# Deep quantum neural networks form Gaussian processes

Diego García-Martín,[1] Martín Larocca,[2,3] and M. Cerezo[1,*]

[1] *Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[2] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[3] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

It is well known that artificial neural networks initialized from independent and identically distributed priors converge to Gaussian processes in the limit of large number of neurons per hidden layer. In this work we prove an analogous result for Quantum Neural Networks (QNNs). Namely, we show that the outputs of certain models based on Haar random unitary or orthogonal deep QNNs converge to Gaussian processes in the limit of large Hilbert space dimension $d$. The derivation of this result is more nuanced than in the classical case due the role played by the input states, the measurement observable, and the fact that the entries of unitary matrices are not independent. An important consequence of our analysis is that the ensuing Gaussian processes cannot be used to efficiently predict the outputs of the QNN via Bayesian statistics. Furthermore, our theorems imply that the concentration of measure phenomenon in Haar random QNNs is much worse than previously thought, as we prove that expectation values and gradients concentrate as $\mathcal{O}\left(\frac{1}{e^d\sqrt{d}}\right)$ – exponentially in the Hilbert space dimension. Finally, we discuss how our results improve our understanding of concentration in $t$-designs.

Neural Networks (NNs) have revolutionized the fields of Machine Learning (ML) and artificial intelligence. Their tremendous success across many fields of research in a wide variety of applications [1–3] is certainly astonishing. While much of this success has come through heuristics, the past few decades have witnessed a significant increase in our theoretical understanding of their inner workings. One of the most interesting results regarding NNs is that fully-connected models with a single hidden layer converge to Gaussian Processes (GPs) in the limit of large number of hidden neurons, when the parameters are initialized from independent and identically distributed (i.i.d.) priors [4]. More recently, it has been shown that i.i.d.-initialized, fully-connected, multi-layer NNs also converge to GPs in the infinite-width limit [5]. Furthermore, other architectures, such as convolutional NNs [6], transformers [7] or recurrent NNs [8] are also GPs under certain assumptions. More than just a mathematical curiosity, the correspondence between NNs and GPs opened up the possibility of performing exact Bayesian inference for regression and learning tasks using wide NNs [4, 9].

With the advent of quantum computers, there has been an enormous interest in merging quantum computing with ML, leading to the thriving field of Quantum Machine Learning (QML) [10–14]. Rapid progress has been made in this field, largely fueled by the hope that QML may provide a quantum advantage in the near-term for some practically-relevant problems. While the prospects for such a practical quantum advantage remain unclear [15], a number of promising analytical results have already been put forward [16–19]. Still, much remains to be known about QML models.

In this work, we contribute to the QML body of knowledge by proving that under certain conditions, the outputs of deep Quantum Neural Networks (QNNs) – i.e., parametrized quantum circuits acting on input states drawn from a training set– converge to GPs in the limit of large Hilbert space dimension (see Fig. 1). Our results are derived for QNNs that are Haar random over the unitary and orthogonal groups. Unlike in the classical case, where the proof of the emergence of GPs stems from the central limit theorem, the situation becomes more intricate in the quantum setting as the entries of the QNN are not independent – the rows and columns of a unitary matrix are constrained to be mutually orthonormal. Hence, our proof strategy boils down to showing that each moment of the QNN's output distribution converges to that of a multivariate Gaussian. In addition, we show that in contrast to classical NNs, the Bayesian distribution of the QNN is inefficient for predicting the model's outputs. We then use our results to provide a precise characterization of the concentration of measure phenomenon in deep random quantum circuits [20–25]. Here, our theorems indicate that the expectation values, as well as the gradients, of Haar random processes concentrate exponentially faster than reported in previous barren plateau studies [20, 21]. Finally, we discuss how our results can be leveraged to study QNNs that are not fully Haar random but instead form $t$-designs, which constitutes a much more practical assumption [26–28].

* cerezo@lanl.gov

## I. GAUSSIAN PROCESSES AND CLASSICAL MACHINE LEARNING

We begin by introducing GPs.

**Definition 1** (Gaussian process)**.** *A collection of random variables* $\{X_1, X_2, \dots\}$ *is a GP if and only if, for every finite set of indices* $\{1, 2, \dots, m\}$*, the vector* $(X_1, X_2, \dots, X_m)$ *follows a multivariate Gaussian distribution, which we denote as* $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*. Said otherwise, every linear combination of* $\{X_1, X_2, \dots, X_m\}$ *follows a univariate Gaussian distribution.*

In particular, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is determined by its $m$-dimensional mean vector $\boldsymbol{\mu} = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_m])$, and its $m \times m$ dimensional covariance matrix with entries $(\boldsymbol{\Sigma})_{\alpha\beta} = \mathrm{Cov}[X_\alpha, X_\beta]$.

GPs are extremely important in ML since they can be used as a form of kernel method to solve learning tasks [4, 9]. For instance, consider a regression problem where the data domain is $\mathscr{X} = \mathbb{R}$ and the label domain is $\mathscr{Y} = \mathbb{R}$. Instead of finding a single function $f : \mathscr{X} \to \mathscr{Y}$ which solves the regression task, a GP instead assigns probabilities to a set of possible $f(x)$, such that the probabilities are higher for the "more likely" functions. Following a Bayesian inference approach, one then selects the functions that best agree with some set of empirical observations [9, 14].

Under this framework, the output over the distribution of functions $f(x)$, for $x \in \mathscr{X}$, is a random variable. Then, given a set of training samples $x_1, \dots, x_m$, and some covariance function $\kappa(x, x')$, Definition 1 implies that if one has a GP, the outputs $f(x_1), \dots, f(x_m)$ are random variables sampled from some multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. From here, the GP is used to make predictions about the output $f(x_{m+1})$ (for some new data instance $x_{m+1}$), given the previous observations $f(x_1), \dots, f(x_m)$. Explicitly, one constructs the joint distribution $P(f(x_1), \dots, f(x_m), f(x_{m+1}))$ from the averages and the covariance function $\kappa$, and obtains the sought-after "predictive distribution" $P(f(x_{m+1})|f(x_1), \dots, f(x_m))$ via marginalization. The power of the GP relies on the fact that this distribution usually contains less uncertainty than $P(f(x_{m+1})) = \mathcal{N}(\mathbb{E}[f(x_{m+1})], \kappa(x_{m+1}, x_{m+1}))$ (see the Methods).

## II. HAAR RANDOM DEEP QNNS FORM GPS

In what follows we consider a setting where one is given repeated access to a dataset $\mathscr{D}$ containing pure quantum states $\{\rho_i\}_i$ on a $d$-dimensional Hilbert space. We will make no assumptions regarding the origin of these states, as

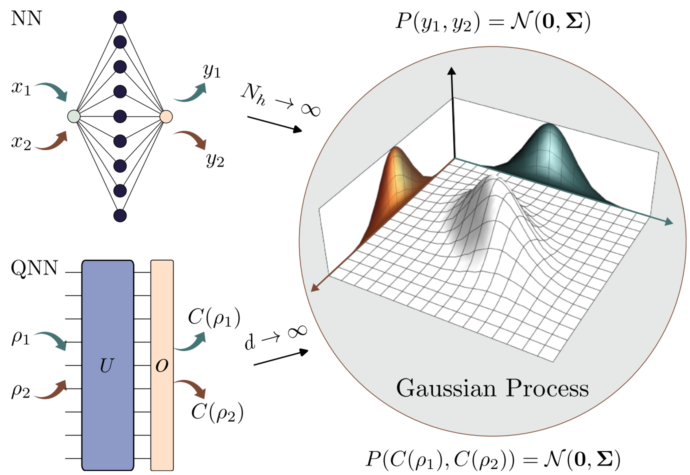



Figure 1. **Schematic of our main results.** It is well known that certain classical NNs with $N_h$ neurons per hidden layer become GPs when $N_h \to \infty$. That is, given inputs $x_1$ and $x_2$, and corresponding outputs $y_1$ and $y_2$, then the joint probability $P(y_1, y_2)$ is a multivariate Gaussian $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. In this work, we show that a similar result holds under certain conditions for deep QNNs in the limit of large Hilbert space dimension, $d \to \infty$. Now, given quantum states $\rho_1$ and $\rho_2$, $C(\rho) = \mathrm{Tr}[U\rho U^\dagger O]$ is such that $P(C(\rho_1), C(\rho_2)) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

they can correspond to classical data encoded in quantum states [29, 30], or quantum data obtained from some quantum mechanical process [31, 32]. Then, we assume that the states are sent through a deep QNN, denoted $U$. While in general $U$ can be parametrized by some set of trainable parameters $\boldsymbol{\theta}$, we leave such dependence implicit for the ease of notation. At the output of the circuit one measures the expectation value of a traceless Hermitian operator taken from a set $\mathscr{O} = \{O_j\}_j$ such that $\mathrm{Tr}[O_j O_{j'}] = d\delta_{j,j'}$ and $O_j^2 = \mathbb{1}$, for all $j, j'$ (e.g., Pauli strings). We denote the QNN outputs as

$$C_j(\rho_i) = \mathrm{Tr}\left[U\rho_i U^\dagger O_j\right] . \qquad (1)$$

Then, we collect these quantities over some set of states from $\mathscr{D}$ and some set of measurements from $\mathscr{O}$ in a vector

$$\mathscr{C} = (C_j(\rho_i), \dots, C_{j'}(\rho_{i'}), \dots) . \qquad (2)$$

As we will show below, in the large-$d$ limit $\mathscr{C}$ converges to a GP when the QNN unitaries $U$ are sampled according to the Haar measure on the degree-$d$ unitary $\mathbb{U}(d)$ or orthogonal $\mathbb{O}(d)$ groups (see Fig. 1). We will henceforth use the notation $\mathbb{E}_{\mathbb{U}(d)}$ and $\mathbb{E}_{\mathbb{O}(d)}$ to respectively denote Haar averages over $\mathbb{U}(d)$ and $\mathbb{O}(d)$. Moreover, we assume that when the circuit is sampled from $\mathbb{O}(d)$, the states in $\mathscr{D}$ and the measurement operators in $\mathscr{O}$ are real valued.

### A. Moment computation in the large-$d$ limit

As we discuss in the Methods, we cannot rely on simple central-limit-theorem arguments to show that $\mathscr{C}$ forms a GP. Hence, our proof strategy is based on computing all the moments of the vector $\mathscr{C}$ and showing that they asymptotically match those of a multivariate Gaussian distribution. To conclude the proof we show that these moments unequivocally determine the distribution, for which we can use Carleman's condition [33, 34]. We refer the reader to the Supplemental Information (SI) for the detailed proofs of the results in this manuscript.

First, we present the following lemma.

**Lemma 1.** *Let $C_j(\rho_i)$ be the expectation value of a Haar random QNN as in Eq. (1). Then for any $\rho_i \in \mathscr{D}$, $O_j \in \mathscr{O}$,*

$$\mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i)] = \mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)] = 0\,. \tag{3}$$

*Moreover, for any pair of states $\rho_i, \rho_{i'} \in \mathscr{D}$ and operators $O_j, O_{j'} \in \mathscr{O}$ we have*

$$\mathrm{Cov}_{\mathbb{U}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})] = \mathrm{Cov}_{\mathbb{O}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})] = 0\,,$$

*if $j \neq j'$ and*

$$\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'} = \frac{d}{d^2-1}\left(\mathrm{Tr}[\rho_i\rho_{i'}] - \frac{1}{d}\right), \tag{4}$$

$$\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'} = \frac{2(d+1)}{(d+2)(d-1)}\left(\mathrm{Tr}[\rho_i\rho_{i'}]\left(1-\frac{1}{d+1}\right) - \frac{1}{d+1}\right), \tag{5}$$

*if $j = j'$. Here, we have defined $\boldsymbol{\Sigma}^{G}_{i,i'} = \mathrm{Cov}_G[C_j(\rho_i)C_j(\rho_{i'})]$, where $G = \mathbb{U}(d), \mathbb{O}(d)$.*

Lemma 1 shows that the expectation value of the QNN outputs is always zero. More notably, it indicates that the covariance between the outputs is null if we measure different observables (even if we use the same input state and the same circuit). This implies that the distributions $C_j(\rho_i)$ and $C_{j'}(\rho_{i'})$ are independent if $j \neq j'$. That is, knowledge of the measurement outcomes for one observable and different input states does not provide any information about the outcomes of other measurements, at these or any other input states. Therefore, in what follows we will focus on the case where $\mathscr{C}$ contains expectation values for different states, but the same operator. In this case, Lemma 1 shows that the covariances will be positive, zero, or negative depending on whether $\mathrm{Tr}[\rho_i\rho_{i'}]$ is larger, equal, or smaller than $\frac{1}{d}$, respectively.

We now state a useful result.

**Lemma 2.** *Let $\mathscr{C}$ be a vector of $k$ expectation values of a Haar random QNN as in Eq. (2), where one measures the same operator $O_j$ over a set of $k$ states $\rho_1, \ldots, \rho_k \in \mathscr{D}$. In the large-$d$ limit, if $k$ is odd then $\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right] = \mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right] = 0$. Moreover, if $k$ is even and if a) $\mathrm{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $i, i'$, we have*

$$\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right] = \frac{1}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\{t,t'\}\in\sigma}\mathrm{Tr}[\rho_t\rho_{t'}] = \frac{\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]}{2^{k/2}}\,, \tag{6}$$

*where the summation runs over all the possible disjoint pairing of indexes in the set $\{1,2,\ldots,k\}$, $T_k$, and the product is over the different pairs in each pairing; while if b) $\mathrm{Tr}[\rho_i\rho_{i'}] = 0$ for all $i, i'$, we have*

$$\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right] = \frac{k!}{2^{k/2}(k/2)!}\frac{1}{d^k} = \frac{\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]}{2^{k/2}}\,. \tag{7}$$

Using Lemma 2 as our main tool, we will be able to prove that deep QNNs form GPs for different types of datasets. In Table I we present a summary of our main results.

| Dataset. For all $\rho_i, \rho_{i'} \in \mathscr{D}$: | GP | Correlation | Statement |
|---|---|---|---|
| $\mathrm{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ | Yes | Positive | Theorem 1 |
| $\mathrm{Tr}[\rho_i\rho_{i'}] = \frac{1}{d}$ | Yes | Null | Theorem 2 |
| $\mathrm{Tr}[\rho_i\rho_{i'}] = 0$ | Yes | Negative | Theorem 3 |

Table I. **Summary of main results.** In the first column we present conditions for the states in the dataset under which the deep QNN's outputs form GPs. In the remaining columns we report the correlation in the GP variables and the associated theorem where the main result is stated. In all cases we assume that we measure the same operator $O_j$ for all $\rho_i, \rho_{i'} \in \mathscr{D}$. In Theorem 5, we extend some of these results to the cases where the conditions are only met on average when sampling states over $\mathscr{D}$.

### B. Positively correlated GPs

We begin by studying the case when the states in the dataset satisfy $\mathrm{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $\rho_i, \rho_{i'} \in \mathscr{D}$. According to Lemma 1, this implies that the variables are positively correlated. In the large $d$ limit, we can derive the following theorem.

**Theorem 1.** *Under the same conditions for which Lemma 2(a) holds, the vector $\mathscr{C}$ forms a GP with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix given by $\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'} = \frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'}}{2} = \frac{\mathrm{Tr}[\rho_i\rho_{i'}]}{d}$.*

Theorem 1 indicates that the covariances for the orthogonal group are twice as large as those arising from the unitary group. In Fig. 2, we present results obtained by numerically simulating a unitary Haar random QNN for a system of $n = 18$ qubits. The circuits were sampled using known results for the asymptotics of the entries of unitary matrices [33]. In the left panels of Fig. 2, we show the corresponding two-dimensional GP obtained for two initial states that satisfy $\mathrm{Tr}[\rho_i, \rho_{i'}] \in \Omega(1)$. Here, we see that the variables are positively correlated in accordance with the prediction in Theorem 1.

The fact that the outputs of deep QNNs form GPs unravels a deep connection between QNNs and quantum kernel methods. While it has already been pointed out that QNN-based QML constitutes a form of kernel-based learning [35], our results materialize this connection for the case of Haar random circuits. Notably, we can recognize that the kernel arising in the GP covariance matrix is proportional to the Fidelity kernel, i.e., to the Hilbert-Schmidt inner product between the data states [35–37]. Moreover, since the predictive distribution of a GP can be expressed as a function of the covariance matrix (see the Methods), and thus of the kernel entries, our results further cement the fact that quantum models such as those in Eq. (1) are functions in the reproducing kernel Hilbert space [35].

### C. Uncorrelated GPs

We now consider the case when $\mathrm{Tr}[\rho_i \rho_{i'}] = \frac{1}{d}$ for all $\rho_i, \rho_{i'} \in \mathscr{D}$. We find the following result.

**Theorem 2.** *Let $\mathscr{C}$ be a vector of $k$ expectation values of operators in $\mathscr{O}$ over a set of $k$ states $\rho_1, \ldots, \rho_k \in \mathscr{D}$. If $\mathrm{Tr}[\rho_i \rho_{i'}] = \frac{1}{d}$ for all $i, i'$, then in the large $d$-limit $\mathscr{C}$ forms a GP with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and diagonal covariance matrix*

$$\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'} = \frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'}}{2} = \begin{cases} \frac{1}{d} & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases} . \tag{8}$$

In the right panel of Fig. 2, we plot the GP corresponding to two initial states such that $\mathrm{Tr}[\rho_i, \rho_{i'}] = \frac{1}{d}$. In this case, the variables appear uncorrelated as predicted by Theorem 2.

### D. Negatively correlated GPs

Here we study the case when $\mathrm{Tr}[\rho_i \rho_{i'}] = 0$ for all $\rho_i, \rho_{i'} \in \mathscr{D}$. We prove the following theorem.

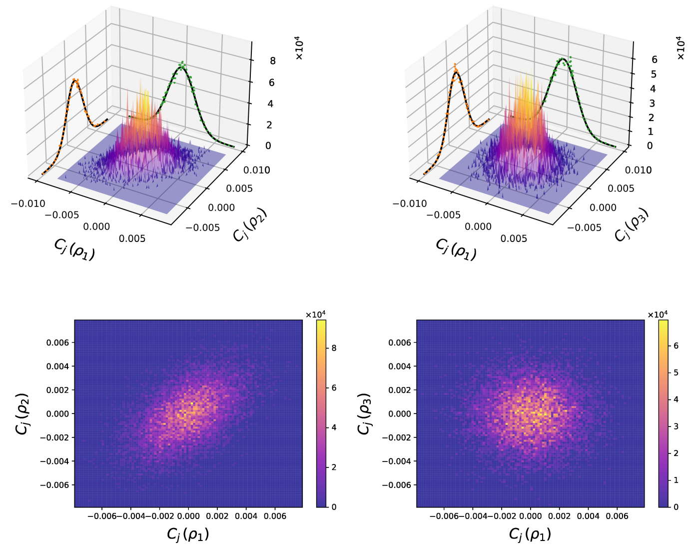


Figure 2. **Two-dimensional GPs.** We plot the joint probability density function, as well as its scaled marginals, for the measurement outcomes at the output of a unitary Haar random QNN acting on $n = 18$ qubits. The measured observable is $O_j = Z_1$, where $Z_1$ denotes the Pauli $z$ operator on the first qubit. Moreover, the input states are: $\rho_1 = |0\rangle\langle 0|^{\otimes n}$ and $\rho_2 = |\mathrm{GHZ}\rangle\langle \mathrm{GHZ}|$ with $|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$, for the left panel; $\rho_1$ and $\rho_3 = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = \frac{1}{\sqrt{d}}\,|0\rangle^{\otimes n} + \sqrt{1 - \frac{1}{d}}\,|1\rangle^{\otimes n}$ for the right panel. In both cases we took $10^4$ samples.

**Theorem 3.** *Under the same conditions for which Lemma 2(b) holds, the vector $\mathscr{C}$ forms a GP with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix*

$$\boldsymbol{\Sigma}^{\mathbb{U}(d)}_{i,i'} = \begin{cases} \frac{1}{d+1} & \text{if } i = i' \\ -\frac{1}{(d^2-1)} & \text{if } i \neq i' \end{cases} , \tag{9}$$

*and*

$$\boldsymbol{\Sigma}^{\mathbb{O}(d)}_{i,i'} = \begin{cases} \frac{2}{d+1} & \text{if } i = i' \\ -\frac{1}{(d+2)(d-1)} & \text{if } i \neq i' \end{cases} . \tag{10}$$

Note that since we are working in the large $d$ limit, we could have expressed the entries of the covariance matrices of Theorem 3 as $\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i} = \frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i}}{2} = \frac{1}{d}$, and $\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'} = \frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i}}{2} = -\frac{1}{d^2}$ for $i \neq i'$. However, we find it convenient to present their full form as it will be important below.

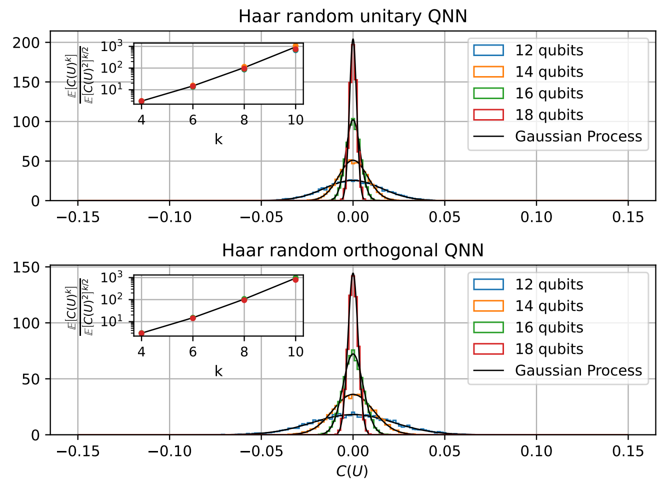



Figure 3. **Probability density function for $C_j(\rho_i)$, for Haar random QNNs and different problem sizes.** We consider unitary and orthogonal QNNs with $n$-qubits, and we take $\rho_i = |0\rangle\langle 0|^{\otimes n}$, and $O_j = Z_1$. The colored histograms are built from $10^4$ samples in each case, and the solid black lines represent the corresponding Gaussian distributions $\mathcal{N}(0, \sigma^2)$, where $\sigma^2$ is given in Corollary 1. The insets show the numerical versus predicted value of $\mathbb{E}[C_j(\rho_i)^k]/\mathbb{E}[C_j(\rho_i)^2]^{k/2}$. For a Gaussian distribution with zero mean, such quotient is $\frac{k!}{2^{k/2}(k/2)!}$ (solid black line).

## E. Deep QNN outcomes, and their linear combination

In this section, and the following ones, we will study the implications of Theorems 1, 2 and 3. Unless stated otherwise, the corollaries we present can be applied to all considered datasets (see Table I).

First, we study the univariate probability distribution $P(C_j(\rho_i))$.

**Corollary 1.** *Let $C_j(\rho_i)$ be the expectation value of a Haar random QNN as in Eq. (1). Then, for any $\rho_i \in \mathscr{D}$ and $O_j \in \mathscr{O}$, we have*

$$P(C_j(\rho_i)) = \mathcal{N}(0, \sigma^2)\,, \tag{11}$$

*where $\sigma^2 = \frac{1}{d}, \frac{2}{d}$ when $U$ is Haar random over $\mathbb{U}(d)$ and $\mathbb{O}(d)$, respectively.*

Corollary 1 shows that when a single state from $\mathscr{D}$ is sent through the QNN, and a single operator from $\mathscr{O}$ is measured, the outcomes follow a Gaussian distribution with a variance that vanishes inversely proportional with the Hilbert space dimension. This means that for large problems sizes, we can expect the results to be extremely concentrated around their mean (see below for more details). In Fig. 3 we compare the predictions from Corollary 1 to numerical simulations. We find that the simulations match very closely our theoretical results for both the unitary and the orthogonal groups. Moreover, we can observe that the standard deviation for orthogonal Haar random QNNs is larger than that for unitary ones. In Fig. 3 we also plot the quotient $\frac{\mathbb{E}[C_j(\rho_i)^k]}{\mathbb{E}[C_j(\rho_i)^2]^{k/2}}$ obtained from our numerics, and we verify that it follows the value $\frac{k!}{2^{k/2}(k/2)!}$ of a Gaussian distribution.

At this point, it is worth making an important remark. According to Definition 1, if $\mathscr{C}$ forms a GP, then any linear combination of its entries will follow a univariate Gaussian distribution. In particular, if $\{C_j(\rho_1), C_j(\rho_2), \ldots, C_j(\rho_m)\} \subseteq \mathscr{C}$, then $P(C_j(\widetilde{\rho}))$ with $\widetilde{\rho} = \sum_{i=1}^m c_i \rho_i$ will be equal to $\mathcal{N}(0, \widetilde{\sigma}^2)$ for some $\widetilde{\sigma}$. Note that the real-valued coefficients $\{c_i\}_{i=1}^m$ need not be a probability distribution, meaning that $\widetilde{\rho}$ is not necessarily a quantum state. The previous then raises an important question: What happens if $\widetilde{\rho} \propto \mathbb{1}$? A direct calculation shows that $C_j(\widetilde{\rho}) = \sum_{i=1}^d C_j(c_i\rho_i) \propto \mathrm{Tr}\left[U \mathbb{1} U^\dagger O_j\right] = \mathrm{Tr}[O_j] = 0$. How can we then unify these two perspectives? On the one hand $C_j(\widetilde{\rho})$ should be normally distributed, but on the other hand we know that it is always constant. To solve this issue, we note that the only dataset we considered where the identity can be constructed is the one where $\mathrm{Tr}[\rho_i \rho_{i'}] = 0$ for all $i, i'$[1]. In that case, we can leverage Theorem 3 along with the identity $\widetilde{\sigma}^2 = \mathrm{Var}_G[\sum_{i=1}^d C_j(\rho_i)] = \sum_{i,i'} \mathrm{Cov}_G[C_j(\rho_i), C_j(\rho_{i'})]$ to explicitly prove that $\mathrm{Var}_G[\sum_{i=1}^d C_j(\rho_i)] = 0$ (for $G = \mathbb{U}(d), \mathbb{O}(d)$). Hence, we find a zero-variance Gaussian distribution, i.e., a delta distribution in the QNN's outcomes (as expected).

