

Out of Tune: Demystifying Noise-Effects on Quantum Fourier Models

Maja Franz^{1†}, Melvin Strobl^{2†}, Leonid Chaichenets²,
Eileen Kuehn², Achim Streit², Wolfgang Mauerner^{1, 3}

¹Technical University of Applied Sciences Regensburg, Germany.

²Karlsruhe Institute of Technology, Germany.

³Siemens AG, Technology, Munich, Germany.

Contributing authors: maja.franz@othr.de; melvin.strobl@kit.edu;

[†]These authors contributed equally to this work.

Abstract

Variational quantum algorithms have received substantial theoretical and empirical attention. As the underlying Variational Quantum Circuits (VQCs) can be represented by Fourier series that contain an exponentially large spectrum in the number of input features, hope for quantum advantage remains. Nevertheless, it remains an open problem if and how Quantum Fourier Models (QFMs) can concretely outperform classical alternatives, as the eventual sources of non-classical computational power (for instance, the role of entanglement) are far from being fully understood. Likewise, hardware noise continues to pose a challenge that will persist also along the path towards fault tolerant quantum computers.

In this work, we study VQCs with Fourier lenses, which provides possibilities to improve their understanding, while also illuminating and quantifying constraints and challenges. We seek to elucidate critical characteristics of QFMs under the influence of noise. Specifically, we undertake a systematic investigation into the impact of noise on the Fourier spectrum, expressibility, and entangling capability of QFMs through extensive numerical simulations and link these properties to training performance. The insights may inform more efficient utilisation of quantum hardware and support the design of tailored error mitigation and correction strategies.

Decoherence imparts an expected and broad detrimental influence across all ansätze. Nonetheless, we observe that the severity of these deleterious effects varies among different model architectures, suggesting that certain configurations may exhibit enhanced robustness to noise and show computational utility.

Keywords: Quantum Machine Learning, Fourier Analysis, Expressibility, Entanglement

1 Introduction

Given the remarkable advancements in machine learning, significant optimism has been directed towards Quantum Machine Learning (QML). However, such enthusiasm has frequently been tempered by the prevailing limitations of contemporary quantum hardware [1], together with algorithmic shortcomings [2]. The exact capabilities of QML are not yet fully understood, but considerable opportunities to achieve advantage over classical approaches remain. As the concept of QML (see the introductory review by Schuld and Petruccione [3]) has received considerable attention in the literature during the last years, crucial limitations and trade-offs have been identified [4], especially regarding trainability challenges [5] and model complexity constraints [6]. Despite the progress from Noisy Intermediate Scale Quantum (NISQ) devices [7] towards Fault Tolerant Quantum Computing (FTQC), noise and imperfections will continue to influence QML performance and properties in the foreseeable future, even when considering the best possible use of already available hardware resources [8].

In this work, we aim to shed light on the impact of noise on key properties of Variational Quantum Circuits (VQCs), the core computational resource of QML. Specifically, we focus on Quantum Fourier Models (QFMs), a representation of VQCs as truncated Fourier series. We emphasise that our results are hardware agnostic, that is, they are not only applicable to error-corrected architectures, but also to intermediate Quantum Processing Units (QPUs) and algorithms (*e.g.* as in Refs.[9]–[15]), until full FTQCs are available.

Common metrics used to assess VQCs include expressibility and entangling capability [16], which generally determine how effectively a VQC can explore the Hilbert space. This exploration is crucial because it indicates the capacity of a VQC to learn various functions, with the solutions to these functions residing in this Hilbert space. However, a highly expressive VQC can face challenges such as the Barren-Plateau (BP) problem [5], where the optimisation landscape becomes difficult to navigate. Therefore, a successful VQC must strike a balance: It needs to be sufficiently expressive so that the portion of the Hilbert space it accesses contains potential solutions, while maintaining a manageable optimisation landscape. Another way to gauge this property is by examining the number and values of the QFM’s Fourier coefficients [17], which dictate the types of functions the QFM can learn. Conversely, if an ansatz lacks sufficient expressibility, the *trainable Hilbert space* it accesses is highly unlikely to contain viable solutions.

Entangling capability is another metric to determine the amount of entanglement a VQC can create using optimal parameters. This metric can be viewed as the *quantumness* a particular circuit can achieve and is therefore crucial to set it apart from classical surrogates or VQCs without entanglement.

In this work, we determine the Fourier spectral properties, expressibility, and entangling capability of commonly used ansätze under the influence of noise using extensive numerical simulations. Furthermore we link these characteristics to the performance of the QFMs in machine learning tasks by conducting trainings on synthetically generated regression problems.

Our approach aims to provide a nuanced perspective on the factors that contribute to the aptitude of QML for specific learning tasks. The main contributions of this work

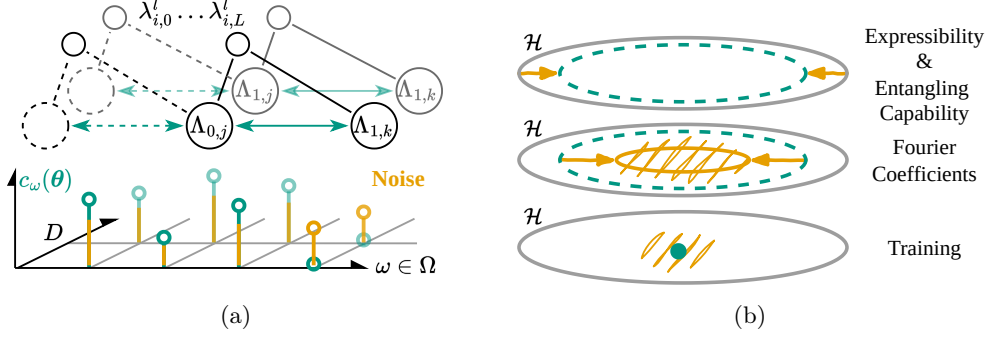


Fig. 1: (a) Quantum spectrum tree [17] where Λ is constituted by the sum of the individual eigenvalues λ of each generator in the D -dimensional input encoding. The spectrum Ω is determined by the differences between each Λ where each coefficient $c_\omega(\theta)$ depends on a parameter vector θ . Noisy coefficients generally have a reduced magnitude, which can lead to non-trainable frequency components. (b) Implications of the left considerations on Hilbert space level (solid grey), denoted by \mathcal{H} : Expressibility and entangling capability limits depend on the level of noise (arrows) and the accessible space (dashed) of the model. Additionally, noise further reduces this space when viewed through the Fourier lens, as it reduces the ability to change frequency components while adding untrainable frequency artifacts (hatched outside of solid orange). Similarly, in a training scenario, noise causes a shift away from the target space as frequency components get affected differently by noise.

are illustrated in Fig. 1. Additionally, we provide a consistent collection of notation on the QFM characteristics, together with a software reproduction package [18] that encompasses the computation of the relevant metrics using our *QML-Essentials* Python framework [19].

The rest of this paper is structured as follows: Before we review related work in Sec. 3, we introduce the concept behind QFMs in Sec. 2. We then present our methods in Sec. 4, where we introduce the noise models used in this work, discuss how we measure coefficients, entangling capability and expressibility, and define a synthetic dataset on which we train the QFMs. Sec. 5 presents our main contribution, which is the numerical results on the computation of the aforementioned characteristics in VQCs. Here we measure Fourier coefficients, expressibility, and entangling capability in a quantitative manner through random parameter sampling, and also in a synthetic training setting. In Sec. 6 we provide a discussion and finally conclude in Sec. 7.

2 Quantum Fourier Models

We consider an n -qubit quantum circuit whose expectation value, measured on each qubit (n -local) using the observable \mathcal{M} , is given by

$$f(\mathbf{x}, \boldsymbol{\theta}) = \langle 0|^{\otimes n} U^\dagger(\mathbf{x}, \boldsymbol{\theta}) \mathcal{M} U(\mathbf{x}, \boldsymbol{\theta}) |0\rangle^{\otimes n}, \quad (1)$$

which is parametrised by the D -dimensional input $\mathbf{x} = (x_1, \dots, x_D) \in \mathbb{R}^D$ and the trainable parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{L+1}) \in [\mathbb{R}^K]^{L+1}$ with L layers of K parameters per layer¹.

We build the eventual unitary $U(\mathbf{x}, \boldsymbol{\theta})$ as products of multiple layers ($\ell \in [1, L]$) of consecutive encoding unitaries $S^{(\ell)}(\mathbf{x})$ and trainable unitaries $W^{(\ell)} := W^{(\ell)}(\boldsymbol{\theta}_\ell)$ as follows:

$$U(\mathbf{x}) = W^{(L+1)} S^{(L)}(\mathbf{x}) W^{(L)} \dots W^{(2)} S^{(1)}(\mathbf{x}) W^{(1)} \quad (2)$$

where an additional last trainable layer $W^{(L+1)}$ is added. To allow for an arbitrary circuit structure and placement of unitary and encoding gates, we construct $S^{(\ell)}(\mathbf{x})$ to comprise combinations of D unitaries $S_i^{(\ell)}(\mathbf{x})$, each of which encodes one single input feature x_i :

$$S_i^{(\ell)}(\mathbf{x}) = \exp(-i \mathbf{e}_i^T \mathbf{x} G_i^{(\ell)}) = \exp(-i x_i G_i^{(\ell)}) \quad (3)$$

with Hermitian generator $G_i^{(\ell)}$ and standard basis vector \mathbf{e}_i .

As shown in the seminal contributions in Refs. [20], [21], such an architecture allows for rewriting Eq. 1 as a partial Fourier series

$$\begin{aligned} f(\mathbf{x}, \boldsymbol{\theta}) &= \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} c_{\boldsymbol{\omega}}(\boldsymbol{\theta}) e^{i \boldsymbol{\omega}^T \mathbf{x}} \\ &= \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} |c_{\boldsymbol{\omega}}(\boldsymbol{\theta})| (\cos(\varphi(\boldsymbol{\theta})) \cos(\boldsymbol{\omega}^T \mathbf{x}) - \sin(\varphi(\boldsymbol{\theta})) \sin(\boldsymbol{\omega}^T \mathbf{x})), \end{aligned} \quad (4)$$

where $\boldsymbol{\Omega}$ contains the unique frequencies resulting from the eigenvalues of each generating encoding Hamiltonian $G_i^{(\ell)}$. The corresponding complex Fourier coefficients are given by $\{c_{\boldsymbol{\omega}}(\boldsymbol{\theta})\} = \{|c_{\boldsymbol{\omega}}(\boldsymbol{\theta})| e^{i \varphi(\boldsymbol{\theta})}\}$.

Following the definition in Ref. [17], the resulting set of frequencies is

$$\boldsymbol{\Omega} = \left\{ \times_{i=1}^D \{ \Lambda_{i, \mathbf{j}} - \Lambda_{i, \mathbf{k}} \} \mid \mathbf{j}, \mathbf{k} \in \llbracket 1, d_\ell \rrbracket_{\ell=1}^L \right\}, \quad (5)$$

where the bracket $\llbracket \cdot \rrbracket$ defines the product of the intervals $[\cdot]$, s.t. \mathbf{j} (and \mathbf{k}) is a multi-index (j_1, j_2, \dots, j_L) with each $j_\ell \in [1, d_\ell]$. The corresponding $\Lambda_{\mathbf{j}}$ presents the sum of all d_ℓ eigenvalues of the ℓ -th generator grouped across all layers, given by

$$\Lambda_{i, \mathbf{j}} = \lambda_{i, j_1} + \dots + \lambda_{i, j_L}. \quad (6)$$

Here, Eq. 5 restates a finding from Ref. [22] where it was shown that the set of frequencies is constituted by the Cartesian product over all the individual sets that

¹Without loss of generality, we choose the same number of parametrised gates for each layer.

stem from each input encoding. While there are no constraints on the position of each input encoding unitary inside the circuit, it is important to note that the useable set of frequencies only extends if all data encoding unitaries are a mutually orthogonal set of functions.

Another descriptive metric of the Fourier spectrum that also provides an upper bound for the expressibility was introduced in Ref. [17] in form of the frequency redundancy which is defined by the size of the redundancy generator

$$R(\omega) = \{(\mathbf{j}, \mathbf{k}) \mid \times_{i=1}^D \{\Lambda_{i,\mathbf{j}} - \Lambda_{i,\mathbf{k}}\} = \omega \wedge \mathbf{j}, \mathbf{k} \in \llbracket 1, d_\ell \rrbracket_{\ell=1}^L\}. \quad (7)$$

Considering single qubit Pauli-encodings, by increasing the number of layers, the number of distinct frequencies $|\Omega|$ grows linearly, whereas the number of redundancies $\sum_{\omega \in \Omega} |R(\omega)|$ grows exponentially. Notably, it was also shown in Ref. [20], that the exact same result can be achieved by increasing the number of qubits instead, such that $|\Omega| \sim (nL)^D$. Generally, the increasing number of frequencies stem from the gaps between the eigenvalues of the feature map generating Hamiltonians as visualised in Fig. 1a.

