Quantum advantage without exponential concentration: Trainable kernels for symmetry-structured data

Laura J. Henderson[1], Kerstin Beer[*2], Salini Karuvade[3], Riddhi Gupta[1], Angela White[1], and Sally Shrapnel[1]

[1]*School of Mathematics and Physics, The University of Queensland, Brisbane, Australia*
[2]*School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia*
[3]*School of Physics, The University of Sydney, Sydney, Australia*

Abstract

Quantum kernel methods promise enhanced expressivity for learning structured data, but their usefulness has been limited by kernel concentration and barren plateaus. Both effects are mathematically equivalent and suppress trainability. We analytically prove that covariant quantum kernels tailored to datasets with group symmetries avoid exponential concentration, ensuring stable variance and guaranteed trainability independent of system size. Our results extend beyond prior two-coset constructions to arbitrary coset families, broadening the scope of problems where quantum kernels can achieve advantage. We further derive explicit bounds under coherent noise models—including unitary errors in fiducial state preparation, imperfect unitary representations, and perturbations in group element selection—and show through numerical simulations that the kernel variance remains finite and robust, even under substantial noise. These findings establish a family of quantum learning models that are simultaneously trainable, resilient to coherent noise, and linked to classically hard problems, positioning group-symmetric quantum kernels as a promising foundation for near-term and scalable quantum machine learning.

1 Introduction

The interdisciplinary field at the intersection of quantum computing and machine learning - quantum machine learning (QML) - seeks to harness quantum mechanical resources to perform learning algorithms. Broadly, QML research falls into three main categories [1]: (i) quantum algorithms that accelerate classical ML tasks [2, 3, 4]; (ii) classical ML techniques used to analyze and characterize quantum systems [5, 6, 7]; and (iii) quantum devices employed to learn from quantum data [8, 9, 10].

A powerful technique of classical machine learning are kernel methods [11]. These map input data into high-dimensional feature spaces and evaluate inner products between data points there. The so-called kernel trick allows to learn complex patterns without the computational burden of directly working in high dimensions. Quantum kernel [12, 13, 14, 15, 16] methods extend this idea by using quantum computers to evaluate these inner products in exponentially large Hilbert space: classical data is encoded into quantum states via quantum feature maps, and the overlap between these states is computed by the quantum device which offers enhanced expressivity.

As quantum hardware is advancing, there is growing interest in identifying problems for which QML can offer a clear computational advantage over classical methods. One promising approach is to construct a trainable problem instance inspired by quantum algorithms. In the following case, this instance is derived from the discrete logarithm problem (DLP) and thus inherits its classical computational hardness. The authors of [17, 18] introduce a novel class of quantum kernels tailored for datasets exhibiting group symmetries. These kernels use unitary representations of the underlying group to construct feature maps that are covariant under the group action. This enables the algorithm to learn symmetries present in the data. The authors demonstrate, both theoretically and experimentally (on a 27-qubit superconducting quantum processor), that such covariant kernels can be used to effectively classify data defined on coset-structured tasks that are challenging for classical models unless the symmetry is explicitly encoded. This family of quantum kernels is especially notable because it extends the kernel studied in [17], which was capable of learning the task inspired by the DLP.

---

*Correspondence to: kerstin.beer@mq.edu.au

Despite these promises, the trainability of QML methods remains a key challenge. For quantum kernel methods, a major obstacle is kernel concentration [19], where kernel values concentrate exponentially and the classifier loses expressive power. Interestingly, kernel concentration is closely related to barren plateaus in variational quantum circuits [20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]; the presence of one implies the other [19, 16]. While various strategies have been proposed to mitigate barren plateaus [31, 32, 21, 33, 34, 35, 36] an equally desirable long-term goal is to design QML approaches that inherently avoid kernel concentration and its associated trainability issues.

In this work we analytically prove that quantum kernel methods tailored to data with coset group structure do not exhibit barren plateaus, ensuring trainability regardless of system size. Furthermore, we show that model remain trainable in the presence of various sources of coherent noise, including errors in fiducial state preparation, imperfect unitary representations, and perturbations in group element selection. Our numerical simulations support these results, demonstrating stable kernel variance and effective classification across a range of parameters. Together, these findings characterize a family of quantum learning models that are provably trainable, resilient to coherent noise, and grounded in classically difficult problems, thereby establishing group-symmetric quantum kernels as a promising framework for both near-term and scalable quantum machine learning.

In Section 2, we present the methods and theoretical tools underpinning our analysis. Section 2.1 introduces covariant quantum kernels tailored for data with group structure, while Section 2.2 reviews the broader framework of quantum kernels. Section 2.3 discusses key trainability challenges, focusing on barren plateaus and exponential concentration effects. Section 3 contains our main results. In Section 3.1, we derive analytical expressions for the variance of the coset quantum kernel. Section 3.2 reports numerical simulations validating these results. Section 3.3 extends the analysis to noisy systems, examining the impact of errors in the fiducial state (Section 3.3.1), the unitary representation of group elements (Section 3.3.2), and the selection of group elements (Section 3.3.3), as well as providing numerical simulations of these errors (Section 3.3.4). Finally, Section 4 discusses the implications of our findings, their limitations, and potential directions for future research.

2 Methods

In this section, we describe the techniques and theoretical tools underlying our approach. We begin with an overview of covariant quantum kernels designed for data with group structure, followed by a brief review of quantum kernel methods. We then discuss challenges in trainability, particularly the issue of barren plateaus and concentration phenomena in parameterized quantum circuits.

2.1 Learning on data with group structure

In the following we introduce the class of quantum kernels presented in [18], which are specifically designed to learn datasets with an underlying group structure. Specifically, it is assumed that the data belongs to either of the two cosets obtained from a subgroup $S \subset \mathcal{G}$ of the group $\mathcal{G}$. The cosets are obtained from S by the left or right action of the group elements $c_1, c_2 \in \mathcal{G}$ on the elements of $\mathcal{G}$. Let us denote the dataset by $\mathcal{X} := c_1 S \cup c_2 S$. The task is to assign a label $y \in \{-1, 1\}$ to a datum $x \in \mathcal{X}$, where the classification rule, also referred to as the *concept* [17], is determined by the coset to which x belongs. A special class of quantum kernels called the covariant kernels was introduced in [18] to learn the concept class for problems with coset structure. In particular, the concept class chosen in [17] was based on the discrete log problem (DLP) and it was shown to be a learning problem with the same classical hardness as the DLP itself.

We now discuss the above terms in detail and also extend the learning problem to the case of more than two cosets. To embed the classical data $\mathcal{X} \subset \mathcal{G}$ into the quantum setting we define a unitary representation $D : \mathcal{G} \rightarrow U(2^N)$ of the group $\mathcal{G}$ that acts on N qubits. Given a homogeneous space induced by some subgroup S of $\mathcal{G}$, we define $m \geq 2$ distinct left cosets for the group elements $\boldsymbol{c}_j \in \mathcal{G}$, $1 \leq j \leq m$ by

$$C_j = \boldsymbol{c}_j S,$$

where $\{\boldsymbol{c}_j\}$ is drawn randomly, for instance, according to the Haar measure on $\mathcal{G}$. We also assume that the number of cosets m is independent of n and N. The learning goal is to categorise data $\boldsymbol{x} \in \bigcup_j C_j = \mathcal{X}$ into the correct coset.

The algorithm introduced in [18] to solve the above learning problem requires a reference state $|\psi\rangle$, which is invariant under the action of the subgroup S

$$D_{\boldsymbol{s}} |\psi\rangle = |\psi\rangle \text{ for all } \boldsymbol{s} \in S$$
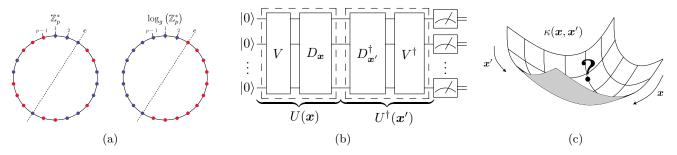
(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

Figure 1: **Overview of the quantum kernel and concept class studied in this work.** (a) An illustration of the problem of learning the concept class, $\mathcal{C}$ [17]. In the $\mathbb{Z}_p^*$ group (where p is prime) the data is scrambled (*left*), but after taking the discrete log (DLOG), the data is easily separated by the concept $\boldsymbol{c}$ (*right*). (b) The quantum circuit for calculating matrix elements of a covariant quantum kernel, $\kappa(\boldsymbol{x}, \boldsymbol{x}')$. Each datum $\boldsymbol{x} \in \mathcal{X}$ is encoded in the unitary $D_{\boldsymbol{x}}$ and V creates the fiducial state. This type of kernel can be used to learn the concept class (a) [18]. (c) An illustration of the landscape of a quantum kernel, $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ as a function of it's inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$. The landscape is described by the kernel variance, Equation (2), and is a measure of the performance of the quantum kernel [19].

and

$$D_{\boldsymbol{x}} |\psi\rangle = D_{\boldsymbol{c}_j \boldsymbol{s}} |\psi\rangle = D_{\boldsymbol{c}_j} D_{\boldsymbol{s}} |\psi\rangle = D_{\boldsymbol{c}_j} |\psi\rangle .$$

for all $\boldsymbol{x} \in \mathcal{X}$. By using this reference state, the classifier in [18] only needs to be able to distinguish between the m states

$$\rho_j = D_{\boldsymbol{c}_j} |\psi\rangle \langle\psi| D_{\boldsymbol{c}_j}^\dagger, \quad 1 \le j \le m.$$

2.2　Quantum kernels

Kernel methods are a powerful classical machine learning technique that enables learning in high-dimensional spaces via inner products, and quantum kernel methods extend this by using quantum computers to compute overlaps in exponentially large Hilbert spaces for greater expressivity. A *quantum kernel* is a function

$$\kappa(x, x') = |\langle \phi(x) | \phi(x') \rangle|^2$$

capturing the similarity between inputs x and x' in a quantum feature space. Here the classical data input points $x, x' \in \mathcal{X}$ are mapped onto quantum states $|\phi(x)\rangle, |\phi(x')\rangle$, respectively, in a Hilbert space $\mathcal{H}$, via the feature map

$$x \mapsto |\phi(x)\rangle = U_\phi(x)|0\rangle.$$

Here $U_\phi(x)$ is a parameterized quantum circuit dependent on the input x, and $|0\rangle$ is the all-zero reference state. Furthermore, the overlap between the two data points x, x' is computed by evaluating the following kernel function on a quantum computer

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = |\langle \phi(x') | \phi(x) \rangle|^2 .$$

In the specific case of learning data with group structure discussed above, one can use the *covariant* quantum kernel given in [18]

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \left| \langle 0^{\otimes N} | V^\dagger D_{\boldsymbol{x}}^\dagger D_{\boldsymbol{x}'} V | 0^{\otimes N} \rangle \right|^2 = \left| \langle \psi | D_{\boldsymbol{x}}^\dagger D_{\boldsymbol{x}'} | \psi \rangle \right|^2 .$$

Figure 1b illustrates the quantum circuit employed to evaluate this covariant kernel.

