Editors' Suggestion

Models of Quantum Complexity Growth

Fernando G.S.L. Brandão,[1,2,3,4] Wissam Chemissany,[1] Nicholas Hunter-Jones,[1,5,*] Richard Kueng,[1,3,6,†] and John Preskill[1,2,3,4]

[1] *Institute for Quantum Information and Matter, Caltech, Pasadena, California, USA*
[2] *AWS Center for Quantum Computing, Pasadena, California, USA*
[3] *Department of Computing and Mathematical Sciences, Caltech, Pasadena, California, USA*
[4] *Walter Burke Institute for Theoretical Physics, Caltech, Pasadena, California, USA*
[5] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada*
[6] *Institute for Integrated Circuits, Johannes Kepler University Linz, Austria*

(Received 13 January 2021; accepted 21 May 2021; published 29 July 2021)

The concept of quantum complexity has far-reaching implications spanning theoretical computer science, quantum many-body physics, and high-energy physics. The quantum complexity of a unitary transformation or quantum state is defined as the size of the shortest quantum computation that executes the unitary or prepares the state. It is reasonable to expect that the complexity of a quantum state governed by a chaotic many-body Hamiltonian grows linearly with time for a time that is exponential in the system size; however, because it is hard to rule out a shortcut that improves the efficiency of a computation, it is notoriously difficult to derive lower bounds on quantum complexity for particular unitaries or states without making additional assumptions. To go further, one may study more generic models of complexity growth. We provide a rigorous connection between complexity growth and unitary k-designs, ensembles that capture the randomness of the unitary group. This connection allows us to leverage existing results about design growth to draw conclusions about the growth of complexity. We prove that local random quantum circuits generate unitary transformations whose complexity grows linearly for a long time, mirroring the behavior one expects in chaotic quantum systems and verifying conjectures by Brown and Susskind. Moreover, our results apply under a strong definition of quantum complexity based on optimal distinguishing measurements.

DOI: 10.1103/PRXQuantum.2.030316

I. MOTIVATION AND OVERVIEW

The *complexity* of a computation is a measure of the resources needed to perform the computation. In a classical model of computation, the complexity of a Boolean function may be defined as the minimal number of elementary steps needed to evaluate the function. The precise number of steps needed depends on how the model is chosen, but this notion of complexity provides a useful way to quantify the hardness of a computational problem because how the number of steps scales with the size of the input to the problem has only weak dependence on the choice of model. By broad consensus, a computational task is considered to be feasible if its complexity grows no faster than a power of the input size, and intractable otherwise; using this criterion, all classical models of computation agree about which problems are (classically) "easy" and which ones are "hard."

Likewise, we may separate computational tasks into those that are easy or hard for a quantum computer. The circuit model of quantum computation provides a convenient way to quantify quantum complexity—namely, the quantum complexity of a Boolean function is the minimal size of a quantum circuit, which computes the function and outputs the right answer with high success probability. Here by the size of the circuit we mean the number of quantum gates in the circuit. These gates are chosen from a universal set of gates, where each gate in the set is a unitary transformation acting on a constant number of qubits or qudits. Though there are many ways to choose the universal gate set, any set of universal gates can simulate another accurately and efficiently, so that circuit size provides a useful model-independent measure of complexity.

*nickrhj@pitp.ca
†richard.kueng@jku.at

From a physicist's perspective, a quantum computation is a process governed by a local time-dependent Hamiltonian, and an intractable computation is a process that requires a time, which grows superpolynomially with the system size. Such intractable processes are not expected to be observed in nature.

Furthermore, in quantum physics, in contrast to classical digital computation, there is a meaningful notion of complexity not only for processes, but also for quantum states. Starting from a state in which all the bits are set to 0, any string of n classical bits can be prepared by flipping at most n bits. But the time needed to prepare a pure n-qubit quantum state, starting from a product state, even if we are permitted to use any time-dependent Hamiltonian, which is a sum of terms with constant weight and bounded norm, can be exponential in n. In fact, because the volume of the n-qubit Hilbert space is *doubly exponential* in n, while the number of quantum circuits with T gates is merely exponential in T, *most* n-qubit pure quantum states have exponentially large complexity. That is, for a typical pure state in the n-qubit Hilbert space, the time needed to prepare the state with some small constant error δ, starting from a product state, grows exponentially with n. Thus, nearly all quantum states of any macroscopic system will forever be far beyond the grasp of the quantum engineers [1].

While the complexity of quantum *circuits* has long been a foundational concept in quantum information theory [2], appreciation that quantum *state* complexity is an important concept has blossomed relatively recently. For example, the complexity of ground-state wave functions may be used to classify topological phases of matter at zero temperature [3]. Furthermore, a chaotic quantum Hamiltonian H can be usefully characterized by saying that evolution governed by H over a long time period generates highly complex states. A particularly intriguing proposal is that, in the context of the anti-de Sitter/conformal field theory (AdS/CFT) correspondence, the complexity of a quantum state of the boundary theory corresponds to the volume in the bulk geometry, which is hidden behind the event horizon of a black hole [4–7].

When we say a quantum state is highly complex, we mean there is no easy way to prepare the state, but how can we be sure? Perhaps we were not clever enough to think of an ingenious shortcut that prepares the state efficiently. It is not possible in practice to enumerate all the quantum circuits that approximate a specified state to find one of minimal size. For that reason, it is quite difficult to obtain a useful lower bound on the complexity of the quantum state prepared by a specified many-body Hamiltonian in a specified time. It is reasonable to expect that, for a chaotic Hamiltonian H and an unentangled initial state, the complexity grows linearly in time for an exponentially long time, but we do not have the tools to prove it from first principles for any particular H.

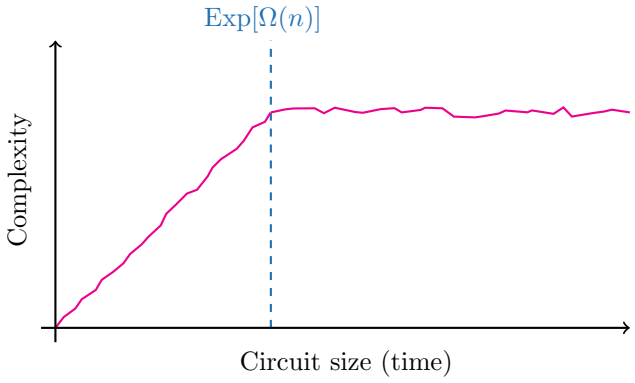


FIG. 1. *Expected complexity growth in random circuits.* Conjecture 1 states that, for random quantum circuits acting on n qubits, the circuit complexity grows linearly with circuit size (time) until it saturates at a value exponentially large in n. Our work provides rigorous evidence supporting this picture for quantum systems with sufficiently large local dimension; see Corollary 5.

One possible approach is to rely on highly plausible complexity theory assumptions to derive nontrivial conclusions about the complexity of states generated by particular circuits or Hamiltonians [8–10]. Another is to consider ensembles of circuits, and to derive lower bounds on complexity, which hold with high probability when samples are selected from these ensembles. We follow the latter approach here, drawing inspiration from recent work by Susskind [8] and Brown and Susskind [7]. These authors state a conjecture about the complexity growth of geometrically local random quantum circuits (see Fig. 1).

Conjecture 1 (Brown and Susskind [7]; Susskind [8]): *Most local random circuits of size T have a complexity that scales linearly in T for an exponentially long time.*

Our goal is to strengthen the evidence supporting this conjecture.

Brown and Susskind provided evidence for this scaling law by means of a simple counting argument; see also Ref. [11]. For a fixed finite set of universal quantum gates, consider the ensemble of all circuits with size T. By definition, this ensemble accurately approximates (to within a specified error δ) all unitary transformations with complexity T or less. Furthermore, the number of circuits increases exponentially with T, and, because the unitary group has a very large volume, it seems reasonable to assume that "collisions" between circuits are rare unless T is very large; that is, that the number of distinct unitary transformations realized by this ensemble (where "distinct" means more than distance δ apart) is comparable to the number of circuits. This means that the number of circuits with size T' is too small to account for more than a small fraction of the unitary transformations realized by

circuits of size T if T' is much smaller than T. In other words, most random circuits with size T have complexity at least T', where T' is comparable to T.

This argument hinges on a crucial assumption, which sounds plausible but is hard to prove: *collisions between circuits of subexponential size are rare*. Collisions certainly occur for any circuit size T, and necessarily become common for circuits of exponential size, where T is comparable to the Hilbert-space dimension so that the exponential of T is comparable to the Hilbert-space volume. Thus an analytic treatment of complexity growth seems like a daunting combinatorial task.

The work [12] provides some rigorous support for Conjecture 1. There, the authors show that local random circuits can "fool" short-measurement procedures. That is, a typical quantum state prepared by a local random circuit of size polynomial in n, acting on an initial product state, cannot be distinguished from a maximally mixed state by any procedure that is much simpler than running the circuit backwards and verifying that the initial product state is recovered. Although not stated in this fashion, the results from Ref. [12] imply that, with high probability, a local random circuit of size T has complexity $\Omega(T^{1/11})$. While this argument rigorously proves a weakened version of Conjecture 1, there are still issues we wish to address:

(i) *Restricted notion of complexity:* The authors implicitly define complexity as the capability of fooling short-measurement protocols. While this operational notion of complexity is well motivated, the actual measurement procedures considered are quite restrictive. In particular, they do not take into account ancilla-assisted measurements—a mainstay of modern quantum information.

(ii) *Collisions are not treated explicitly:* The ensemble of local random circuits of size T defines a probability distribution on the n-qubit unitaries. If we are only interested in specifying unitary transformations up to some specified error δ, collisions occur, so that some unitaries are more likely than others. The arguments in Ref. [12] show that the unitaries sampled from this distribution typically have complexity $\Omega(T^{1/11})$, but do not rule out the possibility that the distribution is highly nonuniform. It is at least a logical possibility, compatible with the findings of Ref. [12], that the ensemble contains only a small number of unitaries, which have high complexity, all of which occur with relatively high probability. To conclude that the ensemble contains many high-complexity unitaries, we need to know more about the properties of the probability distribution governing the ensemble.

(iii) *Polynomial relation between circuit size and complexity:* The relation between circuit size T and expected minimal complexity $T^{1/11}$ is polynomial, not (yet) linear as required by Conjecture 1.

In this work we make progress toward a rigorous proof of Conjecture 1 by developing a general framework that addresses some of the shortcomings of the previously known rigorous evidence in favor of the conjecture [12]. In particular, we define and use a *strong* notion of complexity, which captures the difficulty of distinguishing a given circuit from the most useless possible quantum channel: the completely depolarizing channel $\mathcal{D}(\rho) = [\mathrm{Tr}(\rho)/d]\mathbb{I}$ that maps any state to the maximally mixed state.

Definition 1 (Strong complexity: Informal definition): *The complexity of a quantum circuit U is the minimal circuit size required to implement an ancilla-assisted measurement that is capable of distinguishing $\rho \mapsto U\rho U^\dagger$ from the completely depolarizing channel $\rho \mapsto (1/d)\mathbb{I}$.*

We refer to Sec. II A for a more detailed motivation and a precise statement of this definition. For now, we emphasize that this strong definition of complexity implies other (weaker) definitions, such as the minimal circuit size required to approximate U.

Our first main contribution is a rigorous connection between complexity growth and the notion of *approximate unitary k-designs* [13,14]. We use the notation $\{p_i, U_i\}$ for an ensemble of unitary transformations in which the unitary U_i occurs with probability p_i. A unitary k-design is an ensemble with strong pseudorandom properties; an approximate k-design accurately approximates the first k-moments of the Haar measure on the unitary group. Hence a k-design with large k behaves essentially like a Haar-random ensemble of unitaries, while a small-k-design can be highly structured. For instance, the n-qubit Pauli group forms a 1-design, while the n-qubit Clifford group is a 3-design [15–17]. The design order k allows us to interpolate between these two very different regimes. Intuitively, we would expect that the complexity of a k-design grows with k. Our first technical contribution makes this intuition precise: a linear growth in design implies a linear growth in (strong) complexity.

Theorem 2 (Informal statement): *Let $\{p_i, U_i\}$ be an approximate unitary k-design. Then, a randomly selected (according to the weights) element is very likely to have strong circuit complexity approximately equal to k.*

We refer to Theorem 9 for a more detailed, quantitative statement. This result strengthens the assertions in Ref. [12] by allowing ancilla-assisted measurement procedures. To do so we prove novel bounds on Haar moments, see Sec. II D for details. Our second technical contribution

shows that the k-design property alone severely limits the likelihood of collisions.

Lemma 3: *Let $\{p_i, U_i\}$ be an approximate k-design. Then, the associated weight distribution cannot be too spiky:* $\max_i p_i \lesssim k! d^{-2k}$.

This result formalizes the intuitive idea that giving unusually high weight to some unitaries cannot be compatible with the k-design property, but we are not aware of any precise statements along these lines in the existing literature. Importantly, because Lemma 3 establishes that the distribution is nearly flat, knowing that sampling from a k-design yields a high-complexity unitary with high probability (as stated in Theorem 2) allows us to infer that there must be many distinct high-complexity unitaries in the ensemble. Here our reasoning is based on an approximate version of Laplace's definition of probability: if events are assigned nearly equal probabilities, then the probability of property X is approximately the number of events with property X divided by the total number of events. Together, Theorem 2 and Lemma 3 imply the following corollary.

Corollary 4: *Any approximate k-design contains exponentially many (in k) unitaries that have circuit complexity* $\Omega(k)$.

While Corollary 4 does not by itself strongly constrain how these high-complexity unitary transformations are distributed geometrically within the n-qudit unitary group, we are also able to prove a stronger result: *An approximate k-design contains exponentially many (in k) high-complexity unitaries whose pairwise distance* (i.e., the distance between any pair of unitaries) *is almost maximal in the diamond norm.* This stronger statement rules out the possibility that most of the high-complexity unitaries reside inside a few tightly packed clusters within $U(d)$.

Approximate unitary k-designs are a central concept in quantum information, where their pseudorandom properties have found extensive application across subfields, e.g., state distinguishability [18], decoupling [19], state tomography [20,21], randomized benchmarking [22], equilibration [12] (and references therein), information scrambling [11,23], and many more. As a result, several probabilistic constructions are known. Applying Corollary 4 to any of these constructions establishes a rigorous model for quantum complexity growth. In particular, the following.

(a) *Local random quantum circuits with polynomial design growth:* Ref. [12] proves that the set of all geometrically local circuits of size $T = O(n^2 k^{11})$ forms an approximate unitary k-design [24]. Corollary 4 therefore implies that local circuits of size T contain at least $\exp[\Omega(T^{1/11})]$ elements with strong complexity $\Omega(T^{1/11})$.

(b) *Stochastic quantum Hamiltonians with polynomial design growth:* One can study the growth of complexity in continuous-time models of chaotic dynamics, rather than the discrete-time dynamics embodied by random circuits [25–27]. Stochastic Hamiltonian dynamics, in which a local Hamiltonian fluctuates as a function of time, has been shown to realize approximate k-designs [26] with a relationship between k and the evolution time similar to what was established in Ref. [12] for local random circuits. Further progress achieved in Ref. [27] shows that, for a particular class of stochastic Hamiltonians, evolution time linear in k suffices to generate approximate k-designs for $k = o(\sqrt{n})$. Corollary 4 therefore implies that with high probability the complexity grows linearly in time, at least for a while.

(c) *Local random circuits with linear design growth:* Recently, the results of Ref. [12] were improved using an exact mapping from random circuits to the statistical mechanics of a lattice model [28], showing that local circuits of size $T = O(n^2 k)$ form approximate k-designs in the limit of large local dimension (Hilbert space dimension $d = q^n$ with q large). The q dependence was subsequently improved in Ref. [29] by studying the spectral gap of the moment operator for random quantum circuits. Combined with Corollary 4 this establishes a *linear* relation between circuit size and complexity. Thus we can prove the following statement analogous to Conjecture 1.

Corollary 5: *The set of all local circuits of size T contains at least* $\exp[\Omega(T)]$ *elements with strong complexity* $\Omega(T)$*, provided that the local dimension is sufficiently large:* $q \geq \Omega(k^2)$.

More precise statements of our main results, and a more detailed comparison to previous work, can be found in Sec. II.

II. QUANTUM COMPLEXITY AND UNITARY DESIGNS

A. Operational definitions of complexity

1. State complexity

We consider systems comprised of n qudits with local dimension q: $d = q^n$. Existing works on complexity typically start with identifying a class of states that are *useful* starting states for quantum computations. In this work we take a reverse approach and start with identifying a *useless*

state. The maximally mixed state

$$\rho_0 = \frac{\mathbb{I}}{d}, \tag{1}$$

is unique in the sense that it is invariant under arbitrary unitary evolutions, including any quantum circuit. Intuitively, useful starting states should be as far away from this useless state as possible. If we use trace distance, this intuition is true to some extent. Any pure state $|\psi\rangle\langle\psi|$ obeys

$$\frac{1}{2}\,\||\psi\rangle\langle\psi| - \rho_0\|_1 = 1 - \frac{1}{d}. \tag{2}$$

But this is clearly too coarse for distinguishing the usefulness of different pure states. In order to achieve such a task, we recall the operational interpretation of the trace distance. It corresponds the optimal bias achievable in distinguishing the state $|\psi\rangle\langle\psi|$ from ρ_0 using a single measurement [30,31]. We refer to Appendix B 1b for a more detailed exposition. The optimal measurement achieving this bias is $M = |\psi\rangle\langle\psi|$ and *does* depend on the state in question. Such a measurement may be challenging to implement for states that we would intuitively assign a high complexity to (such as random states) and very easy for states that we consider useful (such as computational basis states). We can interpolate between these extreme cases by limiting the resources available to implement distinguishing measurements. Let $\mathbb{H}_d$ denote the space of $d \times d$ Hermitian matrices. For fixed $r \in \mathbb{N}$, we consider the class of measurements $\mathsf{M}_r(d) \subset \mathbb{H}_d$ that can be implemented by combining (at most) r 2-local gates from a fixed, universal gate set $\mathsf{G} \subset U(4)$. We refer to Appendix B 2 for further details and justification. The maximal bias achievable for quantum states (QS) with such a restricted set of measurements is the solution to the following optimization problem:

$$\begin{aligned}\beta^{\sharp}_{\mathrm{QS}}(r, |\psi\rangle) = \text{maximize} \quad & |\mathrm{Tr}\left[M\left(|\psi\rangle\langle\psi| - \rho_0\right)\right]| \\ \text{subject to} \quad & M \in \mathsf{M}_r(d).\end{aligned} \tag{3}$$

We may decompose the true optimal measurement as $|\psi\rangle\langle\psi| = U|0\rangle\langle 0|U^\dagger$ for some $U \in U(d)$. The unitary U may be approximated to arbitrary precision by 2-local circuits chosen from a universal gate set [32]. This ensures

$$\beta^{\sharp}_{\mathrm{QS}}(r, |\psi\rangle) \to \frac{1}{2}\||\psi\rangle\langle\psi| - \rho_0\|_1 = 1 - \frac{1}{d} \quad \text{as } r \to \infty. \tag{4}$$

For simple states, like computational basis states, this convergence happens rapidly, while generic states require exponentially large circuit sizes. This observation is the motivation for the following definition of complexity.

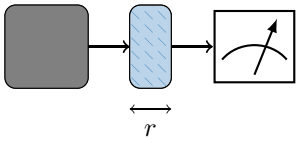


FIG. 2. *Pictographic illustration of strong state complexity (Definition 2).* A blackbox either outputs a (known) pure state $\rho = |\psi\rangle\langle\psi|$, or the maximally mixed state $\rho_0 = (1/d)\mathbb{I}$. The task is to correctly guess which one it produced by applying a preprocessing circuit V (blue line pattern) of limited size r and performing a simple measurement (right). We say that $|\psi\rangle$ has *strong state complexity* less than r if the probability of correctly distinguishing both possibilities is close to optimal.

Definition 2 (Strong state complexity): *Fix $r \in \mathbb{N}$ and $\delta \in (0, 1)$. We say that a pure state $|\psi\rangle$ has strong δ-state complexity at most r if and only if*

$$\beta^{\sharp}_{QS}(r, |\psi\rangle) \geq 1 - \frac{1}{d} - \delta, \tag{5}$$

which we denote as $\mathcal{C}_\delta(|\psi\rangle) \leq r$.

This definition has a ready operational interpretation that is illustrated in Fig. 2. The following result relates it to more traditional definitions.

Lemma 6: *Suppose that $|\psi\rangle \in \mathbb{C}^d$ obeys $\mathcal{C}_\delta(|\psi\rangle) \geq r + 1$ for some $\delta \in (0, 1)$ and $r \in \mathbb{N}$. Then,*

$$\min_{size(V)\leq r} \frac{1}{2}\left\||\psi\rangle\langle\psi| - V|0\rangle\langle 0|V^\dagger\right\|_1 > \sqrt{\delta}, \tag{6}$$

i.e., it is impossible to accurately produce $|\psi\rangle$ with fewer than r elementary gates.

The converse is false in general. To see this, select a generic state $|\widetilde{\psi}\rangle$ on $(n-1)$ qudits and set $|\psi\rangle = |0\rangle \otimes |\widetilde{\psi}\rangle$. Then, the quantity in Eq. (6) is dominated by the (traditional) complexity of $|\widetilde{\psi}\rangle$, which may be very high. Nonetheless, the simple distinguishing measurement $M = |0\rangle\langle 0| \otimes \mathbb{I}$ (ignore everything but the first qudit) achieves

$$\begin{aligned}\mathrm{Tr}\left[M(|\psi\rangle\langle\psi| - \rho_0)\right] &= \mathrm{Tr}\left[|0\rangle\langle 0|\left(|0\rangle\langle 0| - \frac{1}{q}\mathbb{I}\right)\right] \\ &= 1 - \frac{1}{q},\end{aligned} \tag{7}$$

which is high, especially for large local dimension q. This example highlights that Definition 2 is indeed more stringent than traditional definitions of state complexity.

Proof of Lemma 6. By contraposition. Let $\mathsf{G}_r \subset U(d)$ denote the class of unitary circuits that are comprised of at most r 2-local gates chosen from a universal gate

set G. Suppose there exists a size-r circuit $V \in \mathsf{G}_r$ such that $\frac{1}{2}\left\||\psi\rangle\langle\psi| - V|0\rangle\langle 0|V^\dagger\right\|_1 \leq \sqrt{\delta}$. The state difference in question has rank two, which allows for explicitly computing the trace distance: $\frac{1}{2}\||\psi\rangle\langle\psi| - V|0\rangle\langle 0|V^\dagger\|_1 = \sqrt{1-|\langle 0|V^\dagger|\psi\rangle|^2}$. The assumption is therefore equivalent to $|\langle 0|V^\dagger|\psi\rangle|^2 \geq 1-\delta$ and we conclude

$$\begin{aligned}\beta^{\sharp}_{\mathrm{QS}}(r,|\psi\rangle) &\geq \mathrm{Tr}\left[V|0\rangle\langle 0|V^\dagger(|\psi\rangle\langle\psi| - \rho_0)\right] \\ &= |\langle 0|V^\dagger|\psi\rangle|^2 - \frac{1}{d} \geq 1 - \frac{1}{d} - \delta, \end{aligned} \tag{8}$$

because $V|0\rangle\langle 0|V^\dagger \in \mathsf{M}_r$. This in turn implies $\mathcal{C}_\delta(|\psi\rangle) \leq r$ and the claim follows. ■

2. Unitary complexity

We define the complexity of unitary channels $\mathcal{U}(\rho) = U\rho U^\dagger$ in a fashion similar to state complexity. We start with identifying the completely depolarizing channel as the most *useless* channel conceivable:

$$\mathcal{D}(\rho) = \rho_0 = \frac{\mathbb{I}}{d} \quad \text{for all states } \rho. \tag{9}$$

The *diamond distance* between $\mathcal{D}$ and any unitary channel is close to maximal:

$$\frac{1}{2}\|\mathcal{U} - \mathcal{D}\|_\diamond = 1 - \frac{1}{d^2}. \tag{10}$$

As detailed in Appendix B 1c, the diamond distance also has an appealing operational definition [33]. It corresponds to the maximal bias achievable for the task of distinguishing $\mathcal{U}$ from $\mathcal{D}$ with a single channel use. The optimal strategy may involve a quantum memory. Choose a state in the doubled Hilbert space $|\phi\rangle\langle\phi|$, with $|\phi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and input one half into the unknown channel, while the other half remains unchanged in the quantum memory. Subsequently, perform a two-outcome measurement on the output state to distinguish both channels.

An optimal strategy for distinguishing $\mathcal{U}$ from $\mathcal{D}$ corresponds to choosing a maximally entangled (Bell) state $|\Omega\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ as input and measuring $M = (U \otimes \mathbb{I})|\Omega\rangle\langle\Omega|(U^\dagger \otimes \mathbb{I})$. Equivalently, choose $(U^\dagger \otimes \mathbb{I})|\Omega\rangle$ as input and measure $M = |\Omega\rangle\langle\Omega|$ on the output. Similar to the state complexity argument, the optimal input state, or the optimal outcome measurement (or both) depend on the unitary $U \in U(d)$ describing the channel $\mathcal{U}$. This may be challenging to implement, especially if U corresponds to a complicated circuit. We restrict apparatus power by bounding the total circuit sizes that are allowed to implement such a measurement procedure. Let $\mathsf{G}_{r'} \subset U(d^2)$ be the set of all unitary circuits on $2n$ qudits (register+memory) that are comprised of at most r' elementary gates. Likewise, let $\mathsf{M}_{r''} \subset \mathbb{H}_d^{\otimes 2}$ denote the class of all two-outcome measurements on $2n$ qudits that require circuit size at most r'' to implement. The optimal bias for quantum channels (QC) achievable under such restrictions is

$$\begin{aligned}\beta^{\sharp}_{\mathrm{QC}}(r,U) = \text{maximize} \quad & \big|\mathrm{Tr}\big\{M\big[(\mathcal{U}\otimes\mathcal{I})(|\phi\rangle\langle\phi|) \\ & - (\mathcal{D}\otimes\mathcal{I})(|\phi\rangle\langle\phi|)\big]\big\}\big| \\ \text{subject to} \quad & M \in \mathsf{M}_{r'},\ |\phi\rangle = V|0\rangle, \\ & V \in \mathsf{G}_{r''},\ r = r' + r'', \end{aligned} \tag{11}$$

where the identity channel $\mathcal{I} : \mathbb{H}_d \to \mathbb{H}_d$ indicates that the memory is left unchanged. As r increases, more complicated measurements and state preparations become possible. At some point this will include ever more precise approximations of the optimal measurement [32]:

$$\beta^{\sharp}_{\mathrm{QC}}(r,U) \longrightarrow \frac{1}{2}\|\mathcal{U} - \mathcal{D}\|_\diamond = 1 - \frac{1}{d^2} \quad \text{as } r \to +\infty. \tag{12}$$

Similar to the state case, the rate of convergence does depend on the complexity of the unknown unitary U. This is the basis for our operational definition of unitary complexity.