## F. Predictive power of the deep QNN's GP

Let us now study the predictive distribution of the QNN's GP. We consider a scenario where we send $k$ states $\rho_1, \ldots, \rho_k \in \mathscr{D}$ to the QNN and measure the same operator $O_j$ at its output. Moreover, we assume that there exists some (statistical) noise in the measurement process, so that

[1] This follows from the fact that if $\mathscr{D}$ contains a complete basis then for any $\widetilde{\rho} \in \mathscr{D}^\perp$, one has that if $\mathrm{Tr}[\widetilde{\rho}\rho_i] = 0$ for all $\rho_i \in \mathscr{D}$, then $\widetilde{\rho} = 0$. Here, $\mathscr{D}^\perp$ denotes the kernel of the projector onto the subspace spanned by the vectors in $\mathscr{D}$.

we actually estimate the quantities $y(\rho_i) = C_j(\rho_i) + \varepsilon_i$, where the noise terms $\varepsilon_i$ are assumed to be independently drawn from the same distribution $P(\varepsilon_i) = \mathcal{N}(0, \sigma_N^2)$. For simplicity, we assume that the noise is given by finite sampling and that $\sigma_N^2 = \frac{1}{N}$, where $N \in \mathcal{O}(\text{poly}(\log(d)))$ is the number of shots used to estimate each $y(\rho_i)$. We then prove the following result.

**Theorem 4.** *Consider a GP obtained from a Haar random QNN. Given the set of observations $(y(\rho_1), \dots, y(\rho_m))$ obtained from $N \in \mathcal{O}(\text{poly}(\log(d)))$ measurements, then the predictive distribution of the GP is trivial:*

$$P(C_j(\rho_{m+1})|C_j(\rho_1), \dots, C_j(\rho_m)) = P(C_j(\rho_{m+1})) = \mathcal{N}(0, \sigma^2)\,,$$

*where $\sigma^2$ is given by Corollary 1.*

Theorem 4 shows that by spending only a poly-logarithmic-in-$d$ number of measurements, one cannot use Bayesian statistical theory to learn any information about new outcomes given previous ones.

### G. Concentration of measure

In this section we show that Corollary 1 provides a more precise characterization of the concentration of measure and the barren plateau phenomena for Haar random circuits than that found in the literature [20–25]. First, it implies that deep orthogonal QNNs will exhibit barren plateaus, a result not previously known. Second, we recall that in standard barren plateau analyses, one only looks at the first two moments of the distribution of cost values $C_j(\rho_i)$ (or, similarly, of gradient values $\partial_\theta C_j(\rho_i)$). Then one uses Chebyshev's inequality, which states that for any $c > 0$, the probability $P(|X| \geqslant c) \leqslant \frac{\text{Var}[X]}{c^2}$, to prove that $P(|C_j(\rho_i)| \geqslant c)$ and $P(|\partial_\theta C_j(\rho_i)| \geqslant c)$ are in $\mathcal{O}(\frac{1}{d})$ [21, 25]. However, having a full characterization of $P(C_j(\rho_i))$ allows us to compute tail probabilities and obtain a much tighter bound. For instance, for $U$ being Haar random over $\mathbb{U}(d)$, we find

**Corollary 2.** *Let $C_j(\rho_i)$ be the expectation value of a Haar random QNN as in Eq. (1). Assuming that there exists a parametrized gate in $U$ of the form $e^{-i\theta H}$ for some Pauli operator $H$, then*

$$P(|C_j(\rho_i)| \geqslant c),\, P(|\partial_\theta C_j(\rho_i)| \geqslant c) \in \mathcal{O}\left(\frac{1}{c e^{dc^2}\sqrt{d}}\right).$$

Corollary 2 indicates that the QNN outputs, and their gradients, actually concentrate with a probability which vanishes exponentially with $d$. In an $n$-qubit system, where $d = 2^n$, then $P(|C_j(\rho_i)| \geqslant c)$ and $P(|\partial_\theta C_j(\rho_i)| \geqslant c)$ are doubly exponentially vanishing with $n$. The tightness of our bound arises from the fact that Chebyshev's inequality is loose for highly narrow Gaussian distributions. Corollary 2 also implies that the narrow gorge region of the landscape [25], i.e., the fraction of non-concentrated $C_j(\rho_i)$ values, also decreases exponentially with $d$.

In the Methods we furthermore show how our results can be used to study the concentration of functions of QNN outcomes, e.g., standard loss functions used in the literature, like the mean-squared error.

### H. Implications for $t$-designs

We now note that our results allow us to characterize the output distribution for QNN's that form $t$-designs, i.e., for QNNs whose unitary distributions have the same properties up to the first $t$ moments as sampling random unitaries from $\mathbb{U}(d)$ with respect to the Haar measure. With this in mind, one can readily see that the following corollary holds.

**Corollary 3.** *Let $U$ be drawn from a $t$-design. Then, under the same conditions for which Theorems 1, 2 and 3 hold, the vector $\mathscr{C}$ matches the first $t$ moments of a GP.*

Corollary 3 extends our results beyond the strict condition of the QNN being Haar random to being a $t$-design, which is a more realistic assumption [26–28]. In particular, we can study the concentration phenomenon in $t$-designs: using an extension of Chebyshev's inequality to higher order moments leads to $P(|C_j(\rho_i)| \geqslant c)$, $P(|\partial_\theta C_j(\rho_i)| \geqslant c) \in \mathcal{O}\left(\frac{(2\lfloor\frac{t}{2}\rfloor)!}{2^{\lfloor\frac{t}{2}\rfloor}(dc^2)^{\lfloor\frac{t}{2}\rfloor}(\lfloor\frac{t}{2}\rfloor)!}\right)$ (see the SI for a proof). Note that for $t = 2$ we recover the known concentration result for barren plateaus, but for $t \geqslant 4$ we obtain new polynomial-in-$d$-tighter bounds.

### I. Generalized datasets

Up to this point we have derived our theorems by imposing strict conditions on the overlaps between every pair of states in the dataset. However, we can extend these results to the cases where the conditions are only met on average when sampling states over $\mathscr{D}$.

**Theorem 5.** *The results of Theorems 1 and 2 will hold, on average, if $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}} \text{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\text{poly}(\log(d))}\right)$ and $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}} \text{Tr}[\rho_i\rho_{i'}] = \frac{1}{d}$, respectively.*

As discussed in the Methods section these extensions make our results more practical as $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}} \text{Tr}[\rho_i\rho_{i'}] \in$

$\Omega\left(\frac{1}{\text{poly}(\log(d))}\right)$ holds on standard multi-class classification settings [29, 32], while $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}\operatorname{Tr}[\rho_i\rho_{i'}]=\frac{1}{d}$ holds when the dataset is composed of Haar random states.

## III. DISCUSSION AND OUTLOOK

In this manuscript we have shown that under certain conditions, the output distribution of deep Haar random QNNs converges to a Gaussian process in the limit of large Hilbert space dimension. While this result had been conjectured in [13], a formal proof was still lacking. We remark that although our result mirrors its classical counterpart –that certain classical NNs form GPs–, there exist nuances that differentiate our findings from the classical case. For instance, we need to make assumptions on the states processed by the QNN, as well as on the measurement operator. Moreover, some of these assumptions are unavoidable, as Haar random QNNs will not necessarily always converge to a GP. As an example, we have that if $O_i$ is a projector onto a computational basis state, then one recovers a Porter-Thomas distribution [38]. Ultimately, these subtleties arise because the entries of unitary matrices are not independent. In contrast, classical NNs are not subject to this constraint.

It is worth noting that our theorems have further implications beyond those discussed here. We envision that our methods and results will be useful in more general settings where Haar random unitaries / $t$-designs are considered, such as quantum information scramblers and black holes [24, 39, 40], many-body physics [41], quantum decouplers and quantum error correction [42]. Finally, we leave for future work several potential generalizations of our results. For instance, one could envision proving a general result that combines our Theorems 1, 2, and 3 into a single setting. Moreover, it could be interesting to study if GPs arise in other architectures such as quantum convolutional neural networks [43], or re-uploading circuits [30, 44], among others.

## IV. METHODS

### A. Infinitely-wide neural networks as Gaussian processes

Here we will briefly review the seminal work of Ref. [4], which proved that artificial NNs with a single infinitely-wide hidden layer form GPs. Our main motivation for reviewing this result is that, as we will see below, the simple technique used in its derivation cannot be directly applied to the quantum case.

For simplicity let us consider a network consisting of a single input neuron, $N_h$ hidden neurons, and a single output neuron (see Fig. 1). The input of the network is $x\in\mathbb{R}$, and the output is given by

$$f(x)=b+\sum_{l=1}^{N_h}v_l h_l(x)\,, \tag{12}$$

where $h_l(x)=\phi(a_l+u_l x)$ models the action of each neuron in the hidden layer. Specifically, $u_l$ is the weight between the input neuron and the $l$-th hidden neuron, $a_l$ is the respective bias and $\phi$ is some (non-linear) activation function such as the hyperbolic tangent or the sigmoid function. Similarly, $v_l$ is the weight connecting the $l$-th hidden neuron to the output neuron, and $b$ is the output bias. From Eq. (12) we can see that the output of the NN is a weighted sum of the hidden neurons' outputs plus some bias.

Next, let us assume that the $v_l$ and $b$ are taken i.i.d. from a Gaussian distribution with zero mean and standard deviations $\sigma_v/\sqrt{N_h}$ and $\sigma_b$, respectively. Likewise, one can assume that the hidden neuron weights and biases are taken i.i.d. from some Gaussian distributions. Then, in the limit of $N_h\to\infty$, one can conclude via the central limit theorem that, since the NN output is a sum of infinitely many i.i.d. random variables, then it will converge to a Gaussian distribution with zero mean and variance $\sigma_b^2+\sigma_v^2\mathbb{E}[h_l(x)^2]$. Similarly, it can be shown that in the case of multiple inputs $x_1,\dots,x_m$ one gets a multivariate Gaussian distribution for $f(x_1),\dots,f(x_m)$, i.e., a GP [4].

Naively, one could try to mimic the technique in Ref. [4] to prove our main results. In particular, we could start by noting that $C_j(\rho_i)$ can always be expressed as

$$C_j(\rho_i)=\sum_{k,k',r,r'=1}^{d}u_{kk'}\rho_{k'r}u^*_{r'r}o_{r'k}\,, \tag{13}$$

where $u_{kk'}$, $u^*_{r'r}$, $\rho_{k'r}$ and $o_{r'l}$ are the matrix entries of $U$ and $U^\dagger$, $\rho$ and $O$, respectively. Although Eq. (13) is a summation over a large number of random variables, we cannot apply the central limit theorem (or its variants) here, since the matrix entries $u_{kk'}$ and $u^*_{rr'}$ are not independent [33].

In fact, the correlation between the entries in the same row, or column, of a Haar random unitary are of order $\frac{1}{d}$, while those in different rows, or columns, are of order $\frac{1}{d^2}$. This small, albeit critical, difference makes it such that we cannot simply use the central limit theorem to prove that $\mathscr{C}$ converges to a GP. Instead, we need to rely on the techniques described in the main text.

### B. Learning with the Gaussian process

In this section we will review the basic formalism for learning with Gaussian processes. Let $\boldsymbol{C}$ be a Gaussian process. Then, by definition, given a collection of inputs $\{x_i\}_{i=1}^m$, $\boldsymbol{C}$ is determined by its $m$-dimensional mean vector $\boldsymbol{\mu}$, and its $m \times m$-dimensional covariance matrix $\boldsymbol{\Sigma}$. In what follows we will assume that the mean of $\boldsymbol{C}$ is zero, and that the entries of its covariance matrix are expressed as $\kappa(x_i, x_{i'})$. That is,

$$P\left(\begin{pmatrix} C(x_1) \\ \vdots \\ C(x_m) \end{pmatrix}\right) = \mathcal{N}\left(\boldsymbol{\mu} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \kappa(x_1, x_1) & \cdots & \kappa(x_1, x_m) \\ \vdots & & \vdots \\ \kappa(x_m, x_1) & \cdots & \kappa(x_m, x_m) \end{pmatrix}\right).$$

The previous allows us to know that, *a priori*, the distribution of values for any $f(x_i)$ will take the form

$$P(C(x_i)) = \mathcal{N}(0, \sigma_i^2)\,, \tag{14}$$

with $\sigma_i^2 = \kappa(x_i, x_i)$.

Now, let us consider the task of using $m$ observations, which we will collect in a vector $\boldsymbol{y}$, to predict the value at $x_{m+1}$. First, if the observations are noiseless, then $\boldsymbol{y} = (y(x_1), \cdots, y(x_m))$ is equal to $\boldsymbol{C} = (C(x_1), \cdots, C(x_m))$. That is, $\boldsymbol{C} = \boldsymbol{y}$. Here, we can use the fact that $C$ forms a Gaussian process to find [9, 45]

$$\begin{aligned} P(C(x_{m+1})|\boldsymbol{C}) &= P(C(x_{m+1})|C(x_1), C(x_2), \dots, C(x_m)) \\ &= \mathcal{N}\left(\mu(C(x_{m+1})), \sigma^2(C(x_{m+1}))\right)\,, \end{aligned} \tag{15}$$

where $\mu(C(x_{m+1}))$ and $\sigma^2(C(x_{m+1}))$ respectively denote the mean and variance of the associated Gaussian probability distribution, and which are given by

$$\mu(C(x_{m+1})) = \boldsymbol{m}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{C} \tag{16}$$

$$\sigma^2(C(x_{m+1})) = \sigma_{m+1}^2 - \boldsymbol{m}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{m}\,. \tag{17}$$

The vector $\boldsymbol{m}$ has entries $\boldsymbol{m}_i = \kappa(x_{m+1}, x_i)$. We can compare Eqs. (14) and (15) to see that using Bayesian statistics to obtain the predictive distribution of $P(C(x_{m+1})|\boldsymbol{C})$ shifts the mean from zero to $\boldsymbol{m}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{C}$ and the variance is decreased from $\sigma_{m+1}^2$ by a quantity $\boldsymbol{m}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{m}$. The decrease in variance follows from the fact that we are incorporating knowledge about the observations, and thus decreasing the uncertainty.

In a realistic scenario, we can expect that noise will occur during our observation procedure. For simplicity we model this noise as Gaussian noise, so that $y(x_i) = C(x_i) + \varepsilon_i$, where the noise terms $\varepsilon_i$ are assumed to be independently drawn from the same distribution $P(\varepsilon_i) = \mathcal{N}(0, \sigma_N^2)$. Now, since we have assumed that the noise is drawn independently, we know that the likelihood of obtaining a set of observations $\boldsymbol{y}$ given the model values $\boldsymbol{C}$ is given by $P(\boldsymbol{y}|\boldsymbol{C}) = \mathcal{N}(\boldsymbol{C}, \sigma_N^2 \mathbb{1})$. In this case, we can find the probability distribution [9, 45]

$$\begin{aligned} P(C(x_{m+1})|\boldsymbol{C}) &= \int d\boldsymbol{C} P(x_{m+1}|\boldsymbol{C}) P(\boldsymbol{C}|\boldsymbol{y}) \\ &= \int d\boldsymbol{C} P(C(x_{m+1})|\boldsymbol{C}) P(\boldsymbol{y}|\boldsymbol{C}) P(\boldsymbol{C})/P(\boldsymbol{y}) \\ &= \mathcal{N}\left(\widetilde{\mu}(C(x_{m+1})), \widetilde{\sigma}^2(C(x_{m+1}))\right)\,, \end{aligned} \tag{18}$$

where now we have

$$\widetilde{\mu}(C(x_{m+1})) = \boldsymbol{m}^T \cdot (\boldsymbol{\Sigma} + \sigma_N^2 \mathbb{1})^{-1} \cdot \boldsymbol{C} \tag{19}$$

$$\widetilde{\sigma}^2(C(x_{m+1})) = \sigma_{m+1}^2 - \boldsymbol{m}^T \cdot (\boldsymbol{\Sigma} + \sigma_N^2 \mathbb{1})^{-1} \cdot \boldsymbol{m}\,. \tag{20}$$

In the first and the second equality we have used the explicit decomposition of the probability, along with Bayes and marginalization rules. We can see that the probability is still governed by a Gaussian distribution but where the inverse of $\boldsymbol{\Sigma}$ has been replaced by the inverse of $\boldsymbol{\Sigma} + \sigma_N^2 \mathbb{1}$.

### C. Concentration of functions of QNN outcomes

In the main text we have evaluated the distribution of QNN outcomes and their linear combinations. However, in many cases one is also interested in evaluating a function of the elements of $\mathscr{C}$. For instance, in a standard QML setting the QNN outcomes are used to compute some loss function $\mathcal{L}(\mathscr{C})$ which one wishes to optimize [10–14]. While we do not aim here at exploring all possible relevant functions $\mathcal{L}$, we will present two simple examples that showcase how our results can be used to study the distribution of $\mathcal{L}(\mathscr{C})$, as well as its concentration.

First, let us consider the case when $\mathcal{L}(C_j(\rho_i)) = C_j(\rho_i)^2$. It is well known that given a random variable with a Gaussian distribution $\mathcal{N}(0, \sigma^2)$, then its square follows a Gamma distribution $\Gamma(\frac{1}{2}, 2\sigma^2)$. Hence, we know that $P(\mathcal{L}(C_j(\rho_i))) = \Gamma(\frac{1}{2}, 2\sigma^2)$. Next, let us consider the case when $\mathcal{L}(C_j(\rho_i)) = (C_j(\rho_i) - y_i)^2$ for $y_i \in [-1, 1]$. This case is relevant for supervised learning as the mean-squared error loss function is composed of a linear combination of such terms. Here, $y_i$ corresponds to the label associated to the state $\rho_i$. We can exactly compute all the moments of $\mathcal{L}(C_j(\rho_i))$ as

$$\mathbb{E}_G[\mathcal{L}(C_j(\rho_i))^k] = \sum_{r=0}^{2k} \binom{2k}{r} \mathbb{E}_G[C_j(\rho_i)^r](-y_i)^{2k-r}\,, \tag{21}$$

for $G = \mathbb{U}(d), \mathbb{O}(d)$. We can then use Lemma 2 to obtain

$$\mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i)^r] = \frac{r!}{d^{r/2} 2^{r/2} (r/2)!} = \frac{\mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)^r]}{2^{r/2}}\,,$$

if $r$ is even, and $\mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i)^r] = \mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)^r] = 0$ if $r$ is odd. We obtain

$$\mathbb{E}_{\mathbb{U}(d)}[\mathcal{L}(C_j(\rho_i))^k] = \frac{2^k}{(-d)^k} M\left(-k, \frac{1}{2}, -\frac{dy^2}{2}\right), \quad (22)$$

with $M$ the Kummer's confluent hypergeometric function.

Furthermore, we can also study the concentration of $\mathcal{L}(C_j(\rho_i))$ and show that $P\left(|\mathcal{L}(C_j(\rho_i)) - \mathbb{E}_{\mathbb{U}(d)}(\mathcal{L}(C_j(\rho_i)))| \geqslant c\right)$, where the average $\mathbb{E}_{\mathbb{U}(d)}(\mathcal{L}(C_j(\rho_i))) = y_i^2 + \frac{1}{d}$, is in $\mathcal{O}\left(\frac{1}{|\sqrt{c}+y_i| e^{d|\sqrt{c}+y_i|^2}\sqrt{d}}\right)$.

### D. Motivation for the generalized datasets

In Theorem 5 we generalized the results of Theorems 1 and 2 to hold on average when a) $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}} \operatorname{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right)$ and b) $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}} \operatorname{Tr}[\rho_i\rho_{i'}] = \frac{1}{d}$, respectively. Interestingly, these two cases have practical relevance. Let us start with Case a). Consider a multiclass classification problem, where each state $\rho_i$ in $\mathscr{D}$ belongs to one of $Y$ classes, with $Y \in \mathcal{O}(1)$, and where the dataset is composed of an (approximately) equal number of states from each class. That is, for each $\rho_i$ we can assign a label $y_i = 1, \dots, Y$. Then, we assume that the classes are well separated in the Hilbert feature space, a standard and sufficient assumption for the model to be able to solve the learning task [29, 32]. By well separated we mean that

$$\operatorname{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right), \quad \text{if} \quad y_i = y_{i'}, \quad (23)$$

$$\operatorname{Tr}[\rho_i\rho_{i'}] \in \mathcal{O}\left(\frac{1}{2^n}\right), \quad \text{if} \quad y_i \neq y_{i'}. \quad (24)$$

In this case, it can be verified that for any pair of states $\rho_i$ and $\rho_{i'}$ sampled from $\mathscr{D}$, one has $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}[\operatorname{Tr}[\rho_i\rho_{i'}]] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right)$.

Next, let us evaluate Case b). Such situation arises precisely if the sates in $\mathscr{D}$ are Haar random states. Indeed, we can readily show that

$$\begin{aligned}
\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}[\operatorname{Tr}[\rho_i\rho_{i'}]] &= \mathbb{E}_{\rho_i,\rho_{i'}\sim\text{Haar}}[\operatorname{Tr}[\rho_i\rho_{i'}]] \\
&= \int_{\mathbb{U}(d)} d\mu(U)d\mu(V) \operatorname{Tr}\left[U\rho_0 U^\dagger V\rho_0' V^\dagger\right] \\
&= \int_{\mathbb{U}(d)} d\mu(U) \operatorname{Tr}\left[U\rho_0 U^\dagger \rho_0'\right] \\
&= \frac{\operatorname{Tr}[\rho_0]\operatorname{Tr}[\rho_0']}{d} \\
&= \frac{1}{d}. \quad (25)
\end{aligned}$$

Here, in the first equality we have used that sampling Haar random pure states $\rho_i$ and $\rho_{i'}$ from the Haar measure is equivalent to taking two reference states $\rho_0$ and $\rho_0'$ and evolving them with Haar random unitaries. In the second equality we have used the left-invariance of the Haar measure, and in the third equality we have explicitly performed the integration (see the SI).

### E. Sketch of the proof of our main results

Since our main results are mostly based on Lemmas 1 and 2, we will here outline the main steps used to prove these Lemmas. In particular, to prove them we need to calculate, in the large $d$ limit, quantities of the form

$$\mathbb{E}_G\left[\operatorname{Tr}\left[U^{\otimes k}\Lambda(U^\dagger)^{\otimes k} O^{\otimes k}\right]\right], \quad (26)$$

for arbitrary $k$, and for $G = \mathbb{U}(d), \mathbb{O}(d)$. Here, the operator $\Lambda$ is defined as $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$, where the pure states $\rho_i$ belong to $\mathscr{D}$, and where $O$ is an operator in $\mathscr{O}$. The first moment ($k=1$), $\boldsymbol{\mu}$, and the second moments ($k=2$), $\boldsymbol{\Sigma}_{i,i'}^G$ can be directly computed using standard formulas for integration over the unitary and orthogonal groups (see the SI). This readily recovers the results in Lemma 1. However, for larger $k$ a direct computation quickly becomes intractable, and we need to resort to asymptotic Weingarten calculations. More concretely, let us exemplify our calculations for the unitary group and for the case when the states in the dataset are such that $\operatorname{Tr}\left[\rho_i\rho_{i_{i'}}\right] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right)$ for all $\rho_i, \rho_{i'} \in \mathscr{D}$. As shown in the SI, we can prove the following lemma.

**Lemma 3.** *Let $X$ be an operator in $\mathcal{B}(\mathcal{H}^{\otimes k})$, the set of bounded linear operators acting on the $k$-fold tensor product of a $d$-dimensional Hilbert space $\mathcal{H}$. Let $S_k$ be the symmetric group on $k$ items, and let $P_d$ be the subsystem permuting representation of $S_k$ in $\mathcal{H}^{\otimes k}$. Then, for large Hilbert space dimension ($d \to \infty$), the twirl of $X$ over $\mathbb{U}(d)$ is*

$$\begin{aligned}
\mathbb{E}_{\mathbb{U}(d)}[U^{\otimes k}X(U^\dagger)^{\otimes k}] =& \frac{1}{d^k}\sum_{\sigma\in S_k} \operatorname{Tr}[XP_d(\sigma)]P_d(\sigma^{-1}) \\
&+ \frac{1}{d^k}\sum_{\sigma,\Pi\in S_k} c_{\sigma,\pi} \operatorname{Tr}[XP_d(\sigma)]P_d(\pi),
\end{aligned}$$

*where the constants $c_{\sigma,\Pi}$ are in $\mathcal{O}(1/d)$.*

We recall that the subsystem permuting representation of a permutation $\sigma \in S_k$ is

$$P_d(\sigma) = \sum_{i_1,\dots,i_k=0}^{d-1} |i_{\sigma^{-1}(1)},\dots,i_{\sigma^{-1}(k)}\rangle\langle i_1,\dots,i_k|. \quad (27)$$

Lemma 3 implies that (26) is equal to

$$\begin{aligned}
&\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\left[U^{\otimes k}\Lambda(U^\dagger)^{\otimes k}O^{\otimes k}\right]\right] \\
&\quad= \frac{1}{d^k}\sum_{\sigma\in S_k}\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big] \\
&\quad+ \frac{1}{d^k}\sum_{\sigma,\Pi\in S_k} c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]\,.
\end{aligned} \tag{28}$$

We now note that, by definition, since $O$ is traceless and such that $O^2=\mathbb{1}$, then $\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]=0$ for odd $k$ (and for all $\sigma$). This result implies that all the odd moments are exactly zero, and also that the non-zero contributions in Eq. (28) for the even moments come from permutations consisting of cycles of even length. We remark that as a direct consequence, the first moment, $\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\big[U\rho_i U^\dagger O\big]\right]$, is zero for any $\rho_i\in\mathscr{D}$, and thus we have $\boldsymbol{\mu}=\mathbf{0}$. To compute higher moments, we show that $\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]=d^r$ if $k$ is even and $\sigma$ is a product of $r$ disjoint cycles of even length. The maximum of $\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]$ is therefore achieved when $r$ is maximal, i.e., when $\sigma$ is a product of $k/2$ disjoint transpositions (cycles of length two), leading to $\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]=d^{k/2}$. Then, we look at the factors $\mathrm{Tr}[\Lambda P_d(\sigma)]$ and include them in the analysis. We have that for all $\pi$ and $\sigma$ in $S_k$,

$$\begin{aligned}
&\frac{1}{d^k}\Big|(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big] \\
&+c_{\sigma^{-1},\pi}\,\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big])\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]\Big| \in \mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right).
\end{aligned} \tag{29}$$

Moreover, since $\mathrm{Tr}[\rho_i\rho_{i'}]\in\Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all pair of states $\rho_i,\rho_{i'}\in\mathscr{D}$, it holds that

$$\frac{1}{d^k}\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big]\in\Omega\left(\frac{1}{d^{k/2}}\right), \tag{30}$$

if $\sigma$ is a product of $k/2$ disjoint transpositions, and

$$\begin{aligned}
&\frac{1}{d^k}\Big|\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big]+ \\
&\qquad\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big]\,\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]\Big|\in\mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right),
\end{aligned} \tag{31}$$

for any other $\sigma$. We remark that if $\sigma$ consist only of transpositions, then it is its own inverse, that is, $\sigma=\sigma^{-1}$.

