







# Beyond Gates: Pulse Level Quantum Fourier Models

Melvin Strobl<sup>†</sup>, Maja Franz<sup>§</sup>, Lukas Scheller<sup>†</sup>,  
Eileen Kuehn<sup>†</sup>, Wolfgang Mauerer<sup>§¶</sup>, Achim Streit<sup>†</sup>

<sup>†</sup> Karlsruhe Institute of Technology, Germany, {melvin.strobl, lukas.scheller, eileen.kuehn, achim.streit}@kit.edu

<sup>§</sup> Technical University of Applied Science Regensburg, Germany, {maja.franz, wolfgang.mauerer}@othr.de

<sup>¶</sup> Siemens AG, Foundational Technology, Munich, Germany

**Abstract**—In the domain of variational quantum algorithms, quantum Fourier models (QFMs) provide a mathematically well defined structure for quantum machine learning (QML). There has been a substantial amount of work on the scalability and trainability of such models showcasing the potential but also the limitations for the prospective application of QFMs. However, much less is known in the context of pulse-level quantum computing, where the microwave parameters that implement unitary operations on the hardware are used to perform computations directly instead of through the interface of quantum circuits. In this work, we evaluate QFMs through the lens of pulse parameters and link metrics such as expressibility and Fourier coefficient correlation (FCC) to this extended set of variational parameters. We show that while control over pulse shapes does not significantly alter the global expressibility or structural correlations of an ansatz, it fundamentally alters the local optimisation landscape. For composite gates, independent pulse scalings replace a single logical angle by multiple independently tunable sub-angles. This relaxes the rigid monomial couplings induced by the gate-level parameterisation, and provides gradient descent with higher-dimensional escape routes, decoupling local parameter constraints and significantly boosting performance during training. Following an analytical proof, we show numerical results validating our theory on training a QFM with an exponential (ternary) feature map on a Fourier series with the same frequencies.

**Index Terms**—Quantum Computing, Quantum Machine Learning, Quantum Fourier Models, Pulse Level Quantum Computing

## I. INTRODUCTION

In the field of quantum machine learning (QML), parameterised quantum circuits (PQCs) are generally used when training models conceptually similar to classical neural networks [27, 4]. Among these, the class of data-reuploading models [29], also known as quantum Fourier models (QFMs), are especially interesting as they impose a mathematically well defined structure on the PQC [35], allowing us to study trainability and dequantisability [19, 41]. The trainability of PQCs is generally bottlenecked by two distinct phenomena: Barren Plateaus (BPs) [32] (vanishing gradients globally) and sub-optimal local minima (traps in the loss landscape). While much literature focuses on avoiding BPs via structural constraints [6], less attention is given to navigating the highly non-convex local landscapes of structured models like QFMs.

In the picture of such QFMs, the variance of coefficients, richness of the spectrum and its bias [7] play an important role for the expressivity of the model [26].

Furthermore, correlations between frequency components coin a fingerprint, unique for each ansatz [40] and indicate a limiting factor when training QFMs.

Aforementioned works have been focused on the simulation of such models using gate-based circuit structures. However, running any quantum algorithm on a real device requires (i) compilation into a series of supported basis gates (ii) application of pulse sequences to the quantum system, describing these gates [17]. As the focus of this work lies on the fundamental optimisation landscape and trainability of QFMs, we restrict our analysis to the coherent pulse dynamics, leaving hardware-specific noise [14, 24] and calibration overheads for future empirical studies.

In this picture, pulse sequences are subject to additional parameters, independent of the trainable- or input parameters typically seen in QFMs. Fine-tuning of such parameters is subject to quantum optimal control (QOC) and carried out by the hardware provider. Given that one has access to such parameters, they can be treated as an extended set of parameters in the QML context.

The literature on quantum-classical algorithms like QML and its applications is abundant [3, 5, 45]; while the integration of pulse-level parameters into QML architectures has been explored previously, it is much a much less explored aspect. In Ref. [22], the authors presented an approach of optimising parametrised pulse characteristics for desired operations and ansätze, showcasing advantages in expressibility, entangling capability and efficiency with practical benefits demonstrated in various applications. Magann et al. [23] provide a more in-depth overview on how optimal control within PQCs necessitates sufficient control resources without which the potential benefits offered by pulse-level control may be limited. More recently, Acedo et al. [1] formulated a pulse-level equivalent of data reuploading models (QFMs) showcasing noise robustness on superconducting transmon processors.

In this work, we aim to contribute to a more theoretical understanding of the pulse-level representation of QFMs and show that access to pulse parameters can activate or unsuppress coefficients associated with frequencies already present in the encoded frequency set, and can enlarge the algebraic support relative to gate-based QFMs and thus significantly improve

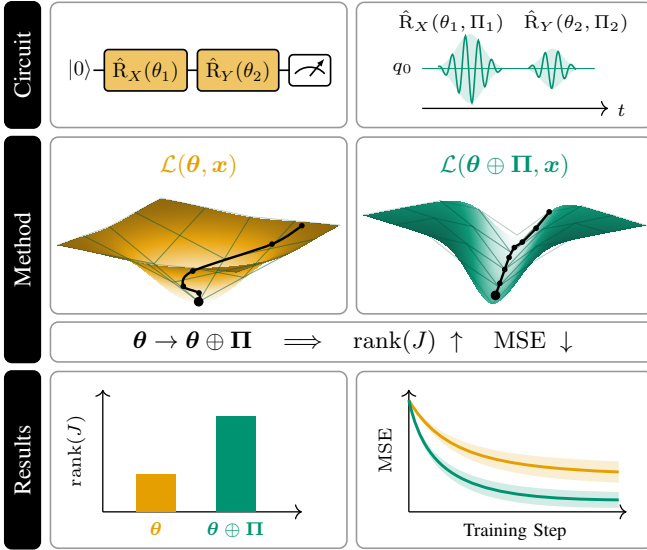


Fig. 1: Illustration of our contribution. The optimisation trajectory of a QFM in  $\mathcal{L}$  can be improved by extending the **gate parameters**  $\theta$  through access to **pulse parameters**  $\Pi$ . This allows to escape local minima and increases the rank of the coefficient Jacobian  $\text{rank}(J)$ . Practically, the combined optimisation of  $\theta \oplus \Pi$  consistently lowers the mean squared error (MSE) towards a global optimum for all ansätze examined in this work.

the training of such models, even beyond over-parametrised ansätze obtained through decomposition into basis gates.

Analytical findings are supported through numerical experiments for which we provide the source code for reproduction purposes [25] on Github<sup>1</sup> and in Ref. [38]. We illustrate the core contribution of our work in Fig. 1.

## II. METHOD

This section describes the theoretical foundations and methodology of this work. We generally denote vectors by bold symbols; for example,  $\theta$  denotes the vector with components  $\theta_k$ .

### A. Quantum Fourier Models

Using the same framework as in Ref. [40], we define a QFM with  $L$  layers by

$$\hat{U}(\mathbf{x}, \theta) = \hat{W}^{(L+1)}(\theta) \hat{S}(\mathbf{x}) \hat{W}^{(L)}(\theta) \cdots \hat{W}^{(2)}(\theta) \hat{S}(\mathbf{x}) \hat{W}^{(1)}(\theta).$$

Here,  $\hat{W}^{(\ell)}$  and  $\hat{S}$  are  $n$ -qubit trainable and encoding unitaries, respectively, which depend on trainable parameters  $\theta$  and input parameters  $\mathbf{x}$ .

Using the notation of Ref. [35], the expectation value of such a model can be written as a truncated Fourier series,

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

<sup>1</sup><https://github.com/cirKITers/pulse-level-quantum-fourier-models>

where  $\Omega$  denotes the set of unique frequency vectors and  $c_{\omega}(\theta)$  is the complex-valued Fourier coefficient associated with  $\omega$ . An analytical expression for the Fourier coefficients was given by Wiedmann et al. [43]:

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

Here,  $p_{s, c}(\omega)$  collects all contributions to the coefficient at frequency  $\omega$ , while  $s, c$  and  $s', c'$  count the numbers of sine and cosine factors associated with the input and trainable parameters, respectively. The constants  $k_{s, c, s', c'}$  depend on the chosen basis functions.

Extending the notation of Ref. [43], it is useful to distinguish three related notions of frequency support: (a) The *encoded frequency set* is the set of frequencies determined by the eigenvalue gaps of the data-encoding generators – the set  $\Omega$  appearing in Eq. 1. Since pulse parameters enter only through the trainable unitaries, they do not change this encoded frequency set. (b) For a fixed family of ansätze  $\mathcal{A}$ , not every frequency in  $\Omega$  needs to have a non-vanishing coefficient. We therefore define the *algebraic coefficient support*

$$\Omega_{\text{supp}}^{\mathcal{A}} := \{\omega \in \Omega \mid c_{\omega}(\mathbf{p}) \neq 0\},$$

where  $\mathbf{p}$  denotes the trainable parameters of the chosen implementation. For a gate-level implementation  $\mathbf{p} = \theta$ , whereas for a pulse-level implementation  $\mathbf{p} = (\theta, \Pi)$ , or equivalently  $\mathbf{p} = (\theta, \lambda)$  when only effective pulse scalings are considered (cf. Sec. II-C). Frequencies in  $\Omega \setminus \Omega_{\text{supp}}^{\mathcal{A}}$  are allowed by the feature map but vanish for the chosen ansatz, for example due to algebraic cancellations or symmetries. (c) For a parameter-sampling distribution  $\mu$  and threshold  $\tau > 0$ , we define the set of *high-variance coefficients* by

$$\Omega_{\text{act}}^{\mathcal{A}}(\mu, \tau) := \left\{ \omega \in \Omega_{\text{supp}}^{\mathcal{A}} \mid \text{Var}_{\mathbf{p} \sim \mu} [c_{\omega}(\mathbf{p})] > \tau \right\}. \quad (3)$$

Saying that pulse parameters activate frequencies (cf. Sec. III-B) means they can increase the variance of coefficients already contained in the encoded frequency set  $\Omega$ , but do not introduce new encoded frequencies.