Definition 3 (Strong unitary complexity): *Fix $r \in \mathbb{N}$ and $\delta \in (0,1)$. We say that a unitary $U \in U(d)$ has strong δ-unitary complexity at most r if and only if*

$$\beta^{\sharp}_{QC}(r,U) \geq 1 - \frac{1}{d^2} - \delta, \tag{13}$$

which we denote as $\mathcal{C}_\delta(U) \leq r$.

The operational motivation for this definition is sketched in Fig. 3. Strong unitary complexity (Definition 3) is more stringent than traditional definitions that use approximation errors in some norm. But the comparison between the two is not quite as straightforward as in the state complexity case. This is because, the optimal strategy for distinguishing $\mathcal{U}$ from $\mathcal{D}$ involves a maximally entangled (Bell) input state $|\Omega\rangle\langle\Omega|$, as well as a corresponding two-outcome measurement. In the following statement, we explicitly allow such input states and measurements in the distinguishability protocol. Although mild—relatively short circuits allow for transforming computational basis states into Bell states [34]—this assumption does further increase the power of the measurements we are allowed to make. Our main technical results, most notably Theorem 9, do take this into account and apply to this slightly stronger notion of strong unitary complexity.

Lemma 7: *Consider a setup that contains maximally entangled inputs and measurements and suppose that*

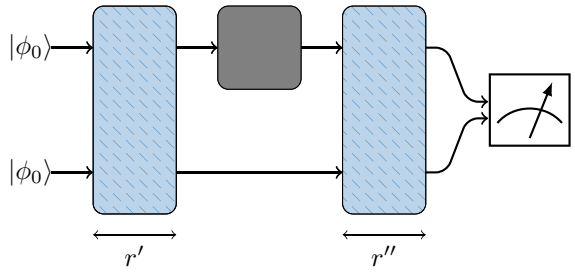


FIG. 3. *Pictographic illustration of strong unitary complexity (Definition 3).* A blackbox (center) takes quantum states as inputs and applies either a unitary channel $\mathcal{U}(\rho) = U\rho U^\dagger$, or the depolarizing channel $\mathcal{D}(\rho) = \rho_0 = \mathbb{I}/d$. The task is to correctly guess which evolution occurred. The rules of the game allow short pre- and postprocessing circuits (blue line patterns) that may involve a quantum memory. The final guess must be based on a simple measurement (right). We say that U has complexity less than $r = r' + r''$ if the probability of correctly distinguishing both options is close to optimal.

$U \in U(d)$ obeys $\mathcal{C}_\delta(U) \geq r + 1$ for some $\delta \in (0, 1)$, $r \in \mathbb{N}$. Then,

$$\min_{size(V)\leq r} \frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_\diamond > \sqrt{\delta}, \tag{14}$$

i.e., it is impossible to accurately approximate U by circuits comprised of fewer than r elementary gates.

Again, the converse relation is false in general.

Proof of Lemma 7. By contraposition. Assume there exists $V \in U(d)$ with $\text{size}(V) \leq r$ such that $\frac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond \leq \sqrt{\delta}$. Then,

$$\begin{aligned}\sqrt{\delta} &\geq \frac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond \geq \frac{1}{2}\left\|(U\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(U^\dagger\otimes\mathbb{I})\right. \\ &\quad \left. - (V\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(V^\dagger\otimes\mathbb{I})\right\|_1 \\ &= \sqrt{1 - |\langle\Omega|V^\dagger U\otimes\mathbb{I}|\Omega\rangle|^2}, \end{aligned}\tag{15}$$

as the second expression involves a trace distance of two pure states, which can be computed explicitly. Next, note that $M = (V\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(V^\dagger\otimes\mathbb{I})$ is a legitimate distinguishing measurement, because $\text{size}(V) \leq r$ and we explicitly include the Bell measurement. Likewise, the input state $|\Omega\rangle\langle\Omega|$ is also allowed and produces a maximally mixed state when completely depolarized: $\mathcal{D}\otimes\mathcal{I}(|\Omega\rangle\langle\Omega|) = \rho_0^{\otimes 2}$ (this is why we need Bell states) ensures

$$\begin{aligned}\beta^\sharp_{\text{QC}}(r, U) &\geq \text{Tr}\Big\{(V\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(V^\dagger\otimes\mathbb{I}) \\ &\quad \times \left[(U\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(U^\dagger\otimes\mathbb{I}) - \rho_0^{\otimes 2}\right]\Big\} \\ &= \left|\langle\Omega|V^\dagger U\otimes\mathbb{I}|\Omega\rangle\right|^2 - \langle\Omega|V^\dagger\rho_0 V\otimes\rho_0|\Omega\rangle \\ &\geq 1 - \delta^2 - \frac{1}{d^2}. \end{aligned}\tag{16}$$

■

B. Approximate unitary designs

The concept of *unitary k-designs* [13,14] provides an interpolation between two extreme cases: (i) small collections of highly structured unitaries that form the basic building blocks of quantum-computing devices (e.g., local Pauli gates, or elements of the Clifford group). (ii) generic (Haar random) unitaries that lack any sort of structure and require circuits of exponential size to approximate.

Roughly speaking, an ensemble $\mathcal{E} = \{p_i, U_i\}$ of unitaries is a unitary k-design if it exactly reproduces the first k moments of the Haar measure over the unitary group. More precisely, given the twirling channels $\mathcal{T}_U^{(k)}(X) = \int dU U^{\otimes k} X (U^\dagger)^{\otimes k}$ and $\mathcal{T}_\mathcal{E}^{(k)}(X) = \sum_i p_i U_i^{\otimes k} X (U_i^\dagger)^{\otimes k}$, an ensemble $\mathcal{E}$ is a unitary design with order k if

$$\mathcal{T}_\mathcal{E}^{(k)}(X) = \mathcal{T}_U^{(k)}(X), \tag{17}$$

for all X in the k-fold tensor product. Although seemingly abstract, this notion captures important physical concepts. 1-designs are in one-to-one correspondence with unitary operator frames, while 2-designs sufficiently capture the notion of *scrambling* [11,23].

Unitary k-designs are known to exist for any dimension d and any order k. Nevertheless, explicit constructions are notoriously difficult to find. This challenge can be overcome by relaxing the notion of a k-design. Indeed, for most applications it is sufficient to ensure that Eq. (17) is only approximately true, see Definition 4 in the Appendix for a precise statement. Several conventions for choosing an appropriate distance measure $\|\cdot\|$ have been put forth, but here we opt for the diamond distance $\|\cdot\|_\diamond$, which quantifies the distinguishability of two ensembles.

In contrast to exact k-designs, several explicit constructions for approximate k-designs have been established [12, 26–28,35,36], including local random circuits and various Brownian circuits and stochastic quantum Hamiltonians. These constructions allow us to relate abstract insights about complexity growth in designs to concrete random circuit models.

C. Complexity by design

This section presents our main technical contributions.

1. State complexity growth

Theorem 8: *Consider the set of (pure) states in $d = q^n$ dimensions that results from applying all unitaries associated with an ϵ-approximate $2k$-design to a fixed (but*

arbitrary) starting state $|\psi_0\rangle$. *Then, this set contains at least*

$$\binom{d+k-1}{k}\left[\frac{1}{1+\epsilon}-2d(n+1)^r|\mathsf{G}|^r\left(\frac{16k^2}{d(1-\delta)^2}\right)^k\right],$$

distinct states that obey $\mathcal{C}_\delta(|\psi\rangle)\geq r+1$ *each. Qualitatively, this number is of order* $(d/k)^k$ *as long as r obeys*

$$r\lesssim\frac{k[n-2\log(k)]}{\log(n)}.$$

Because of collisions, the emphasis on distinct is justified; two or more distinct unitaries can lead to the same final state.

2. Unitary complexity growth

Theorem 9: *A discrete approximate* $2k$*-design in* $d=q^n$ *dimensions contains at least*

$$\frac{d^{2k}}{k!}\left[\frac{1}{1+\epsilon}-3d^2n^{2r}|\mathsf{G}|^r\left(\frac{1024k^4}{d(1-\delta)^2}\right)^k\right],$$

distinct unitaries that obey $\mathcal{C}_\delta(U)\geq r+1$ *each. Qualitatively, this number is of order* $(d^2/k)^k$ *as long as r obeys*

$$r\lesssim\frac{k[n-4\log(k)]}{\log(n)}.$$

D. Moment bounds

Both Theorems 8 and 9 follow from an initial probabilistic statement combined with relatively straightforward counting arguments. These probabilistic statements highlight that it is very unlikely to distinguish random k-design elements from their average with a fixed measurement procedure. Markov's inequality—$\Pr[S\geq\tau]=\Pr\left[S^k\geq\tau^k\right]\leq\mathbb{E}[S^k]/\tau^k$ for non-negative random variables S—reduces this probabilistic assertion to a question about moment growth. The larger the moments we can control, the stronger this assertion becomes. Designs appropriately capture this feature: a k-design accurately approximates Haar-random moments up to order k. This is why designs with growing k become increasingly complex.

For state complexity, the associated Haar-moment computation is relatively straightforward:

$$\mathbb{E}_{|\psi\rangle}\left(\left\{\mathrm{Tr}(M|\psi\rangle\langle\psi|)-\mathbb{E}_{|\psi\rangle}\left[\mathrm{Tr}(M|\psi\rangle\langle\psi|)\right]\right\}^k\right)\leq\left(\frac{k^2}{d}\right)^{k/2},\tag{18}$$

for any fixed measurement M, see e.g., Corollary 24 below.

However, such simple moments do not cover strong unitary complexity. Quantum channels allow for more sophisticated measurement procedures that render the associated Haar-moment computations nontrivial. Our main technical contribution is a novel bound that addresses this setting.

Theorem 10: *Fix a bipartite input state* $|\phi\rangle\in\mathbb{C}^d\otimes\mathbb{C}^d$ *and measurement* M *of compatible dimension. For* U *chosen uniformly from the Haar measure, we have*

$$\mathbb{E}_U\left[\left(Tr\left[M\left(U\otimes\mathbb{I}\right)|\phi\rangle\langle\phi|\left(U^\dagger\otimes\mathbb{I}\right)\right]-\mathbb{E}_U\left\{Tr\left[M\left(U\otimes\mathbb{I}\right)|\phi\rangle\langle\phi|\left(U^\dagger\otimes\mathbb{I}\right)\right]\right\}\right)^k\right]\leq\frac{C_k(k!)^2}{d^{k/2}},$$

where $C_k=[1/(k+1)]\binom{2k}{k}<4^k/k$ *denotes the* k*th Catalan number.*

This bound is considerably more general than existing ones in the literature. Reference [12], for instance, utilizes Eq. (18) only. We establish this result by combining Schur-Weyl duality [37,38] with Weingarten calculus [39,40] and auxiliary arguments from tensor network theory [41,42] and convex optimization [43,44]. We believe that the dimensional scaling in the final bound is tight, but there may be room for further improving the k-dependent constants. In particular, we do not know if the Catalan number is necessary, or merely an artifact of our proof technique.

E. Relation to previous work

Quantum complexity has recently become a popular subject in high-energy physics. A considerable amount of attention has been devoted to understanding the complexity accumulated after an exponentially long time. Works by Susskind and Aaronson [4,8,9] point to an intriguing connection to theoretical computer science: unless PSPACE $\subseteq$ BQP/poly (a hypothetical relation between different computational complexity classes that is widely believed to be false), the circuit complexity of certain Hamiltonian evolutions $U=\exp(-iHt)$ achieves superpolynomial values for exponentially long time scales t. In a similar vein, Bohdanowicz and Brandão [10] constructed a family of Hamiltonians that provably achieves superpolynomial complexity in exponential time, unless PSPACE = BQP.

These arguments address late-time complexity and therefore do not yield insights regarding early-time complexity growth. In this regard, relations between complexity growth and approximate k-designs have recently been pointed out in Refs. [11,45]. Specifically, Ref. [11] defined a notion of the complexity of generating an ensemble of unitaries and gave a lower bound on the ensemble complexity in terms of the distance to forming a unitary design. They also argued that the lower bound of the complexity of a k-design is linear in k. Our arguments and results may be regarded as a substantial refinement of these ideas.

The notion of strong complexity put forward in our work has its conceptual roots in quantum information. Encompassing this mindset is the statement from Ref. [46]: "most states are too entangled to be useful as computational resources." At the core of this argument is the following observation. Haar-random pure states are so highly entangled that *local* measurements yield almost uniformly random outcomes. In turn, any quantum device that relies on local measurements and uses known, but Haar-random, states could be efficiently simulated by tossing classical coins! This prevents any genuine quantum advantage for computation.

Strong state complexity (Definition 2) may be thought as a formal version of this observation. Measuring the maximally mixed state ρ_0 always results in a uniform outcome distribution. Moreover, Ref. [46] makes essential use of the fact that the measurements are constrained to be "simple" (in their case: local measurements augmented by classical postprocessing). The core of their argument may be summarized as follows: low complexity measurements do not allow for distinguishing a Haar-random state from the maximally mixed state. We present a variant of this argument in Appendix A 1 below.

While Ref. [46] considers only Haar-random pure states, similar arguments have been established for states that are less generic, see e.g., Ref. [12, Section 3]. Although not stated at this level of generality, Ref. [12, Corollary 10] effectively points out that states generated by approximate k-designs fool short quantum circuits: with high probability they cannot be distinguished from the maximally mixed state by means of any measurement with small circuit size. They also extend this result to circuits [12, Corollary 11]. With high probability, a randomly selected (according to the weights) k-design element cannot be approximated by any short-sized circuit V in the sense that $\|U - V\|_\infty$ is small.

The second main result of our work, Theorem 9, improves upon this result in two ways. Firstly, the strong unitary complexity (Definition 3) is more stringent than their more traditional definition. While Theorem 9 does imply [12, Corollary 11], the converse is not necessarily true.

Secondly, and more importantly, both Corollaries 10 and 11 in Ref. [12] are probabilistic. While this is enough to deduce average-case behavior, a strong quantitative statement about the number of k-design elements with high circuit complexity remains beyond the scope of these assertions. A worst-case caricature may help to illustrate this subtlety. Suppose that the weights accompanying a unitary k-design are extremely spiky. A single high-complexity unitary, say $U_1 \in U(d)$ is accompanied by an exceedingly large weight $p_1 \simeq 1$, while all other design unitaries U_i have low complexity and almost vanishing weights $p_i \simeq 0$. Such a weight distribution would not contradict the assertion of Ref. [12, Corollary 11]. The single high-complexity circuit occurs with high probability (over the weights). Nonetheless, the hypothetical k-design contains only a single high-complexity element.

Here we overcome this issue by explicitly ruling out the possibility of such extreme cases ever occurring. The definition of an approximate k-design alone implies that the weights cannot be too spiky, see Lemma 3. This bound on the weights allows us to convert probabilistic (average case) statements into quantitative ones. Not only does the average circuit complexity grow linearly with the order k of an approximate design, the absolute number of distinct circuits that have high complexity must also grow *exponentially* with k.

Interest in state complexity has been stimulated by its potential role in quantum gravity and the AdS/CFT correspondence; see Sec. IV for further discussion. Recently, the relevance to holographic duality of *computational pseudorandomness* has been emphasized. Specifically, the authors of Ref. [47] argue that one can construct two *mixed* quantum states on the boundary (A and B) such that both A and B can be efficiently prepared, yet A and B cannot be distinguished from maximally mixed states by polynomial-size quantum circuits. Furthermore, the corresponding bulk states (A' and B') *can* be distinguished efficiently from one another. This observation indicates that the holographic dictionary, which relates bulk and boundary states must have high computational complexity.

We stress that this concept of *pseudorandom quantum states*, which can be efficiently prepared yet cannot be distinguished from random by computationally bounded observers, is applicable to mixed states, or ensembles of pure states, but not to individual pure quantum states. If a particular pure state can be prepared efficiently by a quantum circuit, that state can always be distinguished efficiently from a maximally mixed state by running the circuit backwards. An ensemble of pure states can be pseudorandom only if it contains superpolynomially many pure states, where the observer who draws a sample from the ensemble and attempts to distinguish this sampled state from a maximally mixed state has no information about which sample was chosen. In contrast, in our definition of complexity for pure states, the observer is permitted to use a different distinguishing circuit for each possible pure state. On the other hand, the existence of pseudorandom quantum states [48] indicates that, for mixed states, our definition of state complexity, namely the computational cost of *distinguishing* the state from a maximally mixed state, can differ substantially from another natural definition, the computational cost of *preparing* the state.

III. COMPLEXITY GROWTH IN RANDOM CIRCUITS

The rigorous statements put forward in Theorems 8 and 9 gain additional meaning when applied to concrete

examples. The literature contains several proofs of design growth in random circuits. Combining these with our rigorous insights yields a number of concrete models for complexity growth.

A. Local random circuits

For concreteness, we focus here on systems comprised of n qubits, i.e., $q = 2$ and $d = 2^n$. Let $\mathsf{G} \subset U(4)$ be a (finite) universal gate set containing inverses, i.e., $g^{-1} = g^\dagger \in \mathsf{G}$ whenever $g \in \mathsf{G}$. We can generate G-*local random circuits* by sequentially applying a random gate $g \in \mathsf{G}$ to a randomly selected pair of neighboring qubits. Repeating this procedure independently for T steps results in random circuits of size T. We refer to the application of each gate as a time step, such that size T circuits are of depth T and have thus evolved to time T. Intuitively, the larger T, the more random the circuit becomes. A seminal result by Brandão, Harrow, and Horodecki confirms this intuition in a precise sense.

Theorem 11 (Corollary 7 in Ref. [12]): *Fix $d = 2^n$, $\epsilon > 0$, $k \leq \sqrt{d}$, and let $\mathsf{G} \subset U(4)$ be a universal gate set containing inverses [49]. Then, the set of all G-local random circuits of size T forms an ϵ-approximate k-design if*

$$T \geq Cn\lceil \log_2(k)\rceil^2 k^{9.5}\left[nk + \log\left(1/\epsilon\right)\right], \tag{19}$$

where $C > 0$ is a (large) constant, which depends on G.

We emphasize that the weights associated with each unitary in this ensemble are defined implicitly by this random procedure. Several different T-sized circuits may give rise to the same final unitary, say U_1, while another one, say U_2, may exclusively be obtained from a single circuit geometry. The weights associated with U_1 and U_2 take into account this behavior, i.e., $p_1 \geq 2p_2$ for our example. However, the fact that the entire ensemble still forms an approximate k-design limits potential fluctuations. This in turn imposes lower bounds on the minimal number of distinct unitaries and severely limits the potential for collisions. It cannot be too likely that two or more different random circuits coincide. These features were conjectured by Brown and Susskind [7, Sec. 6.5] who, in turn, base their counting argument that relates circuit size and complexity on an extreme version of this conjecture: *collisions do not occur at all*. One of the main results of this work is rigorous proof for a weaker version of their conjectured relation between circuit size and complexity. It is an immediate consequence of Theorems 9 and 11.

Corollary 12 (Polynomial relation between circuit size and circuit complexity for local random circuits): *Fix $\delta \in (0,1)$, $r \leq 2^{n/2}$ and set $T \geq Cn^2\left[\log_2(n)r/n\right]^{11}$. Then, the set of all G-local circuits of size T contains at least $\tilde{C}n^r$ unitaries that obey $\mathcal{C}_\delta(U) > r$. Here, $C, \tilde{C} > 0$ are constants that implicitly depend on δ and G.*

This result establishes a polynomial relation between the size T of G-local circuits and the strong δ-unitary complexity that may be achieved in such a model [50]. The relation $T \simeq r^{11}$ is a consequence of Theorem 11, which features a similar relation between the degree $2k$ of an approximate $2k$-design and the circuit size T required to implement it. This relation between complexity and circuit size can certainly be improved, which we soon discuss, but there are fundamental limits: a lower bound on the design depth for random circuits is known. A converse result (Proposition 8 in Ref. [12]) states that for $\epsilon \leq 1/4$ and $k \leq d^{1/2}$, the size of random circuits on n qudits must be at least

$$T \geq \frac{2kn\log q}{q^4\log k} \quad \text{to form an } \epsilon\text{-approximate } k\text{-design.} \tag{20}$$

See Appendix C 10 for a rederivation of this claim.

B. Relating two conjectures

Fix $q = 2$, $d = 2^n$ (n qubits) and suppose that the aforementioned lower bound were not only necessary, but also (approximately) sufficient: G-local circuits of size $T \simeq 2nk/\log_2(n)$ generate (sufficiently accurate) approximate $2k$-designs. Under this assumption, G-local random circuits of size T contain at least $d^{2k}/(k!)^2$ elements with circuit complexity $r \simeq T$. If we also assume $k \leq \sqrt{d}$ [$\log_2(k) \leq n/2$], then this bound can be simplified further as

$$\begin{aligned}\frac{d^{2k}}{(k!)^2} &= 2^{2nk - 2\log_2(k!)} \gtrsim 2^{2k[n-\log_2(k)]} \\ &\gtrsim 2^{nk} \simeq 2^{\log_2(n)T} \geq 2^T.\end{aligned} \tag{21}$$

This is essentially Conjecture 1: the set of all G-local circuits of size T contains an exponentially growing set of elements with complexity $r \simeq T$. This observation provides a relation between Conjecture 1 (linear growth in complexity) to an existing conjecture in quantum information [12].

Conjecture 13 (Linear growth in design): *G-local circuits on n qubits of size $T = O(n^2k)$ form approximate k-designs.*

To achieve a linear growth in complexity it suffices to have a linear growth in design.

C. Linear growth in design for local random circuits at large local dimension

We again consider a $1d$ system comprised of n qudits of local dimension q, with total dimension $d = q^n$, and

evolve the system by a random circuit consisting of local 2-site unitaries drawn Haar randomly from $U(q^2)$. The results of Ref. [12] also ensure that such random circuits form approximate k-designs when the size is $O(n^2k^{11})$. Although Conjecture 13, a linear design growth in G-local random circuits with local qubits, is currently out of reach, progress was made recently in Ref. [28], improving the k dependence for Haar-local random circuits in the limit of large local dimension and giving a linear growth in the circuit size to form a unitary k-design.

Theorem 14 ([28]): *Random quantum circuits on n qudits of local dimension q form approximate unitary k-designs when the circuit size is $T = O(n^2k)$ for some $q > q_0$, where q_0 depends on the size of the circuit [51].*

The approach of Ref. [28] was to consider the frame potential, capturing the 2-norm distance to forming an approximate design, and make use of an exact statistical mechanical mapping [52,53] in order to write the frame potential as the partition function of a triangular lattice model. The contributions to the partition function can be interpreted as domain walls in the lattice model. In the limit of large q, Ref. [28] showed that only a simple sector of domain walls contribute, allowing for the calculation of the k-design circuit size. More precisely, by computing the single domain-wall terms and showing that the multidomain wall terms contribute at subleading order in $1/q$, it was proved that local random circuits exhibit a linear growth in design for some $q > q_0$, where q_0 depends on the circuit size T and moment k.

Theorem 14 and Corollary 12 allow us to establish Conjecture 1 for local random circuits with Haar-random 2-site unitaries in the limit of large q.

Corollary 15 (Linear complexity growth): *Given the set of local random circuits of size T at large q, most circuits have strong complexity $\Omega(T)$, i.e., growing linearly in T for a long time.*

Although Theorem 9 still applies for local Haar random quantum circuits, giving a lower bound on the number of distinct unitaries with high complexity, its meaning becomes less clear when we have a continuous ensemble. We can consider an ensemble of finite cardinality by constructing an ε-covering of the set of random circuits. We review the notion of an ε-covering in Appendix C 10 and give a bound on the cardinality of a covering for local random circuits. Constructing a coarse net then shows that exponentially many random quantum circuits, with Haar-random 2-site unitaries, have high complexity.

Recently, an improvement was made in the q dependence of Theorem 14. By studying the spectral gap of the moment operator for random quantum circuits, and using Knabe bounds to bound the spectral gap, it was proven in Ref. [29] that one requires only the local dimension to be $q \geq \Omega(k^2)$ to form unitary designs. While that work explicitly studied circuits with Haar-random 2-local gates, the seminal result in Ref. [54] that the spectral gap is k independent for any set of universal gates G (containing inverses and comprised of algebraic entries), guarantees that the circuit size required to form a k-design for G-local circuits changes only by a constant. This allows us to extend the result to random quantum circuits instead comprised of 2-local gates randomly chosen from G.

Theorem 16 ([29]): *G-local random quantum circuits on n qudits of local dimension q form approximate unitary k-designs for $T \geq O(n^2k)$ when $q \geq 6k^2$.*

Therefore, Theorem 16 and Corollary 12 immediately establish Conjecture 1 for G-local random quantum circuits for $q \geq 6k^2$.

Lastly, we emphasize that we do not prove linear complexity growth up to time scales of order d. While taking a large enough q will ensure linear design growth for times exponential in n, such a limit still pushes the true exponential time scales of interest, $t \sim d = q^n$, out of reach. Proving an optimal design growth for local random circuits away from the large q limit would allow us to better probe late-time complexity.