It immediately follows that for fixed $k$ and $d\to\infty$, the second sum in Eq. (28) is suppressed at least inversely proportional to the dimension of the Hilbert space with respect to the first one (i.e. exponentially in the number of qubits for QNNs made out of qubits). Likewise, the contributions in the first sum in (28) coming from permutations that are not the product of $k/2$ disjoint transpositions are also suppressed inversely proportional to the Hilbert space dimension. Therefore, in the large $d$ limit we arrive at

$$\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\left[U^{\otimes k}\Lambda(U^\dagger)^{\otimes k}O^{\otimes k}\right]\right]=\frac{1}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\{t,t'\}\in\sigma}\mathrm{Tr}[\rho_t\rho_{t'}]\,, \tag{32}$$

where we have defined as $T_k\subseteq S_k$ the set of permutations which are exactly given by a product of $k/2$ disjoint transpositions. Note that this is precisely the statement in Lemma 2.

From here we can easily see that if every state in $\Lambda$ is the same, i.e., if $\rho_i=\rho$ for $i=1,\dots,k$, then $\mathrm{Tr}[\rho_t\rho_{t'}]=1$ for all $t,t'$, and we need to count how many terms are there in Eq. (32). Specifically, we need to count how many different ways there exist to split $k$ elements into pairs (with $k$ even). A straightforward calculation shows that

$$\sum_{\sigma\in T_k}\prod_{\{t,t'\}\in\sigma}1=\frac{1}{(k/2)!}\binom{k}{2,2,\dots,2}=\frac{k!}{2^{k/2}(k/2)!}\,. \tag{33}$$

Thus, we arrive at

$$\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\left[U^{\otimes k}\Lambda(U^\dagger)^{\otimes k}O^{\otimes k}\right]\right]=\frac{1}{d^{k/2}}\frac{k!}{2^{k/2}(k/2)!}\,. \tag{34}$$

Identifying $\sigma^2=\frac{1}{d}$ implies that the moments $\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\left[U\rho U^\dagger O\right]^k\right]$ exactly match those of a Gaussian distribution $\mathcal{N}(0,\sigma^2)$.

To prove that these moments unequivocally determine the distribution of $\mathscr{C}$, we use Carleman's condition.

**Lemma 4** (Carleman's condition, Hamburger case [34])**.** *Let $\gamma_k$ be the (finite) moments of the distribution of a random variable $X$ that can take values on the real line $\mathbb{R}$. These moments determine uniquely the distribution of $X$ if*

$$\sum_{k=1}^{\infty}\gamma_{2k}^{-1/2k}=\infty\ . \tag{35}$$

Explicitly, we have

$$\begin{aligned}
\sum_{k=1}^{\infty}\left(\frac{1}{d^k}\frac{(2k)!}{2^k k!}\right)^{-1/2k} &= \sqrt{2d}\sum_{k=1}^{\infty}\left((2k)\cdots(k+1)\right)^{-1/2k} \\
&\geqslant \sum_{k=1}^{\infty}\left((2k)^k\right)^{-1/2k} \\
&= \sum_{k=1}^{\infty}\frac{1}{\sqrt{2k}}=\infty\,.
\end{aligned} \tag{36}$$

Hence, according to Lemma 4, Carleman's condition is satisfied, and $P(C_j(\rho_i)$ is distributed following a Gaussian distribution.

A similar argument can be given to show that the moments of $\mathscr{C}$ match those of a GP. Here, we need to compare Eq. (32) with the $k$-th order moments of a GP, which are provided by Isserlis theorem [46]. Specifically, if we want to compute a $k$-th order moment of a GP, then we have that $\mathbb{E}[X_1 X_2 \cdots X_k] = 0$ if $k$ is odd, and

$$\mathbb{E}[X_1 X_2 \cdots X_k] = \sum_{\sigma \in T_k} \prod_{\{t,t'\}\in\sigma} \mathrm{Cov}[X_t, X_{t'}]\,, \tag{37}$$

if $k$ is even. Clearly, Eq. (32) matches Eq. (37) by identifying $\mathrm{Cov}[X_t, X_{t'}] = \frac{\mathrm{Tr}[\rho_t \rho_{t'}]}{d}$. We can again prove that these moments uniquely determine the distribution of $\mathscr{C}$ from the fact that since its marginal distributions are determinate via Carleman's condition (see above), then so is the distribution of $\mathscr{C}$ [34]. Hence, $\mathscr{C}$ forms a GP.


## ACKNOWLEDGEMENTS

We acknowledge Francesco Caravelli, Frédéric Sauvage, Lorenzo Leone, and Cinthia Huerta for useful conversations. D.G-M. was supported by the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory (LANL) under project number 20230049DR. M.L. acknowledges support by the Center for Nonlinear Studies at Los Alamos National Laboratory (LANL). M.C. acknowledges support by the LDRD program of LANL under project number 20230527ECR. This work was also supported by LANL ASC Beyond Moore's Law project.



[1] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions, Journal of big Data **8**, 1 (2021).

[2] D. Khurana, A. Koli, K. Khatter, and S. Singh, Natural language processing: State of the art, current trends and challenges, Multimedia tools and applications **82**, 3713 (2023).

[3] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland1, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, Highly accurate protein structure prediction with alphafold, Nature **596**, 583 (2021).

[4] R. M. Neal, Priors for infinite networks, Bayesian learning for neural networks , 29 (1996).

[5] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein, Deep neural networks as gaussian processes, arXiv preprint arXiv:1711.00165 (2017).

[6] R. Novak, L. Xiao, Y. Bahri, J. Lee, G. Yang, D. A. Abolafia, J. Pennington, and J. Sohl-dickstein, Bayesian deep convolutional networks with many channels are gaussian processes, in *International Conference on Learning Representations* (2019).

[7] J. Hron, Y. Bahri, J. Sohl-Dickstein, and R. Novak, Infinite attention: Nngp and ntk for deep attention networks, in *International Conference on Machine Learning* (PMLR, 2020) pp. 4376–4386.

[8] G. Yang, Wide feedforward or recurrent neural networks of any architecture are gaussian processes, Advances in Neural Information Processing Systems **32** (2019).

[9] C. E. Rasmussen, C. K. Williams, *et al.*, *Gaussian processes for machine learning*, Vol. 1 (Springer, 2006).

[10] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[11] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, Nature Computational Science 10.1038/s43588-022-00311-3 (2022).

[12] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nature Reviews Physics **3**, 625–644 (2021).

[13] J. Liu, F. Tacchino, J. R. Glick, L. Jiang, and A. Mezzacapo, Representation learning via quantum neural tangent kernels, arXiv preprint arXiv:2111.04225 (2021).

[14] M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers* (Springer, 2021).

[15] M. Schuld and N. Killoran, Is quantum advantage the right goal for quantum machine learning?, arXiv preprint arXiv:2203.01340 (2022).

[16] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, arXiv preprint arXiv:2109.11676 (2021).

[17] E. R. Anschuetz, H.-Y. Hu, J.-L. Huang, and X. Gao, Interpretable quantum advantage in neural sequence learning, arXiv preprint arXiv:2209.14353 (2022).

[18] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, Nature Computational Science **1**, 403 (2021).

[19] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).
[20] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications **9**, 1 (2018).
[21] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nature Communications **12**, 1 (2021).
[22] C. O. Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).
[23] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, Entanglement devised barren plateau mitigation, Physical Review Research **3**, 033090 (2021).
[24] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, Barren plateaus preclude learning scramblers, Physical Review Letters **126**, 190501 (2021).
[25] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, Quantum Science and Technology **7**, 045015 (2022).
[26] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Communications in Mathematical Physics **291**, 257 (2009).
[27] A. Harrow and S. Mehraban, Approximate unitary $t$-designs by short random quantum circuits using nearest-neighbor and long-range gates, arXiv preprint arXiv:1809.06957 (2018).
[28] J. Haferkamp, Random quantum circuits are approximate unitary $t$-designs in depth $O\left(nt^{5+o(1)}\right)$, arXiv preprint arXiv:2203.16571 (2022).
[29] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, Quantum embeddings for machine learning, arXiv preprint arXiv:2001.03622 (2020).
[30] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).
[31] L. Schatzki, A. Arrasmith, P. J. Coles, and M. Cerezo, Entangled datasets for quantum machine learning, arXiv preprint arXiv:2109.03400 (2021).
[32] M. Larocca, F. Sauvage, F. M. Sbahi, G. Verdon, P. J. Coles, and M. Cerezo, Group-invariant quantum machine learning, PRX Quantum **3**, 030341 (2022).
[33] D. Petz and J. Réffy, On asymptotics of large haar distributed unitary matrices, Periodica Mathematica Hungarica **49**, 103 (2004).
[34] C. Kleiber and J. Stoyanov, Multivariate distributions and the moment problem, Journal of Multivariate Analysis **113**, 7 (2013).
[35] M. Schuld, Quantum machine learning models are kernel methods, arXiv preprint arXiv:2101.11020 (2021).
[36] 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 **567**, 209 (2019).
[37] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, Exponential concentration and untrainability in quantum kernel methods, arXiv preprint arXiv:2208.11060 (2022).
[38] C. E. Porter and R. G. Thomas, Fluctuations of nuclear reaction widths, Physical Review **104**, 483 (1956).
[39] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, Journal of High Energy Physics **9**, 120 (2007).
[40] S. F. Oliviero, L. Leone, S. Lloyd, and A. Hamma, Black hole complexity, unscrambling, and stabilizer thermal machines, arXiv preprint arXiv:2212.11337 (2022).
[41] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, Physical Review X **8**, 021014 (2018).
[42] W. Brown and O. Fawzi, Decoupling with random quantum circuits, Communications in mathematical physics **340**, 867 (2015).
[43] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nature Physics **15**, 1273 (2019).
[44] S. Jerbi, L. J. Fiderer, H. Poulsen Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, Quantum machine learning beyond kernel methods, Nature Communications **14**, 517 (2023).
[45] R. Mukherjee, F. Sauvage, H. Xie, R. Löw, and F. Mintert, Preparation of ordered states in ultra-cold gases using bayesian optimization, New Journal of Physics **22**, 075001 (2020).
[46] L. Isserlis, On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, Biometrika **12**, 134 (1918).
[47] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, Communications in Mathematical Physics **264**, 773 (2006).
[48] F. Caravelli, F. L. Traversa, and M. Di Ventra, Complex dynamics of memristive circuits: Analytical results and universal slow relaxation, Physical Review E **95**, 022140 (2017).
[49] Z. Puchala and J. A. Miszczak, Symbolic integration with respect to the haar measure on the unitary groups, Bulletin of the Polish Academy of Sciences Technical Sciences **65**, 21 (2017).
[50] B. Collins, Moments and cumulants of polynomial random variables on unitarygroups, the itzykson-zuber integral, and free probability, International Mathematics Research Notices **2003**, 953 (2003).

## SUPPLEMENTAL INFORMATION FOR "*DEEP QUANTUM NEURAL NETWORKS FORM GAUSSIAN PROCESSES*"

In this Supplemental Information (SI) we present detailed proofs of our main results. First, in Supp. Info. A we introduce preliminary definitions that will be used throughout the rest of this SI. In Supp. Info. B we review the basics of the Weingarten calculus that will allow us to compute averages when sampling Haar random unitaries over $\mathbb{U}(d)$ and $\mathbb{O}(d)$. Next, in Supp. Info. C and D we respectively present results for twirling over the unitary and orthogonal groups. Then, Supp. Info E contains the proof of Lemma 1, while Supp. Info. F contains that of Lemma 2. In Supp. Info. G we derive Corollary 1. From here, we use these results to prove Theorems 1, 2, 3, and 4 in Supp. Info. H, I, J, and K, respectively. Supp. Info. L contains a proof for Corollary 2, and Supp. Info. M a proof for Corollary 3. Finally, we present the derivation of Theorem 5 in Supp. Info. N.

### Supp. Info. A: Preliminaries

#### 1. Useful definitions

Let $\mathcal{H}$ be a $d$-dimensional Hilbert space. We denote as $\mathcal{B}(\mathcal{H})$ the space of bounded linear operators in $\mathcal{H}$, and by $GL(\mathcal{H})$ the general linear group of $d\times d$ complex non-singular matrices, which contains all invertible linear transformations on $\mathcal{H}$. We now introduce the following definitions.

**Representation.** Let $G$ be a group. A *representation* $R$ of $G$ on $\mathcal{H}$ is a group homomorphism $R: G\rightarrow GL(\mathcal{H})$. Given $R$, we recall that we can always build the *k-fold tensor representation* of $G$, acting on $\mathcal{H}^{\otimes k}$ as $R(g)^{\otimes k}$ for any $g\in G$. It is not hard to see that if $R$ is a valid representation, then its $k$-fold tensor product is also a representation.

**Character.** A useful way of characterizing representations of a group $G$ is through their associated *characters*. Given a representation $R$, the character $\chi: G\rightarrow \mathbb{C}$ is a function that associates to each element of $g\in G$ a complex number

$$\chi(g)=\mathrm{Tr}[R(g)]\,. \tag{A1}$$

**Commutant.** Given some representation $R$ of $G$, we define its *k-th order commutant*, denoted as $\mathcal{C}^{(k)}(G)$, to be the vector subspace of the space of linear operators on $\mathcal{H}^{\otimes k}$ that commutes with $R(g)^{\otimes k}$ for all $g$ in $G$. That is,

$$\mathcal{C}^{(k)}(G)=\{A\in\mathcal{B}(\mathcal{H}^{\otimes k})\quad|\quad [A,R(g)^{\otimes k}]=0\,,\quad \forall g\in G\}\,. \tag{A2}$$

Then, we denote as $\mathcal{S}^{(k)}(G)$ a basis for $\mathcal{C}^{(k)}(G)$, which consists of $D$ elements.

**Useful operators.** Let us now introduce two useful operators acting on $\mathcal{H}^{\otimes 2}$. The first is the SWAP operator, whose action is to permute the two copies of $\mathcal{H}$. In matrix notation, we have

$$\mathrm{SWAP}=\sum_{i_1,i_2=0}^{d-1}|i_2,i_1\rangle\langle i_1,i_2|\,, \tag{A3}$$

and one can readily verify that for any $A$, $B$ in $\mathcal{B}(\mathcal{H}^{\otimes 2})$ one has

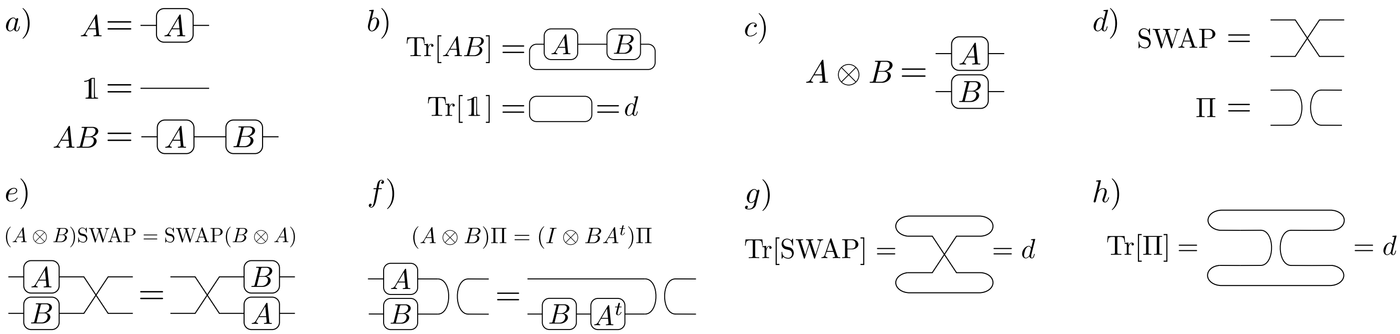
$$(A\otimes B)\mathrm{SWAP}=\mathrm{SWAP}(B\otimes A)\,. \tag{A4}$$

The second operator, which we denote as $\Pi$, acts on $\mathcal{H}^{\otimes 2}$ as

$$\Pi=\sum_{i_1,i_2=0}^{d-1}|i_1,i_1\rangle\langle i_2,i_2|\,. \tag{A5}$$

One can see that $\Pi$ is proportional to the projector onto the maximally-entangled Bell state between the two copies of $\mathcal{H}$. That is, defining as $|\Phi^+\rangle$ the Bell state between the two Hilbert spaces,

$$|\Phi^+\rangle=\frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}|i,i\rangle\,, \tag{A6}$$




Supplemental Figure 1. **Matrix and operations as tensor-network diagrams.** a) An operator $A$ acting on $\mathcal{H}$ is a $d \times d$ matrix which represents a linear map from $\mathcal{H} \to \mathcal{H}$. Such a mapping can be pictured as a node with two $d$-dimensional edges. The $d$-dimensional identity operator is simply a line. The product of the matrices is drawn by a node connected to another node. b) When taking the trace of a matrix, we connect its output edge with its input one. The trace of the identity, visualized as a closed loop, is equal to $d$. c) The tensor product between matrices is represented as their tensors stacked one on top of the other. d) We show the tensor representation of the SWAP and $\Pi$ operators respectively defined in Eqs. (A3) and (A5). e) and f) Respectively show Eqs. (A4) and (A7) . In g) and h) we use the tensor representation to compute the traces of the SWAP and $\Pi$ operators.

then one has $\Pi = d\,|\Phi\rangle\langle\Phi|$. However, we note that $\Pi$ is not a projector itself as $\Pi^2 = d\Pi$. The identification of $\Pi$ with the Bell state shows that $\Pi$ satisfies the so-called *ricochet* property,

$$(A \otimes B)\Pi = (\mathbb{1} \otimes BA^T)\Pi = (AB^T \otimes \mathbb{1})\Pi\,, \tag{A7}$$

where $A^T$ denotes the transpose of $A$.

For the remainder of this SI, we will use standard tensor-network diagram notation to visually represent matrix operations (see Fig. 1).

## Supp. Info. B: Weingarten calculus

Let $G$ be a compact Lie group, and let $R$ be the fundamental (unitary) representation of $G$. That is, $R(g) = g$. Given an operator $X \in \mathcal{B}(\mathcal{H}^{\otimes k})$, we consider the task of computing its *twirl*, $\mathcal{T}_G^{(k)}[X]$, with respect to the $k$-fold tensor representation of $G$. That is,

$$\mathcal{T}_G^{(k)}[X] = \int_G d\mu(g) g^{\otimes k} X (g^\dagger)^{\otimes k}\,, \tag{B1}$$

where $d\mu(g)$ is the volume element of the Haar measure. Now, we can use the following proposition.

**Supplemental Proposition 1.** *Let $\mathcal{T}_G^{(k)}[X]$ be the twirl of an operator $X$ in $\mathcal{B}(\mathcal{H}^{\otimes k})$ with respect to a continuous unitary group $G$ acting on $\mathcal{H}$. Then, we have*

$$\mathcal{T}_G^{(k)}[X] \in \mathcal{C}^{(k)}(G)\,. \tag{B2}$$

*Proof.* First, let us note that if an operator $A$ belongs to $\mathcal{C}^{(k)}(G)$, then $h^{\otimes k} A (h^\dagger)^{\otimes k} = A$ for all $h \in G$. Then, let us compute

$$
\begin{aligned}
h^{\otimes k}\mathcal{T}_G^{(k)}[X](h^\dagger)^{\otimes k} &= \int_G d\mu(g) h^{\otimes k} g^{\otimes k} X (g^\dagger)^{\otimes k}(h^\dagger)^{\otimes k} \\
&= \int_G d\mu(g) (hg)^{\otimes k} X ((hg)^\dagger)^{\otimes k} \\
&= \int_G d\mu(g) (g)^{\otimes k} X (g^\dagger)^{\otimes k} \\
&= \mathcal{T}_G^{(k)}[X]\,,
\end{aligned} \tag{B3}
$$

where in the third line we have used the left-invariance of the Haar measure. That is, we have used the fact that for any integrable function $f(g)$ and for any $h \in G$, we have

$$
\int_G d\mu(g) f(hg) = \int_G d\mu(g) f(gh) = \int_G d\mu(g) f(g)\,. \tag{B4}
$$

Thus, we have shown that $\mathcal{T}_G^{(k)}[X] \in \mathcal{C}^{(k)}(G)$. □

From Supplemental Proposition 1, it follows that $\mathcal{T}_G^{(k)}[X]$ can be expressed as

$$
\mathcal{T}_G^{(k)}[X] = \sum_{\mu=1}^{D} c_\mu(X) P_\mu\,, \quad \text{with} \quad P_\mu \in \mathcal{S}^{(k)}(G)\,. \tag{B5}
$$

Hence, in order to solve Eq. (B5) one needs to determine the $D$ unknown coefficients $\{c_\mu(X)\}_{\mu=1}^D$. This can be achieved by finding $D$ equations to form a linear system problem. In particular, we note that the change $X \to P_\nu X$ for some $P_\nu \in \mathcal{S}^{(k)}(G)$ leads to

$$
\begin{aligned}
\mathcal{T}_G^{(k)}[P_\nu X] &= \int_G d\mu(g) g^{\otimes k} P_\nu X (g^\dagger)^{\otimes k} \\
&= \int_G d\mu(g) P_\nu g^{\otimes k} X (g^\dagger)^{\otimes k} \\
&= \sum_{\mu=1}^{D} c_\mu(X) P_\nu P_\mu\,,
\end{aligned} \tag{B6}
$$

where in the second line we have used the fact that $P_\nu$ belongs to the commutant $\mathcal{C}^{(k)}(G)$. Then, taking the trace on both sides of Eq. (B6) leads to

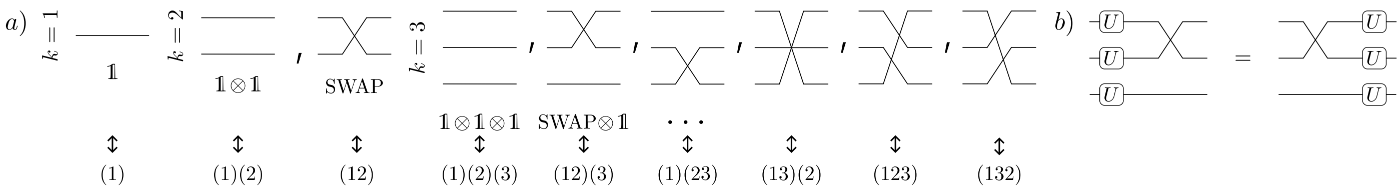
$$
\mathrm{Tr}[P_\nu X] = \sum_{\mu=1}^{D} \mathrm{Tr}[P_\nu P_\mu] c_\mu(X)\,, \tag{B7}
$$

where we used the fact that $\mathrm{Tr}\Big[\mathcal{T}_G^{(k)}[X]\Big] = \mathrm{Tr}[X]$. Repeating Eq. (B7) for all $P_\nu$ in $\mathcal{S}^{(k)}(G)$ leads to $D$ equations. Thus, we can find the vector of unknown coefficients $\boldsymbol{c}(X) = (c_1(X), \dots, c_D(X))$ by solving

$$
A \cdot \boldsymbol{c}(X) = \boldsymbol{b}(X)\,, \tag{B8}
$$

where $\boldsymbol{b}(X) = (\mathrm{Tr}[XP_1], \dots, \mathrm{Tr}[XP_D])$. Here, $A$ is a $D \times D$ symmetric matrix, known as the Gram matrix, whose entries are $(A)_{\nu\mu} = \mathrm{Tr}[P_\nu P_\mu]$. By inverting the $A$ matrix, we can then find

$$
\boldsymbol{c}(X) = A^{-1} \cdot \boldsymbol{b}(X)\,. \tag{B9}
$$




Supplemental Figure 2. **Elements of $S_k$.** a) We show the elements of $S_k$ for $k = 1, 2, 3$. Here we can see that the SWAP operator in Eq. (A3) is a transposition that permutes two subsystems. Below each element of $S_k$ we also show its cycle decomposition. b) We explicitly show that an element of $S_3$ commutes with $U^{\otimes 3}$ for any $U \in \mathbb{U}(d)$.

The matrix $A^{-1}$ is known as the Weingarten matrix. We refer the reader to Ref. [47] for additional details on the Weingarten matrix.

From the previous, we can see that computing the twirl $\mathcal{T}_G^{(k)}[X]$ in Eq. (B1) requires calculating the matrix $A$ and inverting it (if such inverse exists). In what follows, we will consider the cases when $G$ is the unitary or orthogonal group. For those cases we will present, in Supp. Info. C and D respectively, a simple explicit decomposition for the twirl into the $k$-th order commutant as in Eq. (B5), which holds asymptotically in the limit of large $d$.

## Supp. Info. C: The unitary group

In this section we present a series of results that will allow us to compute quantities of the form

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C_j(\rho_i)\right] = \mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} \mathrm{Tr}\big[U\rho_i U^\dagger O_j\big]\right]. \tag{C1}$$

### 1. Twirling over the unitary group

We begin by recalling that the standard representation of the unitary group of degree $d$, which we denote as $\mathbb{U}(d)$, is the group formed by all $d \times d$ unitary matrices acting on a $d$-dimensional Hilbert space $\mathcal{H}$. That is

$$\mathbb{U}(d) = \{U \in GL(\mathcal{H}) \quad | \quad UU^\dagger = U^\dagger U = \mathbb{1}\} \subset GL(\mathcal{H}), \tag{C2}$$

where $GL(\mathcal{H})$ is the general linear group, i.e., the group of all invertible matrices acting on $\mathcal{H}$. Here, we have used the standard notation $R(U) = U$ for all the elements of the unitary group.

From the Schur-Weyl duality, we know that a basis for the $k$-th order commutant of $\mathbb{U}(d)$ is the representation $P_d$ of the Symmetric group $S_k$ that permutes the $d$-dimensional subsystems in the $k$-fold tensor product Hilbert space, $\mathcal{H}^{\otimes k}$. That is, for a permutation $\sigma \in S_k$,

$$P_d(\sigma) = \sum_{i_1,\ldots,i_k=0}^{d-1} |i_{\sigma^{-1}(1)},\ldots,i_{\sigma^{-1}(k)}\rangle\langle i_1,\ldots,i_k|\,. \tag{C3}$$

Hence,

$$\mathcal{S}^{(k)}(\mathbb{U}(d)) = \{P_d(\sigma)\}_{\sigma\in S_k}, \tag{C4}$$

and we note that $\mathcal{S}^{(k)}(\mathbb{U}(d))$ contains $k!$ elements. In Sup. Fig. 2 we show the tensor representation of the elements in $S_k$ for $k = 1, 2, 3$, as well as an explicit illustrative example which showcases that the elements of $S_k$ commute with $U^{\otimes k}$.