The coefficients $c_\omega(\theta)$, parametrised by the trainable parameters in the VQC, are solely determined by the circuit architecture (also including the input encoding structure [23]). The way parameters act on the coefficients is non trivial with an analytical description of this relation as published in Refs. [24], [25].

3 Related Work

Considering noise in the context of QFMs involves a multitude of relevant aspects, ranging from *Fourier analysis* through *trainability*, *dequantisation*, *expressibility*, and *entanglement*. In this section, we review the main body of relevant related work.

The seminal work of Pérez-Salinas *et al.* [21] introduced the *data reuploading* technique that enables a non-linear transformation between input and output of a VQC with unitary transformations. A follow-up study by Schuld, Sweke, and Meyer [20] sparked a whole subfield of research within QML by establishing how the spectrum of a QFM changes with the input encoding. Casas and Cervera-Lierta [22] extended the idea to multidimensional input features and found that for some ansätze, the size of the spectrum grows faster than the available degrees in a Hilbert space. A fixed encoding strategy, as used in the aforementioned papers, results in an evenly spaced spectrum. Jaderberg *et al.* [26] found that by adding a trainable parameter to the input, distances between the frequencies can be modified. The expressibility of QFM and the redundancies of eigenvalues that effectively determine the resulting spectrum was linked by Mhiri *et al.* [17], which also demonstrated that the variance of a frequencies coefficients linearly depends on the number of its redundancies. Nemkov, Kiktenko, and Fedorov [24] introduced an analytical description between Fourier coefficients and trainable parameters. Wiedmann, Periyasamy, and Scherer [25] use this analytical description to argue that certain coefficients (frequencies) can vanish depending on the trainable parameters, which leads to a reduction of the accessible spectrum.

While there has been a significant effort to build mathematical formulations and find analytical relations around QFMs, the trainability of such models has to be regarded

as well. One of the fundamental problems is the Barren-Plateau (BP) phenomenon, introduced in Ref. [27]. Generally, it describes the exponential decay of gradients in a VQC caused by (i) the circuit expressibility; (ii) degree of entanglement; (iii) locality of the observable [5], [28] and (iv) noise [5], [29].

Simply avoiding BPs might look like an obvious solution, but it is known that circuits without a BP can be classically simulated, eliminating any potential quantum advantage [30] for a majority of use cases. Also, Ref. [23] showed that the encoding strategy plays a significant role in the trainability of QML models in terms of avoiding BPs.

To assess whether QML offers advantages over classical Machine Learning (ML), one approach is to employ dequantisation techniques that seek to understand if classical alternatives constructed from a quantum circuit can match or surpass QML performance. Schreiber, Eisert, and Meyer [31] demonstrated that such surrogate models can outperform QFMs for small problem instances, but this becomes intractable as the number of Fourier coefficients grows exponentially with input features. To address this issue, Fontana *et al.* [32] and Landman *et al.* [33] proposed approximating VQC outcomes by trimming frequencies or using Random Fourier Features (RFFs). Sweke *et al.* [34] established that efficient dequantisation for regression is possible with RFFs if the spectrum of the QFM is polynomially concentrated, requiring a polynomial number of frequencies.

The aforementioned references consider VQC, optimisation algorithm and training data as a whole, which is reasonable from an end-to-end point of view of QML. However, this holistic way of considering QML makes it hard to assess the influence of the actual VQC *structures* that are the core computational resources. Characterising the properties of VQCs to identify suitable ansätze for variational quantum algorithms is a common approach, while the implications are often left as open questions in the field. In this context, Sim, Johnson, and Aspuru-Guzik [16] conducted an extensive study to measure the expressibility (*cf.* Sec. 4.3) and entangling capability (*cf.* Sec. 4.4) of different ansätze. Expressibility, which indicates how effectively an ansatz can explore the Hilbert space, has known relations to redundancies in the Fourier spectrum [17].

The role of entanglement in QML and other variational algorithms remains far from fully understood. When dealing with input data that is inherently quantum in nature, entanglement is often considered as the key resource for successful learning [35]. Wang *et al.* [36] demonstrated that increasing entangling capability up to a certain threshold can lead to improved model performance. However, when the input data is classical, empirical studies, such as those in Refs. [37], [38], have shown that a low-entanglement circuit can perform similarly well, or even better than highly entangled ones on specific learning tasks. Furthermore, Joch, Uhrig, and Fauseweh [39] highlighted the potential downside of excessive entanglement in learning scenarios, suggesting that high entangling capability can lead to the BP phenomenon. This indicates a critical threshold beyond which entanglement may no longer be beneficial for learning tasks. To our knowledge, there is currently no established connection in previous work between the entangling capability of a circuit and its Fourier spectrum, to which we contribute in this work.

Given that fault-tolerance is not yet fully in reach (and does not automatically imply zero error), the influence of noise on QFMs remains an important open question. Dalzell, Hunter-Jones, and Brandão [40] showed that on random quantum circuits, weak local noise translates into global white noise after measurement. More precisely, incoherent local errors get scrambled by random quantum circuits causing the output to approach a uniform distribution. Fontana *et al.* [41] measured the effect of noise on the Fourier spectrum to make suggestions for noise mitigation and diagnostics. By cutting or filtering certain noise-induced coefficients or frequencies, noiseless landscapes can be approximately reconstructed. The authors focus on the Quantum Approximate Optimisation Algorithm (QAOA) and Variational Quantum Eigensolver (VQE), for which they derive a uniform contraction of the Fourier coefficients when applying certain types of decoherent noise. Apart from Ref. [41], we are not aware of other publications that investigate the direct influence of noise on the Fourier spectrum and other properties of VQCs, especially in the QML domain. With this article, we aim to contribute to filling this research gap.

4 Method

In the following, we present our methods, which encompass the noise models employed, along with the notations and metrics utilised to assess the characteristics of QFMs. These characteristics include Fourier coefficients, expressibility, and entangling capability. Additionally, we detail the synthetic dataset constructed for training these QFMs.

4.1 Noise

We follow a standard approach of modelling decoherent noise [1], [7], [42] by three mechanisms, namely (i) damping noise (*i.e.*, environmental effects, decoherence) encompassing Amplitude Damping (AD) and Phase Damping (PD); (ii) decoherent gate errors, including Bit Flip (BF), Phase Flip (PF) and Depolarisation (DP) noise and (iii) State Preparation and Measurement (SPAM) errors. In addition to the aforementioned, we also consider Coherent Gate Errors (CGEs). Details on the modelling of the different noise types are available in [Appendix A](#).

4.2 Coefficients

Based on the definition of a QFM from [Eq. 4](#), we investigate the coefficients of such a model after parametrisation using the Fast Fourier Transform (FFT). Given the correct number of frequencies, which is in the simplified scenario of Pauli-encoding and a single encoding layer equal to the number of qubits n , this transformation yields a numerical estimate of the coefficients. Note that this set of coefficients is a numerical approximation based on the expectation value of the model given a range of input samples $x \in \mathcal{X}$ and a fixed parameter vector θ . As stated in [Eq. 4](#), these coefficients depend on the parameters. For each ansatz we assume that this set of coefficients is characterised by the mean and the standard deviation which is what we evaluate in the following numerical experiments.

We calculate the covariance matrices as

$$\text{Cov}_c = \begin{pmatrix} \text{Cov}(\text{Re}(c), \text{Re}(c)) & \text{Cov}(\text{Re}(c), \text{Im}(c)) \\ \text{Cov}(\text{Im}(c), \text{Re}(c)) & \text{Cov}(\text{Im}(c), \text{Im}(c)) \end{pmatrix}. \quad (8)$$

To estimate the model spectrum, we sample the parameter space Θ . We estimate the mean value of a coefficient contributing to the frequency ω ,

$$\mu_c(\omega) = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} |\text{FFT}_{\mathcal{X}}(f(\cdot, \theta))(\omega)|, \quad (9)$$

and the relative standard deviation

$$\sigma_c(\omega) = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} \frac{\sqrt{|\text{FFT}_{\mathcal{X}}(f(\cdot, \theta))(\omega) - \mu_c(\omega)|^2}}{\mu_c(\omega)}. \quad (10)$$

$\text{FFT}_{\mathcal{X}}$ represents the discrete Fourier transform over $x \in \mathcal{X}$. Analytical coefficients can be obtained by expanding the expectation value using trigonometric polynomials [24], [25]. Prior to our numerical experiments, we use the analytical method to cross-validate the results obtained by the FFT. This is to ensure a sufficient number of frequencies that has to be estimated in advance to satisfy the Shannon-Nyquist sampling theorem, *i.e.* $|\mathcal{X}| \geq 2 \max(\Omega)$.

4.3 Expressibility

To determine expressibility, we utilise the Kullback-Leibler (KL) divergence between the distributions (i) obtained by sampling from the Haar integral $\int_{\text{Haar}} d\psi (|\psi\rangle\langle\psi|)^{\otimes t}$ of a state t -design and (ii) obtained from the model $\int_{\theta} d\theta (|\psi_{\theta}\rangle\langle\psi_{\theta}|)^{\otimes t}$ [16], [43]:

$$D_{\text{KL}} \left(\hat{P}_{\text{Model}}(F; \theta) \| P_{\text{Haar}}(F) \right). \quad (11)$$

Here, the fidelity $F = |\langle\psi_{\phi} | \psi_{\phi}\rangle|^2$ is based on the state overlap whereas the distributions of state overlaps is then $p(F)$.

This metric yields zero if $\hat{P}_{\text{Model}}(F; \theta) = P_{\text{Haar}}(F)$, meaning the states sampled from the QFM are Haar distributed. For the least expressive case, that is, the empty circuit $U = \mathbb{1}$, the KL divergence becomes $\ln(n_{\text{bins}})$ where n_{bins} describes the number of bins that are used for discretising the probability distribution using a histogram. For the remainder of this work, we refer to the expressibility as the inverse of KL divergence.

4.4 Entangling Capability

There are different ways to calculate the entangling capability of a VQC (as there are many different ways of quantifying entanglement [44]).

4.4.1 Meyer-Wallach Measure

The Meyer-Wallach (MW) entangling capability [45], [46] of a state $|\psi\rangle$ is defined as

$$Q(|\psi\rangle) = 2 \left(1 - \frac{1}{n} \sum_{k=0}^{n-1} \text{Tr} [\rho_k^2] \right), \quad (12)$$

based on partial density matrices ρ_k , where k denotes the subsystem (*i.e.*, qubit index).

This metric has the property that if $\text{Tr} [\rho_j^2] = 1 \quad \forall j$, implying $Q = 0$, $|\psi\rangle$ is a product state whereas $Q = 1$ iff $\text{Tr} [\rho_k^2] = 1/2 \quad \forall k$ meaning the state is maximally mixed. Notable, this metric is restricted to pure states, but cannot be used for mixed states as they occur in decoherent noisy circuits (despite working with mixed states, we need the measure nonetheless, as will become clear in the next section).

4.4.2 Entanglement of Formation

The Entanglement of Formation (EF) [47] can be used as a metric on mixed states. Following Ref. [44], it is usually defined as

$$Q(|\psi\rangle) := \inf \left\{ \sum_i p_i E(|\psi_i\rangle \langle \psi_i|) : \rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| \right\}, \quad (13)$$

and represents the average entanglement over all pure state decompositions of a density matrix ρ . The entanglement for pure states is then calculated using the MW measure as introduced in Eq. 12. Finding a pure state decomposition however is a non-trivial task. In this work we utilise an eigenvalue and eigenstate decomposition of the density matrix ρ . Based on the resulting decomposition, we proceed with calculating the entanglement of each eigenstate while weighting it by its eigenvalue as depicted in the left part of the right hand side of Eq. 13.

4.5 Synthetic Training Dataset

To analyse the behaviour of QFMs during optimisation, we employ random Fourier series as objective functions for regression tasks:

$$f'(\mathbf{x}) = \sum_{\boldsymbol{\omega} \in \Omega} c'_{\boldsymbol{\omega}} e^{i\boldsymbol{\omega}^T \mathbf{x}} \quad \text{with} \quad c'_{\boldsymbol{\omega}} = a r_{\boldsymbol{\omega}} \left| \sum_{\boldsymbol{\omega} \in \Omega} r_{\boldsymbol{\omega}} e^{i\boldsymbol{\omega}^T \mathbf{x}} \right|^{-1}. \quad (14)$$

Here, normalisation factor a and $r_{\boldsymbol{\omega}} \sim \text{Uniform}([0, 1])$ are uniformly randomly sampled complex (or real-valued for $\boldsymbol{\omega} = \mathbf{0}$) numbers that satisfy the symmetry constraints of a real-valued Fourier series. The normalisation factor, set to $a = 0.5$ in our experiments, restricts the values of $f'(\boldsymbol{\omega})$ to the interval $[-a, a]$, such that they are attainable by the QFM. The number of frequencies in the objective Fourier series is set to match the number of expected frequencies in the QFM $|\Omega|$ and the input data consists of evenly spaced points $\mathbf{x} = (x_1, \dots, x_s)$ in the interval $[-\pi, \pi]$, where $s = |\Omega|$. While we only

focus on one-dimensional input data ($D = 1$) in the training experiments of this work, the approach can be extended to also generating higher-dimensional training datasets.