D. Stochastic quantum Hamiltonians

There also exist continuous-time models of chaotic dynamics, analogous to random circuits, which scramble in $O(\log n)$ time [25]. In a similar spirit to models of random walks on the unitary group, one can define a one-parameter family of Hamiltonians, which generate a time-dependent unitary evolution. The Hamiltonian on n qubits at a time step s is given by a sum of random all-to-all 2-body interactions, meaning we sum over all possible 1- and 2-local interactions with independently chosen Gaussian random couplings

$$H_s = \sum_{i<j} \sum_{\alpha,\beta} J_{s,i,j,\alpha,\beta} S_i^\alpha S_j^\beta, \tag{22}$$

where S_i^α is a Pauli operator acting on site i with $\alpha = \{0,1,2,3\}$. The couplings are each drawn independently from a Gaussian distribution with zero mean and variance σ^2. Not only are the couplings random in space, but are further chosen randomly at each time step s. The time evolution to time t is then given by

$$U_t = \prod_{s=1}^{t} e^{-iH_s \delta t}, \tag{23}$$

where we consider the continuum limit $\delta t \to 0$ with the variance of the couplings scaling as $\sigma^2 = J/\delta t$ so that the

interactions strength increases proportionally to the inverse time step and where J is a constant.

It was shown in Ref. [26], using similar techniques to Ref. [12], that these stochastic quantum Hamiltonians (also called Brownian circuits) form k-designs in polynomial time.

Theorem 17 (Corollary 10 in Ref. [26]): *For $d = 2^n$ and $\epsilon > 0$, the ensemble of time evolutions by stochastic Hamiltonians in Eq. (22), forms an ϵ-approximate k-design for times*

$$t \geq C\lceil \log_2(k)\rceil^2 k^{9.5}[nk + \log(1/\epsilon)], \tag{24}$$

where $C > 0$ is a constant.

For the Brownian circuit models, the constant prefactor C depends on the local dimension, here chosen to be 2, but also on the interaction strength of the couplings J, $C \sim 1/J$, meaning if the interactions are stronger then the depth required to form a design decreases accordingly.

We can again use the polynomial relation between complexity and design to discuss complexity growth. Theorems 9 and 17 together give that Brownian circuits have a complexity growing polynomially in time as $\Omega(t^{1/11})$.

E. Nearly time-independent Hamiltonian dynamics

There is another random quantum circuitlike construction of a time-dependent Hamiltonian with varying couplings over discrete time steps. This "nearly time-independent" model of Ref. [27] forms k-designs in a circuit size $O(n^2k)$, for moments up to $k = o(\sqrt{n})$, achieving the nearly optimal lower bound with a linear growth in design for a short time.

Consider a $1d$ system of n qudits, with $d = q^n$, and define a time-dependent set of random couplings

$$\mathcal{J}(t,g) = \left\{\lambda/(\lfloor t/2\rfloor + 1), \lambda \in [-g/2, g/2]\right\}, \tag{25}$$

where λ is drawn uniformly at random from the interval. We now generate two ensembles of Hamiltonians with time-dependent couplings

$$\begin{aligned}\mathcal{E}_Z(t) &= \left\{ -\sum_{j<k} h_{jk} Z_j Z_k - \sum_j b_j Z_j \right\},\\ \mathcal{E}_X(t) &= \left\{ -\sum_{j<k} h_{jk} X_j X_k - \sum_j b_j X_j \right\},\end{aligned} \tag{26}$$

with $h_{jk} \in \mathcal{J}(t,h)$ and $b_j \in \mathcal{J}(t,b)$, and where $h = \lfloor t/2\rfloor/2$ and $b = \lfloor t/2\rfloor + 1/2$. We then define the time evolution of our system: we evolve by an X-type Hamiltonian $H_X \sim \mathcal{E}_X$ at even time steps and a Z-type Hamiltonian $H_Z \sim \mathcal{E}_Z$ at odd time steps. Then the unitary time evolutions form an ϵ-approximate k-design for $k = o(n^{1/2})$, after T time steps, where

$$T \geq [k + 1/2 + (1/n)\log_2(1/\epsilon)], \tag{27}$$

where each time step involves $O(n^2)$ gates.

This construction forms unitary k-designs almost linearly in time, with the caveat that the time scale is limited to approximately $\sqrt{n}$. Thus we get a linear growth in design at early times, but not exponentially in n. Consequently, this implies a linear growth in complexity at (very) early times.

F. Comment on time-independence

We discuss a few explicit models of complexity growth in systems that are random in both space and time. As we emphasize, one of our results is that a polynomial growth in design implies a polynomial growth in complexity (Corollary 4). Thus, the random circuit and Brownian circuit models, which form approximate k-designs in poly(k) depth, are also explicit examples of systems with a long-time polynomial growth in complexity.

But for an ensemble of time evolutions to form a k-design, randomness in time is likely essential. For instance, consider an ensemble of time evolutions generated by an ensemble of Hamiltonians, $\mathcal{E}_t = \{e^{-iHt},\ H \in \mathcal{E}_H\}$, where $\mathcal{E}_H$ could be a disordered spin system or any ensemble of random Hermitian matrices. The rigid structure of eigenvalues then prohibits the late-time Haar randomness.

Interestingly, the Gaussian unitary ensemble (GUE), an ensemble of $d \times d$ random Hermitian matrices with a unitarily-invariant measure, does come close to an approximate k-design in 2-norm for moments $k \ll d$ at a specific time scale $t \sim \sqrt{d}$ [45]. But at later times, the 2-norm distance between the ensemble of unitaries generated by GUE Hamiltonians and the Haar ensemble becomes large. More generally, one expects that any ensemble of unitary evolutions generated by time-independent Hamiltonians will not form a k-design at late times. A general argument for this is as follows [11], under the exponential map $\lambda \to e^{i\lambda t}$, the eigenvalues of a Hamiltonian are distributed as time-evolving phases on the unit circle. In the limit $t \to \infty$, the phases become uncorrelated and uniformly distributed, unlike the correlated and logarithmically repelling eigenvalues of Haar-random unitaries. Thus, to understand the complexity growth of (ensembles of) time-independent Hamiltonian evolution, we would need to look beyond their design properties for an alternative approach, for instance, by studying the approximate invariance of the ensemble [45,55].

IV. COMPLEXITY IN HOLOGRAPHIC SYSTEMS

Much of the recent interest in quantum complexity in the high-energy literature has centered on the conjectured relation between complexity growth and the long-time growth of black-hole interiors [4,5,56]. More specifically in the context of the AdS/CFT correspondence, the region behind the horizon of an eternal AdS-Schwarzschild black hole grows linearly in time for an exponential time ($t \sim e^n$). The holographic picture is a two-sided geometry connected by a wormhole, where the throat of the wormhole is growing in time. The claim is that the quantum complexity of the dual CFT state is the long-time linearly increasing quantity, which captures the wormhole growth. There have been a number of proposals for what bulk quantity actually computes the complexity, including the volume and action of the AdS wormhole. The complexity/volume conjecture states that the computational complexity of the boundary state is equal to the volume of the wormhole. More precisely, the complexity of time-evolved thermofield double state of the two boundary CFTs is equal to spatial volume behind the horizon of the two-sided geometry on a maximal time slice [5]. The "complexity equals action" conjecture posits that the action computed on a certain region of the bulk geometry, which extends behind the horizon (the Wheeler-DeWitt patch), is the quantity, which is dual to the complexity [6,57]. A lot of progress has been made checking these conjectures and studying complexity growth in holographic systems, see, for instance, [58–64].

In this work we rigorously compute the complexity growth in a number of random circuit models, by relating the growth in design to the growth of complexity, and are able to prove a linear growth in complexity for local random circuits in the limit of large local dimension (albeit, not for an exponentially long time). As we mention, the connection between unitary designs and quantum complexity will likely not inform complexity growth in holography as evolution by time-independent Hamiltonians will not converge to approximate designs. Thus, to study complexity growth in holography we need to explore properties beyond the Haar randomness of the evolution.

A. Strong complexity in the bulk

We briefly discuss why we believe our proposed strong definition of complexity (in terms of a distinguishing measurement), is congruent with expectations from the bulk and might be more suited for holography than the standard definition in terms of the circuit complexity.

One feature we expect complexity growth will exhibit in holography, and fast scrambling systems more generally, is the switchback effect [5]. Consider a time-evolved local operator $\mathcal{O}(t) = e^{-iHt}\mathcal{O}e^{iHt}$ (sometimes called a precursor), where $\mathcal{O}$ might be a single-site Pauli. For such an operator, we anticipate a delay in the onset of the linear complexity growth. For the traditional definition of complexity, consider the minimal circuit approximating the evolution operator e^{-iHt}. The reason for this delay is the exact cancellation of gates outside the lightcone of the spreading operator. Once the operator grows to be the size of the system (more precisely, to have support on weight n Pauli operators) after a time scale called the scrambling time, we expect the complexity of $\mathcal{O}(t)$ to begin its long time linear growth. Such an effect is also present in the bulk for both complexity-volume and action conjectures. This feature is also present in complexity growth of $\mathcal{O}(t)$ under the strong definition of complexity in Definition 2. To be concrete, consider a system of n qubits and the evolved state $e^{-iHt}\mathcal{O}e^{iHt}|\psi_0\rangle$, where H is a chaotic but local Hamiltonian and we take $|\psi_0\rangle$ to be an unentangled product state. Prior to the scrambling time, the optimal measurement to distinguish the evolving state from the maximally mixed state is a simple measurement of a qubit outside the lightcone of the evolving operator. It is not until the scrambling time, when operator has grown to have support on all sites, that the complexity of the distinguishing measurement begins to grow.

Another interesting expectation from holographic systems, where the strong and weak definitions of complexity differ, is that of one-clean qubit. This is essentially the argument given in Lemma 6, to prove that measurement complexity is a stronger definition than standard circuit complexity. Consider a simple initial state $|\psi_0\rangle$, which has been evolved for an exponential time such that $|\psi(t)\rangle$ is maximally complex. If we add a single unentangled qubit to the state $|\psi(t)\rangle \otimes |0\rangle$, then the minimal circuit complexity will be unchanged, but maximal potential complexity increases and the complexity of the state can continue to grow for a long time until it saturates at the new maximal value. For the complexity of a distinguishing measurement, adding a single clean qubit resets the complexity to an order-one value, as the optimal measurement is simply the projection onto the clean qubit. Reference [7] proposed the notion of uncomplexity as the difference of the complexity of a state or unitary from its maximal complexity and suggested an interpretation in the bulk as the total spacetime volume accessible to an infalling observer. Uncomplexity can be thought of as a resource to do useful computation. As we describe, our strong definition of complexity directly encodes this potential for useful quantum computation.

B. Entanglement growth by design

The suggestion that complexity be the dual of the long-time geometric growth in the bulk was motivated by the observation that the wormhole grows long past the times cales at which entropic quantities saturate. Given that we discuss long-time growth in complexity from a long-time growth in design, it is worth commenting on the saturation of entropies after a short growth in design order.

The entanglement entropies for k-designs were studied in Ref. [65]. Specifically, they looked at the Rényi-α entropies of a density matrix ρ: $S^{(\alpha)}(\rho) = [1/(1-\alpha)]\log[\mathrm{Tr}(\rho^\alpha)]$. For any state, the Rényis are bounded above and below by the min-entropy $S_{\min}(\rho) := \lim_{\alpha\to\infty} S^{(\alpha)}(\rho) = -\log(\|\rho\|_\infty)$ [66]. For an n-qubit system, consider the reduced density matrix $\rho_A = \mathrm{Tr}_{\bar{A}}|\psi\rangle\langle\psi|$ on a subsystem A consisting of half the qubits, so that $d_A = d_{\bar{A}}$. Reference [65] showed that for states $|\psi\rangle$ drawn from a $(k > \log d)$ design, the min-entropy of ρ_A is nearly maximal. Therefore, all entropies are nearly maximal once the design order is $k \approx n$. Considering then the time-evolved states of a fast-scrambling system, which forms unitary designs linearly in time, all entropies will saturate at a time of order n. Our arguments ensure complexity growth of approximate k-designs well beyond this entropy saturation threshold.

V. DISCUSSION

We rigorously establish a growth of the quantum complexity in the time evolution of a number of models. We prove that with overwhelming probability, an element sampled from an approximate unitary k-design has a strong complexity that scales at least linearly in k. Moreover, we can count the elements of a design of a given complexity and show that there are at least an exponential number (in k) of distinct unitaries with this complexity. Using the known relations between the evolution time and the design order k thereby establishes a lower bound on the growth of quantum complexity. Specifically, for random quantum circuits we make substantial progress on conjectures by Brown and Susskind and, using a recently established linear relation between the circuit size and design order, prove a linear growth of quantum complexity.

A number of open questions remain. For one, the results in Refs. [28,29] required taking the local dimension q to be large in a k-dependent manner. For local qubits, $T = O(n^2k^{11})$ is still the best known design depth. A proof of a linear design growth for random quantum circuits on qubits up to exponentially high moments would prove a linear growth of complexity for exponentially long times. In this work we largely focus on time-dependent evolution, but the original discussion of a long-time linear complexity growth in holographic systems was focused on time-independent Hamiltonian evolution. It remains to be seen if one can prove anything about the complexity $\mathcal{C}_\delta(e^{-iHt})$ for a specific many-body Hamiltonian H. Lastly, we largely focus on the growth regime for complexity. Nevertheless, there are a number of interesting questions at exponentially late times, when $t \geq d^2$ and complexity saturates at its maximal value.

As we emphasize, our results hold for a new and stronger notion of quantum complexity, defined in terms of optimal distinguishing measurements. We believe strong complexity to be more aptly suited for complexity in holography than circuit complexity, mirroring expectations from the bulk. More broadly, it would be interesting to explore the implications of our strong definition of complexity for quantum error correction and topological order.

ACKNOWLEDGMENTS

The authors thank Dorit Aharonov, Thom Bohdanowicz, Elizabeth Crosson, Felix Haehl, Aram Harrow, Tomas Jochym-O'Connor, Hugo Marrochio, Grant Salton, Eugene Tang, Thomas Vidick, and Beni Yoshida for inspiring discussions and valuable feedback. We also thank the anonymous reviewers for detailed comments and suggestions. All authors acknowledge funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant No. PHY-1733907). J.P. is supported in part by DOE Award No. DE-SC0018407 and by the Simons Foundation It from Qubit Collaboration. R.K. is supported in part by the Office of Naval Research (Award No. N00014-17-1-2146) and the Army Research Office (Award No. W911NF121054). N.H.J. thanks the IQIM at Caltech, McGill University, and UC Berkeley for their hospitality during the completion of this work. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

APPENDIX A: PROOF OF THE MAIN RESULTS

1. Motivating example computations for Haar-random states

In this section, we provide valuable intuition by analyzing the complexity of Haar-random states using concentration of measure (Levy's lemma). The results presented in the main text will follow by replacing Haar-random states and unitaries with approximate k-designs and measure concentration with moment bounds. Moment bounds, however, are considerably weaker than measure concentration. This, in particular, affects constants and subleading contributions.

a. Most states have high complexity

The Hilbert space of n qudits is enormous, $d = q^n$. Nonetheless, only a tiny fraction of all possible (pure) quantum states seems to be useful for quantum computation, see, e.g., Ref. [46]. Strong state complexity (Definition 2) captures this curious aspect. In order to get a quantitative handle on the set of all pure states we endow it with the uniform measure $d\psi$ that is induced by the Haar measure on the unitary group $U(d)$. Then, random pure states $|\psi\rangle\langle\psi|$ behave like the maximally mixed state $\rho_0 = \mathbb{I}/d$ in expectation. This behavior extends to the

outcome statistics of arbitrary (fixed) measurements:

$$\mathbb{E}_{|\psi\rangle}\left[\mathrm{Tr}\left(M|\psi\rangle\langle\psi|\right)\right] = \mathrm{Tr}\left(M\mathbb{E}_{|\psi\rangle}\left[|\psi\rangle\langle\psi|\right]\right) = \mathrm{Tr}\left(M\rho_0\right). \tag{A1}$$

Concentration of measure (Levy's lemma) ensures that deviations from this average case behavior are exponentially suppressed in concrete instances:

$$\begin{aligned}&\Pr\{|\mathrm{Tr}\left[M(|\psi\rangle\langle\psi| - \rho_0)\right]| \geq \tau\}\\ &\quad\leq 2\exp\left(-\frac{d\tau^2}{9\pi^3}\right) \quad \text{for any } \tau \geq 0.\end{aligned} \tag{A2}$$

We refer to Proposition 29 in Appendix D below for a proof of this well-known result. We can combine this assertion with a union bound (Boole's inequality) to conclude for any $r \in \mathbb{N}$ and $\delta \in (0,1)$

$$\begin{aligned}&\Pr\left[\mathcal{C}_\delta(|\psi\rangle) \leq r\right]\\ &= \Pr\left\{\max_{M\in\mathsf{M}_r} |\mathrm{Tr}\left[M(|\psi\rangle\langle\psi| - \rho_0)\right]| \geq 1 - d^{-1} - \delta\right\}\\ &\leq 2|\mathsf{M}_r|\exp\left(-\frac{d(1-d^{-1}-\delta)^2}{9\pi^3}\right)\\ &\leq 2.0072|\mathsf{M}_r|\exp\left(-\frac{d(1-\delta)^2}{9\pi^3}\right).\end{aligned} \tag{A3}$$

Suppose that M_r arises from combining at most r elements of a fixed universal gate set $\mathsf{G} \subset U(q^2)$. A naive counting argument reveals $|\mathsf{M}_r| \leq 2d(n+1)^r|\mathsf{G}|^r$ and we refer to Appendix B 2 below for details. We conclude that the $\Pr\left[\mathcal{C}_\delta(|\psi\rangle) \leq r\right]$ remains exponentially suppressed (in $d = q^n$) until

$$r \simeq \frac{q^n}{\log(n)}. \tag{A4}$$

To summarize, a random state is exceedingly likely to have an exponentially large strong δ-state complexity.

The Haar measure has another desirable property. It is fair in the sense that it assigns the same (infinitesimal) weight to each pure state. Such perfectly flat probability distributions allow for turning the probabilistic statement, Eq. (A3), into a quantitative one. From the definition of probability, $\Pr\left[C_\delta(|\psi\rangle) \leq r\right]$ corresponds to the ratio of low-complexity states over all states. Thus, Eq. (A3) ensures that the fraction of low-complexity states remains exponentially tiny until $r \simeq q^n/\log(n)$. In other words, *most pure states have exponentially large complexity.*

b. Most high-complexity states are far apart

In the previous subsection, we saw that concentration of measure, Eq. (A2), allows us to conclude that most quantum states have exponentially high state complexity. This argument, however, does not (yet) tell us anything about the geometric separation between high-complexity states. In principle, a large fraction of high-complexity states could result from tiny perturbations of only a few well-separated core states that have high complexity each. In other words, high-complexity states could come in few tightly packed clusters, in which case the effective number of high-complexity regions could still be comparatively small.

The probabilistic method [67] allows us to prove that extreme clustering cannot occur: *there are exponentially many high-complexity states whose pairwise distance is almost maximal.*

We show this statement by induction based on two features of Haar-random states. Firstly, we use the main result from the previous subsection. Choose $r \lesssim q^n/\log(n)$ such that Eq. (A3) ensures

$$\Pr\left[\mathcal{C}_\delta(|\psi\rangle) \leq r\right] \leq 2.0072|\mathsf{M}_r|\exp\left(-\frac{d(1-\delta)^2}{9\pi^3}\right) < \frac{1}{2}. \tag{A5}$$

The parameter r is chosen such that Haar-random states exceed this complexity with probability (at least) $1/2$. Concentration of measure also implies that a Haar-random state is very likely to be far away from any fixed state $|\phi\rangle\langle\phi|$. For any $\Delta \in (0,1)$,

$$\begin{aligned}&\Pr\left[\frac{1}{2}\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \leq 1-\Delta\right]\\ &\quad= \Pr\left[|\langle\psi|\phi\rangle|^2 \geq \Delta^2\right] \leq 3\exp\left(-\frac{\Delta^2 d}{9\pi^3}\right).\end{aligned} \tag{A6}$$

This bound readily follows from Eq. (A2) (set $M = |\phi\rangle\langle\phi|$ and perform elementary modifications).

The first step in our inductive argument is simple. Equation (A5) asserts that the probability of Haar randomly sampling a low complexity (at most r) state is smaller than $1/2$. This is equivalent to stating that the probability of Haar randomly sampling a high complexity (larger than r) is at least $1/2$. Importantly, this assertion implies that such a state exists, because the probability of sampling one is strictly positive. Choose one such state $|\phi_1\rangle$ as the first state in our list.

To construct the second state in our list, we refine this probabilistic existence argument. The probability of Haar randomly sampling a low-complexity state *or* a state that

is too close to $|\phi_1\rangle$ is bounded by

$$\begin{aligned}\Pr\left[\mathcal{C}_\delta(|\psi\rangle) \le r \cup \frac{1}{2}\||\psi\rangle\langle\psi| - |\phi_1\rangle\langle\phi_1|\|_1 \le 1-\Delta\right] \\ \le \Pr\left[\mathcal{C}_\delta(|\psi\rangle) \le r\right] + \Pr \\ \times \left[\frac{1}{2}\||\psi\rangle\langle\psi| - |\phi_1\rangle\langle\phi|\|_1 \le 1-\Delta\right] \\ < \frac{1}{2} + 3\exp\left(-\frac{\Delta^2 d}{9\pi^3}\right). \end{aligned} \tag{A7}$$

By contraposition, the probability of sampling a state that has high complexity *and* is simultaneously far away from $|\phi_1\rangle$ is at least $\frac{1}{2} - 3\exp[-(\Delta^2 d/9\pi^3)] > 0$. This implies the existence of such a state. Choose one such state $|\phi_2\rangle$ and append it to the list: $\{|\phi_1\rangle, |\phi_2\rangle\}$.

We can now inductively repeat this probabilistic existence argument and iteratively append distant high-complexity states to the list $\{|\phi_1\rangle, \ldots, |\phi_N\rangle\}$. This construction only breaks down once the list cardinality N counterbalances exponential suppression: $\frac{1}{2} - 3N\exp[-(\Delta^2 d/9\pi^3)] \le 0$, or equivalently $N \ge \frac{1}{6}\exp[(\Delta^2 d/9\pi^3)]$. Beyond this threshold, we cannot use simple union bounds and concentration of measure to ensure existence of another list element. Such a threshold, however, scales exponentially in the Hilbert-space dimension: the list $\{|\phi_1\rangle, \ldots, |\phi_N\rangle\}$ contains $N = \frac{1}{6}\exp[(\Delta^2 d/9\pi^3)]$ high-complexity states whose pairwise trace distance is at least $1-\Delta$.

We conclude this subsection with providing a bit of context. Existence proofs based on strictly positive probabilities date back to Erdős who developed them to solve several important problems in graph theory. Today, this technique is known as the *probabilistic method* and does constitute an important tool in applied math, combinatorics, and theoretical computer science [67].

2. Proof of Theorem 8

Haar-random states result from applying a Haar-random unitary $U \in U(d)$ to an arbitrary starting state, say $|\psi_0\rangle$. Now suppose that this unitary U is not chosen from the Haar measure, but from an approximate $2k$-design. By definition, this ensures that the first $2k$ moments of $|\psi\rangle\langle\psi| = U|\psi_0\rangle\langle\psi_0|U^\dagger$ accurately approximate the corresponding Haar moments. While this is too coarse to deduce exponential concentration, Eq. (A2), (this would require matching behavior for *all* moments), polynomial concentration arguments do apply. Fix a measurement $M \in \mathbb{H}_d$ and let $\bar{M} = M - [\mathrm{Tr}(M)/d]\mathbb{I}$ denote its traceless part. Markov's inequality then implies that for any $\tau > 0$

$$\begin{aligned}\Pr\{\left|\mathrm{Tr}\left[M(|\psi\rangle\langle\psi| - \rho_0)\right]\right| \ge \tau\} \\ = \Pr\{\left[\mathrm{Tr}\left(\bar{M}|\psi\rangle\langle\psi|\right)\right]^{2k} \ge \tau^{2k}\} \\ \le \tau^{-2k}\mathbb{E}\left[\mathrm{Tr}\left(\bar{M}|\psi\rangle\langle\psi|\right)^{2k}\right]. \end{aligned} \tag{A8}$$

The final expectation value corresponds to a moment of order $2k$. This is the largest moment that still approximately exhibits Haar-random behavior. Explicit bounds can be derived by exploiting this similarity and we refer to Corollary 24 below for a technical derivation:

$$\Pr\{|\mathrm{Tr}\left[M(|\psi\rangle\langle\psi| - \rho_0)\right]| \ge \tau\} \le (1+\epsilon)\left(\frac{2k}{\tau\sqrt{d}}\right)^{2k}. \tag{A9}$$

Qualitatively, this deviation bound is proportional to d^{-k} and becomes ever more stringent as the design order $2k$ increases. We can now combine this tail bound with a union bound and a counting argument for the measurement set M_r in a fashion analogous to the Haar-random case. For any $r \in \mathbb{N}$ and any $\delta \in (0,1)$ this yields

$$\begin{aligned}\Pr\left[\mathcal{C}_\delta(|\psi\rangle) \le r\right] &\le |\mathsf{M}_r|(1+\epsilon)\left(\frac{2k}{\sqrt{d}\left(1-d^{-1}-\delta\right)}\right)^{2k} \\ &\le 2(1+\epsilon)d(n+1)^r|\mathsf{G}|^r\left(\frac{16k^2}{d(1-\delta)^2}\right)^k, \end{aligned} \tag{A10}$$

where we tacitly assume $(1-\delta) \ge 2d^{-1}$ in the last step. Qualitatively, this probability remains tiny until

$$r \simeq \frac{(k-1)n - 2k\log(k)}{\log(n) + \log(|\mathsf{G}|)} \simeq \frac{k[n - 2\log(k)]}{\log(n)}, \tag{A11}$$

provided that $n \ge |\mathsf{G}|$ and $k < d/2$. We can compare this to the complexity of Haar-random states in Eq. (A4). Note that the two coincide when we consider designs of exponentially large degree. So far, this is a purely probabilistic statement. In contrast to the Haar-uniform case it is *a priori* not clear whether it is possible to transform it into a quantitative one. The reason for this is twofold: (i) the weights p_j associated with different elements from an approximate $2k$-design are typically *not* uniform. This nonuniformity extends to the distribution over the different states $|\psi_i\rangle$; (ii) collisions in the state generation: two (or more) distinct design unitaries can produce the same state.