To exemplify how one can use the previous result to compute twirls over the unitary group let us consider the $k=1$ and $k=2$ case. First, let $k=1$. Here, we can readily see in Fig. (2) that $S_1$ contains a single element, and its representation is given by $\{\mathbb{1}\}$. As such, the basis of the commutant contains one element

$$\mathcal{S}^{(1)}(\mathbb{U}(d)) = \{\mathbb{1}\}\,. \tag{C5}$$

The fact that the commutant is trivial also follows from the fact that the representation of $\mathbb{U}(d)$ with $k=1$ copies is irreducible. From here, we can build the $1\times 1$ Gram matrix $A=\begin{pmatrix} d \end{pmatrix}$, so that the Weingarten matrix is $A^{-1}=\begin{pmatrix} \frac{1}{d} \end{pmatrix}$, and $c_1(X)=\mathrm{Tr}[X]/d$. Hence,

$$\mathcal{T}^{(1)}_{\mathbb{U}(d)}[X] = \frac{\mathrm{Tr}[X]}{d}\mathbb{1}\,. \tag{C6}$$

Next, we consider the case of $k=2$. Now the basis of the commutant contains two elements (see Supp. Fig. (2))

$$\mathcal{S}^{(2)}(\mathbb{U}(d)) = \{\mathbb{1}\otimes\mathbb{1}, \mathrm{SWAP}\}\,. \tag{C7}$$

The ensuing Gram Matrix is

$$A = \begin{pmatrix} d^2 & d \\ d & d^2 \end{pmatrix}\,, \tag{C8}$$

and the Weingarten matrix

$$A^{-1} = \frac{1}{d^2-1}\begin{pmatrix} 1 & \frac{-1}{d} \\ \frac{-1}{d} & 1 \end{pmatrix}\,. \tag{C9}$$

Now we find

$$\begin{pmatrix} c_1(X) \\ c_2(X) \end{pmatrix} = \frac{1}{d^2-1}\begin{pmatrix} 1 & \frac{-1}{d} \\ \frac{-1}{d} & 1 \end{pmatrix}\cdot\begin{pmatrix} \mathrm{Tr}[X] \\ \mathrm{Tr}[X\mathrm{SWAP}] \end{pmatrix} = \frac{1}{d^2-1}\begin{pmatrix} \mathrm{Tr}[X]-\frac{\mathrm{Tr}[X\mathrm{SWAP}]}{d} \\ \mathrm{Tr}[X\mathrm{SWAP}]-\frac{\mathrm{Tr}[X]}{d} \end{pmatrix}\,. \tag{C10}$$

Hence,

$$\mathcal{T}^{(2)}_{\mathbb{U}(d)}[X] = \frac{1}{d^2-1}\left(\mathrm{Tr}[X]-\frac{\mathrm{Tr}[X\mathrm{SWAP}]}{d}\right)\mathbb{1}\otimes\mathbb{1} + \frac{1}{d^2-1}\left(\mathrm{Tr}[X\mathrm{SWAP}]-\frac{\mathrm{Tr}[X]}{d}\right)\mathrm{SWAP}\,. \tag{C11}$$

For more general $k$ the process of building the Gram matrix and inverting it can become quite cumbersome as the matrix $A$ will be a $k!\times k!$ dimensional matrix. However, since we are interested in the large $d$-limit, we can use the following result (presented in the main text as Lemma 3).

**Supplemental Theorem 1.** *Let $X$ be an operator in $\mathcal{B}(\mathcal{H}^{\otimes k})$, the twirl of $X$ over $\mathbb{U}(d)$, as defined in Eq. (B1) is*

$$\mathcal{T}^{(k)}_G[X] = \frac{1}{d^k}\sum_{\sigma\in S_k}\mathrm{Tr}[XP_d(\sigma)]P_d(\sigma^{-1}) + \frac{1}{d^k}\sum_{\sigma,\Pi\in S_k} c_{\sigma,\pi}\,\mathrm{Tr}[XP_d(\sigma)]P_d(\pi)\,, \tag{C12}$$

*where the constants $c_{\sigma,\Pi}$ are in $\mathcal{O}(1/d)$.*

In order to prove Supplemental Theorem 1, we find it convenient to recall the following definitions.

**Supplemental Definition 1** (Permutation Cycle)**.** *Let $\sigma$ be a permutation belonging to $S_k$. A permutation cycle $c$ is a set of indices $\{i_m, \sigma(i_m), \sigma(\sigma(i_m)), \dots\}$ that are closed under the action of $\sigma$.*

Equipped with Supplemental Definition 1, we can now introduce the cycle decomposition of a permutation.

**Supplemental Definition 2** (Cycle Decomposition)**.** *Given a permutation $\sigma \in S_k$, its cycle decomposition is an expression of $\sigma$ as a product of disjoint cycles*

$$\sigma = c_1 \cdots c_r \,. \tag{C13}$$

We will henceforth refer to those indexes which are not permuted, i.e., which are contained in length-one cycles, as fixed points. Moreover, we note that while it is usually standard to drop in the notation of Eq. (C13) the cycles of length one, (i.e., the cycles where an element is left unchanged), we assume that Eq. (C13) contains *all* cycles, including those of length one (see Sup. Fig. 2). We also remark that the cycle decomposition is unique, up to permutations of the cycles (since they are disjoint) and up to cyclic shifts within the cycles (since they are cycles). For example, we can express some permutation $\sigma \in S_5$ as $(145)(23)$ or as $(23)(451)$, that is, as the composition of a length-two cycle and a length-three cycle. Here we also recall that length-two cycles are also known as *transpositions*.

While the cycle notation is useful to identify each element in the symmetric group, we will be interested in counting how many cycles, and of what length, are contained in each $\sigma \in S_k$. This can be addressed by defining the cycle type.

**Supplemental Definition 3** (Cycle type)**.** *Given a permutation $\sigma \in S_k$, its cycle type $\nu(\sigma)$ is a vector of length $k$ whose entries indicate how many cycles of each length are present in the cycle decomposition of $\sigma$. That is,*

$$\nu(\sigma) = (\nu_1, \cdots, \nu_k)\,, \tag{C14}$$

*where $\nu_j$ denotes the number of length-$j$ cycles in $\sigma$.*

The cycle type of every element in $S_k$ is unique. In the previous example, $\nu((145)(23)) = (0,1,1,0,0)$. Similarly, expressing the identity element as $e = (1)(2)(3)(4)(5)$, it is easy to see that its cycle type is $\nu(e) = (5,0,0,0,0)$.

These definitions allow us to prove the following proposition.

**Supplemental Proposition 2.** *Let $S_k$ be the symmetric group, and let $P_d$ be the representation which permutes subsystems of the $k$-fold tensor product of a $d$-dimensional Hilbert space as in Eq. (C3). Then, the character of an element $\sigma \in S_k$ is*

$$\chi(\sigma) = \mathrm{Tr}[P_d(\sigma)] = d^{\|\nu(\sigma)\|_1} = d^r\,, \tag{C15}$$

*where $\nu(\sigma)$ is the cycle type of $\sigma$ as defined in Supplemental Definition 3, and $r$ is the number of cycles in the cycle decomposition of $\sigma$ as in Supplemental Definition 2.*

*Proof.* Let us begin by re-writing Eq. (C3) in term of the cycles decomposition of $\sigma$ as in Supplemental Definition 2. That is, given $\sigma = c_1 \cdots c_r$, we write

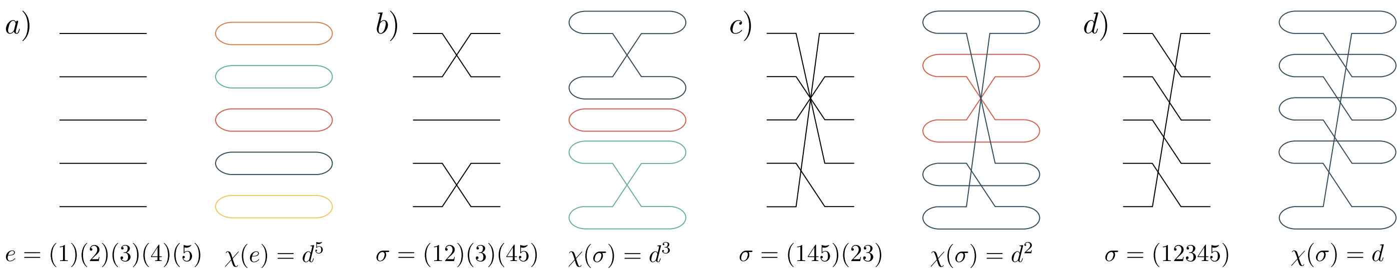
$$P_d(\sigma) = \bigotimes_{\alpha=1}^{r} \left( \sum_{i_1,\dots,i_{|c_\alpha|}=0}^{d-1} |i_{c_\alpha^{-1}(1)},\dots,i_{c_\alpha^{-1}(|c_\alpha|)}\rangle\langle i_1,\dots,i_{|c_\alpha|}| \right)\,,$$

where $|c_i|$ denotes the length of the $c_i$ cycle. Then, the character of $\sigma$ is

$$\chi(\sigma) = \mathrm{Tr}[P_d(\sigma)] = \prod_{\alpha=1}^{r} \left( \sum_{i_1\dots,i_{|c_\alpha|}=0}^{d-1} \mathrm{Tr}\Big[|i_{c_\alpha^{-1}(1)},\dots,i_{c_\alpha^{-1}(|c_\alpha|)}\rangle\langle i_1,\dots,i_{|c_\alpha|}|\Big] \right)\,. \tag{C16}$$

Here we can use the fact that

$$\mathrm{Tr}\Big[|i_{c_\alpha^{-1}(1)},\dots,i_{c_\alpha^{-1}(|c_\alpha|)}\rangle\langle i_1,\dots,i_{|c_\alpha|}|\Big] = \prod_{\beta=1}^{|c_\alpha|} \delta_{i_\beta, i_{c_\alpha^{-1}(\beta)}}\,, \tag{C17}$$




Supplemental Figure 3. **Computing the character of elements of $S_5$.** In all panels we present an elements of $S_5$ on the left, and then compute its trace on the right to obtain its character (see Eq. (A1)). Here we can verify that, as indicated by Proposition 2, the character is given by $d$ raised to the number of cycles in $\sigma$.

which means that for any cycle $c_\alpha$, independently of its length, one has

$$\sum_{i_1\dots,i_{|c_\alpha|}=0}^{d-1} \mathrm{Tr}\Big[|i_{c_\alpha^{-1}(1)},\dots,i_{c_\alpha^{-1}(|c_\alpha|)}\rangle\langle i_1,\dots,i_{|c_\alpha|}|\Big] = \sum_{i_1\dots,i_{|c_\alpha|}=0}^{d-1}\prod_{\beta=1}^{|c_\alpha|}\delta_{i_\beta,i_{c_\alpha^{-1}(\beta)}} = d\,. \tag{C18}$$

Replacing in Eq. (C16) leads to

$$\chi(\sigma) = \prod_{\alpha=1}^{r} d = d^r = d^{\|\nu(\sigma)\|_1}\,. \tag{C19}$$

□

In Sup. Fig. 3 we show an example where we compute the character for elements of $S_5$ and verify that they are indeed equal to $d^{\|\nu(\sigma)\|_1}$.

Supplemental Proposition 2 implies the following result.

**Supplemental Proposition 3.** *The character of any $\sigma \in S_k$ is uniquely maximized by the identity element $e \in S_k$, in which case it is equal to $\chi(e) = d^k$.*

*Proof.* First, let us note that the identity element $e$ is composed of $k$ 1-cycles. Thus, according to Supplemental Proposition 2, $\chi(e) = d^k$. Next let us note that by definition, all remaining elements in $S_k \setminus \{e\}$ must contain at least some cycle that is not a 1-cycle. This implies $\chi(\sigma) \leqslant d^{k-1}$ for any $\sigma \in S_k \setminus \{e\}$. In fact, it is not hard to see that the elements in $S_k$ whose character is exactly equal to $d^{k-1}$ are those composed of a single transposition. Consequently, the function $\chi : S_k \to \mathbb{C}$ has a unique maximum at $\sigma = e$. □

From here, we can show that the following result holds.

**Supplemental Proposition 4.** *Let $P_d$ be the subsystem-permuting representation of $S_k$ as defined in Eq. (C3). Given a pair of permutations $\sigma, \pi \in S_k$, then*

$$\mathrm{Tr}[P_d(\sigma)P_d(\pi)]\begin{cases} = d^k & \text{if } \pi = \sigma^{-1} \\ \leqslant d^{k-1} & \text{else}\end{cases}\,. \tag{C20}$$

*Proof.* Let us begin by noting that since $S_k$ forms a group, then for any $\sigma$ and $\pi$ in $S_k$, $\xi := \sigma\pi$ is also in $S_k$. Moreover, since $P_d$ is a representation, we have

$$\mathrm{Tr}[P_d(\sigma)P_d(\pi)] = \mathrm{Tr}[P_d(\sigma\pi)] = \mathrm{Tr}[P_d(\xi)] = \chi(\xi)\,. \tag{C21}$$

As such, we now need to ask the question, how large can the character $\chi(\xi)$ be? Supplemental Proposition 3 indicates that the character is maximal for the identity element. If $\xi = e$, this implies by the uniqueness of the inverse that $\pi = \sigma^{-1}$ (or equivalently $\sigma = \pi^{-1}$). In this case we find $\mathrm{Tr}[P_d(\sigma)P_d(\pi)] = \chi(e) = d^k$. However, if $\pi \neq \sigma^{-1}$, then $\xi \neq \mathbb{1}$ and via Supplemental Proposition 3, we have $\chi(\xi) \leqslant d^{k-1}$. □

Let us now go back to computing the twirl $\mathcal{T}_G^{(k)}[X]$. First, we find it convenient to reorder the basis $\mathcal{S}^{(k)}(\mathbb{U}(d))$ such that its first element are the representation of permutations which are their own inverse, i.e., $\sigma=\sigma^{-1}$ . Next, we order the rest of the elements $\sigma\neq\sigma^{-1}$ by placing $P_d(\sigma)$ next to $P_d(\sigma^{-1})$. We recall that the elements such that $\sigma=\sigma^{-1}$ are known as involutions and must consist of a product of disjoint transpositions plus fixed points. It is well known that the number of involutions is given by $I_k=\sum_{\eta=0}^{\lfloor\frac{k}{2}\rfloor}\binom{k}{2\eta}(2\eta-1)!!$.

Then, the following result holds.

**Supplemental Proposition 5.** *The $A$ matrix, of dimension $k!\times k!$, can be expressed as*

$$A=d^k(\widetilde{A}+\frac{1}{d}B)\,. \tag{C22}$$

*Here we defined*

$$\widetilde{A}=\mathbb{1}_{I_k}\bigoplus_{j=1}^{\frac{k!-I_k}{2}}\begin{pmatrix}0&1\\1&0\end{pmatrix}\,, \tag{C23}$$

*where here $\mathbb{1}_{I_k}$ denotes the $I_k\times I_k$ dimensional identity. Moreover, the matrix $B$ is such that its entries are $\mathcal{O}(1)$.*

More visually, the matrix $\widetilde{A}$ is of the form

$$\widetilde{A}=\begin{pmatrix}\underbrace{\boxed{\begin{matrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&\cdots&0&1\end{matrix}}}_{I_k\times I_k}&&&\Large 0\\&\begin{matrix}0&1\\1&0\end{matrix}&&\\&&\ddots&\\\Large 0&&&\begin{matrix}0&1\\1&0\end{matrix}\end{pmatrix}.$$

And we remark the fact that $\widetilde{A}$ its is own inverse. That is, $\widetilde{A}^{-1}=\widetilde{A}$.

*Proof.* Let us recall that the entries of the matrix $A$ are of the form $A_{\nu\mu}=\mathrm{Tr}[P_\nu P_\mu]$ where $P_\nu,P_\mu\in\mathcal{S}^{(k)}(\mathbb{U}(d))$. From Supplemental Proposition 4 we know that

$$A=\begin{pmatrix}d^k&a_{1,2}&\cdots&a_{1,I_k}&&&&&\\a_{2,1}&d^k&\cdots&a_{2,I_k}&&&&&\\\vdots&\vdots&\ddots&\vdots&&&&&\\a_{I_k,1}&\cdots&a_{I_k,I_k-1}&d^k&&&&&\\&&&&a_{I_k+1,I_k+1}&d^k&&&\\&&&&d^k&a_{I_k+2,I_k+2}&&&\\&&&&\cdots&&\ddots&&\\&&&&&&&a_{k!-1,k!-1}&d^k\\&&&&&&&d^k&a_{k!,k!}\end{pmatrix},$$

where the matrix elements $a_{ij} \leqslant d^{k-1}$. This allows us to express the matrix $A$ as

$$A = d^k(\widetilde{A} + \frac{1}{d}B)\,, \tag{C24}$$

where the entries in $B$ are at most equal to 1.

□

Here we present the following lemma proved in [48].

**Supplemental Lemma 1.** *Assume that the matrices $M$ and $\Omega + M$ are invertible. Then,*

$$(\Omega + M)^{-1} = M^{-1} - (\mathbb{1} + M^{-1}\Omega)M^{-1}\Omega M^{-1}\,. \tag{C25}$$

Using Supplemental Lemma 1, setting $M = \widetilde{A}$, $\Omega = \frac{1}{d}B$ and noting that $A \propto (\widetilde{A} + \frac{1}{d}B)$ always has inverse [47, 49] we find

$$A^{-1} = \frac{1}{d^k}\left(\widetilde{A} - \frac{1}{d}(\mathbb{1} + \frac{1}{d}\widetilde{A}B)\widetilde{A}B\widetilde{A}\right) = \frac{1}{d^k}\left(\widetilde{A} + C\right)\,, \tag{C26}$$

where we have defined

$$C = \frac{1}{d}(\mathbb{1} + \frac{1}{d}\widetilde{A}B)\widetilde{A}B\widetilde{A}\,. \tag{C27}$$

It is easy to verify that the matrix entries of $C$ are in $\mathcal{O}(1/d)$. Combining the previous result with Eqs. (B5) and (B9) leads to

$$\mathcal{T}_G^{(k)}[X] = \frac{1}{d^k}\sum_{\sigma\in S_k}\mathrm{Tr}[XP_d(\sigma)]P_d(\sigma^{-1}) + \frac{1}{d^k}\sum_{\sigma,\Pi\in S_k} c_{\sigma,\pi}\,\mathrm{Tr}[XP_d(\sigma)]P_d(\pi)\,, \tag{C28}$$

where the $c_{\sigma,\pi}$ are the matrix entries of $C$ as defined in Eq. (C27). This is then precisely the statement of Supplemental Theorem 1.

### 2. Computing expectation values of twirled operators

Let us consider an expectation value of the form

$$C(\rho_i) = \mathrm{Tr}\big[U\rho_i U^\dagger O\big]\,, \tag{C29}$$

where $\rho_i$ is a pure quantum state and $O$ be some traceless quantum operator such that $O^2 = \mathbb{1}$. Next, let us consider the task of estimating expectation values of the form

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k}\mathrm{Tr}\big[U\rho_i U^\dagger O\big]\right]\,. \tag{C30}$$

Here, we will show that in the large $d$ limit, the following theorem holds.

**Supplemental Theorem 2.** *Let $\rho_i$ for $i = 1, \dots, k$ be a set of pure quantum states such that $\mathrm{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $i, i'$, and let $O$ be some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then let us define $T_k \subseteq S_k$ the set of $k/2$ disjoint transpositions. That is, for any $\sigma \in T_k$, its cycle decomposition is $\sigma = c_1 \cdots c_{k/2}$ where each $c_\alpha$ is a transposition for all $\alpha = 1, \dots, k/2$. Then, in the large $d$ limit we have*

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{2}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\alpha=1}^{k/2}\mathrm{Tr}\big[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\big]\,. \tag{C31}$$

To prove this theorem, let us first re-write

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} \operatorname{Tr}\big[U\rho_i U^\dagger O\big]\right] = \mathbb{E}_{\mathbb{U}(d)}\left[\operatorname{Tr}\big[U^{\otimes k}\Lambda\,(U^\dagger)^{\otimes k} O^{\otimes k}\big]\right], \tag{C32}$$

where $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$. Explicitly,

$$\begin{aligned}\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i))\right] &= \int_{\mathbb{U}(d)} d\mu(U) \operatorname{Tr}\big[U^{\otimes k}\Lambda\,(U^\dagger)^{\otimes k} O^{\otimes k}\big]] \\ &= \operatorname{Tr}\left[\left(\int_{\mathbb{U}(d)} d\mu(U) U^{\otimes k}\Lambda\,(U^\dagger)^{\otimes k}\right) O^{\otimes k}\right].\end{aligned} \tag{C33}$$

Here we can use Supplemental Theorem 1 to find

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{1}{d^k}\sum_{\sigma\in S_k} \operatorname{Tr}[\Lambda P_d(\sigma)] \operatorname{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big] + \frac{1}{d^k}\sum_{\sigma,\Pi\in S_k} c_{\sigma,\pi} \operatorname{Tr}[\Lambda P_d(\sigma)] \operatorname{Tr}\big[P_d(\pi)O^{\otimes k}\big]. \tag{C34}$$

Now, let us prove the following proposition.

**Supplemental Proposition 6.** *Let $O$ be a traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then we have $\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big] = 0$ for any $\sigma \in S_k$ if $k$ is odd, and $\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big] = d^r$ if $k$ is even and $\sigma$ is a product of $r$ disjoint cycles of even length. The maximum of $\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big]$ is therefore achieved when $\sigma$ is a product of $k/2$ disjoint transpositions, leading to $\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big] = d^{k/2}$.*

*Proof.* We first consider the case of $k$ being odd. Let us express $\sigma$ in its cycle decomposition $\sigma = c_1 \cdots c_r$ as in Supplemental Definition 2. We have that

$$\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big] = \prod_{\alpha=1}^{r} \operatorname{Tr}\Big[P_d(c_\alpha)O^{\otimes |c_\alpha|}\Big]. \tag{C35}$$

Because $k$ is odd, we know that there must exist at least one cycle acting on an odd number of subsystems in the right hand side of Eq. (C35). Let us assume that this occurs for the cycle $c_{\alpha'}$. Then, we will have

$$\operatorname{Tr}\Big[P_d(c_\alpha)O^{\otimes |c_\alpha|}\Big] = \operatorname{Tr}\Big[O^{|c_\alpha|}\Big] = \operatorname{Tr}[O] = 0. \tag{C36}$$

Here we have used the fact that $O^2 = \mathbb{1}$, and hence, since $|c_\alpha|$ is odd, we have $O^{|c_\alpha|} = O$.

Next, let us consider the case of $k$ being even. We know from Eq. (C35) that if $\sigma$ contains any cycle acting on an odd number of subsystems, then $\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big]$ will be equal to zero. This means that only the permutations $\sigma$ composed entirely of cycles acting on even number of subsystems will have non-vanishing trace. If this is the case, we will have

$$\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big] = \prod_{\alpha=1}^{r} \operatorname{Tr}\Big[P_d(c_\alpha)O^{\otimes |c_\alpha|}\Big] = \prod_{\alpha=1}^{r} d = d^r. \tag{C37}$$

This follows from the fact that if $|c_\alpha|$ is even, then $O^{|c_\alpha|} = \mathbb{1}$. Moreover, Eq. (C37) will be maximized for the case when $r$ is largest, which corresponds to the case when $\sigma$ is a product of $k/2$ disjoint transpositions. For this special case one finds

$$\operatorname{Tr}\big[P_d(\sigma)O^{\otimes k}\big] = d^{k/2}. \tag{C38}$$

$\square$

Next, let us prove the following result.

**Supplemental Proposition 7.** *Let $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$ be a tensor product of $k$ pure states. Then $\left|\mathrm{Tr}[\Lambda P_d(\sigma)] + \mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right]\right| \leqslant 2$ for all $\sigma \in S_k$.*

*Proof.* Let us again decompose $\sigma$ in its cycle decomposition $\sigma = c_1 \cdots c_r$ as in Supplemental Definition 2. Then, we will have

$$\mathrm{Tr}[\Lambda P_d(\sigma)] = \prod_{\alpha=1}^{r} \mathrm{Tr}\left[\rho_{c_\alpha(1)} \cdots \rho_{c_\alpha(|c_\alpha|)}\right]. \tag{C39}$$

Here we find it convenient to define $\Lambda_1^\alpha = \rho_{c_\alpha(1)}$ and $\Lambda_2^\alpha = \rho_{c_\alpha(2)} \cdots \rho_{c_\alpha(|c_\alpha|)}$. We note that in general $\mathrm{Tr}[\Lambda_1^\alpha \Lambda_2^\alpha]$ is a complex number (as $\Lambda_2^\alpha$ is not necessarily Hermitian). However, it is not hard to see that

$$\mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right] = \prod_{\alpha=1}^{r} \mathrm{Tr}\left[\Lambda_1^\alpha (\Lambda_2^\alpha)^\dagger\right]. \tag{C40}$$

Hence,

$$\left|\mathrm{Tr}[\Lambda P_d(\sigma)] + \mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right]\right| = 2\left|\mathrm{Re}[\mathrm{Tr}[\Lambda P_d(\sigma)]]\right| \leqslant 2\left|\mathrm{Tr}[\Lambda P_d(\sigma)]\right|. \tag{C41}$$

Here we have used the fact that the conjugate of a product of complex numbers is the product of the complex conjugates, plus the fact that for any $z \in \mathbb{C}$, $|\mathrm{Re}[z]| \leqslant |z|$.

Let us now use the Matrix Holder inequality

$$|\mathrm{Tr}[\Lambda_1 \Lambda_2]| \leqslant \sqrt{\mathrm{Tr}[(\Lambda_1^\alpha)^2]}\sqrt{\mathrm{Tr}[(\Lambda_2^\alpha)(\Lambda_2^\alpha)^\dagger]}. \tag{C42}$$

We explicitly find

$$\mathrm{Tr}\left[(\Lambda_1^\alpha)^2\right] = \mathrm{Tr}\left[\rho_{c_\alpha(1)}^2\right] = 1, \tag{C43}$$

and

$$\mathrm{Tr}\left[(\Lambda_2^\alpha)(\Lambda_2^\alpha)^\dagger\right] = \mathrm{Tr}\left[\rho_{c_\alpha(2)} \cdots \rho_{c_\alpha(|c_\alpha|)}\rho_{c_\alpha(|c_\alpha|)} \cdots \rho_{c_\alpha(2)}\right] = \prod_{\eta=2}^{|c_\alpha|-1} \mathrm{Tr}\left[\rho_{c_\alpha(\eta)}\rho_{c_\alpha(\eta+1)}\right] \leqslant 1, \tag{C44}$$

where we have used the fact that the states $\rho_{i_\gamma}$ are pure. Combining the previous results leads to

$$\left|\mathrm{Tr}[\Lambda P_d(\sigma)] + \mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right]\right| \leqslant 2, \tag{C45}$$

for any $\sigma \in S_k$. □

With Supplemental Propositions 6 and 7 we can now state the following result.

