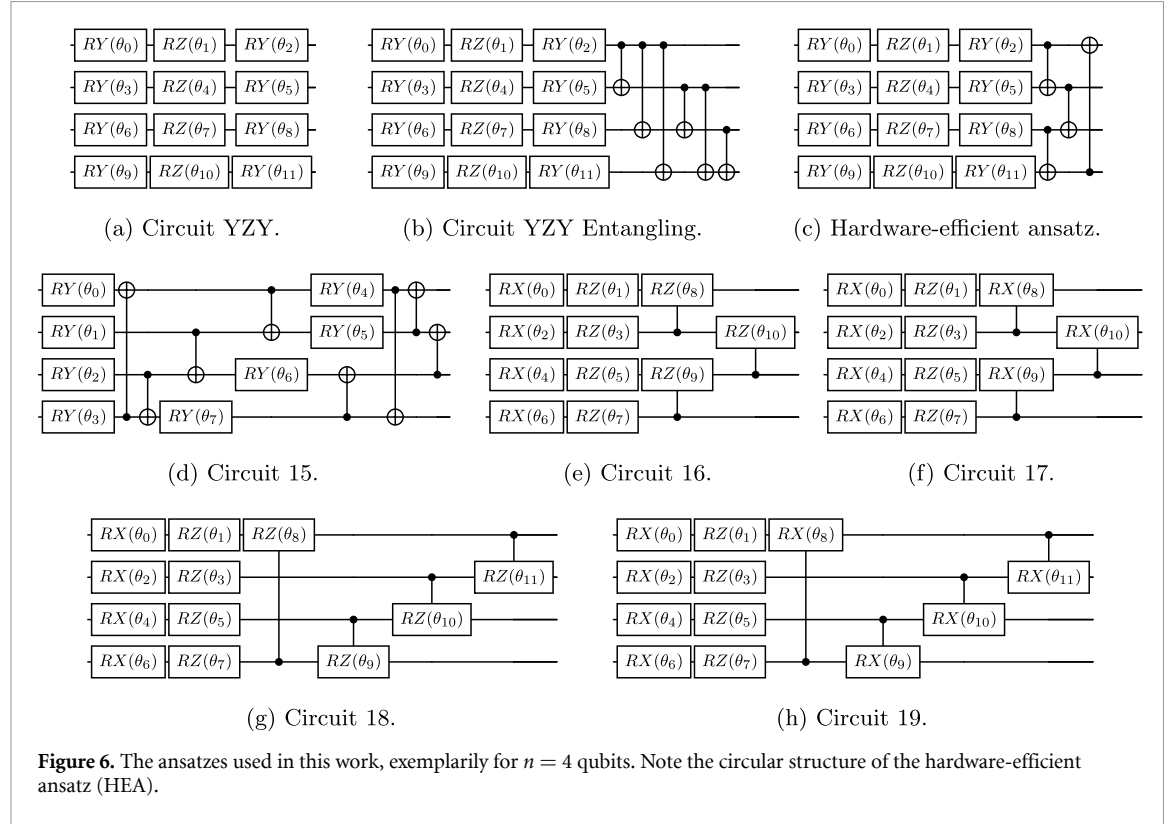
This work was supported by the Hartree National Centre for Digital Innovation, a collaboration between the Science and Technology Facilities Council and IBM.

Data availability statement

No new data were created or analysed in this study.

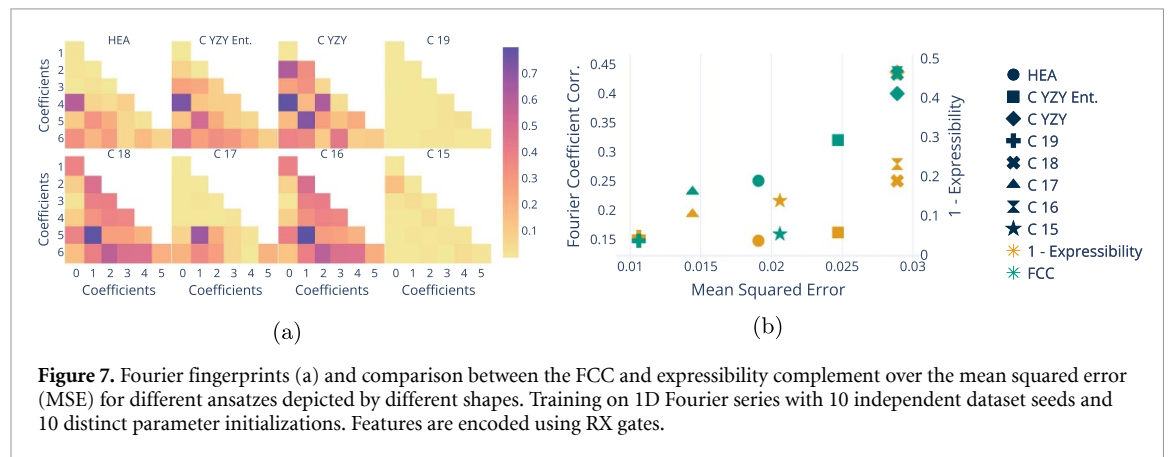
Appendix A. Ansatzes

Figure 6 shows the ansatzes used in this work. Specifically, we depict only a single variational layer and do not include the FM.



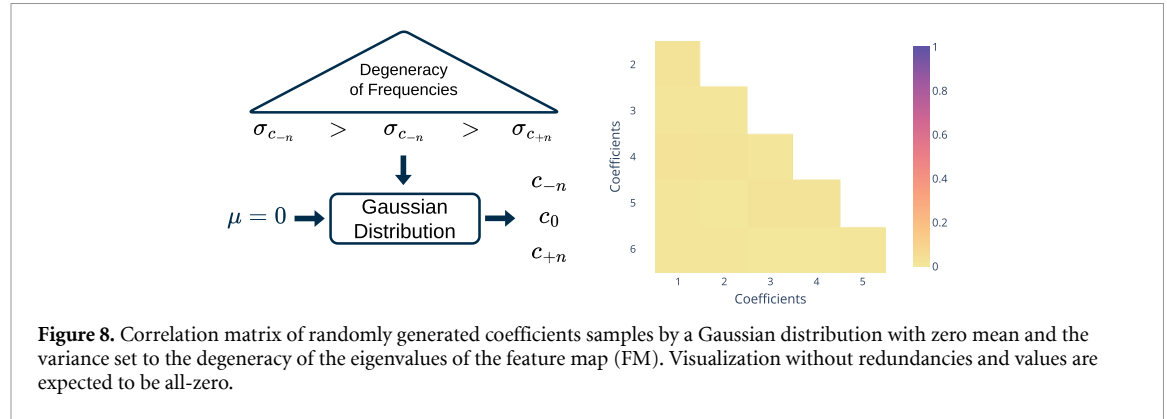
Appendix B. Effect of rotational axis on FCC

In figure 3(a) we utilize RY gates in the FM of the models as we observed an absence of imaginary parts in *Circuit 15* for RX encoding gates. For the sake of completeness, we provide the Fourier fingerprints and comparison between the FCC and expressibility in figures 7(a) and (b) respectively when the model is trained using RX encoding gates.



