

Design of expressive and trainable pulse-based quantum machine learning models

Han-Xiao Tao , Xin Wang , and Re-Bing Wu 

Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing 100084, China



(Received 10 November 2025; accepted 9 February 2026; published 30 March 2026)

Pulse-based quantum machine learning (QML) has emerged as a promising paradigm in quantum artificial intelligence due to its exceptional hardware efficiency. For real-world applications, it is crucial that pulse-based models are both expressive and trainable. Previous studies have shown that pulse-based models operating under dynamical symmetry can be effectively trained, benefiting from a favorable loss landscape devoid of barren plateaus. However, this approach may lead to uncontrollability, which can limit expressivity if the model is poorly designed. In this paper, we explore the conditions necessary for pulse-based QML models to maintain both expressivity and trainability. We present a key requirement involving the system's initial state, the measurement observable, and the underlying dynamical symmetry Lie algebra, supported by numerical simulations. Our results provide a comprehensive framework for the design of practical pulse-based QML models that strike a good balance between expressivity and trainability.

DOI: [10.1103/gvj4-kqwd](https://doi.org/10.1103/gvj4-kqwd)

I. INTRODUCTION

Quantum computing holds substantial promise for improving the efficiency of both training and inference in machine learning tasks [1–4]. However, in the noisy intermediate-scale quantum (NISQ) era, the lack of error correction necessitates that quantum machine learning (QML) works in conjunction with classical computers [5,6] through variational quantum models. In this framework, quantum models execute the inference task, while the classical computer manages the training.

In the literature, variational QML models predominantly employ gate-based parametrized quantum circuits, also known as quantum neural networks (QNNs) [7,8]. Since all gates are ultimately implemented through control pulses on the relevant physical qubits, recent proposals suggest that QML models may be parametrized directly by these continuous-time control pulses instead of discrete-time gate parameters [9–16]. In contrast to gate-based models, pulse-based models are grounded in physical systems without abstraction, making them more accessible for hardware-efficient implementation [17–24].

Within the finite coherence time, pulse-based QML models are in principle more expressive than their gate-based counterparts because they function as *infinitely* deep data-reuploading QNNs [25]. This advantage has been demonstrated in our previous study [25], where we also proved that pulse-based models achieve full expressivity (i.e., the ability to approximate arbitrary functions [26–29]) when the underlying

quantum system is ensemble controllable. This criterion provides a control-oriented framework for designing pulse-based models, analogous to the design of universal quantum computers [30]. Nonetheless, in systems with a large number of qubits, the requirement for controllability becomes excessively stringent as it inevitably leads to barren plateaus within the loss landscape [31,32]. This phenomenon causes the loss gradients (or, equivalently, the loss variance [33,34]) to decay exponentially in the asymptotic time limit [35] and in the limit of many qubits, thereby making optimization impractical for large-scale systems [8,36–41]. Such trade-off between expressivity and trainability has been observed in gate-based models, where the loss function concentrates sharply as the circuit becomes more expressive [37,42,43].

To alleviate the conflict between expressivity and trainability, controllability must be sacrificed in the design of pulse-based models. Larocca *et al.* [31] reveal that trainability is contingent on the degree of controllability characterized by the size and symmetry of its dynamical Lie algebra. The findings indicate that a symmetry-restricted model has the potential to circumvent barren plateaus, thus providing a practical diagnostic approach for the design of QML models. Building upon this insight, a Lie-algebraic theory for barren plateaus has been developed, based on which an exact variance formula is derived for the loss function. It is demonstrated that the variance scales inversely with the dimension of the dynamical Lie algebra, provided that either the initial state or observable is contained within the Lie algebra [44,45]. Diaz *et al.* [46] extend beyond the Lie-algebraic framework by analyzing parametrized matchgate circuits. By decomposing the operator space into Lie group modules, they derive an exact variance formula for loss functions applicable to arbitrary initial states and observables, demonstrating that barren plateaus can emerge from generalized global operators and are dependent on module dimensions rather than solely on the dimension of the dynamical Lie algebra.

*Contact author: rbwu@tsinghua.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

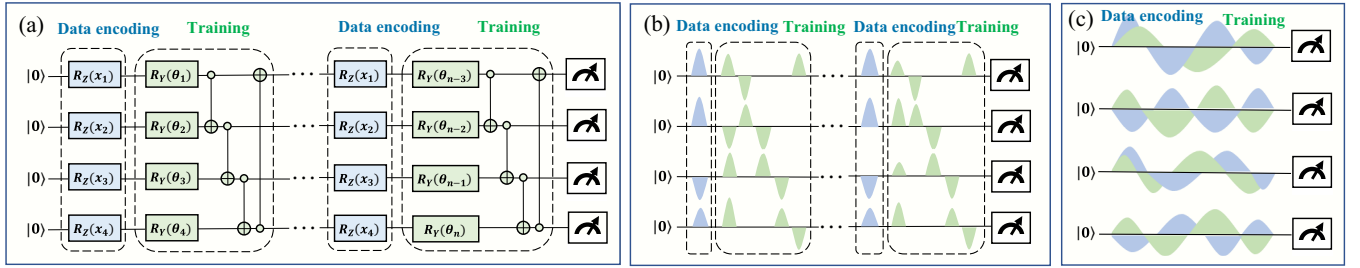


FIG. 1. Schematics of QML models that infer the learning output via a parametrized quantum transformation and subsequent measurement. The model is trained by an external classical optimizer that is not depicted herein for brevity. (a) A four-qubit gate-based data-reuploading QNN model, (b) the intermediate-level pulse-based model transcribed from the circuit [53–56], and (c) the generalized pulse-based model. In principle, the pulse-based model can be continuously and simultaneously applied without waiting for one another, allowing for compression to reduce inference time.

The existing works provide a Lie-algebraic framework for evaluating the trainability of uncontrollable pulse-based QML models induced by dynamical symmetry, but the expressivity of these models remains ambiguous. In the present study, we demonstrate that an ill-suited selection of either the initial state or the measurement observable can result in the loss of expressivity, and conduct a systematic analysis on the expressivity of pulse-based QML models within the context of dynamical symmetry. In contrast to the Fourier-series analysis typically used to assess the expressivity of data-reuploading gate-based models [47–51], we utilize the polynomial expansion of nonlinear functions epitomized by the pulse-based model based on Dyson series analysis (also known as Fliess series in control theory for input-output analysis of nonlinear systems [52]). This analysis yields a Lie-algebraic tool for the design of expressive pulse-based models with dynamical symmetry.

The structure of the paper is as follows: Section II introduces essential background on pulse-based QML models and the existing results on the trainability under dynamical symmetry. Section III analyzes the expressivity of pulse-based models through the Dyson-series expansion from which a necessary condition is derived. In Sec. IV, numerical experiments are conducted to demonstrate the expressivity of pulse-based models with dynamical symmetry, and the balance between expressivity and trainability. Finally, Sec. V draws the conclusion and future perspectives.

II. PULSE-BASED QML MODELS WITH DYNAMICAL SYMMETRY

We first illustrate the potential advantages of pulse-based QML models through a four-qubit system. In the gate-based data-reuploading model shown in Fig. 1(a), the input $\mathbf{x} = (x_1, x_2, x_3, x_4)$ is repeatedly uploaded using the encoding circuits $U(\mathbf{x}) = R_z(x_1) \oplus \dots \oplus R_z(x_4)$. The trainable circuit blocks comprise $R_y(\theta_k)$ ($k = 1, \dots, n$) rotations about the y axis, interspersed with the data-uploading circuit blocks. The final learning outcome is inferred by the expectation-value measurement

$$f(\mathbf{x}, \Theta) = \langle \psi(\mathbf{x}, \Theta) | M | \psi(\mathbf{x}, \Theta) \rangle \quad (1)$$

of a specific observable M , where $|\psi(\mathbf{x}, \Theta)\rangle$ is the output state of the circuit, to approximate multivariable functions by optimizing the parameters $\Theta = \{\theta_1, \dots, \theta_n\}$.

Figure 1(b) illustrates that the physical implementation of the quantum circuit model consistently compiles individual quantum gates into time-varying control pulses on the physical qubits. The QML process dictated by these control pulses can be taken as a pulse-based model implemented at the physical level. Furthermore, as shown in Fig. 1(c), the pulses can be continuously and simultaneously applied without waiting for one another, allowing for compression to reduce inference time. Such generalized pulse-based model is free from a prior circuit design, featuring more tunable parameters within the finite coherence time, and can achieve greater expressive power than gate-based models [25].