To evaluate the performance of a QFM $f(\mathbf{x}, \boldsymbol{\theta})$, we use Mean Squared Error (MSE) $\text{MSE}(f', f) = \frac{1}{s} \sum_{i=1}^s (f'(x_i) - f(x_i, \boldsymbol{\theta}))^2$ as metric. Additionally, we introduce a difference metric $\Delta c_{\boldsymbol{\omega}}$ to assess how closely a QFM approximates the exact Fourier coefficients of the target function for each $\boldsymbol{\omega}$:

$$\Delta c_{\boldsymbol{\omega}} = ||c_{\boldsymbol{\omega}}(\boldsymbol{\theta})| - |c'_{\boldsymbol{\omega}}||. \quad (15)$$

5 Numerical Results

We consider a circuit with the structure introduced in [Eq. 1](#) with $L = 1$ layers and $n \in [3 \dots 6]$ qubits. In the coefficient experiments, where we sample parameters, $D \in \{1, 2\}$ input features are encoded, while $D = 1$ in the training experiments. By convention, such a model has two trainable layers $U(\mathbf{x}) = W^2 S^1(\mathbf{x}) W^1$. However, for calculating expressibility and entanglement we discard encoding gates and the second trainable layer to make the results more consistent with [Ref. \[16\]](#). For the trainable part, we investigate the ansätze depicted in [Fig. 14](#) of [Appendix B](#). We use the Strongly-Entangling Ansatz (SEA) as introduced in [Ref. \[48\]](#) and a circular structure for the Hardware-Efficient Ansatz (HEA). Both of these ansätze are commonly-used within the QML-community: The SEA presents an ansatz that utilises comparatively many entangling gates and trainable parameters; the HEAs uses fewer gates and its depth only grows as a constant with the number of qubits, which is a desirable property for NISQ circuits. Circuit 15 (C15) and Circuit 19 (C19) stem from [Ref. \[16\]](#) and are chosen based on their different structural properties (*cf.* [Fig. 3](#)) and distinguishable expressibility and entangling capability. In particular, C15 presents an ansatz that only uses one type of rotation gate, which is potentially easier to implement on quantum hardware. C19 is chosen as a representative of ansätze that use controlled rotation gates as an entangling gate, instead of CNOT gates, which may provide an additional degree of variability in the entangling capability (*cf.* [Sec. 5.3](#)).

In the following experiments, we apply the types of noise described in [Sec. 4.1](#) to each ansatz with seven linearly increasing noise levels from 0% to 3% in steps of 0.5%. The trainable parameters are randomly uniformly sampled from $[0 \dots 2\pi]$. For each parameter sample, we perform n local measurements on all qubits and then collect and average the expectation value over all samples. To ensure stability against parameter initialisation, we use different seeds upon model initialisation. Throughout all our experiments, we use 250 samples per parameter value for each seed, which is scaled exponentially with the number of qubits if not stated otherwise (*e.g.*, for a three qubit circuit with 10 parameters, we sample $10 \times 250 \times 2^3$ times for each seed). While we use ten different initialisation seeds for the training, all other experiments in [Sec. 5.1](#), [Sec. 4.3](#) and [Sec. 5.3](#) are conducted with five seeds.

In addition to sampling parameters randomly, we also train QFMs on ten generated Fourier series (*cf.* [Sec. 4.5](#)), while applying the aforementioned types of noise and circuits, using a noise level of 3%. Specifically, we observe as time series during the

course the training, (i) the MSE; (ii) the absolute coefficient values $\{|c_\omega|\}$; (iii) the difference Δc_ω to the target coefficients and (iv) the entangling capability

Our results are fully reproducible [49], with the code available in Ref. [18].

5.1 Coefficients

We consider the effect of noise on the Fourier coefficients of a QFM. As the standard deviation across seeds is small (*i.e.*, $< 10\%$ for $\mu_c(\omega)$), figures show the mean of $\mu_c(\omega)$, $\sigma_c(\omega)$ or Cov_c over the five seeds. Samples for a coefficient with a mean $\mu_c(\omega)$ below 1×10^{-14} are taken to be zero, as the corresponding frequency ω is effectively not contained in the spectrum.

5.1.1 Input Encoding

As the encoding strategy substantially impacts coefficient values [23], we first investigate in three different rotational Pauli encodings around X , Y , and Z axis to encode one-dimensional data ($D = 1$) in a noiseless setting and observe how this not only changes the spectrum, but also the real and imaginary part. Fig. 2 shows that all circuits suffer from a strong decay in the absolute coefficient value as well as in the corresponding relative standard deviation with increasing frequency. The magnitude of this decay varies with circuit structure, that is C19 exhibits lower $\mu_c(\omega)$ at higher frequencies than the other ansätze. These findings are in line with Ref. [17]. Generally speaking, the ansatz determines the spectral properties of the QFM with $c_\omega(\theta)$ where the exact correlation between θ and ω varies. Furthermore, the HEA with four to six qubits, and C15 with five qubits, resulted in $\mu_c(\omega)$ smaller than 1×10^{-14} for some frequencies, not exhibiting a full spectrum. However, we note that the absolute coefficient for individual parameter samples is above the threshold, indicating that a full spectrum is possible with these ansätze, but unlikely due to the sampling resolution in parameter space.

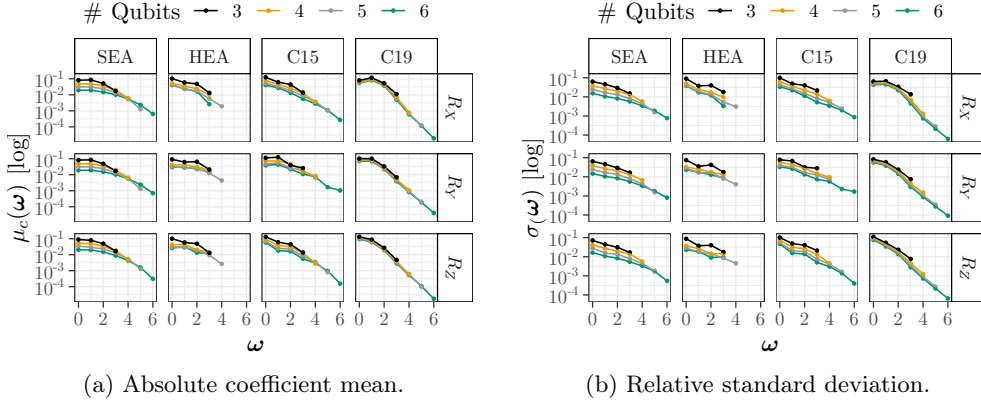


Fig. 2: Absolute coefficient mean $\mu_c(\omega)$ and the corresponding relative standard deviation $\sigma_c(\omega)$ for $[3 \dots 6]$ qubits and $R_{\{X,Y,Z\}}$ encodings over frequencies.

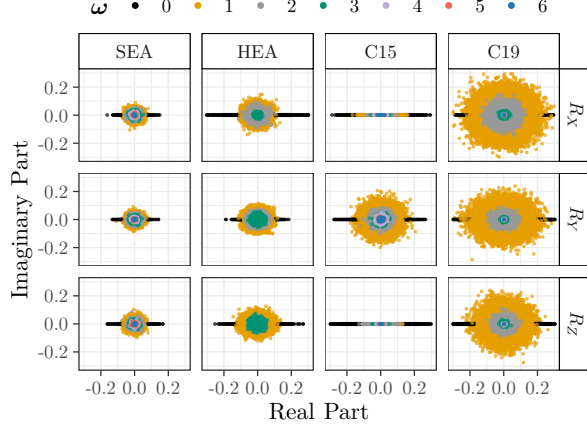


Fig. 3: Coefficients, separated into real and imaginary parts for a circuit with six qubits and different single qubit Pauli-encodings. The individual frequency components are colour-coded.

While there is no obvious difference in $\mu_c(\omega)$ throughout the different encoding strategies, things change when we look at the real and imaginary parts of the individual coefficients $c_\omega(\theta)$ separately, as depicted in Fig. 3. Here, C15 exhibits an imaginary part in the coefficients, only for the case of an R_Y encoding, missing a degree of freedom in the imaginary part for the R_X and R_Z encoding. Therefore, in subsequent experiments on one-dimensional inputs, we use the R_Y encoding for the results presented in the main text.

5.1.2 Impact of Noise on the Real- and Imaginary Parts

Fig. 3 suggests that coefficients are evenly distributed in the real and imaginary part, with no clear correlation (apart from C15 with R_X and R_Y encoding). To corroborate this observation, we next compute the elements of the coefficients covariance matrix of Eq. 8. We calculate coefficients for noiseless samples and for samples subjected to each noise type at 3% probability. Since all circuit sizes yield consistent results, we focus on six-qubit circuits and only on frequencies $\omega \in \{0, 1, \max\}$, where $\omega = \max$ refers to the highest occurring frequency in the noiseless case.

The results shown in Fig. 4, align with those in Fig. 3, demonstrating that the variances of both the real and imaginary parts of the coefficients are approximately equally high across configurations. However, the covariance between these components is significantly smaller or even zero. This suggests that the real and imaginary parts of a coefficient are not strongly correlated, allowing us to consider only the absolute value of the coefficients in subsequent experiments. While additional variance experiments are detailed in Sec. 5.1.3, Fig. 4 highlights that decoherent gate errors significantly reduce the variance of the coefficients, regardless of circuit structure. In contrast, other noise types appear to have no substantial impact on variance.

As shown in Fig. 3 coefficients are centred around zero for all configurations in a noiseless setting. Generally, this also applies to noisy scenarios with the exception of the AD channel, which shifts the coefficients of $\omega = 0$ towards positive values, as shown in Fig. 5. This indicates that the offset due to the Z -basis measurement is skewed when applying an AD channel.

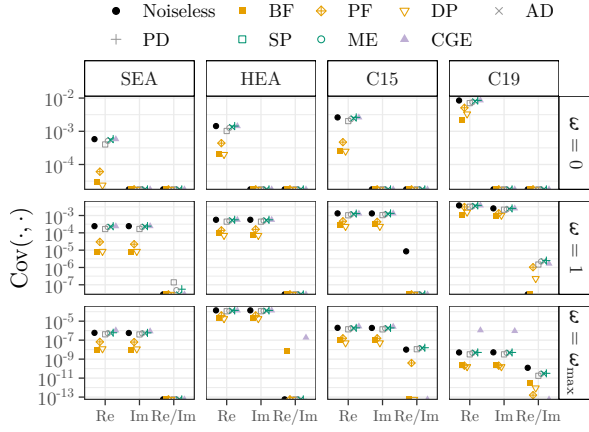


Fig. 4: Elements of the coefficients covariance matrix in noiseless and noisy (3% of each noise type) settings for six qubits. Points on the bottom border are zero. $\omega = \omega_{\max}$ denotes the maximum frequency in the noiseless spectrum, that is $\omega = 4$ for the HEA and $\omega = 6$ for the remaining ansätze.

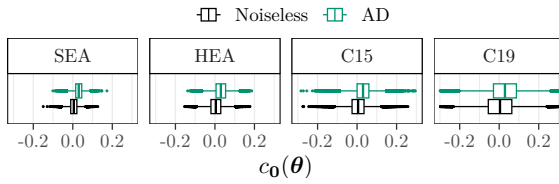


Fig. 5: Real coefficient for $\omega = 0$ in a noiseless setting vs. applying 3% of AD for circuits with six qubits.

5.1.3 Impact of Noise on the Absolute Value

Next, we examine how noise affects the coefficient mean and standard deviation across different ansätze for one-dimensional inputs ($D = 1$) using the R_Y encoding. The results are shown in Fig. 6 for each type of noise, evaluated on six-qubit circuits. Additionally, we computed the coefficient values with an identical setup for the R_X encoding, which we summarise in Appendix C. These results demonstrate that there are many similarities, yet also fundamental differences for the coefficients using R_X and R_Y encodings, highlighting the importance of the encoding strategy selection. Furthermore, we conducted the same set of experiments for circuits with three to five qubits, which provided similar outcomes throughout encodings, but accordingly with fewer frequencies. A selection of the results is provided in Appendix D while the supplementary material in Ref. [18] provides full result sets.

Fig. 6a shows that decoherent gate errors generally lead to an exponential decay of coefficient mean for all considered ansätze, which is in line with the findings for uniform coefficient contraction under noise made in Ref. [32]. For the SPAM and damping noise the contraction is weaker, or not observable at all for $\mu_c(\omega)$.

