Supplemental Information for "A Lie algebraic theory of Barren Plateaus for Deep Parameterized Quantum Circuits"

Michael Ragone,[1,*] Bojko N. Bakalov,[2,*] Frédéric Sauvage,[3] Alexander F. Kemper,[4] Carlos Ortiz Marrero,[5,6] Martín Larocca,[3,7] and M. Cerezo[8,†]

[1] *Department of Mathematics, University of California Davis, Davis, California 95616, USA*
[2] *Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695, USA*
[3] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695, USA*
[5] *AI & Data Analytics Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA*
[6] *Department of Electrical & Computer Engineering, North Carolina State University, Raleigh, North Carolina 27607, USA*
[7] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[8] *Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

This Supplemental Information contains additional details and complete proofs of our main results.

I. PROOF OF PROPOSITION 1

In this section, we give the proof of Proposition 1, which we recall for convenience of the reader:

Proposition 1. *Let O be in $i\mathfrak{g}$ such that $O=\sum_{i=1}^k O_i$ respect the decomposition Eq. (6) of $\mathfrak{g}$ into simple and abelian components. Then, for any ρ, the expectation and variance of the loss function split into sums as*

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho,O)] &= \sum_{i=1}^k \mathbb{E}_{G_i}[\ell_i(\rho,O_i)],\\ \mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho,O)] &= \sum_{i=1}^k \mathrm{Var}_{G_i}[\ell_i(\rho,O_i)],\end{aligned} \tag{1}$$

where $\ell_i(\rho,O_i)=\mathrm{Tr}\big[(1,\dots,U_i,\dots,1)\rho(1,\dots,U_i,\dots,1)^\dagger O_i\big]$ corresponds to the loss in the G_i-component of G.

While in the main text we have omitted the parameter dependence of U_i and ℓ_i for ease of notation, in the proof below we will make it explicit again. That is, we will denote $U_i := U_i(\boldsymbol{\theta}_i)$ and $\ell_i := \ell_{\boldsymbol{\theta}_i}$.

Proof. For simplicity of notation, it suffices to demonstrate the case $k=2$, as the general case is analogous. Since $O\in i\mathfrak{g}$ and $U(\boldsymbol{\theta})\in G$, we have

$$O=O_1+O_2, \quad U(\boldsymbol{\theta})=(U_1(\boldsymbol{\theta}_1),U_2(\boldsymbol{\theta}_2)).$$

Here $\boldsymbol{\theta}_i$ ($i=1,2$) are coordinates on the Lie group $G_i=e^{\mathfrak{g}_i}$, and $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ are coordinates on the Cartesian product $G=G_1\times G_2$. The loss function is then

$$\ell_{\boldsymbol{\theta}}(\rho,O)=\mathrm{Tr}\big[\rho U(\boldsymbol{\theta})^\dagger O_1 U(\boldsymbol{\theta})\big]+\mathrm{Tr}\big[\rho U(\boldsymbol{\theta})^\dagger O_2 U(\boldsymbol{\theta})\big]. \tag{2}$$

* The first two authors contributed equally.
† cerezo@lanl.gov

Notice that since $O_1 \in \mathfrak{g}_1$, it commutes with every element of G_2, and so

$$\begin{aligned}\mathrm{Tr}\big[\rho U(\boldsymbol{\theta})^\dagger O_1 U(\boldsymbol{\theta})\big] &= \mathrm{Tr}\big[\rho\,(U_1(\boldsymbol{\theta}_1), U_2(\boldsymbol{\theta}_2))^\dagger O_1 (U_1(\boldsymbol{\theta}_1), U_2(\boldsymbol{\theta}_2))\big] \\ &= \mathrm{Tr}\big[\rho\,(U_1(\boldsymbol{\theta}_1)^\dagger, U_2(\boldsymbol{\theta}_2)^\dagger)(1, U_2(\boldsymbol{\theta}_2)) O_1 (U_1(\boldsymbol{\theta}_1), 1)\big] \\ &= \mathrm{Tr}\big[\rho\,(U_1(\boldsymbol{\theta}_1), 1)^\dagger O_1 (U_1(\boldsymbol{\theta}_1), 1)\big] \\ &=: \ell_{\boldsymbol{\theta}_1}(\rho, O_1).\end{aligned}$$

Hence, the loss function is expressed as a sum of independent random variables depending on $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$:

$$\ell_{\boldsymbol{\theta}}(\rho, O) = \ell_{\boldsymbol{\theta}_1}(\rho, O_1) + \ell_{\boldsymbol{\theta}_2}(\rho, O_2). \tag{3}$$

Since the expectation (respectively, variance) of a sum of independent random variables is equal to the sum of expectations (respectively, variances), we obtain:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] &= \mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}_1}(\rho, O_1)] + \mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}_2}(\rho, O_2)], \\ \mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)) &= \mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}_1}(\rho, O_1)] + \mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}_2}(\rho, O_2)].\end{aligned} \tag{4}$$

But since we are working with products of the normalized Haar measure and $\ell_{\boldsymbol{\theta}_i}(\rho, O_i)$ is constant with respect to $\boldsymbol{\theta}_j$ for $j \neq i$, we have $\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}_i}(\rho, O)] = \mathbb{E}_{\boldsymbol{\theta}_i}[\ell_{\boldsymbol{\theta}_i}(\rho, O_i)]$, and likewise for the variance.

II. PROOF OF THEOREM 1 AND OF EQ. (28)

Let us now present the proof of Theorem 1, which we restate for convenience.

Theorem 1. *Suppose that $O \in i\mathfrak{g}$ or $\rho \in i\mathfrak{g}$, where the DLA $\mathfrak{g}$ is as in Eq. (6). Then the mean of the loss function vanishes for the semisimple component $\mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_{k-1}$ and leaves only abelian contributions:*

$$\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \mathrm{Tr}[\rho_{\mathfrak{g}_k} O_{\mathfrak{g}_k}]\,. \tag{5}$$

Conversely, the variance of the loss function vanishes for the center $\mathfrak{g}_k$ and leaves only simple contributions:

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \sum_{j=1}^{k-1} \frac{\mathcal{P}_{\mathfrak{g}_j}(\rho)\mathcal{P}_{\mathfrak{g}_j}(O)}{\dim(\mathfrak{g}_j)}\,. \tag{6}$$

Proof. By symmetry, we can assume that $O \in i\mathfrak{g}$. Due to Proposition 1 of the main text, both the expectation and variance split as sums over the components of $\mathfrak{g}$. Therefore, it is enough to consider the cases where $\mathfrak{g}$ is simple or abelian. The abelian case follows from the next lemma, which is obvious but is included here for completeness.

Lemma 1. *Let either $[O, \mathfrak{g}] = 0$ or $[\rho, \mathfrak{g}] = 0$. Then the loss function $\ell_{\boldsymbol{\theta}}(\rho, O)$ is constant. In particular, its mean and variance are given by:*

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] &= \mathrm{Tr}[\rho O]\,, \\ \mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] &= 0\,.\end{aligned} \tag{7}$$

Proof. For concreteness, suppose that $[O, \mathfrak{g}] = 0$. Then O commutes with G. Hence, for any $U(\boldsymbol{\theta}) \in G$,

$$\ell_{\boldsymbol{\theta}}(\rho, O) = \mathrm{Tr}\big[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger O\big] = \mathrm{Tr}\big[\rho U(\boldsymbol{\theta})^\dagger O U(\boldsymbol{\theta})\big] = \mathrm{Tr}[\rho O].$$

This function is constant with respect to $\boldsymbol{\theta}$ and the result follows.