We will also use the coefficient-space map

$$C_{\mathcal{A}} : \mathcal{P} \longrightarrow \mathbb{C}^{|\Omega|}, \quad C_{\mathcal{A}}(\mathbf{p}) := (c_{\omega}(\mathbf{p}))_{\omega \in \Omega}, \quad (4)$$

where  $\mathcal{P}$  is the parameter domain of the chosen implementation. Training against a target coefficient vector  $C^* = (c_{\omega}^*)_{\omega \in \Omega}$  can be written as

$$\mathcal{L}(\mathbf{p}) = \|C_{\mathcal{A}}(\mathbf{p}) - C^*\|_2^2.$$

For real-valued expectation values, the Fourier coefficients obey the Hermitian symmetry  $c_{-\omega} = \overline{c_{\omega}}$ , and the same constraint must be imposed on the target coefficients  $c_{\omega}^*$ .

The spectrum of a QFM is determined by eigenvalue differences of the generators appearing in the feature map [19]. For each encoded input component  $x_i$ , let

$$\begin{aligned}\Omega_i &= \{\Lambda_{i,\mathbf{j}} - \Lambda_{i,\mathbf{k}} \mid \mathbf{j}, \mathbf{k} \in \llbracket 1, d_i \rrbracket^L\}, \\ \tilde{\Omega} &= \Omega_1 \times \dots \times \Omega_D,\end{aligned}\quad (5)$$

with multi-indices  $\mathbf{j} = (j_1, \dots, j_L)$  and  $\mathbf{k} = (k_1, \dots, k_L)$ .

$$\Lambda_{i,\mathbf{j}} = \sum_{\ell=1}^L \lambda_{j_\ell}^{(i,\ell)}.$$

is the sum of eigenvalues selected by the multi-index  $\mathbf{j}$  across the  $L$  occurrences of the  $i$ -th encoding generator.

Naturally, frequency-dependent redundancies arise when different eigenvalue-gap combinations generate the same frequency [19, 16]. Following Mhiri et al. [26], we denote the corresponding *frequency generator* by  $R(\omega)$ . These redundancies induce a feature-map-dependent weighting over  $\Omega$ , and the size of the generator controls the variance of the coefficients via [26]

$$\begin{aligned}\text{Var}_\theta[c_\omega(\theta)] &\leq \text{Var}_{\text{Haar}}[c_\omega] \\ &+ \left( \frac{C_1 \varepsilon}{d^2} + \frac{C_2 \varepsilon}{d(d+1)} \right) |R(\omega)| \\ &+ C_2 \frac{\varepsilon^2}{d^2} |R(\omega)|^2.\end{aligned}\quad (6)$$

Here, the  $\varepsilon$ -distance to a 2-design quantifies how well the ensemble of unitaries generated by uniformly sampling  $\theta$  covers the unitary group. This directly connects to the notion of *expressibility* [37]: highly expressible ansätze cover the unitary space more uniformly, corresponding to  $\varepsilon \rightarrow 0$ . Accordingly, for highly expressible models, the coefficient variance approaches the Haar-random limit, whereas restricted expressibility tightens the bound through the multiplicities encoded in  $R(\omega)$ . We refer to [Apx. B](#) for details on the computation of this metric.

In addition to expressibility, one may quantify global dependencies between Fourier coefficients through the Fourier coefficient correlation (FCC). Since this metric is not needed for the arguments below, we defer its precise definition to [Apx. C](#).

### B. Pulse Level Implementation of Unitary Gates

Parameterised gates are the basic building blocks of trainable quantum circuits. During transpilation, such gates are decomposed into basis gates supported by the target hardware. Each basis gate is then realised by a control pulse sequence implementing the corresponding unitary evolution. It is useful for the below calculations to consider this in the wider context of a system-level description of a QPU: For many platforms, notably including superconducting Transmons, the system Hamiltonian can be modelled as  $\hat{H}(t) = \hat{H}_{\text{static}} + \hat{H}_{\text{drive}}(t) + \hat{H}_{\text{dynamic}}(t)$ , where  $\hat{H}_{\text{static}}$  describes unperturbed energy levels of the qubits and their constant couplings;  $\hat{H}_{\text{drive}}(t)$  represents external pulses applied to manipulate the qubit states (this is where pulse

parameters  $\mathbf{\Pi}$  enter), and  $\hat{H}_{\text{dynamic}}(t)$  accounts for time-varying couplings or environmental interactions, often used to model noise or specific tunable couplers.

We denote a single control pulse by  $S_j(t, \mathbf{\Pi})$ , where  $\mathbf{\Pi}$  collects the pulse-shape parameters and  $t$  denotes time. A standard scalar decomposition model (assuming no complex amplitudes or more than one quadrature) is

$$S_j(t, \mathbf{\Pi}) = E_j(t, \mathbf{\Pi}) \cos(\omega_c t + \phi_c),$$

where  $E_j(t, \mathbf{\Pi})$  is the pulse envelope, and  $\omega_c$  and  $\phi_c$  are carrier frequency and carrier phase, respectively. Combining the control signals with the control Hamiltonians  $\hat{H}_j$  yields the time-dependent Hamiltonian

$$\hat{H}(t, \mathbf{\Pi}) = \hat{H}_{\text{static}} + \sum_j S_j(t, \mathbf{\Pi}) \hat{H}_j, \quad (7)$$

which governs the evolution of a state  $|\psi(t)\rangle$ .

For a time-independent Hamiltonian, the evolution from  $t_0$  to  $t_0 + \delta t$  is (using the convention  $\hbar = 1$ )  $|\psi(t_0 + \delta t)\rangle = e^{-i\hat{H}\delta t} |\psi(t_0)\rangle$ . For the time-dependent Hamiltonian in [Eq. 7](#), the time evolution of a quantum state is given by ( $\mathcal{T}$  denotes time ordering [8, 33])

$$|\psi(t_0 + \delta t)\rangle = \mathcal{T} \exp \left( -i \int_{t_0}^{t_0 + \delta t} \hat{H}(t, \mathbf{\Pi}) dt \right) |\psi(t_0)\rangle. \quad (8)$$

Typically, this equation must be solved numerically. To balance computational economy and result accuracy, we perform the required computations, but using the common rotating-wave approximation (RWA) as detailed below. We refer to [Ref. \[11\]](#) for implementation details.

### C. Pulse Parameters in Quantum Fourier Models

Combining [Sec. II-A](#) and [Sec. II-B](#), we now study the effect of treating pulse parameters  $\mathbf{\Pi}$  as additional trainable variables during QFM optimisation. Since these parameters appear only in the trainable unitaries, they do not alter frequencies in the exponent of [Eq. 1](#); they only modify Fourier coefficients. This allows us to compare gate-level and pulse-level optimisation while keeping the represented frequency set fixed.

Among our numerical results, it is useful to consider a simplified fixed-axis pulse-area model to guide intuition on the observed phenomena, and to motivate our chosen diagnostics. We employ the canonical RWA [17] to absorb a rapidly oscillating carrier into an effective drive Hamiltonian. Gate actions are then governed primarily by pulse envelope and phase, while the generator  $\tilde{H}_j$  is fixed by the rotation axis. Pulse parameters then act as effective angle scalings.

To isolate the main mechanism, consider first a single-axis rotation in the effective rotating frame. Assume that the corresponding effective Hamiltonian takes the form

$$\hat{H}_{\text{eff}}(t, \theta, \mathbf{\Pi}) = \theta E(t, \mathbf{\Pi}) \tilde{H}_j,$$

where  $\tilde{H}_j$  is time independent.<sup>2</sup> Then the resulting unitary (assuming fixed-axis resonant control and commuting effective Hamiltonians  $[\hat{H}_{\text{eff}}(t), \hat{H}_{\text{eff}}(t')] = 0$ )<sup>3</sup> is

$$\begin{aligned}\hat{U}(\theta, \mathbf{\Pi}) &= \exp\left(-i\theta \int_0^{\delta t} E(\tau, \mathbf{\Pi}) \tilde{H}_j d\tau\right) \\ &= \exp\left(-i\theta \lambda(\mathbf{\Pi}) \tilde{H}_j\right),\end{aligned}\quad (9)$$

with the effective pulse-area factor

$$\lambda(\mathbf{\Pi}) := \int_0^{\delta t} E(\tau, \mathbf{\Pi}) d\tau.$$

For a rectangular envelope  $E(t, A) = A$  on  $[0, \delta t]$ , this reduces to  $\lambda(\mathbf{\Pi}) = A \delta t$ . Any constant prefactors arising from calibration or the RWA can be absorbed into  $\lambda(\mathbf{\Pi})$ .

Thus, for a single basis rotation, the pulse parameter acts only through the effective angle  $\theta \lambda(\mathbf{\Pi})$ . In this simplified setting, optimising  $\mathbf{\Pi}$  amounts to a reparameterisation of the original gate parameter. This remains valid if  $\delta t$  itself is included in  $\mathbf{\Pi}$ : its effect is still mediated through the scalar factor  $\lambda(\mathbf{\Pi})$ .