The relative standard deviation in Fig. 6b remains constant for all noise types except CGEs for all frequencies and ansätze, while frequencies $\omega \neq 0$ generally lead to higher $\sigma_c(\omega)$. Since $\sigma_c(\omega)$ is proportional to $\mu_c(\omega)$, this implies that the absolute variance also decreases exponentially with increasing noise level under the influence of decoherent gate errors. Interestingly, in case of the SEA with BF applied, $\sigma_c(\omega)$ for $\omega = 0$ increases with increasing noise level, which likely is also a consequence of it

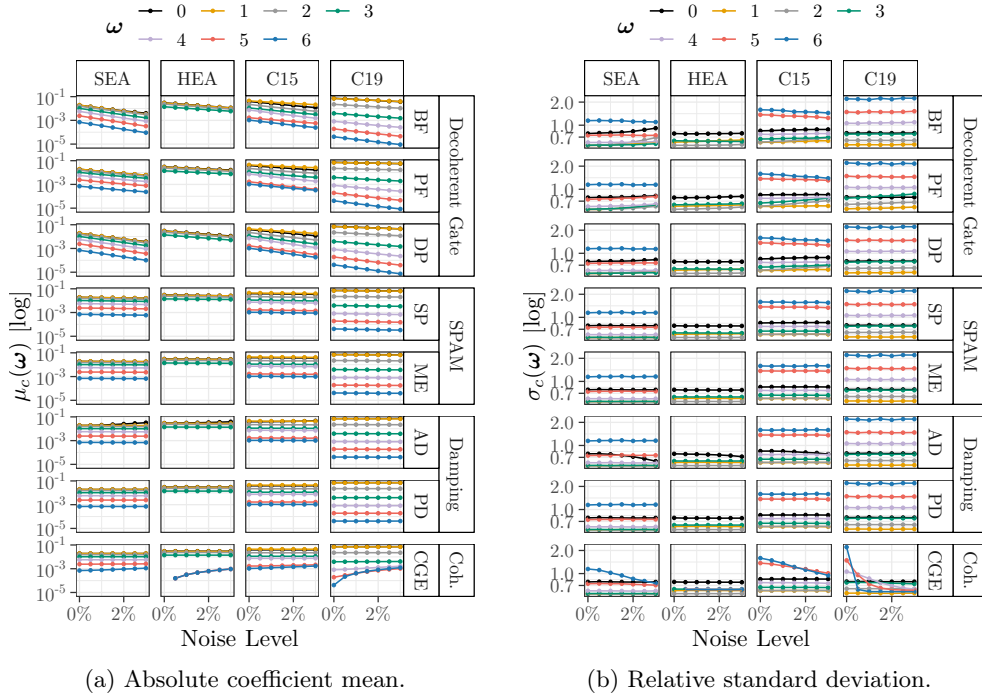


Fig. 6: Absolute coefficient mean $\mu_c(\omega)$ and the corresponding relative standard deviation $\sigma_c(\omega)$ over noise levels for various types of noise and six-qubit circuits that use R_Y encoding.

being the relative standard deviation to the corresponding $\mu_c(\omega)$, which in comparison only slightly increases.

Generally SPAM and damping noise have minimal effect on both, coefficient mean and relative standard deviation. While the influence of CGE is similar on the lower frequency coefficients, we observe that for the two highest frequencies $\mu_c(\omega)$ increases, while $\sigma_c(\omega)$ rapidly decreases with increasing noise level, especially for the HEA and C19 (see [Sec. 5.1.4](#) for a detailed discussion).

Given substantial similarities across different noise types and frequencies, we present a curated selection of combined results for $D \in \{1, 2\}$ across all qubit numbers and ansätze in [Appendix D](#). The comprehensive results are available in our supplementary material in [Ref. \[18\]](#).

5.1.4 Effect of Coherent Noise

As shown in [Eq. 16](#), the CGE on the trainable gates ϵ_θ only affects the magnitude of each c_ω . However, the CGE on the encoding gates ϵ_x alters the frequency component as the differences in the eigenvalues Λ do not fully overlap, yielding a higher number of unique frequency components while the redundancies $R(\omega)$ are reduced. This effect becomes visible when increasing the frequency resolution Δf by increasing the window

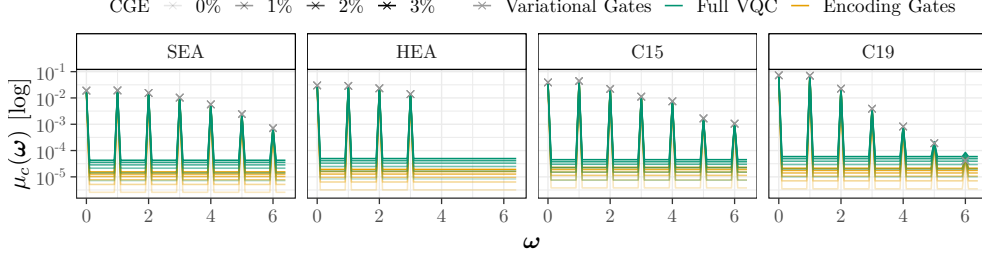


Fig. 7: Absolute coefficient mean $\mu_c(\omega)$ for 6 qubits and R_Y encodings over frequencies for different noise levels $[0, \dots, 3]\%$ and different ansätze. Noise is applied either on the encoding gates (ϵ_x), the trainable gates (ϵ_θ) or on the full VQC (ϵ_x and ϵ_θ).

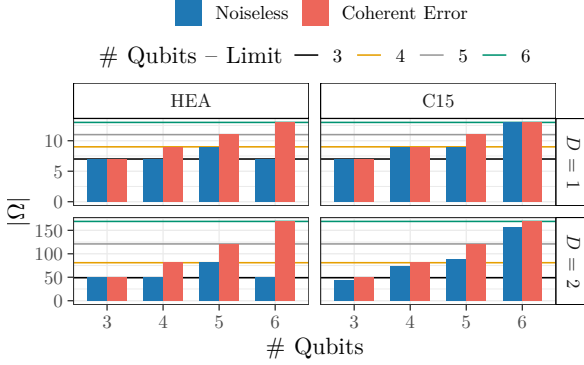


Fig. 8: Number of frequencies in the spectrum with and without applying a CGE for D -dimensional inputs and only those circuits where the maximum possible number of frequencies is not achieved in the noiseless case.

length of the FFT. In Fig. 7 we can observe the effect of CGE on the spectrum of different 6-qubit ansätze with a window length being 10 times the required minimum according to the sampling theorem. Here, we differentiate between applying the CGE on (i) the encoding gates (ϵ_x); (ii) the trainable gates (ϵ_θ) or (iii) on the full VQC (ϵ_x and ϵ_θ).

While the absolute mean value of the integer-valued frequencies $[0, 1, \dots, 6]$ remains mostly constant for all noise levels, we can observe that the magnitude of non-integer frequencies in between increases. Notably, only the input CGE causes this effect, which is amplified in combination with but not solely caused by the CGE acting on the variational gates.

This observation is in line with an approach introduced in [26], where trainable parameters were added to the input encoding, increasing the number of differences between the summation of eigenvalues Λ and therefore also increasing the number of frequencies. While the referenced work used this effect to adjust the frequencies individually, the CGE can be viewed similarly while changing the frequencies randomly and therefore, on average, causing a uniform increase of non-integer frequencies.

The impact of CGE on the frequency spectrum also becomes evident when comparing the number of frequencies with and without coherent noise, as illustrated in Fig. 8. In particular, coefficients that are computed to be zero in the FFT under noiseless

conditions become non-zero in the presence of CGE. Therefore, the increase in higher-frequency coefficients, as shown in Fig. 6 and Fig. 16, is a result of these spectral shifts, which lead to a more uniform distribution of coefficients. This also applies to the HEA and C15 that do not result in a complete frequency spectrum, likely due to limitations in the ansatz structure and input encodings, which restrict the set of achievable eigenvalue differences (*cf.* Eq. 5). While CGEs seems to solve this problem, we would like to point out that although the circuit under CGE contains more frequencies, this does not mean that they are individually tunable (*cf.* Sec. 5.1.5). This can result in a spectrum that is significantly correlated, a phenomenon that is undesirable in the context of most learning problems.

5.1.5 Coefficients during Training

To measure the training performance under noisy conditions, we utilise the setup and metrics described in Sec. 4.5 for the synthetic regression tasks. Training is performed using the Adam optimiser [50] with a learning rate of 0.01 over 1000 training steps. The MSE over the course of the training, shown in Fig. 9, demonstrates that all configurations converge although the performance between circuits differs: The HEA and C19 fail to converge towards the optimum for all tested problem- and model initialisations, while C15 and the SEA achieve lower MSEs, with the SEA performing best on this metric in a noiseless setting.

The detrimental effect of noise is particularly evident for the decoherent gate errors across all ansätze, resulting in higher MSEs, while the other noise types appear to be less significant. It is evident that the distance in the MSE between the noiseless case and the QFMs that suffer from decoherent gate errors increases as the MSE in the noiseless case decreases. This results in a similar MSE of ≈ 0.02 , when decoherent gate errors are applied throughout all ansätze. Other types of noise result in minor deviations from the noiseless case.

The trend of the MSE can be attributed to how closely the coefficients of the QFM approach the target coefficients. In Appendix E, we show the exact coefficients for one of the problem instances considered over the course of the training where we note, that noise causes a shift of the frequencies as they are affected differently by noise. Fig. 10 shows the mean absolute coefficient difference Δc_ω between the QFMs and the

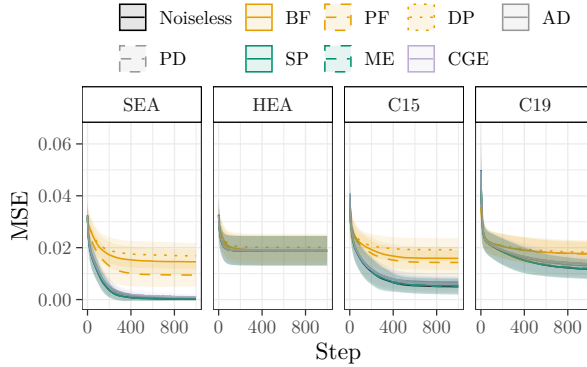


Fig. 9: The MSE over ten parameter initialisation seeds and ten randomly generated problem instances during training. Lines represent the mean, and shaded areas show the standard deviation over the 10×10 configurations.

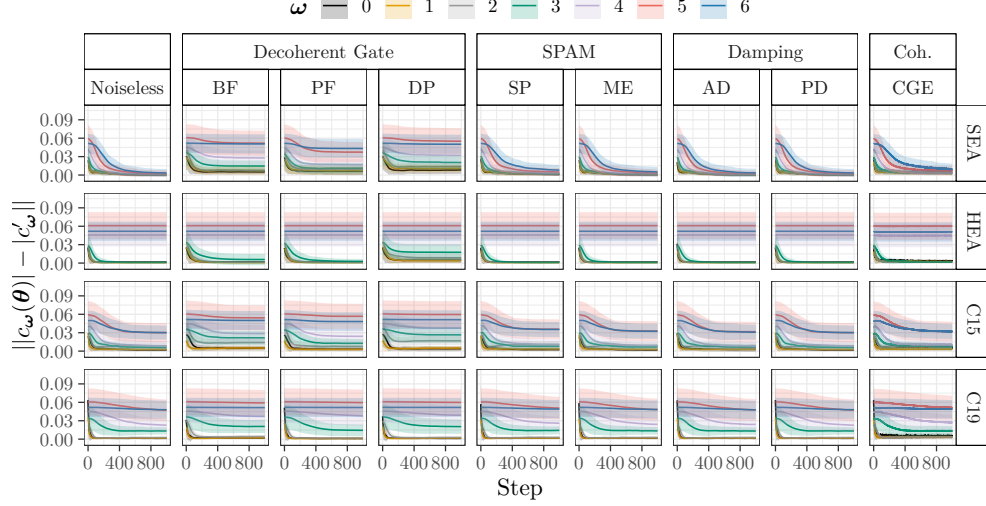


Fig. 10: The difference Δc_ω between the target and learned coefficients $c_\omega(\theta)$ and c'_ω respectively of the QFMs averaged over ten model parameter initialisation seeds and ten randomly generated problem instances during training. Lines represent the mean, and shaded areas show the standard deviation over the 10×10 configurations.

objective Fourier series over all problem seeds. Considering the noiseless case, while Δc_ω for lower ω converges faster to zero across all ansätze, the ansätze, which achieve a lower MSE, also reach a lower Δc_ω for all ω . Nevertheless, higher frequencies are also crucial for the performance, as indicated by the HEA, which achieves a low Δc_ω for $\omega \in [0 \dots 4]$, however with a constant high Δc_ω for the remaining two highest frequencies, as it does not achieve a full spectrum, resulting in an overall higher MSE. The large values of Δc_ω for higher frequencies across all ansätze, except the SEA, indicate that the Fourier coefficients are not fully tunable for every ansatz, even in a noiseless setting.