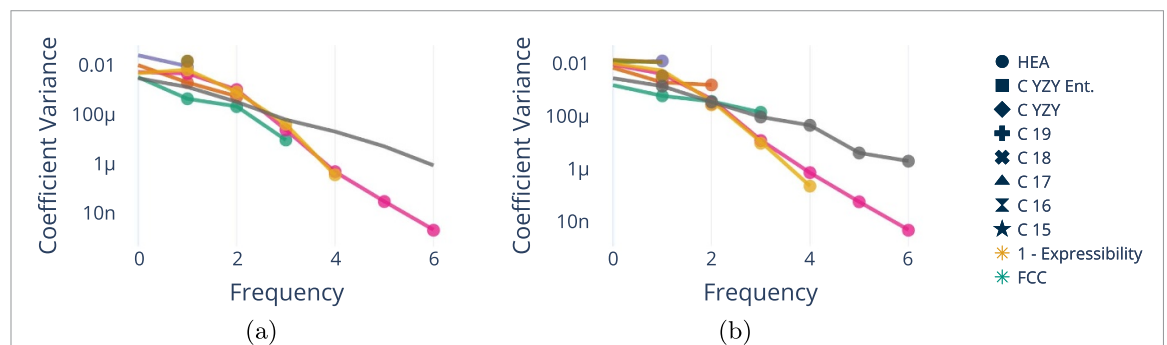
Appendix C. Impact of the variance of coefficients in the FCC

The variance of the Fourier coefficients of quantum Fourier models (QFMs) is linked to the redundancy of the frequency, as observed in [12]. For the FMs studied in this work, this means that the variance of the coefficients decays exponentially towards higher frequencies. In this section, we consider whether this can lead to spurious correlations between coefficients which are independent but have the same redundancy. This is particularly important, since the variance of the coefficients enters in the denominator of our definition of correlation in equation (5). To numerically validate this, we randomly sample a set of coefficients from a Gaussian distribution with zero mean and the variance set to the degeneracy of the eigenvalues of the FM. Applying the Pearson method r we obtain the correlation between the coefficients samples, which are shown in figure 8 and can be seen to be all-zero, as expected.



It is important that the FCC is independent of the variance of the coefficients, since, in practice, this varies in magnitude at different frequencies for different circuits. We can observe this effect in figure 9, where we show the real- and imaginary part of the variance decaying exponentially towards higher frequencies. The correlations are computed using $500 \cdot |\theta| \cdot 2^n \cdot D$ parameter samples with 10 different initial seeds on $n = 6$ qubit ansatzes. This finding is in line with a main statement from [12]. Here, specifically, we find that only *Circuit 15* and *Circuit 19* are exhibiting $\text{Var}(|c_\omega|) > 10^{-10}$ for all frequencies evaluated. Furthermore, we can see that *Circuit 15* does not exhibit an imaginary part when utilizing RX gates for the FM.

Notably, we observe that the mean value of the coefficients remains zero, i.e. that the absolute value of the coefficients is, independent of the frequency, zero-centered.



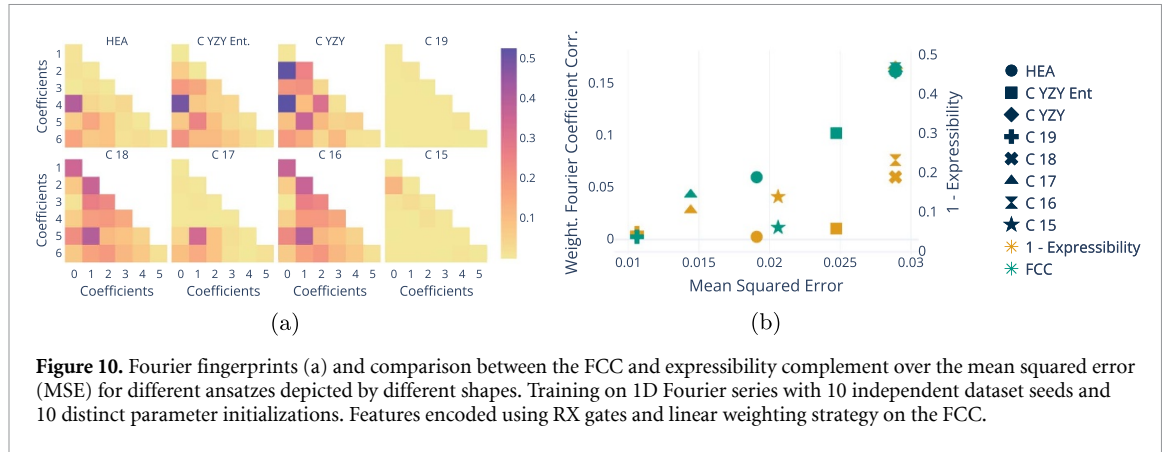


Figure 10. Fourier fingerprints (a) and comparison between the FCC and expressibility complement over the mean squared error (MSE) for different ansatzes depicted by different shapes. Training on 1D Fourier series with 10 independent dataset seeds and 10 distinct parameter initializations. Features encoded using RX gates and linear weighting strategy on the FCC.

Appendix D. Weighting the FCC

Considering the findings of appendix C, the FCC as introduced in equation (6) does not take into account the degeneracy of the spectrum towards higher frequencies. To compensate for this, we perform additional experiments where a weighting strategy is applied to the fingerprints, giving higher frequencies a lower weight for correlations. Following this approach, we can restate the FCC as

$$\overline{\text{FCC}}_{\Theta} := \frac{1}{|\Omega|} \sum_{\omega, \omega' \in \Omega} |r_{\Theta}(\omega, \omega') w(\omega, \omega')| \quad (9)$$

introducing the weighting function $w(\omega, \omega')$ which depends on the frequency pair. In the results depicted in figure 10 we applied a linear weighting $w(\omega, \omega') \sim (\omega + \omega')^{-1}$.