Fortunately, the defining properties of designs ensure that these deviations cannot be too radical: the weights associated with *distinct* states $|\psi_i\rangle$ must obey $q_j \le (1+\epsilon)\binom{d+k-1}{k}^{-1}$—see Lemma 21 in Appendix C 6 below (or, equivalently, Lemma 3 in the main text). This extra condition does allow for drawing quantitative conclusions. Recall that the probability of an event E is the expected

value of its indicator function $\mathbb{1}\{E\}$. Therefore,

$$\Pr[\mathcal{C}_\delta(|\psi\rangle) > r] = \sum_j q_j \mathbb{1}\{\mathcal{C}_\delta(|\psi\rangle) > r\} \leq (1+\epsilon)\binom{d+k-1}{k}^{-1} \sum_j \mathbb{1}\{\mathcal{C}_\delta(|\psi\rangle) > r\}. \tag{A12}$$

The sum on the rhs is simply the cardinality N of the set of states $|\psi\rangle$ with δ-state complexity greater than r and the lhs is $1 - \Pr[\mathcal{C}_\delta(|\psi\rangle) \leq r]$. Substituting the bound, Eq. (A10), into this expression establishes the claim:

$$N \geq \binom{d+k-1}{k} \times \left[\frac{1}{1+\epsilon} - 2d(n+1)^r|\mathsf{G}|^r\left(\frac{16k^2}{d(1-\delta)^2}\right)^k\right]. \tag{A13}$$

3. Proof of Theorem 9

The proof is largely analogous to the proof of Theorem 8. Fix a measurement $M \in \mathbb{H}_d \otimes \mathbb{H}_d$ and an input state $|\phi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. Recall that the bias of distinguishing a unitary channel $\mathcal{U} : \mathbb{H}_d \to \mathbb{H}_d$ from the depolarizing channel $\mathcal{D}$ via this measurement procedure is $\mathrm{Tr}[M(\mathcal{U}\otimes\mathcal{I} - \mathcal{D}\otimes\mathcal{I})(|\phi\rangle\langle\phi|)]$. Moreover, the depolarizing channel corresponds to the Haar average over all unitary channels: $\mathbb{E}_U(\mathcal{U}) = \mathcal{D}$, see, e.g., Lemma 26 in Appendix C 9 below. Now suppose that the corresponding unitary $U \in U(d)$ is chosen randomly from an ϵ-approximate $2k$-design. Markov's inequality yields

$$\Pr\{|\mathrm{Tr}[M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)] - \mathrm{Tr}[M\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)]| \geq \tau\} \leq \tau^{-2k}\,\mathbb{E}\Big(\{\mathrm{Tr}[M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)] - \mathrm{Tr}[M\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)]\}^{2k}\Big). \tag{A14}$$

The final expectation value corresponds to the highest $2k$-design moment that still approximates Haar-random behavior. Our main technical contribution in Theorem 10 establishes tight bounds on such Haar-random moments. These generalize approximate $2k$-design ensembles $\mathcal{E}$ in a relatively straightforward fashion:

$$\mathbb{E}_\mathcal{E}\Big(\{\mathrm{Tr}[M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)] - \mathrm{Tr}[M\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)]\}^{2k}\Big) \leq \frac{[(2k)!]^2}{d^k}\left(C_{2k} + \frac{\epsilon}{(2k)!d^{3k}}\right). \tag{A15}$$

See Corollary 23 in Appendix C 8 below for a precise statement and proof. Next, we emphasize that the crude bound $|\mathsf{M}_r| \leq (2d^2+1)n^{2r}|\mathsf{G}|^r$ applies to circuit measurements. Combining the above concentration inequality with a union bound over all measurements $M \in \mathsf{M}_r$ ensures

$$\Pr[\mathcal{C}_\delta(U) \leq r] \leq 3\left(C_{2k} + \frac{\epsilon}{(2k)!d^{3k}}\right)d^2n^{2r}|\mathsf{G}|^r\left(\frac{64k^4}{d(1-\delta)^2}\right)^k, \tag{A16}$$

where we tacitly assume $(1-\delta) \geq 2d^{-1}$. Qualitatively, this probability remains tiny until

$$r \lesssim \frac{(k-2)[n-4k\log(k)]}{\log(n)+\log|\mathsf{G}|} \simeq \frac{k[n-4\log(k)]}{\log(n)}, \tag{A17}$$

provided that $n \geq |\mathsf{G}|$ and $k \leq d/2$. The definition of an approximate $2k$-design also imposes constraints on the weight fluctuations. Lemma 3 asserts that weights associated with distinct ensemble unitaries must obey $p_j \leq (1+\epsilon)(k!/d^{2k})$. This approximate flatness allows us to turn the probabilistic statement from above into a quantitative one:

$$\Pr[\mathcal{C}_\delta(U) > r] = \sum_j p_j \mathbb{1}\{\mathcal{C}_\delta(U) > r\} \leq (1+\epsilon)\frac{k!}{d^{2k}}\sum_j \mathbb{1}\{\mathcal{C}_\delta(U) > r\}. \tag{A18}$$

The sum on the right counts the cardinality N of distinct unitaries with δ-unitary complexity at least $r+1$, while the lhs may be lower bounded by Eq. (A16):

$$N \geq \frac{d^{2k}}{k!}\left[\frac{1}{1+\epsilon} - 3d^2n^{2r}|\mathsf{G}|^r\left(\frac{1024k^4}{d(1-\delta)^2}\right)^k\right]. \tag{A19}$$

4. Distant and distinct design elements

We show that unitary and state designs contain an exponential number $[\Omega(d^k)]$ of distinct high-complexity elements. But to really capture the ergodic nature of chaotic evolution over the unitary group, these distinct high-complexity elements should be pairwise far apart. Here we address this subtlety and show that unitary and state designs contain an exponential number of distant high-complexity unitaries or states.

a. Distant and distinct state design elements

Consider an element drawn at random from an ϵ-approximate spherical k-design $|\psi\rangle$. Equation (A10) gives that the probability the state has δ-state complexity less than r, $\mathcal{C}_\delta(|\psi\rangle) \leq r$, is bounded to be $O(d^{-k})$ when $r \lesssim kn$. We can also show that the probability an element drawn at random from an ϵ-approximate spherical k-design is close to a fixed reference state $|\phi\rangle$ is polynomially suppressed in

k. Choose $\Delta \in (0,1)$ and combine $\frac{1}{2}\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 = \sqrt{1-|\langle\psi,\phi\rangle|^2}$ with Markov's inequality to conclude

$$\begin{aligned}\Pr\left[\frac{1}{2}\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \le 1-\Delta\right] \\ = \Pr\left[|\langle\psi,\phi\rangle|^2 \ge \Delta^2\right] = \Pr\left[|\langle\psi,\phi\rangle|^{2k} \ge \Delta^{2k}\right] \\ \le \Delta^{-2k}\mathbb{E}_{|\psi\rangle}\left[|\langle\psi,\phi\rangle|^{2k}\right] \le \frac{1+\epsilon}{\Delta^{2k}}\binom{d+k-1}{k}^{-1}.\end{aligned} \tag{A20}$$

The last inequality follows from a k-design moment bound similar to Eq. (18). We refer to the proof of Lemma 21 in Appendix C 6 below for a detailed derivation. Qualitatively, this bound is of order $O(d^{-k})$. We can now use a union bound to limit the probability of a random k-design state to have either low complexity *or* to be close to the reference state,

$$\begin{aligned}\Pr\left[\mathcal{C}_\delta(|\psi\rangle) \le r \cup \frac{1}{2}\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \le 1-\Delta\right] \\ \le \Pr\left[\mathcal{C}_\delta(|\psi\rangle) \le r\right] + \Pr \\ \times\left[\frac{1}{2}\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \le 1-\Delta\right] \\ \le 2(1+\epsilon)dn^r|\mathsf{G}|^r\left(\frac{16k^2}{d(1-\delta)^2}\right)^k \\ + \frac{1+\epsilon}{\Delta^{2k}}\binom{d+k-1}{k}^{-1}.\end{aligned} \tag{A21}$$

As long as $r \lesssim nk$, this bound is also of order $O(d^{-k})$ and, in turn, strictly smaller than one. We know that if the probability of the state having low complexity or being close to our fixed state is strictly less than 1, then there is a nonzero probability of a design element that is of high complexity and is far away from the fixed state. Simply stated, if $\Pr[A \cup B] < 1$ then $\Pr[\bar{A} \cap \bar{B}] > 0$.

We can iterate this procedure to construct a set of high-complexity states that are pairwise separated. As long as the probability that the design element is of low complexity or is close to all elements of the set is less than one, then there exists a design element, which is of high complexity and far away from all other design elements in the set. To construct the set $\{|\psi_1\rangle,\ldots,|\psi_N\rangle\}$, we simply need that

$$\begin{aligned}\Pr\left[\mathcal{C}_\delta(|\psi_N\rangle \le r \bigcup_{i=1}^{N-1}\frac{1}{2}\||\psi_N\rangle\langle\psi_N| - |\psi_i\rangle\langle\psi_i|\|_1 \le 1-\Delta\right] \\ < 1.\end{aligned} \tag{A22}$$

A union bound then converts this requirement into the following sufficient condition on the set cardinality N:

$$N < \Delta^{2k}\binom{d+k-1}{k}\left[\frac{1}{1+\epsilon} - 2dn^r|\mathsf{G}|^r\left(\frac{16k^2}{d(1-\delta)^2}\right)^k\right]. \tag{A23}$$

For constant $\Delta \in (0,1)$, this threshold is exponential as long as the complexity obeys $r \lesssim k$,

$$N \approx O(d^k) \quad \text{for } \mathcal{C}_\delta(|\psi\rangle) \le r \approx k. \tag{A24}$$

We note the similarity of this bound to the bound derived for the number of distinct design elements.

a. Distant and distinct unitary design elements

Now we consider a unitary U drawn from an ϵ-approximate unitary k-design $\mathcal{E}$. Equation (A16) bounds the probability of the unitary having δ-unitary complexity less than r, $\mathcal{C}_\delta(U) \le r$, to be $O(d^{-2k})$ when the complexity is roughly $r \lesssim nk$.

Randomly chosen k-design elements also tend to land far away from any fixed unitary. For some $V \in U(d)$ and $\Delta \in (0,1)$, Markov's inequality implies

$$\begin{aligned}\Pr\left[|\mathrm{Tr}(U^\dagger V)|^2 \ge d^2\Delta^2\right] = \Pr\left[|\mathrm{Tr}(U^\dagger V)|^{2k} \ge d^{2k}\Delta^{2k}\right] \\ \le \frac{\mathbb{E}_{\mathcal{E}}\left[|\mathrm{Tr}(U^\dagger V)|^{2k}\right]}{d^{2k}\Delta^{2k}} \le \frac{1+\epsilon}{\Delta^{2k}}\frac{k!}{d^{2k}},\end{aligned} \tag{A25}$$

where the last inequality follows from a k-design moment bound. We refer to the proof of Lemma 20 in Appendix C 6 below for a detailed derivation. Next, we apply a trick from the proof of Lemma 7 in the main text: $|\mathrm{Tr}(U^\dagger V)|^2 \ge d^2\Delta^2$ is a necessary condition for $\|\mathcal{U} - \mathcal{V}\|_\diamond < 1-\Delta$. This allows us to conclude

$$\Pr\left[\|\mathcal{U} - \mathcal{V}\|_\diamond \le 1-\Delta\right] \le (1+\epsilon)\frac{k!}{d^{2k}}\frac{1}{\Delta^{2k}}. \tag{A26}$$

Qualitatively, this is of order $O(d^{-2k})$.

We now have all the ingredients in place to repeat the argument from the state case. The probability of sampling a unitary that has either low complexity *or* is close to any reference unitary V is

$$\begin{aligned}\Pr\left[\mathcal{C}_\delta(U) \le r \cup \|\mathcal{U} - \mathcal{V}\|_\diamond \le 1-\Delta\right] \\ \le 3(1+\epsilon)d^2n^{2r}|\mathsf{G}|^r\left(\frac{1024k^4}{d(1-\delta)^2}\right)^k + \frac{1+\epsilon}{\Delta^{2k}}\frac{k!}{d^{2k}},\end{aligned} \tag{A27}$$

according to a union bound. This is on the order of $O(d^{-2k}) < 1$ as long as the complexity $r \lesssim nk$. By contraposition, this ensures that there exists a design element U_1

that has both high complexity *and* is far away from V. We can use this insight to iteratively construct a set of N high-complexity design unitaries with large pairwise distances. Explicitly, to construct a set of unitaries $\{U_1,\ldots,U_N\}$, we need that

$$\Pr\left[\mathcal{C}_\delta(U_N) \le r \bigcup_{i=1}^{N-1} \|\mathcal{U}_N - \mathcal{U}_i\|_\diamond \le 1-\Delta\right] < 1. \quad \text{(A28)}$$

A union bound relates this condition to a sufficient upper bound on the set cardinality N:

$$N < \Delta^{2k}\frac{d^{2k}}{k!}\left[\frac{1}{1+\epsilon} - 3d^2n^{2r}|\mathsf{G}|^r\left(\frac{1024k^4}{d(1-\delta)^2}\right)^k\right]. \quad \text{(A29)}$$

This threshold is exponential as long as the complexity obeys $r \lesssim k$:

$$N \approx O(d^{2k}) \quad \text{for } \mathcal{C}_\delta(|\psi\rangle) \le r \approx k. \quad \text{(A30)}$$

APPENDIX B: CONCEPTUAL BACKGROUND AND CONTRIBUTIONS

1. Distinguishing states and channels

This conceptual section will review one fundamental question in probability theory, as well as two quantum generalizations. We refer to Refs. [33,68] for details. The underlying question is the following: *what is the best strategy to distinguish two (biased) coins based on a single toss?* More precisely, we consider the following game: there are two identically looking coins with different biases towards coming up heads when being tossed. These biases are known to the player. A referee then picks one of these coins uniformly at random and hands it to the player. The player is allowed to perform a single toss. Based on the result she must guess which coin she obtained and wins if this guess was correct.

a. Distinguishing classical probability distributions

Consider two (discrete) d-variate random variables. Then, we may represent the associated probability distributions by d-dimensional vectors $p,q \in \mathbb{R}^d$, which are entrywise positive ($p_i, q_i \ge 0$) and whose entries sum up to one. Likewise, a collection of events $E_1,\ldots,E_m$ can be also represented by vectors $e_1,\ldots,e_m \in \mathbb{R}^d$ that are entrywise non-negative and obey the following normalization condition: $\sum_{i=1}^m e_i = \vec{1}$. Here, $\vec{1} = (1,\ldots,1)^T \in \mathbb{R}^d$ denotes the all-ones vector. The probability of observing the event associated with index i is

$$\Pr[i] = \langle e_i, p\rangle. \quad \text{(B1)}$$

The properties of probability and event vectors then assure $\Pr[i] \ge 0$ and $\sum_{i=1}^m \Pr[i] = 1$. Let us now return to the motivating question: what is the best strategy to distinguish two random variables, characterized by known probability vectors p and q in the single-shot limit? This is a binary question and without loss of generality we can restrict our attention to binary events. Let e_1 denote the event that leads us to guess that we observe the first random variable. The complementary event $e_2 = \vec{1} - e_1$ is then fully characterized as well. Under the additional assumption that either random variable is handed to us with equal prior probability, the probability of success becomes

$$\begin{aligned} p_{\text{cl}} &= \frac{1}{2}\Pr[1|p] + \frac{1}{2}\Pr[2|q] = \frac{1}{2}\left(\langle e_1,p\rangle + \langle e_2,q\rangle\right) \\ &= \frac{1}{2}\left(\langle e_1,p-q\rangle + \langle \vec{1},q\rangle\right) = \frac{1}{2} + \frac{1}{2}\langle e_1,p-q\rangle. \end{aligned} \quad \text{(B2)}$$

This expression may now be optimized over all possible events e_1 in order to determine the optimal guessing strategy. The only constraints on e_1 are non-negativity and normalization. Together, they demand $0 \le e_1 \le \vec{1}$, where the inequality signs are to be understood componentwise. The resulting optimization problem is a *linear program* [44,69]

$$\begin{aligned} &\text{maximize} && \frac{1}{2} + \langle e_1, p-q\rangle \\ &\text{subject to} && \vec{1} \ge e_1 \ge 0, \end{aligned} \quad \text{(B3)}$$

and can be solved in a computationally tractable way. In fact, this problem is simple enough to solve analytically. The optimal e_1 is the indicator function for $p_i \ge q_i$, i.e., $e_i = \mathbb{1}\{p_i \ge q_i\}$. This is the *maximum-likelihood estimator* from statistics. Opt for the distribution that is most likely to produce the outcome that has been observed. This choice achieves an optimal success probability of

$$p_{\text{cl}}^\sharp = \frac{1}{2} + \frac{1}{4}\|p-q\|_{\ell_1}. \quad \text{(B4)}$$

Note that a success probability of $1/2$ can be trivially achieved by mere guessing. The remaining factor (multiplied by 2)

$$\beta_{\text{cl}}^\sharp = \frac{1}{2}\|p-q\|_{\ell_1} = \frac{1}{2}\sum_{i=1}^d |p_i - q_i|, \quad \text{(B5)}$$

is called the *bias* and corresponds to the *total variational distance* between p and q.

b. Distinguishing quantum states

It is useful to think of quantum states ρ as matrix generalizations of probability vectors. Similarly, positive operator-valued measurements (POVM) with m outcomes

are characterized by a collection of positive semidefinite (PSD) matrices $\{M_i\}_{i=1}^m \in \mathbb{H}_d$ that sum up to the identity matrix $\mathbb{I}$. Born's rule states that the probability of observing certain outcomes is

$$\Pr[i] = \mathrm{Tr}(M_i\rho) \quad \text{for all } 1 \leq i \leq m. \tag{B6}$$

This may be viewed as a noncommutative analog of the classical probability rule in Eq. (B1). One may also adapt the distinguishability game to the quantum setting: what is the probability of correctly distinguishing two quantum states ρ, σ by performing a single measurement? Once more, this is a binary question. We can without loss restrict attention to two-outcome measurements: M_1 and $M_2 = \mathbb{I} - M_1$. We associate the first outcome with opting for ρ while the second outcome flags σ. Similar to the classical case, the probability of success is

$$p_{\mathrm{QS}} = \frac{1}{2} + \frac{1}{2}(M_1, \rho - \sigma), \tag{B7}$$

which corresponds to a bias of $\beta_{\mathrm{QS}} = (M_1, \rho - \sigma)$. We may now optimize over all possible measurements M_1 to obtain the best bias possible:

$$\begin{aligned} \beta_{\mathrm{QS}}^{\sharp} = &\text{maximize} \quad (M_1, \rho - \sigma) \\ &\text{subject to} \quad \mathbb{I} \succeq M_1 \succeq 0. \end{aligned} \tag{B8}$$

The constraint denotes the positive semidefinite order ($A \succeq B$ if and only if $A - B$ is positive semidefinite). This is a semidefinite program [44,69] that is simple enough to solve analytically. The optimal measurement M_1 corresponds to the orthogonal projection onto the positive range of $\rho - \sigma$. The associated optimal bias is

$$\beta_{\mathrm{QS}}^{\sharp} = \frac{1}{2}\|\rho - \sigma\|_1, \tag{B9}$$

which is the *trace distance* of the density matrices ρ and σ. This result is known as the *Holevo-Helstrom theorem* [30,31].

Example 1: *Choose* $\rho = |\psi\rangle\langle\psi|$ *and* $\sigma = \rho_0 = (1/d)\mathbb{I}$. *Then, the (unique) optimal measurement is* $M_1 = |\psi\rangle\langle\psi|$ *and achieves a bias of*

$$\beta_{QS}^{\sharp} = \frac{1}{2}\||\psi\rangle\langle\psi| - \rho_0\|_1 = 1 - \frac{1}{d}. \tag{B10}$$

c. Distinguishing quantum channels

Quantum channels describe evolutions of quantum-mechanical systems. They are linear maps $\mathcal{A} : \mathbb{H}_d \to \mathbb{H}_{d'}$ that map density operators to density operators of potentially different dimension d'.

Suppose that we wish to distinguish two channels, say $\mathcal{A}$ and $\mathcal{B}$ based on a single channel use. For instance, input a concrete quantum state and perform a measurement on the outcome state. This indicates more freedom to maximize the probability of correct distinction by optimizing over potential input states and measurements of the channel output. The laws of quantum mechanics allow for further improving this strategy. It is possible to entangle the input state with a quantum memory: $\rho_{\mathrm{in}} \in \mathbb{H}_d \otimes \mathbb{H}_d$. We then apply the channel to the first quantum system, while the second one is left unchanged in the memory. A final two-outcome measurement $M_1 \in \mathbb{H}_{d'} \otimes \mathbb{H}_d$ on both output and memory state potentially reveals additional information. The outcome state depends on the channel in question. *A priori* there are two possibilities. Either $\rho_{\mathrm{out}} = \mathcal{A} \otimes \mathcal{I}(\rho_{\mathrm{in}})$, or $\rho_{\mathrm{out}} = \mathcal{B} \otimes \mathcal{I}(\rho_{\mathrm{in}})$. Here, $\mathcal{I}(X) = X$ denotes the identity channel acting trivially on the memory. The probability of correctly distinguishing these states—and thus the underlying channels—with a single measurement $M_1 \in \mathbb{H}_{d'} \otimes \mathbb{H}_d$ becomes

$$p_{\mathrm{QC}} = \frac{1}{2} + \mathrm{Tr}\{M_1[\mathcal{A} \otimes \mathcal{I}(\rho_{\mathrm{in}}) - \mathcal{B} \otimes \mathcal{I}(\rho_{\mathrm{in}})]\}. \tag{B11}$$

We may now optimize over all degrees of freedom to maximize the value of p_{QC}. Optimizing the measurement M_1 results in a bias that is proportional to the trace distance of the outcome states. Because of convexity, optimization over potential input states can without loss of generality be restricted to pure states:

$$\beta_{\mathrm{QC}}^{\sharp} = \frac{1}{2}\max_{|\psi\rangle\langle\psi|} \left\|\mathcal{A} \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \mathcal{B} \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\right\|_1. \tag{B12}$$

This optimal bias is called the *diamond distance* between channels $\mathcal{A}$ and $\mathcal{B}$ [70].

It defines a distance measure between quantum channels that is more complicated than the trace distance between quantum states and the total variational distance between classical probability distributions, respectively. It can be difficult to compute it analytically, but does admit a computationally tractable reformulation (as a semidefinite program) [71–73].

Example 2: *Consider a unitary channel* $\mathcal{U}(\rho) = U\rho U^{\dagger} \in \mathbb{H}_d$ *and the completely depolarizing channel* $\mathcal{D}(\rho) = [Tr(\rho)/d]\mathbb{I} \in \mathbb{H}_d$. *Then,*

$$\frac{1}{2}\|\mathcal{U} - \mathcal{D}\|_{\diamond} = 1 - \frac{1}{d^2}, \tag{B13}$$

and optimal strategies are based on maximally entangling the input with the memory: Let $|\Omega\rangle = (1/\sqrt{d})\sum_{i=1}^{d}|i\rangle \otimes |i\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ *be the maximally entangled (Bell) state. Set* $\rho_{in} = |\Omega\rangle\langle\Omega|$ *and measure* $M_1 = (U^{\dagger} \otimes \mathbb{I})|\Omega\rangle\langle\Omega|(U \otimes \mathbb{I})$.

It is easy to check that this strategy achieves the diamond distance in Eq. (B13). Proving optimality is less trivial. For instance, this claim follows from relating the diamond distance to another norm that is easier to compute. We refer to Ref. [74, Theorem 7] and Ref. [75] for details.

2. Conceptual contributions

a. Cornering "easy" unitary transformations

Fix $d = q^n$. The evolution of a closed, d-dimensional quantum-mechanical system is unitary: $\mathcal{U}(\rho) = U\rho U^\dagger$ with $U \in U(d)$. While evolutions may represent natural processes, they can also be engineered to perform certain tasks, such as quantum computing. Scalability of quantum computing hinges on the important observation that complicated evolutions (quantum gate architectures) can be decomposed into sequences of simple building blocks. A universal gate set $\mathsf{G} \subset U(q^2)$ acting on two (neighboring) qudits forms such a basic set of building blocks. For technical reasons, we assume that G contains the identity (doing nothing), as well as inverses: $g \in \mathsf{G}$ implies $g^\dagger \in \mathsf{G}$.

Universality then means that any unitary $U \in U(d)$ may be accurately approximated by a finite sequence of r unitaries chosen from G. We refer to Fig. 4 for an illustrative example. Such decompositions into sequences of elementary gates provide us with a notion of simplicity. Intuitively, a quantum cicuit V is simple if it may be generated by a G-local circuit of short size. In contrast to depth, size counts the total number of elementary gates in a circuit. For $r \in \mathbb{N}$ we define

$$\mathsf{G}_r := \{V \in U(d) : V \text{ is generated by a} \times \mathsf{G}\text{-local circuit of size } \leq r\}. \tag{B14}$$

We set $\mathsf{G}_0 = \{\mathbb{I}\}$ and the following inclusion relation follows from $\mathbb{I} \in \mathsf{G}$:

$$\mathsf{G}_0 \subseteq \mathsf{G}_1 \subseteq \cdots \subseteq \mathsf{G}_r. \tag{B15}$$

The cardinality of G_r may be bounded by a simple counting argument:

$$\left|\mathsf{G}_r\right| \leq \left(n|\mathsf{G}|\right)^r = \log_q(d)^r |\mathsf{G}|^r. \tag{B16}$$

The fact that G is a universal gate set ensures that G_r becomes dense in $U(d)$ provided that $r \to \infty$. *A priori* G_r depends on the particular choice of universal gate set G. However, the Solayev-Kitaev theorem also asserts that other universal gate sets can be accurately compiled at the cost of a constant overhead only [32].

b. Cornering "easy" measurements

The conceptual question underlying our definition of complexity is binary. Are we facing a pure state (unitary channel), or a maximally mixed state (depolarizing channel)? This allows us to restrict attention to two-outcome measurements, where we associate one outcome with each possibility.