From now on, let us restrict to the case where $\mathfrak{g}$ is simple. By definition, the expectation of the loss function is

$$\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \int_G d\mu(U(\boldsymbol{\theta}))\ \mathrm{Tr}\big[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger O\big], \tag{8}$$

where the integral is taken with respect to the Haar measure on G. Using the linearity of the trace, we can rewrite this as

$$\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \mathrm{Tr}\Big[\mathcal{M}_G^{(1)}(\rho) O\Big] = \Big\langle \mathcal{M}_G^{(1)}(\rho), O \Big\rangle, \tag{9}$$

where the first moment operator $\mathcal{M}_G^{(1)}$ is defined by (cf. Eq. (19) of the main text):

$$\mathcal{M}_G^{(1)}(\rho) = \int_G d\mu(U(\boldsymbol{\theta}))\, U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger. \tag{10}$$

On the right side of Eq. (9), we have expressed the trace in terms of the Hilbert–Schmidt inner product $\langle A, B\rangle = \mathrm{Tr}\big[A^\dagger B\big]$, using that $\mathcal{M}_G^{(1)}(\rho)$ is Hermitian. If we use the cyclicity of trace in Eq. (8) to replace $\mathrm{Tr}\big[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger O\big]$ with $\mathrm{Tr}\big[\rho U(\boldsymbol{\theta})^\dagger O U(\boldsymbol{\theta})\big]$ in Eq. (8) and the invariance of the Haar measure under $U \mapsto U^\dagger$, we can also derive

$$\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \mathrm{Tr}\Big[\rho \mathcal{M}_G^{(1)}(O)\Big] = \Big\langle \rho, \mathcal{M}_G^{(1)}(O) \Big\rangle. \tag{11}$$

More generally, let us recall the t-moment operator (cf. Eq. (19)):

$$\mathcal{M}_G^{(t)}(A) = \int_G d\mu(U(\boldsymbol{\theta}))\, U(\boldsymbol{\theta})^{\otimes t} A (U(\boldsymbol{\theta})^\dagger)^{\otimes t}. \tag{12}$$

Here A is any linear operator acting on the t-th tensor power of the Hilbert space $\mathcal{H}^{\otimes t}$. Below, we will apply it in the particular case $t = 2$ and $A = O \otimes O$. A similar calculation as above, using the linearity and cyclicity of the trace, shows that the linear operator $\mathcal{M}_G^{(t)}$ is self-adjoint with respect to the Hilbert–Schmidt inner product on $\mathfrak{gl}(\mathcal{H}^{\otimes t})$:

$$\begin{aligned}
\Big\langle \mathcal{M}_G^{(t)}(A), B \Big\rangle &= \int_G d\mu(U(\boldsymbol{\theta}))\ \mathrm{Tr}\big[U(\boldsymbol{\theta})^{\otimes t} A^\dagger (U(\boldsymbol{\theta})^\dagger)^{\otimes t} B\big] \\
&= \int_G d\mu(U(\boldsymbol{\theta}))\ \mathrm{Tr}\big[A^\dagger (U(\boldsymbol{\theta})^\dagger)^{\otimes t} B U(\boldsymbol{\theta})^{\otimes t}\big] \\
&= \Big\langle A, \mathcal{M}_G^{(t)}(B) \Big\rangle.
\end{aligned} \tag{13}$$

The next lemma, which follows from the left-invariance of the Haar measure, is the key ingredient in the proof of the theorem. The result of the lemma is well known and forms the foundation for the Weingarten Calculus (see Ref. [1] and the references therein).

Lemma 2. *The t-moment operator is an orthogonal projection from the space of linear operators $\mathfrak{gl}(\mathcal{H}^{\otimes t})$ onto its subspace $\mathfrak{gl}(\mathcal{H}^{\otimes t})^G$ of G-invariants, where G acts on $\mathfrak{gl}(\mathcal{H}^{\otimes t})$ by conjugation: $A \mapsto g^{\otimes t} A (g^\dagger)^{\otimes t}$ for $g \in G$.*

Proof. By the left-invariance of the Haar measure, we have:

$$g^{\otimes t}\mathcal{M}_G^{(t)}(A)(g^\dagger)^{\otimes t} = \int_G d\mu(U(\boldsymbol{\theta}))\, (gU(\boldsymbol{\theta}))^{\otimes t} A (g(U(\boldsymbol{\theta}))^\dagger)^{\otimes t} = \mathcal{M}_G^{(t)}(A). \tag{14}$$

Hence, $\mathcal{M}_G^{(t)}(A)$ is G-invariant, i.e., fixed by every $g \in G$. Conversely, if A is G-invariant, then it commutes with the action of G, and we have $\mathcal{M}_G^{(t)}(A) = A$ as in the proof of Lemma 1. Therefore, $\mathcal{M}_G^{(t)}$ is a projection. Since $\mathcal{M}_G^{(t)}$ is self-adjoint, it is an orthogonal projection.

The Haar measure on a compact Lie group G is both left and right invariant. Similarly to Eq. (14), we also have

$$\mathcal{M}_G^{(t)}\left(g^{\otimes t} A (g^\dagger)^{\otimes t}\right) = \int_G d\mu(U(\boldsymbol{\theta}))\, (U(\boldsymbol{\theta})g)^{\otimes t} A ((U(\boldsymbol{\theta})g)^\dagger)^{\otimes t} = \mathcal{M}_G^{(t)}(A). \tag{15}$$

This means that if we view $\mathcal{M}_G^{(t)}$ as a linear transformation

$$\mathcal{M}_G^{(t)} \colon \mathfrak{gl}(\mathcal{H}^{\otimes t}) \to \mathfrak{gl}(\mathcal{H}^{\otimes t})^G, \tag{16}$$

then it is a homomorphism of representations of G, i.e., it commutes with the G-actions on both sides of the equation. Moreover, due to the unitarity and complete reducibility of the representation of G on $\mathfrak{gl}(\mathcal{H}^{\otimes t})$, the latter splits as a direct sum of $\mathfrak{gl}(\mathcal{H}^{\otimes t})^G$ and its orthogonal complement.

Now let us get back to the proof of the theorem in the case when $\mathfrak{g}$ is simple and $O \in i\mathfrak{g}$. Our construction of the DLA $\mathfrak{g}$ as a subalgebra of $\mathfrak{gl}(\mathcal{H})$ is justified by the fact that any nontrivial representation of a simple Lie algebra is faithful. Moreover, we are assuming that the Lie group $G = e^{\mathfrak{g}}$ has a unitary representation on $\mathcal{H}$, which means that $\mathfrak{g} \subseteq \mathfrak{u}(\mathcal{H})$. As $\mathfrak{g}$ is a subalgebra of $\mathfrak{gl}(\mathcal{H})$, it is also a subrepresentation for the adjoint action of G. Then the restriction of $\mathcal{M}_G^{(1)}$ to $\mathfrak{g}$ gives a homomorphism

$$\mathcal{M}_G^{(1)} \colon \mathfrak{g} \to \mathfrak{g}^G. \tag{17}$$

But $\mathfrak{g}^G$ is the center of $\mathfrak{g}$, which is trivial. Therefore, $\mathcal{M}_G^{(1)}(O) = 0$, proving that the expectation of the loss function is zero.

Now we find the variance of the loss function. Using that $\mathrm{Tr}[A \otimes B] = \mathrm{Tr}[A]\,\mathrm{Tr}[B]$, we have:

$$\begin{aligned}
\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] &= \mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)^2] \\
&= \int_G d\mu(U(\boldsymbol{\theta}))\, \left(\mathrm{Tr}\big[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger O\big]\right)^2 \\
&= \int_G d\mu(U(\boldsymbol{\theta}))\, \mathrm{Tr}\big[U(\boldsymbol{\theta})^{\otimes 2} \rho^{\otimes 2} (U(\boldsymbol{\theta})^\dagger)^{\otimes 2} O^{\otimes 2}\big] \\
&= \mathrm{Tr}\Big[\mathcal{M}^{(2)}(\rho^{\otimes 2}) O^{\otimes 2}\Big].
\end{aligned} \tag{18}$$

As above, we can express the last equation in terms of the Hilbert–Schmidt inner product:

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \Big\langle \mathcal{M}^{(2)}(\rho^{\otimes 2}), O^{\otimes 2} \Big\rangle = \Big\langle \rho^{\otimes 2}, \mathcal{M}^{(2)}(O^{\otimes 2}) \Big\rangle. \tag{19}$$

Recall that $O \in i\mathfrak{g}$, so $O^{\otimes 2} \in \mathfrak{g} \otimes \mathfrak{g}$. Since $\mathfrak{g} \subset \mathfrak{gl}(\mathcal{H})$, we have $\mathfrak{g} \otimes \mathfrak{g} \subset \mathfrak{gl}(\mathcal{H}) \otimes \mathfrak{gl}(\mathcal{H}) \cong \mathfrak{gl}(\mathcal{H}^{\otimes 2})$ is a subrepresentation for the adjoint action of G. As above, the restriction of the homomorphism $\mathcal{M}_G^{(2)}$ to $\mathfrak{g}^{\otimes 2}$ gives a homomorphism

$$\mathcal{M}_G^{(2)} \colon \mathfrak{g} \otimes \mathfrak{g} \to (\mathfrak{g} \otimes \mathfrak{g})^G, \tag{20}$$

which is an orthogonal projection. The following lemma is well known (see, e.g., [2], Proposition 6.54 and the discussion after it).