In figure 10(a), we can observe that, due to the weighting, correlations between coefficients with a higher frequency index are being reduced when directly comparing to figure 7(a). Furthermore, in figure 10(b) the FCC of *Circuit YZY* and *Circuit 17* adjusts slightly, such that the relation to the (MSE) becomes more linear in general. Identical experiments with Pauli-RY encoding showed a less prominent effect and are therefore omitted in this section.

Appendix E. Existence of coefficient correlations

While the existence of correlations between coefficient of the Fourier expansion of a QFM is visible from the numerical experiments we provide in this work, the work from [11, 41] provides the foundation for a more formal proof. Therefore we first restate a key finding from [11], i.e. the formula for the exact coefficients dependent on θ :

$$c_{\omega}(\theta) = \sum_{\substack{s, c \in \mathbb{N}_0^d \\ s', c' \in \mathbb{N}_0^w}} \frac{k_{s, c, s', c'} (-i)^{\sum_{j=1}^d s_j}}{2^{\sum_{j=1}^d (s_j + c_j)}} \times p_{s, c}(\omega) \prod_{k=1}^w \sin(\theta_k)^{s'_k} \cos(\theta_k)^{c'_k}. \quad (10)$$

Here $p(s, c, \omega)$ defines the sum of all coefficients belonging to ω , with s, c and s', c' counting the number of sine and cosine contributions by the input- and trainable parameters respectively. The constant $k_{s, c, s', c'}$ depends on the basis function following the notation in [11].

Proposition. Let $\theta \in [0, 2\pi)^w$ be fixed across ω . Then, even if $p(s, c, \omega)$ and $p(s, c, \omega')$ have disjoint supports, the functions $c_{\omega}(\theta)$ and $c_{\omega'}(\theta)$ can exhibit nonzero correlation due to their dependence on the shared basis functions $\{\sin(\theta_k)^{s'_k} \cos(\theta_k)^{c'_k}\}$, i.e. if we consider a single parameter $\theta_k \in \theta$ per frequency ω , the correlations between different $c_{\omega}(\theta)$ already exist due to the right-hand side of equation (10).

The dependence on a shared basis function, however, does not imply that all ansatzes have equal distribution of correlations. Depending on the circuit, $p(s, c, \omega)$ can vary, resulting in an ansatz specific FCC. As already stated in section 2.1, an ansatz with absolutely no correlation would be required to have an exponential number of parameters to account for the equally sized number of coefficients, rendering it untrainable in practice.

Appendix F. Numerical errors

Following the results of figure 3(a), we provide a summary of the numerical errors in table 1 as standard deviation $\sigma_m = \sqrt{\mathbb{E}[(\bar{m} - \mu(\bar{m}))^2]}$ of mean measurements $\bar{m}$ (expressibility, FCC and (MSE)) for different ansatzes. Similarly, we provide the errors for the 2D Fourier series (see figure 3(b)) and the HEP results (see figure 5(b)) in tables 2 and 3 respectively. For all scenarios, we observe a low standard deviation consistent for all ansatzes and measurements.

Table 1. Standard deviation of mean measurements(expressibility, FCC and (MSE)) as depicted in figure 3(a) across all seeds for different ansatzes.

σ	C 15	C 16	C 17	C 18	C 19	C YZY	C YZY Ent.	HEA
Expr.	3.6×10^{-03}	5.4×10^{-03}	3.8×10^{-03}	3.8×10^{-03}	1.4×10^{-03}	6.6×10^{-03}	1.8×10^{-03}	1.8×10^{-03}
FCC	9.6×10^{-04}	1.4×10^{-03}	1.6×10^{-03}	1.2×10^{-03}	9.3×10^{-04}	1.7×10^{-03}	1.2×10^{-03}	1.7×10^{-03}
MSE	1.0×10^{-03}	2.4×10^{-03}	6.6×10^{-05}	4.4×10^{-17}	1.5×10^{-03}	3.3×10^{-17}	2.4×10^{-03}	1.5×10^{-03}

Table 2. Standard deviation of mean measurements(expressibility, FCC and (MSE)) as depicted in figure 3(b) across all seeds for different ansatzes.

σ	C 15	C 16	C 17	C 18	C 19	C YZY	C YZY Ent.	HEA
Expr.	3.6×10^{-03}	5.4×10^{-03}	3.8×10^{-03}	3.7×10^{-03}	1.4×10^{-03}	7.6×10^{-03}	2.8×10^{-03}	1.8×10^{-03}
FCC	1.3×10^{-04}	2.9×10^{-04}	1.7×10^{-04}	2.4×10^{-04}	6.4×10^{-04}	2.5×10^{-04}	1.4×10^{-04}	1.6×10^{-04}
MSE	2.9×10^{-05}	0.0	2.2×10^{-05}	1.4×10^{-19}	3.5×10^{-05}	0.0	4.6×10^{-06}	2.0×10^{-05}

Table 3. Standard deviation of mean measurements(expressibility, FCC and (MSE)) as depicted in figure 5(b) across all seeds for different ansatzes.

σ	C 15	C 16	C 17	C 18	C 19	C YZY	C YZY Ent.	HEA
Expr.	3.6×10^{-03}	5.4×10^{-03}	3.8×10^{-03}	3.7×10^{-03}	1.4×10^{-03}	7.6×10^{-03}	2.8×10^{-03}	1.8×10^{-03}
FCC	1.3×10^{-04}	2.9×10^{-04}	1.7×10^{-04}	2.4×10^{-04}	6.4×10^{-04}	2.5×10^{-04}	1.4×10^{-04}	1.6×10^{-04}
MSE	2.9×10^{-05}	3.3×10^{-05}	3.0×10^{-05}	2.8×10^{-05}	2.4×10^{-05}	2.5×10^{-05}	3.0×10^{-05}	5.0×10^{-05}