**Supplemental Proposition 8.** *Let $O$ be a traceless Hermitian operator such that $O^2 = \mathbb{1}$. Let $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$ be a tensor product of $k$ pure states $\rho_{i_\gamma}$. Then, for all $\pi$ and $\sigma$ in $S_k$*

$$\frac{1}{d^k}\left|\left(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)] + c_{\sigma^{-1},\pi}\,\mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right]\right)\mathrm{Tr}\left[P_d(\pi) O^{\otimes k}\right]\right| \in \mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right). \tag{C46}$$

*Proof.* We begin by assuming, without loss of generality, that $|c_{\sigma^{-1},\pi}| \leqslant |c_{\sigma,\pi}|$. Thus, we have

$$\frac{1}{d^k}\left|\left(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)] + c_{\sigma^{-1},\pi}\,\mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right]\right)\mathrm{Tr}\left[P_d(\pi) O^{\otimes k}\right]\right| \leqslant \frac{|c_{\sigma,\pi}|}{d^k}\left|\left(\mathrm{Tr}[X P_d(\sigma)] + \mathrm{Tr}\left[\Lambda P_d(\sigma^{-1})\right]\right)\mathrm{Tr}\left[P_d(\pi) O^{\otimes k}\right]\right|.$$

Then, from Supplemental Propositions 6 and 7 we find

$$\frac{|c_{\sigma,\pi}|}{d^k}\left|(\mathrm{Tr}[\Lambda P_d(\sigma)]+\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big])\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]\right| \leqslant \frac{|c_{\sigma,\pi}|}{d^k}2d^{k/2}=\frac{2|c_{\sigma,\pi}|}{d^{k/2}}\,. \tag{C47}$$

Since by definition, $|c_{\sigma,\pi}|\in\mathcal{O}(1/d)$, we have

$$\left|\frac{1}{d^k}(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)]+c_{\sigma^{-1},\pi}\,\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big])\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]\right|\in\mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right)\,. \tag{C48}$$

□

Finally, consider the following proposition.

**Supplemental Proposition 9.** *Let $O$ be a traceless Hermitian operator such that $O^2=\mathbb{1}$. Let $\Lambda=\rho_1\otimes\cdots\otimes\rho_k$ be a tensor product of $k$ pure states $\rho_i$ such that $\mathrm{Tr}[\rho_i\rho_{i'}]\in\Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $i_\gamma,i_{\gamma'}$. Then,*

$$\frac{1}{d^k}\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]\in\Omega\left(\frac{1}{d^{k/2}}\right) \tag{C49}$$

*if $\sigma$ is a product of $k/2$ disjoint transpositions, and*

$$\frac{1}{d^k}\left|\mathrm{Tr}[|\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big]+\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big]\,\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]\right|\in\mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right) \tag{C50}$$

*for any other $\sigma$.*

*Proof.* We start by considering the case when $\sigma$ is a product of $k/2$ disjoint transpositions. We know from Supplemental Propositions 6 that $\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]=d^{k/2}$. Then, we have that

$$\mathrm{Tr}[RP_d(\sigma)]=\prod_{\alpha=1}^{k/2}\mathrm{Tr}\big[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\big]\in\Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)\,, \tag{C51}$$

where we have used the fact that $\mathrm{Tr}[\rho_i\rho_{i'}]\in\Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $\rho_i,\rho_{i'}$. Thus, we know that

$$\frac{1}{d^k}\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]\in\Omega\left(\frac{1}{d^{k/2}}\right)\,. \tag{C52}$$

Next, let us consider the case of $\sigma$ not being a product of $k/2$ disjoint transpositions. Here, we have from Supplemental Propositions 6

$$\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]=\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big]=d^r\,, \tag{C53}$$

with $r\leqslant\frac{k}{2}-1$. Then, the following chain of inequalities holds

$$\begin{aligned}\frac{1}{d^k}\left|\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big]+\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big]\,\mathrm{Tr}\big[P_d(\sigma)O^{\otimes k}\big]\right| &= \frac{1}{d^{\frac{k+2}{2}}}\left|\mathrm{Tr}[\Lambda P_d(\sigma)]+\mathrm{Tr}\big[\Lambda P_d(\sigma^{-1})\big]\right|\\ &\leqslant\frac{2}{d^{\frac{k+2}{2}}}\in\mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right)\,.\end{aligned} \tag{C54}$$

Where in the last line we have used Supplemental Proposition 7. □

To finish the proof of Supplemental Theorem 2 we simply combine Supplemental Propositions 8 and 9 and note that in the large $d$ limit we get

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{\gamma=1}^{k} C_\gamma\right] = \frac{1}{d^{k/2}} \sum_{\sigma\in T_k} \prod_{\alpha=1}^{k/2} \mathrm{Tr}\left[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\right], \tag{C55}$$

where we have defined as $T_k \subseteq S_k$ the set of permutations which are exactly given by a product of $k/2$ disjoint transpositions.

Here we can additionally prove the following corollary from Supplemental Theorem 2.

**Supplemental Corollary 1.** *Let $C(\rho_i) = \mathrm{Tr}\left[U\rho_i U^\dagger O\right]$, with $O$ be some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then,*

$$\mathbb{E}_{\mathbb{U}(d)}\left[C(\rho_i)^k\right] = \frac{k!}{d^{k/2}2^{k/2}(k/2)!}. \tag{C56}$$

*Proof.* The proof of Supplemental Corollary 1 simply follows from Supplemental Theorem 2 by noting that there are

$$\frac{1}{(k/2)!}\binom{k}{2,2,\ldots,2} = \frac{k!}{2^{k/2}(k/2)!}, \tag{C57}$$

elements in $T_k$. $\square$

### 3. Orthogonal states

Here we will prove the following theorem for the special case when the states in $\Lambda$ are all mutually orthogonal.

**Supplemental Theorem 3.** *Let $\rho_i$ for $i = 1,\ldots,k$ be a set of pure and mutually orthogonal quantum states, and let $O$ be some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then, in the large $d$ limit we have*

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{k!}{2^{k/2}(k/2)!}\frac{1}{d^k}. \tag{C58}$$

Going back to Eq. (C34), which we recall for convenience, we have

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{1}{d^k}\sum_{\sigma\in S_k} \mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\left[P_d(\sigma^{-1})O^{\otimes k}\right] + \frac{1}{d^k}\sum_{\sigma,\Pi\in S_k} c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\left[P_d(\pi)O^{\otimes k}\right], \tag{C59}$$

and we can see that all the terms in the first summation are zero. This follows from the fact that $\mathrm{Tr}[\Lambda P_d(\sigma)] = \prod_{\alpha=1}^{r}\mathrm{Tr}\left[\rho_{c_\alpha(1)}\cdots\rho_{c_\alpha(|c_\alpha|)}\right] = 0$ for all $\sigma \neq e$ (where we recall that $R(e) = \mathbb{1}^{\otimes k}$) as all the states are orthogonal. Moreover, for the case of $\sigma \neq e$ one has $\mathrm{Tr}\left[P_d(e^{-1})O^{\otimes k}\right] = \mathrm{Tr}\left[P_d(e)O^{\otimes k}\right] = \mathrm{Tr}[O]^k = 0$. Thus, one must here study the terms coming from the second summation.

Following a similar argument as the one previously given, we see that the only terms that survive in the second summation are those of the form $\mathrm{Tr}[\Lambda P_d(e)]\,\mathrm{Tr}\left[P_d(\pi)O^{\otimes k}\right] = \mathrm{Tr}\left[P_d(\pi)O^{\otimes k}\right]$. Next, let us prove the following result.

**Supplemental Proposition 10.** *The term $\frac{1}{d^k}c_{e,\pi}\,\mathrm{Tr}\left[P_d(\pi)O^{\otimes k}\right]$ is maximized when $\pi$ is a product of $k/2$ disjoint transpositions.*

*Proof.* Let us start by using known results for the asymptotics of the Weingarten functions [50]. We know that in the large $d$ limit $\frac{c_{e,\pi}}{d^k} = \frac{1}{d^{k+|\pi|}}$, where $|\pi|$ is the smallest number of transpositions that $\pi$ is a product of. We can easily compute $|\pi|$ by noting that if $\pi$ has $r$ cycles which we denote as $c_1,\dots,c_r$ following Supplemental Definition 2. Then, since each cycle can be decomposed as $|c_r|-1$ transpositions we have $|\pi| = \sum_{\alpha=1}^{r} |c_\alpha| - 1$. Note that if $\pi$ is a product of $k/2$ disjoint transpositions then $|\pi| = \sum_{\alpha=1}^{k/2} = \frac{k}{2}$, while if $\pi$ is a single $k$-cycle, then $|\pi| = \sum_{\alpha=1}^{1}(k-1) = k-1$. More generally, we have that $|\pi| = \sum_{\alpha=1}^{r} |c_\alpha| - 1 = k - r$. Combining this result with Supplemental Proposition 6, we can use the fact that since $\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]$ is maximized for $\pi$ is a product of $k/2$ disjoint transpositions (leading to $\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big] = d^{k/2}$). Assuming that all cycles are of even length, we have

$$\frac{1}{d^k} c_{e,\pi}\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big] = \begin{cases} \frac{d^{2r}}{d^{2k}}\,, & \text{if } r \leqslant \frac{k}{2}\,,\\ \frac{d^{2r-e_r}}{d^{2k}}\,, & \text{if } r \geqslant \frac{k}{2} \text{ and } r = \frac{k}{2} + e_r \end{cases} . \tag{C60}$$

Then the max of $\frac{1}{d^k} c_{e,\pi}\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]$ is achieved when $\pi$ is a product of $k/2$ disjoint transpositions. □

Using Supplemental Proposition 10, along with the fact that there are $\frac{(k)!}{2^{k/2}(k/2)!}$ product of $k/2$ disjoint transpositions leads to the proof of Supplemental Theorem 3,

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{k!}{2^{k/2}(k/2)!}\frac{1}{d^k}\,. \tag{C61}$$

Next, let us consider the case when $\Lambda$ contains $k_1$ of the same state $\rho_1$, $k_2$ of the same state $\rho_2$, and so on. In total, we assume that $\Lambda$ contains $q$ different states and that $\sum_{\beta=1}^{q} k_\beta = k$. Moreover, we denote as $K_1$ as the set of indexes $k_\beta$ equal to one, and $K_2$ as the set of indexes $k_\beta$ larger or equal than 2. That is

$$k_\beta \in \begin{cases} K_1\,, & \text{if } k_\beta = 1\,,\\ K_2\,, & \text{if } k_\beta \geqslant 2\,. \end{cases} \tag{C62}$$

We henceforth assume that there is at least one $k_\beta$ which is larger than 2, so that $K_2 \neq \emptyset$. Next, let us denote as $R = \sum_{k_\beta \in K_2} \lfloor \frac{k_\beta}{2} \rfloor$. Note that $1 \leqslant R \leqslant \frac{k}{2}$ where the upper bound is reached if each $k_\beta = 2$, and the lower bound when $K_2$ contains a single index. That is, when just a single state in $\Lambda$ is repeated a single time. Finally, let us define the subsets $T_{k_1},\dots,T_{k_q} \subseteq S_k$ of transpositions where $T_{k_\beta}$ pairs copies of the same state. Here we can prove that

**Supplemental Theorem 4.** *Let $\rho_i$ for $i=1,\dots,q$ be a set of pure and mutually orthogonal states. Then, let $\Lambda$ contains $k_1$ copies of $\rho_1$, $k_2$ copies of $\rho_2$, and so on. In total, we assume that $\Lambda$ contains $q$ different states and that $\sum_{\beta=1}^{q} k_\beta = k$. Moreover, we assume that $O$ is some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then, in the large $d$ limit we have*

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{d^{\sum_{k_\beta \in K_2} \lfloor \frac{k_\beta}{2} \rfloor}}{d^k}\left(\sum_{k_\beta \% 2 = 1} k_\beta \frac{(2\lfloor k_\beta/2 \rfloor)!}{2^{\lfloor k_\beta/2 \rfloor}(\lfloor [k_\beta/2 \rfloor)!} + \sum_{k_\beta \% 2 = 0} \frac{(k_\beta)!}{2^{k_\beta/2}([k_\beta/2)!}\right)\,. \tag{C63}$$

Note that if there exists a $k_\beta = k$, i.e., all the states are the same, then we recover the result in Supplemental Corollary 1.

*Proof.* Let us consider Eq. (C34), which we (again) recall here

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{1}{d^k}\sum_{\sigma \in S_k} \mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big] + \frac{1}{d^k}\sum_{\sigma,\Pi \in S_k} c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]\,, \tag{C64}$$

We already know that $\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big]$ is maximal when it is composed of $k/2$ disjoint transpositions. If $\left(\bigcup_{\beta=1}^{q} T_\beta\right) \cap T_k \neq \emptyset$ then there will be terms in the first summation which are non-zero. In this case, we will have that for large $d$

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{1}{d^k} \sum_{\sigma \in \left(\bigcup_{\beta=1}^{q} T_\beta\right)\cap T_k} \mathrm{Tr}[\Lambda P_d(\sigma)]\,\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big] = \frac{1}{d^{k/2}} \sum_{\sigma \in \left(\bigcup_{\beta=1}^{q} T_\beta\right)\cap T_k} = \frac{1}{d^{k/2}} \prod_{\beta=1}^{q} \frac{(k_\beta)!}{2^{k_\beta/2}(k_\beta/2)!}\,, \tag{C65}$$

where we have used the fact that $\mathrm{Tr}[\Lambda P_d(\sigma)]$ can be expressed as a product of terms of the form $\mathrm{Tr}[\rho_\beta \rho_\beta] = 0$, and where we have replaced $\mathrm{Tr}\big[P_d(\sigma^{-1})O^{\otimes k}\big] = d^{k/2}$. Moreover, since $\sum_{\sigma \in \left(\bigcup_{\beta=1}^{q} T_\beta\right)\cap T_k}$ is the number of ways in which we can pair all the states in $\Lambda$ such that a state is always paired to itself. Clearly, this requires that all $k_\beta$ are even, and therefore we can simple express $\sum_{\sigma \in \left(\bigcup_{\beta=1}^{q} T_\beta\right)\cap T_k} = \prod_{\beta=1}^{q} \frac{(k_\beta)!}{2^{k_\beta/2}(k_\beta/2)!}$.

However, if $\left(\bigcup_{\beta=1}^{q} T_\beta\right) \cap T_k = \emptyset$, or alternatively, if there is some $k_\beta$ which is odd (i.e., $K_1 \neq \emptyset$) then all the terms in the first summation will be zero, and we need to consider the second summation. Now, the terms in the second summation that will be non-zero are the ones where $\pi$ is composed of cycles of even length, and where $\sigma$ is composed of cycles which only connect a state with itself. Again, we can use known results for the asymptotics of the Weingarten functions [50] to have that in the large $d$ limit $\frac{c_{\sigma,\pi}}{d^k} = \frac{1}{d^{k+|\sigma\pi|}}$, where now $|\sigma\pi|$ is the smallest number of transpositions that the product of $\sigma$ and $\pi$ is a product of. Therefore, $\frac{1}{d^k} c_{e,\pi}\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big] = \frac{d^{k/2}}{d^{k+|\sigma\pi|}}$ will be the largest when $\sigma$ and $\pi$ are composed of the largest possible number of transpositions on the same set of indexes (as in this case $\sigma\pi$ will contain the most 1-cycles). In particular, we have that

$$\frac{1}{d^k} c_{e,\pi}\,\mathrm{Tr}\big[P_d(\pi)O^{\otimes k}\big] = \frac{d^{k/2+R}}{d^{3k/2}}\,. \tag{C66}$$

Hence, we have

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{d^{k/2+R}}{d^{3k/2}} M\,, \tag{C67}$$

where $M$ denotes the number of ways in which one can pair the same states with themselves. In particular, we can compute

$$M = \left(\prod_{k_\beta \% 2=1} k_\beta \frac{(2\lfloor k_\beta/2\rfloor)!}{2^{\lfloor k_\beta/2\rfloor}(\lfloor k_\beta/2\rfloor)!} \prod_{k_\beta \% 2=0} \frac{(k_\beta)!}{2^{k_\beta/2}(k_\beta/2)!}\right), \tag{C68}$$

where the first term arises from the cases when $k_\beta \geqslant 2$ and odd, and the second term when $k_\beta \geqslant 2$ and even. Thus, we find

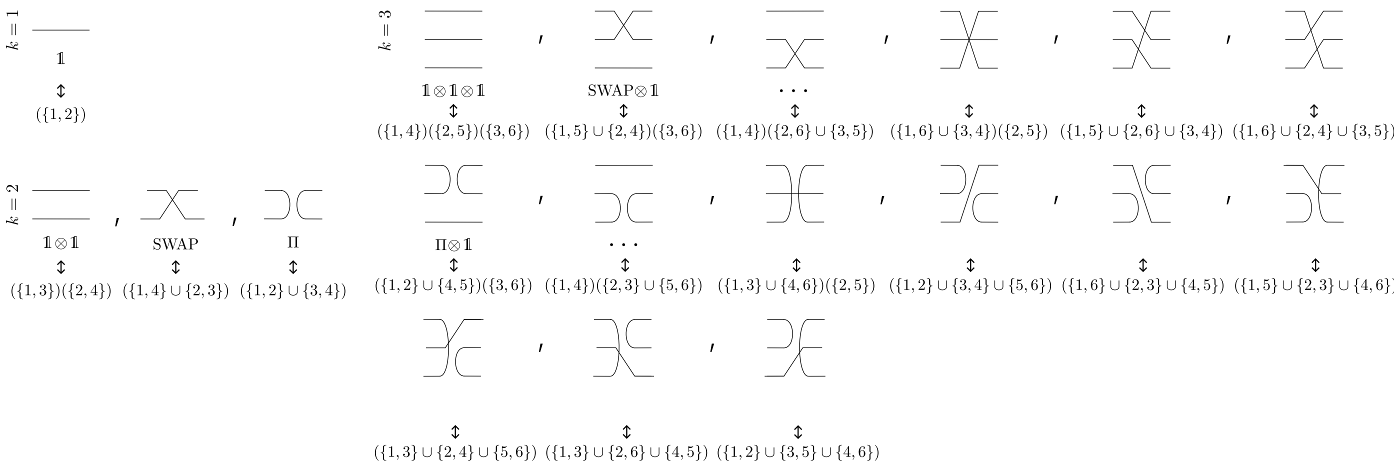
$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{d^{k/2+\sum_{k_\beta\in K_2}\lfloor \frac{k_\beta}{2}\rfloor}}{d^{3k/2}} \left(\prod_{k_\beta \% 2=1} k_\beta \frac{(2\lfloor k_\beta/2\rfloor)!}{2^{\lfloor k_\beta/2\rfloor}(\lfloor k_\beta/2\rfloor)!} \prod_{k_\beta \% 2=0} \frac{(k_\beta)!}{2^{k_\beta/2}(k_\beta/2)!}\right). \tag{C69}$$

□

**Supp. Info. D: The Orthogonal group**

In this section we will present a series of results that will allow us to compute quantities of the form

$$\mathbb{E}_{\mathbb{O}(d)}\left[\prod_{i=1}^{k} C_j(\rho_i)\right] = \mathbb{E}_{\mathbb{O}(d)}\left[\prod_{i=1}^{k} \mathrm{Tr}\big[U\rho_i U^\dagger O_j\big]\right]. \tag{D1}$$




Supplemental Figure 4. **Elements of $\mathfrak{B}_k$.** We show the elements of $\mathfrak{B}_k$ for $k = 1, 2, 3$. Here we can see that the SWAP and $\Pi$ in Eq. (A3) and Eq. (A5) are operators in $B_2$. Below each element of $\mathfrak{B}_k$ we also show its cycle decomposition.

### 1. Twirling over the orthogonal group

The standard representation of the orthogonal group of degree $d$, which we denote as $\mathbb{O}(d)$, is the group consisting of all $d \times d$ orthogonal matrices with real entries. That is,

$$\mathbb{O}(d) = \{U \in GL(\mathcal{H}) \quad | \quad UU^\dagger = U^\dagger U = UU^T = U^T U = \mathbb{1}\} \subset GL(\mathcal{H})\,, \tag{D2}$$

where $U^T$ denotes the transpose of $U$, and the entries of $U$ are real, i.e., $U_{ij} \in \mathbb{R}$. We note that we have employed again the standard notation $R(U) = U$ for all the elements of the orthogonal group. For this group, a basis for the $k$-th order commutant is given by a representation $F_d$ of the Brauer algebra $\mathfrak{B}_k(d)$ acting on the $k$-fold tensor product Hilbert space, $\mathcal{H}^{\otimes k}$. That is,

$$\mathcal{S}^{(k)}(\mathbb{O}(d)) = \{F_d(\sigma)\}_{\sigma \in \mathfrak{B}_k(d)}. \tag{D3}$$

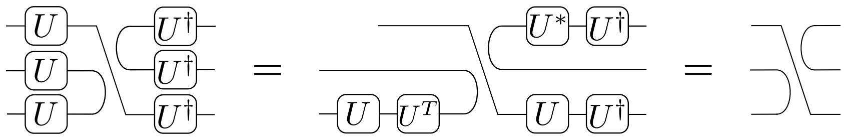
Here we recall that the Brauer algebra is composed of all possible pairings on a set of $2k$ items. That is, given a set of $2k$ items, the elements of $\mathfrak{B}_k(d)$ correspond to all possible ways of splitting them in pairs. Hence, the basis of the commutant, $\mathcal{S}^{(k)}(\mathbb{O}(d))$, contains $\frac{(2k)!}{2^k k!}$ elements. For the sake of illustration, the tensor representation of the Brauer algebra for $k = 1, 2, 3$ is depicted in Sup. Fig. 4, and in Sup. Fig. 5 we explicitly show that an element of $\mathfrak{B}_3$ commutes with $U^{\otimes 3}$ for any $U$ in $\mathbb{O}(d)$.

An element $\sigma \in \mathfrak{B}_k(d)$ can be completely specified by $k$ disjoint pairs, as $\sigma = \{\lambda_1, \sigma(\lambda_1)\} \cup \cdots \cup \{\lambda_k, \sigma(\lambda_k)\}$. Moreover, we find it convenient to also define for any $\sigma$ its *transpose* as $\sigma^T = \{\lambda_1 + k, \sigma(\lambda_1) + k\} \cup \cdots \cup \{\lambda_k + k), \sigma(\lambda_k) + k\}$ where the sum is taken mod$(k)$. Note that $\forall \sigma \in \mathfrak{B}_k(d)$, then $\sigma^T \in \mathfrak{B}_k(d)$. Let us now consider an explicit example and write down an element of $\mathfrak{B}_7(d)$. For instance, consider $\sigma = (\{1, 2\} \cup \{8, 9\})(\{3, 5\} \cup \{4, 10\} \cup \{11, 12\})(\{6, 14\} \cup \{7, 13\})$, where the parenthesis correspond to cycles (as defined below in Supplemental Definition 4). We present this element in Sup. Fig. 6(a). In bra-ket notation, this is equivalent to

$$F_d(\sigma) = \sum_{i_1, i_2 = 0}^{d-1} |i_1, i_1\rangle \langle i_2, i_2| \otimes \sum_{i_3, i_4, i_5 = 0}^{d-1} |i_3, i_4, i_3\rangle \langle i_4, i_5, i_5| \otimes \sum_{i_6, i_7 = 0}^{d-1} |i_6, i_7\rangle \langle i_7, i_6|\,. \tag{D4}$$

where $F_d(\sigma)$ is a representation of the Brauer algebra element $\sigma$. Here, we find $\sigma^T = (\{1, 2\} \cup \{8, 9\})(\{3, 11\} \cup \{4, 5\} \cup \{10, 12\})(\{6, 14\} \cup \{7, 13\})$ (see Sup. Fig. 6(b)) and

$$F_d(\sigma^T) = F_d(\sigma)^T\,. \tag{D5}$$



Supplemental Figure 5. **Commutation relation.** We explicitly show that an element of $B_3$ commutes with $U^{\otimes 3}$ for any $U\in\mathbb{O}(d)$. Here we use the ricochet property of Eq. (A7) plus the fact that $UU^\dagger = UU^T = U^*U^\dagger = \mathbb{1}$.

More generally, given an element $\sigma\in\mathfrak{B}_k(d)$, we can express it as

$$F_d(\sigma) = \sum_{i_1,\ldots,i_{2k}=0}^{d-1} |i_{k+1}, i_{k+2},\ldots,i_{2k}\rangle\langle i_1,i_2,\ldots,i_k| \prod_{\gamma=1}^{k} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}}\,. \tag{D6}$$

It is important to stress that, in contrast to the $k$-th order commutant of the unitary group, $\mathcal{S}^{(k)}(\mathbb{O}(d))$ is not a group itself but a $\mathbb{Z}(d)$-algebra. This implies that when we multiply two elements in $\mathfrak{B}_k(d)$, we don't necessarily obtain an element of $\mathfrak{B}_k(d)$ but rather an element of $\mathfrak{B}_k(d)$ times an integer power of $d$. Diagrammatically, this means that when we connect (multiply) two diagrams, closed loops can appear. Then, the power to which the factor $d$ is raised is equal to the number of closed loops. We illustrate this fact in Sup. Fig. 6(c). Furthermore, not every element in $\mathfrak{B}_k(d)$ has an inverse. We remark that the symmetric group $S_k$ is a subalgebra of the Brauer algebra. We will denote as $B_k = \mathfrak{B}_k(d)\backslash S_k$ the elements in $\mathfrak{B}_k(d)$ that do not belong to $S_k$, and recall that the elements of $B_k$ do not have an inverse.