Lemma 3. *Let $\mathfrak{g}$ be a simple subalgebra of $\mathfrak{u}(\mathcal{H}) \cong \mathfrak{u}(N)$, where $\mathcal{H} \cong \mathbb{C}^N$ is an N-dimensional Hilbert space. Then the space of G-invariants $(\mathfrak{g} \otimes \mathfrak{g})^G$ is 1-dimensional and spanned by the split quadratic Casimir element*

$$C = \sum_{j=1}^{\dim(\mathfrak{g})} B_j \otimes B_j, \tag{21}$$

where $\{B_j\}_{j=1}^{\dim(\mathfrak{g})}$ is an orthonormal basis for $i\mathfrak{g}$ with respect to the Hilbert–Schmidt inner product.

Proof. As $\mathfrak{g} \subseteq \mathfrak{u}(N)$ and is simple, it is a compact real Lie algebra (i.e., the Lie group $G \subseteq \mathbb{U}(N)$ is compact); hence, its complexification $\mathfrak{g}_\mathbb{C}$ is a simple Lie algebra whenever $\mathfrak{g}$ is simple (see e.g. [3], Chapters IV and VI).

It is well known that for any simple Lie algebra $\mathfrak{g}_\mathbb{C}$ over $\mathbb{C}$, the space of G-invariants $(\mathfrak{g}_\mathbb{C} \otimes \mathfrak{g}_\mathbb{C})^G = \mathbb{C}C$, where C is given by Eq. (21) and $\{B_j\}_{j=1}^{\dim(\mathfrak{g})}$ is an orthonormal basis for $\mathfrak{g}_\mathbb{C}$ with respect to some nondegenerate symmetric G-invariant bilinear form $(\cdot,\cdot)$. Moreover, C does not depend on the choice of basis for $\mathfrak{g}_\mathbb{C}$. These claims follow from Schur's Lemma, because the adjoint action of $\mathfrak{g}_\mathbb{C}$ on itself is irreducible and after we identify $\mathfrak{g}_\mathbb{C} \otimes \mathfrak{g}_\mathbb{C} \cong \mathfrak{gl}(\mathfrak{g}_\mathbb{C})$ as G-representations, the G-invariant elements correspond to homomorphisms $\mathfrak{g}_\mathbb{C} \to \mathfrak{g}_\mathbb{C}$, which are only scalar multiples of the identity.

We take $(A, B) = \mathrm{Tr}[AB]$ to be the trace form (and again by Schur's Lemma, any two such forms are scalar multiples of each other). Now the proof follows from the observation that $\langle A, B\rangle = \mathrm{Tr}[AB]$ for all $A, B \in i\mathfrak{g}$.

Using that $(\mathfrak{g} \otimes \mathfrak{g})^G = \mathbb{R}C$, we can compute $\mathcal{M}^{(2)}(O^{\otimes 2})$ by taking the orthogonal projection of $O^{\otimes 2} \in \mathfrak{g} \otimes \mathfrak{g}$ onto C:

$$\mathcal{M}^{(2)}(O^{\otimes 2}) = \frac{\langle C, O^{\otimes 2}\rangle}{\langle C, C\rangle} C. \tag{22}$$

Combining this with Eq. (19), we get:

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{\langle \rho^{\otimes 2}, C\rangle \langle C, O^{\otimes 2}\rangle}{\langle C, C\rangle} = \frac{\langle C, \rho^{\otimes 2}\rangle \langle C, O^{\otimes 2}\rangle}{\langle C, C\rangle}\,, \tag{23}$$

where for the last equality we used that all elements of $\mathfrak{g} \otimes \mathfrak{g}$ are Hermitian.

Now we compute each of the inner products in the above formula. For the denominator, we have

$$\langle C, C\rangle = \sum_{i,j=1}^{\dim(\mathfrak{g})} \langle B_i \otimes B_i, B_j \otimes B_j\rangle = \sum_{i,j=1}^{\dim(\mathfrak{g})} \langle B_i, B_j\rangle \langle B_i, B_j\rangle = \dim(\mathfrak{g}). \tag{24}$$

For the other two terms, we apply the next lemma, which follows from simple linear algebra.

Lemma 4. *For any Hermitian operator $H \in i\mathfrak{u}(\mathcal{H})$, we have that*

$$\langle C, H^{\otimes 2}\rangle = \mathcal{P}_\mathfrak{g}(H) \tag{25}$$

is the $\mathfrak{g}$-purity of H as defined in Eq. (7) of the main text.

Proof. The orthogonal projection of H into $\mathfrak{g}_\mathbb{C}$ is given by

$$H_\mathfrak{g} = \sum_{j=1}^{\dim(\mathfrak{g})} \langle B_j, H\rangle B_j, \tag{26}$$

where $\{B_j\}_{j=1}^{\dim(\mathfrak{g})}$ is any orthonormal basis for $\mathfrak{g}_\mathbb{C}$ with respect to the Hilbert–Schmidt inner product $\langle\cdot,\cdot\rangle$. Then

$$\langle H_\mathfrak{g}, H_\mathfrak{g}\rangle = \sum_{j=1}^{\dim(\mathfrak{g})} |\langle B_j, H\rangle|^2. \tag{27}$$

Since the restriction of $\langle\cdot,\cdot\rangle$ to the space $i\mathfrak{u}(\mathcal{H})$ of Hermitian operators is positive definite, we can pick the basis vectors B_j to be all in $i\mathfrak{g}$. This gives that $H_\mathfrak{g} \in i\mathfrak{g}$. Hence, $\langle H_\mathfrak{g}, H_\mathfrak{g}\rangle = \mathrm{Tr}\big[H_\mathfrak{g}^2\big]$, which proves the equality in Eq. (7).

On the other hand, when all $B_j \in i\mathfrak{g}$, we have

$$\langle C, H^{\otimes 2}\rangle = \sum_{j=1}^{\dim(\mathfrak{g})} \langle B_j \otimes B_j, H^{\otimes 2}\rangle = \sum_{j=1}^{\dim(\mathfrak{g})} \langle B_j, H\rangle^2 = \langle H_\mathfrak{g}, H_\mathfrak{g}\rangle\,,$$

because in this case $\langle B_j, H\rangle \in \mathbb{R}$.

Plugging the result of the above lemma in Eq. (23) completes the proof of Theorem 1.

We end this section with a proof of Eq. (28). Let us pick an orthonormal basis $\{B_j\}_{j=1}^{\dim(\mathfrak{g})}$ for $i\mathfrak{g}$, so that it contains as a subset the basis $\{H_j\}_{j=1}^{\dim(\mathfrak{h})}$ for the Cartan subalgebra $i\mathfrak{h}$ and the B_j that are not in $i\mathfrak{h}$ are complex linear combinations of raising and lowering operators (i.e., root vectors of $\mathfrak{g}_{\mathbb{C}}$ with respect to its Cartan subalgebra $\mathfrak{h}_{\mathbb{C}}$). Hence, such $B_j \notin i\mathfrak{h}$ are sums of strictly triangular matrices and have zero diagonal with respect to the basis $\{|v\rangle\}$. On the other hand, by assumption, ρ is diagonal in this basis; therefore, $\mathrm{Tr}\Big[B_j^\dagger \rho\Big] = 0$ for $B_j \notin i\mathfrak{h}$. Then Eq. (26) reduces to a sum over the basis of $i\mathfrak{h}$:

$$\rho_{\mathfrak{g}} = \sum_{j=1}^{\dim(\mathfrak{h})} \mathrm{Tr}\Big[H_j^\dagger \rho\Big] H_j = \rho_{\mathfrak{h}}. \tag{28}$$

Now recall that $H_j^\dagger = H_j$, and compute using Eqs. (26), (27):

$$H_j \rho = \sum_v r_v H_j |v\rangle \langle v| = \sum_v r_v \lambda_v(H_j) |v\rangle \langle v| , \tag{29}$$

from where

$$\mathrm{Tr}[H_j \rho] = \sum_v r_v \lambda_v(H_j) = \lambda_\rho(H_j). \tag{30}$$

Plugging this into Eq. (28) proves Eq. (28).

III. PROOF OF COROLLARY 1

Here we present the derivation of Corollary 1, which we now recall.

Corollary 1. *Let* $\|O\|_2^2 \leqslant 2^n$. *If either* $\dim(\mathfrak{g})$, $1/\mathcal{P}_{\mathfrak{g}}(\rho)$ *or* $1/\mathcal{P}_{\mathfrak{g}}(O)$ *is in* $\Omega(b^n)$ *with* $b > 2$, *then the loss has a BP.*

Proof. Let us first consider the case where $\dim(\mathfrak{g}) \in \Omega(b^n)$. Then, irrespective of whether O or ρ are in $i\mathfrak{g}$, we will have $\mathcal{P}_{\mathfrak{g}}(O) \leqslant \mathrm{Tr}\big[O^2\big] = \|O\|_2^2 \leqslant 2^n$ and $\mathcal{P}_{\mathfrak{g}}(\rho) \leqslant 1$. Hence,

$$\begin{aligned} \mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] &= \frac{1}{\dim(\mathfrak{g})} \mathrm{Tr}\big[O_{\mathfrak{g}}^2\big] \mathrm{Tr}\big[\rho_{\mathfrak{g}}^2\big] \\ &\leqslant \frac{1}{\dim(\mathfrak{g})} 2^n \in \mathcal{O}\left(\frac{1}{c^n}\right) , \end{aligned} \tag{31}$$

with $c = b/2 > 1$, and hence the loss has a BP.