2.3　Exponential concentration and barren plateaus

As QML gains attention for its potential to outperform classical methods, a key challenge remains its limited trainability due to barren plateaus. For a model with loss function $l(\rho, O)$, defined as the variance of the expectation value of the observable O over the quantum state ρ, we say the method reaches a *barren plateau* when the variance of the loss over parameters $\boldsymbol{\theta}$ vanishes exponentially in the number of qubits N as

$$\mathrm{Var}_{\boldsymbol{\theta}} \left[\ell_{\boldsymbol{\theta}}(\rho, O) \right] \in \mathcal{O}\big(1/b^N\big),$$

3

for some constant $b > 1$ [37].

Furthermore, the exponential concentration of a quantum kernel $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ in the number of qubits N is defined as

$$\Pr\left[\left|\kappa(\boldsymbol{x}, \boldsymbol{x}') - \mu\right| \geq \delta\right] \leq \frac{\beta}{\delta^2}, \quad \beta \in \mathcal{O}(1/b^N) \tag{1}$$

$$\mu = \mathbb{E}_{\boldsymbol{x}, \boldsymbol{x}'}\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right]$$

for some $b > 1$ and where $\mathbb{E}_{\boldsymbol{x}, \boldsymbol{x}'}\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right]$ is the expectation value of the kernel [19]. The authors of [19] show that if the kernel landscape (illustrated in Figure 1c) becomes exponentially flat, meaning that if μ is exponentially small ($\mu \in \mathcal{O}(1/b^N)$), then the resulting model (after a polynomial number of measurements) does not depend on the input data.

There is an equivalence between barren plateaus in variational quantum circuits and the exponential concentration of quantum machine learning kernels [19, 16], showing that the presence of one implies the other. Finally, we note that Chebyshev's inequality allows us to cast the existence of exponential concentration (equation (1)) in terms of the variance of kernel values

$$\mathrm{Var}_{\boldsymbol{x}, \boldsymbol{x}'}\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right] \in \mathcal{O}(1/b'^N) \tag{2}$$

for some $b' > 1$ [19, 16]. It is this definition of exponential concentration that we will use for the rest of the paper.

3   Results

This section presents our main theoretical and numerical results. We derive variance expressions for covariant quantum kernels, analyze their robustness to errors coming from various sources of coherent noise, and identify conditions under which kernel concentration is avoided. Simulations support the theory across relevant regimes.

3.1   Variance of the coset quantum kernel

While such a kernel was already proposed in [18] for the case of two cosets, we extend the definition here so that it can be applied to datasets containing more than two cosets. In practice, each kernel evaluation still compares states from only two cosets at a time, but when multiple cosets are present in the data, the resulting kernel matrix naturally includes entries for all pairwise comparisons, leading to a larger matrix. For now, we consider the idealized case without noise. The impact of noise will be addressed in Section 3.3. By leveraging the properties of the fiducial state, the only kernel values are those that depend on the hidden group elements, $\boldsymbol{c}_i$:

$$\kappa(\boldsymbol{c}_i \boldsymbol{s}_a, \boldsymbol{c}_i \boldsymbol{s}_b) = \left|\langle \psi | D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_i} | \psi \rangle\right|^2 = 1$$

$$\kappa(\boldsymbol{c}_i \boldsymbol{s}_a, \boldsymbol{c}_j \boldsymbol{s}_b) = \left|\langle \psi | D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} | \psi \rangle\right|^2 =: \alpha_{i,j} = \alpha_{j,i} < 1 \quad i \neq j \tag{3}$$

for any $s_a, s_b \in S$.

Since each coset has the same number of elements, each kernel matrix will always have mn^2 occurrences of $\kappa = 1$ and $2n^2$ occurrences of each $\kappa = \alpha_{i,j}$ (after accounting for the fact that $\alpha_{i,j} = \alpha_{j,i}$), where $n = |S|$ and $m \geq 2$ the number of distinct left cosets. When considering the distribution of kernel values, we will exclude the mn trivial kernel values that appear in the diagonal, $\kappa(\boldsymbol{c}_i \boldsymbol{s}_a, \boldsymbol{c}_i \boldsymbol{s}_a) = 1$ [19]. Now the expectation value of the kernel value is

$$\mathbb{E}_{\boldsymbol{x}, \boldsymbol{x}'}\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right] = \frac{1}{(mn)^2 - mn}\left[m(n^2 - n) + 2n^2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \alpha_{i,j}\right]$$

$$= \frac{1}{m(mn - 1)}\left[m(n - 1) + 2n \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \alpha_{i,j}\right]$$

and the variance will be

$$\mathrm{Var}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right]$$

$$= \frac{1}{(mn)^2 - mn}\left[m(n^2-n)\left(1 - \mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right]\right)^2 + 2n^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\left(\alpha_{i,j} - \mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right]\right)^2\right]$$

$$= \frac{1}{m^2(mn-1)^2}\left[m(mn-1)\left(m(n-1) + 2n\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\alpha_{i,j}^2\right) - \left(m(n-1) + 2n\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\alpha_{i,j}\right)^2\right].$$

Since each $\alpha_{i,j}$ is smaller than 1, we note that neither the expected value nor the variance of the kernel is vanishing. This becomes clearer as we consider the large N limit below.

In the limit of large N, $\alpha_{i,j}$ can be approximated by its expected value taken over all possible choices of c_i, c_j, where these two group elements are sampled randomly according to some chosen measure on $\mathcal{G}$. Hence the expected value of $\alpha_{i,j}$ depends on both the measure chosen as well as on the unitary representation of $\mathcal{G}$. However, as noted above, for any nontrivial choice of this measure we have $\alpha_{i,j} < 1$ and hence the relevant quantities are nonvanishing. To make our argument more concrete, we consider the special case where c_i, c_j are sampled according to Haar measure defined on the N-qubit unitary group. For large N, we can reasonably approximate

$$\alpha_{i,j} \approx \frac{1}{2^N}$$

for all $i \neq j$ (see Appendix A.1 for details). Then,

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right] \approx \frac{1}{mn-1}\left((n-1) + n(m-1)\frac{1}{2^N}\right) \xrightarrow[N\to\infty]{} \frac{n-1}{mn-1}$$

and

$$\mathrm{Var}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right] \approx \frac{n(n-1)(m-1)}{(mn-1)^2}\left(1 - \frac{1}{2^N}\right)^2 \xrightarrow[N\to\infty]{} \frac{n(n-1)(m-1)}{(mn-1)^2}. \tag{4}$$

Finally, we note that in the case that $n \to \infty$ as $N \to \infty$, as is the case in the protocol proposed in [18] on which we base our numerical simulations, then

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right] \xrightarrow[N\to\infty]{} \frac{1}{m}$$

and

$$\mathrm{Var}_{\boldsymbol{x},\boldsymbol{x}'}\left[\kappa(\boldsymbol{x},\boldsymbol{x}')\right] \xrightarrow[N\to\infty]{} \frac{m-1}{m^2}. \tag{5}$$

That is, even if the data is partitioned into Haar-randomly chosen cosets, there is so much structure in the problem that this group-symmetric learning problem is free of exponential concentration. This is desirable because it ensures that the kernel retains the ability to meaningfully distinguish between different input data points. If the variance of the kernel values were to vanish, all inputs would appear nearly identical in the quantum feature space, rendering the kernel uninformative for learning tasks. This situation is analogous to the barren plateau problem in variational algorithms, where gradients vanish and optimization becomes ineffective. While this is an interesting theoretical result, it remains somewhat idealized. To move toward a more realistic setting, we introduce noise to the cosets in Section 3.3. Finally, we end this section with the following remark.

Remark 3.1. We note that the same results will hold if S is only a subset of a group and does not form a subgroup, provided that (i) a fiducial state can be found so that $s\,|\psi\rangle = |\psi\rangle$ for all $s \in S$ and (ii) each set $C_i = c_i S$ is disjoint so that if $x \in C_i$ then $x \notin C_j$ for all $j \neq i$.

3.2 Numerical results

Our analytical results establish the theoretical foundation for the absence of barren plateaus in the considered setting. We now show that these findings are further supported by numerical simulations, which confirm the predicted behavior across a range of system sizes and parameters.

We perform classical simulations of the protocol proposed in [18] for chain graph states of 2 to 10 qubits. We provide a review of this protocol in Appendix A.2. The group $\mathcal{G}$ under consideration comprises arbitrary single-qubit rotations on all N qubits and the subgroup S_{graph} is generated by the graph stabilizers of the form given in

equation (6). However, it is crucial to note here that rather than working with the full subgroup with $|S_{\text{graph}}| = 2^N$, we work with a subset of the same. That is, we take S to be the set of generators of S_{graph} so that $|S| = n = N$ (see equation (7)). This guarantees that our sets $C_i = c_i S$ are disjoint and the size of the dataset $\mathcal{X}$ does not scale unfavorably with N. Following Remark 3.1, our analytical results still hold for this modified setting, while at the same time keeping the computational cost down. The simulated values closely match the theoretical predictions for approximately $N > 6$, confirming the expected behavior in the asymptotic regime: Figure 2 shows that if neither errors nor kernel alignment are included, this problem reduces to finding the distance between two random states, and the variance (through basic simulation) appears to asymptote to $(m-1)/m^2 \in \mathcal{O}(1)$ (equation (5)) with increasing N, which is consistent with our theoretical results.
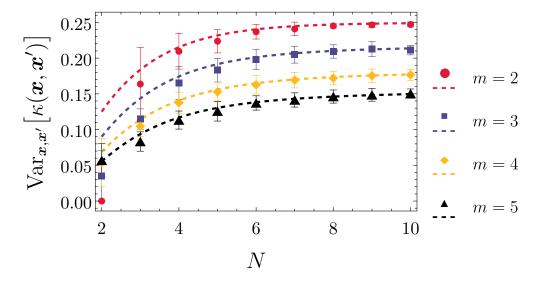

Figure 2: An average of 100 simulated values of variance in the kernel, $\text{Var}_{\boldsymbol{x},\boldsymbol{z}}\left[\kappa(\boldsymbol{x},\boldsymbol{z})\right]$, (dots) compared to the theoretical variance (dashed lines) as a function of the number of qubits, N, for various number of disjoint sets, m, for perfect encoding. The kernel is constructed from training data, which consists of half of the data, randomly selected for each trial, such that each coset is represented at least once. Error bars represent one standard deviation.