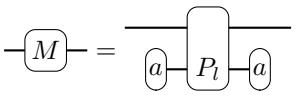
Two-outcome measurements always assume the following form: $(M, \mathbb{I} - M)$, where M obeys $\mathbb{I} \succeq M \succeq 0$. Measuring a quantum state $\rho \in \mathbb{H}_d$ results in two potential outcomes, say "yes" and "no." The probability of observing either is characterized by Born's rule (B6):

$$\begin{aligned}\Pr[\text{"yes"}] &= \operatorname{Tr}(M\rho) \quad \text{and} \quad \Pr[\text{"no"}] \\ &= \operatorname{Tr}[(\mathbb{I} - M)\rho] = 1 - \Pr[\text{"yes"}].\end{aligned} \tag{B17}$$

A *projective* two-outcome measurement is one for which M is an orthogonal projection:

$$M = VP_lV^\dagger, \quad \text{with } P_l = \sum_{i=1}^{l} |i\rangle\langle i| \text{ and } V \in U(d). \tag{B18}$$

Here $l \in [d]$ characterizes the rank of the measurement M and V is a unitary basis change to the eigenbasis of M. *Naimark's theorem*, see, e.g., Refs. [33,76], provides a powerful connection between arbitrary two-outcome measurements M and projective measurements of the form Eq. (B18). Every two-outcome measurement on $\rho \in \mathbb{H}_d$ corresponds to a projective measurement on $\rho \otimes |a\rangle\langle a| \in \mathbb{H}_d \otimes \mathbb{H}_2$, where $|a\rangle\langle a| \in \mathbb{H}_2$ is an ancilla system prepared in a pure state $|a\rangle \in \mathbb{C}^2$. Pictorially (see Appendix C 3 for an introduction of wiring diagrams),



(B19)

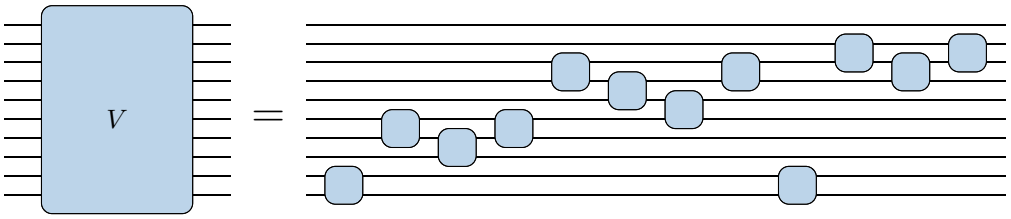


FIG. 4. *Illustration of elementary gate decompositions.* A unitary V on $n = 10$ qudits is comprised of 12 geometrically local 2-qudit gates at random positions, i.e., size$(V) = 12$.

Based on this reformulation of general two-outcome measurements, we model limited resources in the following way:

1. The ancilla state $|a\rangle \in \mathbb{C}^2$ corresponds to a (fixed) simple state, e.g., $|a\rangle = |0\rangle$.
2. The unitary $V \in U(2d)$ must be feasible to implement. More concretely we assume that it is comprised of at most r 2-qudit gates chosen from a (fixed) universal gate set $\mathsf{G} \subset U(q^2)$.
3. The projective measurement $P_l = \sum_{i=1}^{l} |i\rangle\langle i|$ is diagonal in the computational basis.

For fixed $r \in \mathbb{N}$ (circuit size for V), this framework defines the following class of measurements:

$$\mathsf{M}_r = \left\{ \mathrm{Tr}_2 \left(\mathbb{I} \otimes |a\rangle\langle a| V P_{l'} V^\dagger \right) : V \in \mathsf{G}_r, l' \in [2d] \right\} \subset \mathbb{H}_d. \tag{B20}$$

Here, $\mathrm{Tr}_2 : \mathbb{H}_d \otimes \mathbb{H}_2 \to \mathbb{H}_d$ denotes the partial trace. By construction, this set is finite and obeys

$$|\mathsf{M}_r| \leq 2d \left| \mathsf{G}_r \right| \leq 2d \left[\log_q(d) + 1 \right]^r |\mathsf{G}|^r = 2d(n+1)^r |\mathsf{G}|^r. \tag{B21}$$

The last equality is contingent on $d = q^n$ (n qudits). The set M_r captures all two-outcome measurements in Hilbert-space dimension d that can be implemented by using a single ancilla qubit, as well as circuits of size at most r.

We can readily extend this family of two-outcome measurements to quantum channel discrimination. But there we need to take into account an additional quantum memory whose dimension is also d (see, e.g., Fig. 3). So, the two-outcome measurement must act on a composite system with dimension $\dim\left(\mathbb{C}^d \otimes \mathbb{C}^d\right) = d^2$. For technical reasons, we also include a single Bell measurement $(|\Omega\rangle\langle\Omega|, \mathbb{I} - \Omega\rangle\langle\Omega|) \subset \mathbb{H}_d^{\otimes 2} \simeq \mathbb{H}_{d^2}$ with $|\Omega\rangle = (1/\sqrt{d}) \sum_{i=1}^{d} |i\rangle \otimes |i\rangle$ in the definition. This implies that the total number of elementary projective measurements is $2d^2 + 1$ and we conclude

$$\mathsf{M}_r = \left\{ \mathrm{Tr}_2 \left(\mathbb{I} \otimes |a\rangle\langle a| V P_{l'} V^\dagger \right) : V \in \mathsf{G}_r, l' \in \left[2d^2\right] \right\} \cup \left\{ V|\Omega\rangle\langle\Omega|V^\dagger : V \in \mathsf{G}_r \right\} \subset \mathbb{H}_d^{\otimes 2}. \tag{B22}$$

This modification simplifies the proof of Lemma 7 and is comparatively benign. Assuming $d = q^n$ (n qudits), a simple counting argument reveals

$$|\mathsf{M}_r| \leq (2d^2+1)|\mathsf{G}_r| \leq (2d^2+1)(2n+1)^r |\mathsf{G}|^r. \tag{B23}$$

APPENDIX C: TECHNICAL BACKGROUND AND CONTRIBUTIONS

1. Notation and basic facts from matrix analysis

Endow the vector space $\mathbb{C}^d$ with the standard inner product $\langle x|y\rangle$. A pure quantum state is a vector $\psi \in \mathbb{C}^d$ normalized to (Euclidean) unit length, i.e., $\langle \psi, \psi \rangle = 1$. We succinctly denote this by identifying normalized vectors with kets:

$$|\psi\rangle \quad \text{denotes } \psi \in \mathbb{C}^d \quad \text{with } \langle\psi|\psi\rangle = 1. \tag{C1}$$

Let $\mathbb{H}_d$ denote the space of Hermitian $d \times d$ matrices. This is a real-valued subspace of the space of all (complex-valued) $d \times d$ matrices $\mathbb{M}_d$. Fix an orthonormal basis $|1\rangle, \ldots, |d\rangle$ of $\mathbb{C}^d$. Then, the trace of a matrix X is $\mathrm{Tr}(X) = \sum_{i=1}^{d} \langle i|X|i\rangle$. The trace is cyclic, i.e., $\mathrm{Tr}(XY) = \mathrm{Tr}(YX)$ and forms the basis for defining the Schatten p-norms. In particular,

$$\begin{aligned} \|X\|_1 &= \mathrm{Tr}(|X|), |X| = \sqrt{X^2} \quad \text{(trace norm)}, \\ \|X\|_2 &= \sqrt{\mathrm{Tr}(X^2)} \quad \text{(Frobenius norm)}, \\ \|X\|_\infty &= \max_{|y\rangle} |\langle y|X|y\rangle| \quad \text{(operator norm)}. \end{aligned} \tag{C2}$$

Schatten-norms obey the following order relations:

$$\begin{aligned} \|X\|_\infty \leq \|X\|_2 \leq \|X\|_1 \quad \text{and} \quad \|X\|_1 \leq \sqrt{d}\|X\|_2 \\ \leq d\|X\|_\infty \quad \text{for all } X \in \mathbb{H}_d. \end{aligned} \tag{C3}$$

A variant of Hölder's inequality applies to traces of inner products, see, e.g., Ref. [77, Ex. IV.2.12]:

$$|\mathrm{Tr}(XY)| \leq \|X\|_1 \|Y\|_\infty \quad \text{for all } X, Y \in \mathbb{H}_d. \tag{C4}$$

The trace corresponds to a full index contraction. Partial contractions are possible for tensor products and *partial traces* are concrete examples. For $X, Y \in \mathbb{H}_d$ define

$$\mathrm{Tr}_1 (X \otimes Y) = \mathrm{Tr}(X) Y \quad \text{and} \quad \mathrm{Tr}_2 (X \otimes Y) = \mathrm{Tr}(Y) X, \tag{C5}$$

and extend this definition linearly to the tensor product $\mathbb{H}_d^{\otimes 2} \simeq \mathbb{H}_{d^2}$. This definition naturally extends to tensor products of higher order. The following tight bound connects partial traces and operator norms:

$$\begin{aligned} \max\left\{ \|\mathrm{Tr}_1(X)\|_\infty, \|\mathrm{Tr}_2(X)\|_\infty \right\} \\ \leq d\|X\|_\infty \quad \text{for all } X \in \mathbb{H}_d^{\otimes 2}. \end{aligned} \tag{C6}$$

A matrix $X \in \mathbb{H}_d$ is PSD if $\langle y|X|y\rangle \geq 0$ for all $y \in \mathbb{C}^d$. We denote this feature by $X \succeq 0$. Positive semidefiniteness is preserved under partial traces:

$$X \in \mathbb{H}_d^{\otimes 2}, X \succeq 0 \quad \text{implies } \mathrm{Tr}_1(X) \succeq 0, \mathrm{Tr}_2(X) \succeq 0. \tag{C7}$$

The trace norm of PSD matrices is particularly simple: $\|X\|_1 = \mathrm{Tr}(X)$ whenever $X \succeq 0$.

2. Convex geometry and optimization

The main technical contributions of this paper are based on bounds that follow from a fundamental argument in convex optimization. Comprehensive references for convex geometry and optimization include Refs. [43,44]. A function $f : \mathbb{H}_d \to \mathbb{R}$ is *convex* if

$$f\left[\tau X + (1-\tau)Y\right] \leq \tau f(X) + (1-\tau) f(Y) \quad \text{for all } X, Y \in \mathbb{H}_d, \tau \in [0,1]. \tag{C8}$$

Linear transformations in the argument preserve this feature. Similarly, a set $\mathcal{K} \subseteq \mathbb{H}_d$ is *convex* if

$$X, Y \in \mathcal{K} \quad \text{imply } \tau X + (1-\tau)Y \in \mathcal{K} \quad \text{for all } \tau \in [0,1]. \tag{C9}$$

Let $\mathcal{K} \subseteq \mathbb{H}_d$ be a convex set. A point $X \in \mathcal{K}$ is an *extreme point* if $Y, Z \in \mathcal{K}$ and $X = \tau Y + (1-\tau)Z$ for some $\tau \in (0,1)$ necessarily imply $Y = Z = X$. Extreme points form the boundary of a convex set.

Example 3: *The set of all quantum states in $\mathbb{H}_d$ is the convex hull (i.e., the set of all convex combinations) of pure states:*

$$\{\rho \in \mathbb{H}_d : Tr(\rho) = 1, \rho \succeq 0\} = conv\left\{|\psi\rangle\langle\psi| : |\psi\rangle \in \mathbb{C}^d\right\}. \tag{C10}$$

All extreme points are pure states.

Fact 18 (Convex functions achieve their maximum at an extreme point): *Let $\mathcal{K} \subseteq \mathbb{H}_d$ be a convex set and let $f : \mathcal{K} \to \mathbb{R}$ be a convex function. Then, there exists an extreme point $X_\sharp$ of $\mathcal{K}$ such that*

$$\max_{X \in \mathcal{K}} f(X) \leq f(X_\sharp). \tag{C11}$$

This result justifies the presentation of the diamond distance in Eq. (B12). The function $X \mapsto \|\mathcal{A} \otimes \mathcal{I}(X) - \mathcal{B} \otimes \mathcal{I}(X)\|_1$ is convex (norms are convex and the channel acts like a linear transformation of the argument) and pure states are the extreme points of the set of all quantum states. Hence,

$$\max_{\rho} \|\mathcal{A} \otimes \mathcal{I}(\rho) - \mathcal{B} \otimes \mathcal{I}(\rho)\|_1 = \max_{|\psi\rangle\langle\psi|} \|\mathcal{A} \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \mathcal{B} \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1. \tag{C12}$$

The following technical result will prove highly valuable for establishing bounds on very general Haar moments.

Lemma 19: *Fix $A \in \mathbb{H}_d$ PSD ($A \succeq 0$). Then, the function $h(X) = Tr(XAXA)$ is non-negative and convex for all $X \in \mathbb{H}_d$.*

Proof. Apply an eigenvalue decomposition: $A = U(\sum_{i=1}^d \alpha_i |i\rangle\langle i|)U^\dagger$. The assumption that A is PSD ensures $\alpha_1, \ldots, \alpha_d \geq 0$. Next, fix $X \in \mathbb{H}_d$ arbitrary, set $\tilde{X} = U^\dagger X U$ and compute

$$\mathrm{Tr}(XAXA) = \sum_{i,j=1}^d \alpha_i \alpha_j \left|\langle i|\tilde{X}|j\rangle\right|^2 \geq 0. \tag{C13}$$

This establishes non-negativity of $h(X)$. For convexity, fix $X, Y \in \mathbb{H}_d$ and $\tau \in [0,1]$. Set $\bar{\tau} = 1 - \tau$ and note that $\tau\bar{\tau} = \tau - \tau^2 = \bar{\tau} - \bar{\tau}^2 \geq 0$. Non-negativity moreover implies $h(X - Y) \geq 0$ and we can readily deduce convexity:

$$\begin{aligned} h(\tau X + \bar{\tau} Y) &= \tau^2 \mathrm{Tr}(XAXA) + 2\tau\bar{\tau}\mathrm{Tr}(XAYA) + \bar{\tau}^2 \mathrm{Tr}(YAYA) \\ &= \tau h(X) - \tau\bar{\tau}\left[\mathrm{Tr}(XAXA) - 2\mathrm{Tr}(XAYA)\right. \\ &\quad \left. + \mathrm{Tr}(YAYA)\right] + \bar{\tau} h(Y) \\ &= \tau h(X) - \tau\bar{\tau} h(X - Y) + \bar{\tau} h(Y) \leq \tau h(X) + \bar{\tau} h(Y). \end{aligned} \tag{C14}$$

■

3. Wiring calculus

Wiring diagrams, sometimes also known as tensor network diagrams, provide a graphical way for computing contractions between tensors. Here we provide only a brief overview and refer to the recent survey [41] and lecture notes [68] for a detailed introduction. The wiring formalism associates a box with every tensor and a line emanating from the box with every index. Connected lines represent contracted indices. More precisely, we place contravariant indices of a tensor on the left of the box and covariant ones on the right. Table I contains all the essential rules necessary for the scope of this work.

Importantly lines can be bent at will without changing the value of an equation [78]. For instance, let $\rho = |\psi\rangle\langle\psi| \in \mathbb{H}_d$ be a pure quantum state and suppose that $M \in \mathbb{H}_d$ is measurement. We can then represent Born's rule pictographically as

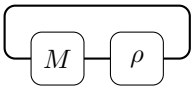
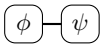
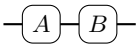
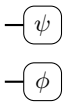
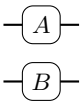
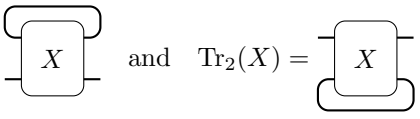
$\mathrm{Tr}(M\rho) =$  $= \langle\psi|M|\psi\rangle$. (C15)

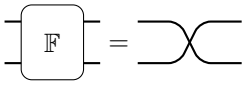
TABLE I. Basic building blocks of wiring calculus.

ket vector	$\|\psi\rangle \in \mathbb{C}^d$	
bra vector	$\langle\phi\| \in \left(\mathbb{C}^d\right)^* \simeq \mathbb{C}^d$	
inner product (contraction)	$\langle\phi\|\psi\rangle$	
matrix	$A \in \mathbb{M}_d$	
matrix product of $A, B \in \mathbb{M}_d$	$AB \in \mathbb{M}_d$	
matrix trace (contraction)	$\mathrm{Tr}(A) \in \mathbb{C}$	
tensor product (vectors)	$\|\psi\rangle \otimes \|\phi\rangle \in (\mathbb{C}^d)^{\otimes 2}$	
tensor product (matrices)	$A \otimes B \in \mathbb{H}_d^{\otimes 2}$	

Partial traces also assume a simple form. For $X \in \mathbb{H}_d \otimes \mathbb{H}_d$

$$\mathrm{Tr}_1(X) = \quad \text{and} \quad \mathrm{Tr}_2(X) = \tag{C16}$$


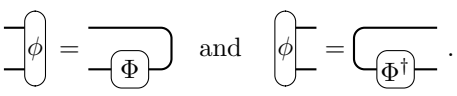
Wiring calculus is exceptionally well suited to keep track of *flip operators*. Define $\mathbb{F}|i\rangle| \otimes |j\rangle = |j\rangle \otimes |i\rangle$ via its action on computational basis elements and extend this definition linearly to $\mathbb{C}^d \otimes \mathbb{C}^d$. Then,

$$= \tag{C17}$$


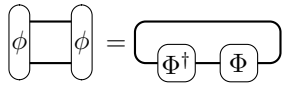
Vectorization is a linear map vec: $\mathbb{M}_d \to \mathbb{C}^d \otimes \mathbb{C}^d$ defined by its action on computational basis elements

$$|\mathrm{vec}\left(|i\rangle\langle j|\right)\rangle := |i\rangle \otimes |j\rangle, \tag{C18}$$

and linearly extended to all of $\mathbb{M}_d$. In wiring calculus, $|\phi\rangle = |\mathrm{vec}(\Phi)\rangle$ corresponds to bending the right (covariant) index of a matrix A to the left (into a contravariant one):

$$= \quad \text{and} \quad = \tag{C19}$$


It is easy to see that vectorization is an isometry:

$$\langle\phi|\phi\rangle = = = \mathrm{Tr}\left(\Phi^\dagger\Phi\right) = \|\Phi\|_2^2 . \tag{C20}$$


4. Random unitaries and *k*-designs

Here we introduce a few essential concepts from quantum information theory, including a discussion of random unitaries and the notion of a design. First, recall that the Haar measure is the unique left and right invariant measure on the unitary group $U(d)$. We are often interested in moments of the Haar ensemble. Consider an operator X acting on the k-fold Hilbert space $(\mathbb{C}^d)^{\otimes k}$, the k-fold channel, or k-fold twirl, of the operator with respect to the Haar measure on the unitary group is

$$\mathcal{T}_U^{(k)}(X) = \int dU\, U^{\otimes k}(X)U^{\dagger\otimes k}. \tag{C21}$$

Similarly, we can average an operator over an ensemble of unitaries $\mathcal{E} = \{p_i, U_i\}$, a weighted subset of the full unitary group. The k-fold channel with respect to $\mathcal{E}$ is

$$\mathcal{T}_\mathcal{E}^{(k)}(X) = \sum_i p_i U_i^{\otimes k}(X)U_i^{\dagger\otimes k}, \tag{C22}$$

here written for a discrete ensemble, but such an ensemble might be discrete or continuous.

Unitary k-designs. We are often interested in how well an average over an ensemble captures an average over the full unitary group, i.e., how random the ensemble is with respect to the Haar measure on $U(d)$. A *unitary k-design* is an ensemble of unitaries $\mathcal{E} = \{p_i, U_i\}$, for which the k-fold twirl equals its Haar-random counterpart:

$$\mathcal{T}_\mathcal{E}^{(k)}(X) = \mathcal{T}_U^{(k)}(X) \quad \text{for all } X \in \mathbb{H}_d^{\otimes k}. \tag{C23}$$

This means that the ensemble $\mathcal{E}$ exactly captures the first k moments of the Haar ensemble. Unitary operator bases, such as the n-qubit Pauli group, form an exact 1-design. But very little is known about the construction of exact designs for higher k, with the notable exception of $k = 3$ and the n-qubit Clifford group [15–17]. We return to this point when discussing approximate designs.

Schur-Weyl duality. Many of the important analytic expressions for Haar averages rely on *Schur-Weyl duality* [37,38], a deep connection between irreducible representations (irreps) of the unitary group $U(d)$ and the symmetric group S_k. First, when thinking about k-fold Hilbert spaces, there is a useful set of operators that acts on this space, namely permutations of the k copies. A permutation

operator P_σ acts on the computational basis of $(\mathbb{C}^d)^{\otimes k}$ as

$$P_\sigma |i_1, \ldots, i_k\rangle = |i_{\sigma^{-1}(1)}, \ldots, i_{\sigma^{-1}(k)}\rangle. \tag{C24}$$

This action can be extended linearly to all of $(\mathbb{C}^d)^{\otimes k}$. Schur-Weyl duality is the statement that an operator acting on $(\mathbb{C}^d)^{\otimes k}$ commutes with all k-fold unitaries $U^{\otimes k}$ if and only if it is a linear combination of permutation operators

$$U^{\otimes k} X U^{\dagger \otimes k} = X \quad \longleftrightarrow \quad X = \sum_{\sigma \in S_k} c_\sigma P_\sigma. \tag{C25}$$

Many of the exact expressions for Haar moments and random unitary averages in the following subsection follow directly from this powerful result.

5. Haar integration over the unitary group

We now introduce the general formalism for integrating arbitrary moments of random unitaries over the full unitary group with respect to the Haar measure, often referred to as Weingarten calculus. Note that the k-fold twirl in Eq. (C21) describes a linear operator on the tensor product space $\mathbb{H}_d^{\otimes k}$. The associated matrix representation is called the kth moment operator, written as $O_U^{(k)} = \int dU\, U^{\otimes k} \otimes \bar{U}^{\otimes k}$, where $\bar{U}$ denotes the complex conjugate. Weingarten calculus [40,79] provides exact expressions for individual matrix elements of the moment operator:

$$\int dU\, U_{i_1 j_1} \ldots U_{i_k j_k} \bar{U}_{\ell_1 m_1} \ldots \bar{U}_{\ell_k m_k} = \sum_{\sigma,\tau \in S_k} \delta_\sigma(\vec{\imath}|\vec{\ell})\delta_\tau(\vec{\jmath}|\vec{m}) \mathcal{W}g(\sigma^{-1}\tau, d), \tag{C26}$$

where we sum over elements of the permutation group S_k and define a contraction of indices with respect to a permutation $\sigma \in S_k$ as

$$\delta_\sigma(\vec{\imath}|\vec{\jmath}) := \prod_{s=1}^{k} \delta_{i_s j_{\sigma(s)}} = \delta_{i_1 j_{\sigma(1)}} \ldots \delta_{i_k j_{\sigma(k)}}. \tag{C27}$$

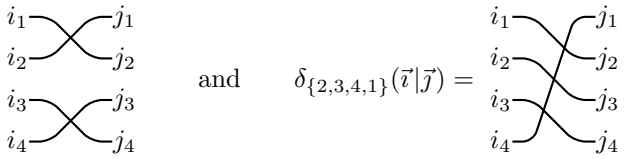
Mixed moments of U and $\bar{U}$, i.e., averages of $U^{\otimes k} \otimes \bar{U}^{\otimes k'}$ with $k \neq k'$, vanish identically.

It is often convenient to interpret the index contraction $\delta_\sigma(\vec{\imath}|\vec{\jmath})$ as a permutation operator acting on the computational basis of the k-fold space,

$$\delta_\sigma(\vec{\imath}|\vec{\jmath}) = P_\sigma. \tag{C28}$$

For instance, two examples of contractions for $k = 4$ are

$$\delta_{\{2,1,4,3\}}(\vec{\imath}|\vec{\jmath}) = \quad \text{and} \quad \delta_{\{2,3,4,1\}}(\vec{\imath}|\vec{\jmath}) = \tag{C29}$$



The weight associated to a given contraction is called the Weingarten function, $\mathcal{W}g(\sigma, d)$. It is a function on elements of S_k and admits an expansion in terms of characters of the symmetric group

$$\mathcal{W}g(\sigma, d) = \frac{1}{k!} \sum_{\lambda \vdash k} \frac{f_\lambda \chi_\lambda(\sigma)}{c_\lambda(d)}, \tag{C30}$$

where we sum over the integer partitions of k that label the irreps of S_k; $\chi_\lambda(\sigma)$ is an irreducible character of λ, and f_λ is the dimension of the irrep λ. The polynomial in the denominator is defined as

$$c_\lambda(d) = \prod_{(i,j)\in\lambda} (d + j - 1), \tag{C31}$$

where we take a product over the coordinates (i, j) of the Young diagram of λ. Writing λ as an integer partition of k, with elements λ_i, the product is taken over i from 1 to $\ell(\lambda)$, the length of the partition, and j from 1 to λ_i. The expression for the Weingarten function in Eq. (C30), is valid for $k \geq d$ by restricting the sum over partitions of length $\ell(\lambda) \leq d$ [such that the polynomial $c_\lambda(d)$ in the denominator is free of zeroes].

The Weingarten functions depend only on the cycle type of the permutation, where the cycle type of $\sigma \in S_k$ is an integer partition of k. We end this brief exposition by listing the first few unitary Weingarten functions, labeled by cycle type. For $k = 1$, $\mathcal{W}g[(1), d] = (1/d)$, and for $k = 2$, we have

$$\mathcal{W}g[(1,1), d] = \frac{1}{d^2 - 1}, \quad \text{and}$$
$$\mathcal{W}g[(2), d] = -\frac{1}{d(d^2 - 1)}. \tag{C32}$$

k-fold twirl over $U(d)$. The k-fold twirl, Eq. (C21), of an operator over the unitary group can be written using Eq. (C26) as

$$\mathcal{T}_U^{(k)}(X) = \mathbb{E}_U\left[U^{\otimes k}(X)U^{\dagger\otimes k}\right] = \sum_{\sigma,\tau\in S_k} \mathcal{W}g(\sigma^{-1}\tau, d)P_\sigma \mathrm{Tr}(XP_\tau). \tag{C33}$$

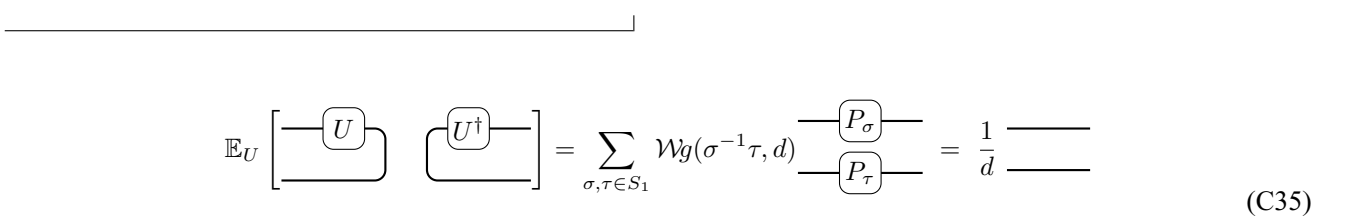
This expression equivalently follows from noting that, by the invariance of the Haar measure, the k-fold twirl $\mathcal{T}_U^{(k)}$ is invariant both under k-fold unitary conjugation and under k-fold conjugation of X.