Second, let us analyze the case where $1/\mathcal{P}_{\mathfrak{g}}(\rho)$ is in $\Omega(b^n)$. Irrespective of the size of the DLA, we will have $\frac{\mathrm{Tr}[O_{\mathfrak{g}}^2]}{\dim(\mathfrak{g})} \leqslant 2^n$. Thus, we have $\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}(\frac{1}{c^n})$ with $c = b/2 > 1$.

Finally, if $1/\mathcal{P}_{\mathfrak{g}}(O)$ is in $\Omega(b^n)$, we have $\frac{\mathrm{Tr}[\rho_{\mathfrak{g}}^2]}{\dim(\mathfrak{g})} \leqslant 1$, thus $\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}(\frac{1}{b^n})$ with $b > 2$.

IV. PROOF OF THEOREM 2

In this section, we present a derivation of Theorem 2. We will frequently be working with the Schatten p-norms on the spaces of operators $\mathfrak{gl}(\mathcal{H}^{\otimes t})$, where $p = 1, \infty$. Both are spectral norms: given the singular values $s_i(A) \geqslant 0$ of A, we may express these norms as

$$\|A\|_1 = \sum_i s_i(A), \qquad \|A\|_\infty = \max_i s_i(A). \tag{32}$$

This induces corresponding operator norms for linear maps $\mathcal{M}\colon \mathfrak{gl}(\mathcal{H}^{\otimes t}) \to \mathfrak{gl}(\mathcal{H}^{\otimes t})$ by the usual

$$\|\mathcal{M}\|_p = \sup_{\substack{A\neq 0 \\ A\in\mathfrak{gl}(\mathcal{H}^{\otimes t})}} \frac{\|\mathcal{M}(A)\|_p}{\|A\|_p}. \tag{33}$$

Theorem 2 is the $t = 2$ case of the following theorem:

Theorem 2. *The ensemble of unitaries $\mathcal{E}_L \subseteq G$ generated by an L-layered circuit $U(\boldsymbol{\theta})$ will form an ϵ-approximate G t-design, when the number of layers L is*

$$L \geqslant \frac{\log(1/\epsilon)}{\log\left(1/\left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty\right)}, \tag{34}$$

where $\mathcal{E}_1$ is the ensemble generated by a single layer $U_1(\boldsymbol{\theta}_1)$.

Note that to the authors' knowledge, this choice of L need not be uniform in t.

Proof. First, recall Eq. (12), the definition of the t-moment operator $\mathcal{M}^{(t)}_{\mathcal{E}_L}\colon \mathfrak{gl}(\mathcal{H}^{\otimes t}) \to \mathfrak{gl}(\mathcal{H}^{\otimes t})$ corresponding to an ensemble $\mathcal{E}_L$

$$\mathcal{M}^{(t)}_{\mathcal{E}_L}(A) = \int_{\mathcal{E}_L} d\mu(U)\, U^{\otimes t} A (U^\dagger)^{\otimes t}, \qquad A \in \mathfrak{gl}(\mathcal{H}^{\otimes t}). \tag{35}$$

Also recall that $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ is self-adjoint with respect to the Hilbert–Schmidt inner product on $\mathfrak{gl}(\mathcal{H}^{\otimes t})$, and so $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ is diagonalizable on $\mathfrak{gl}(\mathcal{H}^{\otimes t})$ with real eigenvalues.

Since $\mathcal{E}_L \subset G$, all elements of $\mathcal{E}_L$ fix the G-invariants $\mathfrak{gl}(\mathcal{H}^{\otimes t})^G$. Hence, as in the proofs of Lemmas 1 and 2, the operator $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ acts as the identity on $\mathfrak{gl}(\mathcal{H}^{\otimes t})^G$. Since it is self-adjoint, the orthogonal complement $\mathfrak{p}$ is also an invariant subspace, and so we may write the orthogonal direct sum

$$\mathfrak{gl}(\mathcal{H}^{\otimes t}) = \mathfrak{gl}(\mathcal{H}^{\otimes t})^G \oplus \mathfrak{p}. \tag{36}$$

Recall from Lemma 2 that $\mathcal{M}^{(t)}_G$ is the orthogonal projection onto $\mathfrak{gl}(\mathcal{H}^{\otimes t})^G$. Thus, $\mathcal{A}^{(t)}_{\mathcal{E}_L} = \mathcal{M}^{(t)}_{\mathcal{E}_L} - \mathcal{M}^{(t)}_G$ acts as zero on $\mathfrak{gl}(\mathcal{H}^{\otimes t})^G$ and as $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ on $\mathfrak{p}$.

Now we investigate the eigenvalues of $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ on $\mathfrak{p}$.

Lemma 5. *Every eigenvalue λ of $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ has $|\lambda| \leqslant 1$.*

Proof. Since $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ is self-adjoint, its singular values are equal to the absolute values of its eigenvalues, and so every eigenvalue has $|\lambda| \leqslant \left\|\mathcal{M}^{(t)}_{\mathcal{E}_L}\right\|_\infty$ by Eq. (32). Let $A \in \mathfrak{gl}(\mathcal{H}^{\otimes t})$ have $\|A\|_\infty = 1$. Then we have

$$\left\|\mathcal{M}^{(t)}_{\mathcal{E}_L}(A)\right\|_\infty \leqslant \int_{\mathcal{E}_L} d\mu(U)\, \left\|U^{\otimes t} A (U^\dagger)^{\otimes t}\right\|_\infty = \int_{\mathcal{E}_L} d\mu(U)\, 1 = 1,$$

where we used that the Schatten norms are unitarily invariant.

We want to show that all eigenvalues λ of $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ on $\mathfrak{p}$ have $|\lambda| < 1$. We prove that first in the case of a single layer.

Lemma 6. *The single-layer ensemble $\mathcal{E}_1 := \{U_1(\boldsymbol{\theta}_1) : \boldsymbol{\theta}_1 \in \mathbb{R}^K\}$ with measure $d\mu_1$ satisfies $|\lambda| < 1$ for all eigenvalues λ of $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ on $\mathfrak{p}$.*

Proof. Let $A \in \mathfrak{p}$ with $\|A\|_\infty = 1$. Since A is not G-invariant, there exists a $\tilde{U} \in G$ such that for some $\epsilon > 0$,

$$\left|\left\langle A, \tilde{U}^{\otimes t} A (\tilde{U}^\dagger)^{\otimes t} \right\rangle\right| < 1 - \epsilon.$$

Note that Corollary 3.2.6 of [4] assures us that since G compact, there exists a uniform choice of depth $\tilde{L}$ such that any $\tilde{U} \in G$ can be written as

$$\tilde{U} = U_{\tilde{L}}(\tilde{\boldsymbol{\theta}}_{\tilde{L}}) U_{\tilde{L}-1}(\tilde{\boldsymbol{\theta}}_{\tilde{L}-1}) \cdots U_1(\tilde{\boldsymbol{\theta}}_1). \tag{37}$$

Use $(\tilde{\mathcal{E}_L}, d\tilde{\mu})$ to denote the ensemble generated by the depth $\tilde{L}$ circuit. Since the map $\tilde{U} \mapsto \left\langle A, \tilde{U}^{\otimes t}(A)(\tilde{U}^\dagger)^{\otimes t} \right\rangle$ is a smooth map from $G \to \mathbb{C}$, there exists an open neighborhood $N \subseteq G$ of $\tilde{U}$, on which

$$\left|\left\langle A, V^{\otimes t} A (V^\dagger)^{\otimes t} \right\rangle\right| < 1 - \frac{\epsilon}{2} \qquad \text{for all } \ A \in N.$$

As the choice of depth $\tilde{L}$ is uniform for all elements in G, Eq. (37) assures us that $N \cap \operatorname{supp}(d\tilde{\mu})$ is not measure 0. This then means that

$$\left|\left\langle A, \mathcal{M}^{(t)}_{\tilde{\mathcal{E}_L}}(A) \right\rangle\right| < 1.$$

Thus any eigenvalue λ of $\mathcal{M}^{(t)}_{\tilde{\mathcal{E}_L}}\big|_{\mathfrak{p}}$ must have $|\lambda| < 1$. But a straightforward calculation reveals that $\mathcal{M}^{(t)}_{\tilde{\mathcal{E}_L}} = (\mathcal{M}^{(t)}_{U_1})^{\tilde{L}}$, and so any eigenvalue λ of $\mathcal{M}^{(t)}_{U_1}\big|_{\mathfrak{p}}$ has $|\lambda| < 1$.