In Figure 3, we plot a heat map of the kernel matrix of two instances of our numerical simulations, where we show in both cases there remain many non-diagonal kernel entries where $\kappa = 1$. This highlights our main theoretical result, where the absence of kernel concentration results from the fact that the number $\left(\frac{n-1}{mn-1}\right)$ of off-diagonal kernel entries with $\kappa = 1$ approaches a constant $(1/m)$ as n becomes large.

A potential concern arises from the construction of the kernel using only the training data, which is subsequently applied to the testing data. If the dataset $\mathcal{X}$ is randomly partitioned, the training subset may not be uniformly distributed across all cosets. As a result, the variance will deviate from the expected $(m-1)/m^2$. However, our numerical simulations confirm that, on average, the variance remains consistent with theoretical expectations across random data splits, where training data consist of half the total data such that at least one element from each coset appears in the training data (Figure 2).

3.3 Extension to Noisy Systems

Simulating noise allows us to evaluate the robustness and practical viability of quantum algorithms. In this section, we focus on coherent errors as a first step to showing robustness under realistic operating conditions.

We assume that any errors will take the form

$$\kappa(\boldsymbol{x}_{a,i}, \boldsymbol{x}_{b,i}) = 1 - \gamma_{ab,i}$$
$$\kappa(\boldsymbol{x}_{a,i}, \boldsymbol{x}_{b,j}) = \alpha + \delta_{ab,ij},$$

where $x_{a,i} := c_a s_i$, $\gamma_{ab,i} > 0$ and $\delta_{ab,ij} \in \mathbb{R}$ and we assume that $\alpha_{i,j} = \alpha$ for all $1 \leq i, j \leq m$ with $i \neq j$. For notational convenience, we vectorize $\boldsymbol{\gamma} \in \mathbb{R}^{mn(n-1)}$ and $\boldsymbol{\delta} \in \mathbb{R}^{m(m-1)n^2}$.
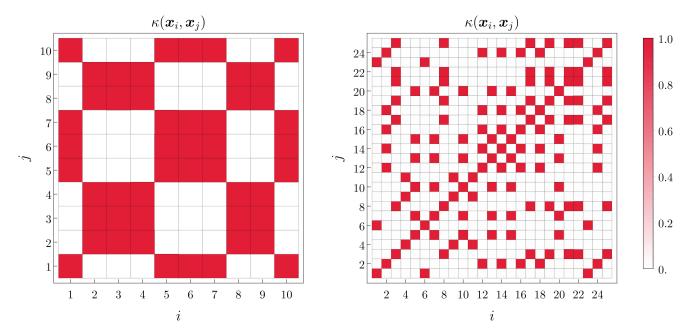
Figure 3: A heat map showing the kernel values for one simulation of $N = 10$ qubits for (*left*) $m = 2$ and (*right*) $m = 5$ disjoint subsets. In both cases, there are many off-diagonal kernel entries where $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = 1$ (shown in red) and do not approach the identity matrix. Note that the white entries are not exactly zero, but correspond to very small values ($< 5.3 \times 10^{-3}$).

This yields an expected kernel value of

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}\big[\kappa(\boldsymbol{x},\boldsymbol{x}')\big] = \frac{n-1}{mn-1}\big(1 - \mathbb{E}[\boldsymbol{\gamma}]\big) + \frac{n(m-1)}{mn-1}\big(\alpha - \mathbb{E}[\boldsymbol{\delta}]\big)$$

where $\mathbb{E}[\boldsymbol{\gamma}]$ and $\mathbb{E}[\boldsymbol{\delta}]$ are the expected values of the $\gamma_{ab,i}$'s and the $\delta_{ab,ij}$'s, respectively, and a variance of

$$\mathrm{Var}_{\boldsymbol{x},\boldsymbol{x}'}\big[\kappa(\boldsymbol{x},\boldsymbol{x}')\big] = \frac{n(m-1)(n-1)}{(mn-1)^2}\Big(\big(1 - \mathbb{E}[\boldsymbol{\gamma}]\big) - \big(\alpha + \mathbb{E}[\boldsymbol{\delta}]\big)\Big)^2 + \frac{(n-1)\,\mathrm{Var}[\boldsymbol{\gamma}] + n(m-1)\,\mathrm{Var}[\boldsymbol{\delta}]}{mn-1}$$

where $\mathrm{Var}[\boldsymbol{\gamma}]$ and $\mathrm{Var}[\boldsymbol{\delta}]$ are the variances in the $\gamma_{ab,i}$'s and the $\delta_{ab,ij}$'s, respectively.

With that we establish a lower bound on the variance, showing that as long as $1 - \mathbb{E}[\boldsymbol{\gamma}]$ remains sufficiently large, the kernel does not concentrate. Intuitively, this is expected: as long as $1 - \gamma$ remains, on average, well-separated from $\alpha + \delta$, the variance remains non-negligible, preventing concentration of the kernel.

In the following sections, we examine how different sources of coherent noise affect the kernel values, focusing on errors in the fiducial state, the unitary representation of the group elements, and the group element selection process. In all three cases, we assume that the evaluated kernel values are still the result of the overlap of pure states, $\kappa(\boldsymbol{x}, \boldsymbol{x}') = |\langle \tilde{\phi}_{\boldsymbol{x}} | \tilde{\phi}_{\boldsymbol{x}'} \rangle|^2$ and therefore any noise will map pure states to pure states. More specifically, given that the initial state $|0^{\otimes N}\rangle \in \mathcal{H}$, we will assume that any operator $T : \mathcal{H} \to \mathcal{H}^*$ (where $\mathcal{H}^*$ is the dual to $\mathcal{H}$) that acts on $|0^{\otimes N}\rangle$, including the operators that describe coherent noise, is a bounded linear operator, $T \in B(\mathcal{H})$. We use the operator norm (Appendix A.3) as a metric on $B(\mathcal{H})$.

3.3.1 Errors in the Fiducial State

Suppose there is an implementation error in the operator V, which prepares the fiducial state $|\psi\rangle$ from the all-zero state $|0\rangle^{\otimes N}$. That is, there exists bounded linear operators $W_k \in B(\mathcal{H})$ is such that

$$\|V - W_k\|_{\mathrm{op}} \leq \epsilon.$$

Then, for any indices k, ℓ and group elements strings $\boldsymbol{c}_i, \boldsymbol{c}_j, \boldsymbol{s}_a, \boldsymbol{s}_b$, it follows that (to order ϵ^2)

$$\left| \Big\langle 0^{\otimes N} \,\Big|\, W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_i} D_{\boldsymbol{s}_b} W_\ell \,\Big|\, 0^{\otimes N} \Big\rangle \right|^2 \geq 1 - 4\epsilon + 2\epsilon^2$$

$$\alpha - 4\sqrt{\alpha}\epsilon + 2(2 - \sqrt{\alpha})\epsilon^2 \leq \left| \Big\langle 0^{\otimes N} \,\Big|\, W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b} W_\ell \,\Big|\, 0^{\otimes N} \Big\rangle \right|^2 \leq \alpha + 4\sqrt{\alpha}\epsilon + 2(2 + \sqrt{\alpha})\epsilon^2.$$

Appendix A.4 contains the complete derivation and Appendix A.5 discusses the simulations of the errors.

3.3.2 Error in the unitary representation of the elements of the group

Now, we assume that there is an error in the unitary operator $D_{\boldsymbol{x}_{a,i}}$ such that

$$\left\| D_{\boldsymbol{x}_{a,i}} - A_{a,i} \right\|_{\mathrm{op}} = \left\| D_{\boldsymbol{s}_a} D_{c_i} - A_{a,i} \right\|_{\mathrm{op}} \leq \epsilon$$

where $A_{a,i} \in B(\mathcal{H})$.

We can show

$$\left| \left\langle \psi \left| A_{a,i}^\dagger A_{b,i} \right| \psi \right\rangle \right|^2 \geq 1 - 4\epsilon + 2\epsilon^2$$

$$\alpha - 4\sqrt{\alpha}\epsilon + 2(2 - \sqrt{\alpha})\epsilon^2 \leq \left| \left\langle \psi \left| A_{a,i}^\dagger A_{b,j} \right| \psi \right\rangle \right|^2 \leq \alpha + 4\sqrt{\alpha}\epsilon + 2(2 + \sqrt{\alpha})\epsilon^2.$$

The derivation of the bound is very similar to that for errors in the fiducial state and can be found in the Appendix A.6.

3.3.3 Error in the selection of the group elements

Assume there exists a random perturbation satisfying

$$d(\boldsymbol{e}_{a,i}, \mathbb{1}) := \left\| D_{\boldsymbol{e}_{a,i}} - \mathbb{1} \right\|_{\mathrm{op}} < \epsilon,$$

such that the unitary operator $D_{\boldsymbol{x}_{a,i}}$ admits the decomposition

$$D_{\boldsymbol{x}_{a,i}} = D_{\boldsymbol{e}_{a,i}} D_{\boldsymbol{q}_{a,i}} = D_{\boldsymbol{e}_{a,i}} D_{\boldsymbol{c}_i} D_{\boldsymbol{s}_a},$$

as described in [18].

We prove to order ϵ^2 that

$$\left| \left\langle \psi \left| D_{\boldsymbol{c}_\pm}^\dagger D_{\boldsymbol{e}_i}^\dagger D_{\boldsymbol{e}_j} D_{\boldsymbol{c}_\pm} \right| \psi \right\rangle \right|^2 \geq 1 - \epsilon^2$$

$$\alpha - 4\sqrt{\alpha}\epsilon + 4\epsilon^2 \leq \left| \left\langle \psi \left| D_{\boldsymbol{c}_\pm}^\dagger D_{\boldsymbol{e}_k}^\dagger D_{\boldsymbol{e}_\ell} D_{\boldsymbol{c}_\mp} \right| \psi \right\rangle \right|^2 \leq \alpha + 4\sqrt{\alpha}\epsilon + 4\epsilon^2.$$

See Appendix A.7 for the derivation and Appendix A.8 for a discussion on the simulation of these errors.