We also note that the k-fold twirl of a permutation operator is $\mathcal{T}_U^{(k)}(P_\rho) = P_\rho$. Equation (C33), then gives that $\mathcal{W}g(\sigma^{-1}\tau, d)\mathrm{Tr}(P_\tau P_\rho) = \delta_{\sigma,\rho}$. Viewed as a matrix equation, the matrix of Weingarten functions $\mathcal{W}g_{(k)}$ is the pseudoinverse of the $k! \times k!$ matrix $G_{(k)}$ of inner products of permutation operators P_σ (the Gram matrix of P_σ's). The elements of $G_{(k)}$ are the inner products between permutation operators, $\mathrm{Tr}(P_\sigma P_\tau) = d^{\ell(\sigma^{-1}\tau)}$, where $\ell(\sigma^{-1}\tau)$ simply counts the number of closed cycles in the permutation product (equivalently, the length of the cycle type of the product):

$$\mathcal{W}g_{(k)} = G_{(k)}^{-1} \quad \text{with } \mathcal{W}g_{(k)} = \left[\mathcal{W}g(\sigma^{-1}\tau, d)\right]_{\sigma,\tau\in S_k} \quad \text{and} \quad G_{(k)} = \left[\mathrm{Tr}(P_\sigma P_\tau)\right]_{\sigma,\tau\in S_k}. \tag{C34}$$

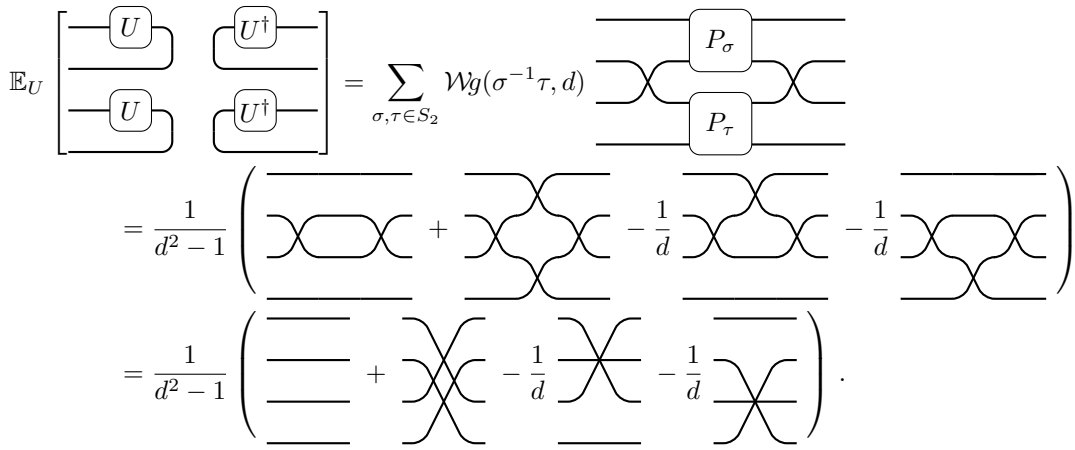
For more discussion on this, see Refs. [11,79]. The matrix inverse exists for $k \leq d$. Although elegant, this derivation of the Weingarten functions quickly becomes intractable as we need to invert a $k! \times k!$ matrix. The representation theoretic definition in Eq. (C30) is straightforward to use in computing high moments.

Wiring diagrams for the first few Haar moments. To set up the calculations that will follow in the next section, we explicitly write out the wiring diagrams in the first two moments, detailing the index contractions one must take. For $k = 1$, we simply have



(C35)

For $k = 2$, we sum over elements of S_2, separately permuting the internal and external indices as



(C36)

Moments of traces. We can use the formalism introduced above to compute a few simple expressions averaged over the unitary group, which will be of use in later sections. Consider the $2k$th moment of the trace of a random unitary, $|\mathrm{Tr}(U)|^{2k}$, which we integrate over the unitary group as

$$\mathbb{E}_U\left[|\mathrm{Tr}(U)|^{2k}\right] = \sum_{\sigma,\tau\in S_k} \mathcal{W}g(\sigma^{-1}\tau, d)\mathrm{Tr}(P_\sigma P_\tau), \tag{C37}$$

with $\mathrm{Tr}(P_\sigma P_\tau) = d^{\ell(\sigma\tau)}$. View this as a matrix equation, and recall that for $k \leq d$ the Weingarten functions are the inverse of the inner products Eq. (C34). Then, we simply have the trace of the identity matrix, a sum over S_k:

$$\mathbb{E}_U\left[|\mathrm{Tr}(U)|^{2k}\right] = k!. \tag{C38}$$

This quantity is essentially the same as the frame potential [14], a quantity that quantifies the 2-norm distance between an ensemble of unitaries $\mathcal{E}$ and the Haar ensemble. The frame potential for any ensemble is lower bounded by this Haar value.

Averages of pure states. Consider a Haar random state $|\psi\rangle = U|0\rangle$, with $|0\rangle \in \mathbb{C}^d$ and $U \in U(d)$, and take the k-fold average with respect to the unitary group. Then,

$$\begin{aligned}\mathcal{T}_U^{(k)}(|\psi\rangle\langle\psi|^{\otimes k}) &= \sum_{\sigma,\tau\in S_k} \mathcal{W}g(\sigma^{-1}\tau, d) P_\sigma \mathrm{Tr}(P_\tau |\psi\rangle\langle\psi|^{\otimes k}) \\ &= \sum_{\sigma,\tau\in S_k} \mathcal{W}g(\sigma^{-1}\tau, d) P_\sigma,\end{aligned} \tag{C39}$$

as permuting and contracting the pure state moments is the same for any permutation. This also follows from Schur-Weyl duality by noting that the k-fold average is invariant under k-fold unitary conjugation and may thus be expressed as a sum of permutations. Fixing σ above, the sum over τ just gives the sum over Weingarten functions, which is

$$\sum_{\tau\in S_k} \mathcal{W}g(\tau, d) = \frac{1}{k!}\binom{k+d-1}{k}^{-1}. \tag{C40}$$

Equivalently, we can fix this coefficient by taking the trace of Eq. (C39). Thus we find that the k-fold average of a pure state is

$$\mathcal{T}_U^{(k)}(|\psi\rangle\langle\psi|^{\otimes k}) = \binom{k+d-1}{k}^{-1} \Pi_{\mathrm{sym}}, \tag{C41}$$

where $\Pi_{\mathrm{sym}} = (1/k!)\sum_{\sigma\in S_k} P_\sigma$ is the projector onto the symmetric subspace and $\binom{k+d-1}{k}$ is the corresponding dimension.

A similar calculation is to consider the moments of the expectation value of a conjugated operator $\langle\psi|U^\dagger M U|\psi\rangle$, where $|\psi\rangle \in \mathbb{C}^d$ and a Hermitian operator $M \in \mathbb{H}_d$. We find

$$\begin{aligned}&\mathbb{E}_U\left[|\langle\psi|U^\dagger M U|\psi\rangle|^k\right] \\ &\quad= \sum_{\sigma,\tau\in S_k} \mathcal{W}g(\sigma^{-1}\tau, d)\mathrm{Tr}(P_\sigma|\psi\rangle\langle\psi|)\mathrm{Tr}(P_\tau M^{\otimes k}).\end{aligned} \tag{C42}$$

Again, as permuting and contracting tensor products of a pure state just gives one, for any τ the σ sum is just a sum over Weingarten functions. Using Eq. (C40) and recalling the definition of the projector onto the symmetric subspace, we conclude

$$\mathbb{E}_U\left[\langle\psi|U^\dagger M U|\psi\rangle^k\right] = \binom{d+k-1}{k}^{-1} \mathrm{Tr}\left(\Pi_{\mathrm{sym}} M^{\otimes k}\right). \tag{C43}$$

6. Approximate k-designs and bounds on weight distributions

Weingarten calculus is a powerful tool. It characterizes twirls over the diagonal representation of the unitary group for arbitrary tensor powers $k \in \mathbb{N}$. In turn, this formula allows for computing moments of random variables that involve Haar random unitaries. These then can be used to establish *generic* features, such as concentration of measure. However, full control of *all* moments comes at a price. It is excessively difficult to sample unitaries directly from the Haar measure. Simple dimension counting highlights that circuits of exponential size are required to implement a Haar-random unitary circuit on n qudits.

The notion of k-designs introduced in Appendix C 4 addresses this issue by allowing one to interpolate between Haar-random ($k = \infty$) and highly structured ($k = 1$) ensembles. Unfortunately, very few explicit constructions of k-designs are known. This lack of efficient constructions can be overcome by relaxing the defining property of a k-design.

Definition 4 (Approximate k-design): *Fix $k \in \mathbb{N}$ and $\epsilon > 0$. A unitary ensemble $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$ is an ϵ-approximate (unitary) k-design if the associated twirling channel $\mathcal{T}_\mathcal{E}^{(k)}(X) = \sum_{i=1}^n p_i U_i^{\otimes k} X (U_i^\dagger)^{\otimes k}$ obeys*

$$\left\|\mathcal{T}_\mathcal{E}^{(k)} - \mathcal{T}_U^{(k)}\right\|_\diamond \leq \frac{k!}{d^{2k}}\epsilon. \tag{C44}$$

Here, $\mathcal{T}_U^{(k)}$ denotes the twirl over the full unitary group (C33) (with respect to the Haar measure).

This definition readily extends to ensembles of infinite cardinality. Several different definitions of approximate k-designs can be found in the literature. By and large these differ in terms of the metric that is used to quantify closeness. We define an approximate design up to additive error, but choose ϵ to scale with d in a manner that mimics relative error, similar to the strong definition of a design used in Ref. [12]. This will also simplify exposition considerably.

The approximate k-design property imposes severe restrictions on associated distribution of weights and the ensemble size.

Lemma 20 (Restatement of Lemma 3): *Let* $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$ *be an* ϵ*-approximate* k*-design for* $U(d)$*. Then,*

$$\max_{1\le j\le N} p_j \le (1+\epsilon)\frac{k!}{d^{2k}} \quad \textit{and} \quad N \ge \frac{d^{2k}}{(1+\epsilon)k!}. \tag{C45}$$

Lower bounds on approximate k-design cardinality are known, see, e.g., Ref. [12, Lemma 26] for a similar result. We are not aware of any weight bounds in the literature.

We also consider orbits of approximate k-designs $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$. Fix $|x\rangle \in \mathbb{C}^d$ arbitrary and define $|y_i\rangle = U_i|x\rangle$ for $i \in [N]$. Doing so results in a weighted set of unit vectors. These sets are called approximate complex-projective k-designs [18,80]. They approximately reproduce the first k moments of the uniform distribution on the complex unit sphere. Lower bounds on the cardinality of exact spherical k-designs are known, see, e.g., Ref. [20], but we are not aware of any statement that bounds the associated weights.

Lemma 21: *Let* $\{q_i, |y_i\rangle\}_{i=1}^{N'} \subset \mathbb{C}^d$ *be the weighted set of distinct states contained in an orbit of an* ϵ*-approximate* k*-design. Then,*

$$\max_{j\in[N']} q_j \le (1+\epsilon)\binom{d+k-1}{k}^{-1} \quad \textit{and} \quad N' \ge \frac{1}{1+\epsilon}\binom{d+k-1}{k}. \tag{C46}$$

The emphasis on distinct states is justified. Two or more distinct unitaries can give rise to the same state.

Proof of Lemma 20. Fix $j \in [N] = \{1,\dots,N\}$ and use Eq. (C38) to conclude

$$\begin{aligned}\sum_{i=1}^N p_i \left|\mathrm{Tr}(U_j^\dagger U_i)\right|^{2k} &= \mathbb{E}_{\mathcal{E}}\left[|\mathrm{Tr}(U_j^\dagger U)|^{2k}\right] \le k! \\ &+ \underbrace{\mathbb{E}_{\mathcal{E}}\left[|\mathrm{Tr}(U_j^\dagger U)|^{2k}\right] - \mathbb{E}_U\left[|\mathrm{Tr}(U_j^\dagger U)|^{2k}\right]}_{\Delta}.\end{aligned} \tag{C47}$$

The approximate k-design property implies that the mismatch on the rhs remains small. Let $|\Omega\rangle = (1/\sqrt{d})\sum_{i=1}^d |i\rangle \otimes |i\rangle$ denote the maximally entangled state. Then, $\mathrm{Tr}(U) = d\langle\Omega|U\otimes\mathbb{I}|\Omega\rangle$ and we apply Definition 4 to bound

$$\begin{aligned}\Delta &= \mathbb{E}_{\mathcal{E}}\left[|\mathrm{Tr}(U_j^\dagger U)|^{2k}\right] - \mathbb{E}_U\left[|\mathrm{Tr}(U_j^\dagger U)|^{2k}\right] \\ &= d^{2k}\langle\Omega|^{\otimes k}\left(\mathbb{E}_{\mathcal{E}}\left\{\left[(U\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(U\otimes\mathbb{I})^\dagger\right]^{\otimes k}\right\}\right. \\ &\quad \left. - \mathbb{E}_U\left\{\left[(U\otimes\mathbb{I})|\Omega\rangle\langle\Omega|(U\otimes\mathbb{I})^\dagger\right]^{\otimes k}\right\}\right)|\Omega\rangle^{\otimes k} \\ &\le d^{2k}\left\|\mathbb{E}_{\mathcal{E}}\left\{[\mathcal{U}\otimes\mathcal{I}(|\Omega\rangle\langle\Omega|)]^{\otimes k}\right\}\right. \\ &\quad \left. - \mathbb{E}_U\left\{[\mathcal{U}\otimes\mathcal{I}(|\Omega\rangle\langle\Omega|)]^{\otimes k}\right\}\right\|_\infty \\ &\le d^{2k}\left\|\mathcal{T}_{\mathcal{E}}^{(k)} - \mathcal{T}_U^{(k)}\right\|_\diamond \le \epsilon k!.\end{aligned} \tag{C48}$$

Combining both arguments implies $\sum_{i=1}^N p_i \left|\mathrm{Tr}(U_j^\dagger U_i)\right|^{2k} \le (1+\epsilon)k!$. This allows us to conclude

$$\begin{aligned}(1+\epsilon)k! &\ge \sum_{i=1}^N p_i \left|\mathrm{Tr}\left(U_j^\dagger U_i\right)\right|^{2k} \\ &= \sum_{i\ne j} p_i \left|\mathrm{Tr}\left(U_j^\dagger U_i\right)\right|^{2k} + p_j \left|\mathrm{Tr}\left(U_j^\dagger U_j\right)\right|^{2k} \ge p_j d^{2k},\end{aligned} \tag{C49}$$

for $j \in [N]$ arbitrary. The lower bound on the cardinality N is an immediate consequence of this weight restriction:

$$1 = \sum_{i=1}^N p_i \le \sum_{i=1}^N (1+\epsilon)\frac{k!}{d^{2k}} = N(1+\epsilon)\frac{k!}{d^{2k}}. \tag{C50}$$

■

Proof of Lemma 21. The argument is very similar to the proof of Lemma 20. Fix $j \in [N']$, set $M = |y_j\rangle\langle y_j|$ and use Eq. (C43) to conclude

$$\begin{aligned}\sum_{i=1}^{N'} q_i \left|\langle y_j, y_i\rangle\right|^{2k} &= \sum_{i=1}^N p_i \left|\langle y_j|U_i|x\rangle\right|^2 = \mathbb{E}_{\mathcal{E}}\left[\langle x|UMU^\dagger|x\rangle\right] \\ &= \binom{d+k-1}{k}^{-1}\mathrm{Tr}\left(\Pi_{\mathrm{sym}}M^{\otimes k}\right) \\ &\quad + \underbrace{\mathrm{Tr}\left(M^{\otimes k}\{\mathbb{E}_{\mathcal{E}}\left[(U|x\rangle\langle x|U^\dagger)^{\otimes k}\right] - \mathbb{E}_U\left[(U|x\rangle\langle x|U^\dagger)^{\otimes k}\right]\}\right)}_{\Delta}.\end{aligned} \tag{C51}$$

Next, observe that the Haar average obeys $\mathrm{Tr}\left(\Pi_{\mathrm{sym}}M^{\otimes k}\right) = \mathrm{Tr}\left(\Pi_{\mathrm{sym}}|y_j\rangle^{\otimes k}\right) = 1$. The approximate k-design property in addition implies that the deviation from this ideal value remains small. The matrix Hoelder inequality asserts

$$\begin{aligned}\Delta &= \mathrm{Tr}\left(M^{\otimes k}\{\mathbb{E}_{\mathcal{E}}\left[(U|x\rangle\langle x|U^{\dagger})^{\otimes k}\right]-\mathbb{E}_{U}\left[(U|x\rangle\langle U^{\dagger})^{\otimes k}\right]\}\right)\\ &\leq \|M^{\otimes k}\|_{\infty}\left\|\mathcal{T}_{\mathcal{E}}^{(k)}\left[(|x\rangle\langle x|)^{\otimes k}\right]-\mathcal{T}_{U}^{(k)}\left[(|x\rangle\langle x|)^{\otimes k}\right]\right\|_{1}\\ &\leq \|M\|_{\infty}^{k}\left\|\mathcal{T}_{\mathcal{E}}^{(k)}-\mathcal{T}_{U}^{(k)}\right\|_{\diamond}\leq\epsilon\frac{k!}{d^{2k}}\leq\binom{d+k-1}{k}^{-1}\epsilon,\end{aligned} \tag{C52}$$

because $\|M\|_{\infty}=\||y_j\rangle\langle\|_{\infty}=1$. This allows us to conclude

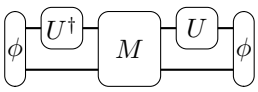
$$(1+\epsilon)\binom{d+k-1}{k}^{-1}\geq\sum_{i=1}^{N'}q_i\left|\langle y_j,y_i\rangle\right|^{2k}=q_j\left|\langle y_j,y_j r\rangle\right|^{2k}+\sum_{i\neq j}q_j\left|\langle y_j,y_i\rangle\right|^{2k}\geq q_j, \tag{C53}$$

for any $j\in\left[N'\right]$. Both weight and cardinality bound readily follow from this assertion. ■

7. A general moment bound for Haar-random unitaries

Theorem 22 (Detailed restatement of Theorem 10): *Fix* $|\phi\rangle\in(\mathbb{C}^d)^{\otimes 2}$ *and* $M\in\mathbb{H}_d^{\otimes 2}$ *such that* $\mathbb{I}\succeq M\succeq 0$. *Set*

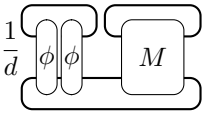
$$S_U(M,\phi):=\mathrm{Tr}\big(M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\big)=$$



, (C54)

where $U\in U(d)$ *is chosen uniformly from the Haar measure. Then,*

$$\mu(M,\phi):=\mathbb{E}_U\left[S_U(M,\phi)\right]=\frac{1}{d}$$



$$=\mathrm{Tr}\left(M\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right), \tag{C55}$$

where $\mathcal{D}(X)=[Tr(X)/d]\mathbb{I}$ *is the depolarizing channel. Moreover, the following bounds apply to all centered moments of order* $k=1,\ldots,d^{2/3}$*:*

$$\mathbb{E}_U\Big\{\left[S_U(M,\phi)-\mu(M,\phi)\right]^k\Big\}\leq\frac{C_k(k!)^2}{d^{k/2}}. \tag{C56}$$

Here, $C_k=[1/(k+1)]\binom{2k}{k}$ *is the* kth *Catalan number.*

8. Moment bounds for approximate designs

Corollary 23: *With the same assumptions in Theorem 22, but suppose that* $U\in U(d)$ *is chosen from an* ϵ*-approximate unitary* k*-design* $\mathcal{E}$*. Then,*

$$\mathbb{E}_{\mathcal{E}}\left[\left(\underbrace{\phi\ U^{\dagger}\ M\ U\ \phi}_{S_U(M,\phi)}-\underbrace{\frac{1}{d}\ \phi\ \phi\ M}_{\mu(M,\phi)}\right)^k\right]\leq\frac{(k!)^2}{d^{k/2}}\left(C_k+\frac{\epsilon}{k!d^{3k/2}}\right). \tag{C57}$$

Proof. We can rewrite random variable and (Haar) expectation as

$$S_U(M,\phi)=\mathrm{Tr}\big[M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\big]\quad\text{and}\quad\mu(M,\phi)=\mathrm{Tr}\left[\frac{\mathbb{I}}{d}\otimes\mathrm{Tr}_1(M)\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right]. \tag{C58}$$

Combine them to obtain

$$\bar{S}_U(M,\phi) = S_U(M,\phi) - \mu(M,\phi) = \mathrm{Tr}\left[\tilde{M}\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right], \tag{C59}$$

where $\tilde{M} = M - (1/d)\mathbb{I}\otimes \mathrm{Tr}_1(M) \in \mathbb{H}_d\otimes\mathbb{H}_d$ is a traceless difference of two PSD matrices. Next, fix $k\in\mathbb{N}$ and compare the kth centered moment to its Haar-averaged counterpart:

$$\mathbb{E}_{\mathcal{E}}\left[\bar{S}_U(M,\phi)^k\right] \le \mathbb{E}_U\left[\bar{S}_U(M,\phi)^k\right] + \underbrace{\{\mathbb{E}_{\mathcal{E}}\left[\bar{S}_U(M,\phi)^k\right] - \mathbb{E}_U\left[\bar{S}_U(M,\phi)^k\right]\}}_{\Delta}. \tag{C60}$$

The first contribution is bounded by Theorem 22 and the approximate k-design property (Definition 4) ensures that the mismatch Δ remains controlled:

$$\begin{aligned}\Delta &= \mathrm{Tr}\left(\tilde{M}^{\otimes k}\left\{\mathbb{E}_{\mathcal{E}}\left[(\mathcal{U}\otimes\mathcal{I})^{\otimes k}\right] - \mathbb{E}_U\left[(\mathcal{U}\otimes\mathcal{I})^{\otimes k}\right]\right\}\left[(|\phi\rangle\langle\phi|)^{\otimes k}\right]\right)\\ &\le \|\tilde{M}^{\otimes k}\|_\infty \left\|\left(\mathbb{E}_{\mathcal{E}}\left[\mathcal{U}^{\otimes k}\otimes\mathcal{I}\right] - \mathbb{E}_U\left[\mathcal{U}^{\otimes k}\otimes\mathcal{I}\right]\right)\left[(|\phi\rangle\langle\phi|)^{\otimes k}\right]\right\|_1\\ &\le \|\tilde{M}\|_\infty^k \left\|\mathbb{E}_{\mathcal{E}}\left[\mathcal{U}^{\otimes k}\right] - \mathbb{E}_U\left[\mathcal{U}^{\otimes k}\right]\right\|_\diamond = \|\tilde{M}\|_\infty^k \left\|\mathcal{T}_{\mathcal{E}}^{(k)} - \mathcal{T}_U^{(k)}\right\|_\diamond \le \|\tilde{M}\|_\infty^k \frac{k!}{d^{2k}}\epsilon.\end{aligned} \tag{C61}$$

Finally, use the fact that $\tilde{M}$ is the difference of two PSD matrices to conclude

$$\begin{aligned}\|\tilde{M}\|_\infty &\le \max\left\{\|M\|_\infty, \|\frac{1}{d}\mathbb{I}\otimes\mathrm{Tr}_1(M)\|_\infty\right\}\\ &= \max\left\{\|M\|_\infty, \frac{1}{d}\|\mathrm{Tr}_1(M)\|_\infty\right\} \le 1,\end{aligned} \tag{C62}$$

where we also use Eq. (C6). ■

Corollary 24 (Moments of *k*-design orbits): *For* $|x\rangle\in\mathbb{C}^d$ *and a measurement* $M\in\mathbb{H}_d$ $(\mathbb{I}\succeq M\succeq 0)$ *define*

$$\bar{Q}_U(M,x) = \langle x|U^\dagger M U|x\rangle - \frac{Tr(M)}{d}, \tag{C63}$$

where U *is sampled from an* ϵ*-approximate* k*-design. Then,*

$$\begin{aligned}\mathbb{E}_{\mathcal{E}}\left[\bar{Q}_U(M,x)^k\right] &\le \binom{d+k-1}{k}^{-1}\left(d^{k/2}+\epsilon\right)\\ &\le (1+\epsilon)\left(\frac{k^2}{d}\right)^{k/2}.\end{aligned} \tag{C64}$$

Proof. Let $\bar{M} = M - [\mathrm{Tr}(M)/d]\mathbb{I}$ denote the traceless part of M and note that this reformulation cannot increase the operator norm: $\|\bar{M}\|_\infty \le \|M\|_\infty \le 1$. Moreover,

$$\begin{aligned}\mathbb{E}_{\mathcal{E}}\left[\bar{Q}_U(M,x)^k\right] \le{}& \mathbb{E}_U\left[\bar{Q}_U(M,x)^k\right]\\ &+ \underbrace{\mathbb{E}_{\mathcal{E}}\left[\bar{Q}_U(M,x)^k\right] - \mathbb{E}_U\left[\bar{Q}_U(M,x)^k\right]}_{\Delta},\end{aligned} \tag{C65}$$

and $\Delta \le \|\bar{M}\|_\infty^k \binom{d+k-1}{k}^{-1}\epsilon$ follows from arguments that are analogous to the ones presented in the proof of Lemma 21. Next, apply Eq. (C43) to the remaining Haar expectation:

$$\begin{aligned}\mathbb{E}_U\left[\bar{Q}_U(\bar{M},x)\right] &= \mathbb{E}_U\left[\langle x|U^\dagger M U|x\rangle^k\right]\\ &= \binom{d+k-1}{k}^{-1}\mathrm{Tr}\left(\Pi_{\mathrm{sym}}\bar{M}^{\otimes k}\right).\end{aligned} \tag{C66}$$

This trace can be bounded using $\mathrm{tr}(\bar{M}) = 0$, $\mathrm{tr}(\bar{M}^l) \le \mathrm{tr}(\bar{M}^2)^{l/2}$ for $l\ge 2$ and $\mathrm{tr}(\bar{M}^2) = \|\bar{M}\|_2^2 \le \|M\|_2^2$, see, e.g., Ref. [21, Lemma 17]:

$$\mathrm{Tr}\left(\Pi_{\mathrm{sym}}\bar{M}^{\otimes k}\right) \le \|\bar{M}\|_2^k \le \|M\|_2^k \le d^{k/2}\|M\|_\infty^k \le d^{k/2}. \tag{C67}$$

■

9. Proof of the general moment bound

This section is devoted to proving the general moment bound presented in Theorem 22 in Appendix C 7.

a. Reformulation and basic norm bounds

Fix $M \in \mathbb{H}_d \otimes \mathbb{H}_d$ PSD with $\|M\|_\infty \leq 1$ and a state $|\phi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. Use the vectorization correspondence $|\phi\rangle = \mathrm{vec}(\Phi)$ with $\Phi \in \mathbb{M}_{d\times d}$ to rewrite the random variable defined in Theorem 22:

$$S_U(M,\phi) = \quad = \quad = . \tag{C68}$$

Here, we implicitly define $M_\Phi := (\mathbb{I} \otimes \Phi^\dagger) M (\mathbb{I} \otimes \Phi)$. Also, recall that vectorization is an isometry, i.e., $\|\Phi\|_2 = \langle\phi|\phi\rangle = 1$. The following auxiliary result bounds the 2-norm of M_Φ and its partial contractions.