Now to finish the proof of the theorem, we combine the results of Lemmas 5 and 6. It follows that for the single layer ensemble $\mathcal{E}_1$, the operator $\mathcal{A}^{(t)}_{\mathcal{E}_1} = \mathcal{M}^{(t)}_G - \mathcal{M}^{(t)}_{\mathcal{E}_1}$ is diagonalizable with all eigenvalues λ satisfying $|\lambda| < 1$. Therefore, its operator norm $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty < 1$. For a circuit of depth L, we may use the following Lemma 7 to see that $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|_\infty = \left(\left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty\right)^L$, so we can pick L as in Eq. (34) to grant our result.

Lemma 7. *Let $\mathcal{E}_L \subseteq G$ be the ensemble of unitaries generated by an L-layered circuit $U(\boldsymbol{\theta})$. Then*

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} = \underbrace{\mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \cdots \circ \mathcal{A}^{(t)}_{\mathcal{E}_1}}_{L \ \text{times}}. \tag{38}$$

In particular, $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|_\infty = \left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty^L$.

Proof. Let us more explicitly write $\mathcal{A}^{(t)}_{\mathcal{E}_L}$:

$$\mathcal{A}^{(t)}_{\mathcal{E}_L}(\cdot) = \mathcal{M}^{(t)}_{\mathcal{E}_L}(\cdot) - \mathcal{M}^{(t)}_G(\cdot) = \int_{\mathcal{E}_L} d\mu(U(\boldsymbol{\theta}))\, U(\boldsymbol{\theta})^{\otimes t}(\cdot)(U(\boldsymbol{\theta})^\dagger)^{\otimes t} - \int_G d\mu(U(\boldsymbol{\theta}))\, U(\boldsymbol{\theta})^{\otimes t}(\cdot)(U(\boldsymbol{\theta})^\dagger)^{\otimes t}. \tag{39}$$

Taking the composition of two such linear maps leads to

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} \circ \mathcal{A}^{(t)}_{\mathcal{E}_L} = \mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_{\mathcal{E}_L} + \mathcal{M}^{(t)}_G \circ \mathcal{M}^{(t)}_G - \mathcal{M}^{(t)}_G \circ \mathcal{M}^{(t)}_{\mathcal{E}_L} - \mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_G. \tag{40}$$

Then, the following hold

$$\mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_{\mathcal{E}_L} = \mathcal{M}^{(t)}_{\mathcal{E}_{2L}}, \qquad \mathcal{M}^{(t)}_G \circ \mathcal{M}^{(t)}_G = \mathcal{M}^{(t)}_G \circ \mathcal{M}^{(t)}_{\mathcal{E}_L} = \mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_G = \mathcal{M}^{(t)}_G. \tag{41}$$

The first equality simply states that the multiplication of two distributions from L-layered circuits is equal to the distribution of a $2L$-layered circuit, while the latter ones follows from the left- and right-invariance of the Haar measure. Hence, we have

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} \circ \mathcal{A}^{(t)}_{\mathcal{E}_L} = \mathcal{A}^{(t)}_{\mathcal{E}_{2L}} \,. \tag{42}$$

In particular, we obtain

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} = \underbrace{\mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \cdots \circ \mathcal{A}^{(t)}_{\mathcal{E}_1}}_{L \text{ times}} , \tag{43}$$

where $\mathcal{A}^{(t)}_{\mathcal{E}_1}$ corresponds to the single layer ensemble $\mathcal{E}_1$.

The final claim that $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|_\infty = \left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty^L$ follows by noting that $\mathcal{A}^{(t)}_{\mathcal{E}_1}$ is self-adjoint, and so its singular values are exactly the absolute values of its eigenvalues. Take the largest singular value $|\lambda|$ and a corresponding eigenvector A of $\mathcal{A}^{(t)}_{\mathcal{E}_1}$. Then by Eq. (43), A is an eigenvector of $\mathcal{A}^{(t)}_{\mathcal{E}_L}$ of largest singular value $|\lambda|^L$.

A. Example of a use-case of Theorem 2

In this section we present an example where we can use Theorem 2 to predict the number of layers needed for an L-layered circuit to form an ϵ-approximate 2-design.

Let S be a subgroup of $\mathbb{U}(d)$, the unitary group of degree $d = 2^n$ acting on $\mathcal{H}$, and let $\{P_j\}_{j=1}^D$ be a basis for the commutant of the t-fold tensor representation $\mathcal{H}^{\otimes t}$ of S. That is, $[P_j, U^{\otimes t}] = 0$ for all $U \in S$. Then, we know that (see, e.g., [5]):

$$\mathcal{M}^{(t)}_S = \int_S d\mu(U)\, (U \otimes U^*)^{\otimes t} = \sum_{i,j=1}^{D} W^{-1}_{ij} |P_i\rangle\rangle\langle\langle P_j| \,. \tag{44}$$

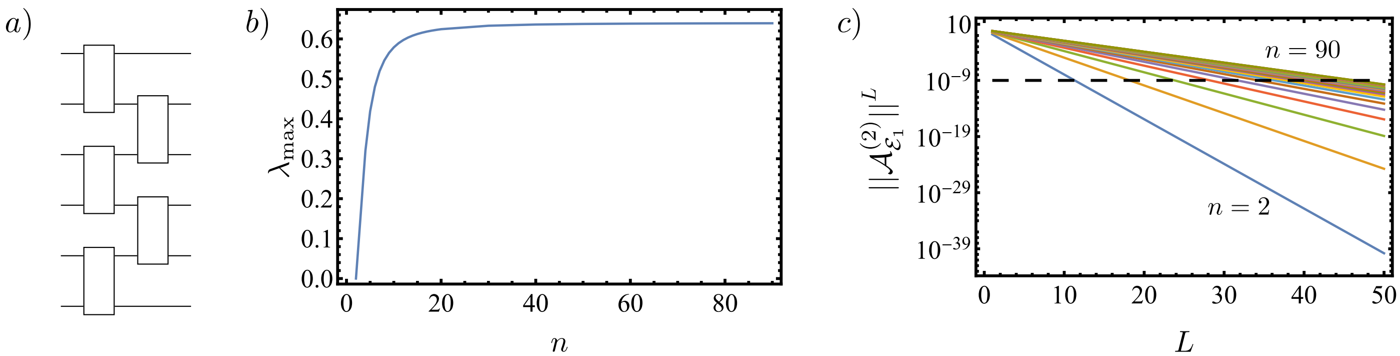
Here W^{-1} is the so-called Weingarten matrix, the inverse of the Gram matrix W whose entries $W_{ij} = \mathrm{Tr}\left[P_i^\dagger P_j\right]$ are given by the Hilbert–Schmidt inner product of the commutant's basis elements. Moreover, we have denoted as $|A\rangle\rangle$ the standard vectorization of a linear operator A. Namely, given some $A \in \mathcal{L}(\mathcal{H})$ such that $A = \sum_{\mu,\nu=1}^{2^n} A_{\mu\nu} |\mu\rangle\langle\nu|$, then $|A\rangle\rangle = \sum_{\mu,\nu=1}^{2^n} A_{\mu\nu} |\mu\rangle \otimes |\nu\rangle$. Hence, if we can characterize the basis of the commutant of the t-fold tensor representation of S, then we can find a closed form expression for $\mathcal{M}^{(t)}_S$. Finally, note that $\mathcal{M}^{(t)}_S$ is an operator acting on $2t$ copies of the Hilbert space $\mathcal{H}$, and hence is of dimension $2^{2nt} \times 2^{2nt}$.