For more general envelopes, the dependence on  $\mathbf{\Pi}$  need not reduce exactly to a single scalar factor at the laboratory-frame level. However, in the effective single-axis description the dominant effect is still a pulse-dependent rescaling of the rotation angle, and the argument below carries through with only notational modifications. For example, using a textbook calculation [2] for a Gaussian envelope  $E(t, A, \sigma) = A \exp\left(-1/2 \left(\frac{t-t_c}{\sigma}\right)^2\right)$ , we find

$$\begin{aligned}\lambda(A, \sigma) &= \int_0^{\delta t} E(t, A, \sigma) dt \\ &= A\sigma \sqrt{\frac{\pi}{2}} \left[ \text{erf}\left(\frac{\delta t - t_c}{\sqrt{2}\sigma}\right) + \text{erf}\left(\frac{t_c}{\sqrt{2}\sigma}\right) \right] \\ &\approx A\sigma \sqrt{2\pi}\end{aligned}\quad (10)$$

$$\approx A\sigma \sqrt{2\pi}\quad (11)$$

where the approximation holds for a centered pulse ( $t_c = \delta t/2$ ) that is contained in the pulse window ( $\delta t \gg \sigma$ ).

The situation changes for composite gates. There, a single logical gate parameter  $\theta$  controls several sub-rotations, while the pulse parameters of the sub-gates can vary independently. This distinction is illustrated by the following example.

<sup>2</sup>The scalar pulse-area model is appropriate when pulse parameters modify only the amplitude or duration of a resonant fixed-axis control (*i.e.*, the pulse changes only the total rotation angle). In more general pulse models that consider, for instance, detuning, time-dependent phases, leakage or crosstalk, pulse parameters can affect several independent physical contributions to the implemented operation. In that case,  $\lambda(\mathbf{\Pi})$  would need to be replaced by a vector, which we do not consider in the present paper. Pulse optimisation could then change not only the size of the rotation, but also the effective axis or phase, or introduce multi-qubit interaction terms generated by the pulse.

<sup>3</sup>Laboratory-frame Hamiltonians do not commute at different times. In the rotating frame with the RWA applied, a resonant fixed-phase single-axis drive is described by an effective Hamiltonian  $\hat{H}_{\text{eff}}(t) = \Omega(t)G$ , where  $G$  is time independent. Only in this effective description is the time ordering trivial, and the pulse affects the gate primarily through the pulse area  $\int \Omega(t) dt$ .

**Example 1** (Composite gate with independent pulse parameters). Assume the basis gate set  $\{\hat{R}_X, \hat{R}_Y, \hat{R}_Z, C\hat{X}\}$  and the decomposition

$$C\hat{R}_X(\theta) = \hat{R}_Z(\pi/2) \cdot \hat{R}_Y(\theta/2) \cdot C\hat{X} \cdot \hat{R}_Y(-\theta/2) \cdot C\hat{X} \cdot \hat{R}_Z(-\pi/2).$$

Let the  $m$ -th basis gate (out of  $M$  sub-gates) carry its own pulse parameter  $\mathbf{\Pi}_m$ , and let  $\lambda_m(\mathbf{\Pi}_m)$  denote the corresponding effective scaling. Then the implemented unitary may be written as

$$\hat{U}_{\text{CRX}}(\theta, \{\mathbf{\Pi}_m\}_{m=1}^M) = \prod_{m=1}^M U_m(\alpha_m(\theta), \mathbf{\Pi}_m),$$

where  $\alpha_m(\theta)$  is the sub-gate angle induced by the global logical parameter  $\theta$ . Using  $e^{-i\frac{\gamma}{2}P} = \cos(\frac{\gamma}{2})\mathbb{1} - i\sin(\frac{\gamma}{2})P$ , each sub-gate contributes factors of the form

$$\sin(\alpha_m(\theta)\lambda_m(\mathbf{\Pi}_m)/2), \quad \cos(\alpha_m(\theta)\lambda_m(\mathbf{\Pi}_m)/2).$$

Since a single  $\theta$  controls several distinct angles  $\alpha_m(\theta)$ , while the  $\lambda_m(\mathbf{\Pi}_m)$  are independent, it is generically not reducible to a single reparametrised angle. In particular, on the control branch  $|0\rangle$ , the standard decomposition yields

$$\hat{R}_Z(\pi/2) \hat{R}_Y(\theta/2) \hat{R}_Y(-\theta/2) \hat{R}_Z(-\pi/2) = \mathbb{1}.$$

If the two  $\hat{R}_Y$  sub-gates are independently rescaled by  $\lambda_1$  and  $\lambda_2$ , then the same branch becomes

$$\begin{aligned}\hat{R}_Z(\pi/2) \hat{R}_Y(\theta/2\lambda_1) \hat{R}_Y(-\theta/2\lambda_2) \hat{R}_Z(-\pi/2) \\ = \hat{R}_X(-\theta/2(\lambda_1 - \lambda_2)).\end{aligned}$$

Hence, if  $\lambda_1 \neq \lambda_2$ , the operation on the  $|0\rangle$  branch is no longer the identity. The resulting controlled operation therefore leaves the original  $C\hat{R}_X$  gate family, even though it remains block diagonal in the computational basis. This demonstrates that pulse-level flexibility in a composite implementation spans a larger local unitary family than simple angle reparameterisation of the logical gate.

More generally, if  $U(\theta, \mathbf{\Pi})$  is a pulse-level realisation of a single-axis basis rotation, then it is unitarily equivalent to  $\hat{R}_j(\theta\lambda(\mathbf{\Pi}))$ . By contrast, for a composite logical gate

$$\hat{U}_{\text{comp}}(\theta, \{\mathbf{\Pi}_m\}_{m=1}^M) = \prod_{m=1}^M U_m(g_m(\theta), \mathbf{\Pi}_m),$$

the Pauli decomposition contains terms of the form

$$\prod_{m=1}^M \sin(g_m(\theta)\lambda_m(\mathbf{\Pi}_m))^{a_m} \cos(g_m(\theta)\lambda_m(\mathbf{\Pi}_m))^{b_m},$$

with some sub-gate specific factors  $g_m$ ,  $a_m$  and  $b_m$ . Whenever  $M \geq 2$ , at least two sub-gates act on different axes, and the  $\lambda_m(\mathbf{\Pi}_m)$  are independently adjustable, these factors are not reducible to a function of a single effective angle.

To clarify this mechanism further, let a QFM contain  $w$  trainable basis rotations, and let the  $k$ -th of these gates be implemented in the effective single-axis form

$$\hat{U}_k(\theta_k, \mathbf{\Pi}_k) = \exp(-i\theta_k \lambda_k(\mathbf{\Pi}_k) \tilde{H}_k),$$

where  $\tilde{H}_k$  is the corresponding generator. Consider the effective angle  $\varphi_k := \theta_k \lambda_k(\mathbf{\Pi}_k)$ , then the Fourier coefficients retain the monomial structure of Ref. [43] and may be written as

$$c_\omega(\boldsymbol{\theta}, \mathbf{\Pi}) = \sum_{(s,c,s',c')} \left( \kappa_{s,c,s',c'} p_{s,c}(\omega) \times \prod_{k=1}^w \sin(\varphi_k)^{s'_k} \cos(\varphi_k)^{c'_k} \right), \quad (12)$$

where  $\kappa_{s,c,s',c'}$  absorb constants and phases from Eq. 2.

In particular, the map

$$(\boldsymbol{\theta}, \mathbf{\Pi}) \mapsto \boldsymbol{\phi} = (\varphi_1, \dots, \varphi_w)$$

has image in  $\mathbb{R}^w$ , so basis-gate pulse parameters alone do not enlarge the reachable coefficient set; they only change its parametrisation. Equivalently, the Jacobian

$$\frac{\partial \boldsymbol{\phi}}{\partial (\boldsymbol{\theta}, \mathbf{\Pi})}$$

has rank at most  $w$  in a  $(w + |\mathbf{\Pi}|)$ -dimensional parameter space and therefore possesses a non-trivial kernel.

Now let a logical trainable gate  $k$  be implemented as a composite gate

$$\hat{U}_{\text{comp},k}(\theta_k, \{\mathbf{\Pi}_{k,m}\}_{m=1}^{M_k}) = \prod_{m=1}^{M_k} \hat{U}_{k,m}(g_{k,m}(\theta_k), \mathbf{\Pi}_{k,m}).$$

Then its Pauli decomposition contains factors of the form

$$\prod_{m=1}^{M_k} \sin(g_{k,m}(\theta_k) \lambda_{k,m}(\mathbf{\Pi}_{k,m}))^{a_{k,m}} \times \cos(g_{k,m}(\theta_k) \lambda_{k,m}(\mathbf{\Pi}_{k,m}))^{b_{k,m}}. \quad (13)$$

These terms are usually not reducible to a function of a single effective logical angle. Therefore, at the logical-gate level, independent pulse parameters can relax the rigid monomial coupling present in the gate-based representation.

#### D. Effects on Learning

In standard gate-based QML, escaping poor stationary points often requires over-parameterisation, for instance by increasing circuit depth [20]. The above consideration shows that pulse-level control provides a hardware-native analogue of this approach: for basis gates it changes the optimisation parameterisation, whereas for composite gates it can introduce additional local directions in coefficient space without changing the encoded frequency set.