3.3.4 Numerical simulation of errors

The theoretical errors calculated in the previous subsections represent the worst-case scenarios. In order to explore a slightly more experimentally relevant scenario, we numerically simulate errors in the evaluation of the quantum kernel, which are shown in Figure 4. Once again following the protocol proposed in [18] for chain graph states of 2 to 10 qubits, we simulate errors in the fiducial state (left plot) and perturbations in the group elements (right plot) with errors bounded by 0.9 in the operator norm (see Appendices A.5 and A.8 for details). We find that although the error is large enough for the variance to show visible deviations from the ideal value, the kernel does not concentrate; its variance approaches a constant value as the number of qubits (N) increases. Although this constant value is less than the ideal theoretical value given by Equation (5), it is much larger than the value given by the worst-case theoretical values above. We also find that errors in the fiducial state lead to a larger reduction of variance than perturbations in the group elements even though the allowed range of errors is the same in both cases. This is in line with the theoretical calculations, where the latter case results in a smaller error in the $\kappa(x_{\boldsymbol{a,i}}, x_{\boldsymbol{b,i}})$ kernel elements. Finally we note, that as we do not simulate learning with these kernels, only the distribution of their entries, we do not comment on the effect of the simulated noise on their learning performance.

4 Discussion

Our results highlight the power of quantum kernel methods in learning data with an underlying group structure, where classical approaches may struggle. We analytically demonstrated that the associated quantum kernel does
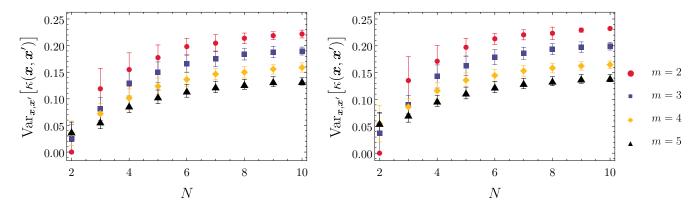
Figure 4: An average of 100 simulated values of variance in the kernel with errors, $\text{Var}_{\boldsymbol{x},\boldsymbol{z}}\left[\kappa(\boldsymbol{x},\boldsymbol{z})\right]$, as a function of the number of qubits, N, for various number of disjoint sets, m, with imperfect kernel evaluation. The *left* figure simulates an error in the fiducial state with $\|V - W_k\|_{\text{op}} \leq 0.9$ and the *right* figure simulates an error of the group elements with a perturbation of $\|D_{\boldsymbol{e}_{a,i}} - \mathbb{1}\|_{\text{op}} \leq 0.9$. The kernel is constructed from training data, which consists of half of the data, randomly selected for each trial, such that each coset is represented at least once. In both cases, the variance still approaches a positive constant with increasing number of qubits, $\mathcal{O}(1)$, showing the absence of kernel concentration. Error bars represent one standard deviation.

not suffer from barren plateau phenomena. This stands in contrast to many variational quantum algorithms, where exponentially vanishing gradients can severely limit translatability as system size increases.

In addition, we rigorously analyzed the impact of noise on our method. We found that this quantum kernel construction is remarkably robust to common types of coherent noise—both analytically and through extensive numerical simulations. This robustness further strengthens the case for implementing our approach on near-term noisy intermediate-scale quantum (NISQ) devices. Unlike many quantum algorithms whose performance degrades significantly under realistic noise models, our method continues to perform reliably, demonstrating both stability and resilience. To better understand the robustness of the method, we examine how different noise sources influence the kernel values. These include errors in the preparation of the fiducial state, inexact unitary implementations, and deviations in the selection of group elements.

An additional strength of our kernel family lies in its connection to earlier constructions. It generalizes a kernel previously shown to learn the concept class $\mathcal{C}$, which was based on the discrete logarithm problem (DLP). Since $\mathcal{C}$ inherits the classical hardness of the DLP, this places our construction in a particularly interesting category: quantum models that can efficiently learn tasks tied to classically intractable problems.

Together, these results position our quantum kernel method as a theoretically sound, resistant to coherent noise, and practically viable approach to learning structured data with quantum-enhanced models.

Competing interests: All authors declare no financial or non-financial competing interests.
Data and code availability: All scripts and data are available upon reasonable request and will be made publicly available at time of publishing.

References

[1] E. Aïmeur, G. Brassard, and S. Gambs, "Machine Learning in a Quantum World," in *Advances in Artificial Intelligence* (L. Lamontagne and M. Marchand, eds.), (Berlin, Heidelberg), pp. 431–442, Springer Berlin Heidelberg, 2006.

[2] G. D. Paparo, V. Dunjko, A. Makmal, M. A. Martin-Delgado, and H. J. Briegel, "Quantum Speedup for Active Learning Agents," *Physical Review X*, vol. 4, no. 3, p. 031002, 2014.

[3] M. Schuld, I. Sinayskiy, and F. Petruccione, "The quest for a Quantum Neural Network," *Quantum Inf. Process.*, vol. 13, no. 11, pp. 2567–2586, 2014.

[4] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum Perceptron Models," 2016.

[5] N. B. Lovett, C. Crosnier, M. Perarnau-Llobet, and B. C. Sanders, "Differential Evolution for Many-Particle Adaptive Quantum Metrology," *Phys. Rev. Lett.*, vol. 110, no. 22, p. 220501, 2013.

[6] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, no. 6325, pp. 602–606, 2017.

[7] M. Tiersch, EJ. Ganahl, and H. J. Briegel, "Adaptive quantum computation in changing environments using projective simulation," *Scientific reports*, vol. 5, no. 1, pp. 1–18, 2015.

[8] U. Alvarez-Rodriguez, L. Lamata, P. Escandell-Montero, J. D. Martín-Guerrero, and E. Solano, "Supervised quantum learning without measurements," *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017.

[9] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, "Quantum boltzmann machine," *Physical Review X*, vol. 8, p. 021050, May 2018.

[10] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, "Training deep quantum neural networks," *Nature Communications*, vol. 11, no. 1, pp. 1–6, 2020.

[11] F. Bach, "Information Theory With Kernel Methods," *IEEE Trans. Inform. Theory*, vol. 69, pp. 752–775, Feb. 2023.

[12] M. Schuld, "Supervised quantum machine learning models are kernel methods," Jan. 2021.

[13] S. Jerbi, L. J. Fiderer, H. Poulsen Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, "Quantum machine learning beyond kernel methods," *Nat Commun*, vol. 14, p. 517, Jan. 2023.

[14] A. Deshpande, M. Hinsche, S. Najafi, K. Sharma, R. Sweke, and C. Zoufal, "Dynamic parameterized quantum circuits: Expressive and barren-plateau free," Nov. 2024.

[15] C. Wood, S. Shrapnel, and G. J. Milburn, "A Kerr kernel quantum learning machine," Apr. 2024.

[16] P. Kairon, J. Jäger, and R. V. Krems, "Equivalence between exponential concentration in quantum machine learning kernels and barren plateaus in variational algorithms," Jan. 2025.

[17] Y. Liu, S. Arunachalam, and K. Temme, "A rigorous and robust quantum speed-up in supervised machine learning," *Nat. Phys.*, vol. 17, pp. 1013–1017, Sept. 2021.

[18] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, "Covariant quantum kernels for data with group structure," *Nat. Phys.*, vol. 20, pp. 479–483, Mar. 2024.

[19] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, "Exponential concentration and untrainability in quantum kernel methods," Aug. 2022.

[20] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nat Commun*, vol. 9, p. 4812, Nov. 2018.

[21] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nat Commun*, vol. 12, p. 1791, Mar. 2021.

[22] 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*, vol. 12, p. 6961, Nov. 2021.

[23] M. Cerezo and P. J. Coles, "Higher order derivatives of quantum neural networks with barren plateaus," *Quantum Sci. Technol.*, vol. 6, p. 035006, June 2021.

[24] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, "Effect of barren plateaus on gradient-free optimization," *Quantum*, vol. 5, p. 558, Oct. 2021.

[25] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, "Barren Plateaus Preclude Learning Scramblers," *Phys. Rev. Lett.*, vol. 126, p. 190501, May 2021.

[26] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, "Entanglement-Induced Barren Plateaus," *PRX Quantum*, vol. 2, p. 040316, Oct. 2021.

[27] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, "Entanglement devised barren plateau mitigation," *Phys. Rev. Research*, vol. 3, p. 033090, July 2021.

[28] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, "Absence of Barren Plateaus in Quantum Convolutional Neural Networks," *Phys. Rev. X*, vol. 11, p. 041011, Oct. 2021.

[29] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, "Equivalence of quantum barren plateaus to cost concentration and narrow gorges," *Quantum Sci. Technol.*, vol. 7, p. 045015, Aug. 2022.

[30] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, "Trainability of Dissipative Perceptron-Based Quantum Neural Networks," *Phys. Rev. Lett.*, vol. 128, p. 180505, May 2022.

[31] G. Verdon, M. Broughton, J. R. McClean, K. J. Sung, R. Babbush, Z. Jiang, H. Neven, and M. Mohseni, "Learning to learn with quantum neural networks via classical neural networks," July 2019.

[32] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, "An initialization strategy for addressing barren plateaus in parametrized quantum circuits," *Quantum*, vol. 3, p. 214, Dec. 2019.

[33] A. V. Uvarov and J. D. Biamonte, "On barren plateaus and cost function locality in variational quantum algorithms," *J. Phys. A: Math. Theor.*, vol. 54, p. 245301, May 2021.

[34] T. Volkoff and P. J. Coles, "Large gradients via correlation in random parameterized quantum circuits," *Quantum Sci. Technol.*, vol. 6, p. 025008, Jan. 2021.

[35] M. Bilkis, M. Cerezo, G. Verdon, P. J. Coles, and L. Cincio, "A semi-agnostic ansatz with variable structure for quantum machine learning," *Quantum Mach. Intell.*, vol. 5, p. 43, Dec. 2023.

[36] A. Sannia, F. Tacchino, I. Tavernelli, G. L. Giorgi, and R. Zambrini, "Engineered dissipation to mitigate barren plateaus," Oct. 2023.

[37] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, "A Lie Algebraic Theory of Barren Plateaus for Deep Parameterized Quantum Circuits," *Nat Commun*, vol. 15, p. 7172, Aug. 2024.

[38] D. Petz and J. Réffy, "On asymptotics of large Haar distributed unitary matrices," *Periodica Mathematica Hungarica*, vol. 49, pp. 103–117, Sept. 2004.

[39] R. Bhatia, *Matrix Analysis*, vol. 169 of *Graduate Texts in Mathematics*. New York, NY: Springer New York, 1997.

[40] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 1 ed., Apr. 2018.

A   Appendix

A.1   Discussion of the $\alpha_{i,j}$

In the following we discuss the value of $\alpha_{i,j}$, the overlap between two random pure states defined in Equation (3) in Section 3.1, where we derive the variance expressions for covariant quantum kernels.