Next, let us recall from the main text that we have factorized the parametrized quantum circuit into layers as $U(\boldsymbol{\theta}) = U_L(\boldsymbol{\theta}_L) U_{L-1}(\boldsymbol{\theta}_{L-1}) \cdots U_1(\boldsymbol{\theta}_1)$, and that each layer $U_l(\boldsymbol{\theta}_l)$ is given by $U_l(\boldsymbol{\theta}_l) = \prod_{k=1}^K U_k(\theta_{l,k})$. Hence, if the parameters in each U_k are independently sampled, it is not hard to see that the t-th moment operator of a single layer can be expressed as the product of the t-th moment operators of each U_k. That is,

$$\mathcal{M}^{(t)}_{\mathcal{E}_1} = \prod_{k=1}^{K} \mathcal{M}^{(t)}_{\mathcal{E}_{1,k}} \,, \tag{45}$$

where we have denoted as $\mathcal{E}_{1,k}$ the distribution of unitaries obtained from each $U_k(\theta_{l,k})$. As such, if we assume that each $\mathcal{E}_{1,k}$ forms a subgroup, and that we sample it over its respective Haar measure, we will have via Eq. (44) that

$$\mathcal{M}^{(t)}_{\mathcal{E}_1} = \sum_{i_1,j_1=1}^{D_1} \cdots \sum_{i_K,j_K=1}^{D_K} \left((W_1^{-1})_{i_1 j_1} \cdots (W_K^{-1})_{i_K j_K} \langle\langle P_{j_1} | P_{i_2} \rangle\rangle \cdots \langle\langle P_{j_{K-1}} | P_{i_K} \rangle\rangle \right) |P_{i_1}\rangle\rangle\langle\langle P_{j_K}| \,. \tag{46}$$



Supp. Fig. 1. **Layers needed to become an ϵ-approximate 2-design.** (a) Hardware efficient ansatz where two-qubit gates act in a brick-like fashion on neighboring pairs of qubits. We assume that each gate is independently and randomly sampled from $\mathbb{SU}(4)$. (b) Largest eigenvalue of $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ as a function of the number of qubits. (c) Each curve shows $(\mathcal{A}^{(2)}_{\mathcal{E}_1})^L$ for different values of n. The thick horizontal dashed line is plotted at 10^{-9}. Vertical axis is shown in a logarithmic scale.

Here, we denoted as $\{P_{j_k}\}_{j_k=1}^{D_k}$ be a basis for the commutant of the t-fold tensor representation of $\mathcal{E}_{1,k}$, and as W_k^{-1} its associated Weingarten matrix. Moreover, we recall that $\langle\langle A|B\rangle\rangle = \text{Tr}\big[A^\dagger B\big]$. Equation (46) indicates that we can compute the single layer operator $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ by decomposing the layer into unitaries whose ensembles form groups. In its lower form, such procedure can always be taken to the single-gate level where each $U_k(\theta_{l,k}) = e^{-iH_k\theta_{l,k}}$ with $H_k^2 = \mathbb{1}$, and hence where its associated distribution will be a representation of $\mathbb{U}(1)$.

For instance, let us consider a hardware efficient circuit composed of alternating layers of two-qubit gates acting on neighboring qubits (see Fig. 1(a)). Moreover, let us assume that each two-qubit gate forms an independent local 2-design over $\mathbb{SU}(4)$. If we consider a gate acting on qubits i and $i+1$, we will have [5]

$$\mathcal{M}^{(2)}_{\mathbb{SU}(4)} = \frac{1}{15}\left(|\mathbb{I}_i\mathbb{I}_{i+1}\rangle\rangle\langle\langle\mathbb{I}_i\mathbb{I}_{i+1}| + |\mathbb{S}_i\mathbb{S}_{i+1}\rangle\rangle\langle\langle\mathbb{S}_i\mathbb{S}_{i+1}| - \frac{1}{4}\left(\mathbb{S}_i\mathbb{S}_{i+1}| + |\mathbb{S}_i\mathbb{S}_{i+1}\rangle\rangle\langle\langle\mathbb{I}_i\mathbb{I}_{i+1}|\right)\right) . \tag{47}$$

where $\mathbb{I}_i$ and $\mathbb{S}_i$ denote the identity and SWAP operators acting on the two copies of the i-th qubit Hilbert spaces. By defining a basis for the operator space spanned from the non-orthogonal elements

$$\big\{|\mathbb{I}_i\mathbb{I}_{i+1}\rangle\rangle, |\mathbb{I}_i\mathbb{S}_{i+1}\rangle\rangle, |\mathbb{S}_i\mathbb{I}_{i+1}\rangle\rangle, |\mathbb{S}_i\mathbb{S}_{i+1}\rangle\rangle\big\}, \tag{48}$$

we can express $\mathcal{M}^{(2)}_{\mathbb{SU}(4)}$ as a 4×4 matrix

$$\mathcal{M}^{(2)}_{\mathbb{SU}(4)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2/5 & 0 & 0 & 2/5 \\ 2/5 & 0 & 0 & 2/5 \\ 2/5 & 0 & 0 & 2/5 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{49}$$

One can readily verify that $\mathcal{M}^{(2)}_{\mathbb{SU}(4)} \circ \mathcal{M}^{(2)}_{\mathbb{SU}(4)} = \mathcal{M}^{(2)}_{\mathbb{SU}(4)}$ and that $\text{Tr}\Big[\mathcal{M}^{(2)}_{\mathbb{SU}(4)}\Big] = 2$, as expected from the fact that it is a projector onto a subspace of dimension 2. Assuming for simplicity that n is even (the calculations below can be readily generalized to the case when n is odd), we find that the single layer moment operator is

$$\mathcal{M}^{(t)}_{\mathcal{E}_1} = \left(\mathbb{1} \otimes \left(\mathcal{M}^{(2)}_{\mathbb{SU}(4)}\right)^{\otimes(n/2-1)} \otimes \mathbb{1}\right) \circ \left(\mathcal{M}^{(2)}_{\mathbb{SU}(4)}\right)^{\otimes(n/2)} , \tag{50}$$

where $\mathbb{1}$ denotes the 2×2 identity matrix. Note that the previous has allowed us to effectively express $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ as a $2^n \times 2^n$ matrix, which is much smaller than its vanilla form of $16^n \times 16^n$. Moreover, this approach follows very closely manuscripts in the literature where Haar average properties are computed by mapping objects such as $\mathcal{M}_{\mathcal{E}_1}^{(t)}$ to Markov-like mixing process matrices [6–9].

To compute the norm of the operator $\mathcal{A}_{\mathcal{E}_1}^{(2)} = \mathcal{M}_{\mathcal{E}_1}^{(2)} - \mathcal{M}_{G}^{(2)}$ we can use the fact that deep hardware efficient ansatz are universal [10], meaning that $\mathfrak{g} = \mathfrak{su}(d)$, and hence $G = \mathbb{SU}(d)$. From here, we can express

$$\mathcal{M}_{\mathbb{SU}(d)}^{(2)} = \frac{1}{4^n - 1}\left(|\mathbb{I}\rangle\rangle\langle\langle\mathbb{I}| + |\mathbb{S}\rangle\rangle\langle\langle\mathbb{S}| - \frac{1}{2^n}\left(|\mathbb{I}\rangle\rangle\langle\langle\mathbb{S}| + |\mathbb{S}\rangle\rangle\langle\langle\mathbb{I}|\right)\right), \tag{51}$$

where $|\mathbb{I}\rangle\rangle = \bigotimes_{i=1}^n |\mathbb{I}_i\rangle\rangle$ and $|\mathbb{S}\rangle\rangle = \bigotimes_{i=1}^n |\mathbb{S}_i\rangle\rangle$. Following a similar procedure such as that used to build an efficient representation of $\mathcal{M}_{\mathcal{E}_1}^{(t)}$ in Eq. (50), we can express $\mathcal{M}_{\mathbb{SU}(d)}^{(2)}$ and thus $\mathcal{A}_{\mathcal{E}_1}^{(2)}$ as a $2^n \times 2^n$ matrix. Combining this with the fact that $\mathcal{A}_{\mathcal{E}_1}^{(2)}$ can be represented efficiently as a matrix product operator, and the fact that the eigenvector associated to its largest eigenvalue admits an efficient representation as a bond dimension 4 matrix product state, we can easily obtain the largest eigenvalue for $\mathcal{A}_{\mathcal{E}_1}^{(2)}$.

For instance, in Fig. 1(b) we plot the largest eigenvalue $\lambda_{\max}$ of $\mathcal{A}_{\mathcal{E}_1}^{(2)}$ as a function of the number of qubits n that $U(\boldsymbol{\theta})$ acts on. We have considered system sizes $n = 2, \ldots, 90$. Here we can see that for $n = 2$, $\lambda_{\max} = 0$, as expected. Moreover, as n increases, the value of $\lambda_{\max}$ appears to saturate at a value ≈ 0.639. Then, in Fig. 1(c) we plot $\left\|\mathcal{A}_{\mathcal{E}_L}^{(2)}\right\| = \left(\left\|\mathcal{A}_{\mathcal{E}_1}^{(2)}\right\|\right)^L$ for different values of L. Here, we find that as the number of layers increases, the norm of the expressiveness superoperator $\left\|\mathcal{A}_{\mathcal{E}_L}^{(2)}\right\|$ decreases exponentially. Indeed, we can now answer questions such as: "*How many layers are needed for the circuit to become an $\epsilon = 10^{-9}$ approximate design?*" For the number of qubits considered, we show that $L = 47$ will suffice.