The pulse-based model implemented with an n -qubit quantum device can be characterized as a controlled quantum system:

$$i\dot{\psi}(t; \mathbf{x}) = -[H_E(t, \mathbf{x}) + H_C(t)]|\psi(t; \mathbf{x})\rangle, \quad (2)$$

in which the encoding Hamiltonian

$$H_E(t, \mathbf{x}) = \sum_{k=1}^m x_k \theta_k(t) H_k \quad (3)$$

encodes the input variable $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{X} \subseteq \mathbb{R}^m$ into the dynamical evolution of $|\psi(t; \mathbf{x})\rangle \in \mathbb{C}^{2^n}$. The control pulses $\theta_k(t)$ are either tunable or, as is typically done in angle-encoding schemes, fixed to a constant value. The control Hamiltonian

$$H_C(t) = \sum_{k=m+1}^{m+p} \theta_k(t) H_k \quad (4)$$

is introduced to enhance the expressivity of the pulse-based model. The time-variant functions $\theta_1(t), \dots, \theta_{m+p}(t)$ are control pulses that can be trained to fit the dataset. Similar to gate-based models, the output is inferred through the expectation-value measurement

$$f(\mathbf{x}, \Theta) = \langle \psi(T; \mathbf{x}) | M | \psi(T; \mathbf{x}) \rangle, \quad (5)$$

with $\Theta = \{\theta_1(t), \dots, \theta_{m+p}(t), t \in [0, T]\}$. In practice, certain control pulses may be constant or exhibit limited tunability due to hardware constraints.

Let $\mathfrak{g} = \{iH_1, \dots, iH_{m+p}\}_{LA}$ be the Lie algebra generated by data-encoding and control Hamiltonians in Eq. (2), and

$$\mathfrak{g} = \mathfrak{c} \oplus \mathfrak{g}_1 \oplus \dots \oplus \mathfrak{g}_k$$

be its decomposition as a direct sum of commuting ideals, where $\mathfrak{c}$ is the center of $\mathfrak{g}$ and $\mathfrak{g}_1 \oplus \dots \oplus \mathfrak{g}_k$ is the semisimple part. Assume that either the initial state $\rho = |\psi_0\rangle\langle\psi_0|$ or the observable M belongs to the Lie algebra $\mathfrak{g}$ (a relaxed condition can be found in Ref. [46]), and the QNN is sufficiently deep (i.e., a sufficiently long time duration T for pulse-based QNN) to establish a 2-design over the Lie group $e^{\mathfrak{g}}$ [44,45]. The variance of $f(\mathbf{x}, \Theta)$ can be estimated as follows [32,44]:

$$\text{Var}[f(\mathbf{x}, \Theta)] = \sum_{j=1}^k \frac{P_{\mathfrak{g}_j}(\rho)P_{\mathfrak{g}_j}(M)}{\dim(\mathfrak{g}_j)}, \quad (6)$$

where $\dim(\mathfrak{g})$ denotes the dimension of the Lie algebra $\mathfrak{g}$, and

$$P_{\mathfrak{g}}(H) \stackrel{\text{def}}{=} \sum_{j=1}^{\dim(\mathfrak{g})} |\text{Tr}(B_j^\dagger H)|^2$$

denotes the square norm of the orthogonal projection (with respect to the Hilbert-Schmidt inner product $\langle A, B \rangle \stackrel{\text{def}}{=} \text{Tr}(A^\dagger B)$) of H onto $\mathfrak{g}_{\mathbb{C}}$ (the complexification of $\mathfrak{g}$). $\{B_j\}_{j=1}^{\dim(\mathfrak{g})}$ is an orthonormal basis for $\mathfrak{g}$.

Equation (6) indicates that the model becomes untrainable when the underlying control system is fully controllable, because in such case $\dim(\mathfrak{g}) = 4^n$ with $\mathfrak{g} = \mathfrak{su}(2^n)$, yielding exponentially decreasing variance as n increases. Therefore, nontrivial dynamical symmetry must be introduced to constrain the quantum state evolution to a manifold whose dimension does not increase exponentially with n .

III. EXPRESSIVITY OF QML MODELS IN THE PRESENCE OF DYNAMICAL SYMMETRY

The pulse-based model (2) is said to be expressive if, for any function $f_0(\mathbf{x})$ and a specified error threshold $\epsilon > 0$, there exists a suitable control pulse Θ and a scaling factor θ_0 such that the measurement-induced function (5) can approximate the function with the desired precision, i.e., $\|\theta_0 f(\mathbf{x}, \Theta) - f_0(\mathbf{x})\| < \epsilon$ for all $\mathbf{x}$.

In a previous study [25], we demonstrated that a pulse-based model is expressive when the underlying system is ensemble controllable (i.e., any $|\psi(T; \mathbf{x})\rangle$ as a function of $\mathbf{x}$ can be realized within a sufficiently long time T). This condition guarantees expressivity for arbitrary selections of the initial state $|\psi_0\rangle$ and the measurement observable M . However, as previously noted, it is overly stringent for trainability.

In the following, we demonstrate that such uncontrollable pulse-based models can be expressive when the initial state and measurement observable are appropriately selected. To derive the required condition, we expand the output function (5) into a polynomial series of $\mathbf{x}$.

To facilitate the following derivation and for more generality, we rewrite Eq. (2) as a Liouville equation:

$$\dot{\rho}(t) = \mathcal{L}(t)\rho(t), \quad \rho(0) = |\psi_0\rangle\langle\psi_0|, \quad (7)$$

where $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$ is the density matrix of the system, and

$$\mathcal{L}(t) = \sum_{j=1}^m x_j \theta_j(t) \mathcal{L}_j + \sum_{j=m+1}^{m+p} \theta_j(t) \mathcal{L}_j, \quad (8)$$

with the Liouvillian defined as $\mathcal{L}_j X = -i[H_j, X]$ for $j = 1, \dots, m+p$.

Then, we apply the Dyson-series expansion to the solution of the Liouville equation, yielding

$$\begin{aligned} \rho(t) = & \rho(0) + \int_0^t \mathcal{L}(t_1)\rho(0)dt_1 \\ & + \int_0^t \mathcal{L}(t_1)dt_1 \int_0^{t_1} \mathcal{L}(t_2)\rho(0)dt_2 + \dots \end{aligned}$$

Expanding the series into polynomial terms of $x_1, \dots, x_m$, we obtain the following polynomial expansion:

$$\begin{aligned} f(\mathbf{x}, \Theta) = & \text{tr}[\rho(T)M] \\ = & \sum_{k_1, \dots, k_m \geq 0} C_{k_1, \dots, k_m}(\Theta) x_1^{k_1} \dots x_m^{k_m}, \end{aligned}$$

where

$$\begin{aligned} C_{k_1, \dots, k_m}(\Theta) = & \sum_{n=k_1+\dots+k_m}^{\infty} \sum_{(j_1, \dots, j_n) \in \mathbb{J}_{k_1, \dots, k_m}} \\ & \times c_{j_1 \dots j_n}(\Theta) \langle \psi_0 | \mathcal{L}_{j_1} \dots \mathcal{L}_{j_n} M | \psi_0 \rangle. \end{aligned} \quad (9)$$

The index set $\mathbb{J}_{k_1, \dots, k_m}$ comprises tuples $(j_1, \dots, j_n)$ ($n \geq k_1 + \dots + k_m$) wherein each index j in $\{1, \dots, m\}$ appears k_j times. The coefficients $c_{j_1 \dots j_n}(\Theta)$ are determined by the control pulses as follows:

$$\begin{aligned} c_{j_1 \dots j_n}(\Theta) = & \int_0^T \theta_{j_1}(t_1) dt_1 \int_0^{t_1} \theta_{j_2}(t_2) dt_2 \\ & \dots \int_0^{t_{n-1}} \theta_{j_n}(t_n) dt_n. \end{aligned} \quad (10)$$

The polynomial expansion clearly demonstrates the inherent nonlinearity in the model induced by a continuous-limit data-reuploading mechanism [25]. It is evident that the pulse-based model can approximate any continuous function if the coefficients $\{C_{k_1, \dots, k_m}(\Theta)\}$ in the polynomial series are fully adjustable by tuning the control pulses $\theta_1(t), \dots, \theta_{m+p}(t)$. According to Eq. (9), the adjustability of these coefficients depends on both the control pulses, through $c_{j_1 \dots j_n}(\Theta)$ terms, and the initial state and observable through $\langle \psi_0 | \mathcal{L}_{j_1} \dots \mathcal{L}_{j_n} M | \psi_0 \rangle$ terms.