As Eq. 13 shows, pulse parameters can partially decouple terms that are rigidly linked at the logical-gate level. For non-rectangular envelopes, additional pulse-shape dependence generally produces further correction terms; this does not alter the encoded spectrum, but it can increase the local flexibility of the coefficient map.

**Remark 1.** *Pulse parameters need not substantially change global metrics such as expressibility, even when they improve trainability. For the family of ansätze considered here, their primary effect is to modify the parameterisation of the trainable unitary family and, for composite gates, to enlarge local tangent directions accessible during optimisation. Accordingly, pulse-level control can smooth the local loss landscape without necessarily changing the global set of representable frequencies.*

#### E. Training and Trainability

We consider a target function  $f^*$  whose frequency support matches that of the QFM, with coefficients  $C^*$  (cf. Sec. II-A) drawn randomly from the unit disk. In this subsection, we restrict to the one-dimensional input case for simplicity. Assuming that the frequencies are orthogonal on  $[0, 2\pi]$  (for example,  $\Omega \subset \mathbb{Z}$ ), Parseval's identity gives

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \frac{1}{2\pi} \int_0^{2\pi} |f(x, \boldsymbol{\theta}) - f^*(x)|^2 dx \\ &= \sum_{\omega \in \Omega} |c_\omega(\boldsymbol{\theta}) - c_\omega^*|^2. \end{aligned} \quad (14)$$

Thus, training can be viewed as matching the model coefficients to the target coefficients in Fourier space.

Following Eq. 2 and Ref. [26], different frequencies generally can depend on the same gate parameter  $\theta_k$ , which creates coupled constraints across coefficient directions. Using the coefficient-space map from Eq. 4, the gate-level Jacobian is

$$J_\theta := \frac{\partial C_{\mathcal{A}}}{\partial \boldsymbol{\theta}} \in \mathbb{C}^{|\Omega| \times |\boldsymbol{\theta}|}. \quad (15)$$

Since the trainable parameters are real while the Fourier coefficients are complex, all ranks and tangent-space dimensions below are understood over  $\mathbb{R}$  with  $\mathbb{C}^{|\Omega|} \cong \mathbb{R}^{2|\Omega|}$ .

If we define the residual vector

$$\mathbf{r} := C_{\mathcal{A}}(\boldsymbol{\theta}) - C^*, \quad r_\omega = c_\omega(\boldsymbol{\theta}) - c_\omega^*,$$

then a first-order critical point of  $\mathcal{L}$  satisfies

$$\text{Re} \left( \mathbf{r}^\dagger \frac{\partial C_{\mathcal{A}}}{\partial \theta_k} \right) = \sum_{\omega \in \Omega} \text{Re} \left( r_\omega^* \frac{\partial c_\omega}{\partial \theta_k} \right) = 0 \quad \forall k.$$

This condition is automatically satisfied at a global minimum, where  $C_{\mathcal{A}}(\boldsymbol{\theta}) = C^*$ . More generally, it is also satisfied at non-global stationary points whenever the residual is orthogonal to all tangent directions generated by the columns of  $J_\theta$ .

To build geometric intuition, assume that  $|\Omega| > |\boldsymbol{\theta}|$ . Then the coefficient-space map

$$C_{\mathcal{A}} : \mathbb{R}^{|\boldsymbol{\theta}|} \longrightarrow \mathbb{C}^{|\Omega|} \cong \mathbb{R}^{2|\Omega|}$$

has an image that, locally and at regular points, is an immersed manifold  $\mathcal{M}$  of dimension

$$\dim T_c \mathcal{M} = \text{rank}_{\mathbb{R}} J_\theta \leq |\boldsymbol{\theta}|.$$

Local optimisation is equivalent to moving on this coefficient manifold to reduce the Euclidean distance to the target point  $C^*$ . Because the coefficients are trigonometric polynomials in the parameters,

$$c_\omega(\boldsymbol{\theta}) = \sum_\ell \kappa_\ell \prod_{k=1}^w \sin(\theta_k)^{s'_{\ell,k}} \cos(\theta_k)^{c'_{\ell,k}},$$

the manifold  $\mathcal{M}$  is generally highly curved and may exhibit self-intersections, singular points, or narrow tangent directions.

As in [Sec. II-C](#), two distinct cases arise when pulse parameters are introduced and we optimise

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Pi}) := \|\mathbf{c}(\boldsymbol{\theta}, \boldsymbol{\Pi}) - \mathbf{c}^*\|^2.$$

*a) Basis gates:* If each trainable gate remains an effective single-axis basis rotation, then the reachable coefficient manifold does not change; only its parameterisation changes. By the chain rule,

$$\frac{\partial \mathcal{L}}{\partial \Pi_k} = \frac{\partial \mathcal{L}}{\partial \varphi_k} \theta_k \frac{\partial \lambda_k}{\partial \Pi_k}, \quad (16)$$

where  $\varphi_k = \theta_k \lambda_k(\Pi_k)$ . Hence, pulse parameters act as parameter-dependent rescaling or preconditioning of the gradient field (given that  $\theta_k \neq 0$  and  $\partial \lambda_k / \partial \Pi_k \neq 0$  as otherwise, the pulse parameters provide no first-order direction). They can improve conditioning and alter optimisation trajectories, but they do not enlarge the set of reachable coefficients.

*b) Composite gates:* If a logical gate is implemented by several basis gates with independently adjustable pulse scalings, then additional tangent directions can appear in coefficient space. Let  $\boldsymbol{\lambda}$  denote the collection of effective pulse scalings with  $J_\lambda(\boldsymbol{\lambda}) := \frac{\partial C_A}{\partial \boldsymbol{\lambda}}$  and define the pulse-level Jacobian

$$J_{\text{ext}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) := \frac{\partial C_A}{\partial(\boldsymbol{\theta}, \boldsymbol{\lambda})} = [J_\theta \quad J_\lambda]. \quad (17)$$

The tangent spaces accessible to first-order optimisation are<sup>4</sup>

$$T_{\text{gate}} = \text{Im}(J_\theta), \quad T_{\text{pulse}} = \text{Im}(J_{\text{ext}}).$$

Consequently, the increase in local tangent-space dimension is

$$\dim T_{\text{pulse}} - \dim T_{\text{gate}} = \text{rank}_{\mathbb{R}} J_{\text{ext}} - \text{rank}_{\mathbb{R}} J_\theta. \quad (18)$$

This implies that pulse parameters provide genuinely new local search directions when

$$\text{rank}_{\mathbb{R}} J_{\text{ext}} > \text{rank}_{\mathbb{R}} J_\theta.$$

Equivalently, at least one pulse derivative  $\partial C_A / \partial \lambda_{k,m}$  must have a component outside  $\text{Im}(J_\theta)$ .

<sup>4</sup>The gate-level tangent space is the space of all first-order coefficient variations induced by infinitesimal changes of the gate parameters. More precisely, at a point  $\theta_0$ , let  $\theta(t) = \theta_0 + tv$  with  $v \in \mathbb{R}^{|\theta|}$ . Then

$$C_A(\theta(t)) = C_A(\theta_0) + tJ_\theta(\theta_0)v + \mathcal{O}(t^2).$$

Thus the accessible first-order directions in coefficient space are

$$T_{\text{gate}}(\theta_0) := \{J_\theta(\theta_0)v \mid v \in \mathbb{R}^{|\theta|}\} = \text{Im} J_\theta(\theta_0).$$

Since the coefficients are complex-valued and the trainable parameters are real, this tangent space is understood as a real linear subspace of  $\mathbb{C}^{|\Omega|} \simeq \mathbb{R}^{2|\Omega|}$ .

If  $q$  independent pulse-scaling variables are introduced, the general upper bound is

$$0 \leq \text{rank}_{\mathbb{R}} J_{\text{ext}} - \text{rank}_{\mathbb{R}} J_\theta \leq \min\{q, 2|\Omega| - \text{rank}_{\mathbb{R}} J_\theta\}.$$

For a composite gate with  $M_g$  independently scaled sub-gates, a common rescaling of all sub-rotations is often equivalent to changing the original logical gate angle. In such cases, at most  $M_g - 1$  directions per composite gate can be genuinely new. However, the actual increase is not determined solely by the number of sub-gates or rotation axes; it is always given by the rank difference in [Eq. 18](#).

Composite gates can therefore enlarge the local tangent space seen by gradient descent even though the encoded frequency set  $\Omega$  is unchanged. In practice, most ansätze contain both types of gates. Basis-gate pulse parameters alter the optimisation metric on the same manifold, while composite-gate pulse parameters can add genuinely new local search directions.

Because [Eq. 14](#) is the squared distance from  $\mathbf{c}^*$  to the coefficient manifold, stationary points occur when the residual vector is orthogonal to the tangent space. Enlarging the tangent space therefore makes it less likely that a non-zero residual remains orthogonal to all available descent directions.

**Example 2.** Consider a logical gate with a single parameter  $\theta$  and decomposition

$$\hat{U}_{\text{std}}(\theta) = \hat{R}_X(a\theta)\hat{R}_Y(b\theta).$$

In the standard parameterisation, the pair of effective sub-angles  $(a\theta, b\theta)$  lies on a one-dimensional line in  $\mathbb{R}^2$ , as the ratio of angles remains invariant with scaling  $\theta$ . If the two sub-gates acquire independent pulse scalings, the implemented gate becomes

$$\hat{U}_{\text{pulse}}(\theta, \lambda_1, \lambda_2) = \hat{R}_X(a\theta\lambda_1)\hat{R}_Y(b\theta\lambda_2),$$

so the pair of effective sub-angles becomes  $(a\theta\lambda_1, b\theta\lambda_2)$ . Near generic points with  $\theta \neq 0$ , this spans a two-dimensional region rather than a one-dimensional line, as  $\lambda_i$  can be independently scaled. Thus, pulse parameters create local directions that are unavailable in the logical gate parameter alone. In terms of the above considerations, the change is from rank 1 to rank 2 in the Jacobian.