For that, let $|0\rangle$ state be represented by the column vector $(1,0,\ldots,0)^T$. Any vector distributed according to Haar measure can be written as $|\phi\rangle := U|0\rangle$ and $|\psi\rangle := V|0\rangle$ for some Haar-randomly sampled unitaries U,V. Now

$$
\mathbb{E}_{\phi,\varphi\sim\mathbb{C}^{2^N}} = \left[\,|\langle\phi\,|\,\varphi\rangle|^2\,\right] = \mathbb{E}_{U,V\sim\mathbb{C}^{2^N\times2^N}} = \left[\,|\langle 0\,|\,U^\dagger V\,|\,0\rangle|^2\,\right].
$$

Note that Haar measure is invariant under left and right multiplication, therefore $W := U^\dagger V$ is yet another Haar random unitary. We can now rewrite the overlap as

$$
\mathbb{E}_{\phi,\varphi\sim\mathbb{C}^{2^N}} = \mathbb{E}_{W\sim\mathbb{C}^{2^N\times2^N}}\,\left|W_{11}^2\right| = \int_0^1 (2^N-1)w(1-w)^{2^N-2}\mathrm{d}w = \frac{1}{2^N},
$$

since the entry $|W_{11}|^2$ of a Haar-random unitary matrix [38] $W\in\mathbb{C}^{d\times d}$ for $d := 2^N$ is distributed according to a Beta distribution:

$$
|W_{11}|^2 \sim \mathrm{Beta}(1,d-1).
$$

This yields the probability density function

$$
P(w) = (d-1)(1-w)^{d-2}, \quad w\in[0,1]\,.
$$

This result follows from the fact that the columns of a Haar-random unitary are uniformly distributed unit vectors in $\mathbb{C}^d$, and the squared moduli of their components follow a Dirichlet distribution. The marginal distribution of a single component squared modulus is therefore Beta-distributed.

The expected value of a beta distribution is

$$
\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}.
$$

With parameters $\alpha = 1$ and $\beta = d-1$, it is reasonable to approximate, in the limit of a large number of qubits, that

$$
\alpha_{i,j} \approx \frac{1}{2^N}
$$

for all $i\neq j$.

A.2   Classifying cosets of $\mathrm{SU}(2)^{\otimes N}$

In this section we briefly review the protocol proposed in [18] which we use to generate the numerical simulation shown in Figures 2 and 4.

The chosen group is $\mathcal{G} = \mathrm{SU}(2)^{\otimes N}$, which is the group of single qubit rotations on N qubits. The natural unitary representation is a rotation $D(\theta_1,\theta_2,\theta_3)$ on each qubit where θ_1, θ_2, and θ_3 are Euler angles

$$
D(\theta_1,\theta_2,\theta_3) = \exp\left(-\frac{\mathrm{i}\theta_1}{2}X\right)\exp\left(-\frac{\mathrm{i}\theta_2}{2}Z\right)\exp\left(-\frac{\mathrm{i}\theta_3}{2}X\right) = R_x(\theta_1)R_z(\theta_2)R_x(\theta_3).
$$

The choice of subgroup is a Pauli stablizer graph state on N qubits. For a graph $G = (E,V)$, the graph-stabilizer is

$$
S_{\mathrm{graph}} = \left\langle\left\{X_j\bigotimes_{k:(j,k)\in E}Z_k\right\}_{j\in V}\right\rangle. \tag{6}
$$

The graph state of the same graph $G = (E,V)$ is

$$
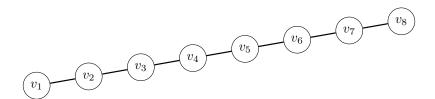|G\rangle = \prod_{(j,k)\in E}\mathrm{CZ}_{(j,k)}\,|+\rangle^{\otimes V}
$$

Figure 5: A chain graph of 8 vertices.

where $\mathrm{CZ}_{(j,k)}$ is the controlled-Z interaction between qubits j and k. The state $|G\rangle$ is the simultaneous $(+1)$-eigenvalue eigenvector of all of the elements of S_{graph}:

$$s\,|G\rangle = |G\rangle \quad \text{for all } s \in S_{\mathrm{graph}}.$$

Therefore, $|G\rangle$ matches the properties we require for a fiducial state.

For simplicity, we choose the chain graph (Figure 5), so our subgroup is the graph-stabilizer group whose generators are given by

$$S := \left\{ X_1 Z_2 \mathbb{1}_3 \cdots \mathbb{1}_N, Z_1 X_2 Z_3 \cdots \mathbb{1}_N, \mathbb{1}_1 Z_2 X_3 Z_4 \cdots \mathbb{1}_N, \ldots, \mathbb{1}_1 \mathbb{1}_2 \mathbb{1}_3 \cdots Z_{N-1} X_N \right\}. \tag{7}$$

For this particular choice of graph-stabilizer group, the fiducial state is

$$|\psi\rangle = \mathrm{CZ}_{(1,2)} \mathrm{CZ}_{(2,3)} \mathrm{CZ}_{(3,4)} \cdots \mathrm{CZ}_{(N-1,N)} \left(R_y \left(\frac{\pi}{2} \right) |0\rangle \right)^{\otimes N} \tag{8}$$

and the full quantum circuit for calculating the kernel value $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ in this case is given by Figure 6.

Finally, we note that for ease of numerical simulations, we work with the generators of the group (Equation (7)) to form the set S, so that $n = N$, rather than the full group S_{graph}, which has 2^N elements. As a consequence, our simulated data consists of the union of m disjoint sets, $\mathcal{X} = \bigcup_{i=1}^{m} \boldsymbol{c}_j S$, rather than m cosets. To see that the sets are indeed disjoint, consider the full subgroup, S_{graph} and note that $S \subset S_{\mathrm{graph}}$. So, if $\boldsymbol{x} \in \boldsymbol{c}_i S$ then $\boldsymbol{x} \in \boldsymbol{c}_i S_{\mathrm{graph}}$. Since cosets are disjoint then $\boldsymbol{x} \notin \boldsymbol{c}_j S_{\mathrm{graph}}$ for all $j \neq i$, and therefore $\boldsymbol{x} \notin \boldsymbol{c}_j S$. Additionally, given the fiducial state Equation (8), then

$$\boldsymbol{s}\,|\psi\rangle = |\psi\rangle \text{ for all } \boldsymbol{s} \in S_{\mathrm{graph}} \implies \boldsymbol{s}\,|\psi\rangle = |\psi\rangle \text{ for all } \boldsymbol{s} \in S.$$

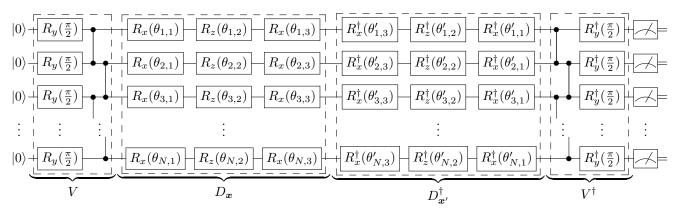Therefore, our simulated kernel will have its entries given by Equation (3).

Figure 6: The specific circuit of the quantum kernel proposed in [18] to classify cosets of $SU(2)^{\otimes N}$.

A.3 Properties of Operator Norms

Here, we present some properties of the operator norm which we will use in calculations in later appendices.

Let $A : V \to W$ be a linear operator between normed spaces. The *operator norm* of A is (for $V \neq \{0\}$)

$$\|A\|_{\text{op}} = \sup \left\{ \frac{\|Av\|}{\|v\|} \ : \ v \neq 0 \text{ and } v \in V \right\}$$

where $\|v\|$ is the norm in V and $\|Av\|$ is the norm in W [39].

Statement: The first useful observation of the operator norm is that if $\|A - B_1\|_{\text{op}} \leq \epsilon$ and $\|A - B_2\|_{\text{op}} \leq \epsilon$ then

$$\|B_2 - B_1\|_{\text{op}} \leq 2\epsilon. \tag{9}$$

Proof:

$$\begin{aligned}
\|B_1 - B_2\|_{\text{op}} &= \|B_1 - A + A - B_2\|_{\text{op}} \\
&\leq \|B_1 - A\|_{\text{op}} + \|A - B_2\|_{\text{op}} \\
&\leq \epsilon + \epsilon \\
&= 2\epsilon.
\end{aligned}$$

Statement: Next, notice that if $\|A_1 - B_1\|_{\text{op}} \leq \epsilon$ and $\|A_2 - B_2\|_{\text{op}} \leq \epsilon$ then

$$\|A_1 A_2 - B_1 B_2\|_{\text{op}} \leq \epsilon \min \left\{ \|A_1\|_{\text{op}} + \|B_2\|_{\text{op}} , \|A_2\|_{\text{op}} + \|B_1\|_{\text{op}} \right\}. \tag{10}$$

Proof [40]:

$$\begin{aligned}
\|A_1 A_2 - B_1 B_2\|_{\text{op}} &= \|A_1 A_2 - A_1 B_2 + A_1 B_2 - B_1 B_2\|_{\text{op}} \\
&\leq \|A_1 A_2 - A_1 B_2\|_{\text{op}} + \|A_1 B_2 - B_1 B_2\|_{\text{op}} \\
&\leq \|A_1\|_{\text{op}} \|A_2 - B_2\|_{\text{op}} + \|B_2\|_{\text{op}} \|A_1 - B_1\|_{\text{op}} \\
&\leq \epsilon \big(\|A_1\|_{\text{op}} + \|B_2\|_{\text{op}} \big).
\end{aligned}$$

A similar argument can be used to show that

$$\|A_1 A_2 - B_1 B_2\|_{\text{op}} \leq \epsilon \big(\|A_2\|_{\text{op}} + \|B_1\|_{\text{op}} \big).$$

Combining these two bounds gives Equation (10).

Statement: Finally, notice that $\|A - B\|_{\text{op}} \leq \epsilon$ then

$$\|A\|_{\text{op}} - \epsilon \leq \|B\|_{\text{op}} \leq \|A\|_{\text{op}} + \epsilon, \tag{11}$$

and in particular if A is unitary so $\|A\|_{\text{op}} = 1$, then

$$1 - \epsilon \leq \|B\|_{\text{op}} \leq 1 + \epsilon. \tag{12}$$

Proof:

$$\epsilon \geq \|A - B\|_{\text{op}} \geq \Big| \|A\|_{\text{op}} - \|B\|_{\text{op}} \Big| \quad \text{(Reverse triangle inequality)}.$$

Statement: Next, we write down three relationships between the operator norm and the overlap between states (vectors in an an inner product space of unit norm). The first relationship is

$$\|A\|_{\text{op}} \geq |\langle \phi \, | \, A \, | \, \psi \rangle| \tag{13}$$

for states $|\psi\rangle$ and $|\phi\rangle$.