Observing that the coefficient $C_{k_1, \dots, k_m}(\Theta)$ corresponding to the monomial $x_1^{k_1} \dots x_m^{k_m}$ vanishes when $\langle \psi_0 | \mathcal{L}_{j_1} \dots \mathcal{L}_{j_n} M | \psi_0 \rangle = 0$ holds for all $(j_1, \dots, j_n) \in \mathbb{J}_{k_1, \dots, k_m}$. If such circumstance occurs for some $(k_1, \dots, k_m)$, the model cannot approximate functions containing the monomial $x_1^{k_1} \dots x_m^{k_m}$ in their polynomial expansions, and thus loses expressivity.

If such circumstance does not occur for all $(k_1, \dots, k_m)$, the expressivity of the model further depends on the adjustability of the coefficients $c_{j_1 \dots j_n}(\Theta)$. Apparently, if they are mutually independent and fully adjustable, the model will be expressive because all $C_{k_1, \dots, k_m}(\Theta)$ can be freely tuned. Unfortunately, the

mutual independence of $c_{j_1 \dots j_n}(\Theta)$ does not hold, as it can be negated by the counterexample

$$c_{j_1 j_2}(\Theta) + c_{j_2 j_1}(\Theta) = c_{j_1}(\Theta) c_{j_2}(\Theta), \quad \forall j_1, j_2.$$

Nevertheless, the above analysis still provides a necessary condition summarized in the following theorem.

Theorem 1. Given a sufficiently long duration T , the pulse-based QML model (2) can approximate any function $f: \mathcal{X} \rightarrow \mathbb{R}$ only if there does not exist $k_1, \dots, k_m \geq 0$ such that $\langle \psi_0 | \mathcal{L}_{j_1} \dots \mathcal{L}_{j_n} M | \psi_0 \rangle = 0$ for all $(j_1, \dots, j_n) \in \mathbb{J}_{k_1, \dots, k_m}$.

Theorem 1 provides a necessary condition for the expressivity of pulse-based QML models under dynamical symmetry, which ensures that no monomial term in the Dyson expansion is structurally vanishing due to improper choice of initial state or observable.

Although the condition involves an infinite set of multi-indices, it can be verified in finitely many steps due to the finite-dimensional nature of the operator space. Specifically, the operators $\mathcal{L}_j$ act on the space of linear operators over the n -qubit Hilbert space, which is finite-dimensional. For each monomial $x_1^{k_1} \dots x_m^{k_m}$, the associated subspace

$$\mathcal{S}_{k_1, \dots, k_m} = \{ \mathcal{L}_{j_1} \dots \mathcal{L}_{j_n} M \mid (j_1, \dots, j_n) \in \mathbb{J}_{k_1, \dots, k_m} \}$$

is contained within this finite-dimensional space. As the total degree $K = \sum_i k_i$ increases, the recursive construction of these subspaces can generate only finitely many distinct subspaces. Consequently, the sequence eventually becomes periodic: There exists a finite $K_{\max}$ such that for all $K \geq K_{\max}$, each $\mathcal{S}_{k_1, \dots, k_m}$ with total degree K is contained in (or equal to) a subspace already encountered for some smaller total degree. This periodicity reduces the verification of Theorem 1's condition to checking a finite set of multi-indices up to $K_{\max}$.

To facilitate practical verification, we introduce a recursive procedure for constructing the relevant subspaces. Let

$$\mathfrak{M}(\mathcal{S}) = \text{span} \{ \mathcal{L}_{\alpha_1} \dots \mathcal{L}_{\alpha_\ell} X \mid X \in \mathcal{S}, \\ m+1 \leq \alpha_i \leq m+p, \ell \in \mathbb{N} \}$$

denote the submodule generated from an operator set $\mathcal{S}$ using the control Liouvillians. Starting from $\mathcal{S}_{0, \dots, 0} = \mathfrak{M}(M)$, higher-indexed subspaces can be built recursively via

$$\mathcal{S}_{k_1, \dots, k_i+1, \dots, k_m} = \mathfrak{M}(\mathcal{L}_i \mathcal{S}_{k_1, \dots, k_i, \dots, k_m}).$$

In practice, this recursive procedure can be implemented numerically. Beginning with $\mathcal{S}_{0, \dots, 0}$, we increment the indices k_i until no new subspaces are generated. It suffices to verify Theorem 1's condition for all $\mathcal{S}_{k_1, \dots, k_m}$ generated up to that point. A detailed description of this verification algorithm is provided in Algorithm 1. This algorithmic approach makes the condition of Theorem 1 practically verifiable for any given pulse-based model with dynamical symmetry.

IV. NUMERICAL SIMULATIONS

In this section, we demonstrate through numerical simulations the expressivity and trainability of pulse-based QML models under selected dynamic symmetries.

ALGORITHM 1. Expressivity verification.

Input:

- Encoding Hamiltonians: $\{H_1, \dots, H_m\}$
- Control Hamiltonians: $\{H_{m+1}, \dots, H_{m+p}\}$
- Initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$
- Observable M

Output: Returns **True** if the condition of Theorem 1 is satisfied; otherwise, returns **False**

Construct matrix representations of Liouvillians:

$\mathcal{L}_j X = -i[H_j, X]$, $j = 1, \dots, m+p$;
Initialize a queue $Q \leftarrow \{(0, \dots, 0)\}$; // queue of multi-indices to be processed
Initialize visited set $V \leftarrow \{(0, \dots, 0)\}$; // set of visited multi-indices
Initialize a set $\mathcal{S}_{\text{all}}$ to store all generated $\mathcal{S}$ sets;
Initialize $\mathcal{S}_{(0, \dots, 0)} \leftarrow \mathfrak{M}(M)$;
if $\langle \psi_0 | A | \psi_0 \rangle = 0$ for all $A \in \mathcal{S}_{(0, \dots, 0)}$ then
| return **False**

end

while Q is not empty do

Pop $(k_1, \dots, k_m)$ from Q ;
for $i = 1$ to m do
| $(k'_1, \dots, k'_m) \leftarrow (k_1, \dots, k_i + 1, \dots, k_m)$;
| if $(k'_1, \dots, k'_m) \notin V$ then
| | $\mathcal{S}_{\text{prev}} \leftarrow \mathcal{S}_{(k_1, \dots, k_m)}$;
| | $\mathcal{S}_{\text{new}} \leftarrow \mathfrak{M}(\mathcal{L}_i \mathcal{S}_{\text{prev}})$;
| | if $\langle \psi_0 | A | \psi_0 \rangle = 0$ for all $A \in \mathcal{S}_{\text{new}}$ then
| | | return **False**
| | end
| | $\mathcal{S}_{(k'_1, \dots, k'_m)} \leftarrow \mathcal{S}_{\text{new}}$;
| | if $\mathcal{S}_{\text{new}}$ is already in $\mathcal{S}_{\text{all}}$ then
| | | Add $(k'_1, \dots, k'_m)$ to V without pushing to Q ;
| | | continue
| | end
| | Add $\mathcal{S}_{\text{new}}$ to $\mathcal{S}_{\text{all}}$;
| | Push $(k'_1, \dots, k'_m)$ into Q ;
| | Add $(k'_1, \dots, k'_m)$ to V ;
end

end

return **True**

Given a training dataset $\{(\mathbf{x}^{(k)}, y^{(k)})\}$ that contains N samples, the model is trained by minimizing the loss function

$$L(\Theta, \theta_0) = N^{-1} \sum_{k=1}^N \|\theta_0 f(\mathbf{x}^{(k)}, \Theta) - y^{(k)}\|^2, \quad (11)$$

where $\Theta = \{\theta_1(t), \dots, \theta_{m+p}(t)\}$. The scaling factor θ_0 is introduced to adapt the function represented by the model to the range of the dataset. Without loss of generality, the domain of $\mathbf{x}$ is always set as $\mathcal{X} = [-1, 1]^m$ [25].

In the numerical simulations, the control pulses are chosen in a piecewise-constant form. Each control pulse is divided into K subpulses, analogous to the number of layers in gate-based QNN models, with the subpulse amplitudes used to

parametrize the model. In all simulations, the sampling period is set to $\Delta t = T/K = 0.1$, and the model is trained by Adam optimizer [57] with a learning rate of 0.05. The simulations are conducted using MATLAB on a computer equipped with two 10-core Xeon CPUs and 208 GB RAM.

A. Expressivity of pulse-based models with dynamical symmetry

For the convenience of model description, we adopt the following notation:

$$\sigma_\alpha^{(k)} = \mathbb{I}_2 \otimes \cdots \otimes \sigma_\alpha \otimes \cdots \otimes \mathbb{I}_2, \quad \alpha = x, y, z,$$

for Pauli operators located on the k th site in an n -qubit system.