First, let us consider the case of $k=1$. As shown in Sup. Fig. 4, $\mathfrak{B}_1(d)$ contains a single element whose representation is given by $\{\mathbb{1}\}$. As such, we recover the same result as for the unitary group, where the Gram matrix is $A=\begin{pmatrix} d\end{pmatrix}$, and thus

$$\mathcal{T}^{(1)}_{\mathbb{O}(d)}[X] = \frac{\mathrm{Tr}[X]}{d}\mathbb{1}\,. \tag{D7}$$

Next, we consider the case of $k=2$. Now $\mathfrak{B}_2(d)$ contains three elements (see Sup. Fig. 4) whose representation is given by

$$\mathcal{S}^{(2)}(\mathbb{O}(d)) = \{\mathbb{1}\otimes\mathbb{1}, \mathrm{SWAP}, \Pi\}\,, \tag{D8}$$

where $\Pi$ was defined in Eq. (A5). The ensuing Gram Matrix is

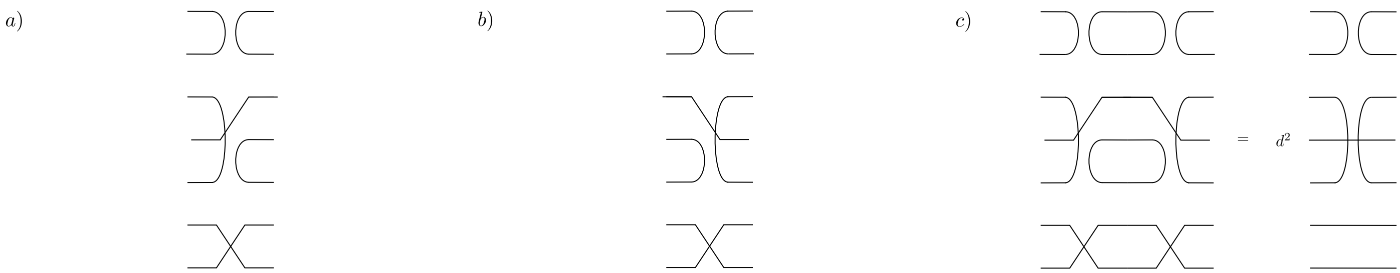
$$A = \begin{pmatrix} d^2 & d & d\\ d & d^2 & d\\ d & d & d^2\end{pmatrix}\,, \tag{D9}$$

and the Weingarten matrix

$$A^{-1} = \frac{1}{d(d+2)(d-1)}\begin{pmatrix} d+1 & -1 & -1\\ -1 & d+1 & -1\\ -1 & -1 & d+1\end{pmatrix}\,. \tag{D10}$$

Thus, we find

$$\begin{pmatrix} c_1(X)\\ c_2(X)\\ c_3(X)\end{pmatrix} = \frac{1}{d(d+2)(d-1)}\begin{pmatrix} d+1 & -1 & -1\\ -1 & d+1 & -1\\ -1 & -1 & d+1\end{pmatrix}\cdot\begin{pmatrix} \mathrm{Tr}[X]\\ \mathrm{Tr}[X\mathrm{SWAP}]\\ \mathrm{Tr}[X\Pi]\end{pmatrix}\,. \tag{D11}$$




$\sigma = (\{1,2\} \cup \{8,9\})(\{3,5\} \cup \{4,10\} \cup \{11,12\})(\{6,14\} \cup \{7,13\})$ $\sigma^T = (\{1,2\} \cup \{8,9\})(\{3,11\} \cup \{4,5\} \cup \{10,12\})(\{6,14\} \cup \{7,13\})$ $\sigma\sigma^T$

Supplemental Figure 6. **Element and its transpose.** In panels a) and b) we respectively present two elements of $\mathfrak{B}_7$, $\sigma$, and its transpose $\sigma^T$. Then, in panel c) we present their composition $\sigma\sigma^T$. Here we can verify that $\mathfrak{B}_k$ is a $\mathbb{Z}(d)$-algebra, as the multiplications of two elements of $\mathfrak{B}_k(d)$ is an element of $\mathfrak{B}_k(d)$ times an integer power of $d$.

Hence,

$$\begin{aligned}\mathcal{T}^{(2)}_{\mathbb{O}(d)}[X] =& \frac{1}{d(d+2)(d-1)}\left((d+1)\operatorname{Tr}[X] - \operatorname{Tr}[X\mathrm{SWAP}] - \operatorname{Tr}[X\Pi]\right)\mathbb{1}\otimes\mathbb{1} \\ &+ \frac{1}{d(d+2)(d-1)}\left(-\operatorname{Tr}[X] + (d+1)\operatorname{Tr}[X\mathrm{SWAP}] - \operatorname{Tr}[X\Pi]\right)\mathrm{SWAP} \\ &+ \frac{1}{d(d+2)(d-1)}\left(-\operatorname{Tr}[X] - \operatorname{Tr}[X\mathrm{SWAP}] + (d+1)\operatorname{Tr}[X\Pi]\right)\Pi\,. \end{aligned} \tag{D12}$$

Similarly to the unitary case, building the Gram matrix for large $k$ can be quite cumbersome. However, since we are interested in the large $d$-limit, we can use the following result.

**Supplemental Theorem 5.** *Let $X$ be an operator in $\mathcal{B}(\mathcal{H}^{\otimes k})$, then for large Hilbert space dimension $d$, the twirl of $X$ over $\mathbb{O}(d)$, as defined in Eq. (B1) is*

$$\mathcal{T}^{(k)}_G[X] = \frac{1}{d^k}\sum_{\sigma\in\mathfrak{B}_k}\operatorname{Tr}[XP_d(\sigma)]P_d(\sigma^T) + \frac{1}{d^k}\sum_{\sigma,\Pi\in\mathfrak{B}_k} c_{\sigma,\pi}\operatorname{Tr}[XP_d(\sigma)]P_d(\pi)\,, \tag{D13}$$

*where the constants $c_{\sigma,\Pi}$ are in $\mathcal{O}(1/d)$.*

In order to prove Supplemental Theorem 5, we recall the following definitions.

**Supplemental Definition 4** (Cycle)**.** *Let $\sigma$ be an element belonging to $\mathfrak{B}_k(d)$. A cycle $c$ is a set of indices $\{i_m, \overline{\sigma(i_m)}, \sigma(\overline{\sigma(i_m)}), \overline{\sigma(\overline{\sigma(i_m)})}, \sigma(\overline{\sigma(\overline{\sigma(i_m)})}), \ldots\}$ that are closed under the action of $\sigma$. Here, we use the notation $\overline{i_m}$ to denote the opposite of $i_m$, i.e., $\overline{i_m} = i_m + k$ if $i_m \leqslant k$ and $\overline{i_m} = i_m - k$ if $i_m > k$. Moreover, we will refer to the number of indices in the cycle divided by two as the length of the cycle.*

**Supplemental Definition 5** (Cycle Decomposition)**.** *Given an element $\sigma$ belonging to $\mathfrak{B}_k(d)$, its cycle decomposition is an expression of $\sigma$ as a product of disjoint cycles*

$$\sigma = c_1 \cdots c_r\,. \tag{D14}$$

We will refer to those indexes contained in length-one cycles (such that $\sigma(i_m) = \overline{i_m}$) as fixed points. Moreover, as we did for the unitary group, we assume that Eq. (D14) contains *all* cycles, including those of length one (see Sup. Fig. 2). We remark that Supplemental Definitions 4 and 5 generalize and include as particular cases Supplemental Definitions 1 and 2 respectively. As in the case of the elements of $S_k$, the cycle decomposition is unique, up to permutations of the cycles (since they are disjoint) and up to cyclic shifts within the cycles (since they are cycles).

We will be interested in counting how many cycles, and of what length, are contained in each $\sigma \in \mathfrak{B}_k(d)$. To that end, we introduce the definition of the cycle type.

**Supplemental Definition 6** (Cycle type)**.** *Given a $\sigma \in \mathfrak{B}_k(d)$, its cycle type $\nu(\sigma)$ is a vector of length $k$ whose entries indicate how many cycles of each length are present in the cycle decomposition of $\sigma$. That is,*

$$\nu(\sigma) = (\nu_1, \cdots, \nu_k)\,, \tag{D15}$$

*where $\nu_j$ denotes the number of length-$j$ cycles in $\sigma$.*

We now introduce the following propositions,

**Supplemental Proposition 11.** *The character of an element $\sigma \in \mathfrak{B}_k(d)$ is*

$$\chi(\sigma) = \mathrm{Tr}[\sigma] = d^{\|\nu(\sigma)\|_1} = d^r\,, \tag{D16}$$

*where $\nu(\sigma)$ is the cycle type of $\sigma$ as defined in Supplemental Definition 6, and $r$ is the number of cycles in the cycle decomposition of $\sigma$ as in Supplemental Definition 5.*

*Proof.* Let us begin by re-writing Eq. (D6) in term of the cycles decomposition of $\sigma$ as in Supplemental Definition 5. That is, given $\sigma = c_1 \cdots c_r$, we write

$$F_d(\sigma) = \bigotimes_{\alpha=1}^{r} \left( \sum_{i_{\lambda_1^\alpha},\dots,i_{\lambda_{2|c_\alpha|}^\alpha}=0}^{d-1} |i_{\lambda^\alpha_{|c_\alpha|+1}}, \dots, i_{\lambda^\alpha_{2|c_\alpha|}}\rangle\langle i_{\lambda^\alpha_1}, \dots, i_{\lambda^\alpha_{|c_\alpha|}}| \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\alpha_\gamma}, i_{\sigma(\alpha_\gamma)}} \right)\,, \tag{D17}$$

where $|c_\alpha|$ denotes the length of the $c_\alpha$ cycle and we used the notation $c_\alpha = \bigcup_{\gamma=1}^{|c_\alpha|}\{\alpha_\gamma, \sigma(\alpha_\gamma)\}$. Then, the character of $\sigma$ is

$$\chi(\sigma) = \mathrm{Tr}[F_d(\sigma)] = \prod_{\alpha=1}^{r} \left( \sum_{i_{\lambda_1^\alpha},\dots,i_{\lambda_{2|c_\alpha|}^\alpha}=0}^{d-1} \mathrm{Tr}\left[ |i_{\lambda^\alpha_{|c_\alpha|+1}}, \dots, i_{\lambda^\alpha_{2|c_\alpha|}}\rangle\langle i_{\lambda^\alpha_1}, \dots, i_{\lambda^\alpha_{|c_\alpha|}}| \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\alpha_\gamma}, i_{\sigma(\alpha_\gamma)}} \right] \right)\,. \tag{D18}$$

We now compute

$$\mathrm{Tr}\left[ |i_{\lambda^\alpha_{|c_\alpha|+1}}, \dots, i_{\lambda^\alpha_{2|c_\alpha|}}\rangle\langle i_{\lambda^\alpha_1}, \dots, i_{\lambda^\alpha_{|c_\alpha|}}| \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\alpha_\gamma}, i_{\sigma(\alpha_\gamma)}} \right] = \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\alpha_\gamma}, i_{\sigma(\alpha_\gamma)}} \prod_{\beta=1}^{|c_\alpha|} \delta_{i_{\lambda_\beta}, i_{\lambda_{|c_\alpha|+\beta}}}\,, \tag{D19}$$

which implies that for any cycle $c_\alpha$, independently of its length, we have

$$\sum_{i_{\lambda_1^\alpha},\dots,i_{\lambda_{2|c_\alpha|}^\alpha}=0}^{d-1} \mathrm{Tr}\left[ |i_{\lambda^\alpha_{|c_\alpha|+1}}, \dots, i_{\lambda^\alpha_{2|c_\alpha|}}\rangle\langle i_{\lambda^\alpha_1}, \dots, i_{\lambda^\alpha_{|c_\alpha|}}| \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\alpha_\gamma}, i_{\sigma(\alpha_\gamma)}} \right] = \sum_{i_{\lambda_1^\alpha},\dots,i_{\lambda_{2|c_\alpha|}^\alpha}=0}^{d-1} \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\alpha_\gamma}, i_{\sigma(\alpha_\gamma)}} \prod_{\beta=1}^{|c_\alpha|} \delta_{i_{\lambda_\beta}, i_{\lambda_{|c_\alpha|+\beta}}} = d\,, \tag{D20}$$

where we used the fact that a cycle either has no fixed points or is itself a fixed point. Replacing in Eq. (D18), we obtain

$$\chi(\sigma) = \prod_{\alpha=1}^{r} d = d^r = d^{\|\nu(\sigma)\|_1}\,. \tag{D21}$$

$\square$

**Supplemental Proposition 12.** *Let $\sigma$ be an element of $\mathfrak{B}_k(d)$. The number of cycles $||\nu(\sigma)||_1$ is at most $k-l$, where $l$ is the number of indices pairs $\left(i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}\right)$ such that $\lambda_\gamma, \sigma(\lambda_\gamma) \leqslant k$. Moreover, this maximum is uniquely achieved when for every pair $\lambda_\gamma, \sigma(\lambda_\gamma) \leqslant k$ there exists a pair $\lambda_{\gamma'}, \sigma(\lambda_{\gamma'}) > k$ such that $\lambda_\gamma = \overline{\lambda_{\gamma'}}$ and $\sigma(\lambda_\gamma) = \overline{\sigma(\lambda_{\gamma'})}$, or $\sigma(\lambda_\gamma) = \overline{\lambda_{\gamma'}}$ and $\gamma = \overline{\sigma(\lambda_{\gamma'})}$, and the rest of indices are fixed points.*

*Proof.* Using Supplemental Definition 4, we first have that every fixed point is a cycle. Then, if for a pair $\lambda_\gamma, \sigma(\lambda_\gamma) \leqslant k$ there exists a pair $\lambda'_\gamma, \sigma(\lambda'_\gamma) > k$ such that $\lambda_\gamma = \overline{\lambda_{\gamma'}}$ and $\sigma(\lambda_\gamma) = \overline{\sigma(\lambda_{\gamma'})}$, or $\sigma(\lambda_\gamma) = \overline{\lambda_{\gamma'}}$ and $\lambda_\gamma = \overline{\sigma(\lambda_{\gamma'})}$, then those two pairs form a cycle, as the sequences

$$\lambda_\gamma,\ \sigma(\lambda_\gamma) = \overline{\sigma(\lambda_{\gamma'})},\ \sigma(\lambda_{\gamma'}),\ \sigma(\sigma(\lambda_{\gamma'})) = \lambda_{\gamma'},\ \lambda_\gamma\,, \tag{D22}$$

or

$$\lambda_\gamma,\ \sigma(\lambda_\gamma) = \overline{\lambda_{\gamma'}},\ \lambda_{\gamma'},\ \sigma(\lambda_{\gamma'}) = \overline{\lambda_\gamma},\ \lambda_\gamma\,, \tag{D23}$$

are closed under $\sigma$. Therefore, there are $l$ such cycles plus $k-2l$ fixed points, which add up to $k-l$ cycles.

Then, we note that given pair of indices such that $\lambda_\gamma, \sigma(\lambda_\gamma) \leqslant k$, a cycle containing them must consist of at least four indices. This is so because $\overline{\sigma(\lambda_\gamma)} \neq \lambda_\gamma, \sigma(\lambda_\gamma)$. By direct inspection, it is clear that the only possible sequences that lead to cycles of four indices are (D22) or (D23). Therefore, if the conditions stated above are not satisfied, $||\nu(\sigma)||_1 < k-l$. $\square$

**Supplemental Proposition 13.** *Let $\sigma$ be an element of $B_k$ and $\pi$ an element of $\mathfrak{B}_k(d)$. Then, it holds that*

$$\mathrm{Tr}[\sigma\pi] \begin{cases} = d^k & \textit{if } \pi = \sigma^T \\ \leqslant d^{k-1} & \textit{else} \end{cases}\,, \tag{D24}$$

*Proof.* Let us start with $\pi = \sigma^T$. We have

$$\begin{aligned}
\sigma\pi &= \sum_{\substack{i_1,\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{2k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \left\langle i_1, i_2, \ldots, i_k \middle| i'_{k+1}, i'_{k+2}, \ldots, i'_{2k}\right\rangle \prod_{\gamma=1}^{k} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\beta=1}^{k} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \\
&= \sum_{\substack{i_1,\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \left\langle i_1, i_2, \ldots, i_k \middle| i_1, i_2, \ldots, i_k\right\rangle \prod_{\gamma=1}^{k} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\beta=1}^{k} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \\
&= \sum_{\substack{i_{k+1},\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \sum_{\substack{i_{\lambda_{\gamma_1}},\ldots,i_{\lambda_{\gamma_{2l}}}=0\\ \lambda_\gamma,\sigma(\lambda_\gamma)\leqslant k}}^{d-1} \left\langle i_{\lambda_{\gamma_1}}, \ldots, i_{\lambda_{\gamma_{2l}}} \middle| i_{\lambda_{\gamma_1}}, \ldots, i_{\lambda_{\gamma_{2l}}} \right\rangle \prod_{\substack{\gamma\\ \lambda_\gamma,\sigma(\lambda_\gamma)\leqslant k}} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \\
&\quad \otimes \sum_{\substack{i_{\lambda_{\gamma_{2l+1}}},\ldots,i_{\lambda_{\gamma_{k-2l}}}=0\\ \sigma(\lambda_\gamma)>k}}^{d-1} \left\langle i_{\lambda_{2l+1}}, \ldots i_{\lambda_{k-2l}} \middle| i_{\lambda_{2l+1}}, \ldots i_{\lambda_{k-2l}} \right\rangle \prod_{\substack{\gamma\\ \sigma(\lambda_\gamma)>k}} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\substack{\beta\\ \lambda_\beta\leqslant k}} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \\
&= d^l \sum_{\substack{i_{k+1},\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \prod_{\substack{\gamma\\ \lambda_\gamma,\sigma(\lambda_\gamma)>k}} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\substack{\beta\\ \lambda_\beta,\pi(\lambda_\beta)\leqslant k}} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \prod_{\substack{\omega\\ \lambda_\omega\leqslant k\\ \pi(\omega)>k}} \delta_{i_{\lambda_{\omega+k}}, i'_{\lambda_\omega}} \\
&= d^l \tau\,,
\end{aligned} \tag{D25}$$

where $\tau \in B_k$ and $l$ is the number of indices pairs such that $\lambda_\gamma, \sigma(\lambda_\gamma) \leqslant k$. That is, $l$ is the number of closed loops that appear when multiplying $\pi$ and $\sigma$. Therefore, using Supplemental Propositions 11 and 12 ,

$$\mathrm{Tr}[\sigma\pi] = d^l d^{||\nu(\sigma)||_1} = d^l d^{k-l} = d^k\,. \tag{D26}$$

Now, let us assume that $\pi \neq \sigma^T$. Let us further define $C_k(\sigma,\pi)$ as the set of indices $\{\gamma,\gamma'\}$ such that $i_\gamma, \sigma(\gamma) \leqslant k$, $i'_{\gamma'}, \pi(\gamma') > k$, and such that for every $\gamma \in C_k(\sigma,\pi)$ there exists $\gamma' \in C_k(\sigma,\pi)$ leading to $\gamma = \gamma'$ or $\gamma = \pi(\gamma')$ and $\sigma(\gamma) = \gamma'$ or $\sigma(\gamma) = \pi(\gamma')$. Accordingly, we introduce

$$\tau = \sum_{i_\gamma, i'_{\gamma'}=0}^{d-1} \bigotimes_{\gamma,\gamma' \in C_k(\sigma,\pi)} |i_\gamma\rangle \langle i'_{\gamma'}| \prod_{\gamma \in C_k(\sigma,\pi)} \delta_{i_\gamma, i_{\sigma(\gamma)}} \prod_{\gamma' \in C_k(\sigma,\pi)} \delta_{i'_{\gamma'}, i'_{\pi(\gamma')}}\,, \tag{D27}$$

and write its cycle decomposition as $\tau = c_1 \cdots c_r$ (since $\tau \in \mathfrak{B}_s$ where $s = \sum_\alpha |c_\alpha|$). We then find

$$\begin{aligned}
\sigma\pi &= \sum_{\substack{i_1,\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{2k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \left\langle i_1, i_2, \ldots, i_k \middle| i'_{k+1}, i'_{k+2}, \ldots, i'_{2k}\right\rangle \prod_{\gamma=1}^{k} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\beta=1}^{k} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \\
&= \sum_{\substack{i_1,\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \left\langle i_1, i_2, \ldots, i_k \middle| i_1, i_2, \ldots, i_k\right\rangle \prod_{\gamma=1}^{k} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\beta=1}^{k} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \\
&= \sum_{\substack{i_{k+1},\ldots,i_{2k}=0\\ i'_1,\ldots,i'_{k}=0}}^{d-1} |i_{k+1}, i_{k+2}, \ldots, i_{2k}\rangle \langle i'_1, i'_2, \ldots, i'_k| \bigotimes_{\alpha=1}^{r} \sum_{i_{\lambda_1^\alpha},\ldots,i_{\lambda^\alpha_{|c_\alpha|}}=0}^{d-1} \left\langle i_{\lambda_1^\alpha}, \ldots, i_{\lambda^\alpha_{|c_\alpha|}} \middle| i_{\lambda_1^\alpha}, \ldots, i_{\lambda^\alpha_{|c_\alpha|}} \right\rangle \prod_{\gamma=1}^{|c_\alpha|} \delta_{i_{\lambda^\alpha_\gamma}, i_{\sigma\left(\lambda^\alpha_\gamma\right)}} \\
&\quad \otimes \sum_{\gamma,\gamma'=0}^{d-1} \bigotimes_{\gamma,\gamma' \notin C_k} \langle i_\gamma | i'_{\gamma'}\rangle \prod_{\gamma \notin C_k(\sigma,\pi)} \delta_{i_\gamma, i_{\sigma(\gamma)}} \prod_{\gamma' \notin C_k(\sigma,\pi)} \delta_{i'_{\gamma'}, i'_{\pi(\gamma')}} \prod_{\substack{\gamma\\ \lambda_\gamma, \sigma(\lambda_\gamma) > k}} \delta_{i_{\lambda_\gamma}, i_{\sigma(\lambda_\gamma)}} \prod_{\substack{\beta\\ \lambda_\beta, \pi(\lambda_\beta) \leqslant k}} \delta_{i'_{\lambda_\beta}, i'_{\pi(\lambda_\beta)}} \\
&= d^r \xi\,,
\end{aligned} \tag{D28}$$

where $\xi \in \mathfrak{B}_k(d)$. Thus, using Supplemental Proposition 11 we find $\mathrm{Tr}[\sigma\pi] = d^r d^{||\nu(\xi)||_1}$. Finally, using Supplemental Proposition 12 it follows that when $\pi \neq \sigma^T$ either $r < l$ or $||\nu(\xi)||_1 < k - l$, so that

$$\mathrm{Tr}[\sigma\pi] \leqslant d^{k-1}\,. \tag{D29}$$

$\square$

Let us now go back to computing the twirl $\mathcal{T}^{(k)}_{\mathbb{O}}[X]$. First, analogously to what we did for the case of the unitary group, we reorder the basis $\mathcal{S}^{(k)}(\mathbb{O}(d))$ in such a way that the first element is $F_d(e)$, followed by the elements $F_d(\sigma)$ that fulfill $F_d(\sigma) = F_d(\sigma)^T$, and finally we order the rest of the elements $\sigma \neq \sigma^T$ by placing $F_d(\sigma)$ next to $F_d(\sigma)^T$. Here, we remark that if $\sigma \in S_k$, $F_d(\sigma)^T = F_d(\sigma^T) = F_d(\sigma^{-1})$. We also recall again that the elements such that $\sigma = \sigma^{-1}$ are known as involutions and must consist of a product of disjoint transpositions plus fixed points. The number of involutions is given by $I_k$ (defined above). More generally, we have that for an element $\sigma \in \mathfrak{B}_k(d)$ to satisfy that $F_d(\sigma) = F_d(\sigma)^T$ it must consist of a product of length-two cycles and fixed points. The number of such elements is $\mathfrak{I}_k = \sum_{\eta=0}^{\lfloor \frac{k}{2} \rfloor} 2^\eta \binom{k}{2\eta} (2\eta-1)!!$.

Then, the following result holds.

**Supplemental Proposition 14.** *The $A$ matrix, of dimension $\frac{(2k)!}{2^k k!} \times \frac{(2k)!}{2^k k!}$, can be expressed as*

$$A = d^k (\widetilde{A} + \frac{1}{d} B)\,. \tag{D30}$$

*Here we defined*

$$\widetilde{A} = \mathbb{1}_{\mathfrak{I}_k} \bigoplus_{j=1}^{\left(\frac{(2k)!}{2^{k+1}k!} - \frac{\mathfrak{I}_k}{2}\right)} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \tag{D31}$$

*where $\mathbb{1}_{\mathfrak{I}_k}$ denotes the $\mathfrak{I}_k \times \mathfrak{I}_k$ dimensional identity. Moreover, the matrix $B$ is such that its entries are $\mathcal{O}(1)$.*

More visually, the matrix $\widetilde{A}$ is of the form

$$\widetilde{A} = \begin{pmatrix} \underbrace{\boxed{\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{matrix}}}_{\mathfrak{I}_k \times \mathfrak{I}_k} & & & \Large 0 \\ & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} & & \\ & & \ddots & \\ \Large 0 & & & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \end{pmatrix} .$$

And we remark the fact that $\widetilde{A}$ its is own inverse. That is, $\widetilde{A}^{-1} = \widetilde{A}$.

*Proof.* Let us recall that the entries of the matrix $A$ are of the form $A_{\nu\mu} = \mathrm{Tr}[P_\nu P_\mu]$ where $P_\nu, P_\mu \in \mathcal{S}^{(k)}(\mathbb{O}(d))$. From Supplemental Proposition 13 it follows that

$$A = \begin{pmatrix} d^k & a_{1,2} & \cdots & a_{1,\mathfrak{I}_k} & & & & & & \\ a_{2,1} & d^k & \cdots & a_{2,\mathfrak{I}_k} & & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & & \\ a_{\mathfrak{I}_k,1} & \cdots & a_{\mathfrak{I}_k,\mathfrak{I}_k-1} & d^k & & & & & & \\ & & & & a_{\mathfrak{I}_k+1,\mathfrak{I}_k+1} & d^k & & & & \\ & & & & d^k & a_{\mathfrak{I}_k+2,\mathfrak{I}_k+2} & & & & \\ & & & & \cdots & & \ddots & & & \\ & & & & & & & a_{\frac{(2k)!}{2^k k!}-1,\frac{(2k)!}{2^k k!}-1} & d^k \\ & & & & & & & d^k & a_{\frac{(2k)!}{2^k k!},\frac{(2k)!}{2^k k!}} \end{pmatrix} ,$$

where the matrix elements $a_{ij} \leqslant d^{k-1}$. This allows us to express the matrix $A$ as

$$A = d^k(\widetilde{A} + \frac{1}{d}B)\,, \tag{D32}$$

where the entries in $B$ are at most equal to 1.