Proof: Recall that the induced norm of an inner product space is $\|v\| = \sqrt{\langle v \, | \, v \rangle}$. Therefore, we can bound

$$\begin{aligned}
\|A\|_{\text{op}} &\geq \frac{\|A \, |\psi\rangle \|}{\| \, |\psi\rangle \, \|} \\
&= \|A \, |\psi\rangle \| \\
&= \|A \, |\psi\rangle \| \, \| \, |\phi\rangle \, \| \\
&\geq |\langle \psi \, | \, A \, | \, \phi \rangle| \quad \text{(Cauchy-Schwartz)}.
\end{aligned}$$

Statement: The second relationship states that if $\|A - B\|_{\text{op}} \leq \delta$ then

$$\left| |\langle \phi | A | \psi \rangle| - |\langle \phi | B | \psi \rangle| \right| \leq \delta. \tag{14}$$

Proof:

$$
\begin{aligned}
\delta &\geq \|A - B\|_{\text{op}} \\
&\geq |\langle \phi | (A - B) | \psi \rangle| \\
&= \left| \langle \phi | A | \psi \rangle - \langle \phi | B | \psi \rangle \right| \\
&\geq \left| |\langle \phi | A | \psi \rangle| - |\langle \phi | B | \psi \rangle| \right| \quad \text{(Reverse triangle inequality)}.
\end{aligned}
$$

Statement: The third relationship states that if $\|U_1 - U_2\|_{\text{op}} \leq \delta$ for unitary operators U_1 and U_2, then for any state $|\psi\rangle$

$$\left| \left\langle \psi \left| U_1^\dagger U_2 \right| \psi \right\rangle \right| \geq 1 - \delta^2/2. \tag{15}$$

Proof:

$$
\begin{aligned}
\delta^2 &\geq \|U_1 - U_2\|_{\text{op}}^2 \\
&\geq \|(U_1 - U_2) |\psi\rangle \|^2 \\
&= \left\langle (U_1 - U_2)\psi \middle| (U_1 - U_2)\psi \right\rangle \\
&= \left\langle \psi \left| U_1^\dagger U_1 \right| \psi \right\rangle - \left\langle \psi \left| U_1^\dagger U_2 \right| \psi \right\rangle - \left\langle \psi \left| U_2^\dagger U_1 \right| \psi \right\rangle + \left\langle \psi \left| U_2^\dagger U_2 \right| \psi \right\rangle \\
&= 2\left(1 - \text{Re}\left[\left\langle \psi \left| U_1^\dagger U_2 \right| \psi \right\rangle \right] \right) \\
&\geq 2\left(1 - \left| \left\langle \psi \left| U_1^\dagger U_2 \right| \psi \right\rangle \right| \right) \quad \text{since } \text{Re}[z] \leq |z| \implies -\text{Re}[z] \geq -|z|.
\end{aligned}
$$

Rearranging gives Equation (15).

A.4 Errors in the Fiducial State

In the following we derive the bound discussed in Section 3.3.1 of the main paper. For this, we assume that there is an error in the operator V which takes the all zero state $|0^{\otimes N}\rangle$ to the fiducial state $|\psi\rangle$ such that

$$\|V - W\|_{\text{op}} \leq \epsilon.$$

If this is the only error, then each kernel element will take the form

$$\left| \left\langle 0^{\otimes N} \left| W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b} W_\ell \right| 0^{\otimes N} \right\rangle \right|^2 = \left| \left\langle 0^{\otimes N} \left| W_k^\dagger U W_\ell \right| 0^{\otimes N} \right\rangle \right|^2$$

where $U = D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b}$. We also note that in general $W_k \neq W_\ell$, such as if the kernel is being implemented via the SWAP test [19].

We can bound

$$
\begin{aligned}
&\left| |\langle 0^{\otimes N} | V^\dagger U V | 0^{\otimes N} \rangle| - \left| \left\langle 0^{\otimes N} \left| W_k^\dagger U W_\ell \right| 0^{\otimes N} \right\rangle \right| \right| \\
&\quad \leq \left| \left\langle 0^{\otimes N} | V^\dagger U V | 0^{\otimes N} \right\rangle - \left\langle 0^{\otimes N} \left| W_k^\dagger U W_\ell \right| 0^{\otimes N} \right\rangle \right| \quad \text{(Reverse triangle inequality)} \\
&\quad = \left| \left\langle 0^{\otimes N} \left| \left(V^\dagger U V - W_k^\dagger U W_\ell \right) \right| 0^{\otimes N} \right\rangle \right| \\
&\quad \leq \left\| V^\dagger U V - W_k^\dagger U W_\ell \right\|_{\text{op}} \quad \text{by Equation (13)} \\
&\quad \leq \epsilon \left(\|V^\dagger U\|_{\text{op}} + \|W_\ell\|_{\text{op}} \right) \quad \text{by Equation (10)} \\
&\quad = \epsilon \left(1 + \|W_\ell\|_{\text{op}} \right) \\
&\quad \leq \epsilon(1 + 1 + \epsilon) \quad \text{by Equation (12)} \\
&\quad = 2\epsilon + \epsilon^2.
\end{aligned}
$$

Therefore, in the case of $\boldsymbol{c}_i = \boldsymbol{c}_j$, where $\left|\left\langle 0^{\otimes N}\left|V^\dagger U V\right|0^{\otimes N}\right\rangle\right| = 1$, we have

$$\left|\left\langle 0^{\otimes N}\left|W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_i} D_{\boldsymbol{s}_b} W_\ell\right|0^{\otimes N}\right\rangle\right| \geq 1 - 2\epsilon - \epsilon^2$$

$$\implies \left|\left\langle 0^{\otimes N}\left|W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_i} D_{\boldsymbol{s}_b} W_\ell\right|0^{\otimes N}\right\rangle\right|^2 \geq 1 - 4\epsilon + 2\epsilon^2 + 4\epsilon^3 + \epsilon^4$$

and in the case of $\boldsymbol{c}_i \neq \boldsymbol{c}_j$, where $\left|\left\langle 0^{\otimes N}\left|V^\dagger U V\right|0^{\otimes N}\right\rangle\right| = \sqrt{\alpha}$, we have

$$\sqrt{\alpha} - 2\epsilon - \epsilon^2 \leq \left|\left\langle 0^{\otimes N}\left|W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b} W_\ell\right|0^{\otimes N}\right\rangle\right| \leq \sqrt{\alpha} + 2\epsilon + \epsilon^2$$

$$\implies \alpha - 4\sqrt{\alpha}\epsilon + 2(2 - \sqrt{\alpha})\epsilon^2 + 4\epsilon^3 + \epsilon^4 \leq \left|\left\langle 0^{\otimes N}\left|W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b} W_\ell\right|0^{\otimes N}\right\rangle\right|^2 \leq \alpha + 4\sqrt{\alpha}\epsilon + 2(2 + \sqrt{\alpha})\epsilon^2 + 4\epsilon^3 + \epsilon^4.$$
$$(16)$$

Finally, notice if $\alpha \to 0$, then Equation (16) becomes

$$\left|\left\langle 0^{\otimes N}\left|W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b} W_\ell\right|0^{\otimes N}\right\rangle\right| \leq 2\epsilon + \epsilon^2 \implies \left|\left\langle 0^{\otimes N}\left|W_k^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b} W_\ell\right|0^{\otimes N}\right\rangle\right|^2 \leq 4\epsilon^2 + 4\epsilon^3 + \epsilon^4$$

since an inner product must be non-negative.

A.5   Simulation of errors in the fiducial state

Here we discuss how to simulate the errors in the fiducial state presented in Section 3.3.1 of the main paper. For that, we present the details of the numerical simulation of the kernel matrix when there is an error in the creation of the fiducial state.

Recall, for our simulations, we take the operator V, which takes the all zero state, $|0\rangle^{\otimes N}$ to the fiducial state $|\psi\rangle$ (given by Equation (8))to be

$$V = \bigotimes_{j=1}^{N-1} \mathrm{CZ}_{(j,j+1)} \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right).$$

Consider the new operator

$$W = \bigotimes_{j=1}^{N-1} \mathrm{CZ}_{(j,j+1)} \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2} - \theta_j\right) \tag{17}$$

which is our choice of error in operator V.

The operator norm, see Appendix A.3, between the two operators is

$$\|V - W\|_{\mathrm{op}} = \left\|\bigotimes_{j=1}^{N-1} \mathrm{CZ}_{(j,j+1)} \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) - \bigotimes_{j=1}^{N-1} \mathrm{CZ}_{(j,j+1)} \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2} - \theta_j\right)\right\|_{\mathrm{op}}$$

$$= \left\|\bigotimes_{j=1}^{N-1} \mathrm{CZ}_{(j,j+1)} \left(\bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) - \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2} - \theta_j\right)\right)\right\|_{\mathrm{op}}$$

$$= \left\|\bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) - \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2} - \theta_j\right)\right\|_{\mathrm{op}}.$$

This can be further simplified by using the fact that

$$R_y\left(\frac{\pi}{2} - \theta\right) = R_y\left(\frac{\pi}{2}\right) R_y(-\theta)$$

16

so that

$$\|V - W\|_{\text{op}} = \left\| \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) - \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) R_y\left(-\theta_j\right) \right\|_{\text{op}}$$

$$= \left\| \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) - \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) \bigotimes_{j=1}^{N} R_y\left(-\theta_j\right) \right\|_{\text{op}}$$

$$= \left\| \bigotimes_{j=1}^{N} R_y\left(\frac{\pi}{2}\right) \left(\mathbb{1} - \bigotimes_{j=1}^{N} R_y\left(-\theta_j\right) \right) \right\|_{\text{op}}$$

$$= \left\| \mathbb{1} - \bigotimes_{j=1}^{N} R_y\left(-\theta_j\right) \right\|_{\text{op}} .$$

Aside: If A is normal $(A^\dagger A = A A^\dagger)$, then $A = U D U^\dagger$ where U is a unitary matrix and D is a diagonal matrix of the eigenvalues, $\lambda_j(A)$, of A and

$$\rho(A) = \|A\|_{\text{op}}$$

where

$$\rho(A) = \max\{|\lambda_1(A)|, |\lambda_2(A)|, \ldots, |\lambda_n(A)|\}$$

is the spectral radius of A. Additionally, for normal A,

$$\sigma_j(A) = |\lambda_j(A)|$$

where $\sigma(A)$ are the singular values of A.