We first demonstrate the approximation of univariate functions using a two-qubit model:

$$H[x; \Theta] = x\sigma_z^{(1)}\sigma_z^{(2)} + \theta_1(t)\sigma_x^{(1)} + \theta_2(t)\sigma_x^{(2)}. \quad (12)$$

The system possesses a six-dimensional dynamical symmetry algebra:

$$\mathfrak{g} = \{\sigma_z^{(1)}\sigma_z^{(2)}, \sigma_x^{(1)}, \sigma_x^{(2)}\}_{LA} = \mathfrak{so}(4) \subset \mathfrak{su}(4), \quad (13)$$

and hence is uncontrollable. Choosing the observable $M = \sigma_z^{(1)}\sigma_z^{(2)}$, we can calculate that

$$\begin{aligned} \mathcal{S}_{2k} &= \{\sigma_z^{(1)}\sigma_z^{(2)}, \sigma_y^{(1)}\sigma_z^{(2)}, \sigma_z^{(1)}\sigma_y^{(2)}, \sigma_y^{(1)}\sigma_y^{(2)}\}, \\ \mathcal{S}_{2k+1} &= \{\sigma_x^{(1)}, \sigma_x^{(2)}\}, \quad k = 0, 1, 2, \dots \end{aligned}$$

According to Theorem 1, if the system is initialized in the state $|\psi_0\rangle = |0\rangle \otimes |0\rangle$, as is typically done, all odd power terms of x will vanish because $\langle\psi_0|\mathcal{S}_{2k+1}|\psi_0\rangle = \{0\}$ for all $k \in \mathbb{N}$. This implies that such a model can only approximate even functions. However, if the initial state is altered to $|\psi_0\rangle = (\frac{2}{\sqrt{5}}|0\rangle + \frac{1}{\sqrt{5}}|1\rangle) \otimes |0\rangle$, one can verify that $\langle\psi_0|\mathcal{S}_k|\psi_0\rangle \neq \{0\}$ holds for all $k \in \mathbb{N}$, suggesting that the model has the potential to approximate any univariate function.

To validate the above prediction, we apply the model (12) to approximate the following randomly selected target function:

$$\begin{aligned} f_0(x) &= 2x + 3x^2 + x^3 + 10x^6 \\ &\quad + 8x^7 - 13x^9 + 5x^{10} - 13x^{12}, \end{aligned} \quad (14)$$

from which 200 points are evenly sampled across $\mathcal{X} = [-1, 1]$ to train the model with pulse duration $T = 20$ (corresponding to $K = 200$). The simulation results presented in Fig. 2 indicate that the trained QML model accurately fit the target function when $|\psi_0\rangle = (\frac{2}{\sqrt{5}}|0\rangle + \frac{1}{\sqrt{5}}|1\rangle) \otimes |0\rangle$. However, when the initial state is $|\psi_0\rangle = |0\rangle \otimes |0\rangle$, the model can only approximate the even part of $f_0(x)$ due to the violation of the necessary condition in Theorem 1.

We also test the fitting of other randomly selected functions, which can all be approximated by the model, provided that the pulse duration is sufficiently long. The required pulse durations tend to be longer when the target function exhibits greater curvature or includes many higher-order power terms.

We further examine the approximation capability of pulse-based models for more complex bivariate functions using the

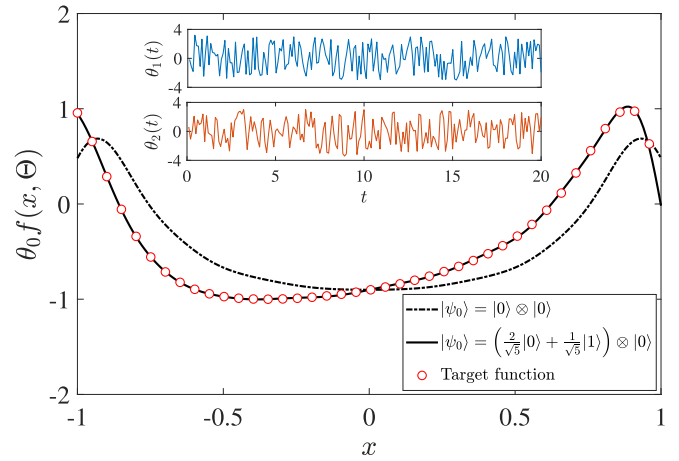


FIG. 2. The approximation of the polynomial function (14) using the two-qubit pulse-based model (12) under $\mathfrak{so}(4)$ dynamical symmetry. The inset figure displays the trained control pulses.

following Hamiltonian:

$$\begin{aligned} H[\mathbf{x}; \Theta] &= x_1\theta_1(t)\sigma_y^{(1)}\sigma_y^{(2)} + x_2\theta_2(t)\sigma_z^{(1)}\sigma_z^{(2)} \\ &\quad + \theta_3(t)\sigma_x^{(1)} + \theta_4(t)\sigma_x^{(2)}. \end{aligned} \quad (15)$$

The system is similar to Eq. (12) with an additional encoding Hamiltonian $\sigma_y^{(1)}\sigma_y^{(2)}$, which also exhibits $\mathfrak{so}(4)$ dynamical symmetry. The observable is chosen as $M = \sigma_z^{(1)} + \sigma_z^{(2)}$, based on which the computation yields the following periodic structure:

$$\mathcal{S}_{k_1, k_2} = \{\sigma_y^{(1)}, \sigma_y^{(2)}, \sigma_z^{(1)}, \sigma_z^{(2)}\} \quad (16)$$

for all cases where $k_1 + k_2$ is even and

$$\mathcal{S}_{k_1, k_2} = \{\sigma_x^{(1)}\sigma_y^{(2)}, \sigma_y^{(1)}\sigma_x^{(2)}, \sigma_x^{(1)}\sigma_z^{(2)}, \sigma_z^{(1)}\sigma_x^{(2)}\} \quad (17)$$

for all cases where $k_1 + k_2$ is odd. Accordingly, the initial state is designated as $|\psi_0\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes (|0\rangle + |1\rangle))$, which can be verified to satisfy the necessary condition in Theorem 1.

In the simulation, the target function is selected as

$$f_0(x_1, x_2) = (x_1^2 + x_2 - 3)^2 + (x_1 + x_2^2 - 1)^2 \quad (18)$$

with $50 \times 50 = 2500$ data points evenly sampled across the domain $\mathcal{X} = [-1, 1] \times [-1, 1]$ and the pulse duration T increased from 0.5 to 4.

For comparison, we train two pulse-based models. In the first case, we set the control pulses $\theta_1(t) = \theta_2(t) \equiv 1$, which considerably limits the tunability of the coefficients $C_{k_1, k_2}(\Theta)$. As illustrated in Fig. 3(a), the fitting progressively aligns with the surface of the target bivariate function as T increases; however, the precision remains limited no matter how T is prolonged. In the second case, we allow $\theta_1(t)$ and $\theta_2(t)$ to vary to enhance the tunability of coefficients $C_{k_1, k_2}(\Theta)$, thereby releasing more expressivity compared to the first case. The fitting results in Fig. 3(b) validate this expectation, as it nearly perfectly matches the target surface when $T = 4$, with a fitting error of $L(\Theta) \leq 10^{-3}$.