When adding noise, there are only minor differences between the performance for the noiseless setting, and SPAM, Damping and CGEs. For the CGE, we notice small variations in the coefficient values between training steps, evident in Fig. 18, while the remaining noise types lead to more smooth curves. While these variations also occur for the higher frequencies of the HEA that do not appear in the noiseless spectrum, they do not contribute to optimising the coefficients towards the target. Decoherent gate errors lead to more drastic performance issues, visible in both, the MSE and Δc_ω .

5.2 Expressibility

The expressibility is calculated as introduced in Sec. 4.3 and quantified using the KL-divergence to the Haar distribution (*cf.* Eq. 11). The experimental results for all ansätze and qubit counts are presented in Fig. 11.

It can be observed that all decoherent gate errors lead to an increase in the KL-divergence, thereby reducing expressibility. This effect is also consistent across SPAM

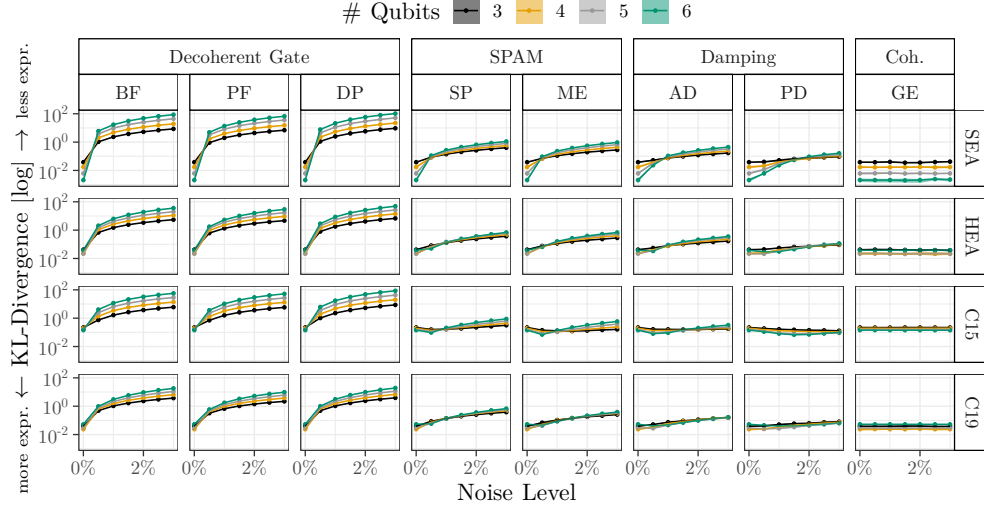


Fig. 11: Average Expressibility (*i.e.*, inverse of the KL divergence) over randomly sampled parameters under the influence of increasing noise levels. The points represent the mean and the shaded areas around it refer to the minimum and maximum across all five seeds, which is negligible in most cases as we chose a sufficiently large number of samples.

and damping noise, though the impact is less pronounced. The observed decrease in expressibility aligns with the reduced variance of coefficients noted in [Sec. 5.1](#) and the connection to frequency redundancies discussed in [Ref. \[17\]](#).

Coherent noise, on the other hand, has no measurable effect on expressibility. This is not surprising since we omit the encoding gates for the CGE to act upon, leaving the system in a coherent state. Effects of shifts in the trainable parameters appear to cancel out over the parameter samples.

In a noiseless environment, the SEA achieves a higher expressibility compared to the other ansätze. However, when coherent noise is applied, the expressibility is quickly equalised. While the effect is consistent across all circuits tested, it can be observed that it becomes more pronounced as the number of qubits increases. This suggests that all forms of decoherent noise degrade the quantum nature of the circuits.

5.3 Entanglement

We first utilise the MW measure as introduced in [Sec. 4.4.1](#) to compute the entanglement of the circuits without noise. Subsequently, the EF (*cf.* [Sec. 4.4.2](#)) is used to measure the entangling capability for the mixed states with increasing noise levels.

The results for the noiseless circuits are shown in [Fig. 12](#). Here, it can be observed, that an increasing qubit count leads to a higher overall value for the entangling capability, while the specific values for the SEA, HEA and C15 are similar. C19, which has controlled rotation gates as entangling gates, results in a lower entangling capability for more than 4 qubits.



Fig. 12: Average MW entanglement over randomly sampled parameters for the considered ansätze in a noiseless environment. The bars represent the mean values with error bars from the minimum to maximum highlighted in red across the five seeds.

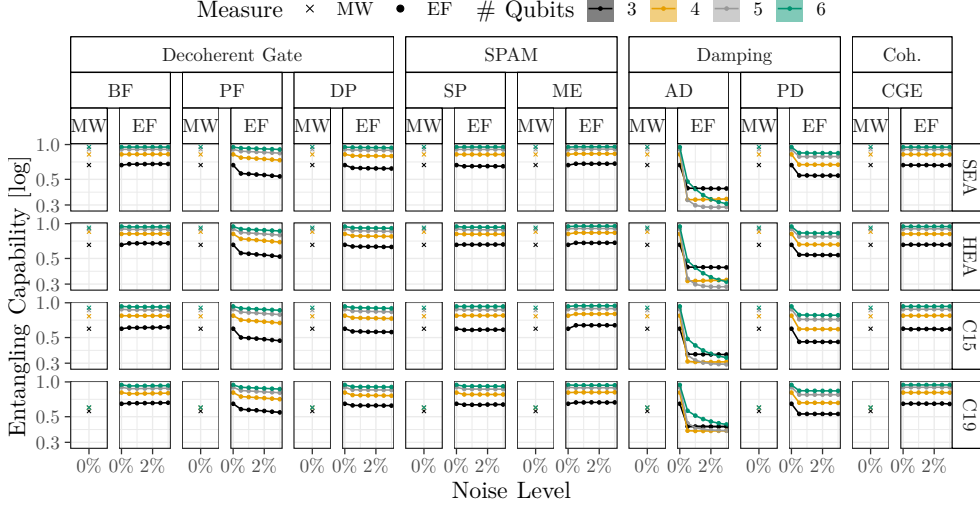


Fig. 13: Average MW entangling capability and EF over randomly sampled parameters under the influence of increasing noise levels. Points represent the mean of five seeds and lines are a linear interpolation to guide the eye. Shaded areas represent the maximum/minimum entanglement across the five seeds. As the difference between seeds is small, it is not visible in the plot.

The results for the mixed states using the EF measure are shown in Fig. 13. Notably, EF results in different values than the MW measure, also for noiseless circuits, especially observable for C19. This observation may be a consequence of the properties of the EF measure: At zero noise-probability for the EF, an eigendecomposition results in only one pure state, over which the MW measure is computed. However, this decomposition is not unique (*cf.* Sec. 4.4), and only one of many possibilities, potentially resulting in different values for the entangling capability. Therefore, we interpret the values in Fig. 13 as upper bounds of the entangling capability, which drop below the MW entangling capability with applied noise in most cases.

Within the decoherent gate-, and SPAM-error groups a similar behaviour can be observed for each ansatz. The values here slightly decrease or remain constant for both decoherent gate- and SPAM-errors, with some exceptions for three and four qubits with BF- and Measurement (ME)-noise, where the entanglement slightly increases from 0%-level noise to 0.5%. Similarly, for PD, the entangling capability also decreases with

noise for all ansätze. For AD-noise entangling capability decreases more drastically for all circuits, and number of qubits with increasing noise level. CGEs do not have an observable impact on the EF. Our results suggest that noise is overall detrimental for the entangling capability, especially for AD-noise.

As we discuss in [Appendix F](#), the entangling capability of a QFM decreases over the course of a training for the considered learning task similarly across all ansätze, indicating no clear correlation between training performance and entangling capability. When adding noise, we measure a similar impact on the overall entangling capability during training as in [Fig. 13](#), which is slightly lower than in the noiseless case for most ansätze.

In the broad and systematic empirical analysis presented in this section, we find that the detailed effect of different types of noise leads to the main general findings for QML outlined as follows:

Result Summary

The influence of noise on properties of QFMs, particularly the Fourier spectrum, expressibility and entangling capability, can be predicted. The effects of each noise type are uniform throughout all tested circuits. More importantly, the structure of a VQC and the input encoding have a crucial impact on these properties, independently of the types of noise applied, also in a noiseless setting. This is a relevant finding for the mid- and long-term of prospective FTQCs.

6 Discussion

This work explores the effects that noise has on the coefficients, expressibility and entangling capability of QFMs. In particular, (i) our work confirms the anti-correlation between coefficients and noise observed in Ref. [32], and demonstrates that the decrease of coefficient values is even exponential for decoherent gate errors. Furthermore, we observed that all of the tested ansätze are less susceptible to damping or SPAM errors. (ii) Using the example of learning a random Fourier series, we demonstrate that decoherent errors have a detrimental influence on the capabilities of QFMs to assume certain Fourier coefficients. This directly impacts the kinds of functions that can be approximated by a QFM. (iii) Our experiments show that the expressibility, measured by the distance to the Haar distribution and the coefficients variance also vanishes (potentially exponentially) with the absolute coefficients value with increasing noise level. (iv) The entangling capability, which is a general indicator of the *quantumness* of a VQC, also decreases when subject to noise.

The ansätze investigated in this work show similar patterns for all of the above. Although our findings are based on a limited circuit size and number of input dimensions (up to six qubits and two features), they demonstrate consistent patterns as the number of qubits increases from three to six. This suggest potential for broader applicability in larger circuits. Additionally, similar results across one-dimensional and two-dimensional inputs offer a good indication for generalisation despite the exponential scaling of frequencies in the spectrum.

Although the observation of the detrimental influence of noise may not be a novel finding [40], [41], to the best of our knowledge our work is the first to show this effect in a holistic, systematic manner on the fundamental computational aspects of VQCs.

As each learning problem in QML has different requirements on the ansatz that is employed, a general *all-fits-one* ansatz is unlikely to exist. Nevertheless, based on our experiments, we can make some statements on the measures of quality for the tested ansätze: (i) The **SEA** is, as the name suggests the ansatz with the highest entangling capability, although the difference to HEA and C15 in that regard is small. Among the tested ansätze, the SEA also utilises most trainable parameters per qubit, resulting in a comparatively higher expressibility in a noiseless setting (*cf.* Fig. 11), which is beneficial for the considered learning tasks. However, the expressibility quickly vanishes when applying noise. The corresponding Fourier spectrum is full (at least for up to two input dimensions), and comparatively uniformly distributed (*cf.* Fig. 2a). (ii) In the **HEA**, the amount by which the depth of the entangling sequence grows, is constant, compared to the other ansätze, where it scales linearly with the number of qubits. This desirable property comes with the cost of lacking the full Fourier spectrum on average, even at one input dimension, which directly impacts the kinds of functions that can be learned by the HEA. (iii) **C15** demonstrates similar behaviour to the SEA and HEA in the scaling of entanglement and expressibility under noise. It further has a unique characteristic, where the imaginary part in the coefficients appears exclusively for a R_Y encoding. As the HEA, C15 does not consistently achieve a full spectrum. (iv) **C19** achieves a full spectrum for up to two input dimensions, as the SEA. However, the coefficients vanish quickly with increasing frequency, even in the noiseless case (*cf.* Fig. 2a), indicating that not all parts of the spectrum may be effectively utilised. Despite having the lowest MW entangling capability, C19 features tunable parameters in the controlled- R_X gates, which could potentially allow for adjusting this property.

7 Conclusion and Outlook

Given the potential for conducting an extensive array of numerical experiments on this subject, the results presented are expected to offer researchers valuable insights into the behaviour of QFM under noisy conditions. However, an analytical correlation between these factors and potential generalisations to larger circuit sizes (both in the number of qubits and the circuit depth) remains to be investigated in future research. Also how different ansätze or encoding strategies, such as the ones presented in Ref. [22] fit into this pattern remains to be explored.

While we acknowledge that noise will lose its importance once FTQC is established, the road to FTQC remains stony, and as not every hardware architecture is suitable for error correction, the effects of noise on QFM are still relevant as the proposed results are hardware agnostic. In subsequent studies, we intend to perform analytical derivations to further extend the results presented and allow more definitive conclusions to be drawn, also for arbitrary ansätze.

Acknowledgements MS, LC, EK and AS acknowledge support by the state of Baden-Württemberg through bwHPC. MF and WM acknowledge support by the German Federal Ministry of Research, Technology and Space (BMFTR), funding program

“Quantum Technologies—From Basic Research to Market”, grant number 13N16092. WM acknowledges support by the High-Tech Agenda of the Free State of Bavaria.