Finally, notice that if A is normal, then

$$A - \mathbb{1} = U D U^\dagger - \mathbb{1} = U D U^\dagger - U U^\dagger = U(D - \mathbb{1})U^\dagger$$

where

$$D - \mathbb{1} = \begin{pmatrix} \lambda_1(A) & 0 & \ldots & 0 \\ 0 & \lambda_2(A) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n(A) \end{pmatrix} - \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} = \begin{pmatrix} \lambda_1(A) - 1 & 0 & \ldots & 0 \\ 0 & \lambda_2(A) - 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n(A) - 1 \end{pmatrix}$$

and it is clear that $(A - \mathbb{1})$ is normal with eigenvalues $\lambda_j(A - \mathbb{1}) = \lambda_j(A) - 1$. And so the square of the jth singular value of $A - \mathbb{1}$ is

$$\sigma_j(A - \mathbb{1})^2 = |\lambda_j(A - \mathbb{1})|^2$$
$$= |\lambda_j(A) - 1|^2$$
$$= (\lambda_j(A) - 1)(\lambda_j^*(A) - 1)$$
$$= 1 + |\lambda_j(A)|^2 - 2\,\text{Re}\left[\lambda_j(A)\right] \tag{18}$$

and if A is unitary then

$$\sigma_j(A - \mathbb{1})^2 = 2 - 2\,\text{Re}\left[\lambda_j(A)\right].$$

Since the operator norm, $\|V - W\|_{\text{op}}$ is of the form $\|U - \mathbb{1}\|_{\text{op}}$ for a unitary operator, U, we can use Equation (18) to bound the maximum possible value that each θ_j can take to keep $\|V - W\|_{\text{op}} \leq \epsilon$.

The two eigenvalues of $R_y(-\theta_j)$ are

$$\lambda_\pm\left(R_y(-\theta_j)\right) = \cos\left(\frac{\theta_j}{2}\right) \pm \text{i}\sin\left(\frac{\theta_j}{2}\right)$$

which we note are the same pair of eigenvalues of $R_y(+\theta_j)$, so for simplicity, we will consider the eigenvalues of $R_y\left(|\theta_j|\right)$. For small angles, $\theta_j = x_j\delta$ with $\delta \ll 1$, the eigenvalues can be approximated as

$$\lambda_{k_j}\left(R_y\left(|\theta_j|\right)\right) = 1 + (-1)^{k_j}\frac{\text{i}\,|x_j|}{2}\delta - \frac{x_j^2}{8}\delta^2 + \mathcal{O}(\delta^3)$$

17

where we now encode the sign ($\pm$) into the value of $k_j \in \{0, 1\}$.

Now, we use the property that if A has m eigenvalues $\lambda_j(A)$ and if B has n eigenvalues $\lambda_k(B)$ then $A \otimes B$ has mn eigenvalues

$$\lambda_{(j,k)}(A \otimes B) = \lambda_j(A)\lambda_k(B)$$

to calculate the the square of the $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)$-th singular value (unordered) of $\bigotimes_{j=1}^{N} R_y(|\theta_j|) - \mathbb{1}$

$$\sigma_{\boldsymbol{k}}^2 \left(\bigotimes_{j=1}^{N} R_y(|\theta_j|) - \mathbb{1} \right)$$

$$= 2 \left\{ 1 - \mathrm{Re} \left[\prod_{j=1}^{N} \lambda_{k_j}\left(R_y(|\theta_j|) \right) \right] \right\}$$

$$= 2 \left\{ 1 - \mathrm{Re} \left[1 + \frac{\mathrm{i}\delta}{2} \sum_{j=1}^{N} (-1)^{k_j} |x_j| - \frac{\delta^2}{8} \sum_{j=1}^{N} x_j^2 - \frac{\delta^2}{4} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (-1)^{k_i + k_j} |x_i||x_j| \right] \right\} + \mathcal{O}(\delta^3)$$

$$= 2 \left(\frac{\delta^2}{8} \sum_{j=1}^{N} x_j^2 + \frac{\delta^2}{4} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (-1)^{k_i + k_j} |x_i||x_j| \right) + \mathcal{O}(\delta^3).$$

This is maximized when the parity of k_i and k_j is the same for all i, j (take either $\boldsymbol{k} = (0, 0, \ldots, 0)$ or $\boldsymbol{k} = (1, 1, \ldots, 1)$), so

$$\sigma_{\max}^2 \left(\bigotimes_{j=1}^{N} R_y(|\theta_j|) - \mathbb{1} \right) = \frac{\delta^2}{4} \left(\sum_{j=1}^{N} x_j^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |x_i||x_j| \right) + \mathcal{O}(\delta^3).$$

Now take $y = \max \{|x_1|, |x_2|, \ldots, |x_N|\}$, so to lowest order in δ.

$$\sigma_{\max}^2 \left(\bigotimes_{j=1}^{N} R_y(|\theta_j|) - \mathbb{1} \right) \leq \frac{\delta^2}{4} \left(\sum_{j=1}^{N} y^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} y^2 \right)$$

$$= \frac{(y\delta)^2}{4} \left(N + 2 \frac{N(N-1)}{2} \right)$$

$$= \frac{N^2}{4} (y\delta)^2.$$

Therefore, if

$$|\theta_j| \leq y\delta = \frac{2\epsilon}{N}$$

for all $j \in \{1, \ldots, N\}$ then

$$\|V - W\|_{\mathrm{op}}^2 = \left\| \bigotimes_{j=1}^{N} R_y(-\theta_y) - \mathbb{1} \right\|_{\mathrm{op}}^2 = \sigma_{\max}^2 \left(\bigotimes_{j=1}^{N} R_y(|\theta_j|) - \mathbb{1} \right) \leq \frac{N^2}{4} (y\delta)^2 = \frac{N^2}{4} \left(\frac{2\epsilon}{N} \right)^2 = \epsilon^2.$$

For each simulation of the kernel matrix, we select the N values of θ_j from a uniform distribution $[-2\epsilon/N, 2\epsilon/N]$ to create W_1 (equation (17)) and another N values of θ_j to create W_2. We simulate the quantum circuit shown in Figure 7, calculate the variance in the non-trivial kernel values, and repeat 100 times, each with new values of θ_j for W_1 and W_2.

A.6 Error in the unitary representation of the elements of the group

We now prove the bound stated in Section 3.3.2 of the main paper. We assume that there is an error in the unitary representation of the group elements so that for each $\boldsymbol{x}_{a,i}$ the operator $A_{a,i}$ is used rather than $D_{\boldsymbol{x}_{a,i}}$ such that

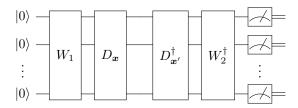$$\left\| D_{\boldsymbol{x}_{a,i}} - A_{a,i} \right\|_{\mathrm{op}} \leq \epsilon.$$

Figure 7: Quantum circuit for calculating the kernel elements when there are errors in the creation of the fiducial state.

If this is the only error, then each kernel element will take the form

$$\left|\left\langle 0^{\otimes N}\,\middle|\,V^\dagger A_{a,i}^\dagger A_{b,j}V\,\middle|\,0^{\otimes N}\right\rangle\right|^2 = \left|\left\langle\psi\,\middle|\,A_{a,i}^\dagger A_{b,j}\,\middle|\,\psi\right\rangle\right|^2.$$

Following the same steps as Appendix A.4, we can bound

$$\left|\left|\left\langle\psi\,\middle|\,D_{\boldsymbol{x}_{a,i}}^\dagger D_{\boldsymbol{x}_{b,j}}\,\middle|\,\psi\right\rangle\right| - \left|\left\langle\psi\,\middle|\,A_{a,i}^\dagger A_{b,j}\,\middle|\,\psi\right\rangle\right|\right| \leq 2\epsilon + \epsilon^2$$

so in the case of $i = j$, where $\left|\left\langle\psi\,\middle|\,D_{\boldsymbol{x}_{a,i}}^\dagger D_{\boldsymbol{x}_{b,i}}\,\middle|\,\psi\right\rangle\right| = 1$, we have

$$\left|\left\langle\psi\,\middle|\,A_{a,i}^\dagger A_{b,i}\,\middle|\,\psi\right\rangle\right|^2 \geq 1 - 4\epsilon + 2\epsilon^2 + 4\epsilon^3 + \epsilon^4$$

and in the case of $i \neq j$, where $\left|\left\langle\psi\,\middle|\,D_{\boldsymbol{x}_{a,i}}^\dagger D_{\boldsymbol{x}_{b,j}}\,\middle|\,\psi\right\rangle\right| = \sqrt{\alpha}$, we have

$$\alpha - 4\sqrt{\alpha}\epsilon + 2(2 - \sqrt{\alpha})\epsilon^2 + 4\epsilon^3 + \epsilon^4 \leq \left|\left\langle\psi\,\middle|\,A_{a,i}^\dagger A_{b,j}\,\middle|\,\psi\right\rangle\right|^2 \leq \alpha + 4\sqrt{\alpha}\epsilon + 2(2 + \sqrt{\alpha})\epsilon^2 + 4\epsilon^3 + \epsilon^4.$$

A.7 Error in the selection of the group elements

In what follows, we present the proof of the bound from Section 3.3.3. As presented in [18], we assume that for each $1 \leq a \leq n$ and $1 \leq i \leq m$ there is a random perturbation $\boldsymbol{e}_{a,i} \in \mathcal{G}$ such that

$$d(\boldsymbol{e}_{a,i}, \mathbb{1}) := \left\|D_{\boldsymbol{e}_{a,i}} - \mathbb{1}\right\|_{\mathrm{op}} < \epsilon$$

and for each $\boldsymbol{q}_{a,i} \in \bigcup_j C_j$ we have an $\boldsymbol{x}_{a,i} \in \mathcal{X}$ such that $\boldsymbol{x}_{a,i} = \boldsymbol{e}_{a,i}\boldsymbol{q}_{a,i}$. Since D is a unitary representation of the group $\mathcal{G}$, we have

$$D_{\boldsymbol{x}_{a,i}} = D_{\boldsymbol{e}_{a,i}}D_{\boldsymbol{q}_{a,i}} = D_{\boldsymbol{e}_{a,i}}D_{\boldsymbol{c}_i}D_{\boldsymbol{s}_a}.$$

So, assuming no more errors, each kernel element will take the form

$$\left|\left\langle 0^{\otimes N}\,\middle|\,V^\dagger D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}D_{\boldsymbol{c}_j}D_{\boldsymbol{s}_b}V\,\middle|\,0^{\otimes N}\right\rangle\right|^2 = \left|\left\langle\psi\,\middle|\,D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}D_{\boldsymbol{c}_j}D_{\boldsymbol{s}_b}\,\middle|\,\psi\right\rangle\right|^2$$

$$= \left|\left\langle\psi\,\middle|\,D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}D_{\boldsymbol{c}_j}\,\middle|\,\psi\right\rangle\right|^2$$

$$= \left|\left\langle\varphi_i\,\middle|\,D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}\,\middle|\,\varphi_j\right\rangle\right|^2$$

where we define $|\varphi_j\rangle := D_{\boldsymbol{c}_j}|\psi\rangle$.