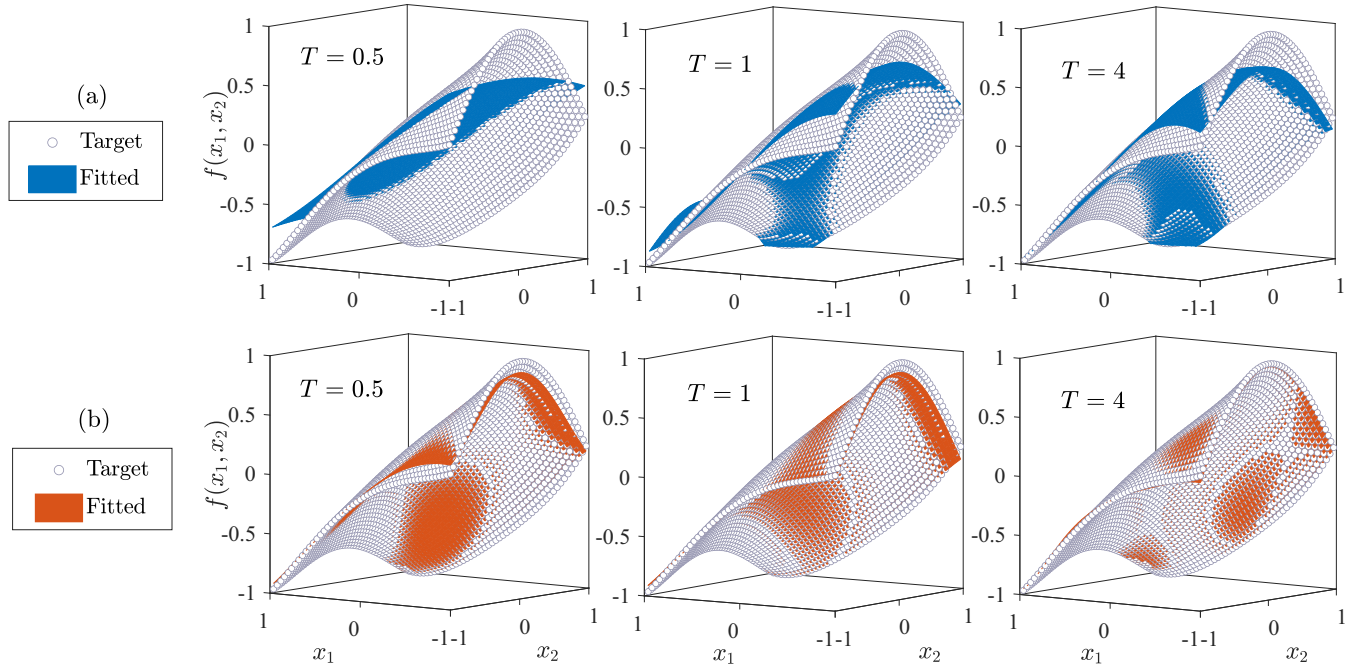


FIG. 3. The approximation results of bivariate function by two-qubit models with (a) $\theta_1(t) = \theta_2(t) = 1$ and (b) $\theta_1(t)$ and $\theta_2(t)$ being freely tunable.

B. The balance between expressivity and trainability

The simulations above demonstrate that pulse-based models can be expressive when uncontrollable due to dynamical symmetry. This establishes a framework for the design of large-scale pulse-based QML models that are both expressive and trainable. Subsequently, we select four classes of multiqubit pulse-based models that exhibit various dynamic symmetries.

Model 1. The first model class represents the standard collective-spin control model, relevant to platforms where dynamics are restricted to a symmetric subspace (e.g., collective modes, encoded qudits, or permutation-symmetric control) [58,59]. This model class possesses $\mathfrak{su}(2)$ dynamical symmetry with the following model Hamiltonian:

$$H_1[x; \Theta] = xJ_z + \theta_1(t)J_x + \theta_2(t)J_y \quad (19)$$

with two control pulses, where

$$\mathfrak{g} = \{iJ_x, iJ_y, iJ_z\}_{LA} = \mathfrak{su}(2)$$

is the 2^n -dimensional unitary irreducible representation of $\mathfrak{su}(2)$ (see the detailed description in the Appendix). In this case, the observable is chosen as $M = \frac{2}{2^n-1}J_z$ (a standard collective magnetization measurement) and the initial state is chosen as $|\psi_0\rangle = |0\rangle^{\otimes n}$.

Model 2. The second model class represents a set of isolated spins under local drives, which is typical in uncoupled superconducting qubits with local control [60]. This model class possesses $\mathfrak{su}(2)^{\oplus n}$ dynamical symmetry with the following noninteracting n -qubit system:

$$H_2[x; \Theta] = x \sum_{k=1}^n \sigma_z^{(k)} + \sum_{k=1}^n [\theta_x^{(k)}(t)\sigma_x^{(k)} + \theta_y^{(k)}(t)\sigma_y^{(k)}] \quad (20)$$

with $2n$ control pulses. The observable is selected as $M = \frac{1}{n} \sum_{k=1}^n \sigma_z^{(k)}$, and the initial state $|\psi_0\rangle = |0\rangle^{\otimes n}$.

Model 3. The third model class is the 2-local nearest-neighbor coupling model selected from Ref. [61], which is representative of spin chain with XY-type couplings (up to local basis rotations). This model class possesses $\mathfrak{so}(n)$ dynamical symmetry. The Hamiltonian of the model is as follows:

$$H_3[x; \Theta] = x\sigma_x^{(1)}\sigma_y^{(2)} + \sum_{k=1}^{n-1} \theta_k(t)\sigma_x^{(k)}\sigma_y^{(k+1)} \quad (21)$$

with $n-1$ control pulses. The observable $M = \sigma_x^{(n-1)}\sigma_y^{(n)}$ and the initial state $|\psi_0\rangle = \frac{1}{2}|0\rangle^{\otimes(n-2)} \otimes (|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle)$.

Model 4. The fourth model class is a fully controllable model utilizing the following circularly coupled n -qubit system in Ref. [25]:

$$H_4[x; \Theta] = x \sum_{k=1}^n \sigma_z^{(k)} + \sum_{k=1}^n [\theta_x^{(k)}(t)\sigma_x^{(k)} + \theta_y^{(k)}(t)\sigma_y^{(k)}] + \theta_z^{(1)}(t)\sigma_z^{(1)}\sigma_z^{(2)} + \dots + \theta_z^{(n)}(t)\sigma_z^{(n)}\sigma_z^{(1)} \quad (22)$$

with $3n$ control pulses, which is an Ising-type architecture with local fields and periodic boundary conditions, commonly used in quantum simulation and hardware-efficient modeling [62]. The observable is selected as $M = \sigma_z^{(1)}$ and the initial state as $|\psi_0\rangle = |0\rangle^{\otimes n}$.

The initial states and measurement observables selected in the four models all satisfy the necessary condition in Theorem 1 for expressivity (see the Appendix for validation). In addition, the observables are selected from the associated Lie algebra $\mathfrak{g}$ for the sake of trainability [44]. The theoretical variance of the measurement output (5) is presented in Table I

TABLE I. The symmetry Lie algebra, dimension, and the variance of the inference output (5) for each model.

Model	$\mathfrak{g}$	$\dim(\mathfrak{g})$	$\text{Var}[f(\mathbf{x}, \Theta)]$
1	$\mathfrak{su}(2)$	3	$\frac{1}{3}$
2	$\mathfrak{su}(2)^{\oplus n}$	$3n$	$\frac{1}{3n}$
3	$\mathfrak{so}(n)$	$\frac{n(n-1)}{2}$	$\frac{2}{n(n-1)}$
4	$\mathfrak{su}(2^n)$	$4^n - 1$	$\frac{1}{2^n + 1}$

(see the Appendix for detailed calculations), which indicates that the variance is inversely proportional to the dimension of the dynamic Lie algebra $\mathfrak{g}$. Note that the variance of $L(\Theta)$ is different from that of $f(\mathbf{x}, \Theta)$ listed in Table I as $L(\Theta)$ is a quadratic function of $f(\mathbf{x}, \Theta)$ and is averaged over a training dataset, whose analytic formula is unavailable. However, their scaling with qubit number is consistent, as can be seen in the simulations.

During the numerical simulations, these models were trained to fit the univariate function (14) using up to six qubits. As can be seen in the inset of Fig. 4(a), each model can approximate the target function with arbitrarily small error when the pulse duration T is sufficiently extended. This indicates that the model expressivity grows with T , and thus provides a measure to quantify the degree of expressivity. Here, we adopt the minimal pulse duration required for the training loss to fall below 10^{-3} to evaluate the expressivity (see the inset for illustration).

Based on this measure, the simulation results in Fig. 4(a) compare the variation of expressivity with respect to system size across the four model classes. It is observed that, except the special case of $\mathfrak{su}(2)$ whose dimensionality is constant, the expressivity can be enhanced by increasing the number of qubits. Furthermore, when the number of qubits is sufficiently large, expressive power increases in the order of $\mathfrak{su}(2)$, $\mathfrak{su}(2)^{\oplus n}$, $\mathfrak{so}(n)$, and $\mathfrak{su}(2^n)$, implying that expressivity can be improved by enlarging the dynamical symmetry Lie algebra.

Regarding the trainability of the models, we estimate the variance of the loss function $L(\Theta)$ and the average magnitude of the partial derivative with respect to a randomly chosen parameter in the numerical simulation. For each model, we randomly pick 1000 sets of control pulses to numerically evaluate $L(\Theta)$ and its variance. Since the variance is also dependent on the pulse duration T , we set T sufficiently long so that the estimated variance gradually converges to a stationary value, corresponding to the condition under which the model approaches a 2-design.