References

- [1] F. Greiwe, T. Krüger, and W. Mauerer, “Effects of imperfections on quantum algorithms: A software engineering perspective,” in *2023 IEEE International Conference on Quantum Software (QSW)*, 2023, pp. 31–42. DOI: [10.1109/QSW59989.2023.00014](https://doi.org/10.1109/QSW59989.2023.00014).
- [2] M. Schuld and N. Killoran, “Is quantum advantage the right goal for quantum machine learning?” *PRX Quantum*, vol. 3, p. 030 101, 3 Jul. 2022. DOI: [10.1103/PRXQuantum.3.030101](https://doi.org/10.1103/PRXQuantum.3.030101).
- [3] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers* (Quantum Science and Technology), en. Cham: Springer International Publishing, 2018. DOI: [10.1007/978-3-319-96424-9](https://doi.org/10.1007/978-3-319-96424-9).
- [4] Z. Zimborás, B. Koczor, Z. Holmes, *et al.*, *Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don’t*, arXiv:2501.05694 [quant-ph], Jan. 2025. DOI: [10.48550/arXiv.2501.05694](https://doi.org/10.48550/arXiv.2501.05694).
- [5] M. Ragone, B. N. Bakalov, F. Sauvage, *et al.*, “A Lie Algebraic Theory of Barren Plateaus for Deep Parameterized Quantum Circuits,” *Nature Communications*, vol. 15, no. 1, p. 7172, Aug. 2024, arXiv:2309.09342 [quant-ph]. DOI: [10.1038/s41467-024-49909-3](https://doi.org/10.1038/s41467-024-49909-3).
- [6] D. Heimann, G. Schönhoff, E. Mounzer, *et al.*, *Learning capability of parametrized quantum circuits*, arXiv:2209.10345 [quant-ph], Mar. 2024. DOI: [10.48550/arXiv.2209.10345](https://doi.org/10.48550/arXiv.2209.10345).
- [7] K. Georgopoulos, C. Emary, and P. Zuliani, “Modeling and simulating the noisy behavior of near-term quantum computers,” *Physical Review A*, vol. 104, no. 6, p. 062 432, Dec. 2021, Publisher: American Physical Society (APS). DOI: [10.1103/physreva.104.062432](https://doi.org/10.1103/physreva.104.062432).
- [8] K. Wintersperger, H. Safi, and W. Mauerer, “Qpu-system co-design for quantum hpc accelerators,” in *Architecture of Computing Systems*, M. Schulz, C. Trinitis, N. Papadopoulou, *et al.*, Eds., Cham: Springer International Publishing, 2022, pp. 100–114.
- [9] S. Thelen, H. Safi, and W. Mauerer, “Approximating under the influence of quantum noise and compute power,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 02, 2024, pp. 274–279. DOI: [10.1109/QCE60285.2024.10291](https://doi.org/10.1109/QCE60285.2024.10291).
- [10] M. Franz, L. Wolf, M. Periyasamy, *et al.*, “Uncovering instabilities in variational-quantum deep Q-networks,” *Journal of the Franklin Institute*, vol. 360, no. 17, pp. 13 822–13 844, 2023. DOI: <https://doi.org/10.1016/j.jfranklin.2022.08.021>.
- [11] M. Franz, T. Winker, S. Groppe, *et al.*, “Hype or heuristic? Quantum reinforcement learning for join order optimisation,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 01, 2024, pp. 409–420. DOI: [10.1109/QCE60285.2024.00055](https://doi.org/10.1109/QCE60285.2024.00055).

- [12] M. Periyasamy, A. Plinge, C. Mutschler, *et al.*, “Guided-spsa: Simultaneous perturbation stochastic approximation assisted by the parameter shift rule,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 01, 2024, pp. 1504–1515. DOI: [10.1109/QCE60285.2024.00177](https://doi.org/10.1109/QCE60285.2024.00177).
- [13] A. Bayerstadler, G. Becquin, J. Binder, *et al.*, “Industry quantum computing applications,” *EPJ Quantum Technology*, vol. 8, no. 1, Nov. 2021. DOI: [10.1140/epjqt/s40507-021-00114-x](https://doi.org/10.1140/epjqt/s40507-021-00114-x).
- [14] C. Carbonelli, M. Felderer, M. Jung, *et al.*, *Challenges for quantum software engineering: An industrial application scenario perspective*, I. Exman, R. Pérez-Castillo, M. Piattini, *et al.*, Eds., Cham, 2024. DOI: [10.1007/978-3-031-64136-7_12](https://doi.org/10.1007/978-3-031-64136-7_12).
- [15] M. Gogeissl, H. Safi, and W. Mauerer, “Quantum data encoding patterns and their consequences,” in *Proceedings of the 1st Workshop on Quantum Computing and Quantum-Inspired Technology for Data-Intensive Systems and Applications*, ser. Q-Data ’24, Santiago, AA, Chile: Association for Computing Machinery, 2024, pp. 27–37. DOI: [10.1145/3665225.3665446](https://doi.org/10.1145/3665225.3665446).
- [16] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, “Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms,” in *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019. DOI: [10.1002/qute.201900070](https://doi.org/10.1002/qute.201900070).
- [17] H. Mhiri, L. Monbroussou, M. Herrero-Gonzalez, *et al.*, *Constrained and Vanishing Expressivity of Quantum Fourier Models*, arXiv:2403.09417, Mar. 2024. DOI: [10.48550/arXiv.2403.09417](https://doi.org/10.48550/arXiv.2403.09417).
- [18] M. Franz and M. Strobl, *Reproduction Package for "Out of Tune: Demystifying Noise-Effects on Quantum Fourier Models"*, Apr. 2025. DOI: [10.5281/zenodo.15211317](https://doi.org/10.5281/zenodo.15211317).
- [19] M. Strobl, M. Franz, E. Kuehn, *et al.*, *Qml essentials – a framework for working with quantum fourier models*, 2025. DOI: [10.48550/arXiv.2506.06695](https://doi.org/10.48550/arXiv.2506.06695).
- [20] M. Schuld, R. Sweke, and J. J. Meyer, “The effect of data encoding on the expressive power of variational quantum machine learning models,” *Physical Review A*, vol. 103, no. 3, p. 032430, Mar. 2021, arXiv: 2008.08605. DOI: [10.1103/PhysRevA.103.032430](https://doi.org/10.1103/PhysRevA.103.032430).
- [21] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, *et al.*, “Data re-uploading for a universal quantum classifier,” *Quantum*, vol. 4, p. 226, Feb. 2020, arXiv: 1907.02085. DOI: [10.22331/q-2020-02-06-226](https://doi.org/10.22331/q-2020-02-06-226).
- [22] B. Casas and A. Cervera-Lierta, “Multidimensional Fourier series with quantum circuits,” *Physical Review A*, vol. 107, no. 6, p. 062612, Jun. 2023, arXiv: 2302.03389. DOI: [10.1103/PhysRevA.107.062612](https://doi.org/10.1103/PhysRevA.107.062612).
- [23] M. C. Caro, E. Gil-Fuster, J. J. Meyer, *et al.*, “Encoding-dependent generalization bounds for parametrized quantum circuits,” *Quantum*, vol. 5, p. 582, Nov. 2021, arXiv: 2106.03880. DOI: [10.22331/q-2021-11-17-582](https://doi.org/10.22331/q-2021-11-17-582).
- [24] N. A. Nemkov, E. O. Kiktenko, and A. K. Fedorov, “Fourier expansion in variational quantum algorithms,” *Physical Review A*, vol. 108, no. 3, p. 032406, Sep. 2023, arXiv:2304.03787 [quant-ph]. DOI: [10.1103/PhysRevA.108.032406](https://doi.org/10.1103/PhysRevA.108.032406).

- [25] M. Wiedmann, M. Periyasamy, and D. D. Scherer, *Fourier Analysis of Variational Quantum Circuits for Supervised Learning*, arXiv:2411.03450, Nov. 2024. DOI: [10.48550/arXiv.2411.03450](https://doi.org/10.48550/arXiv.2411.03450).
- [26] B. Jaderberg, A. A. Gentile, Y. A. Berrada, *et al.*, “Let Quantum Neural Networks Choose Their Own Frequencies,” *Physical Review A*, vol. 109, no. 4, p. 042 421, Apr. 2024, arXiv:2309.03279 [quant-ph]. DOI: [10.1103/PhysRevA.109.042421](https://doi.org/10.1103/PhysRevA.109.042421).
- [27] M. Cerezo, A. Sone, T. Volkoff, *et al.*, “Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits,” *Nature Communications*, vol. 12, no. 1, p. 1791, Mar. 2021, arXiv: 2001.00550. DOI: [10.1038/s41467-021-21728-w](https://doi.org/10.1038/s41467-021-21728-w).
- [28] M. Larocca, S. Thanasilp, S. Wang, *et al.*, “Barren plateaus in variational quantum computing,” en, *Nature Reviews Physics*, pp. 1–16, Mar. 2025, Publisher: Nature Publishing Group. DOI: [10.1038/s42254-025-00813-9](https://doi.org/10.1038/s42254-025-00813-9).
- [29] S. Wang, E. Fontana, M. Cerezo, *et al.*, “Noise-Induced Barren Plateaus in Variational Quantum Algorithms,” *Nature Communications*, vol. 12, no. 1, p. 6961, Nov. 2021, arXiv: 2007.14384. DOI: [10.1038/s41467-021-27045-6](https://doi.org/10.1038/s41467-021-27045-6).
- [30] M. Cerezo, M. Larocca, D. García-Martín, *et al.*, “Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing,” arXiv, Tech. Rep., Dec. 2023, arXiv: 2312.09121. DOI: [10.48550/arXiv.2312.09121](https://doi.org/10.48550/arXiv.2312.09121).
- [31] F. J. Schreiber, J. Eisert, and J. J. Meyer, “Classical surrogates for quantum learning models,” arXiv, Tech. Rep., Jun. 2022, arXiv: 2206.11740. DOI: [10.48550/arXiv.2206.11740](https://doi.org/10.48550/arXiv.2206.11740).
- [32] E. Fontana, M. S. Rudolph, R. Duncan, *et al.*, “Classical simulations of noisy variational quantum circuits,” arXiv, Tech. Rep., Jun. 2023, arXiv: 2306.05400. DOI: [10.48550/arXiv.2306.05400](https://doi.org/10.48550/arXiv.2306.05400).
- [33] J. Landman, S. Thabet, C. Dalyac, *et al.*, “Classically Approximating Variational Quantum Machine Learning with Random Fourier Features,” arXiv, Tech. Rep., Oct. 2022, arXiv: 2210.13200. DOI: [10.48550/arXiv.2210.13200](https://doi.org/10.48550/arXiv.2210.13200).
- [34] R. Sweke, E. Recio-Armengol, S. Jerbi, *et al.*, “Potential and limitations of random Fourier features for dequantizing quantum machine learning,” *Quantum*, vol. 9, p. 1640, Feb. 2025, arXiv:2309.11647 [quant-ph]. DOI: [10.22331/q-2025-02-20-1640](https://doi.org/10.22331/q-2025-02-20-1640).
- [35] K. Sharma, M. Cerezo, Z. Holmes, *et al.*, “Reformulation of the No-Free-Lunch Theorem for Entangled Datasets,” *Physical Review Letters*, vol. 128, no. 7, p. 070 501, Feb. 2022, Publisher: American Physical Society. DOI: [10.1103/PhysRevLett.128.070501](https://doi.org/10.1103/PhysRevLett.128.070501).
- [36] X. Wang, Y. Du, Z. Tu, *et al.*, “Transition role of entangled data in quantum machine learning,” en, *Nature Communications*, vol. 15, no. 1, p. 3716, May 2024, Publisher: Nature Publishing Group. DOI: [10.1038/s41467-024-47983-1](https://doi.org/10.1038/s41467-024-47983-1).
- [37] J. Bowles, S. Ahmed, and M. Schuld, “Better than classical? The subtle art of benchmarking quantum machine learning models,” arXiv, Tech. Rep., Mar. 2024, arXiv: 2403.07059. DOI: [10.48550/arXiv.2403.07059](https://doi.org/10.48550/arXiv.2403.07059).
- [38] T. Rohe, D. Schuman, J. Nüßlein, *et al.*, *The Questionable Influence of Entanglement in Quantum Optimisation Algorithms*, arXiv:2407.17204 [quant-ph], Jul. 2024. DOI: [10.48550/arXiv.2407.17204](https://doi.org/10.48550/arXiv.2407.17204).