In the case of $\boldsymbol{c}_i = \boldsymbol{c}_j$ where $\left|\left\langle\psi\,\middle|\,D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_i}D_{\boldsymbol{s}_b}\,\middle|\,\psi\right\rangle\right| = |\langle\varphi_i|\varphi_i\rangle| = 1$, we first notice that

$$\left\|D_{\boldsymbol{e}_{a,i}} - D_{\boldsymbol{e}_{b,j}}\right\|_{\mathrm{op}} \leq 2\epsilon$$

which follows from Equation (9). Then we can use Equation (15) to bound

$$\left|\left\langle\varphi_i\,\middle|\,D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,i}}\,\middle|\,\varphi_i\right\rangle\right| \geq 1 - \frac{(2\epsilon)^2}{2} = 1 - \frac{\epsilon^2}{2}$$

$$\implies \left|\left\langle\varphi_i\,\middle|\,D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,i}}\,\middle|\,\varphi_i\right\rangle\right|^2 \geq 1 - \epsilon^2 + \frac{\epsilon^4}{4}.$$

19

In the case of $\boldsymbol{c}_i \neq \boldsymbol{c}_j$ where $\left|\langle\psi|D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b}|\psi\rangle\right| = |\langle\varphi_i|\varphi_j\rangle| = \sqrt{\alpha}$, we first notice that (since the operator norm, see Appendix A.3, is an unitarily invariant norm)

$$2\epsilon \geq \left\|D_{\boldsymbol{e}_{a,i}} - D_{\boldsymbol{e}_{b,j}}\right\|_{\mathrm{op}} = \left\|D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{a,i}} - D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}\right\|_{\mathrm{op}} = \left\|\mathbb{1} - D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}\right\|_{\mathrm{op}}$$

so we can now bound

$$
\begin{aligned}
\left|\left|\left\langle\varphi_i\left|D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}\right|\varphi_j\right\rangle\right| - |\langle\varphi_i|\varphi_j\rangle|\right| &\leq \left|\left\langle\varphi_i\left|D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}}\right|\varphi_j\right\rangle - \langle\varphi_i|\varphi_j\rangle\right| \quad \text{(Reverse triangle inequality)} \\
&= \left|\left\langle\varphi_i\left|\left(D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}} - \mathbb{1}\right)\right|\varphi_j\right\rangle\right| \\
&\leq \left\|D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}} - \mathbb{1}\right\|_{\mathrm{op}} \quad \text{by Equation (13)} \\
&\leq 2\epsilon.
\end{aligned}
\tag{19}
$$

Therefore, we have (for $i \neq j$)

$$\sqrt{\alpha} - 2\epsilon \leq \left|\left\langle\psi\left|D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}} D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b}\right|\psi\right\rangle\right| \leq \sqrt{\alpha} + 2\epsilon$$

$$\implies \alpha - 4\sqrt{\alpha}\epsilon + 4\epsilon^2 \leq \left|\left\langle\psi\left|D_{\boldsymbol{s}_a}^\dagger D_{\boldsymbol{c}_i}^\dagger D_{\boldsymbol{e}_{a,i}}^\dagger D_{\boldsymbol{e}_{b,j}} D_{\boldsymbol{c}_j} D_{\boldsymbol{s}_b}\right|\psi\right\rangle\right|^2 \leq \alpha + 4\sqrt{\alpha}\epsilon + 4\epsilon^2.$$

A.8 Simulations of errors in the selection of the group elements

In this section we discuss the details of how we simulated errors in the selection of the group elements, which we presented in Section 3.3.4.

This type of error assumes a random perturbation, $\boldsymbol{e} \in \mathcal{G}$, on each $\boldsymbol{x} \in \mathcal{X}$ such that

$$d(\boldsymbol{e}, \mathbb{1}) := \|D_{\boldsymbol{e}} - \mathbb{1}\|_{\mathrm{op}} < \epsilon.$$

For our choice of group, $\mathcal{G} = \mathrm{SU}(2)^{\otimes N}$, each perturbation can be represented as

$$D_{\boldsymbol{e}} = \bigotimes_{j=1}^N R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}).
\tag{20}$$

Similar to our simulation of errors in the fiducial state, we again use Equation (18) to bound the maximum possible value that each $\theta_{j,i}$ can take to keep $\|D_{\boldsymbol{e}} - \mathbb{1}\|_{\mathrm{op}} \leq \epsilon$.

The eigenvalues of $R_X(\theta_{j,1}) R_Z(\theta_{j,2}) R_X(\theta_{j,3})$ can easily be found to be

$$\lambda_\pm\left(R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3})\right) = \cos\left(\frac{\theta_{j,2}}{2}\right)\cos\left(\frac{\theta_{j,1}+\theta_{j,3}}{2}\right) \pm \mathrm{i}\sqrt{1 - \cos^2\left(\frac{\theta_{j,2}}{2}\right)\cos^2\left(\frac{\theta_{j,1}+\theta_{j,3}}{2}\right)}.$$

For small angles, $\theta_{j,i} = x_{j,i}\delta$ with $\delta \ll 1$, the eigenvalues can be approximated as

$$
\begin{aligned}
&\lambda_{k_j}\left(R_x(x_{1,j}\delta) R_z(x_{j,2}\delta) R_x(x_{j,3}\delta)\right) \\
&= 1 + (-1)^{k_j}\, \mathrm{i}\sqrt{\frac{x_{j,2}^2}{4} + \frac{(x_{j,1}+x_{j,3})^2}{4}}\,\delta - \frac{1}{2}\left(\frac{x_{j,2}^2}{4} + \frac{(x_{j,1}+x_{j,3})^2}{4}\right)\delta^2 + \mathcal{O}(\delta^3) \\
&= 1 + (-1)^{k_j}\, \frac{\mathrm{i}}{2}\sqrt{X_j}\,\delta - \frac{1}{8}X_j\delta^2 + \mathcal{O}(\delta^3)
\end{aligned}
$$

where the sign ($\pm$) is represented as the value of $k_j \in \{0,1\}$ and we define

$$X_j := x_{j,2}^2 + (x_{j,1}+x_{j,3})^2 \geq 0.$$

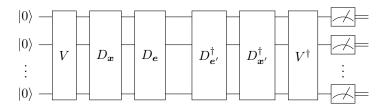Figure 8: Quantum circuit for calculating the kernel elements when there are errors in the selection of group elements.

The square of the $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)$-th singular value (unordered) of $\left(\bigotimes_{j=1}^{N} R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) - \mathbb{1}$ is

$$
\sigma_{\boldsymbol{k}}^2 \left[\left(\bigotimes_{j=1}^{N} R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) - \mathbb{1} \right]
$$

$$
= 2 \left\{ 1 - \mathrm{Re} \left[\prod_{j=1}^{N} \lambda_{k_j} \left(R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) \right] \right\}
$$

$$
= 2 \left(\frac{\delta^2}{8} \sum_{j=1}^{N} X_j + \frac{\delta^2}{4} \sum_{i=1}^{N-1} \sum_{j=i+1}^{n} (-1)^{k_i + k_j} \sqrt{X_i X_j} \right) + \mathcal{O}(\delta^3).
$$

This is maximized when the parity of k_i and k_j is the same for all i, j (take either $\boldsymbol{k} = (0, 0, \ldots, 0)$ or $\boldsymbol{k} = (1, 1, \ldots, 1)$), so

$$
\sigma_{\max}^2 \left[\left(\bigotimes_{j=1}^{N} R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) - \mathbb{1} \right] = \frac{\delta^2}{4} \sum_{j=1}^{N} X_j + \frac{\delta^2}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{X_i X_j} + \mathcal{O}(\delta^3).
$$

Now take $y = \max \left\{ |x_{1,1}|, |x_{1,2}|, |x_{1,3}|, |x_{2,1}|, |x_{2,2}|, |x_{2,3}|, \ldots, |x_{N,1}|, |x_{N,2}|, |x_{N,3}| \right\}$, so

$$
X_j \leq (y + y)^2 + y^2 = 5y^2
$$

and to lowest order in δ

$$
\sigma_{\max}^2 \left[\left(\bigotimes_{j=1}^{N} R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) - \mathbb{1} \right] \leq \frac{\delta^2}{4} \sum_{j=1}^{N} 5y^2 + \frac{\delta^2}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 5y^2
$$

$$
= \frac{5(y\delta)^2}{4} \left(N + 2 \frac{N(N-1)}{2} \right)
$$

$$
= \frac{5N^2}{4} (y\delta)^2.
$$

Therefore, if

$$
|\theta_{j,i}| \leq y\delta = \frac{2\epsilon}{\sqrt{5}N}
$$

then

$$
\|D_{\boldsymbol{e}} - \mathbb{1}\|_{\mathrm{op}}^2 = \left\| \left(\bigotimes_{j=1}^{N} R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) - \mathbb{1} \right\|_{\mathrm{op}}
$$

$$
= \sigma_{\max}^2 \left[\left(\bigotimes_{j=1}^{N} R_x(\theta_{j,1}) R_z(\theta_{j,2}) R_x(\theta_{j,3}) \right) - \mathbb{1} \right]
$$

$$
\leq \frac{5N^2}{4} \left(\frac{2\epsilon}{\sqrt{5}N} \right)^2
$$

$$
= \epsilon^2.
$$

For each $\boldsymbol{x} \in \mathcal{X}$ we select $3N$ values of $\theta_{j,i}$ from a uniform distribution of $\left[-\frac{2\epsilon}{\sqrt{5}N}, \frac{2\epsilon}{\sqrt{5}N}\right]$ to create $D_{\boldsymbol{e}}$ (equation (20)). We then simulate the quantum circuit shown in Figure 8 for each pair $(\boldsymbol{x}, \boldsymbol{x}')$ in the test data and the associated $(\boldsymbol{e}, \boldsymbol{e}')$ and calculate the variance in the non-trivial kernel matrix. This is repeated 100 times, with new values of $\theta_{j,i}$ for each $D_{\boldsymbol{e}}$.