To finish, we note that here we have considered a circuit whose local gates are randomly taken from $\mathbb{SU}(4)$. However, the methods presented can be extended and generalized to other circuit architectures where the gates are sampled from any subgroup. In particular, we can always assume that each gate in the layer is generated by a Pauli and hence samples from a representation of $\mathbb{U}(1)$. We leave the exploration of such cases for future work.

V. PROOF OF THEOREM 3

Here we present a proof for Theorem 3 of the main text, which we recall for convenience:

Theorem 3. *Let $\mathfrak{g} \subseteq \mathfrak{u}(2^n)$ be any dynamical Lie algebra, and suppose either ρ or O are in $i\mathfrak{g}$ with ρ a density matrix. Then the difference between the loss function variance of an L-layered circuit, with ensemble of unitaries $\mathcal{E}_L$, and the variance for a circuit that forms a 2-design over G can be bounded as*

$$|\mathrm{Var}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathrm{Var}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| \leqslant 3\left(\left\|\mathcal{A}_{\mathcal{E}_1}^{(2)}\right\|_\infty\right)^L \|O\|_1^2, \tag{52}$$

where the usual Schatten p-norm and induced Schatten p-norm are given by Eq. (32) and Eq. (33).

Proof. Let us begin by observing that we may split

$$\begin{aligned}|\mathrm{Var}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathrm{Var}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| &= \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - (\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)])^2 + \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - (\mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)])^2\right| \\ &\leqslant \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]^2\right| + \left|(\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)])^2 - (\mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)])^2\right|.\end{aligned} \tag{53}$$

To estimate the first term of Eq. (53), we mimic Eq. (18) to write the second moments in terms of the second moment operators of ensembles $\mathcal{E}$, which again are linear maps $\mathcal{M}_{\mathcal{E}}^{(2)}\colon \mathfrak{gl}(\mathcal{H})^{\otimes 2} \to \mathfrak{gl}(\mathcal{H})^{\otimes 2}$,

$$\begin{aligned}
\left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)]^2 - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]^2\right| &= \left|\mathrm{Tr}\left[\mathcal{M}_{\mathcal{E}_L}^{(2)}(\rho^{\otimes 2})O^{\otimes 2}\right] - \mathrm{Tr}\left[\mathcal{M}_G^{(2)}(\rho^{\otimes 2})O^{\otimes 2}\right]\right| \\
&= \left|\mathrm{Tr}\left[(\mathcal{M}_{\mathcal{E}_L}^{(2)} - \mathcal{M}_G^{(2)})(\rho^{\otimes 2})O^{\otimes 2}\right]\right| \\
&\leqslant \left\|\rho^{\otimes 2}\right\|_1 \left\|O^{\otimes 2}\right\|_1 \left\|\mathcal{A}_{\mathcal{E}_L}^{(2)}\right\|_\infty \\
&\leqslant \left\|O^{\otimes 2}\right\|_1 \left\|\mathcal{A}_{\mathcal{E}_L}^{(2)}\right\|_\infty,
\end{aligned} \tag{54}$$

where we used Hölder's inequality for matrices and that $\left\|\rho^{\otimes 2}\right\|_\infty \leqslant 1$ since ρ a density matrix and so has eigenvalues at most 1.

To estimate the second term of Eq. (53), we use the difference of squares formula $a^2 - b^2 = (a+b)(a-b)$ to write

$$\left|(\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)])^2 - (\mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)])^2\right| = \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)] + \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]\right| \cdot \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)] - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]\right|. \tag{55}$$

Recalling that for any ensemble $\mathcal{E}$ the first moment operator $\mathcal{M}_{\mathcal{E}}^{(1)}\colon \mathfrak{gl}(\mathcal{H}) \to \mathfrak{gl}(\mathcal{H})$ has operator norm $\left\|\mathcal{M}_{\mathcal{E}}^{(1)}\right\|_\infty \leqslant 1$ by Lemma 5, we see that

$$\left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)] + \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]\right| = \left|\mathrm{Tr}\left[(\mathcal{M}_{\mathcal{E}_L}^{(1)} + \mathcal{M}_G^{(1)})(\rho)O\right]\right| \leqslant \|O\|_1 \left\|\mathcal{M}_{\mathcal{E}_L}^{(1)} + \mathcal{M}_G^{(1)}\right\|_\infty \leqslant 2\|O\|_1, \tag{56}$$

again by Hölder's inequality and $\|\rho\|_\infty < 1$. Essentially the same estimate as for the variance shows that

$$\left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)] - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]\right| \leqslant \|O\|_1 \left\|\mathcal{A}_{\mathcal{E}_L}^{(1)}\right\|_\infty. \tag{57}$$

Putting it all together into Eq. (53) and using that $\|O\|_1^2 = \left\|O^{\otimes 2}\right\|_1$, we have that

$$\begin{aligned}
\left|\mathrm{Var}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)] - \mathrm{Var}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]\right| &\leqslant \|O\|_1^2 \left(\left\|\mathcal{A}_{\mathcal{E}_L}^{(2)}\right\|_\infty + 2\left\|\mathcal{A}_{\mathcal{E}_L}^{(1)}\right\|_\infty\right) \\
&= \|O\|_1^2 \left(\left\|\mathcal{A}_{\mathcal{E}_1}^{(2)}\right\|_\infty^L + 2\left\|\mathcal{A}_{\mathcal{E}_1}^{(1)}\right\|_\infty^L\right) \\
&\leqslant 3\|O\|_1^2 \left\|\mathcal{A}_{\mathcal{E}_1}^{(2)}\right\|_\infty^L,
\end{aligned} \tag{58}$$

where in the second line we have used Lemma 7, and in the third line we have used Lemma 8 below to upper bound $\left\|\mathcal{A}_{\mathcal{E}_1}^{(1)}\right\|_\infty \leqslant \left\|\mathcal{A}_{\mathcal{E}_1}^{(2)}\right\|_\infty$.

Lemma 8. *Let $\mathcal{E}$ be an ensemble of unitaries that forms an ϵ-approximate G t-design, i.e., $\left\|\mathcal{M}_{\mathcal{E}}^{(t)} - \mathcal{M}_G^{(t)}\right\|_\infty < \epsilon$. Then $\mathcal{E}$ forms an ϵ-approximate s-design for all $1 \leqslant s \leqslant t$.*

Proof. We show that $\mathcal{E}$ forms an ϵ-approximate $(t-1)$-design, whence the other cases follow. Let $\mathcal{M}_{\mathcal{E}}^{(t)}, \mathcal{M}_G^{(t)}\colon \mathfrak{gl}(\mathcal{H}^{\otimes t}) \to \mathfrak{gl}(\mathcal{H}^{\otimes t})$. We may first note that for any $\rho \in \mathfrak{gl}(\mathcal{H}^{\otimes(t-1)})$, we have

$$\mathcal{M}_{\mathcal{E}}^{(t)}(\rho \otimes \mathbb{1}_{\mathcal{H}}) = \int_{\mathcal{E}} U^{\otimes(t-1)}\rho(U^\dagger)^{\otimes(t-1)} \otimes UU^\dagger\, d\mu = \left(\int_{\mathcal{E}} U^{\otimes(t-1)}\rho(U^\dagger)^{\otimes(t-1)}\, d\mu\right) \otimes \mathbb{1}_{\mathcal{H}} = \mathcal{M}_{\mathcal{E}}^{(t-1)}(\rho) \otimes \mathbb{1}_{\mathcal{H}}, \tag{59}$$

and likewise for the Haar ensemble G. Then computing the norm, we obtain

$$\begin{aligned}
\left\|\mathcal{M}_{\mathcal{E}}^{(t-1)} - \mathcal{M}_G^{(t-1)}\right\|_\infty = \sup_{\substack{\rho \in \mathfrak{gl}(\mathcal{H}^{\otimes(t-1)}) \\ \rho \neq 0}} \frac{\left\|(\mathcal{M}_{\mathcal{E}}^{(t-1)} - \mathcal{M}_G^{(t-1)})(\rho)\right\|_\infty}{\|\rho\|_\infty} &= \sup_{\substack{\rho \in \mathfrak{gl}(\mathcal{H}^{\otimes(t-1)}) \\ \rho \neq 0}} \frac{\left\|(\mathcal{M}_{\mathcal{E}}^{(t)} - \mathcal{M}_G^{(t)})(\rho \otimes \mathbb{1})\right\|_\infty}{\|\rho \otimes \mathbb{1}\|_\infty} \\
&\leqslant \left\|\mathcal{M}_{\mathcal{E}}^{(t)} - \mathcal{M}_G^{(t)}\right\|_\infty \\
&< \epsilon.
\end{aligned} \tag{60}$$

Remark 1. *While we have expressed Theorem 3 with the intent of discussing the variance, one may wonder about bounds quantifying the convergence of deep parameterized quantum circuits to the Haar measure for higher moments. It is not difficult to see that mimicking Eq. (54) for the t^{th} moment yields*

$$\left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho,O)]^t - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho,O)]^t\right| \leqslant \left\|O^{\otimes t}\right\|_1 \left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty^L. \tag{61}$$

Then, to discuss quantities such as the t^{th} cumulant, Lemma 8 guarantees that for any ensemble $\mathcal{E}_L$ and any integer $1 \leqslant s \leqslant t$, we have $\left\|\mathcal{A}^{(s)}_{\mathcal{E}_L}\right\|_\infty \leqslant \left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|_\infty$, which grants control over lower-order moments in terms of the t^{th} moment.