Figure 4(b) illustrates the relation between the stationary variance and the number of qubits. Consistent with the theoretical analysis in Table I, the variance related to $\mathfrak{su}(2^n)$ exhibits an exponential decline as the number of qubits n increases, indicating the presence of a barren plateau. In contrast, the variance associated with the $\mathfrak{su}(2)$ symmetry remains constant owing to its constant dimensionality, while it decreases at a polynomial rate for the $\mathfrak{su}(2)^{\oplus n}$ and $\mathfrak{so}(n)$ symmetries.

The average magnitude of the partial derivative with respect to the chosen parameter is shown in Fig. 4(c), the scaling behavior of which versus qubit number closely mirrors the trend observed in Fig. 4(b) for the sample variance of loss function. Specifically, the average gradient magnitude in the $\mathfrak{su}(2^n)$ case exhibits an exponential decay with increasing system size, indicating that the optimization will become nearly infeasible when the system scale is large. In contrast, for the $\mathfrak{su}(2)$ symmetry the gradient magnitude remains essentially independent of n , while the $\mathfrak{su}(2)^{\oplus n}$ and $\mathfrak{so}(n)$ symmetries display only polynomial decay. As a consequence, these three symmetry-restricted models consistently retain favorable trainability.

To further illustrate the trainability of the different examples, we show in Fig. 5 the training dynamics for the six-qubit instances, where eight independent training curves with random initializations are plotted for each example (all converging to a precision of 10^{-3}). Compared with the losses of $\mathfrak{su}(2)$, $\mathfrak{su}(2)^{\oplus n}$, and $\mathfrak{so}(n)$ cases, the loss of the $\mathfrak{su}(2^n)$

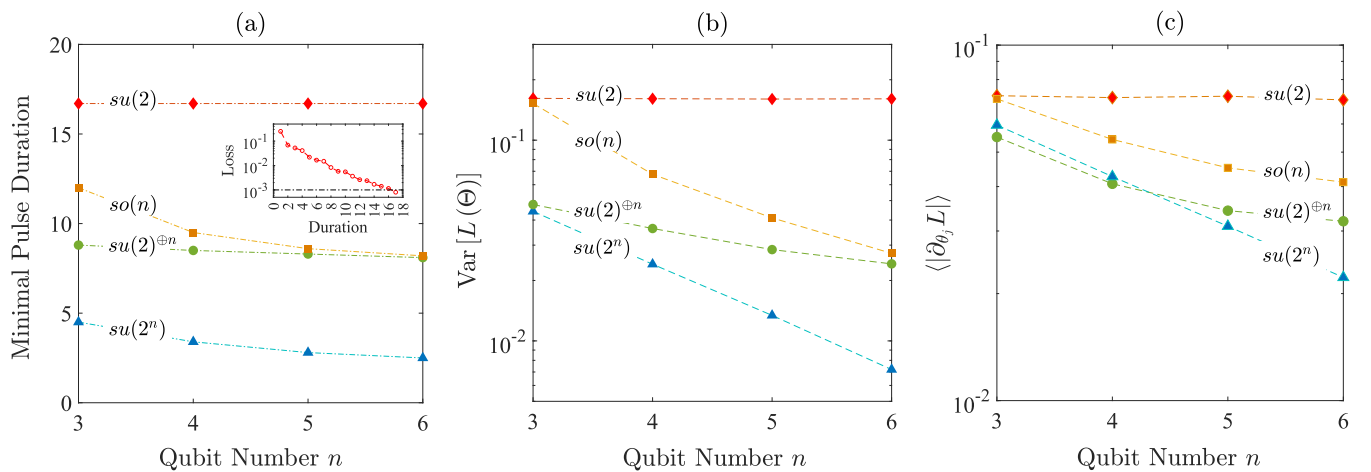


FIG. 4. The comparison of expressivity and trainability between model classes under $\mathfrak{su}(2)$, $\mathfrak{su}(2)^{\oplus n}$, $\mathfrak{so}(n)$, and $\mathfrak{su}(2^n)$ dynamic symmetries: (a) The expressivity (evaluated by the minimal pulse duration when the training loss reaches 10^{-3} , as shown in the inset figure) vs qubit number for each example. (b) The sample variance of loss function vs qubit number for each example. The inset shows how the minimal pulse duration is obtained [Eq. (19)]. (c) The average magnitude of the partial derivative of the loss function vs qubit number for each example.

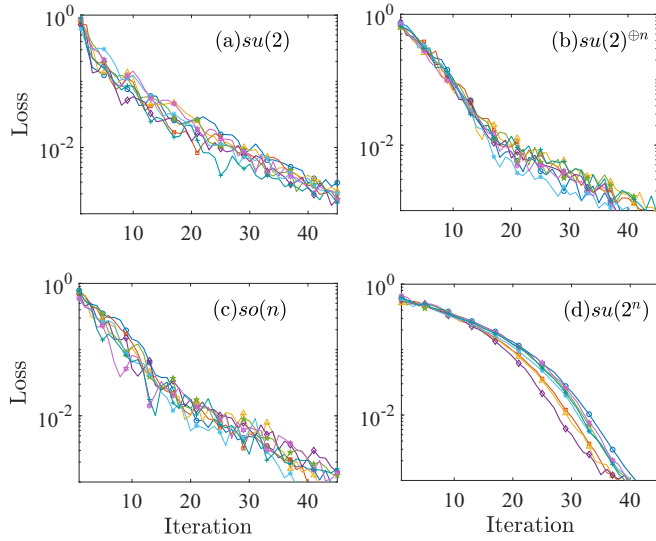


FIG. 5. Training process of loss over iterations for six-qubit models with (a) $\mathfrak{su}(2)$, (b) $\mathfrak{su}(2)^{\oplus n}$, (c) $\mathfrak{so}(n)$, and (d) $\mathfrak{su}(2^n)$ dynamical symmetries.

case decreases more slowly during early iterations due to the small gradients caused by barren plateaus; once this region is escaped, the loss subsequently converges rapidly owing to the strong expressivity of the model. It is noted that the training curve does not exhibit genuine barren plateau behavior because the system dimensionality is not sufficiently large; however, the adverse effects in the beginning stage are still evident.

Taking into account the influence of expressivity, the $\mathfrak{su}(2)^{\oplus n}$ and $\mathfrak{so}(n)$ models possess higher expressivity than the $\mathfrak{su}(2)$ model at $n = 6$, leading to faster overall convergence as shown in Fig. 5. However, despite its substantially higher expressivity, the $\mathfrak{su}(2^n)$ model achieves a convergence speed that is only comparable to that of the $\mathfrak{su}(2)^{\oplus n}$ and $\mathfrak{so}(n)$ models, which further highlights the detrimental impact of barren plateaus on its training process. Combined with the analysis presented in the paper, these results suggest that, as the system size increases, the $\mathfrak{su}(2^n)$ model will become effectively untrainable already at early stages of optimization, whereas the other three symmetry-preserving models can maintain good trainability even at larger system scales.

These observations suggest that barren plateaus can be effectively avoided by restricting the dynamical evolution to submanifolds within the quantum state Hilbert space through engineered dynamical symmetry, thereby achieving a good balance between expressivity and trainability. It is noted that the overall expressivity and trainability are highly dependent on the selection of the Hamilton structure, the initial state, and the observable. For instance, the expressivity of the larger $\mathfrak{so}(n)$ models is inferior to that of the smaller $\mathfrak{su}(2)^{\oplus n}$ models when the number of qubits is small, as the latter has more tunable control pulses. Nonetheless, their rates of decrease align with predictions, indicating that $\mathfrak{so}(n)$ models will become more expressive than $\mathfrak{su}(2)^{\oplus n}$ models when n is sufficiently large.

V. CONCLUSION AND OUTLOOK

In conclusion, our work establishes a practical design framework for pulse-based QML models that balances expressivity and trainability through dynamical symmetry, initial state, and observable selection. The design incorporates Dyson polynomial-series expansion alongside existing Lie-algebraic theories related to the barren plateau phenomenon in quantum systems exhibiting dynamical symmetry. Theoretical analysis and numerical simulations demonstrate the potential to leverage dynamical symmetry to construct expressive and trainable hardware-efficient models on NISQ devices.

Nevertheless, it remains an open question whether the proposed necessary condition is sufficient for the expressivity of QML models with dynamical symmetry. Simulation results indicate that complete expressivity is not always attainable for models that satisfy this condition, particularly when control pulses are constrained. We conjecture that sufficiency is achieved when all control pulses are unrestricted and the pulse duration is sufficiently long, but this requires verification in the future.