Appendix G. Linking the FCC and the MSE

In this subsection, we derive how correlations between coefficients are reflected in the (MSE). For the sake of simplicity, we only consider one-dimensional inputs such that $\mathbf{x} \mapsto x$ and $\boldsymbol{\omega} \mapsto \omega$, while $\boldsymbol{\theta}$ remains a vector of parameters. Consider the target Fourier series $\hat{f}(x) = \sum_{\omega \in \Omega} \hat{c}_\omega \exp^{i\omega x}$ and the truncated Fourier series of the model $f(x, \boldsymbol{\theta}) = \sum_{\omega \in \Omega} c_\omega(\boldsymbol{\theta}) \exp^{i\omega x}$. We now want to minimize a dataset sampled from $x \in \mathcal{X}$ using the (MSE)

$$\begin{aligned}
 \mathcal{L}(x, \boldsymbol{\theta}) &= \frac{1}{|\mathcal{X}|} \int_{x \in \mathcal{X}} |f(x, \boldsymbol{\theta}) - \hat{f}(x)|^2 dx \\
 &= \frac{1}{|\mathcal{X}|} \int_{x \in \mathcal{X}} \left| \sum_{\omega \in \Omega} (c_\omega(\boldsymbol{\theta}) - \hat{c}_\omega) \exp^{i\omega x} \right|^2 dx \\
 &\stackrel{(1)}{=} \sum_{\omega \in \Omega} |c_\omega(\boldsymbol{\theta}) - \hat{c}_\omega|^2
 \end{aligned} \tag{11}$$

where in (1) we assume the input interval $\mathcal{X} = [0, 2\pi)$ such that $\exp^{i\omega x}$ is orthonormal over $\mathcal{X}$ and apply Parseval's identity to simplify. It should be noted that we assume the frequencies in the target function are equal to those of the QFM, which is not the case in general but suffices for the study in this work.

As we want to argue about an arbitrary $\theta \in \Theta$, but $\mathcal{L}$ depends on a specific realization of θ , we must rewrite a statistical form of the loss $\mathbb{E}_\Theta[\mathcal{L}(\theta)]$ based on equation (11) as

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{\omega} (|c_{\omega}(\theta)|^2 + |\hat{c}_{\omega}|^2 - 2\Re(c_{\omega}^*(\theta)\hat{c}_{\omega})) \\ \mathbb{E}_\Theta[\mathcal{L}(\theta)] &\stackrel{(1)}{=} \sum_{\omega \in \Omega} [\mathbb{E}_\Theta |c_{\omega}(\theta)|^2 + |\hat{c}_{\omega}|^2 - 2\Re(\mathbb{E}_\Theta[c_{\omega}(\theta)]\hat{c}_{\omega}^*)] \\ &\stackrel{(2)}{=} \sum_{\omega \in \Omega} [\mathbb{E}_\Theta |c_{\omega}(\theta)|^2 + |\hat{c}_{\omega}|^2] \\ &\stackrel{(3)}{=} \sum_{\omega \in \Omega} [\sigma_{\Theta}^2(c_{\omega}(\theta)) + |\hat{c}_{\omega}|^2]\end{aligned}\tag{12}$$

where in (1) we convert to a statistical loss with $\theta \in \Theta$ and in (2) we assume zero mean based on appendix C, i.e. $\mathbb{E}_\Theta[c_{\omega}(\theta)] = 0$ such that $\Re(\mathbb{E}_\Theta[c_{\omega}(\theta)]\hat{c}_{\omega}^*) = 0$ with $(\cdot)^*$ denoting the complex conjugate and $\Re(\cdot)$ the real part of a complex number.

Notably, $\mathcal{L}$ is separable in ω , therefore the final statement in equation (12) only captures the standard deviation of each coefficient individually, corresponding to the diagonal of the covariance matrix K . The correlation in the off-diagonal elements, however, puts additional constraints on the optimizer, as individual coefficients cannot be adjusted individually. We can capture this in $\mathcal{L}$ as well, by assuming that each coefficient is actually not dependent on a single frequency, but on all other frequencies in Ω instead, as discussed in appendix E. This can be accounted for by substituting $c_{\omega}(\theta) = \sum_{\omega' \in \Omega} \zeta(\omega, \omega'; \theta)$, where $\zeta(\omega, \omega'; \theta)$ is a proxy coefficient, that shares contributions to ω and ω' given the parameter vector θ . The proxy coefficient for $\omega = \omega'$ would then reflect the prior assumption of a fully separable loss, still allowing for a scenario where the coefficients are fully independent, i.e. where $\zeta(\omega, \omega'; \theta) = 0 \forall \omega \neq \omega'$ such that $c_{\omega}(\theta) = \zeta(\omega, \omega; \theta)$.

Given this assumption, we restate the second last equation (12) as follows:

$$\begin{aligned}\mathbb{E}_\Theta[\mathcal{L}(\theta)] &= \sum_{\omega \in \Omega} \left[\mathbb{E}_\Theta \left| \sum_{\omega' \in \Omega} \zeta(\omega, \omega'; \theta) \right|^2 + |\hat{c}_{\omega}|^2 \right] \\ &\stackrel{(1)}{=} \sum_{\omega \in \Omega} \left[\sum_{\omega', \omega'' \in \Omega} [\mathbb{E}_\Theta [\zeta(\omega, \omega'; \theta) \zeta^*(\omega, \omega''; \theta)]] + |\hat{c}_{\omega}|^2 \right] \\ &\stackrel{(2)}{=} \sum_{\omega \in \Omega} \left[\sum_{\omega', \omega'' \in \Omega} r_{\Theta}(c_{\omega'}, c_{\omega''}) \sigma_{\Theta}(c_{\omega'}) \sigma_{\Theta}(c_{\omega''}) + |\hat{c}_{\omega}|^2 \right]\end{aligned}\tag{13}$$

where in (1) we expanded the complex inner product that requires introducing two additional dummy frequencies ω' and ω'' , which can be regarded as directional interactions between frequencies.

From equation (5) we restate the Pearson correlation between coefficients of frequency ω and ω' in a slightly different form as

$$r_{\Theta}(c_{\omega}, c_{\omega'}) = \frac{K_{\Theta}(c_{\omega}, c_{\omega'})}{\sqrt{\sigma_{\Theta}^2(c_{\omega}) \sigma_{\Theta}^2(c_{\omega'})}} \stackrel{\bar{c}_{\omega}=0}{=} \frac{\mathbb{E}_\Theta |c_{\omega}(\theta) c_{\omega'}(\theta)|}{\sigma_{\Theta}(c_{\omega}) \sigma_{\Theta}(c_{\omega'})}\tag{14}$$

with $\mathbb{E}_\Theta(\cdot)$ being the expectation value of $(\cdot)$ over Θ , $\sigma_{\Omega}^2(c_{\omega}) = \text{Var}_{\Omega}(c_{\omega}) = \mathbb{E}_\Theta |c_{\omega}(\theta) c_{\omega'}(\theta)|$ being the standard deviation of c_{ω} , $K_{\Theta}(c_{\omega}, c_{\omega'})$ the covariance between c_{ω} and $c_{\omega'}$ and $\bar{c}_{\omega}$ the mean over Θ . Using the right-hand side of equation (14) we can substitute the inner term in (2), which introduces the Pearson correlation to the overall loss function.