$\square$

Using Supplemental Lemma 1, setting $M = \widetilde{A}$, $\Omega = \frac{1}{d}B$ and noting that $A \propto (\widetilde{A} + \frac{1}{d}B)$ always has inverse [47, 49] we find

$$A^{-1} = \frac{1}{d^k}\left(\widetilde{A} - \frac{1}{d}(\mathbb{1} + \frac{1}{d}\widetilde{A}B)\widetilde{A}B\widetilde{A}\right) = \frac{1}{d^k}\left(\widetilde{A} + C\right)\,, \tag{D33}$$

where we have defined

$$C = \frac{1}{d}(\mathbb{1} + \frac{1}{d}\widetilde{A}B)\widetilde{A}B\widetilde{A}\,. \tag{D34}$$

It is easy to verify that the matrix entries of $C$ are in $\mathcal{O}(1/d)$. Combining the previous result with Eqs. (B5) and (B9) leads to

$$\mathcal{T}_{\mathbb{O}}^{(k)}[X] = \frac{1}{d^k} \sum_{\sigma \in \mathfrak{B}_k(d)} \mathrm{Tr}[X P_d(\sigma)] P_d(\sigma^T) + \frac{1}{d^k} \sum_{\sigma, \Pi \in \mathfrak{B}_k(d)} c_{\sigma,\pi} \mathrm{Tr}[X P_d(\sigma)] P_d(\pi) \,, \tag{D35}$$

where the $c_{\sigma,\pi}$ are the matrix entries of $C$ as defined in Eq. (D34). This is then precisely the statement of Supplemental Theorem 5.

### 2. Computing expectation values of twirled operators

Let us now consider an expectation value of the form

$$C(\rho_i) = \mathrm{Tr}\big[U \rho_i U^\dagger O\big] \,, \tag{D36}$$

where $\rho_i$ is a pure quantum state and $O$ is a traceless quantum operator such that $O^2 = \mathbb{1}$. Next, let us consider the task of estimating expectation values of the form

$$\mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k C(\rho_i) \right] = \mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k \mathrm{Tr}\big[U \rho_i U^\dagger O\big] \right] . \tag{D37}$$

Here, we will show that in the large $d$ limit, the following theorem holds.

**Supplemental Theorem 6.** *Let $\rho_i$ for $i = 1, \ldots, k$ be a set of pure quantum states such that $\mathrm{Tr}[\rho_i \rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $i, i'$, and let $O$ be some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then let us define $\mathfrak{T}_k \subseteq \mathfrak{B}_k(d)$ the set of all possible $k/2$ disjoint cycles of length two. That is, for any $\sigma \in \mathfrak{T}_k$, its cycle decomposition is $\sigma = c_1 \cdots c_{k/2}$ where $c_\alpha$ is a length-two cycle for all $\alpha = 1, \ldots, k/2$. Then, in the large $d$ limit we have*

$$\mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k C(\rho_i) \right] = \frac{1}{d^{k/2}} \sum_{\sigma \in \mathfrak{T}_k} \prod_{\alpha=1}^{k/2} \mathrm{Tr}\big[\rho_{c_\alpha(1)} \rho_{c_\alpha(2)}\big] \,. \tag{D38}$$

To prove this theorem, let us first re-write

$$\mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k C(\rho_i) \right] = \mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k \mathrm{Tr}\big[U \rho_i U^\dagger O\big] \right] = \mathbb{E}_{\mathbb{O}(d)} \left[ \mathrm{Tr}\big[U^{\otimes k} \Lambda \, (U^\dagger)^{\otimes k} O^{\otimes k}\big] \right] , \tag{D39}$$

where $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$. Explicitly,

$$\begin{aligned} \mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k C(\rho_i)) \right] &= \int_{\mathbb{O}(d)} d\mu(U) \, \mathrm{Tr}\big[U^{\otimes k} \Lambda \, (U^\dagger)^{\otimes k} O^{\otimes k}\big]\big] \\ &= \mathrm{Tr}\left[ \left( \int_{\mathbb{O}(d)} d\mu(U) U^{\otimes k} \Lambda \, (U^\dagger)^{\otimes k} \right) O^{\otimes k} \right] . \end{aligned} \tag{D40}$$

Using Supplemental Theorem 5 we readily find

$$\mathbb{E}_{\mathbb{O}(d)} \left[ \prod_{i=1}^k C(\rho_i) \right] = \frac{1}{d^k} \sum_{\sigma \in \mathfrak{B}_k(d)} \mathrm{Tr}[\Lambda P_d(\sigma)] \, \mathrm{Tr}\big[P_d(\sigma^T) O^{\otimes k}\big] + \frac{1}{d^k} \sum_{\sigma, \Pi \in \mathfrak{B}_k(d)} c_{\sigma,\pi} \mathrm{Tr}[\Lambda P_d(\sigma)] \, \mathrm{Tr}\big[P_d(\pi) O^{\otimes k}\big] . \tag{D41}$$

Now, let us prove the following proposition.

**Supplemental Proposition 15.** *Let $O$ be a traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then we have $\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big] = 0$ for any $\sigma \in \mathfrak{B}_k(d)$ if $k$ is odd, and $\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big] = d^r$ if $k$ is even and $\sigma$ is a product of $r$ disjoint cycles of even length. The maximum of $\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big]$ is therefore achieved when $\sigma$ is a product of $k/2$ disjoint cycles of length two, leading to $\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big] = d^{k/2}$.*

*Proof.* We first consider the case of $k$ being odd. Let us express $\sigma$ in its cycle decomposition $\sigma = c_1 \cdots c_r$ as in Definition 5. We have that

$$\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big] = \prod_{\alpha=1}^{r} \mathrm{Tr}\Big[F_d(c_\alpha)O^{\otimes |c_\alpha|}\Big] . \tag{D42}$$

Because $k$ is odd, we know that there must exist at least one cycle acting on an odd number of subsystems in the right hand side of Eq. (D42). Let us assume that this occurs for the cycle $c_{\alpha'}$. Then, we will have

$$\mathrm{Tr}\Big[F_d(c_{\alpha'})O^{\otimes |c_{\alpha'}|}\Big] = \mathrm{Tr}\Big[O^{|c_{\alpha'}|}\Big] = \mathrm{Tr}[O] = 0 . \tag{D43}$$

Here we have used the fact that $O^2 = \mathbb{1}$, $O^T = O$, and hence, since $|c_{\alpha'}|$ is odd, we have $O^{|c_{\alpha'}|} = O$.

Next, let us consider the case of $k$ being even. We know from Eq. (D42) that if $\sigma$ contains any cycle acting on an odd number of subsystems, then $\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big]$ will be equal to zero. This means that only the elements $\sigma$ composed entirely of cycles acting on an even number of subsystems will have non-vanishing trace. If this is the case, we will have

$$\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big] = \prod_{\alpha=1}^{r} \mathrm{Tr}\Big[F_d(c_\alpha)O^{\otimes |c_\alpha|}\Big] = \prod_{\alpha=1}^{r} d = d^r . \tag{D44}$$

This follows from the fact that if $|c_\alpha|$ is even, then $O^{|c_\alpha|} = \mathbb{1}$. Moreover, Eq. (D44) will be maximized for the case when $r$ is largest, which corresponds to the case when $\sigma$ is a product of $k/2$ disjoint length-two cycles. For this special case one finds

$$\mathrm{Tr}\big[F_d(\sigma)O^{\otimes k}\big] = d^{k/2} . \tag{D45}$$

□

Next, let us prove the following proposition.

**Supplemental Proposition 16.** *Let $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$ be a tensor product of $k$ pure states. Then $\big|\mathrm{Tr}[\Lambda F_d(\sigma)] + \mathrm{Tr}\big[\Lambda F_d(\sigma^T)\big]\big| \leqslant 2$ for all $\sigma \in \mathfrak{B}_k(d)$.*

*Proof.* Let us again decompose $\sigma$ in its cycle decomposition $\sigma = c_1 \cdots c_r$ as in Definition 2. Then, we will have

$$\mathrm{Tr}[\Lambda F_d(\sigma)] = \prod_{\alpha=1}^{r} \mathrm{Tr}\big[\rho_{c_\alpha(1)} \cdots \rho_{c_\alpha(|c_\alpha|)}\big] , \tag{D46}$$

where we have used the fact that since all the states are real-valued, then $\rho_i^T = \rho_i$. Here we find it convenient to define $\Lambda_1^\alpha = \rho_{c_\alpha(1)}$ and $\Lambda_2^\alpha = \rho_{c_\alpha(2)} \cdots \rho_{c_\alpha(|c_\alpha|)}$. We note that in general $\mathrm{Tr}[\Lambda_1^\alpha \Lambda_2^\alpha]$ is a complex number (as $\Lambda_2^\alpha$ is not necessarily Hermitian). However, it is not hard to see that

$$\mathrm{Tr}\big[\Lambda F_d(\sigma^T)\big] = \prod_{\alpha=1}^{r} \mathrm{Tr}\big[\Lambda_1^\alpha (\Lambda_2^\alpha)^\dagger\big] . \tag{D47}$$

Hence,

$$\big|\mathrm{Tr}[\Lambda F_d(\sigma)] + \mathrm{Tr}\big[\Lambda F_d(\sigma^T)\big]\big| = 2\,|\mathrm{Re}[\mathrm{Tr}[\Lambda F_d(\sigma)]]| \leqslant 2\,|\mathrm{Tr}[\Lambda F_d(\sigma)]| . \tag{D48}$$

Here we have used the fact that the conjugate of a product of complex numbers is the product of the complex conjugates, plus the fact that for any $z \in \mathbb{C}$, $|\mathrm{Re}[z]| \leqslant |z|$.

Let us now use the Matrix Holder inequality

$$|\mathrm{Tr}[\Lambda_1\Lambda_2]| \leqslant \sqrt{\mathrm{Tr}[(\Lambda_1^\alpha)^2]}\sqrt{\mathrm{Tr}[(\Lambda_2^\alpha)(\Lambda_2^\alpha)^\dagger]}. \tag{D49}$$

We explicitly find

$$\mathrm{Tr}\left[(\Lambda_1^\alpha)^2\right] = \mathrm{Tr}\left[\rho^2_{c_\alpha(1)}\right] = 1\,, \tag{D50}$$

and

$$\mathrm{Tr}\left[(\Lambda_2^\alpha)(\Lambda_2^\alpha)^\dagger\right] = \mathrm{Tr}\left[\rho_{c_\alpha(2)}\cdots\rho_{c_\alpha(|c_\alpha|)}\rho_{c_\alpha(|c_\alpha|)}\cdots\rho_{c_\alpha(2)}\right] = \prod_{\eta=2}^{|c_\alpha|-1}\mathrm{Tr}\left[\rho_{c_\alpha(\eta)}\rho_{c_\alpha(\eta+1)}\right] \leqslant 1\,, \tag{D51}$$

where we have used the fact that the states $\rho_{i_\gamma}$ are pure. Combining the previous results leads to

$$\left|\mathrm{Tr}[\Lambda F_d(\sigma)] + \mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\right| \leqslant 2\,, \tag{D52}$$

for any $\sigma \in \mathfrak{B}_k(d)$. □

With Supplemental Propositions 15 and 16 we can now state the following result.

**Supplemental Proposition 17.** *Let $O$ be a traceless Hermitian operator such that $O^2 = \mathbb{1}$. Let $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$ be a tensor product of $k$ pure states $\rho_{i_\gamma}$. Then, for all $\pi$ and $\sigma$ in $\mathfrak{B}_k(d)$*

$$\frac{1}{d^k}\left|\left(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda F_d(\sigma)] + c_{\sigma^T,\pi}\,\mathrm{Tr}\left[\Lambda P_d(\sigma^T)\right]\right)\mathrm{Tr}\left[F_d(\pi)O^{\otimes k}\right]\right| \in \mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right)\,. \tag{D53}$$

*Proof.* We begin by assuming, without loss of generality, that $|c_{\sigma^T,\pi}| \leqslant |c_{\sigma,\pi}|$. Thus, we have

$$\frac{1}{d^k}\left|\left(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda F_d(\sigma)] + c_{\sigma^T,\pi}\,\mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\right)\mathrm{Tr}\left[F_d(\pi)O^{\otimes k}\right]\right| \leqslant \frac{|c_{\sigma,\pi}|}{d^k}\left|\left(\mathrm{Tr}[\Lambda F_d(\sigma)] + \mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\right)\mathrm{Tr}\left[F_d(\pi)O^{\otimes k}\right]\right|\,.$$

Then, from Supplemental Propositions 15 and 16 we find

$$\frac{|c_{\sigma,\pi}|}{d^k}\left|\left(\mathrm{Tr}[\Lambda F_d(\sigma)] + \mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\right)\mathrm{Tr}\left[F_d(\pi)O^{\otimes k}\right]\right| \leqslant \frac{|c_{\sigma,\pi}|}{d^k}2d^{k/2} = \frac{2|c_{\sigma,\pi}|}{d^{k/2}}\,. \tag{D54}$$

Finally, since from Supplemental Theorem 5 $|c_{\sigma,\pi}| \in \mathcal{O}(1/d)$, we have

$$\left|\frac{1}{d^k}\left(c_{\sigma,\pi}\,\mathrm{Tr}[\Lambda F_d(\sigma)] + c_{\sigma^T,\pi}\,\mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\right)\mathrm{Tr}\left[F_d(\pi)O^{\otimes k}\right]\right| \in \mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right)\,. \tag{D55}$$

□

Finally, consider the following proposition.

**Supplemental Proposition 18.** *Let $O$ be a traceless Hermitian operator such that $O^2 = \mathbb{1}$. Let $\Lambda = \rho_1 \otimes \cdots \otimes \rho_k$ be a tensor product of $k$ pure states $\rho_i$ such that $\mathrm{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $i_\gamma, i_{\gamma'}$. Then,*

$$\frac{1}{d^k}\,\mathrm{Tr}[\Lambda F_d(\sigma)]\,\mathrm{Tr}\left[F_d(\sigma)O^{\otimes k}\right] \in \Omega\left(\frac{1}{d^{k/2}}\right) \tag{D56}$$

*if $\sigma$ is a product of $k/2$ disjoint length-two cycles, and*

$$\frac{1}{d^k}\left|\mathrm{Tr}[\Lambda F_d(\sigma)]\,\mathrm{Tr}\left[F_d(\sigma^T)O^{\otimes k}\right] + \mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\mathrm{Tr}\left[F_d(\sigma)O^{\otimes k}\right]\right| \in \mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right) \tag{D57}$$

*for any other $\sigma$.*

*Proof.* We start by considering the case when $\sigma$ is a product of $k/2$ disjoint length-two cycles. This implies that $\sigma = \sigma^T$. From Supplemental Proposition 15, we know that $\mathrm{Tr}\left[F_d(\sigma)O^{\otimes k}\right] = d^{k/2}$. Then, we find that

$$\mathrm{Tr}[\Lambda F_d(\sigma)] = \prod_{\alpha=1}^{k/2} \mathrm{Tr}\left[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\right] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right), \tag{D58}$$

where we have used the fact that $\mathrm{Tr}\left[\rho_{i_\gamma}\rho_{i_{\gamma'}}\right] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $\gamma, \gamma'$. Thus, it follows that

$$\frac{1}{d^k}\mathrm{Tr}[\Lambda F_d(\sigma)]\,\mathrm{Tr}\left[F_d(\sigma^T)O^{\otimes k}\right] \in \Omega\left(\frac{1}{d^{k/2}}\right). \tag{D59}$$

Next, let us consider the case of $\sigma$ not being a product of $k/2$ disjoint length-two cycles but containing only cycles of even length (if $\sigma$ contains a cycle of odd length, it follows from Supplemental Proposition 15 that $\mathrm{Tr}\left[F_d(\sigma^T)O^{\otimes k}\right] = 0$). Here, we have from Supplemental Proposition 15 that

$$\mathrm{Tr}\left[F_d(\sigma)O^{\otimes k}\right] = \mathrm{Tr}\left[F_d(\sigma)O^{\otimes k}\right] = d^r \tag{D60}$$

with $r \leqslant \frac{k}{2} - 1$. Then, the following chain of inequalities hold

$$\begin{aligned}\frac{1}{d^k}\left|\mathrm{Tr}[\Lambda F_d(\sigma)]\,\mathrm{Tr}\left[F_d(\sigma^T)O^{\otimes k}\right] + \mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\mathrm{Tr}\left[F_d(\sigma)O^{\otimes k}\right]\right| &\leqslant \frac{1}{d^{\frac{k+2}{2}}}\left|\mathrm{Tr}[\Lambda F_d(\sigma)] + \mathrm{Tr}\left[\Lambda F_d(\sigma^T)\right]\right| \\ &\leqslant \frac{2}{d^{\frac{k+2}{2}}} \in \mathcal{O}\left(\frac{1}{d^{\frac{k+2}{2}}}\right).\end{aligned} \tag{D61}$$

Here, we have used Supplemental Proposition 16 for the last inequality. □

To finish the proof of Supplemental Theorem 6 we simply combine Supplemental Propositions 17 and 18 and note that in the large $d$ limit we get

$$\mathbb{E}_{\mathbb{O}(d)}\left[\prod_{\gamma=1}^{k} C_\gamma\right] = \frac{1}{d^{k/2}}\sum_{\sigma\in\mathfrak{T}_k}\prod_{\alpha=1}^{k/2}\mathrm{Tr}\left[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\right], \tag{D62}$$

where we have defined as $\mathfrak{T}_k \subseteq \mathfrak{B}_k(d)$ the set of elements in the Brauer algebra which are exactly given by a product of $k/2$ disjoint length-two cycles. Here we note that every such two-cycle can either be of the form $(\{i, j+k\}\cup\{j, i+k\})$ for $i, j \leqslant k$ or $(\{i, j\}\cup\{i+k, j+k\})$ for $i, j \leqslant k$. Moreover since for either of those two cases the term $\mathrm{Tr}\left[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\right]$ will be equal, we will have redundancies in the summation of Eq. (D62). We can remove this redundancy by only summing over disjoint transpositions, and adding a coefficient $2^{k/2}$ that account for the fact that in each of the $k/2$ length-two cycles, we can choose either $(\{i, j+k\}\cup\{j, i+k\})$ for $i, j \leqslant k$ or $(\{i, j\}\cup\{i+k, j+k\})$ for $i, j \leqslant k$. Then we obtain

$$\mathbb{E}_{\mathbb{O}(d)}\left[\prod_{\gamma=1}^{k} C_\gamma\right] = \frac{2^{k/2}}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\alpha=1}^{k/2}\mathrm{Tr}\left[\rho_{c_\alpha(1)}\rho_{c_\alpha(2)}\right]. \tag{D63}$$

### 3. Orthogonal states

When considering orthogonal states we can derive the following results.

**Supplemental Theorem 7.** *Let $\rho_i$ for $i = 1, \ldots, k$ be a set of pure and mutually orthogonal quantum states, and let $O$ be some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then, in the large $d$ limit we have*

$$\mathbb{E}_{\mathbb{O}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{(k)!}{2^{k/2}(k/2)!}\frac{2^{k/2}}{d^k}. \tag{D64}$$

The proof of this theorem follows similarly to that of Supplemental Theorem 3 but noting that here the terms that contribute are not only the $k/2$ disjoint transpositions, but also any element of $\mathfrak{B}_k(d)$ where any transposition $(\{i, j+k\} \cup \{j, i+k\})$ for $i, j \leqslant k$ is replaced by a term $(\{i, j\} \cup \{i+k, j+k\})$ for $i, j \leqslant k$. Since we have $k/2$ such choices, then we get an extra factor of $2^{k/2}$. Said otherwise, we will get all the contributions of the element of $\mathfrak{B}_k(d)$ which are composed only of cycles of length two.

With a similar argument we can find that the following result holds.

**Supplemental Theorem 8.** *Let $\rho_i$ for $i = 1, \ldots, q$ be a set of pure and mutually orthogonal states. Then, let $\Lambda$ contains $k_1$ copies of $\rho_1$, $k_2$ copies of $\rho_2$, and so on. In total, we assume that $\Lambda$ contains $q$ different states and that $\sum_{\beta=1}^{q} k_\beta = k$. Moreover, we assume that $O$ is some traceless Hermitian operator such that $O^2 = \mathbb{1}$. Then, in the large $d$ limit we have*

$$\mathbb{E}_{\mathbb{O}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{2^{\sum_{k_\beta \in K_2} \lfloor \frac{k_\beta}{2} \rfloor} d^{\sum_{k_\beta \in K_2} \lfloor \frac{k_\beta}{2} \rfloor}}{d^k} \left( \sum_{k_\beta \% 2 = 1} k_\beta \frac{(2\lfloor k_\beta/2 \rfloor)!}{2^{\lfloor k_\beta/2 \rfloor}(\lfloor [k_\beta/2 \rfloor)!} + \sum_{k_\beta \% 2 = 0} \frac{(k_\beta)!}{2^{k_\beta/2}([k_\beta/2)!} \right) . \tag{D65}$$

## Supp. Info. E: Proof of Lemma 1

In this section we present a proof of Lemma 1, which we recall for convenience.

**Lemma 1.** *Let $C_j(\rho_i)$ be the expectation value of a Haar random QNN as in Eq. (1). Then for any $\rho_i \in \mathscr{D}$, $O_j \in \mathscr{O}$,*

$$\mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i)] = \mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)] = 0 . \tag{E1}$$

*Moreover, for any pair of states $\rho_i, \rho_{i'} \in \mathscr{D}$ and operators $O_j, O_{j'} \in \mathscr{O}$ we have*

$$\mathrm{Cov}_{\mathbb{U}(d)}[C_j(\rho_i) C_{j'}(\rho_{i'})] = \mathrm{Cov}_{\mathbb{O}(d)}[C_j(\rho_i) C_{j'}(\rho_{i'})] = 0 ,$$

*if $j \neq j'$ and*

$$\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'} = \frac{d}{d^2-1}\left(\mathrm{Tr}[\rho_i \rho_{i'}] - \frac{1}{d}\right) , \tag{E2}$$

$$\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'} = \frac{2(d+1)}{(d+2)(d-1)}\left(\mathrm{Tr}[\rho_i \rho_{i'}]\left(1 - \frac{1}{d+1}\right) - \frac{1}{d+1}\right) , \tag{E3}$$

*if $j = j'$. Here, we have defined $\boldsymbol{\Sigma}^{G}_{i,i'} = \mathrm{Cov}_G[C_j(\rho_i) C_j(\rho_{i'})]$, where $G = \mathbb{U}(d), \mathbb{O}(d)$.*

*Proof.* We begin by considering expectation values over $\mathbb{U}(d)$. Using Eq. (C6) we can compute

$$\mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i)] = \frac{\mathrm{Tr}[\rho_i]\,\mathrm{Tr}[O_j]}{d} = 0 , \tag{E4}$$

where we have used the fact that $O_j$ is traceless.

Next, we can use Eq. (C11) to find

$$\begin{aligned}
\mathrm{Cov}_{\mathbb{U}(d)}[C_j(\rho_i) C_{j'}(\rho_{i'})] =& \mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i) C_{j'}(\rho_{i'})] \\
=& \frac{1}{d^2-1}\left(\mathrm{Tr}[(\rho_i \otimes \rho_{i'})] - \frac{\mathrm{Tr}[(\rho_i \otimes \rho_{i'})\mathrm{SWAP}]}{d}\right) \mathrm{Tr}[(O_j \otimes O_{j'})(\mathbb{1} \otimes \mathbb{1})] \\
&+ \frac{1}{d^2-1}\left(\mathrm{Tr}[(\rho_i \otimes \rho_{i'})\mathrm{SWAP}] - \frac{\mathrm{Tr}[(\rho_i \otimes \rho_{i'})]}{d}\right) \mathrm{Tr}[(O_j \otimes O_{j'})\mathrm{SWAP}] \\
=& \frac{1}{d^2-1}\left(\mathrm{Tr}[\rho_i \rho_{i'}] - \frac{1}{d}\right) \mathrm{Tr}[O_j O_{j'}] \\
=& \frac{1}{d^2-1}\left(\mathrm{Tr}[\rho_i \rho_{i'}] - \frac{1}{d}\right) \delta_{j,j'} .
\end{aligned}$$

Here we have used the fact that the states are pure plus the fact that, by definition, $\mathrm{Tr}[O_j O_{j'}]=\delta_{j,j'}$. Hence, we find that if $j\neq j'$, then $\mathrm{Cov}_{\mathbb{U}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]=0$, whereas if $j\neq j'$ one obtains

$$\mathrm{Cov}_{\mathbb{U}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]=\frac{1}{d^2-1}\left(\mathrm{Tr}[\rho_i\rho_{i'}]-\frac{1}{d}\right). \tag{E5}$$

Next, let us take expectation values over $\mathbb{O}(d)$. Using Eq. (D7) we again find

$$\mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)]=\frac{\mathrm{Tr}[\rho_i]\,\mathrm{Tr}[O_j]}{d}=0\,. \tag{E6}$$

Next, from Eq. (D12) we obtain

$$\begin{aligned}\mathrm{Cov}_{\mathbb{O}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]=&\mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]\\ =&\frac{1}{d(d+2)(d-1)}\left((d+1)\,\mathrm{Tr}[(\rho_i\otimes\rho_{i'})]-\mathrm{Tr}[(\rho_i\otimes\rho_{i'})\mathrm{SWAP}]-\mathrm{Tr}[(\rho_i\otimes\rho_{i'})\Pi]\right)\mathrm{Tr}[(O_j\otimes O_{j'})(\mathbb{1}\otimes\mathbb{1})]\\ &+\frac{1}{d(d+2)(d-1)}\left(-\mathrm{Tr}[(\rho_i\otimes\rho_{i'})]+(d+1)\,\mathrm{Tr}[(\rho_i\otimes\rho_{i'})\mathrm{SWAP}]-\mathrm{Tr}[(\rho_i\otimes\rho_{i'})\Pi]\right)\mathrm{Tr}[(O_j\otimes O_{j'})\mathrm{SWAP}]\\ &+\frac{1}{d(d+2)(d-1)}\left(-\mathrm{Tr}[(\rho_i\otimes\rho_{i'})]-\mathrm{Tr}[(\rho_i\otimes\rho_{i'})\mathrm{SWAP}]+(d+1)\,\mathrm{Tr}[(\rho_i\otimes\rho_{i'})\Pi]\right)\mathrm{Tr}[(O_j\otimes O_{j'})\Pi]\,.\end{aligned}$$

Using the ricochet property of Eq. (A7) plus the fact that the operators $\rho_i$, $\rho_{i'}$, $O_j$ and $O_{j'}$ are real-valued leads to

$$\begin{aligned}\mathrm{Cov}_{\mathbb{O}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]&=\frac{2}{d(d+2)(d-1)}\left(-1+(d+1)\,\mathrm{Tr}[\rho_i\rho_{i'}]-\mathrm{Tr}[\rho_i\rho_{i'}]\right)\delta_{j,j'}\\ &=\frac{2(d+1)}{(d+2)(d-1)}\left(\mathrm{Tr}[\rho_i\rho_{i'}]\left(1-\frac{1}{d+1}\right)-\frac{1}{d+1}\right)\delta_{j,j'}\,.\end{aligned} \tag{E7}$$

If $j\neq j'$, then $\mathrm{Cov}_{\mathbb{O}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]=0$, whereas if $j\neq j'$ one obtains

$$\mathrm{Cov}_{\mathbb{O}(d)}[C_j(\rho_i)C_{j'}(\rho_{i'})]=\frac{2(d+1)}{(d+2)(d-1)}\left(\mathrm{Tr}[\rho_i\rho_{i'}]\left(1-\frac{1}{d+1}\right)-\frac{1}{d+1}\right). \tag{E8}$$

$\square$

## Supp. Info. F: Proof of Lemma 2

Let us now prove Lemma 2, which we recall for convenience.