This observation leads to the following statement.

**Theorem 1** (Escape directions from non-global stationary points). Consider a QFM trained on a Fourier-compatible target

$$f^*(x) = \sum_\omega c_\omega^* e^{i\omega x},$$

and let  $\boldsymbol{\theta}^*$  be a non-global critical point of the gate-level loss  $\mathcal{L}(\boldsymbol{\theta})$  with residual  $\mathbf{r} \neq 0$ . Assume that, at  $(\boldsymbol{\theta}^*, \boldsymbol{\lambda} = \mathbf{1})$ , the derivatives with respect to at least one composite-gate pulse scaling produce a coefficient-space direction that is not

contained in the span of the gate-level Jacobian columns. Then, for Lebesgue-almost all target coefficient vectors  $\mathbf{c}^*$ , the point  $(\boldsymbol{\theta}^*, \boldsymbol{\lambda} = \mathbf{1})$  is not a critical point of the extended loss  $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ . In particular,

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L} \Big|_{\boldsymbol{\lambda}=\mathbf{1}} \neq 0,$$

so the extended parameterisation provides a first-order escape direction.

Eq. 18 gives the geometric mechanism behind this result: composite-gate pulse parameters can enlarge the local tangent space. Theorem 1 then states that, under a generic non-degeneracy assumption, these additional directions almost surely destroy non-global criticality for randomly drawn targets. A proof is provided in the following.

**Proof 1** (Proof of Theorem 1). At  $(\boldsymbol{\theta}^*, \boldsymbol{\lambda} = \mathbf{1})$ , the gradient along a composite-gate pulse scaling  $\lambda_{k,m}$  is:

$$\frac{\partial \mathcal{L}}{\partial \lambda_{k,m}} \Big|_{\boldsymbol{\lambda}=\mathbf{1}} = 2 \operatorname{Re} \left[ \sum_{\omega} r_{\omega}^* \frac{\partial c_{\omega}}{\partial \lambda_{k,m}} \Big|_{\boldsymbol{\lambda}=\mathbf{1}} \right] \quad (19)$$

For a composite gate with sub-gates on distinct axes,  $\frac{\partial c_{\omega}}{\partial \lambda_{k,m}}$  is not in  $\operatorname{range}(J_{\theta})$  (since it generates cross-axis Pauli components absent in the  $\boldsymbol{\theta}$ -gradient). The condition  $\nabla_{\boldsymbol{\lambda}} \mathcal{L} = 0$  then requires  $\mathbf{r} \perp V$  where  $V = \operatorname{range}(J_{\theta}) + \operatorname{span}\left\{\frac{\partial \mathbf{c}}{\partial \lambda_{k,m}}\right\}$  strictly contains  $\operatorname{range}(J_{\theta})$ .

Since  $\boldsymbol{\theta}^*$  is a local minimum, we have  $\mathbf{r} \perp \operatorname{range}(J_{\theta})$ . The additional condition  $\mathbf{r} \perp \operatorname{span}\left\{\frac{\partial \mathbf{c}}{\partial \lambda_{k,m}}\right\}$  constrains  $\mathbf{r}$  to a strict subspace of  $\operatorname{range}(J_{\theta})^{\perp}$ . For  $\mathbf{c}^*$  drawn from a continuous distribution,  $\mathbf{r} = \mathbf{c}(\boldsymbol{\theta}^*) - \mathbf{c}^*$  lies in this subspace with Lebesgue measure zero.  $\square$

While this provides a formal motivation for pulse-level optimisation, it also increases the number of trainable parameters and hence the optimisation complexity. The exact pulse-parameter counts used in our experiments are listed in Tab. I. Note that, while pulse parameters smooth the landscape by eliminating specific constrained traps inherent to gate-level compilation, they do not guarantee a trap-free landscape unless the system becomes fully over-parametrised.

#### F. Effect on the FCC and Expressibility

Metrics such as the FCC and expressibility are global properties: they are computed by sampling parameters over their full domain to characterise the overall distribution of accessible unitaries or coefficients.

For the family of ansätze and pulse-parameter ranges considered here, the dominant observed effect is local rather than global. We therefore do not expect, and numerically do not observe, a large change in FCC or expressibility.

The same reasoning indicates why improved trainability need not be accompanied by a larger FCC.

The FCC measures normalised linear correlations between coefficients, whereas pulse parameters can increase the variance of individual coefficients without substantially changing the underlying correlation structure. In particular, since the Pearson correlation normalises by the coefficient variances, unsuppressing previously weak coefficients need not increase the global FCC.

A key point is that pulse parameters do not introduce new frequencies. Rather, they can unsuppress existing frequencies by breaking rigid algebraic cancellations inherited from composite gate decompositions; see Sec. II-C. This improves local trainability while leaving the global frequency support unchanged.

Likewise, the variance bound of Mhiri et al. [26] relates coefficient variance to expressibility through an upper bound. A circuit need not saturate that bound. Pulse parameters can therefore move coefficient variances closer to their admissible upper range without implying any substantial change in the global expressibility of the ansatz family.

### III. RESULTS

In this section we provide numerical results to support our findings from Sec. II. Notably, the pulse shape  $E$  is also considered a hyperparameter in QOC and while usually a DRAG [28] pulse is used in practice, we consider a Gaussian pulse described as  $A \exp\left(-1/2 \left(\frac{t-t_c}{\sigma}\right)^2\right)$ , where  $t_c$  is the central pulse time and  $\boldsymbol{\Pi} := \{A, \sigma\}$ , with  $A$  and  $\sigma$  being the pulse amplitude and pulse width, respectively. While we do not constrain the range of values for the pulse parameters in the scope of this work, we acknowledge that for a physical implementation, the amplitude would require bounded optimisation  $A < A_{\max}$  to prevent leakage. Furthermore we choose a set of basis gates as  $\hat{R}_X, \hat{R}_Y, \hat{R}_Z$  and  $C\hat{Z}$ . We refer to Ref. [11] for more details regarding the pulse level simulation implementation. All numerical simulations in this section represent closed-system dynamics without noise or decoherence, to isolate the theoretical impact of the optimisation landscape. Finally, we apply the RWA for all numerical experiments presented in this section; supplementary simulations utilizing the exact interaction-picture dynamics produced nearly identical results.

In the following experiments, we consider a model as defined in Sec. II-A with 3 qubits and a two trainable unitaries. For the encoding we choose a ternary feature map [31] where  $x$  is embedded through Pauli rotations defined as:

$$\hat{S}(x) = \bigotimes_{m=1}^n S(x3^m).$$

where  $n$  depicts the number of qubits. This feature map is chosen as it implements an exponential feature map with a gap free spectrum.

#### A. Training

In this section we present the numerical results for the training of QFMs.

We train a QFM using the Adam optimiser (with analytical gradients) on the MSE loss against a Fourier series dataset. This dataset is generated using the exact same frequencies that the model theoretically represents, with coefficients sampled randomly within the unit disk. Discrete datapoints are sampled according to the Nyquist sampling theorem [36]. The results of this experiment are shown in Fig. 2 for all of the ansätze considered in this work (see Ref. [39] for implementation details), and sorted by their number of pulse parameters. We show the MSE for both the purely gate-based training, the case where we enable access to the pulse parameters and a third scenario in which we decompose all gates into basis gates and add additional trainable parameters to each decomposed gate. This decomposed circuit thus has almost the same number of trainable parameters as in case of the pure pulse simulation while still being fully gate based. The standard deviation over ten seeds for a combined model and data initialisation is reported via error bars. In addition to the MSE, we also report the distance between the rank of the coefficient Jacobian  $\text{rank}(J_\theta) - \text{rank}(J_{\text{ext}})$  (cf. Eq. 17) between the gate-based and pulse-based simulation.

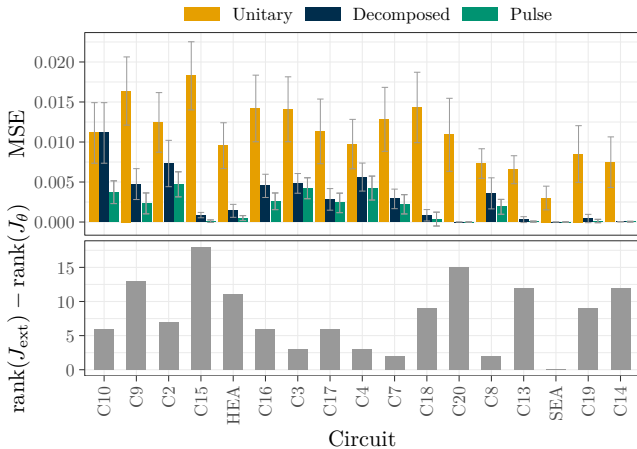


Fig. 2: MSE for all ansätze examined in this work, trained either only using unitary gate parameters (without and with a decomposition into basis gates), or in combination with pulse parameters. Results represent the average over ten different seeds used for both the model and data initialisation. Standard deviation is reported via error bars.