As stated in equation (13), the FCC is the sum of over all frequency pairs of r and therefore linearly related. Therefore, the FCC is also closely related to $\mathcal{L}$ but strongly weighted by the variances of the coefficients. Following the weighting strategy introduced in appendix D, we can set $w(\omega', \omega'') =$

$\sigma_{\Theta}(c_{\omega'})\sigma_{\Theta}(c_{\omega''})$ and therefore forcefully restating the final loss equation as follows:

$$\begin{aligned}\mathbb{E}_{\Theta}[\mathcal{L}(\boldsymbol{\theta})] &= \sum_{\omega \in \Omega} \left[\sum_{\omega', \omega'' \in \Omega} r_{\Theta}(c_{\omega'}, c_{\omega''}) w(\omega', \omega'') + |\hat{c}_{\omega}|^2 \right] \\ &= \sum_{\omega \in \Omega} \left[\overline{\text{FCC}}_{\Theta} |\Omega|^{-1} + |\hat{c}_{\omega}|^2 \right] \\ &= \overline{\text{FCC}}_{\Theta} \sum_{\omega \in \Omega} |\hat{c}_{\omega}|^2.\end{aligned}\tag{15}$$

Here, only the FCC and the contribution of the squared absolute value of each target coefficient remain. The findings from [12] suggest an exponential decay of the variance towards high frequency components, providing a practical hint on estimating σ_{Θ} . This equation analytically summarizes our main contribution of this work; the loss of learning a QFM on a Fourier series with the same number of frequencies is linearly dependent on the (weighted) FCC, i.e. the correlation between the frequency components of the QFM.

Appendix H. Computing the expressibility and complexity estimate

The expressibility is a commonly used metric that quantifies how a given ansatz can cover the Hilbert space spanned by a PQC. It can be calculated using the KL-divergence [27] between the model $\int_{\boldsymbol{\theta}} (|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|)^{\otimes t} d\boldsymbol{\theta}$ and the distribution of the Haar-integral $\int_{\text{Haar}} (|\psi\rangle\langle\psi|)^{\otimes t} d\psi$ of a state t -design with $D_{\text{KL}}(\hat{P}_{\text{Model}}(F; \boldsymbol{\theta}) | P_{\text{Haar}}(F))$. Here $F = |\langle\psi_{\boldsymbol{\theta}} | \psi_{\boldsymbol{\phi}}\rangle|^2$ describes the fidelity between two states.

Generally speaking, an ansatz with a high expressibility is expected to have a high chance of reaching an arbitrary state, as parameters are likely to be sampled in the desired area of the Hilbert space. For this reason, the expressibility can be used as a metric to estimate the performance of a given ansatz. However, it should also be noted that ansatzes with high expressibility are also more prone to the Barren Plateau phenomenon as described in [28].

Calculating the expressibility is a computationally expensive task as it requires simulation of the density matrix. While the distribution of the Haar integral can be cached, the distribution of the model requires sampling the parameters and therefore varies with the number of layers. Therefore, the overall complexity is at $\mathcal{O}(MN^3)$ (matrix-matrix multiplication) where M is the number of parameter samples used to calculate the expressibility and $N = 2^n$ is the dimensionality of the Hilbert space spanned by n qubits. Furthermore, the expressibility can only be calculated in simulation, as access to the density matrix is naturally not available on real hardware.

In comparison, the FCC metric only requires computing the expectation value (matrix-vector multiplication) but adds the overhead of the FFT $\mathcal{O}(K \log(K))$ where K is the number of input samples, resulting in a complexity of the FCC being $\mathcal{O}(MN^2 K \log(K))$. To prevent aliasing, K has to be $\geq 2 \max(\Omega)$, where $\max(\Omega)$ is the highest frequency of the QFM. For the case of a D -dimensional input, the number of samples required is $K \geq (2 \max(\Omega))^D$. Therefore, the overall complexity for computing the FCC of a n -qubit QFM with D -dimensional input and the highest available frequency $\max(\Omega)$ becomes $\mathcal{O}(MN^2 (\max(\Omega))^D \log(\max(\Omega)))$, when just satisfying the sampling theorem. Considering a window length K being power of 2, the complexity of the FFT can further be reduced according to [42] to $K/2 \log_2(K)$ which, however, does not reduce the overall complexity in the asymptotic limit.

In case of having just a single-layered QFM, where the highest frequency is just n [10], it allows us to simplify the expression to $\mathcal{O}(MN^2 n^D \log_2(n))$. In a scenario where the number of qubits is equal to the number of layers, we find that $\max(\Omega) = n^2$ and the FCC complexity becomes $\mathcal{O}(MN^2 n^{2D} \log_2(n))$ which is still easier to compute than the expressibility. The worst case (and unlikely) scenario would be a single qubit QFM with n layers, where the (theoretically) highest frequency is still n , but $N = 2$ and therefore the computational complexity of the expressibility would reduce to $\mathcal{O}(M)$, while the FCC is $\mathcal{O}(M n^D \log_2(n))$. To summarize, a comparison with the expressibility complexity $\mathcal{O}(M 2^{3n})$ shows that in the limit of n , the FCC with $\mathcal{O}(MN^2 (\max(\Omega))^D \log(\max(\Omega)))$ is easier to compute, regardless of D and any (feasible) constraint on the number of layers.

Furthermore, the FCC metric is also more suitable for real hardware as the expectation value can be directly calculated on the device. In contrast, the expressibility requires full-state tomography which is exponentially costly in the number of qubits, whereas the FCC requires only a single measurement basis. The error introduced by shot noise is $1/\sqrt{N_{\text{shots}}}$ with N_{shots} being the number of shots taken and applicable to both measures.