Lemma 25: *Fix a PSD matrix $M \in \mathbb{H}_d^{\otimes 2}$ with $\|M\|_\infty \leq 1$ and a matrix $\Phi \in \mathbb{M}_d$ obeying $\|\Phi\|_2 = 1$. Then, $M_\Phi = (\mathbb{I} \otimes \Phi^\dagger) M (\mathbb{I} \otimes \Phi) \in \mathbb{H}_d^{\otimes 2}$ obeys*

$$\|Tr_1(M_\Phi)\|_2 \leq d \quad \textit{and} \quad \|M_\Phi\|_2 \leq \sqrt{d}, \quad \textit{as well as } \|Tr_2(M_\Phi)\|_2 \leq \sqrt{d}. \tag{C69}$$

Proof. Observe

$$\|\mathrm{Tr}_1(M_\Phi)\|_2^2 = \quad = \mathrm{Tr}\left(\Phi^\dagger \Phi \mathrm{Tr}_1(M) \Phi^\dagger \Phi \mathrm{Tr}_1(M)\right) =: h_1(\Phi^\dagger\Phi)\,. \tag{C70}$$

The function $X \mapsto h_1(X)$ is convex, according to Lemma 19 in Appendix C 2 above [$M \succeq 0$ implies $\mathrm{Tr}_1(M) \succeq 0$]. Moreover, $\rho = \Phi^\dagger\Phi \in \mathbb{H}_d$ is guaranteed to be a quantum state: $\rho = \Phi^\dagger\Phi \succeq 0$ and $\mathrm{Tr}(\rho) = \|\Phi\|_2^2 = 1$. The extreme points of the convex set of all quantum states are pure states. The convex function h_1 achieves its maximum value at such an extreme point (Fact 18 in Appendix C 2) and we infer

$$h_1(\Phi^\dagger\Phi) \leq \max_{\rho \text{ state}} h_1(\rho) = \max_{|\psi\rangle} h_1(|\psi\rangle\langle\psi|) = \max_{|\psi\rangle} \langle\psi|\mathrm{Tr}_1(M)|\psi\rangle^2 = \|\mathrm{Tr}_1(M)\|_\infty^2. \tag{C71}$$

Apply Eq. (C6) to conclude the first estimate: $\|\mathrm{Tr}_1(M)\|_\infty^2 \leq d^2\|M\|_\infty \leq d^2$. The second bound can be derived in a similar fashion. Observe,

$$\|M_\Phi\|_2^2 = \quad = \mathrm{Tr}\left(\Phi^\dagger\Phi \otimes \mathbb{I} M \mathbb{I} \otimes \Phi^\dagger\Phi M\right) = h_2(\Phi^\dagger\Phi)\,. \tag{C72}$$

The function $h_2(X)$ is again convex, because $X \mapsto \mathbb{I} \otimes X$ is a linear transformation and $M \succeq 0$. Moreover, $\rho = \Phi^\dagger\Phi$ is again a quantum state. We infer

$$h_2(\Phi^\dagger\Phi) \leq \max_{\rho \text{ state}} h_2(\rho) = \max_{|\psi\rangle} h_2(|\psi\rangle\langle\psi|) = \max_{|\psi\rangle} \|\mathrm{Tr}_2\left(\mathbb{I} \otimes |\psi\rangle\langle\psi| M\right)\|_2^2\,, \tag{C73}$$

because convex functions achieve their maximum at the boundary of convex sets (Fact 18). Applying the relation $\|X\|_2 \leq \sqrt{d}\|X\|_\infty$ for Schatten norms in $\mathbb{H}_d$, we conclude

$$\|M_\Phi\|_2^2 \leq d \max_{|\psi\rangle} \|\mathrm{Tr}_2\left(\mathbb{I} \otimes |\psi\rangle\langle\psi| M\right)\|_\infty^2 = d\left(\max_{|\psi\rangle,|x\rangle} (\langle x| \otimes \langle\psi|) M (|x\rangle \otimes |\psi\rangle)\right)^2 \leq d\|M\|_\infty^2. \tag{C74}$$

The final bound can be established directly. Set $\rho = \Phi^\dagger\Phi$ and observe

$$\|\mathrm{Tr}_2(M_\Phi)\|_2^2 = \quad = \|\mathrm{Tr}_2\left(\mathbb{I} \otimes \rho M\right)\|_2^2\,. \tag{C75}$$

Apply $\|X\|_2 \leq \sqrt{d}\|X\|_\infty$ to simplify further

$$\begin{aligned}\|\mathrm{Tr}_2(\mathbb{I}\otimes\rho M)\|_2 &\leq \sqrt{d}\|\mathrm{Tr}_2(\mathbb{I}\otimes\rho M)\|_\infty \leq \sqrt{d}\max_{|x\rangle}\left|\langle x|\mathrm{Tr}_2(\mathbb{I}\otimes\rho M)|x\rangle\right| \\ &= \sqrt{d}\max_{|x\rangle}\left|\mathrm{Tr}\left(|x\rangle\langle x|\otimes\rho M\right)\right|. \end{aligned} \tag{C76}$$

Finally, use matrix Hoelder (C4) to infer the advertised bound:

$$\sqrt{d}\max_{|x\rangle}|\mathrm{Tr}\left(|x\rangle\langle x|\otimes\rho M\right)| \leq \sqrt{d}\max_{|x\rangle}\||x\rangle\langle x|\otimes\rho\|_1\|M\|_\infty = \sqrt{d}\|M\|_\infty. \tag{C77}$$

■

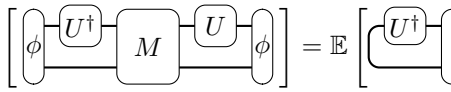
b. Expectation value and centering

The following result is well known in the literature, see, e.g., Ref. [22]. We include a self-contained derivation based on wiring diagrams for the sake of completeness.

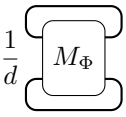
Lemma 26 (Averaging unitary channels produces the depolarizing channel): *Fix a PSD matrix $M \in \mathbb{H}_d^{\otimes 2}$ and $|\phi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. Let $\mathcal{U}(X) = UXU^\dagger$ be a Haar-random unitary channel. Then,*

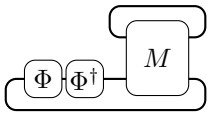
$$\mathbb{E}_U\{Tr\left[M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right]\} = Tr\left[M\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right] \quad \textit{with } \mathcal{D}(\rho) = \frac{Tr(\rho)}{d}\mathbb{I}. \tag{C78}$$

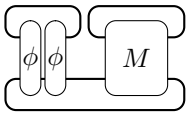
Proof. Averaging over a single unitary U and its adjoint decouples the register in question. Combine this with the reformulation from the previous subsection to conclude

$\mathbb{E}_U\left[\mathrm{Tr}\left(M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right)\right] =$ 

$=$

$= \frac{1}{d}$ 

$= \frac{1}{d}$ 

$= \frac{1}{d}$ 

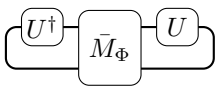
. (C79)

The connection to the depolarizing channel readily follows from $\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|) = (\mathbb{I}/d)\otimes\mathrm{Tr}_2(|\phi\rangle\langle\phi|)$. ■

Corollary 27 (Reformulation of the centered random variable): *Fix $|\phi\rangle \in \mathbb{C}^d\otimes\mathbb{C}^d$ (state) and $M \in \mathbb{H}_d^{\otimes 2}$ such that $\mathbb{I} \succeq M \succeq 0$ (measurement). For channels $\mathcal{U}(X) = UXU^\dagger$ and $\mathcal{D}(X) = [Tr(X)/d]\mathbb{I}$ define*

$$S_U(M,\phi) = Tr\left[M\mathcal{U}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right], \quad \textit{as well as} \quad \mu(M,\phi) = Tr\left[M\mathcal{D}\otimes\mathcal{I}(|\phi\rangle\langle\phi|)\right].$$

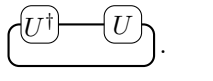
Then, we may rewrite the difference of these variables as

$\bar{S}_U(M,\phi) = S_U(M,\phi) - \mu(M,\Phi) =$ 

, (C80)

where $\bar{M}_\Phi = M_\Phi - [Tr(M_\Phi)/d]\mathbb{I} \in \mathbb{H}_d^{\otimes 2}$ is the traceless part of M_Φ [i.e., $Tr(\bar{M}_\Phi) = 0$].

This reformulation immediately follows from the proof of Lemma 26, provided that we rewrite

$\mu(M,\phi) = \frac{1}{d}\mathrm{Tr}(M_\Phi) = \frac{\mathrm{Tr}(M_\Phi)}{d^2}$ 

. (C81)

c. Bounds on centered moments

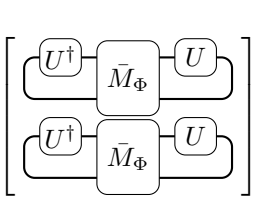
Lemma 28: *With the same assumptions and notation as in Corollary 27, suppose that $U \in U(d)$ is chosen uniformly from the Haar measure. Then, for any $k \leq d^{2/3}$*

$$\mathbb{E}_U\left[\bar{S}_U(M,\phi)^k\right] \leq C_k \frac{(k!)^2}{d^{k/2}}, \tag{C82}$$

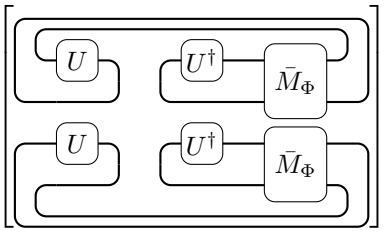
where $C_k = [1/(k-1)]\binom{2k}{k}$ is the kth Catalan number.

Proof. It is instructive to first analyze and understand the second moment:

$$\mathbb{E}_U\left[\bar{S}_U(M,\phi)^2\right] = \mathbb{E}_U$$

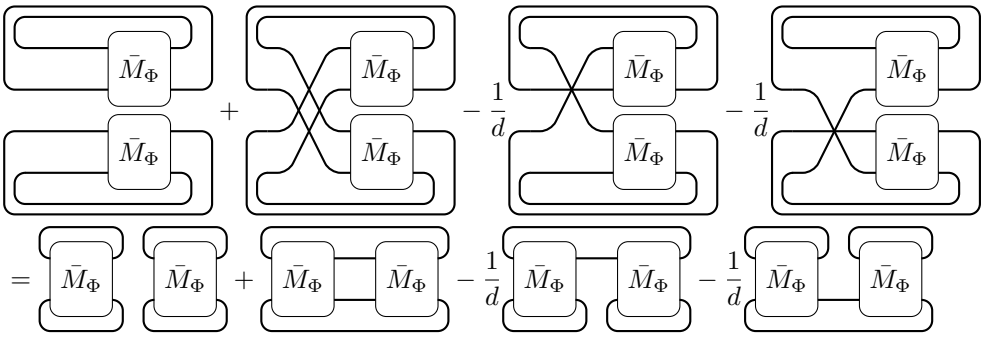


$= \mathbb{E}_U$



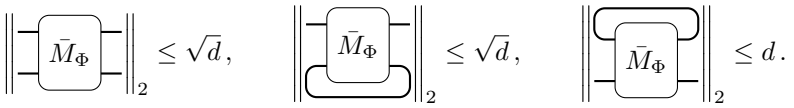
. (C83)

For $k = 2$ there are two permutations: the identity permutation $\mathbb{I} = \{1,2\}$ and swap (or flip) $S = \{2,1\}$. This results in $(k!)^2 = 4$ different contributions to the formula: $(\mathbb{I},\mathbb{I})$, (S,S), $(S,\mathbb{I})$, and $(\mathbb{I},S)$ contribute each. The associated Weingarten functions are $\mathcal{W}g[(1),d] = 1/(d^2-1)$ and $\mathcal{W}g[(2),d] = -[1/d(d^2-1)]$. Ignoring the common factor $1/(d^2-1)$, the individual contributions become



(C84)

Each term is a full contraction that is also called a tensor network [41,42]. There are three possible constituents for each tensor network: $\bar{M}_\Phi$, $\mathrm{Tr}_2(\bar{M}_\Phi)$, and $\mathrm{Tr}_1(\bar{M}_\Phi)$. Importantly, no full self-contractions can contribute to the overall sum, because $\bar{M}_\phi$ is traceless. This ensures that networks with self-contractions—like the first term—evaluate to zero. Moreover, Lemma 25 bounds the 2-norm of each elementary constituent:



(C85)

The final bound is considerably larger than the rest. However, the corresponding contribution in the sum (C84) is also suppressed by an additional dimension factor. This is not a coincidence: term 3 can arise only if the cycle classes of (σ,τ) differ from each other. This feature reflects itself in the Weingarten function. For the second moment, we thus obtain the following simple bound (ignoring signs):

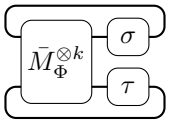
$$\mathbb{E}_U\left[\bar{S}(M,\phi)^2\right] \leq \frac{0+d+d/d+d^2/d}{d^2-1} = \frac{2d+1}{d^2} \leq 4d^{-1}. \tag{C86}$$

It immediately follows from upper bounding individual terms using Eq. (C85).

This general strategy also applies to higher moments. Fix $k \geq 3$ arbitrary. Then, Weingarten calculus implies

$$\mathbb{E}_U\left[\bar{S}_U(M,\phi)^k\right] = \sum_{\sigma,\tau\in S_k} \mathcal{W}g_d(\sigma,\tau) N_{\sigma,\tau} \times \left[\bar{M}_\Phi, \mathrm{Tr}_2(\bar{M}_\Phi), \mathrm{Tr}_1(\bar{M}_\Phi)\right]. \quad \text{(C87)}$$

Here, each $N_{\sigma,\tau}(\cdot)$ indicates a tensor network diagram that combines (at most) three elementary building blocks according to rules that are dictated by the permutations τ and σ:

$N_{\sigma,\tau} =$  . (C88)

We can without loss restrict summation to tensor networks without self-contractions, because $\mathrm{Tr}(\bar{M}_\Phi) = 0$ ensures that such contributions vanish identically. Next, we apply a powerful general bound to individual tensor networks. Reference [42, Proposition 18] states that the value of a tensor network (without self-contractions) is bounded by the product of 2-norms of the individual constituents. For any σ, τ this implies

$$|N_{\sigma,\tau}| = \left|N_{\sigma,\tau}\left[\bar{M}_\Phi, \mathrm{Tr}_2(\bar{M}_\Phi), \mathrm{Tr}_1(\bar{M}_\Phi)\right]\right| \leq \|\bar{M}_\Phi\|_2^{\nu_1} \|\mathrm{Tr}_2(\bar{M}_\Phi)\|_2^{\nu_2} \|\mathrm{Tr}_1(\bar{M}_\Phi)\|_2^{\nu_3}, \quad \text{(C89)}$$

where $\nu_1, \nu_2, \nu_3 \in [k]$ denote the number of times each basic building block occurred in the network. Clearly, $\nu_1 + \nu_2 + \nu_3 = k$ and we can combine this with Eq. (C85) to conclude

$$|N_{\sigma,\tau}| \leq d^{\nu_1/2} d^{\nu_2/2} d^{\nu_3} = d^{k/2+\nu_3/2}. \quad \text{(C90)}$$

The final contribution $d^{\nu_3/2}$ is always counterbalanced by the Weingarten function, i.e., the dangerous terms are always suppressed by powers of $1/d$. As we discuss, the Weingarten functions $\mathcal{W}g(\sigma,d)$ depend only on the cycle type of the permutation σ. The asymptotic behavior is $\mathcal{W}g(\sigma,d) \sim 1/d^{2k-\ell(\sigma)}$, where ℓ is the length of the cycle type, i.e., the number of cycles in the permutation. The leading-order terms are those for which the cycle type is $(1,1,\ldots,1)$, the partition of $2k$ into 1's. For $\mathcal{W}g(\sigma^{-1}\tau,d)$ this corresponds to terms with $\sigma = \tau$. Returning to the problem at hand, we contract the upper indices of the k copies of $\bar{M}_\Phi$ with respect to σ and the lower indices with τ, as shown in Eq. (C88). The leading-order terms are those in which we act similarly on upper and lower indices. In order to generate terms in the tensor-network contraction of M's containing a dangerous contribution, $\mathrm{Tr}_1(\bar{M}_\Phi)$, the lengths of the cycle types of the two permutations must differ by at least one in order to generate a contraction, a length one cycle, in the σ indices:

$$\mathbb{E}_U\left[\bar{S}_U(M,\phi)^k\right] \leq \sum_{\tau,\sigma\in S_k} |\mathcal{W}g(\sigma^{-1}\tau,d)| N_{\sigma,\tau} \leq \sum_{\sigma\in S_k} \mathcal{W}g[(1,\ldots,1),d] d^{k/2} + \sum_{\tau\neq\sigma\in S_k} \mathcal{W}g(\sigma^{-1}\tau,d) d^{k/2+\nu_3/2}. \quad \text{(C91)}$$

Although, the $\mathrm{Tr}_1(\bar{M}_\Phi)$ terms will only contribute at subleading order, they appear with a larger contribution in powers of d. Thus, to rigorously upper bound the expression, we need bounds on the Weingarten functions as well as on the number of terms ν_3 which appear in a given tensor network $N_{\sigma,\tau}$.

Precise upper bounds on the Weingarten functions are known [40,81]. For our purposes, it will be convenient to use the (slightly weaker) bound in Ref. [82], which states that for $k \leq d^{2/3}$

$$|\mathcal{W}g(\sigma,d)| \leq \frac{3}{2}\frac{C_{k-1}}{d^{2k-\ell(\sigma)}}, \quad \text{(C92)}$$

where C_k is the kth Catalan number.

Now we establish that $\nu_3(\sigma,\tau)$, the number of dangerous terms $\mathrm{Tr}_1(\bar{M}_\Phi)$ terms in a given $N_{\sigma,\tau}$, is bounded by the distance between the permutations σ and τ as $\nu_3(\sigma,\tau) \leq 2d(\sigma,\tau)$. First we note a few facts about the symmetric group. $d(\sigma,\tau)$ is defined as the minimal number of transpositions needed to take σ to τ, and defines a distance between the permutations. Specifically, $d(\sigma,\tau)$ is a metric on the Cayley graph of the symmetric group with the generating set of transpositions. The length of the cycle type of a permutation $\sigma \in S_k$ is related to the number of transpositions needed to build σ from the identity permutation as $\ell(\sigma) = k - d(\sigma,\mathbb{I})$. Furthermore, a transposition changes the number of cycles in a permutation by exactly one.

Recalling Eq. (C88), the terms $\mathrm{Tr}_1(\bar{M}_\Phi)$ appear only in a given tensor-network diagram when the permutation σ has a fixed point where τ does not, i.e., there is a self-contraction in the σ indices of $\bar{M}_\Phi^{\otimes k}$ and not the τ indices. This implies that σ has a length one cycle at a point where τ does not. As $d(\sigma,\tau)$ is the minimal number of transpositions required to take σ to τ, and a transposition can only change the number of cycles by exactly 1, then for every two dangerous terms the distance between the permutations σ and τ must increase by at least one. This shows that $\nu_3(\sigma,\tau)$ is bounded as

$$\nu_3(\sigma,\tau) \leq 2d(\sigma,\tau) = 2[k - \ell(\sigma^{-1}\tau)]. \quad \text{(C93)}$$

Returning to the general moment bound, we can apply the bound on Weingarten functions in Eq. (C92) and the bound

on ν_3 to show that

$$\begin{aligned}\mathbb{E}_U\left[\bar{S}_U(M,\phi)^k\right] &\le \sum_{\tau,\sigma\in S_k} |\mathcal{W}g(\sigma^{-1}\tau,d)|N_{\sigma,\tau} \\ &\le \sum_{\tau,\sigma\in S_k} \frac{3}{2}C_{k-1}d^{\ell(\sigma^{-1}\tau)-2k+k/2+\nu_3/2} \\ &\le \sum_{\tau,\sigma\in S_k} \frac{3}{2}C_{k-1}d^{-k/2} \le C_k(k!)^2 d^{-k/2}, \end{aligned} \tag{C94}$$

which establishes the claim. ■

10. ε-coverings of local random circuits

We want to extend our results in Sec. III A on complexity growth to local random circuits, where the gates are chosen Haar randomly from $U(q^2)$. Obviously, the ensemble of size T circuits is continuous and statements about the number of states of a certain complexity become less meaningful. Nevertheless, we can consider an ε-covering of the ensemble of local random quantum circuits (RQCs) in order to make concrete statements about complexity growth.

We say that a set of unitaries V is an ε-covering of a set of unitaries U if for all $U\in\mathsf{U}$ there is some $V\in\mathsf{V}$ such that $\|U(\cdot)U^\dagger - V(\cdot)V^\dagger\|_\diamond \le \varepsilon$.

Consider the set of local random circuits of size T, where again we act on n local qudits with local dimension q and with local gates chosen Haar randomly from $U(q^2)$. Following Lemma 27 from Ref. [12], we can bound the size of an ε-covering of the set $\mathcal{E}_{\mathrm{RQC}}$ size T local RQCs. Approximating each local gate to accuracy ε/T, we construct a covering in diamond norm of each gate with size $\le (10T/\varepsilon)^{q^4}$. For the n^T choices of gates in the circuit, we conclude that there exists an ε-covering $\widetilde{\mathcal{E}}_{\mathrm{RQC}}$ of size T RQCs with cardinality

$$|\widetilde{\mathcal{E}}_{\mathrm{RQC}}| \le n^T\left(\frac{10T}{\varepsilon}\right)^{Tq^4}. \tag{C95}$$

Furthermore, if an ensemble $\mathcal{E}$ forms an ϵ-approximate unitary k-design, then the ε-covering of $\mathcal{E}$ will form an ϵ'-approximate unitary design with $\epsilon' = \epsilon + 2d^{2k}\varepsilon$ (from Proposition 8 in Ref. [12]). Using the lower bound on the cardinality of an approximate design in Lemma 20 and the upper bound on the cardinality of an ε-covering of size T local random circuits in Eq. (C95), means that for $\tilde{\mathcal{E}}_{\mathrm{RQC}}$ to form an approximate design, we must have

$$\frac{1}{1+\epsilon'}\frac{d^{2k}}{k!} \le |\widetilde{\mathcal{E}}_{\mathrm{RQC}}| \le n^T\left(\frac{10T}{\varepsilon}\right)^{Tq^4}. \tag{C96}$$

This gives a lower bound on the size for local random circuits to form k-designs

$$T \ge \frac{2kn\log q}{q^4\log k}. \tag{C97}$$

Therefore, an optimal random circuit implementation of a unitary design will have at least an essentially linear scaling in both n and k.