The results confirm our theory from Sec. II that access to the pulse parameters results in a decrease of the MSE for all ansätze. Moreover, the results indicate a correlation between the number of pulse parameters and decrease of the MSE, thus confirm our expectation that composed gates contribute more significantly than basis gates. The reported MSE of the decomposed circuit falls between the two other cases, but is for almost all circuits close to the pulse simulation, suggesting that over-parameterisation is a significant contributor, while still not being on-par with pulse simulation. For the sake of completeness, we provide the full training history in Apx. E.

## B. Coefficient Variance

In this section we present the numerical results for the activation of frequency components due to access to pulse parameters. For this purpose, we instantiate a QFM identical to the one we used in Sec. III-A. To obtain the coefficients of such a circuit we sweep the input  $x$  in a range from 0 to  $2\pi$  and then repeat this process for 4000 different sets of gate parameters. By applying a normal distribution on the scaler of pulse parameters with unit mean, we can observe the effect on coefficients when gradually increasing the variance  $\sigma_\lambda$ . In Fig. 3 we present the results of this experiment, by reporting the number of active frequencies.

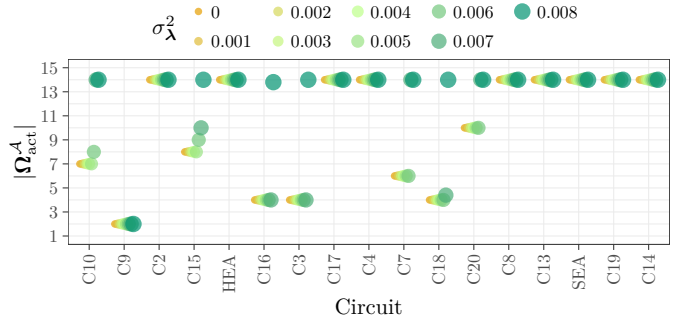


Fig. 3: Count of active frequencies, calculated by the number of high-variance coefficients (see Eq. 3) exceeding a threshold  $\tau > 5 \cdot 10^{-6}$ . Pulse parameter variance is indicated by color.

As the variance of a particular iteration is depicted by color, we can note that different ansätze (sorted by number of pulse parameters) are affected very differently by this distortion. Note that non-integer number of frequencies appear due to the average over 10 different seeds used for the initialisation.

While Fig. 3 demonstrates that access to pulse parameters can increase the variance of coefficients and activates previously suppressed frequencies, Fig. 6 shows that the FCC remains largely invariant. This apparent discrepancy is theoretically consistent as mentioned in Sec. II-F: pulse parameters break exact algebraic cancellations within composite gates (e.g., imperfect cancellation of sub-rotations, cf. Ex. 1), thereby increasing the variance of specific Fourier coefficients. However, the global linear correlations captured by the FCC are governed by the macroscopic circuit topology and the frequency generators  $R(\omega)$ . Because the Pearson correlation normalises variance, the underlying structural dependencies between frequencies remain intact. Consequently, pulse parameters successfully expand the local optimisation manifold (improving trainability) without altering the global structural correlations dictated by the encoding strategy.

**Remark 2.** To build geometric intuition, one can imagine the coefficient manifold as a 3D object and the macroscopic structural correlations (FCC) as its 2D shadow. Introducing pulse parameters adds "texture" to the surface of the object, which increases the variance at the edges of the shadow.

However, because the macroscopic shape of the object (dictated by the circuit topology) remains intact, the statistical correlation between different regions of the shadow remains identical.

#### IV. CONCLUSION & DISCUSSION

In this work we presented theoretical insights as to why pulse parameters break the rigid monomial structure of Fourier coefficients and can enlarge the local tangent space of the coefficient map; equivalently, it can increase the rank of the (real) coefficient Jacobian. This makes an optimisation including the pulse parameters more favourable. Subsequently we showed that these findings are also confirmed through numerical experiments where we trained a QFM on a Fourier series with the same frequencies as the model. Here we observe that access to the pulse parameters consistently improved the training results throughout all ansätze investigated in this work.

While expanding the optimisation space to the pulse-level demonstrates clear theoretical and numerical advantages for trainability, applying these methods to real hardware presents specific trade-offs. On one hand, optimising pulse parameters provides the potential to natively mitigate hardware-specific errors, such as crosstalk or decoherence, by shaping pulses to avoid noise channels which is a capability gate-level compilation lacks. On the other hand, the number of additional parameters increases significantly depending on the specific gates used and their decomposition (*cf.* Tab. I). Thus, a further interesting avenue would be to explore the connection between the effects of noise on QFM [9, 10] and the mitigation of these effects enabled by pulse-level optimisation. As the complexity of training a neural network (and thus a QFM) is linear in the number of parameters [15], the additional complexity added by including pulse parameters is asymptotically negligible for complex models. However, we acknowledge that the increase of parameters requires to consider trade-offs [42] for quantum computing in general and in the noisy intermediatescale quantum (NISQ) era, specifically as overhead required by obtaining gradients through the parameter-shift rule (PSR) [27, 34, 44, 30] is significant.

A common concern when introducing additional parameters to QML models is the exacerbation of BPs, which typically onset when an ansatz becomes too expressive and approaches a 2-design. However, as noted in Sec. II-F, tuning pulse parameters does not fundamentally alter the global expressibility (strict bounds of the unitary volume) dictated by the circuit’s topological structure and encoding strategy [13]. Therefore, extending optimisation to the pulse level is not expected to induce BPs any more than a gate-based architecture, while simultaneously smoothing the local landscape to prevent trapping. Furthermore, while the asymptotic scaling of BPs remains bound by the circuit architecture, the localised gradient magnitudes also depend on the derivative of the pulse-envelope function. However, implementing pulse-level QML on physical QPUs requires extensive low-level access and relies on hardware-in-

the-loop gradient estimation techniques (*e.g.*, pulse-level PSRs), which incur a higher shot-cost overhead. Future work will focus on deploying these models on noisy quantum hardware to empirically evaluate the trade-off between enhanced trainability and operational overhead. While the scaling of training with additional parameters is asymptotically linear, the additional overhead for calculating gradients especially via the PSR is non-trivial. Furthermore, the PSR requires the generator to have two distinct eigenvalues which may not be trivially applicable to arbitrary pulse parameters [21]. In addition to trainability challenges of pulse-level optimisation, we acknowledge that there are physical constraints on the parameters of the pulse (*i.e.* increasing the amplitude  $A$  too much can cause leakage) and various error channels in general that are simplified in our work. While we’re optimistic that these challenges can be overcome in practice, we leave an evaluation inspired by the theoretical and numerical findings shown in this manuscript to future work.

#### ACKNOWLEDGEMENTS

We thank Gabriel Mejía Ruiz for valuable discussions and comments to this work. MS, EK and AS acknowledge support by the state of Baden-Württemberg through bwHPC. LS acknowledges the support by the Doctoral School Karlsruhe School of Elementary and Astroparticle Physics: Science and Technology. MF and WM acknowledge partial support by the German Federal Ministry of Research, Technology and Space (BMFTR), funding program Research Program Quantum Systems, grant number 13N17387, and by the German Research Foundation, grant MA 9739/1-1. WM acknowledges support by the High-Tech Agenda of the Free State of Bavaria.

#### APPENDIX A BASIS GATE SETS

In practice, quantum hardware only supports a limited set of natively executable operations, known as the basis gate set. Any quantum algorithm must be transpiled into a sequence of these basis gates. In our experiments, we consider the basis gate set  $\{\text{CZ, RY, RZ, RX}\}$ . This set satisfies the criteria of universal quantum computation, meaning any unitary operation can be approximated to arbitrary precision. While Tab. I provides an overview of the parameters required for each of the gates utilised in the ansätze of this work, we refer to “Software Between Quantum and Machine Learning - And Down to Pulses” [11] for more details on the implementation.

#### APPENDIX B EXPRESSIBILITY METRIC

The expressibility is evaluated as the KullbackLeibler (KL)-divergence between the fidelity distributions generated by sampling the Haar integral  $\int_{\text{Haar}} d\psi (|\psi\rangle\langle\psi|)^{\otimes t}$  of a state  $t$ -design, a given model  $\int_{\theta} d\theta (|\psi_{\theta}\rangle\langle\psi_{\theta}|)^{\otimes t}$  respectively [37, 18]:

$$D_{\text{KL}}(P_{\text{Model}}(F, \theta) \| P_{\text{Haar}}(F)). \quad (20)$$

Here  $F := |\langle\psi_1|\psi_2\rangle|^2$  is the fidelity defined as the squared overlap of a pair of sampled states  $(|\psi_1\rangle, |\psi_2\rangle)$ , and  $P(F)$

TABLE I: Number of pulse parameters for each gate, including the time parameter. Leaf gates (i.e., hardware-native basis gates requiring no further decomposition) are marked with –.

Gate	$ \mathbf{\Pi} $	Decomposition
$\hat{R}_Z, C\hat{Z}$	1	–
$\hat{R}_X, \hat{R}_Y$	3	–
$\hat{H}$	4	$\hat{R}_Z \cdot \hat{R}_Y$
Rot	5	$\hat{R}_Z \cdot \hat{R}_Y \cdot \hat{R}_Z$
$C\hat{X}$	9	$\hat{H} \cdot C\hat{Z} \cdot \hat{H}$
$C\hat{Y}$	11	$\hat{R}_Z \cdot C\hat{X} \cdot \hat{R}_Z$
$C\hat{R}_Z$	20	$\hat{R}_Z \cdot C\hat{X} \cdot \hat{R}_Z \cdot C\hat{X}$
$C\hat{R}_Y$	24	$\hat{R}_Y \cdot C\hat{X} \cdot \hat{R}_Y \cdot C\hat{X}$
$C\hat{R}_X$	26	$\hat{R}_Z \cdot \hat{R}_Y \cdot C\hat{X} \cdot \hat{R}_Y \cdot C\hat{X} \cdot \hat{R}_Z$

denotes the resulting distribution of fidelities. When the two distributions coincide,  $P_{\text{Model}}(F, \theta) = P_{\text{Haar}}(F)$ , the metric takes the value zero, indicating that the states produced by the quantum Fourier model (QFM) are exactly Haar-distributed.