Appendix I. High-energy physics (HEP)

I.1. Physical foundations

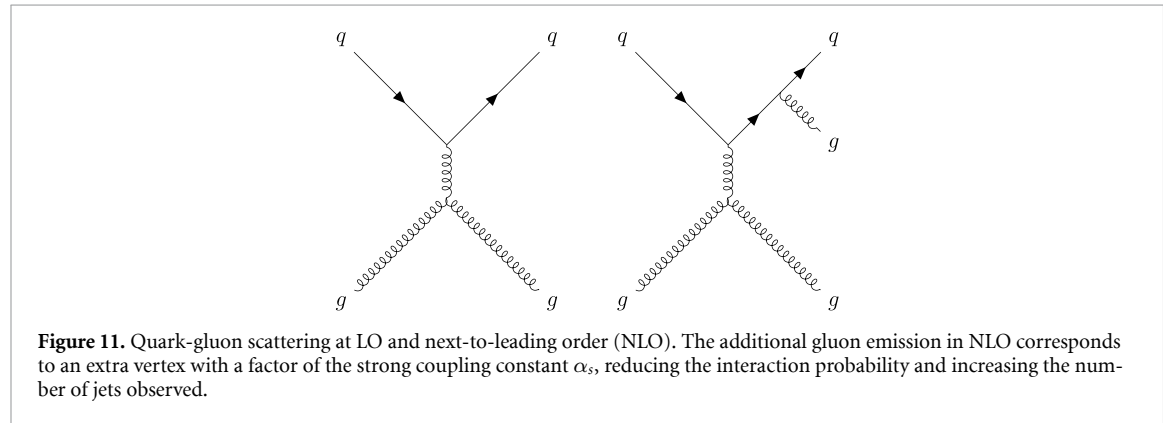
In this paragraph, we want to provide an explanation on the physical foundations of the HEP dataset used in this work. In high-energy proton-proton collisions at the large hadron collider (LHC), the dominant processes involve interactions between constituent partons; quarks and gluons. These quark-gluon scatterings lead to the production of jets, collimated sprays of hadrons that are experimentally observed as signatures of the underlying partonic activity.

At leading order (LO), quark-gluon scattering produces two outgoing partons that hadronize into two jets, corresponding to the final-state quark and gluon, see figure 11 left. The corresponding Feynman diagram features a single gluon exchange vertex.

Beyond LO, at NLO and higher, additional gluon emissions occur, see figure 11 right. Each emission corresponds to an extra vertex in the Feynman diagrams, contributing a factor of the strong coupling constant α_s , which quantifies the interaction strength of quantum chromodynamics (QCD). Since α_s runs with the energy scale, it decreases logarithmically at higher momentum transfers (asymptotic freedom). This running controls the probability of multiple gluon emissions and thus the jet multiplicity.

The presence of three or more jets arises naturally from these additional gluon emissions during the parton shower phase, reflecting the increased complexity of the event. The number of jets and their momentum distribution provide direct insight into the QCD dynamics and the value of α_s at the relevant scale.

The center-of-mass energy E_{cm} plays a central role because it sets the scale for the hard scattering process. Higher E_{cm} implies access to larger momentum transfers, reducing α_s and opening greater phase space for gluon radiation. Consequently, at LHC energies, this results in higher jet multiplicities and broader momentum spectra, making E_{cm} a fundamental metric for studying jet formation and modeling QCD interactions.



I.2. Feature importance analysis

The dataset [37] considered in this work provides the four-vector (p_x, p_y, p_z, E) , the particle id, and charge for each of the two partons as input features. Based on the four-vector one can compute the transverse momentum, pseudo-rapidity and azimuth angle. Due to the geometry of the experimental setup, only the p_z and E are non-zero entries in the input features. For the task of predicting the transverse momentum p_t of the leading jet, we first calculated the Pearson correlation between each of the input features (for both particles) and p_t as provided by the training dataset. This analysis revealed relatively strong correlations with the centre of mass energy E_{CM} and the energy differences ΔE as shown in table 4. Interestingly, the charge and particle id, being the only orthogonal features with respect to the energy, showed a relatively low correlation. This prompted us to focus solely on the energy-based values.

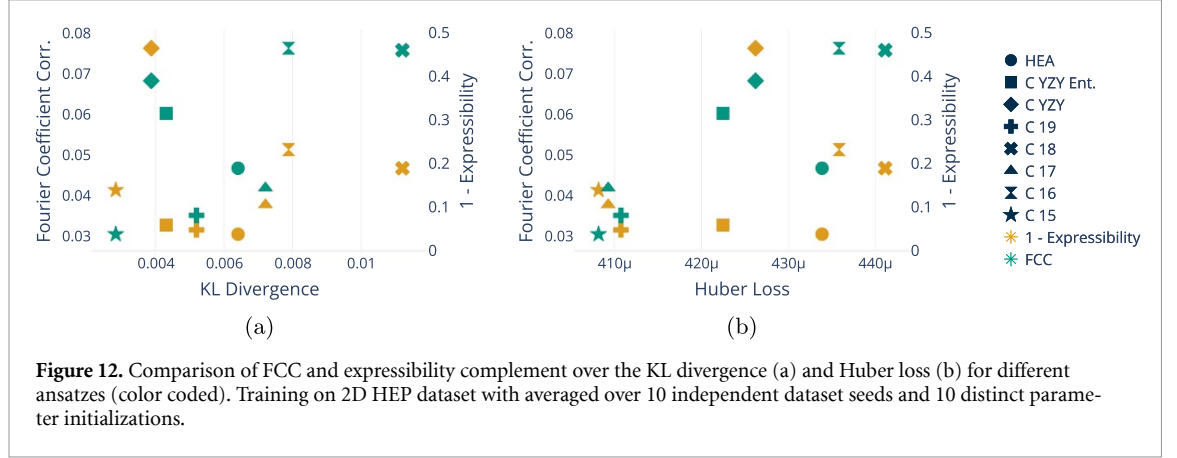
Table 4. Pearson correlation between partons and p_t of leading jet. Features separate for each parton of the collision are shown as tuple (1,2).

E_{CM}	ΔE	Charge (1,2)	Particle ID (1,2)
0.671	0.059	(0.012, 0.011)	(0.004, 0.002)

I.3. Additional numerical results

In the following, we provide additional numerical results for training on the HEP dataset where the setup is identical to section 3.2 but, instead of the (MSE), compare the KLDiv and Huber loss to the FCC and expressibility. The result of this comparison can be seen in figures 12(a) and (b) respectively.

For the KL-divergence, the FCC and expressibility both fail as predictors, although *Circuit 15* and *Circuit 18* still represent the best and worst performing circuits, as already seen in figure 5. In contrast, the results for the Huber loss are almost identical with the results for the (MSE) as shown in figure 5.