We also note that, even if the pulse-based models are expressive and trainable, the loss landscape may contain undesirable local traps induced by dynamical symmetry [63], which can attract the training process to suboptimal models. Although this issue is rarely encountered in our simulations, the possibility remains, and thus the training algorithm needs to be carefully designed.

Besides the trade-off between model expressivity and trainability, an important issue for future investigation is their trade-off between model generalizability, which pertains to the model's performance on unseen data. The overall balance among these three attributes holds significant practical importance and will be the focus of subsequent studies.

ACKNOWLEDGMENT

This work was supported by the Innovation Program for Quantum Science and Technology (No. 2021ZD0300201), NSFC Grant No. 62173201 and the Tsinghua-Toyota Joint Research Fund.

DATA AVAILABILITY

The data are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX: VALIDATION OF EXPRESSIVITY AND VARIANCE CALCULATION FOR THE FOUR EXAMPLES

1. Model 1

The Hamiltonian of the model is as follows:

$$H_1[x; \Theta] = xJ_z + \theta_1(t)J_x + \theta_2(t)J_y, \quad (\text{A1})$$

where $\mathfrak{g} = \{iJ_x, iJ_y, iJ_z\}_{LA} = \mathfrak{su}(2)$ is the 2^n -dimensional unitary irreducible representation of $\mathfrak{su}(2)$. Let $J_{\pm} = J_x \pm iJ_y$ be the raising and lowering operators, and

$$\left\{ |s\rangle \mid s = -\frac{d-1}{2}, -\frac{d-3}{2}, \dots, \frac{d-3}{2}, \frac{d-1}{2} \right\}$$

be the basis states. Under the basis state, we have

$$J_z|s\rangle = s|s\rangle,$$

$$J_{\pm}|s\rangle = \sqrt{\left(\frac{d-1}{2} \mp s\right)\left(\frac{d+1}{2} \pm s\right)}|s \pm 1\rangle.$$

Since the maximum and minimum eigenvalues of J_z are $\pm \frac{2^n-1}{2}$, and the range of the target function is chosen to be $[-1, 1]$, we normalize the observable to $M = \frac{2}{2^n-1}J_z$. According to Theorem 1, it is easy to obtain that

$$\mathcal{S}_k = \{J_x, J_y, J_z\}, \quad k \in N.$$

If the initial state $|\psi_0\rangle = |0\rangle^{\otimes n}$, it can be verified that the model satisfies the condition for expressivity.

On the other hand, the model is uncontrollable and $\dim(\mathfrak{g}) = 3$. Using the orthonormal basis

$$\left\{ \frac{iJ_x}{\sqrt{\langle J_x, J_x \rangle}}, \frac{iJ_y}{\sqrt{\langle J_y, J_y \rangle}}, \frac{iJ_z}{\sqrt{\langle J_z, J_z \rangle}} \right\}$$

of the Lie algebra $\mathfrak{su}(2)$, we have

$$\begin{aligned} \text{Var}[f(\mathbf{x}, \Theta)] &= \frac{1}{3} P_{\mathfrak{g}}(\rho) P_{\mathfrak{g}}(M) \\ &= \frac{1}{3} \left(\frac{2^n-1}{2} \right)^2 \frac{1}{\langle J_z, J_z \rangle} \left(\frac{2}{2^n-1} \right)^2 \langle J_z, J_z \rangle \\ &= \frac{1}{3}. \end{aligned}$$

2. Model 2

The Hamiltonian of the model is as follows:

$$H_2[x; \Theta] = x \sum_{k=1}^n \sigma_z^{(k)} + \sum_{k=1}^n [\theta_x^{(k)}(t) \sigma_x^{(k)} + \theta_y^{(k)}(t) \sigma_y^{(k)}]. \quad (\text{A2})$$

The Lie algebra generated by this model is $\mathfrak{g} = \mathfrak{g}_1 \oplus \dots \oplus \mathfrak{g}_n$, where each $\mathfrak{g}_j = \mathfrak{su}(2)$; thus, $\dim(\mathfrak{g}) = 3n$. In this case,

$$\mathcal{S}_k = \{\sigma_x^{(i)}, \sigma_y^{(i)}, \sigma_z^{(i)}\}_{1 \leq i \leq n}, \quad k \in N.$$

We can verify that $\langle \psi_0 | \mathcal{S}_k | \psi_0 \rangle \neq \{0\}$ for all $k \in N$, which meets the condition in Theorem 1.

Given that the orthonormal basis of $\mathfrak{g}_j$ is $i\{2^{-n/2}\sigma_x^{(j)}, 2^{-n/2}\sigma_y^{(j)}, 2^{-n/2}\sigma_z^{(j)}\}$, we find $M_{\mathfrak{g}_j} = \frac{1}{n}\sigma_z^{(j)}$,

and $\rho_{\mathfrak{g}_j} = 2^{-n}\sigma_z^{(j)}$ for all j . Then, we obtain

$$\begin{aligned} \text{Var}[f(\mathbf{x}, \Theta)] &= \sum_{j=1}^n \frac{1}{3} P_{\mathfrak{g}_j}(\rho) P_{\mathfrak{g}_j}(M) \\ &= \sum_{j=1}^n \frac{1}{3} \times 2^{-n} \times \frac{1}{n^2} \times 2^n \\ &= \frac{1}{3n}. \end{aligned}$$

3. Model 3

The Hamiltonian of the model is as follows:

$$H_3[x; \Theta] = x \sigma_x^{(1)} \sigma_y^{(2)} + \sum_{k=1}^{n-1} \theta_k(t) \sigma_x^{(k)} \sigma_y^{(k+1)}. \quad (\text{A3})$$

We can derive that

$$\mathcal{S}_k = \{\widehat{\sigma_x^{(i)} \sigma_y^{(j)}}\}_{1 \leq i < j \leq n}, \quad k \in N,$$

where $\widehat{\sigma_x^{(i)} \sigma_y^{(j)}} = \sigma_x^{(i)} \sigma_z^{(i+1)} \dots \sigma_z^{(j-1)} \sigma_y^{(j)}$. For the selected initial state $|\psi_0\rangle = |0\rangle^{\otimes(n-2)} \otimes (\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle) \otimes (\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle)$, we can verify that for any $n \geq 3$, this model satisfies the condition in Theorem 1.

As for the trainability of the model, the Lie algebra associated with this model is $\mathfrak{g} = \mathfrak{so}(n)$ that has dimension $\dim(\mathfrak{g}) = \frac{n(n-1)}{2} \in \mathcal{O}(\text{poly}(n))$. Since the orthonormal basis of $\mathfrak{so}(n)$ is

$$\{i2^{-n/2}\widehat{\sigma_x^{(i)} \sigma_y^{(j)}} \mid 1 \leq i < j \leq n\},$$

we can easily get $P_{\mathfrak{g}}(M) = 2^n$ and $P_{\mathfrak{g}}(\rho) = 2^{-n}$. Then, we have

$$\begin{aligned} \text{Var}[f(\mathbf{x}, \Theta)] &= \frac{2}{n(n-1)} P_{\mathfrak{g}}(\rho) P_{\mathfrak{g}}(M) \\ &= \frac{2}{n(n-1)}. \end{aligned}$$

4. Model 4

The model can be verified to possess full expressivity according to the criterion proposed in Ref. [25] because the control Hamiltonians generate the entire $\mathfrak{su}(2^n)$ Lie algebra.

For the calculation of variance, based on the analysis of Ref. [44], we have

$$P_{\mathfrak{g}}(M) = 2^n, \quad P_{\mathfrak{g}}(\rho) = 1 - \frac{1}{2^n},$$

which yields

$$\text{Var}[f(\mathbf{x}, \Theta)] = \frac{1}{2^n + 1}.$$

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).

[2] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).