#### APPENDIX C FOURIER COEFFICIENT CALCULATION

We utilise the definition from Ref. [40] for calculating the Fourier coefficient correlation (FCC) as the average over the correlations of coefficients  $r_{\theta}(\omega, \omega')$  between all frequency pairs parametrised by  $\theta \in \Theta$

$$\text{FCC}_{\Theta} := \frac{1}{|\Omega|} \sum_{\omega, \omega' \in \Omega} |r_{\theta}(\omega, \omega')|. \quad (21)$$

obtained through the Pearson correlation coefficient [12] defined as

$$r_{\theta}(\omega, \omega') = \frac{\sum_{\theta \in \Theta} (c_{\omega}(\theta) - \bar{c}_{\omega})(c_{\omega'}(\theta) - \bar{c}_{\omega'})}{\sqrt{\sum_{\theta \in \Theta} (c_{\omega}(\theta) - \bar{c}_{\omega})^2 \sum_{\theta \in \Theta} (c_{\omega'}(\theta) - \bar{c}_{\omega'})^2}} \quad (22)$$

In the expression above,  $\bar{c}_{\omega}$  and  $\bar{c}_{\omega'}$  are the mean values of the coefficients at frequencies  $\omega$  and  $\omega'$ , respectively. The Pearson coefficient is normalised by the standard deviation of the coefficients at each frequency. When the average correlation across frequency pairs is low, the metric approaches zero; it approaches one for highly correlated ansätze.

#### APPENDIX D FIDELITY AND TRACE DISTANCE

This section highlights the sensitivity of the fidelity and trace distance to the pulse parameter variance. Specifically, we evaluate these two metrics with reference to states produced by gate-level simulation of equivalent circuits. The pulse parameters are then distorted by a scaler which is sampled from a Gaussian distribution of mean 1 and the variance  $\sigma_{\lambda}^2$  being increased incrementally. Fig. 4 shows the result of this evaluation of a pulse parameter variance over 8 steps within a range from 0 to 0.008. It becomes clear that even slight deviations in  $\lambda$  cause the resulting states to diverge rapidly from the target states which is reflected by the fidelity and trace distance going down and up respectively.

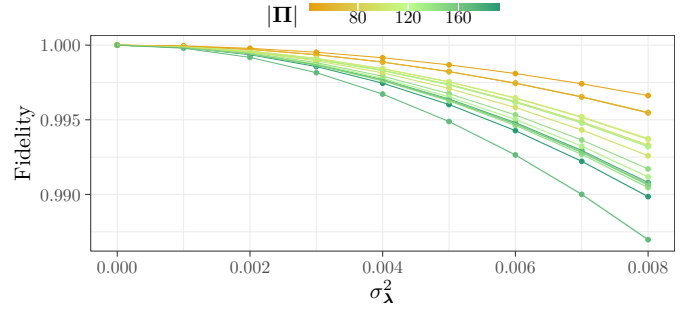


Fig. 4: Fidelity between gate- and pulse-level simulations for different levels of distortion  $\sigma_{\lambda}^2$ , coloured by the number of pulse parameters  $|\mathbf{\Pi}|$ .

#### APPENDIX E ADDITIONAL RESULTS

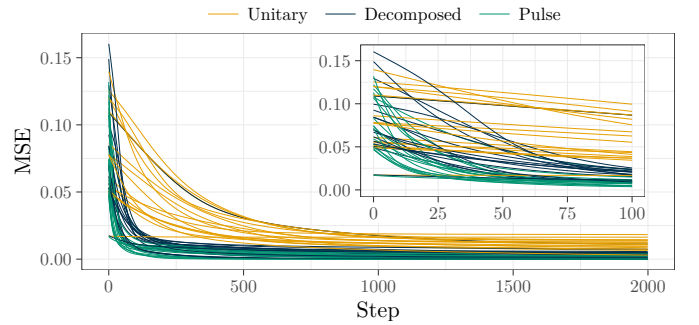


Fig. 5: MSE over the course of training for different ansätze and using either standard or decomposed unitary gates, or pulse gates. The inset shows a larger view of the first 100 steps.

In a similar manner as described in Sec. III-B, we evaluated the FCC and the expressibility metric. Here we've found that both metrics do not change significantly *w.r.t.* the pulse parameter variance, as shown in Fig. 6 for the FCC and the expressibility metric, averaged over ten seeds.

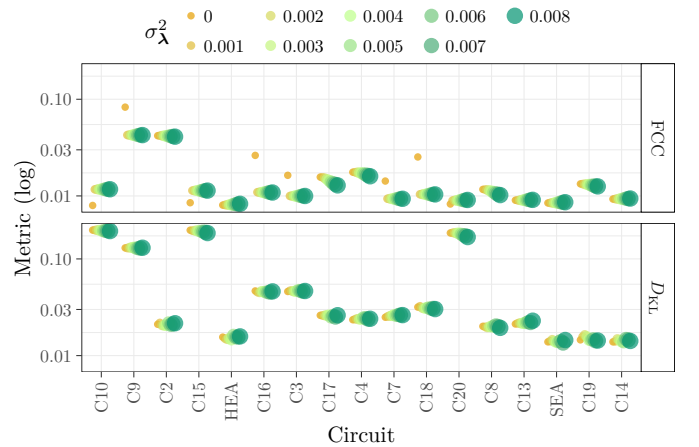


Fig. 6: FCC and KL-divergence  $D_{\text{KL}}$  (Eq. 20, i.e., inverse of the expressibility) over pulse parameter variance  $\sigma_{\lambda}^2$  (coloured) for different ansätze.