**Lemma 2.** *Let $\mathscr{C}$ be a vector of $k$ expectation values of a Haar random QNN as in Eq. (2), where one measures the same operator $O_j$ over a set of $k$ states $\rho_1,\dots,\rho_k\in\mathscr{D}$. In the large-$d$ limit we find that if $k$ is odd then $\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]=\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]=0$. Moreover, if $k$ is even and if a) $\mathrm{Tr}[\rho_i\rho_{i'}]\in\Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$ for all $i,i'$, we have*

$$\begin{aligned}\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]&=\frac{1}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\{t,t'\}\in\sigma}\mathrm{Tr}[\rho_t\rho_{t'}]\\ &=\frac{\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]}{2^{k/2}}\,,\end{aligned} \tag{F1}$$

*where the summation runs over all the possible disjoint pairing of indexes in the set $\{1,2,\dots,k\}$, $T_k$, and the product is over the different pairs in each pairing; while if b) $\mathrm{Tr}[\rho_i\rho_{i'}]=0$ for all $i,i'$, we have*

$$\begin{aligned}\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]&=\frac{k!}{2^{k/2}(k/2)!}\frac{1}{d^k}\\ &=\frac{\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho_1)\cdots C_j(\rho_k)\right]}{2^{k/2}}\,.\end{aligned} \tag{F2}$$

*Proof.* The proof of this lemma follows from results previously derived in this SI. Namely, we simply need to use Eqs. (C55) and (C58) for the unitary group, and Eqs. (D63) and (D62) for the orthogonal group.

□

## Supp. Info. G: Proof of Corollary 1

For ease of calculation, let us first prove Corollary 1 before our main theorems. The statement of Corollary 1 is as follows.

**Corollary 1.** *Let $C_j(\rho_i)$ be the expectation value of a Haar random QNN as in Eq. (1). Then for any $\rho_i \in \mathscr{D}$ and $O_j \in \mathscr{O}$, we have*

$$P(C_j(\rho_i)) = \mathcal{N}(0, \sigma^2)\,, \tag{G1}$$

*where $\sigma^2 = \frac{1}{d}, \frac{2}{d}$ when $U$ is Haar random over $\mathbb{U}(d)$ and $\mathbb{O}(d)$, respectively.*

*Proof.* In order to prove Corollary 1 we need to consider two distinct cases: when we sample over the unitary and orthogonal groups.

We start by taking averages over $\mathbb{U}(d)$. Using Lemma 2 we have that the $k$-th moment (for $k$ even) of the distribution is

$$\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho)^k\right] = \frac{1}{d^{k/2}} \sum_{\sigma\in T_k} \prod_{\{t,t'\}\in\sigma} \mathrm{Tr}[\rho\rho] = \frac{1}{d^{k/2}} \sum_{\sigma\in T_k} \prod_{\{t,t'\}\in\sigma} 1 = \frac{k!}{d^{k/2} 2^{k/2} (k/2)!}. \tag{G2}$$

Clearly, we can verify that these moments match those of a Gaussian distribution, as

$$\frac{\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho)^k\right]}{\mathbb{E}_{\mathbb{U}(d)}\left[C_j(\rho)^2\right]^{k/2}} = \frac{k!}{2^{k/2}(k/2)!}\,. \tag{G3}$$

To prove that these moments uniquely determine the distribution of $\mathscr{C}$, we use Carleman's condition.

**Supplemental Lemma 2** (Carleman's condition, Hamburger case [34])**.** *Let $\gamma_k$ be the (finite) moments of the distribution of a random variable $X$ that can take values on the real line $\mathbb{R}$. These moments determine uniquely the distribution of $X$ if*

$$\sum_{k=1}^{\infty} \gamma_{2k}^{-1/2k} = \infty\ . \tag{G4}$$

In our case, we have

$$\sum_{k=1}^{\infty}\left(\frac{1}{d^k}\frac{(2k)!}{2^k k!}\right)^{-1/2k} = \sqrt{2d}\sum_{k=1}^{\infty}\left((2k)\cdots(k+1)\right)^{-1/2k} \geqslant \sum_{k=1}^{\infty}\left((2k)^k\right)^{-1/2k} = \sum_{k=1}^{\infty}\frac{1}{\sqrt{2k}} = \infty\,, \tag{G5}$$

and so, according to Supplemental Lemma 2, Carleman's condition is satisfied. Therefore, $P(C_j(\rho_i)$ is distributed according to a Gaussian distribution with zero-mean and variance $\sigma^2 = \frac{1}{d}$.

Next, we consider taking averages over $\mathbb{O}(d)$. Using Lemma 2 we have that the $k$-th moment (for $k$ even) of the distribution is

$$\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho)^k\right] = \frac{k!}{d^{k/2}(k/2)!}. \tag{G6}$$

We can again check that these moments match those of a Gaussian distribution, as one has

$$\frac{\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho)^k\right]}{\mathbb{E}_{\mathbb{O}(d)}\left[C_j(\rho)^2\right]^{k/2}}=\frac{k!}{2^{k/2}(k/2)!}\,, \tag{G7}$$

and also that Carleman's condition is satisfied,

$$\sum_{k=1}^{\infty}\left(\frac{(2k)!}{d^k k!}\right)^{-1/2k}=\sqrt{d}\sum_{k=1}^{\infty}\left((2k)\cdots(k+1)\right)^{-1/2k}\geqslant\sum_{k=1}^{\infty}\left((2k)^k\right)^{-1/2k}=\sum_{k=1}^{\infty}\frac{1}{\sqrt{2k}}=\infty\,. \tag{G8}$$

Therefore, $P(C_j(\rho_i)$ is distributed according to a Gaussian distribution with zero mean and variance $\sigma^2=\frac{2}{d}$.

□

## Supp. Info. H: Proof of Theorem 1

Let us now present a proof for Theorem 1, which we recall here.

**Theorem 1.** *Under the same conditions for which Lemma 2(a) holds, the vector $\mathscr{C}$ forms a GP with mean vector $\boldsymbol{\mu}=\mathbf{0}$ and covariance matrix given by $\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'}=\frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'}}{2}=\frac{\mathrm{Tr}[\rho_i\rho_{i'}]}{d}$.*

*Proof.* First, let us recall that a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, is fully defined by its $m$-dimensional mean vector $\boldsymbol{\mu}=(\mathbb{E}[X_1],\dots,\mathbb{E}[X_m])$, and its $m\times m$ dimensional covariance matrix with entries $(\boldsymbol{\Sigma})_{\alpha\beta}=\mathrm{Cov}[X_\alpha,X_\beta]$. From here, any higher order moments can be obtained from Isserlis theorem [46]. Specifically, if we want to compute a $k$-th order moment, then we have $\mathbb{E}[X_1X_2\cdots X_k]=0$ if $k$ is odd, and

$$\mathbb{E}[X_1X_2\cdots X_k]=\sum_{\sigma\in T_k}\prod_{\{t,t'\}\in\sigma}\mathrm{Cov}[X_t,X_{t'}]\,, \tag{H1}$$

if $k$ is even. Here, the summation runs over all the possible pairing of indexes, $T_k$, in the set $\{1,2,\dots,k\}$.

A direct comparison between the results in Lemma 2 and Eq. (H1) shows that indeed the moments of $\mathscr{C}$ match those of a GP with covariance matrix $\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'}=\frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'}}{2}=\frac{\mathrm{Tr}[\rho_i\rho_{i'}]}{d}$. To prove that these moments uniquely determine the distribution of $\mathscr{C}$, we use the fact that since its marginal distributions are determinate via Carleman's condition (see the previous section), then so is the distribution of $\mathscr{C}$ [34]. This result holds for both the unitary and orthogonal groups.

□

## Supp. Info. I: Proof of Theorem 2

Here we present a proof for Theorem 2, which we restate for convenience.

**Theorem 2.** *Let $\mathscr{C}$ be a vector of $k$ expectation values of operators in $\mathscr{O}$ over a set of $k$ states $\rho_1,\dots,\rho_k\in\mathscr{D}$. If $\mathrm{Tr}[\rho_i\rho_{i'}]=\frac{1}{d}$ for all $i,i'$, then in the large $d$-limit $\mathscr{C}$ forms a GP with mean vector $\boldsymbol{\mu}=\mathbf{0}$ and diagonal covariance matrix*

$$\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'}=\frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'}}{2}=\begin{cases}\frac{1}{d} & \text{if } i=i'\\ 0 & \text{if } i\neq i'\end{cases}\,. \tag{I1}$$

*Proof.* This case follows from the fact the variables in the GP process are independent, and individually given by univariate Gaussian variables (see the proof of Corollary 1 above). Thus, their joint distribution is a multivariate Gaussian distribution with entries of the diagonal covariance matrix given by $\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i}=\frac{\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i}}{2}=\frac{1}{d}$. □

### Supp. Info. J: Proof of Theorem 3

We now provide a proof for Theorem 3, whose statement is as follows.

**Theorem 3.** *Under the same conditions for which Lemma 2(b) holds, the vector $\mathscr{C}$ forms a GP with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix*

$$\boldsymbol{\Sigma}^{\mathbb{U}(d)}_{i,i'} = \begin{cases} \frac{1}{d+1} & \text{if } i = i' \\ -\frac{1}{(d^2-1)} & \text{if } i \neq i' \end{cases} , \tag{J1}$$

*and*

$$\boldsymbol{\Sigma}^{\mathbb{O}(d)}_{i,i'} = \begin{cases} \frac{2}{d+1} & \text{if } i = i' \\ -\frac{1}{(d+2)(d-1)} & \text{if } i \neq i' \end{cases} . \tag{J2}$$

*Proof.* In what follows, we will again us the strategy of proving that the moments of $\mathscr{C}$ match those of a GP. From Isserlis theorem [46] (see Eq. (H1)), we know that if $\mathscr{C}$ is indeed a GP, then

$$\mathbb{E}_G[C_j(\rho_1)\cdots C_j(\rho_k)] = \sum_{\sigma\in T_k}\prod_{\{i,i'\}\in\sigma} \mathrm{Cov}_G[C_j(\rho_i)C_j(\rho_{i'})] \,. \tag{J3}$$

for $G = \mathbb{U}(d), \mathbb{O}(d)$

There are three situations we must consider. When the states $\rho_1, \ldots, \rho_k$ are all the same, when they are all different, and when there are $k_1$ of the same state $\rho_1$, $k_2$ of the same state $\rho_2$, and so on. In the last case, we assume that there contains $q$ different states so that $\sum_{\beta=1}^{q} k_\beta = k$.

Let us start with taking averages over $\mathbb{U}(d)$. If all the states are the same we simply use Supplemental Corollary 1 to find

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{\gamma=1}^{k} C_\gamma\right] = \frac{1}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\alpha=1}^{k/2} = \frac{(k)!}{d^{k/2}2^{k/2}(k/2)!} \,, \tag{J4}$$

and we see that the moments match. Then, if all the states are orthogonal, we use Supplemental Theorem 3 to obtain

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{(k)!}{2^{k/2}(k/2)!}\frac{1}{d^k} \,, \tag{J5}$$

which again shows that the moments match. Finally, if there are $k_q$ copies of $\rho_q$, we know from Supplemental Theorem 4 that in the large $d$ limit the moments go as

$$\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{i=1}^{k} C(\rho_i)\right] = \frac{d^{\sum_{k_\beta\in K_2}\lfloor\frac{k_\beta}{2}\rfloor}}{d^k}\left(\prod_{k_\beta\%2=1} k_\beta\frac{(2\lfloor k_\beta/2\rfloor)!}{2^{\lfloor k_\beta/2\rfloor}(\lfloor k_\beta/2\rfloor)!}\prod_{k_\beta\%2=0}\frac{(k_\beta)!}{2^{k_\beta/2}(k_\beta/2)!}\right) . \tag{J6}$$

This equation can be understood as follows. First, we need to count how many copies of the same state $\rho_q$ can we pair at the same time. This is simply given by $\sum_{k_\beta\%2=0}\lfloor\frac{k_\beta}{2}\rfloor$. Next, we see that the first and second term in the multiplications respectively counts how many different ways we have to pair copies of the same state when $k_\beta \geqslant 2$ is odd or even.

Inspecting Isserlis theorem we see what the sum $\sum_{\sigma\in T_k}\prod_{\{i,i'\}\in\sigma}\mathrm{Cov}_G[C_j(\rho_i)C_j(\rho_{i'})]$ will contain terms that are of order $\frac{1}{d}$ and which come from $\mathrm{Cov}_G[C_j(\rho_i)C_j(\rho_i)]$; and terms that are of order $\frac{1}{d^2}$ coming from $\mathrm{Cov}_G[C_j(\rho_i)C_j(\rho_{i'})]$ for $i \neq i'$. One can readily see that one will have a number of terms of the form $\mathrm{Cov}_G[C_j(\rho_i)C_j(\rho_i)]$ equal to $\sum_{k_\beta\%2=0}\lfloor\frac{k_\beta}{2}\rfloor$. As such, Isserlis theorem indicates that the leading order terms of $\mathbb{E}_G[C_j(\rho_1)\cdots C_j(\rho_k)]$ will scale as $\frac{d^{\sum_{k_\beta\%2}\lfloor\frac{k_\beta}{2}\rfloor}}{d^k}$. Then, counting how many such terms exist leads to $\left(\prod_{k_\beta\%2=1} k_\beta\frac{(2\lfloor k_\beta/2\rfloor)!}{2^{\lfloor k_\beta/2\rfloor}(\lfloor k_\beta/2\rfloor)!}\prod_{k_\beta\%2=0}\frac{(k_\beta)!}{2^{k_\beta/2}(k_\beta/2)!}\right)$.

Taken together, the previous results show that the moments of the deep QNN outcomes indeed match those of a GP. Here, we can again prove that these moments uniquely determine the distribution of $\mathscr{C}$ from the fact that since its marginal distributions are determinate via Carleman's condition (see Corollary 1), then so is the distribution of $\mathscr{C}$ [33, 34].

A similar argument can be made for the orthogonal group using Eq. (8). □

### Supp. Info. K: Proof of Theorem 4

In this section we will provide a proof for Theorem 4.

**Theorem 4.** *Consider a GP obtained from a Haar random QNN. Given the set of observations $(y(\rho_1), \dots, y(\rho_m))$ obtained from $N \in \mathcal{O}(\mathrm{poly}(\log(d)))$ measurements, then the predictive distribution of the GP is trivial:*

$$P(C_j(\rho_{m+1})|C_j(\rho_1), \dots, C_j(\rho_m)) = P(C_j(\rho_{m+1})) = \mathcal{N}(0, \sigma^2)\,,$$

*where $\sigma^2$ is given by Corollary 1.*

*Proof.* Let us consider a setting where we are interested in computing the expectation value of an operator $O$ at the output of a Haar random QNN, $U$. That is, we define the quantity $C(\rho) = \mathrm{Tr}[U\rho U O]$. Then, given $m$ quantum states $\rho_1, \cdots, \rho_m$, we want to predict the expectation value $C(\rho_{m+1})$, given the quantities $C(\rho_1), \dots, C(\rho_m)$. As we have seen in the Methods section, in order to make predictions with the Gaussian process, we need the covariances

$$\mathrm{Cov}_{\mathbb{U}(d)}[C(\rho_i), C(\rho_{i'})] = \frac{d}{d^2-1}\left(\mathrm{Tr}[\rho_i\rho_{i'}] - \frac{1}{d}\right)\,, \tag{K1}$$

$$\mathrm{Cov}_{\mathbb{O}(d)}[C(\rho_i), C(\rho_{i'})] = \frac{2(d+1)}{(d+2)(d-1)}\left(\mathrm{Tr}[\rho_i\rho_{i'}]\left(1 - \frac{1}{d+1}\right) - \frac{1}{d+1}\right)\,. \tag{K2}$$

And we recall that in the large-$d$ limit, $P(C(\rho_i)) = \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \frac{1}{d}, \frac{2}{d}$ for the unitary and orthogonal groups, respectively (see Corollary 1). Moreover, we will assume that the expectation values are computed with $N$ shots. This leads to a statistical noise in the observation procedure that is modeled by a zero-mean Gaussian with a variance given by $\sigma_N^2 = \frac{1}{N}$, so that $\frac{1}{\sigma_N^2} \in \mathcal{O}(N)$.

From Eq. (20), we know that we need to invert the matrix $\mathbf{\Sigma} + \sigma_N^2\mathbb{1}$. Using Supplemental Lemma 1, we can then write

$$(\mathbf{\Sigma} + \sigma_N^2\mathbb{1})^{-1} = \frac{1}{\sigma_N^2}\mathbb{1} - \frac{1}{\sigma_N^4}(\mathbb{1} + \frac{1}{\sigma_N^2}\mathbf{\Sigma})\mathbf{\Sigma}\,. \tag{K3}$$

Noting that the absolute value of the matrix elements of $\mathbf{\Sigma}$ are at most in $\mathcal{O}(\frac{1}{d})$, then in the large-$d$ limit we will have

$$(\mathbf{\Sigma} + \sigma_N^2\mathbb{1})^{-1} \sim \frac{1}{\sigma_N^2}\mathbb{1}\,. \tag{K4}$$

Hence, the mean and variance of $P(C(\rho_{m+1})|C(\rho_1), \dots, C(\rho_m))$ will be

$$\mu(C(\rho_{m+1})) = \frac{1}{\sigma_N^2}\boldsymbol{m}^T \cdot \boldsymbol{C}\,, \tag{K5}$$

$$\sigma^2(C(\rho_{m+1})) = \sigma^2 - \frac{1}{\sigma_N^2}\boldsymbol{m}^T \cdot \boldsymbol{m}\,. \tag{K6}$$

Next, we note that the correction $\frac{1}{\sigma_N^2}\boldsymbol{m}^T \cdot \boldsymbol{C} \in \mathcal{O}(\frac{N}{d})$ and $\frac{1}{\sigma_N^2}\boldsymbol{m}^T \cdot \boldsymbol{m} \in \mathcal{O}(\frac{N}{d^2})$, where we have use the fact that $|C(\rho_\gamma)| \leqslant 1$ for all $\rho_\gamma$ and that the absolute values of the entries of $\boldsymbol{m}$ are in $\mathcal{O}(\frac{1}{d})$. If $N \in \mathcal{O}(\mathrm{poly}(\log(d)))$, we have that in the large-$d$ limit

$$\mu(C(\rho_{m+1})) = 0\,, \tag{K7}$$

$$\sigma^2(C(\rho_{m+1})) = \sigma^2\,. \tag{K8}$$

Thus, using a poly-logarithmic-in-$d$ number of shots we cannot use the Gaussian process to learn anything about the probability of $C(\rho_{m+1})$.

$\square$

### Supp. Info. L: Proof of Corollary 2

In this section we present a proof for Corollary 2, which we restate for convenience.

**Corollary 2.** *Let $C_j(\rho_i)$ be the expectation value of a Haar random QNN as in Eq. (1). Assuming that there exists a parametrized gate in $U$ of the form $e^{-i\theta H}$ for some Pauli operator $H$, then*

$$P(|C_j(\rho_i)| \geqslant c),\, P(|\partial_\theta C_j(\rho_i)| \geqslant c) \in \mathcal{O}\left(\frac{1}{ce^{dc^2}\sqrt{d}}\right).$$

*Proof.* Let us start by evaluating the probability $P(|C_j(\rho_i)| \geqslant c)$. Since we know that $C_j(\rho_i)$ follows a Gaussian distribution $\mathcal{N}(0,\sigma^2)$ with $\sigma^2 = \frac{1}{d}$ (see Corollary 1), we can use the equality

$$P(|C_j(\rho_i)| \geqslant c) = \frac{2\sqrt{d}}{\sqrt{2\pi}}\int_c^\infty dx e^{-x^2 d} = \frac{\mathrm{Erfc}\left[c\sqrt{d}\right]}{\sqrt{2}}\,, \tag{L1}$$

where Erfc denotes the complementary error function. Using the fact that for large $x$, $\mathrm{Erfc}[x] \leqslant \frac{e^{-x^2}}{x\sqrt{\pi}}$, we then find

$$P(|C_j(\rho_i)| \geqslant c) \in \mathcal{O}\left(\frac{1}{ce^{dc^2}\sqrt{d}}\right). \tag{L2}$$

Next, let us consider the probability $P(|\partial_\theta C_j(\rho_i)| \geqslant c)$, and let us write again the explicit dependence of $U$ on some set of parameters $\boldsymbol{\theta}$. That is, we write $U(\boldsymbol{\theta})$. Then, let us note that if the parameter $\theta \in \boldsymbol{\theta}$ appears in $U$ as $e^{-i\theta H}$ for some Pauli operator $H$, we can use the parameter shift-rule [12] to compute

$$\partial_\theta C_j(\rho_i) = C_j^+(\rho_i) - C_j^-(\rho_i)\,, \tag{L3}$$

where

$$C_j^\pm(\rho_i) = \mathrm{Tr}\big[U(\boldsymbol{\theta}^\pm)\rho_i U^\dagger(\boldsymbol{\theta}^\pm)O_j\big]\,, \tag{L4}$$

and $\boldsymbol{\theta}^\pm = \boldsymbol{\theta} + \frac{\pi}{4}\hat{e}_\theta$. Here $\hat{e}_\theta$ denotes a unit vector with an entry equal to one in the same position as that of $\theta$. Then, let us define the events $\mathcal{E}_\pm$ as $|C_j^\pm(\rho_i)| > 2c$. Clearly, $|\partial_\theta C_j(\rho_i)| > c$ is a subset of $\mathcal{E}_+ \cup \mathcal{E}_-$. Hence,

$$\begin{aligned}
P(|\partial_\theta C_j(\rho_i)| \geqslant c) &\leqslant P(\mathcal{E}_+ \cup \mathcal{E}_-) \\
&\leqslant P(\mathcal{E}_+) + P(\mathcal{E}_-) \\
&= \mathrm{Erfc}\left[\frac{c\sqrt{d}}{\sqrt{2}}\right] + \mathrm{Erfc}\left[\frac{c\sqrt{d}}{\sqrt{2}}\right] \in \mathcal{O}\left(\frac{1}{ce^{dc^2}\sqrt{d}}\right).
\end{aligned} \tag{L5}$$

In the second inequality we have used the union bound, and in the first equality we used Eqs. (L1) and (L2).

$\square$

### Supp. Info. M: Proof of Corollary 3

Here we provide a proof for Proof of Corollary 3.

**Corollary 3.** *Let $U$ be drawn from a $t$-design. Then, under the same conditions for which Theorems 1, 2 and 3 hold, the vector $\mathscr{C}$ matches the first $t$ moments of a GP.*

*Proof.* The proof follows directly by using Theorems 1, 2 and 3, and the fact that by definition a $t$-design matches the first $t$ moments of a Haar random unitary. Hence, the first $t$-moments of $\mathscr{C}$ match those of a GP under the same conditions under which Theorems 1, 2 and 3 hold. □

Here we also provide a proof for the following equation presented in the main text:

$$P(|C_j(\rho_i)| \geqslant c),\, P(|\partial_\theta C_j(\rho_i)| \geqslant c) \in \mathcal{O}\left(\frac{(2\left\lfloor\frac{t}{2}\right\rfloor)!}{2^{\left\lfloor\frac{t}{2}\right\rfloor}(dc^2)^{\left\lfloor\frac{t}{2}\right\rfloor}\left(\left\lfloor\frac{t}{2}\right\rfloor\right)!}\right). \tag{M1}$$

We start by considering $P(|C_j(\rho_i)| \geqslant c)$ Here we can use the generalization of Chebyshev's inequality to higher-order moments:

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant c\right) \leqslant \frac{\mathbb{E}[|X-\mathbb{E}[X]|^k]}{c^n}, \tag{M2}$$

for $c > 0$ and for $k \geqslant 2$. If $U$ forms a $t$-design, then we can evaluate $\mathbb{E}[C_j(\rho_i)^t]$ by noting that it matches the $t$ first moment of a $\mathcal{N}(0,\sigma^2)$ distribution with $\sigma^2 = \frac{1}{d}$ (see Corollary 1). In particular, we know that since the odd moments are zero, we only need to care about the largest even moment that the $t$-design matches. Hence we can use

$$\mathbb{E}[|C_j(\rho_i)^{2\left\lfloor\frac{t}{2}\right\rfloor}] = \frac{(2\left\lfloor\frac{t}{2}\right\rfloor)!}{2^{\left\lfloor\frac{t}{2}\right\rfloor}\left(\left\lfloor\frac{t}{2}\right\rfloor\right)!}\mathbb{E}[C_j(\rho_i)^2]^{\left\lfloor\frac{t}{2}\right\rfloor} = \frac{(2\left\lfloor\frac{t}{2}\right\rfloor)!}{2^{\left\lfloor\frac{t}{2}\right\rfloor}d^{\left\lfloor\frac{t}{2}\right\rfloor}\left(\left\lfloor\frac{t}{2}\right\rfloor\right)!}. \tag{M3}$$

Combining this with Eq. (M2) leads to the desired result.

One can obtain a similar bound for $\mathbb{E}[|\partial_\theta C_j(\rho_i)^{2\left\lfloor\frac{t}{2}\right\rfloor}]$ following the same steps as the ones used to derive Eq. (L5).

### Supp. Info. N: Proof of Theorem 5

Let us here provide a proof for Theorem 5.

**Theorem 5.** *The results of Theorems 1 and 2 will hold, on average, if* $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}\operatorname{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right)$ *and* $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}\operatorname{Tr}[\rho_i\rho_{i'}] = \frac{1}{d}$*, respectively.*

*Proof.* The proof for Theorem 5 simply follows that of Theorems 1 and 2: anytime we used in these proofs the fact that $\operatorname{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right)$ or $\operatorname{Tr}[\rho_i\rho_{i'}] = \frac{1}{d}$, we replace that by $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}\operatorname{Tr}[\rho_i\rho_{i'}] \in \Omega\left(\frac{1}{\operatorname{poly}(\log(d))}\right)$ and $\mathbb{E}_{\rho_i,\rho_{i'}\sim\mathscr{D}}\operatorname{Tr}[\rho_i\rho_{i'}] = \frac{1}{d}$, respectively. The rest of the proofs are the same. □