- [3] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, *et al.*, Quantum computational advantage via 60-qubit 24-cycle random circuit sampling, *Sci. Bull.* **67**, 240 (2022).
- [4] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, *et al.*, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [5] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
- [6] A. Callison and N. Chancellor, Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond, *Phys. Rev. A* **106**, 010101 (2022).
- [7] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, [arXiv:1802.06002](#).
- [8] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [9] R.-B. Wu, X. Cao, P. Xie, and Y.-x. Liu, End-to-end quantum machine learning implemented with controlled quantum dynamics, *Phys. Rev. Appl.* **14**, 064020 (2020).
- [10] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms, *PRX Quantum* **2**, 010101 (2021).
- [11] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, *Phys. Rev. Res.* **3**, 023092 (2021).
- [12] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, Gate-free state preparation for fast variational quantum eigensolver simulations, *npj Quantum Inf.* **7**, 155 (2021).
- [13] R. De Keijzer, O. Tse, and S. Kokkelmans, Pulse based variational quantum optimal control for hybrid quantum computing, *Quantum* **7**, 908 (2023).
- [14] K. M. Sherbert, H. Amer, S. E. Economou, E. Barnes, and N. J. Mayhall, Parametrization and optimizability of pulse-level variational quantum eigensolvers, *Phys. Rev. Appl.* **23**, 024036 (2025).
- [15] L. Broers and L. Mathey, Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols, *Phys. Rev. Res.* **6**, 013076 (2024).
- [16] Y. Wang, Y. Ding, F. A. Cárdenas-López, and X. Chen, Pulse-based variational quantum optimization and metalearning in superconducting circuits, *Phys. Rev. Appl.* **22**, 024009 (2024).
- [17] X. Pan, X. Cao, W. Wang, Z. Hua, W. Cai, X. Li, H. Wang, J. Hu, Y. Song, D.-L. Deng, *et al.*, Experimental quantum end-to-end learning on a superconducting processor, *npj Quantum Inf.* **9**, 18 (2023).
- [18] A. Melo, N. Earnest-Noble, and F. Tacchino, Pulse-efficient quantum machine learning, *Quantum* **7**, 1130 (2023).
- [19] M. M. Ibrahim, H. Mohammadbagherpoor, C. Rios, N. T. Bronn, and G. T. Byrd, Evaluation of parameterized quantum circuits with cross-resonance pulse-driven entanglers, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
- [20] D. J. Egger, C. Capecchi, B. Pokharel, P. K. Barkoutsos, L. E. Fischer, L. Guidoni, and I. Tavernelli, Pulse variational quantum eigensolver on cross-resonance-based hardware, *Phys. Rev. Res.* **5**, 033159 (2023).
- [21] Z. Liang, J. Cheng, Z. Song, H. Ren, R. Yang, K. Liu, P. Kogge, T. Li, Y. Ding, and Y. Shi, Towards advantages of parameterized quantum pulses, [arXiv:2304.09253](#).
- [22] A. Asthana, C. Liu, O. R. Meitei, S. E. Economou, E. Barnes, and N. J. Mayhall, Leakage reduces device coherence demands for pulse-level molecular simulations, *Phys. Rev. Appl.* **19**, 064071 (2023).
- [23] A. Michel, S. Grijalva, L. Henriot, C. Domain, and A. Browaeys, Blueprint for a digital-analog variational quantum eigensolver using Rydberg atom arrays, *Phys. Rev. A* **107**, 042602 (2023).
- [24] D. Rabinovich, S. Adhikary, E. Campos, V. Akshay, E. Anikin, R. Sengupta, O. Lakhmanskaya, K. Lakhmanskii, and J. Biamonte, Ion-native variational ansatz for quantum approximate optimization, *Phys. Rev. A* **106**, 032418 (2022).
- [25] H.-X. Tao, J. Hu, and R.-B. Wu, Unleashing the expressive power of pulse-based quantum neural networks, *Quantum Mach. Intell.* **7**, 109 (2025).
- [26] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature (London)* **567**, 209 (2019).
- [27] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).
- [28] M. Schuld, Supervised quantum machine learning models are kernel methods, [arXiv:2101.11020](#).
- [29] J. Jäger and R. V. Krems, Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines, *Nat. Commun.* **14**, 576 (2023).
- [30] S. Lloyd, Almost any quantum logic gate is universal, *Phys. Rev. Lett.* **75**, 346 (1995).
- [31] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing barren plateaus with tools from quantum optimal control, *Quantum* **6**, 824 (2022).
- [32] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, Barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
- [33] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, *Quantum Sci. Technol.* **7**, 045015 (2022).
- [34] Q. Miao and T. Barthel, Equivalence of cost concentration and gradient vanishing for quantum circuits: An elementary proof in the Riemannian formulation, *Quantum Sci. Technol.* **9**, 045039 (2024).
- [35] L. Banchi, D. Burgarth, and M. J. Kastoryano, Driven quantum dynamics: Will it blend? *Phys. Rev. X* **7**, 041015 (2017).
- [36] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Trainability of dissipative perceptron-based quantum neural networks, *Phys. Rev. Lett.* **128**, 180505 (2022).
- [37] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).
- [38] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, *Nat. Commun.* **12**, 6961 (2021).

- [39] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Effect of barren plateaus on gradient-free optimization, *Quantum* **5**, 558 (2021).
- [40] M. Cerezo and P. J. Coles, Higher order derivatives of quantum neural networks with barren plateaus, *Quantum Sci. Technol.* **6**, 035006 (2021).
- [41] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, Barren plateaus preclude learning scramblers, *Phys. Rev. Lett.* **126**, 190501 (2021).
- [42] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, *PRX Quantum* **3**, 010313 (2022).
- [43] L. Friedrich and J. Maziero, Quantum neural network cost function concentration dependency on the parametrization expressivity, *Sci. Rep.* **13**, 9978 (2023).
- [44] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo, A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits, *Nat. Commun.* **15**, 7172 (2024).
- [45] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, Characterizing barren plateaus in quantum ansätze with the adjoint representation, *Nat. Commun.* **15**, 7171 (2024).
- [46] N. L. Diaz, D. García-Martín, S. Kazi, M. Larocca, and M. Cerezo, Showcasing a barren plateau theory beyond the dynamical Lie algebra, *arXiv:2310.11505*.
- [47] F. J. Gil Vidal and D. O. Theis, Input redundancy for parameterized quantum circuits, *Front. Phys.* **8**, 297 (2020).
- [48] T. Goto, Q. H. Tran, and K. Nakajima, Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces, *Phys. Rev. Lett.* **127**, 090506 (2021).
- [49] A. Pérez-Salinas, D. López-Núñez, A. García-Sáez, P. Forn-Díaz, and J. I. Latorre, One qubit as a universal approximant, *Phys. Rev. A* **104**, 012405 (2021).
- [50] M. Schuld, R. Sweke, and J. J. Meyer, Effect of data encoding on the expressive power of variational quantum-machine-learning models, *Phys. Rev. A* **103**, 032430 (2021).
- [51] Z. Yu, H. Yao, M. Li, and X. Wang, Power and limitations of single-qubit native quantum neural networks, in *Proceedings of the 36th International Conference on Neural Information Processing Systems* (Curran Associates Inc., New York, 2022).
- [52] A. Isidori, *Nonlinear Control Systems* (Springer Science & Business Media, Springer London, 1995).
- [53] Z. Liang, J. Cheng, H. Ren, H. Wang, F. Hua, Z. Song, Y. Ding, F. T. Chong, S. Han, X. Qian, *et al.*, NAPA: Intermediate-level variational native-pulse ansatz for variational quantum algorithms, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **43**, 1834 (2024).
- [54] D. Meirum and S. H. Frankel, Pansatz: Pulse-based ansatz for variational quantum algorithms, *Front. Quantum Sci. Technol.* **2**, 1273581 (2023).
- [55] Z. Liang, H. Wang, J. Cheng, Y. Ding, H. Ren, Z. Gao, Z. Hu, D. S. Boning, X. Qian, S. Han, *et al.*, Variational quantum pulse learning, in *IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Broomfield, CO, USA, 2022), pp. 556–565.
- [56] Z. Liang, Z. Song, J. Cheng, Z. He, J. Liu, H. Wang, R. Qin, Y. Wang, S. Han, X. Qian, *et al.*, Hybrid gate-pulse model for variational quantum algorithms, in *60th ACM/IEEE Design Automation Conference (DAC)* (IEEE, San Francisco, California, US, 2023), pp. 1–6.
- [57] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv:1412.6980*.
- [58] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Atomic coherent states in quantum optics, *Phys. Rev. A* **6**, 2211 (1972).
- [59] X. Zhao, Y. Zhao, M. Li, T. Li, Q. Liu, S. Guo, and X. Yi, A strategy for preparing quantum squeezed states using reinforcement learning, *Ann. Phys. (NY)* **536**, 2400056 (2024).
- [60] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, *Appl. Phys. Rev.* **6**, 021318 (2019).
- [61] R. Wiersema, E. Kökcü, A. F. Kemper, and B. N. Bakalov, Classification of dynamical Lie algebras of 2-local spin systems on linear, circular and fully connected topologies, *npj Quantum Inf.* **10**, 110 (2024).
- [62] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature (London)* **549**, 242 (2017).
- [63] R.-B. Wu, M. A. Hsieh, and H. Rabitz, Role of controllability in optimizing quantum dynamics, *Phys. Rev. A* **83**, 062306 (2011).