## REFERENCES

- [1] Ignacio B. Acedo, Pablo Rodriguez-Grasa, Pablo Garcia-Azorin, et al. *Pulsed Learning for Quantum Data Re-Uploading Models*. Dec. 2025. DOI: [10.48550/arXiv.2512.10670](https://doi.org/10.48550/arXiv.2512.10670).
- [2] George Arfken. *Mathematical Methods for Physicists*. Third. San Diego: Academic Press, Inc., 1985.
- [3] Andreas and Bayerstadler, Guillaume Becquin, Julia Binder, et al. “Industry quantum computing applications”. In: *EPJ Quantum Technology* 8.1 (Nov. 2021). DOI: [10.1140/epjqt/s40507-021-00114-x](https://doi.org/10.1140/epjqt/s40507-021-00114-x).
- [4] Marcello Benedetti, Erika Lloyd, Stefan Sack, et al. “Parameterized quantum circuits as machine learning models”. In: *Quantum science and technology* 4.4 (2019), p. 043001.
- [5] Cecilia Carbonelli, Michael Felderer, Matthias Jung, et al. “Challenges for Quantum Software Engineering: An Industrial Application Scenario Perspective”. In: *Quantum Software: Aspects of Theory and System Design*. Ed. by Iaakov Exman, Ricardo Perez-Castillo, Mario Piattini, et al. Springer Nature, 2024, pp. 311–335. DOI: [10.1007/978-3-031-64136-7\\_12](https://doi.org/10.1007/978-3-031-64136-7_12).
- [6] M. Cerezo, Akira Sone, Tyler Volkoff, et al. “Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits”. In: *Nature Communications* 12.1 (Mar. 19, 2021), p. 1791. DOI: [10.1038/s41467-021-21728-w](https://doi.org/10.1038/s41467-021-21728-w).
- [7] Callum Duffy and Marcin Jastrzebski. “Spectral Bias in Variational Quantum Machine Learning”. In: *arXiv preprint arXiv:2506.22555* (2025).
- [8] Stavros Efthymiou, Sergi Ramos-Calderer, Carlos Bravo-Prieto, et al. “Qibo: A Framework for Quantum Simulation with Hardware Acceleration”. In: *Quantum Science and Technology* 7.1 (Jan. 2022), p. 015018. DOI: [10.1088/2058-9565/ac39f5](https://doi.org/10.1088/2058-9565/ac39f5).
- [9] Enrico Fontana, Ivan Rungger, Ross Duncan, et al. *Spectral analysis for noise diagnostics and filter-based digital error mitigation*. Tech. rep. arXiv, Nov. 2022. DOI: [10.48550/arXiv.2206.08811](https://doi.org/10.48550/arXiv.2206.08811).
- [10] Maja Franz, Melvin Strobl, Leonid Chaichenets, et al. *Out of Tune: Demystifying Noise-Effects on Quantum Fourier Models*. 2025. DOI: [10.48550/arxiv.2506.09527](https://doi.org/10.48550/arxiv.2506.09527).
- [11] Maja Franz, Melvin Strobl, Jonathan Hunz, Lucas van der Horst, Lukas Scheller, Eileen Kuehn, Achim Streit, and Wolfgang Mauerer. “Software Between Quantum and Machine Learning - And Down to Pulses”. In: *To be published* (2026).
- [12] David Freedman, Robert Pisani, and Roger Purves. “Statistics (international student edition)”. In: *Pisani, R. Purves, 4th edn. WW Norton & Company, New York* (2007).
- [13] Martin Gogeissl, Hila Safi, and Wolfgang 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*. 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).
- [14] Felix Greiwe, Tom Krüger, and Wolfgang Mauerer. “Effects of Imperfections on Quantum Algorithms: A Software Engineering Perspective”. In: *IEEE International Conference on Quantum Software (QSW)*. IEEE, 2023, pp. 31–42. DOI: [10.1109/QSW59989.2023.00014](https://doi.org/10.1109/QSW59989.2023.00014).
- [15] Jeff Heaton. “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep Learning”. In: *Genetic Programming and Evolvable Machines* 19.1 (June 2018), pp. 305–307. DOI: [10.1007/s10710-017-9314-z](https://doi.org/10.1007/s10710-017-9314-z).
- [16] Ben Jaderberg, Antonio A. Gentile, Youssef Achari Berrada, et al. *Let Quantum Neural Networks Choose Their Own Frequencies*. Tech. rep. 4. arXiv, Sept. 2023, p. 042421. DOI: [10.48550/arXiv.2309.03279](https://doi.org/10.48550/arXiv.2309.03279).
- [17] P. Krantz, M. Kjaergaard, F. Yan, et al. “A quantum engineers guide to superconducting qubits”. In: *Applied Physics Reviews* 6.2 (June 2019). DOI: [10.1063/1.5089550](https://doi.org/10.1063/1.5089550).
- [18] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951). Publisher: Institute of Mathematical Statistics, pp. 79–86. DOI: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- [19] Jonas Landman, Slimane Thabet, Constantin Dalyac, et al. *Classically Approximating Variational Quantum Machine Learning with Random Fourier Features*. Tech. rep. arXiv, Oct. 2022. DOI: [10.48550/arXiv.2210.13200](https://doi.org/10.48550/arXiv.2210.13200).
- [20] Martin Larocca, Nathan Ju, Diego García-Martín, et al. “Theory of Overparametrization in Quantum Neural Networks”. In: *Nature Computational Science* 3.6 (June 2023), pp. 542–551. DOI: [10.1038/s43588-023-00467-6](https://doi.org/10.1038/s43588-023-00467-6).
- [21] Jiaqi Leng, Yuxiang Peng, Yi-Ling Qiao, et al. *Differentiable Analog Quantum Computing for Optimization and Control*. Oct. 2022. DOI: [10.48550/arXiv.2210.15812](https://doi.org/10.48550/arXiv.2210.15812).
- [22] Zhiding Liang, Jinglei Cheng, Hang Ren, et al. *NAPA: Intermediate-level Variational Native-pulse Ansatz for Variational Quantum Algorithms*. Jan. 2024. DOI: [10.48550/arXiv.2208.01215](https://doi.org/10.48550/arXiv.2208.01215).
- [23] Alicia B. Magann, Christian Arenz, Matthew D. Grace, et al. “From Pulses to Circuits and Back Again: A Quantum Optimal Control Perspective on Variational Quantum Algorithms”. In: *PRX Quantum* 2.1 (Jan. 2021), p. 010101. DOI: [10.1103/PRXQuantum.2.010101](https://doi.org/10.1103/PRXQuantum.2.010101).
- [24] Stefan Raimund Maschek, Jürgen Schwittalla, Maja Franz, et al. “Make Some Noise! Measuring Noise Model Quality in Real-World Quantum Software”. In: *Proceedings of the IEEE International Conference on Quantum Software (QSW)*. IEEE, 2025, pp. 1–11. DOI: [10.1109/QSW67625.2025.00010](https://doi.org/10.1109/QSW67625.2025.00010).
- [25] Wolfgang Mauerer and Stefanie Scherzinger. “1-2-3 Reproducibility for Quantum Software Experiments”. In: *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).

- [26] Hela Mhiri, Leo Monbroussou, Mario Herrero-Gonzalez, et al. *Constrained and Vanishing Expressivity of Quantum Fourier Models*. Mar. 2024. DOI: [10.48550/arXiv.2403.09417](https://doi.org/10.48550/arXiv.2403.09417).
- [27] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, et al. “Quantum circuit learning”. In: *Physical Review A* 98.3 (2018), p. 032309.
- [28] F. Motzoi, J. M. Gambetta, P. Rebentrost, et al. “Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits”. In: *Physical Review Letters* 103.11 (Sept. 2009), p. 110501. DOI: [10.1103/PhysRevLett.103.110501](https://doi.org/10.1103/PhysRevLett.103.110501).
- [29] Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, et al. “Data re-uploading for a universal quantum classifier”. In: *Quantum* 4 (Feb. 2020), p. 226. DOI: [10.22331/q-2020-02-06-226](https://doi.org/10.22331/q-2020-02-06-226).
- [30] Maniraman Periyasamy, Axel Plinge, Christopher Mutschler, et al. “Guided-SPSA: Simultaneous Perturbation Stochastic Approximation Assisted by the Parameter Shift Rule”. In: *IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2024, pp. 1504–1515. DOI: [10.1109/QCE60285.2024.00177](https://doi.org/10.1109/QCE60285.2024.00177).
- [31] Evan Peters and Maria Schuld. *Generalization despite overfitting in quantum machine learning models*. Tech. rep. arXiv, Sept. 2022. DOI: [10.48550/arXiv.2209.05523](https://doi.org/10.48550/arXiv.2209.05523).
- [32] Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, et al. “A Lie Algebraic Theory of Barren Plateaus for Deep Parameterized Quantum Circuits”. In: *Nature Communications* 15.1 (Aug. 2024), p. 7172. DOI: [10.1038/s41467-024-49909-3](https://doi.org/10.1038/s41467-024-49909-3).
- [33] Tilmann Rothe Santos. “Evaluation of Pulse Level Quantum Fourier Models”. 46.21.02; LK 01. Abschlussarbeit - Bachelor. Karlsruher Institut für Technologie (KIT), 2025. DOI: [10.5445/IR/1000184129](https://doi.org/10.5445/IR/1000184129).
- [34] Maria Schuld, Ville Bergholm, Christian Gogolin, et al. “Evaluating Analytic Gradients on Quantum Hardware”. In: *Physical Review A* 99.3 (Mar. 21, 2019), p. 032331. DOI: [10.1103/PhysRevA.99.032331](https://doi.org/10.1103/PhysRevA.99.032331).
- [35] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. “The effect of data encoding on the expressive power of variational quantum machine learning models”. In: *Physical Review A* 103.3 (Mar. 2021), p. 032430. DOI: [10.1103/PhysRevA.103.032430](https://doi.org/10.1103/PhysRevA.103.032430).
- [36] C.E. Shannon. “Communication in the Presence of Noise”. In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21. DOI: [10.1109/JRPROC.1949.232969](https://doi.org/10.1109/JRPROC.1949.232969).
- [37] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. “Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms”. en. In: *Advanced Quantum Technologies* 2.12 (2019), p. 1900070. DOI: [10.1002/qute.201900070](https://doi.org/10.1002/qute.201900070).
- [38] Melvin Strobl and Maja Franz. *cirKITers/pulse-level-quantum-fourier-models: Initial Release*. Version v1.0.0. Apr. 2026. DOI: [10.5281/zenodo.19847940](https://doi.org/10.5281/zenodo.19847940).
- [39] Melvin Strobl, Maja Franz, Eileen 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).
- [40] Melvin Strobl, M. Emre Sahin, Lucas van der Horst, et al. *Fourier Fingerprints of Ansatzes in Quantum Machine Learning*. 2025. DOI: [10.48550/arxiv.2508.20868](https://doi.org/10.48550/arxiv.2508.20868).
- [41] Ryan Sweke, Erik Recio-Armengol, Sofiene Jerbi, et al. “Potential and limitations of random Fourier features for dequantizing quantum machine learning”. In: *Quantum* 9 (Feb. 2025), p. 1640. DOI: [10.22331/q-2025-02-20-1640](https://doi.org/10.22331/q-2025-02-20-1640).
- [42] Simon Thelen and Wolfgang Mauerer. “Predict and Conquer: Navigating Algorithm Trade-Offs with Quantum Design Automation”. In: *IEEE International Conference on Quantum Computing and Engineering (QCE)*. Los Alamitos, CA, USA: IEEE Computer Society, 2025, pp. 591–602. DOI: [10.1109/QCE65121.2025.00071](https://doi.org/10.1109/QCE65121.2025.00071).
- [43] Marco Wiedmann, Maniraman Periyasamy, and Daniel D. Scherer. *Fourier Analysis of Variational Quantum Circuits for Supervised Learning*. Nov. 2024. DOI: [10.48550/arXiv.2411.03450](https://doi.org/10.48550/arXiv.2411.03450).
- [44] David Wierichs, Josh Izaac, Cody Wang, et al. “General Parameter-Shift Rules for Quantum Gradients”. In: *Quantum* 6 (Mar. 2022), p. 677. DOI: [10.22331/q-2022-03-30-677](https://doi.org/10.22331/q-2022-03-30-677).
- [45] Tao Yue, Wolfgang Mauerer, Shaukat Ali, et al. “Challenges and Opportunities in Quantum Software Architecture”. In: *Software Architecture: Research Roadmaps from the Community*. Ed. by Patrizio Pelliccione, Rick Kazman, Ingo Weber, et al. Cham: Springer Nature Switzerland, 2023, pp. 1–23. DOI: [10.1007/978-3-031-36847-9\\_1](https://doi.org/10.1007/978-3-031-36847-9_1).