- [39] A. Joch, G. S. Uhrig, and B. Fauseweh, *Entanglement-informed Construction of Variational Quantum Circuits*, arXiv:2501.17533 [quant-ph], Jan. 2025. DOI: [10.48550/arXiv.2501.17533](https://doi.org/10.48550/arXiv.2501.17533).
- [40] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, “Random quantum circuits transform local noise into global white noise,” *Communications in Mathematical Physics*, vol. 405, no. 3, 2024. DOI: [10.1007/s00220-024-04958-z](https://doi.org/10.1007/s00220-024-04958-z).
- [41] E. Fontana, I. Rungger, R. Duncan, *et al.*, “Spectral analysis for noise diagnostics and filter-based digital error mitigation,” arXiv, Tech. Rep., Nov. 2022, arXiv: 2206.08811. DOI: [10.48550/arXiv.2206.08811](https://doi.org/10.48550/arXiv.2206.08811).
- [42] S. R. Maschek, J. Schwittalla, M. Franz, *et al.*, “Make some noise! measuring noise model quality in real-world quantum software,” en, in *Proceedings of the IEEE International Conference on Quantum Software (QSW)*, Jul. 2025.
- [43] S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951, Publisher: Institute of Mathematical Statistics. DOI: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- [44] M. B. Plenio and S. Virmani, *An introduction to entanglement measures*, arXiv:quant-ph/0504163, Jun. 2006. DOI: [10.48550/arXiv.quant-ph/0504163](https://doi.org/10.48550/arXiv.quant-ph/0504163).
- [45] D. A. Meyer and N. R. Wallach, “Global entanglement in multiparticle systems,” *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4273–4278, Sep. 2002. DOI: [10.1063/1.1497700](https://doi.org/10.1063/1.1497700).
- [46] G. K. Brennen, “An observable measure of entanglement for pure states of multi-qubit systems,” arXiv, Tech. Rep., Nov. 2003. DOI: [10.48550/arXiv.quant-ph/0305094](https://doi.org/10.48550/arXiv.quant-ph/0305094).
- [47] W. K. Wootters, “Entanglement of Formation of an Arbitrary State of Two Qubits,” *Physical Review Letters*, vol. 80, no. 10, pp. 2245–2248, Mar. 1998, arXiv:quant-ph/9709029. DOI: [10.1103/PhysRevLett.80.2245](https://doi.org/10.1103/PhysRevLett.80.2245).
- [48] M. Schuld, A. Bocharov, K. Svore, *et al.*, “Circuit-centric quantum classifiers,” *Physical Review A*, vol. 101, no. 3, p. 032308, Mar. 2020, arXiv:1804.00633 [quant-ph]. DOI: [10.1103/PhysRevA.101.032308](https://doi.org/10.1103/PhysRevA.101.032308).
- [49] W. Mauerer and S. Scherzinger, “1-2-3 reproducibility for quantum software experiments,” in *2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2022, pp. 1247–1248. DOI: [10.1109/SANER53432.2022.00148](https://doi.org/10.1109/SANER53432.2022.00148).
- [50] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017.
- [51] C. Duffy and M. Jastrzebski, “Spectral bias in variational quantum machine learning,” *arXiv preprint arXiv:2506.22555*, 2025.

A Noise Models

In the following, we describe the utilised methods to model both, decoherent and coherent errors. For the decoherent errors we utilise the established Kraus formalism that models the (possibly non-unitary) effects of noise on a density operator ρ by $\rho \mapsto \sum_i K_i \rho K_i^\dagger$ with Kraus operators $\{K_i\}$.

A.1 Decoherent Gate Errors

We model decoherent gate errors by a combination of Pauli operations $\{I, X, Y, Z\}$ applied with a certain probability after each (ideal and noiseless) quantum gate. For simplicity, we only consider one-qubit channels, summarised in [Tab. 1](#).

A.2 Damping Errors

Damping errors correspond to relaxation and dephasing effects; the probabilities depend on the relaxation time T_1 and dephasing time T_2 of a given QPU implementation (we apply the damping channels at the end of a circuit.). As we do are not interested in vendor-specific details, we simply assume fixed probabilities.

Amplitude Damping (AD) describes the natural decay of the excited state $|1\rangle$ to the ground state $|0\rangle$ due to energy exchange with the environment. Similarly, Phase Damping (PD) describes the transition of a quantum system towards classical behaviour by losing coherence. Corresponding Kraus operators are listed in [Tab. 1](#).

A.3 State Preparation and Measurement Errors

State Preparation and Measurement (SPAM) on a real quantum device can be faulty, just like any other operation. Ideally, the initial state is $|0\rangle^{\otimes n}$ state, and a measurement returns either 0 or 1. However, there is a probability that State-Preparation (SP) is

Noise Type	Kraus Operators
Decoherent Gate	
Bit Flip (BF)	$K_0 = \sqrt{1 - p_{bf}}I, \quad K_1 = \sqrt{p_{bf}}X$
Phase Flip (PF)	$K_0 = \sqrt{1 - p_{pf}}I, \quad K_1 = \sqrt{p_{pf}}Z$
Depolarisation (DP)	$K_0 = \sqrt{1 - p_{dp}}I, \quad K_1 = \sqrt{p_{dp}/3}X,$ $K_2 = \sqrt{p_{dp}/3}Y, \quad K_3 = \sqrt{p_{dp}/3}Z$
Damping	
Amplitude Damping (AD)	$K_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1 - p_{ad}} \end{bmatrix}, \quad K_1 = \begin{bmatrix} 0 & \sqrt{p_{ad}} \\ 0 & 0 \end{bmatrix}$
Phase Damping (PD)	$K_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1 - p_{pd}} \end{bmatrix}, \quad K_1 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{p_{pd}} \end{bmatrix}$

Table 1: Kraus operators for the decoherent gate and damping errors with the probabilities p_{type} for each noise type.

faulty, or that measurements yield 1 instead of 0, and vice versa. We model SP errors with BF applied at the beginning of a circuit with probability p_{sp} . Measurement (ME) errors are applied at the very end of a circuit with probability p_{me} , again using BF errors.

A.4 Coherent Gate Error

Notable, since each operation is applied to the quantum system using an imperfect real QPU, the actual gate operation may deviate from the intended one.

The CGEs are modelled by adding $\epsilon_x \sim \mathcal{N}(0, p_{\text{cge}}^2)$ and $\epsilon_\theta \sim \mathcal{N}(0, p_{\text{cge}}^2)$ to any parameterised rotational input and trainable gate, respectively. The resulting Fourier series therefore becomes

$$f_{\text{cge}}(\mathbf{x} + \epsilon, \theta + \epsilon_\theta) = \sum_{\omega \in \Omega} c_\omega(\theta + \epsilon_\theta) e^{i\omega^T(\mathbf{x} + \epsilon_x)} = \sum_{\omega \in \Omega} c_\omega(\theta + \epsilon_\theta) e^{i(\omega + \epsilon_x)^T \mathbf{x}}. \quad (16)$$

If not stated explicitly, we consider the gate error on the encoding gates constant when calculating the coefficients of the model. The reason for this is, that we want to investigate the spectrum as a property of a model, which would not be possible if the gate error acts independently on the encoding gates, as it presumably causes a shift in the frequency components.

B Ansätze

We use the ansätze as depicted in [Fig. 14](#) in the experiments of our work.

C Coefficient Results for Pauli-RX Encoding

In [Fig. 15](#) we show the absolute coefficient mean $\mu_c(\omega)$ and the corresponding standard deviation $\sigma_c(\omega)$ for six-qubit circuits using R_X encodings, analogous to [Fig. 6](#). While there are many similarities to the R_Y encoding experiments, the main difference is that higher-frequency coefficients for the HEA and C19 increase when applying decoherent gate errors, while they decrease for the R_Y encoding. For the HEA this results in the higher frequencies becoming part of the spectrum, even in they are not present in the noiseless case, similar to when applying CGEs. This is likely to random distortions in the expectation value, leading to an increase of all frequencies in the spectral representation that become more apparent at the low-valued coefficients at higher frequencies. We also observe and discuss this effect in [Appendix D](#).

Additionally, C15 and C19 provide the highest standard deviation at 0% noise level, especially for higher frequencies, which then decays (more rapidly for C19) when decoherent or coherent gate errors are added. For the remaining noise types and ansätze $\sigma_c(\omega)$ remains constant or slightly decreases (with the same exceptions for the SEA as in [Fig. 6b](#)).

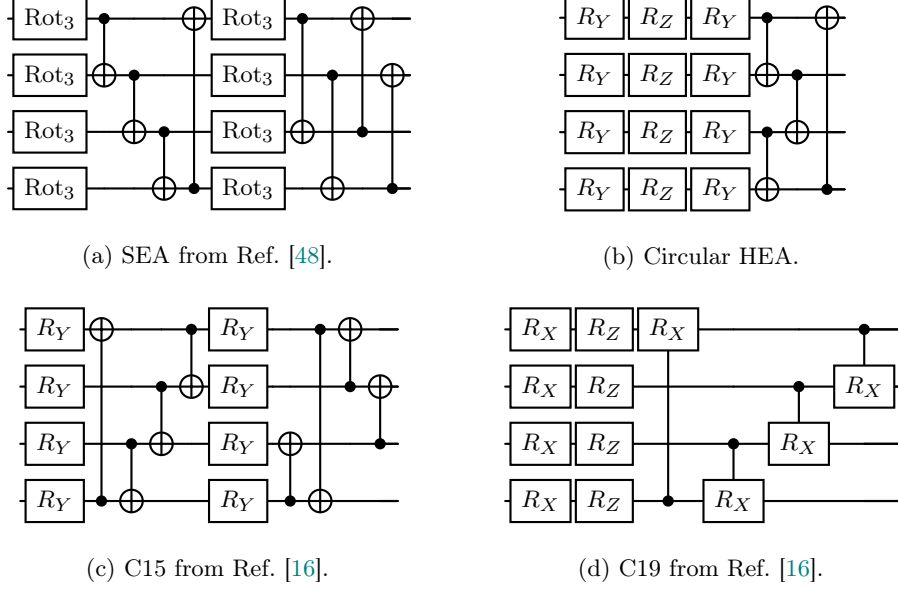


Fig. 14: Ansätze investigated in this work, exemplarily for 4 qubits. Rot_3 represents an arbitrary single qubit rotational gate, which takes three parameters. The single qubit Pauli-rotation gates $R_{\{X,Y,Z\}}$ are parametrised by one parameter each.

D Quantitative Coefficient Results

To assess the impact of noise on circuits with multiple input features, we conduct the same set of experiments detailed in [Sec. 5.1.3](#) and [Appendix C](#), also for $D = 2$, with the difference of only sampling over two instead of five random seeds for all qubit numbers to save computational resources. While different encoding strategies are presented in [Ref. \[22\]](#), this work focuses on a subsequent Pauli-encoding along different rotational axes (*i.e.*, X and Y in our experiments).

Our focus is on the effects of noise on two specific coefficient classes: the zero-frequency coefficient ($\omega = \mathbf{0}$) and the coefficient corresponding to the maximum frequency in all input dimensions in the respective spectrum, determined throughout each noise type ($\omega = \omega_{\max}$). For intermediate frequencies, as illustrated in [Fig. 2a](#), the coefficients decrease. Nevertheless, the effect that noise has on these intermediate coefficient, can also be observed either at $\omega = \mathbf{0}$, or $\omega = \omega_{\max}$. A selection of the full result sets, available in the supplementary material [\[18\]](#), is shown in [Fig. 16](#) and [Fig. 17](#).

Noticeable, DP, as a representative of the class of decoherent gate errors, uniformly causes an exponential decline of $\mu_c(\omega)$ for most coefficients and frequencies, while for higher frequencies, mainly for higher qubit counts and for R_X and $R_X R_Y$ encoding $\mu_c(\omega)$ also increases. This may seem counter-intuitive, but can be explained by a low coefficient mean for these frequencies without noise. Adding noise randomly distorts the expectation value, leading to a uniform increase (on average) of all frequencies in the

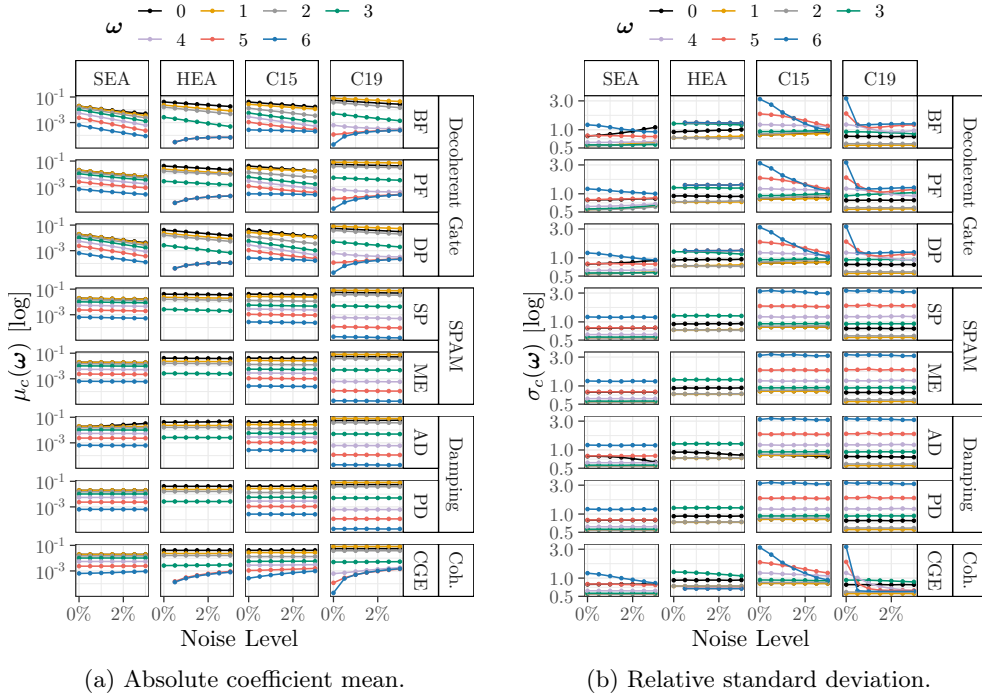


Fig. 15: Analogue to Fig. 6 but for R_X , instead of R_Y encoding. Absolute coefficient mean $\mu_c(\omega)$ and the corresponding relative standard deviation $\sigma_c(\omega)$ over noise levels for various types of noise and six-qubit circuits that use R_X encoding.

spectral representation. The effect is less pronounced for lower frequency components as they typically have a higher mean value. The relative standard deviation $\sigma_c(\omega)$ remains constant or decreases for the special cases where $\mu_c(\omega)$ increases.