VI. NUMERICAL SIMULATIONS

In this section, we present numerical experiments that (i) assess the exactness of Eq. (9) and (ii) probe some non-intuitive aspects revealed by the Lie-algebraic perspective.

In the following, we consider the trainability of circuits composed of parameterized single-qubit Z-rotations and parameterized two-qubit XX-rotations acting over nearest qubit neighbors arranged in a one-dimensional topology. This accounts for a total of $(2n-1)$ parametrized gates per layer of the parametrized quantum circuit. The DLA associated to this circuit is

$$\mathfrak{g} = \left\langle \left(\{iX_jX_{j+1}\}_{j=1}^{n-1}\right) \cup \left(\{iZ_j\}_{j=1}^{n}\right)\right\rangle_{\text{Lie}}. \tag{62}$$

As shown in [11], the DLA is a simple Lie algebra $\mathfrak{g} \cong \mathfrak{so}(2n)$ that has dimension $\dim(\mathfrak{g}) = n(2n-1) \in \Theta(\text{poly}(n))$. An orthogonal basis of $\mathfrak{g}$ is given by

$$i\{\widehat{X_iX_j}, \widehat{X_iY_j}, \widehat{Y_iX_j}, \widehat{Y_iY_j}\}_{1\leqslant i<j\leqslant n}, \quad \text{where} \quad \widehat{A_iB_j} := A_iZ_{i+1}\cdots Z_{j-1}B_j. \tag{63}$$

Given the Cartan subalgebra $\mathfrak{h} = \text{span}_{\mathbb{R}}\{Z_i\}_{i=1}^n$ of $\mathfrak{g}$, one readily identifies the state $|hw\rangle := |0\rangle^{\otimes n}$ as one highest weight vector of the algebra with $\mathfrak{g}$-purity $\mathcal{P}_{\mathfrak{g}}(|hw\rangle\langle hw|) = \dim\mathfrak{h}/2^n = n/2^n$.

In the simulations, we study the variance of the loss for varied system sizes $n \in [3, 15]$, and consider 4 distinct setups each characterized by the choice of the measurement observable O and of the pure initial state $|\psi_0\rangle$. In all cases $O \in i\mathfrak{g}$.

- **Setup 0:** $O = X_pX_{p+1} + Z_p$ (with $p = \lfloor \frac{n}{2} \rfloor$ indexing the middle of the chain) and $|\psi_0\rangle = |hw\rangle$. This illustrates the case where the operator contains a generalized local component and a generalized-nonlocal component; and where the initial state has maximal $\mathfrak{g}$-purity.

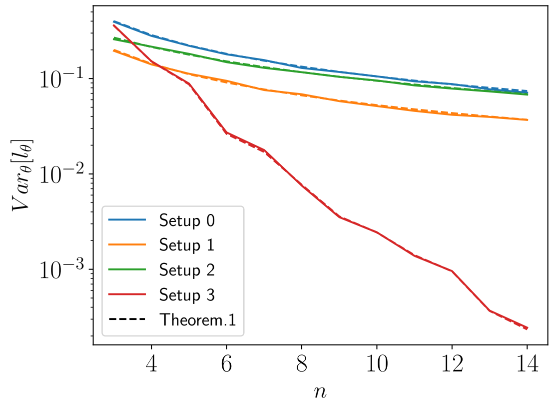
- **Setup 1:** $O = \widehat{X_1Y_n}$ and $|\psi_0\rangle = |hw\rangle$. This illustrates the case where the operator is global (i.e., it acts non trivially on the whole system) but belongs in $i\mathfrak{g}$, and where the initial state has maximal $\mathfrak{g}$-purity.

- **Setup 2:** $O = X_pX_{p+1} + Z_p$ and $|\psi_0\rangle$ is generated by applying a single layer of random single-qubit rotations to $|hw\rangle$.

- **Setup 3:** $O = X_pX_{p+1} + Z_p$ and $|\psi_0\rangle$ is generated by applying n layers of a random circuit composed of two-qubit rotations to $|hw\rangle$.

As we soon discuss further, these last two setups illustrate cases where the $\mathfrak{g}$-purity of the state input to the circuit has been decreased (compared to the original $|hw\rangle$ state), to a different extent, by means of unitary rotations.

For the setups just described, we systematically estimate variances of the cost over a total of 5,000 random initializations of the circuit parameters. For each system size, the variances are estimated for increased number of layers, until they converge to a fixed value (i.e., until the circuit forms an approximate design). In Fig. 2, we report all these estimated variances (solid lines with distinct colors for each of the setups), and also, the theoretical predictions obtained from Eq. (9) in Theorem 1. Over all the setups studied, theory and numerically estimated variances match closely (up to small errors arising from the sample variance uncertainty). Going further, it is verified that measuring a global measurement operator (Setup 1) does not necessarily incur BPs provided that it belong to $i\mathfrak{g}$. Furthermore, we can see



Supp. Fig. 2. **Scaling of the variance of the loss functions for 4 different settings.** We numerically estimate variances of the loss function for a circuit with a DLA given in Eq. (62) for 4 different setups composed of different pairs of measurement observables and initial states. Each setup is described in the main text, and the associated variance corresponds to a solid in the plot. These estimated variances are compared to exact evaluation of Eq. (9) of Theorem 1 (dashed lines).

that under unitary transformation (Setups 2 and 3) the $\mathfrak{g}$-purity of the input states are indeed decreased. Otherwise, these would result in similar variances as Setup 0, as they share the same measurement observable. In particular, we highlight the fact that in Setup 2 we perform single qubit rotations (which do not change the standard purity $\mathrm{Tr}\left[\rho^2\right]$, nor the standard entanglement), but which incur in a small decrease of the variance. Note that in the latter the variance decrease is small, as it is still polynomially vanishing (i.e., this setup does not incur BPs). This result is in contrast to Setup 3, where the circuit applied to $|hw\rangle$ generates enough generalized entanglement to induce a BP, as evidenced by an exponentially decaying variance.

[1] B. Collins, S. Matsumoto, and J. Novak, The weingarten calculus, Notices Of The American Mathematical Society **69**, 734 (2022).

[2] A. Kirillov Jr, *An introduction to Lie groups and Lie algebras*, 113 (Cambridge University Press, 2008).

[3] A. W. Knapp, *Lie Groups Beyond an Introduction*, Vol. 140 (Springer Science & Business Media, 2013).

[4] D. D'Alessandro, *Introduction to Quantum Control and Dynamics*, Chapman & Hall/CRC Applied Mathematics & Nonlinear Science (Taylor & Francis, 2007).

[5] A. A. Mele, Introduction to haar measure tools in quantum information: A beginner's tutorial, arXiv preprint arXiv:2307.08956 (2023).

[6] 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).

[7] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits anticoncentrate in log depth, PRX Quantum **3**, 010333 (2022).

[8] J. Napp, Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze, arXiv preprint arXiv:2203.06174 (2022).

[9] P. Braccia, P. Bermejo, L. Cincio, and M. Cerezo, Computing exact moments of local random quantum circuits via tensor networks, arXiv preprint arXiv:2403.01706 (2024).

[10] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing Barren Plateaus with Tools from Quantum Optimal Control, Quantum **6**, 824 (2022).

[11] E. Kökcü, T. Steckmann, Y. Wang, J. Freericks, E. F. Dumitrescu, and A. F. Kemper, Fixed depth hamiltonian simulation via cartan decomposition, Physical Review Letters **129**, 070501 (2022).