APPENDIX D: CONCENTRATION OF MEASURE FOR HAAR-UNIFORM VECTORS

Proposition 29: *Fix $M\in\mathbb{H}_d$ with $\|M\|_\infty \le 1$ and suppose that $|\psi\rangle\in\mathbb{C}^d$ is chosen uniformly from the complex unit sphere. Then,*

$$\begin{aligned}&Pr\left[|\langle\psi|M|\psi\rangle - \mathbb{E}\left(\langle\psi|M|\psi\rangle\right)| \ge \tau\right] \\ &\quad\le 2\exp\left(-\frac{d\tau^2}{9\pi^3}\right) \quad \textit{for any } \tau \ge 0. \end{aligned} \tag{D1}$$

The proof is standard and we include it in this Appendix for completion. It is based on Levy's lemma, i.e., concentration of measure on the real-unit sphere $\mathbb{S}^{2d-1}\subset\mathbb{R}^{2d}$. A function $f:\mathbb{S}^{2d-1}\to\mathbb{R}$ is L-Lipschitz (with respect to the Euclidean norm $\|\cdot\|_{\ell_2}$ on $\mathbb{R}^{2d}$) if

$$|f(x)-f(y)| \le L\|x-y\|_{\ell_2} \quad \text{for all } x,y\in\mathbb{S}^{2d-1}. \tag{D2}$$

Theorem 30 (Levy's lemma): *Let $f:\mathbb{S}^{2d-1}\to\mathbb{R}$ be a L-Lipschitz function on the unit sphere. Then, the following relation is true if x is chosen uniformly from $\mathbb{S}^{2d-1}$:*

$$Pr\{|f(x)-\mathbb{E}[f(x)]| \ge \tau\} \le 2\exp\left(-\frac{4d\tau^2}{9\pi^3L^2}\right). \tag{D3}$$

Proof of Proposition 29. The complex unit sphere in d dimensions admits an isometric embedding—with respect to the Euclidean norm—onto the real-valued unit sphere $\mathbb{S}^{2d-1}\subseteq\mathbb{R}^{2d}$:

$$|\psi\rangle \mapsto |x\rangle = \mathrm{Re}(|\psi\rangle)\rangle \oplus \mathrm{Im}(|\psi\rangle) \in \mathbb{S}^{2d-1}. \tag{D4}$$

This embedding maps the uniform distribution on the complex unit sphere in $\mathbb{C}^d$ to the uniform distribution on the real-valued unit sphere in $\mathbb{R}^{2d}$. Under this embedding, the function of interest $\langle\psi|M|\psi\rangle$ becomes

$$\begin{aligned}\langle\psi|M|\psi\rangle &= \langle\mathrm{Re}(\psi)|M|\mathrm{Re}(\psi)\rangle \\ &\quad + \langle\mathrm{Im}(\psi)|M|\mathrm{Im}(\psi)\rangle = \langle x|M\oplus M|x\rangle, \end{aligned} \tag{D5}$$

because M is Hermitian. Its expectation is also preserved and Lemma 31 immediately below states that this function

is Lipschitz with constant $2\|M\|_\infty \leq 2$. The claim then readily follows from Levy's lemma (Theorem 30). ■

Lemma 31: *Fix $M \in \mathbb{H}_d$. Then, the following relation is true for any pair of unit-norm vectors $x, y \in \mathbb{S}^{2d-1} \subset \mathbb{R}^{2d}$*

$$|\langle x|M \oplus M|x\rangle - \langle y|M \oplus M|y\rangle| \leq 2\|M\|_\infty \|x - y\|_{\ell_2}. \tag{D6}$$

Proof. Fix $x, y \in \mathbb{S}^{2d-1}$ and apply Hoelder's inequality:

$$\begin{aligned} &|\langle x|M \oplus M|x\rangle - \langle y|M \oplus M|y\rangle|^2 \\ &\quad = \mathrm{Tr}\,[M \oplus M(|x\rangle\langle x| - |y\rangle\langle y|)]^2 \\ &\quad \leq \|M \oplus M\|_\infty^2 \||x\rangle\langle x| - |y\rangle\langle y|\|_1^2. \end{aligned} \tag{D7}$$

The block structure of $M \oplus M$ ensures $\|M \oplus M\|_\infty = \|M\|_\infty$, while the remaining term is the trace norm of a difference of pure states. This can be computed analytically and we obtain

$$\begin{aligned} &\||x\rangle\langle x| - |y\rangle\langle y|\|_1^2 \\ &\quad = 4\left(1 - \langle x, y\rangle^2\right) = 4\,(1 + \langle x, y\rangle)\,(1 - \langle x, y\rangle) \\ &\quad \leq 4\,(2 - 2|\langle x, y\rangle|)\,, \end{aligned} \tag{D8}$$

because $\langle x, y\rangle \leq \|x\|_{\ell_2}\|y\|_{\ell_2} \leq 1$ Finally,

$$\begin{aligned} 2 - 2\langle x, y\rangle &= \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle \\ &= \langle x - y, x - y\rangle = \|x - y\|_{\ell_2}^2, \end{aligned} \tag{D9}$$

and the claim follows. ■

APPENDIX E: DESIGNS AND THE TRADITIONAL DEFINITION OF COMPLEXITY

In the bulk of the paper we focus on a stronger notion of complexity than the standard definition, an operational definition involving the complexity of the distinguishing measurement to differentiate the state from the maximally mixed state. A more traditional definition is often considered in the literature, which involves building a quantum circuit that approximates the state when evolved from an initial state. This intuitive notion of complexity is related to the minimal size of such a circuit.

In this Appendix, we work through the counting arguments in Appendix A for the complexity of elements of a k-design using the more traditional (albeit weaker) definition of complexity. We refer to this as the *weak complexity* of a state or unitary to distinguish it from the operational definitions presented in Sec. II A.

Consider a system of n qudits with local dimension q, such that the total dimension is $d = q^n$. Let $\mathsf{G} \subset U(q^2)$ denote a universal gate set of elementary 2-local gates, and let G_r be the set of circuits of size r built from our gate set G.

Definition 5 (Weak δ-state complexity): *For $\delta \in [0, 1]$, we say that a state $|\psi\rangle$ has δ-state complexity of at most r if there exists a unitary circuit $V \in \mathsf{G}_r$ such that*

$$\begin{aligned} &\frac{1}{2}\left\||\psi\rangle\langle\psi| - V|0\rangle\langle 0|V^\dagger\right\|_1 \leq \delta, \quad \textit{which we denote as} \\ &\quad \times \mathcal{C}'_\delta(|\psi\rangle) \leq r. \end{aligned}$$

We want to be able to make precise statements about the complexity of sets of states. More specifically, if we consider a complex projective design, the requirement that they form a k-design is sufficiently restrictive to deduce a quantitative statement about the complexity of the constituent states.

Theorem 32 (Weak complexity of state designs): *Consider an ϵ-approximate complex projective k-design $\mathcal{E} = \{p_i, |\psi_i\rangle\}_{i=1}^N$. Then there are at least*

$$\frac{d^k}{k!}\frac{1}{1+\epsilon} - \frac{n^r|\mathsf{G}|^r}{(1-\delta^2)^k}, \tag{E1}$$

states with weak δ-state complexity $\mathcal{C}'_\delta(|\psi_i\rangle) > r$.

The number of high complexity states is exponentially large in k for complexity

$$r \lesssim \frac{k(n - \log k)}{\log n}. \tag{E2}$$

Turning now to the complexity of unitaries, the traditional definition of complexity is the minimal size of a circuit, built from our gate set, which approximates that unitary.

Definition 6 (Weak δ-unitary complexity): *For $\delta \in [0, 1]$, we say that a unitary U has δ-unitary complexity of at most r if there exists a circuit $V \in \mathsf{G}_r$ such that*

$$\frac{1}{2}\left\|\mathcal{U} - \mathcal{V}\right\|_\diamond \leq \delta, \quad \textit{which we denote as } \mathcal{C}'_\delta(U) \leq r,$$

where $\mathcal{U}(\rho) = U\rho U^\dagger$ and $\mathcal{V}(\rho) = V\rho V^\dagger$.

Again, we ask if the structure of a unitary k-design allows us to conclude anything about the complexity of unitaries. Once more, we find that we can turn the statement that k-design elements have a certain expected complexity into a quantitative one.

Theorem 33 (Weak complexity of unitary designs): *Consider an ϵ-approximate unitary k-design $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$. Then there are at least*

$$\frac{d^{2k}}{k!}\frac{1}{1+\epsilon} - \frac{n^r|\mathsf{G}|^r}{(1-\delta^2)^k}, \tag{E3}$$

unitaries in $\mathcal{E}$ with weak δ-unitary complexity $\mathcal{C}'_\delta(U_i) > r$.

The number of high-complexity unitaries is again exponentially large in k for complexity less than

$$r \lesssim \frac{k(2n - \log k)}{\log n}. \tag{E4}$$

We now provide details and proofs of the above statements about the complexity of spherical and unitary designs.

1. Weak state complexity for spherical designs

Proof of Theorem 32. First, as stated in Lemma 6, we note that the definition of weak δ-state complexity in Definition 5 is equivalently written as

$$|\langle\psi|V|0\rangle|^2 \geq 1 - \delta^2. \tag{E5}$$

We can show this by first noting that $X := |\psi\rangle\langle\psi| - V|0\rangle\langle 0|V^\dagger$ has rank at most two. Directly computing the eigenvalues of X from

$$\begin{aligned} \mathrm{Tr}(X) &= \lambda_1 + \lambda_2 = 0 \quad \text{and} \\ \mathrm{Tr}(X^2) &= \lambda_1^2 + \lambda_2^2 = 2 - 2|\langle\psi|V|0\rangle|^2, \end{aligned} \tag{E6}$$

we find $\lambda_{1,2} = \pm\sqrt{1 - |\langle\psi|V|0\rangle|^2}$. Then as $\|X\|_1 = |\lambda_1| + |\lambda_2|$ we have that

$$\frac{1}{2}\left\||\psi\rangle\langle\psi| - V|0\rangle\langle 0|V^\dagger\right\|_1 = \sqrt{1 - |\langle\psi|V|0\rangle|^2}, \tag{E7}$$

from which the claim follows.

We want to ask, given some state $|\psi\rangle$ chosen uniformly from an ϵ-approximate spherical k-design, what is the probability that the state has δ complexity at most r: $\mathcal{C}'_\delta(|\psi\rangle) \leq r$? We know that the state will have δ complexity r if there exists a $V \in \mathsf{G}_r$ such that Eq. (E5) holds. A union bound then gives that

$$\begin{aligned} \Pr\left[\mathcal{C}'_\delta(|\psi\rangle) \leq r\right] &= \Pr\left[\bigcup_{V\in\mathsf{G}_r}\left\{|\langle\psi|V|0\rangle|^2 \geq 1-\delta^2\right\}\right] \\ &\leq \sum_{V\in\mathsf{G}_r}\Pr\left[|\langle\psi|V|0\rangle|^2 \geq 1-\delta^2\right]. \end{aligned} \tag{E8}$$

We can bound the probability that a state drawn from a spherical k-design satisfies Eq. (E5) as a straightforward consequence of Markov's inequality:

$$\begin{aligned} &\Pr\left[|\langle\psi|V|0\rangle|^2 \geq 1-\delta^2\right] \\ &\quad= \Pr\left[|\langle\psi|V|0\rangle|^{2k} \geq \left(1-\delta^2\right)^k\right] \\ &\quad\leq \frac{\mathbb{E}_{|\psi\rangle}\left[|\langle\psi|V|0\rangle|^{2k}\right]}{(1-\delta^2)^k} \leq \frac{(1+\epsilon)\binom{d+k-1}{k}^{-1}}{(1-\delta^2)^k}. \end{aligned} \tag{E9}$$

In the last step here, we use Eq. (C43) and proceeding similarly as in the proof of Lemma 21 in Appendix C 6, noting that for a fixed state $|\phi\rangle$ and $|\psi\rangle$ averaged over an ϵ-approximate spherical k-design, we have

$$\mathbb{E}_{|\psi\rangle}\left[|\langle\psi|\phi\rangle|^{2k}\right] \leq (1+\epsilon)\binom{d+k-1}{k}^{-1}. \tag{E10}$$

This claim readily follows from an argument similar to the proof of Lemma 21. Returning to Eq. (E8), we find that the probability that a state in a spherical design has complexity of at most r is

$$\Pr\left[\mathcal{C}'_\delta(|\psi\rangle) \leq r\right] \leq (1+\epsilon)\binom{d+k-1}{k}^{-1}\frac{n^r|\mathsf{G}|^r}{(1-\delta^2)^k}, \tag{E11}$$

using the bound on the expectation and a bound on the cardinality of G_r.

We now turn to proving the primary claim. Negating the above assertion implies that

$$\Pr\left[\mathcal{C}'_\delta(|\psi\rangle) > r\right] \geq 1 - (1+\epsilon)\binom{d+k-1}{k}^{-1}\frac{n^r|\mathsf{G}|^r}{(1-\delta^2)^k}. \tag{E12}$$

Furthermore, we may also write this probability as the expectation of the associated event, which yields

$$\begin{aligned} \Pr\left[\mathcal{C}'_\delta(|\psi\rangle) > r\right] &= \mathbb{E}_{|\psi\rangle}\left[\mathbb{1}\left\{\mathcal{C}'_\delta(|\psi\rangle) > r\right\}\right] \\ &= \sum_i p_i\,\mathbb{1}\left\{\mathcal{C}'_\delta(|\psi_i\rangle) > r\right\} \\ &\leq (1+\epsilon)\binom{d+k-1}{k}^{-1} N, \end{aligned} \tag{E13}$$

where $\mathbb{1}$ is the indicator function, and in the last step we use the bound on the weights of an ϵ-approximate spherical k-design in Lemma 21. N denotes the number of states in the spherical design $|\psi_i\rangle$ with weak δ complexity greater than r. Combining the previous two equations, we find that

$$N \geq \frac{d^k}{k!}\frac{1}{1+\epsilon} - \frac{n^r|\mathsf{G}|^r}{(1-\delta^2)^k}, \tag{E14}$$

which completes the proof. ■

2. Weak unitary complexity for unitary designs

Proof of Theorem 33. We start by noting an equivalent definition of weak δ-unitary complexity as shown in the proof of Lemma 7. A necessary, but in general not sufficient, condition for weak unitary complexity in Definition 6 is

$$\left|\mathrm{Tr}(V^\dagger U)\right|^2 \geq d^2(1-\delta^2). \tag{E15}$$

Now we again ask, given some unitary U chosen uniformly from an ϵ-approximate unitary k-design, what is the probability that it has δ-unitary complexity at most r: $\mathcal{C}'_\delta(U) \leq r$? As this holds if there exists a $V \in \mathsf{G}_r$ such that the channels are close in diamond distance, a union bound then gives that

$$\begin{aligned}\Pr\left[\mathcal{C}'_\delta(U) \leq r\right] &= \Pr\left[\bigcup_{V\in G_r}\left\{\frac{1}{2}\left\|\mathcal{U}-\mathcal{V}\right\|_\diamond \leq \delta\right\}\right] \\ &\leq \sum_{V\in G_r}\Pr\left[\left|\mathrm{Tr}(V^\dagger U)\right|^2 \geq d^2(1-\delta^2)\right],\end{aligned} \tag{E16}$$

using the reformulation above. We can bound the probability that a unitary drawn from a k-design satisfies this condition again by using Markov's inequality:

$$\begin{aligned}&\Pr\left[\left|\mathrm{Tr}(V^\dagger U)\right|^2 \geq d^2(1-\delta^2)\right] \\ &= \Pr\left[\left|\mathrm{Tr}(V^\dagger U)\right|^{2k} \geq d^{2k}(1-\delta^2)^k\right] \\ &\leq \frac{\mathbb{E}_{\mathcal{E}}\left[\left|\mathrm{Tr}(V^\dagger U)\right|^{2k}}{d^{2k}(1-\delta^2)^k} \leq \frac{(1+\epsilon)k!}{d^{2k}(1-\delta^2)^k},\end{aligned} \tag{E17}$$

where in the last step, we use the moments of traces for unitary designs and as in Lemma 20 in Appendix C 6 above find that for a fixed unitary V and a unitary U averaged over an ϵ-approximate unitary k-design, we have

$$\mathbb{E}_{\mathcal{E}}\left[\mathrm{Tr}(V^\dagger U)\right|^{2k}\right] \leq (1+\epsilon)k!. \tag{E18}$$

Returning to the expression above in Eq. (E16), we find that the probability $\mathcal{C}'_\delta(U) \leq r$ is

$$\Pr\left[\mathcal{C}'_\delta(U) \leq r\right] \leq (1+\epsilon)\frac{k!}{d^{2k}}\frac{r^r|\mathsf{G}|^r}{(1-\delta^2)^k}, \tag{E19}$$

using the bound on the expectation and a bound on the cardinality of G_r. Negating the expression gives a lower bound on the probability that a unitary in a k-design has complexity greater than r. Furthermore, we may also write this probability as the expectation

$$\Pr\left[\mathcal{C}'_\delta(U) > r\right] = \sum_i p_i \mathbb{1}\left\{\mathcal{C}'_\delta(U_i) > r\right\} \leq (1+\epsilon)\frac{k!}{d^{2k}}N, \tag{E20}$$

where we use the bound on the unitary design weights in Lemma 20. N denotes the number of untiaries in a k-design with weak δ complexity greater than r. Combining the previous two equations, we find that

$$N \geq \frac{d^{2k}}{k!}\frac{1}{1+\epsilon} - \frac{r^r|\mathsf{G}|^r}{(1-\delta^2)^k}, \tag{E21}$$

which completes the proof. ■

[1] D. Poulin, A. Qarry, R. Somma, and F. Verstraete, Quantum Simulation of Time-Dependent Hamiltonians and the Convenient Illusion of Hilbert Space, Phys. Rev. Lett. **106**, 170501 (2011).

[2] E. Bernstein and U. Vazirani, Quantum complexity theory, SIAM J. Comput. **26**, 1411 (1997).

[3] X. Chen, Z. C. Gu, and X. G. Wen, Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order, Phys. Rev. **B82**, 155138 (2010).

[4] L. Susskind, Computational complexity and black hole horizons, Fortsch. Phys. **64**, 44 (2016), [Fortsch. Phys. **64**, 24 (2016)].

[5] D. Stanford and L. Susskind, Complexity and shock wave geometries, Phys. Rev. **D90**, 126007 (2014).

[6] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, Complexity, action, and black holes, Phys. Rev. **D93**, 086006 (2016).

[7] A. R. Brown and L. Susskind, Second law of quantum complexity, Phys. Rev. **D97**, 086015 (2018).

[8] L. Susskind, Black holes and complexity classes, ArXiv:1802.02175.

[9] S. Aaronson, The complexity of quantum states and transformations: From quantum money to black holes, ArXiv:1607.05256.

[10] T. C. Bohdanowicz and F. G. S. L. Brandão, Universal Hamiltonians for exponentially long simulation, ArXiv:1710.02625.

[11] D. A. Roberts and B. Yoshida, Chaos and complexity by design, JHEP **04**, 121 (2017).

[12] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).

[13] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. **A80**, 012304 (2009).

[14] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, J. Math. Phys. **48**, 052104 (2007).

[15] Z. Webb, The clifford group forms a unitary 3-design, Quantum Info. Comput. **16**, 1379 (2016).

[16] H. Zhu, Multiqubit clifford groups are unitary 3-designs, Phys. Rev. A **96**, 062336 (2017).

[17] R. Kueng and D. Gross, Qubit stabilizer states are complex projective 3-designs, ArXiv:1510.02767.

[18] A. Ambainis and J. Emerson, in *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)* (2007), p. 129.

[19] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner, Decoupling with unitary approximate two-designs, New J. Phys. **15**, 053022 (2013).

[20] A. J. Scott, Tight informationally complete quantum measurements, J. Phys. A: Math. Gen. **39**, 13507 (2006).

[21] R. Kueng, H. Rauhut, and U. Terstiege, Low rank matrix recovery from rank one measurements, Appl. Comput. Harmon. Anal. **42**, 88 (2017).

[22] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B: Quantum Semiclass. Opt **7**, S347 (2005).

[23] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, JHEP **09**, 120 (2007).

[24] Note that here we discuss the size of the circuit, if we parallelize the application of gates, the depth of the circuit required to form an approximate design scales linearly in n.

[25] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Towards the fast scrambling conjecture, JHEP **04**, 022 (2013).

[26] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert, Mixing properties of stochastic quantum Hamiltonians, Commun. Math. Phys. **355**, 905 (2017).

[27] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics, Phys. Rev. **X7**, 021006 (2017).

[28] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits, ArXiv:1905.12053.

[29] J. Haferkamp and N. Hunter-Jones, Improved spectral gaps for random quantum circuits: Large local dimensions and all-to-all interactions, ArXiv:2012.05259.

[30] A. S. Holevo, Optimal quantum measurements, Teoret. Mat. Fiz. **17**, 319 (1973).

[31] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Mathematics in Science and Engineering (Academic Press, New York, NY, 1976).

[32] C. M. Dawson and M. A. Nielsen, The Solovay-Kitaev algorithm, Quantum Info. Comput. **6**, 81 (2006).

[33] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[34] For $q = 2$ a depth-two circuit comprised of n Hadamard gates and n CNOTs suffices.

[35] A. Harrow and S. Mehraban, Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates, ArXiv:1809.06957.

[36] A. W. Harrow and R. A. Low, Efficient quantum tensor product expanders and k-designs, Lect. Notes Comput. Sci. **5687**, 548 (2009).

[37] W. Fulton and J. Harris, *Representation Theory: A First Course*, Graduate Texts in Mathematics (Springer, New York, 1991).

[38] M. Christandl, PhD thesis, University of Cambridge, 2006.

[39] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank, J. Math. Phys. **19**, 999 (1978).

[40] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).

[41] J. C. Bridgeman and C. T. Chubb, Hand-waving and interpretive dance: An introductory course on tensor networks, J. Phys. **A50**, 223001 (2017).

[42] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, Guaranteed recovery of quantum processes from few measurements, Quantum **3**, 171 (2019).

[43] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, Vol. 28 (Princeton University Press, Princeton, NJ, 1970).

[44] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics, Vol. 54 (American Mathematical Society, Providence, RI, 2002).

[45] J. Cotler, N. Hunter-Jones, J. Liu, and B. Yoshida, Chaos, complexity, and random matrices, JHEP **11**, 048 (2017).

[46] D. Gross, S. T. Flammia, and J. Eisert, Most Quantum States are too Entangled to be Useful as Computational Resources, Phys. Rev. Lett. **102**, 190501 (2009).

[47] A. Bouland, B. Fefferman, and U. Vazirani, Computational pseudorandomness, the wormhole growth paradox, and constraints on the AdS/CFT duality, ArXiv:1910.14646.

[48] Z. Ji, Y.-K. Liu, and F. Song, in *Advances in Cryptology—CRYPTO 2018* (Springer, 2018), p. 126.

[49] In addition to containing inverses, Ref. [12] also required that the gate set G be comprised of algebraic entries, but recent results suggest that both these restrictions may be relaxed [83,84].

[50] Recently, Ref. [35] showed that higher-dimensional local random quantum circuits form approximate designs in $O[n^{1/D}\mathrm{poly}(k)]$ depth, with some (high-degree) polynomial dependence on k. Theorem 9 then gives a polynomial growth of complexity for these higher-dimensional circuits.

[51] We note that Ref. [28] computed the circuit depth, whereas the discussion here involves the circuit size, giving an extra factor of n.

[52] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, Phys. Rev. **X8**, 021014 (2018).

[53] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, Phys. Rev. **B99**, 174205 (2019).

[54] J. Bourgain and A. Gamburd, A spectral gap theorem in SU(d), J. Eur. Math. Soc. **14**, 1455 (2012).

[55] J. Cotler and N. Hunter-Jones, Spectral decoupling in many-body quantum chaos, JHEP **12**, 205 (2020).

[56] L. Susskind, Entanglement is not enough, Fortsch. Phys. **64**, 49 (2016).

[57] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, Holographic Complexity Equals Bulk Action? Phys. Rev. Lett. **116**, 191301 (2016).

[58] S. Chapman, H. Marrochio, and R. C. Myers, Complexity of formation in holography, JHEP **01**, 062 (2017).

[59] D. Carmi, R. C. Myers, and P. Rath, Comments on holographic complexity, JHEP **03**, 118 (2017).

[60] M. Alishahiha, Holographic complexity, Phys. Rev. **D92**, 126009 (2015).

[61] D. Carmi, S. Chapman, H. Marrochio, R. C. Myers, and S. Sugishita, On the time dependence of holographic complexity, JHEP **11**, 188 (2017).
[62] P. Caputa, N. Kundu, M. Miyaji, T. Takayanagi, and K. Watanabe, Liouville action as path-integral complexity: From continuous tensor networks to AdS/CFT, JHEP **11**, 097 (2017).
[63] C. A. Agón, M. Headrick, and B. Swingle, Subsystem complexity and holography, JHEP **02**, 145 (2019).
[64] K. Goto, H. Marrochio, R. C. Myers, L. Queimada, and B. Yoshida, Holographic complexity equals which action? JHEP **02**, 160 (2019).
[65] Z.-W. Liu, S. Lloyd, E. Y. Zhu, and H. Zhu, Entanglement, quantum randomness, and complexity beyond scrambling, JHEP **07**, 041 (2018).
[66] To see this, recall the relation between Schatten α-norms in d dimensions: $\|\rho\|_\infty \leq \|\rho\|_\alpha \leq d^{1/\alpha}\|\rho\|_\infty$. This ensures that for any state ρ, we have $S_{\min}(\rho) \leq S^{(\alpha)}(\rho) \leq S_{\min}(\rho) + (\log d/\alpha)$. Note that as we take α to be greater than n, these Rényi entropies concentrate ever sharper around the min-entropy.
[67] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley Series in Discrete Mathematics and Optimization (John Wiley & Sons, Hoboken, NJ, 2016), 4th ed.
[68] R. Kueng, Quantum and classical information processes with tensors (lecture notes), Spring, 2019. Caltech course notes: https://iqim.caltech.edu/classes.
[69] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
[70] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).
[71] J. Watrous, Semidefinite programs for completely bounded norms, Theory Comput. **5**, 217 (2009).
[72] A. Ben-Aroya and A. Ta-Shma, On the complexity of approximating the diamond norm, Quantum Info. Comput. **10**, 77 (2010).
[73] J. Watrous, Simpler semidefinite programs for completely bounded norms, Chic. J. Theoret. Comput. Sci. **8**, 1 (2013).
[74] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, Improving compressed sensing with the diamond norm, IEEE Trans. Inf. Theory **62**, 7445 (2016).
[75] U. Michel, M. Kliesch, R. Kueng, and D. Gross, Comments on improving compressed sensing with the diamond norm—saturation of the norm inequalities between diamond and nuclear norm, IEEE Trans. Inf. Theory **64**, 7443 (2018).
[76] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge Studies in Advanced Mathematics, Vol. 78 (Cambridge University Press, Cambridge, 2002).
[77] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Vol. 169 (Springer-Verlag, New York, 1997).
[78] Technically, this is only true for bending lines an even number of times. A single bend corresponds to transposition, which is basis dependent and not equivalent to conjugation. This subtlety, however, will rarely feature in our arguments.
[79] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability, Int. Math. Res. Not. **2003**, 953 (2003).
[80] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, J. Math. Phys. **45**, 2171 (2004).
[81] B. Collins and S. Matsumoto, Weingarten calculus via orthogonality relations: New applications, Lat. Am. J. Probab. Math. Stat. **14**, 631 (2017).
[82] A. Montanaro, Weak multiplicativity for random quantum channels, Commun. Math. Phys. **319**, 535 (2013).
[83] R. Mezher, J. Ghalbouni, J. Dgheim, and D. Markham, Unitary *t*-designs from *relaxed* seeds, ArXiv:1911.03704.
[84] M. Oszmaniec, A. Sawicki, and M. Horodecki, Epsilon-nets, unitary designs and random quantum circuits, ArXiv:2007.10885.