For the zero-coefficient, we observe that applying an AD leads to an increase in $\mu_c(\omega)$ and a decrease in $\sigma_c(\omega)$. This effect is more noticeable in circuits with a higher number of qubits and input features. While this shift is more evident in the SEA and HEA, it is less noticeable for C15 and C19, suggesting these ansätze potentially exhibit greater resilience against AD. For the remaining frequencies (specifically ω_{\max}), AD has an imperceptible effect. Similarly, SPAM errors (represented by ME), and PD, not detailed in Fig. 16 and Fig. 17, exhibit no noticeable effect for all ω . Overall, these observations align with the proofs on decoherent noise effects in Ref. [41], showing (uniform) contractions for (DP) noise channels.

The relative standard deviation $\sigma_c(\omega)$ for $D = 2$ ($R_X R_Y$ encoding), shown in Fig. 17, demonstrates similar behaviour as observed in Fig. 6b and Fig. 15b, with constant or decreasing values throughout the noise levels, indicating an overall decrease of variance with $\mu_c(\omega)$. Regarding CGEs, their influence is minimal, when ω is small. However, for high frequencies, there is a significant increase in $\mu_c(\omega)$ and decrease in $\sigma_c(\omega)$ with increasing noise levels in configurations with more than three qubits and

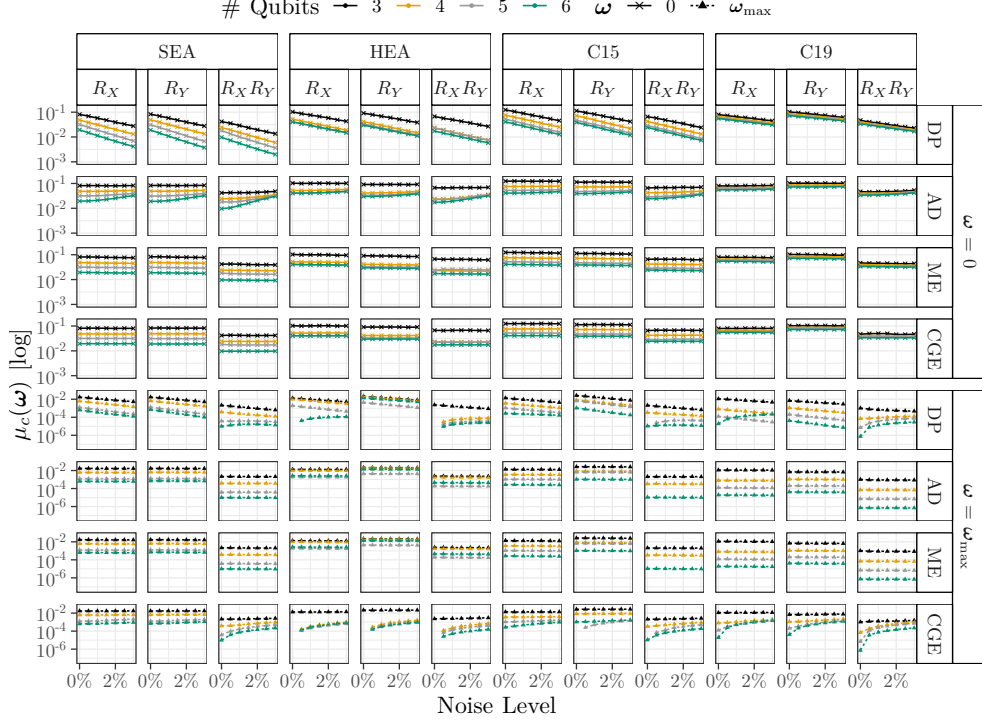


Fig. 16: Absolute coefficient mean $\mu_c(\omega)$ for the lowest frequency $\omega = 0$ and highest frequency $\omega = \omega_{\max}$ in the respective spectrum under the influence of varying noise levels. We considered the one-dimensional encodings R_X and R_Y ($D = 1$), and a two-dimensional encoding $R_X R_Y$ ($D = 2$). The y-axis for each facet row are equal throughout the respective ω , but differs in between.

$D = 2$. This effect, discussed in depth in [Sec. 5.1.4](#), is a consequence of the uniform increase of coefficients across the full spectrum, when applying CGEs.

E Coefficients during Training for Problem Instances

[Fig. 18](#) shows the exact absolute coefficient values over the course of the training described in [Sec. 5.1.5](#) for one of the ten seeds used for the generation of the objective functions. The analogues for the remaining problem seeds are available in our supplementary material in Ref. [\[18\]](#).

In the noiseless scenario, the SEA is capable of learning all coefficients accordingly. It can be observed that the higher frequencies are more prone to the effect of decoherent gate errors compared to lower frequencies, whereas SPAM and DP errors do have a negligible effect. Interestingly, ansätze which do not exhibit a full spectrum, that is the HEA, are rarely affected by errors. The CGEs generally adds high frequent noise on each coefficient magnitude which is feasible as it is modelled by a normal distribution, adding to any gate operation, therefore resulting in white noise in the expectation value.

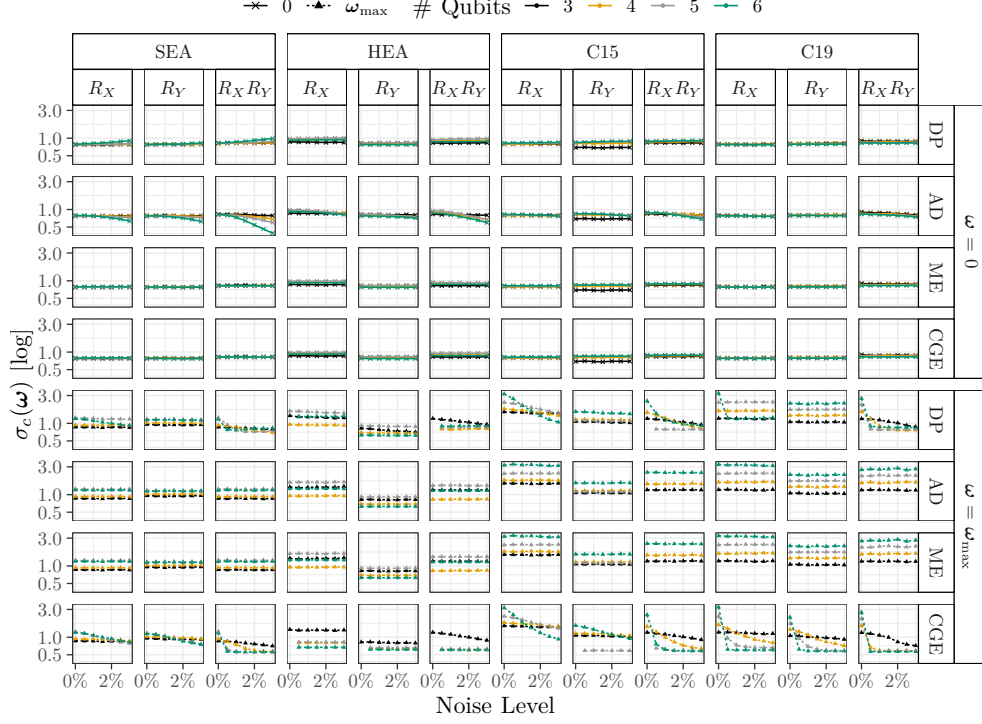


Fig. 17: Relative standard deviation $\sigma_c(\omega)$ for the absolute coefficient mean values from Fig. 16.

In the noiseless scenario, it can be observed that *e.g.* in case of the SEA, lower frequencies are fitted faster than high-frequent coefficients which is in line with an observation made in Ref. [51].

F Entanglement during Training

In Fig. 19 we show the entangling capability over the course of a training. The orders of magnitude for the entangling capability under the influence of the aforementioned noise types are similar to those measured in Sec. 5.3. Over the course of the training we observe a decrease in the entangling capability for most configurations (with the exception of damping errors). While the decrease is more pronounced in some experiments (*e.g.* SEA with DP), the overall trend is consistent across all tested problem and model initialisation seeds for each individually configuration.

As indicated by, for instance, the SEA in the noiseless case, which achieved an MSE near zero (*cf.* Fig. 9) and also converges against a concrete entangling capability, there may exist an optimal degree of entanglement for a learning task in question. However, given that the SEA, HEA and C15 obtain a similar entangling capability value, yet result in rather different MSEs, entangling capability is not a sufficient criterion on which to determine the performance of a QFM.

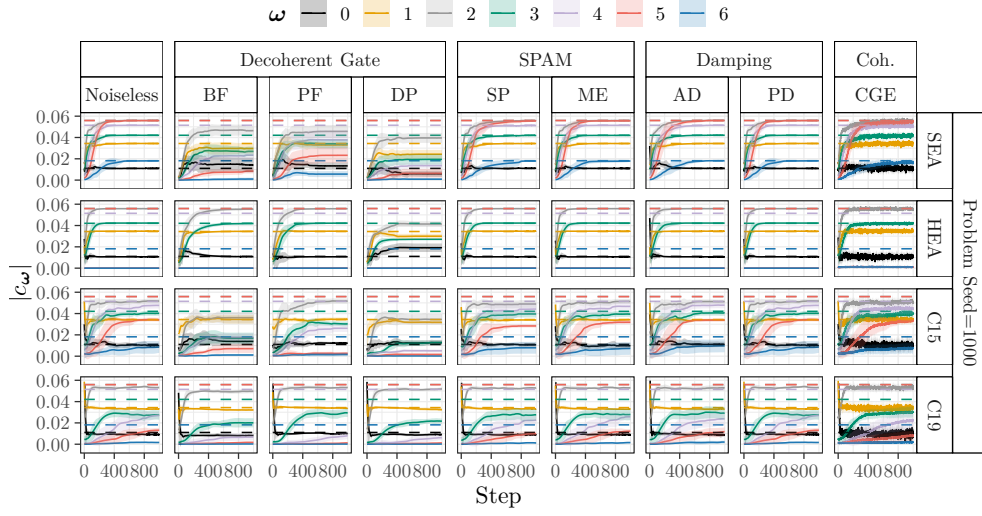


Fig. 18: The learned absolute coefficients of the QFMs averaged over ten parameter model initialisation seeds for a single randomly generated problem instances and different types of noise (3%) during training of 1000 steps. Shaded areas correspond to the standard deviation across the parameter initialisation seeds from the mean. The coefficients of the objective Fourier series are marked with dashed lines.

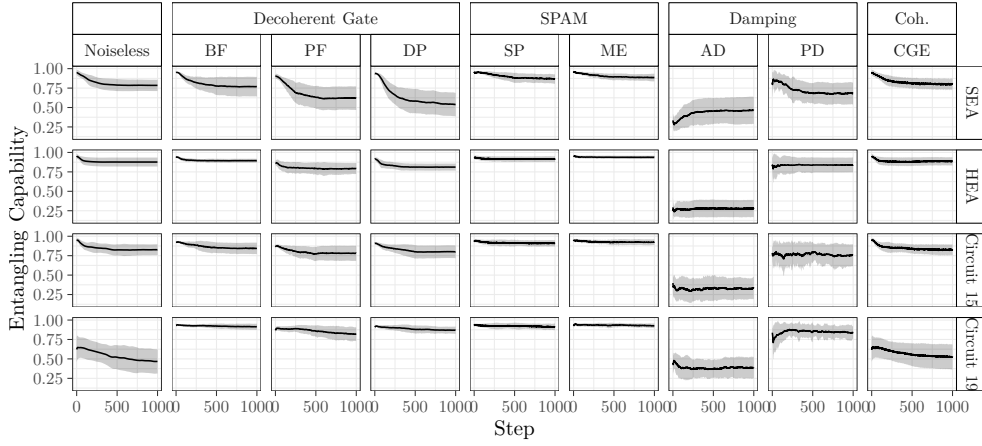


Fig. 19: Entangling capability, assessed with the MW measure for pure states (Noiseless and CGE) and EF for mixed states (decoherent-, SPAM- and damping errors) during training. Lines represent the mean over ten parameter initialisation seeds and ten problem generation seeds. Shaded areas represent the standard deviation.