Appendix J. Sampling

J.1. Shot noise

Considering the expectation value of a QFM $f(\mathbf{x}, \boldsymbol{\theta}) = \langle 0|^{\otimes n} U^\dagger(\mathbf{x}, \boldsymbol{\theta}) \mathcal{M} U(\mathbf{x}, \boldsymbol{\theta}) |0\rangle^{\otimes n}$, it is commonly known that the true error due to a finite number of S shots is given by $1/\sqrt{S}$. Therefore, we can state $\tilde{f}$, which is the expectation value f with some shot noise $\delta(\mathbf{x})$.

$$\tilde{f}(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) \quad \text{with} \quad |\delta(\mathbf{x})| \leq \frac{1}{\sqrt{S}} \quad \forall \mathbf{x}. \quad (16)$$

Due to the linearity of the FFT and $\Delta c_\omega = \tilde{c}_\omega - c_\omega$ being the deviation in the Fourier coefficients caused by shot noise and we can say that $\Delta c_\omega = \sum_{\mathbf{x}} \delta(\mathbf{x}) \exp(i\omega\mathbf{x})$. With $|\delta(\mathbf{x})| \leq \frac{1}{\sqrt{S}}$, we have that

$$|\Delta c_\omega| \leq \sum_{\mathbf{x}} |\delta(\mathbf{x})| \cdot |e^{i\omega\mathbf{x}}| = \sum_{\mathbf{x}} |\delta(\mathbf{x})| \leq \frac{1}{\sqrt{S}} |\mathcal{X}|. \quad (17)$$

Because for a single FFT, we require $|\mathcal{X}|$ sampling points, the error for a single frequency coefficient accumulates linearly while the number of shots reduces this error by $\frac{1}{\sqrt{S}}$.

J.2. Sampling error

The Pearson correlation method requires statistical information of $c_\omega(\boldsymbol{\theta})$ over Θ . For any practical scenario, the actual number of samples M drawn from Θ is therefore important to consider. We assume that $c_\omega(\boldsymbol{\theta})$ is zero-centred (see appendix C) and because $f \in [-1, 1]$, the energy of the Fourier transform is bounded, requiring the coefficients to be bounded as well. The standard error for r given a finite sample size M is given by [43]

$$\sigma_{r_{\omega, \omega'}} \approx \sqrt{\frac{1 - r_{\omega, \omega'}^2}{(M-1)^2}}, \quad (18)$$

where $\sigma_{r_{\omega, \omega'}}$ denotes the deviation from the true Pearson correlation r that converges to 0 in the limit of $M \rightarrow \infty$. As the FCC is just the summation of $r_{\omega, \omega'}$ over all frequency pairs, we can estimate a similar error as

$$\sigma_{\text{FCC}} \approx \frac{1}{|\Omega|} \sum_{\omega, \omega'} \sqrt{\frac{1 - r_{\omega, \omega'}^2}{(M-1)^2}} \stackrel{(1)}{=} \frac{\sqrt{1 - \bar{r}^2}}{|\Omega| (M-1)} \quad (19)$$

where, in (1), we assume independence of the correlation estimates across frequency pairs, i.e. all correlations have a similar variance. Therefore, we can utilize the average correlation $\bar{r}$ of a QFM and the error in the FCC scales as $\sim \frac{1}{M}$ with the number of samples M .

J.3. Variances of coefficients

Practically, we already know that the variance of c_ω decays exponentially with the frequency (see appendix C), which we can model by $\sigma_\Theta^2(c_\omega) \sim \beta e^{-\alpha|\omega|}$ with $\beta > 0$ and $\alpha > 0$ being some constants depending on the specific FM and ansatz. Because of this exponential decay in the denominator of $r_{\omega,\omega'}, \sigma_{\text{FCC}}$ is dominated by the sampling variability for higher frequencies. We can roughly estimate this by $\sigma_{\text{FCC}} \approx \frac{1}{M} e^{\alpha(\|\omega\| + \|\omega'\|)}$. This result implies an exponential requirement on M already, if high frequency accuracy is required. However, if we consider that redundancies decay towards higher frequencies as well [12], we can expect fewer parameters contributing to high-frequent c_ω , which even increases the requirement on the number of parameter samples M .

J.4. Number of qubits

By now, we considered a fixed number of qubits, and therefore also a fixed length of the parameter vector $|\theta\rangle$ (given a particular ansatz). Changing the number of qubits naturally changes the number of parameters and, therefore, the requirement on M . However, because the Hilbert space increases with 2^n , the question arises whether there is an even stronger requirement on M than just compensating for the higher number of parameters. To show that M does not have to scale exponentially with n , we consider an average expectation value $\bar{f}$ as

$$\bar{f}(\mathbf{x}) = \frac{1}{M} \sum_{\theta \in \Theta} f(\mathbf{x}, \theta) \quad \text{with} \quad |\theta| = M. \quad (20)$$

Because the expectation value is bounded in $[-1, 1]$, $\bar{f}$ must be bounded as well, and we can apply Hoeffding's inequality

$$\mathbb{P}\left(\left|\hat{f} - \bar{f}\right| \geq \epsilon\right) \leq 2 \exp(-2M\epsilon^2) \quad (21)$$

which allows us to estimate the probability that an average $\hat{f}$ deviates from the true $\bar{f}$ by less than ϵ .

ORCID iDs

Melvin Strobl  0000-0003-0229-9897
 M Emre Sahin  0000-0002-5996-0407
 Lucas van der Horst  0009-0003-0609-9582
 Eileen Kühn  0000-0002-8034-8837
 Achim Streit  0000-0002-5065-469X
 Ben Jaderberg  0000-0001-9297-0175

References

- [1] Cerezo M, Verdon G, Huang H Y, Cincio L and Coles P J 2022 Challenges and opportunities in quantum machine learning *Nat. Comput. Sci.* **2** 567–76
- [2] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev. A* **98** 032309
- [3] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 Parameterized quantum circuits as machine learning models *Quantum Sci. Technol.* **4** 043001
- [4] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 Supervised learning with quantum-enhanced feature spaces *Nature* **567** 209–12
- [5] Lloyd S, Schuld M, Ijaz A, Izaac J and Killoran N 2020 Quantum embeddings for machine learning (arXiv:2001.03622)
- [6] Nagano L, Miessen A, Onodera T, Tavernelli I, Tacchino F and Terashi K 2023 Quantum data learning for quantum simulations in high-energy physics *Phys. Rev. Res.* **5** 043250
- [7] Jaderberg B, Gentile A A, Ghosh A, Elfving V E, Jones C, Vodola D, Manobianco J and Weiss H 2024 Potential of quantum scientific machine learning applied to weather modeling *Phys. Rev. A* **110** 052423
- [8] Pan X *et al* 2023 Experimental quantum end-to-end learning on a superconducting processor *npj Quantum Inf.* **9** 18
- [9] Chen J *et al* 2025 Quantum ensemble learning with a programmable superconducting processor *npj Quantum Inf.* **11** 83
- [10] Schuld M, Sweke R and Meyer J J 2021 The effect of data encoding on the expressive power of variational quantum machine learning models *Phys. Rev. A* **103** 032430
- [11] Wiedmann M, Periyasamy M and Scherer D D 2024 Fourier analysis of variational quantum circuits for supervised learning (arXiv:2411.03450)

- [12] Mhiri H, Monbroussou L, Herrero-Gonzalez M, Thabet S, Kashefi E and Landman J 2024 Constrained and vanishing expressivity of quantum Fourier models (arXiv:2403.09417)
- [13] Jaderberg B, Gentile A A, Berrada Y A, Shishenina E and Elfving V E 2023 Let quantum neural networks choose their own frequencies (arXiv:2309.03279)
- [14] Duffy C and Jastrzebski M 2025 Spectral bias in variational quantum machine learning (arXiv:2506.22555)
- [15] Thabet S, Monbroussou L, Mamon E Z and Landman J When quantum and classical models disagree: learning beyond minimum norm least square (arXiv:2411.04940)
- [16] Jerbi S, Fiderer L J, Nautrup H P, Kübler J M, Briegel H J and Dunjko V 2023 Quantum machine learning beyond kernel methods *Nat. Commun.* **14** 517
- [17] Landman J, Thabet S, Dalyac C, Mhiri H and Kashefi E 2022 Classically approximating variational quantum machine learning with random Fourier features (arXiv:2210.13200)
- [18] Sweke R, Recio-Armengol E, Jerbi S, Gil-Fuster E, Fuller B, Eisert J and Meyer J J 2025 Potential and limitations of random Fourier features for dequantizing quantum machine learning *Quantum* **9** 1640
- [19] Schuld M, Bocharov A, Svore K and Wiebe N 2020 Circuit-centric quantum classifiers *Phys. Rev. A* **101** 032308
- [20] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 Evaluating analytic gradients on quantum hardware *Phys. Rev. A* **99** 032331
- [21] Wierichs D, Izaac J, Wang C and Lin C Y Y 2022 General parameter-shift rules for quantum gradients *Quantum* **6** 677
- [22] Shannon C E 1949 Communication in the presence of noise *Proc. IRE* **37** 10–21
- [23] Strobl M, Franz M, Kuehn E, Mauerer W and Streit A 2025 QML essentials – A framework for working with quantum Fourier models (arXiv:2506.06695) [quant-ph]
- [24] Freedman D, Pisani R and Purves R 2007 *Statistics (International Student edn)* (WW Norton & Company)
- [25] Sim S, Johnson P D and Aspuru-Guzik A 2019 Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms *Adv. Quantum Technol.* **2** 1900070
- [26] Kingma D P and Ba J Adam: a method for stochastic optimization (arXiv:1412.6980)
- [27] Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79–86
- [28] Ragone M, Bakalov B N, Sauvage F, Kemper A F, Marrero C O, Larocca M and Cerezo M 2024 A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits *Nat. Commun.* **15** 7172
- [29] Franz M, Strobl M, Chaichenets L, Kuehn E, Streit A and Mauerer W 2025 Out of tune: demystifying noise-effects on quantum Fourier models (arXiv:2506.09527 [quant-ph])
- [30] Schwägerl T, Issever C, Jansen K, Khoo T J, Kühn S, Tüysüz C and Weber H 2023 Particle track reconstruction with noisy intermediate-scale quantum computers (arXiv:2303.13249)
- [31] Belis V, Woźniak K A, Puljak E, Barkoutsos P, Dissertori G, Grossi M, Pierini M, Reiter F, Tavernelli I and Vallecorsa S 2024 Quantum anomaly detection in the latent space of proton collision events at the LHC *Commun. Phys.* **7** 334
- [32] Martínez de Lejarza J J, Wu H Y, Kyriienko O, Rodrigo G and Grossi M 2025 Quantum chebyshev probabilistic models for fragmentation functions *Commun. Phys.* **8** 448
- [33] Di Meglio A et al 2024 Quantum computing for high-energy physics: state of the art and challenges *PRX Quantum* **5** 037001
- [34] Ngairangbam V S, Spannowsky M and Takeuchi M 2022 Anomaly detection in high-energy physics using a quantum autoencoder *Phys. Rev. D* **105** 095004
- [35] Tüysüz C, Carminati F, Demirköz B, Dobos D, Fracas F, Novotny K, Potamianos K, Vallecorsa S and Vlimant J R 2020 Particle track reconstruction with quantum algorithms *EPJ Web Conf.* **245** 09013
- [36] Guan W, Perdue G, Pesah A, Schuld M, Terashi K, Vallecorsa S and Vlimant J R 2021 Quantum machine learning in high energy physics *Mach. Learn.: Sci. Technol.* **2** 011003
- [37] Haddad Y and Grossi M 2025 Parton-to-jet mapping proton-proton dataset at 14 TeV Zenodo (<https://doi.org/10.5281/zenodo.15689696>)
- [38] Peters E and Schuld M 2022 Generalization despite overfitting in quantum machine learning models (arXiv:2209.05523)
- [39] Shin S, Teo Y S and Jeong H 2023 Exponential data encoding for quantum supervised learning *Phys. Rev. A* **107** 012422
- [40] Kyriienko O, Paine A E and Elfving V E 2024 Protocols for trainable and differentiable quantum generative modeling *Phys. Rev. Res.* **6** 033291
- [41] Nemkov N A, Kiktenko E O and Fedorov A K 2023 Fourier expansion in variational quantum algorithms *Phys. Rev. A* **108** 032406
- [42] Cooley J W and Tukey J W 1965 An algorithm for the machine calculation of complex Fourier series *Math. Comput.* **19** 297–301
- [43] Bowley A L 1928 The standard deviation of the correlation coefficient *J. Am. Stat. Assoc.* **23** 31–34