

# Randomness and complexity in random complex quantum systems

*im Fachbereich Physik der Freien Universität Berlin eingereichte Dissertation  
zur Erlangung des Grades eines Doktors der Naturwissenschaften*

JONAS HAFERKAMP

*eingereicht im September 2022*

Dahlem Center for Complex Quantum Systems  
Freie Universität Berlin



Reviewers

Prof. Dr. Jens Eisert

Prof. Dr. Piet Brouwer

Date of Disputation: 20<sup>th</sup> March 2023

Name: Haferkamp

Vorname: Jonas

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht. Mit einer Prüfung meiner Arbeit durch ein Plagiatsprüfungsprogramm erkläre ich mich einverstanden.

Datum:

Unterschrift:

## ABSTRACT

---

The interaction of computational complexity and quantum physics touches a wide range of topics from emerging technologies such as quantum computers to the physics of black holes. While tools from quantum information theory can help to answer questions in theoretical computer science, conversely, the ideas developed for analyzing the power of classical computers can shed light on physical phenomena.

Deeply intertwined with both, quantum theory and the theory of complexity, is randomness. Indeed, quantum theory is a probabilistic theory and can predict, in general, only expectation values for observables. In the theory of algorithms, randomness is not only a key design primitive but also indispensable as a proof technique.

In this thesis we make advances at the intersection of randomness, complexity and quantum theory. This includes a mathematical analysis of random ensembles of tensor network states, leading to new results on the average-case complexity of tensor network contraction, contributions to the foundation of verifiable quantum supremacy experiments as well as novel bounds on the generation of quantum pseudorandomness. In particular, we show that unitary  $t$ -designs can be generated with a system-size independent number of non-Clifford resources and that random quantum circuits generate designs in depth  $O(nt^{5+o(1)})$ . These results have numerous applications including the best bounds on the growth of operational notions of quantum circuit complexity. Moreover, we provide a proof of the Brown-Susskind conjecture for the linear growth of exact circuit complexity in random quantum circuits.

The majority of the results in this thesis are theorems published in academic journals. The tools exploited for this analysis range from the concepts of theoretical computer science, such as complexity classes, over ideas from harmonic analysis and Markov chains to the techniques of quantum many-body physics.

In an appendix, this thesis contains several unpublished results such as the first non-trivial upper bounds on moments of the permanent and generation of quantum pseudorandomness with random measurements on the cluster state.



## ZUSAMMENFASSUNG

---

Die Schnittstelle von Komplexitätstheorie und Quantenphysik umfasst Quantentechnologien bis hin zu fundamentalen Fragen über die Physik schwarzer Löcher. Während die Methoden der Quanteninformationstheorie dabei helfen können, Fragen in der Informatik zu beantworten, tragen algorithmische Ideen dazu bei physikalische Phänomene zu erklären.

Sowohl in der Komplexitätstheorie als auch in der Quantentheorie ist das Konzept des Zufalls allgemeingegenwärtig. Für die Entwicklung neuer Algorithmen ist Zufall ein mächtiges Werkzeug und erlaubt oft effiziente Methoden, wo es keine schnellen deterministischen Lösungen gibt. Die Quantentheorie ist inhärent probabilistisch und erlaubt nur Vorhersagen über Erwartungswerte.

In dieser Dissertation machen wir mehrere Fortschritte an der mathematischen Theorie dieser Schnittstelle. Das beinhaltet die Untersuchung von zufälligen Ensembles sogenannter Tensorproduktzustände, Komplexitätsresultate für den typischen Fall von Tensornetzwerken, rigorose Evidenz für verifizierbare Quantenüberlegenheitsexperimente und mehrere neue Schranken auf die Erzeugung von Quantenpseudozufall. Insbesondere zeigen wir, dass unitäre  $t$ -Designs mit einer systemunabhängigen Anzahl an nicht-Clifford Gattern erzeugt werden können und das zufällige Quantenschaltkreise  $t$ -Designs in einer Tiefe von  $O(nt^{5+o(1)})$  erzeugen. Diese Resultate haben zahlreiche Anwendungen und implizieren insbesondere die momentan besten Schranken auf das Wachstum fehlerrobuster Definitionen von Schaltkreiskomplexität. Letztlich enthält diese Dissertation einen Beweis der Brown-Susskind Vermutung für das lineare Wachstum der exakten Schaltkreiskomplexität in zufälligen Quantenschaltkreisen.

Die Mehrzahl der Ergebnisse in dieser Arbeit sind mathematische Theoreme mit rigorosen Beweisen. Die Methoden für diese Analyse rangieren von den Konzepten der theoretischen Informatik, über harmonische Analysis und Markovketten zu den Techniken der Vielteilchentheorie.

In einem Appendix enthält diese Arbeit mehrere unveröffentlichte Resultate wie die ersten nicht-trivialen oberen Schranken auf Momente der Permanente zufälliger Matrizen so wie die Erzeugung von Pseudozufall mit zufälligen Messungen auf Cluster-Zuständen.





# CONTENTS

---

1	INTRODUCTION	13
1.1	Extended outline.	15
2	“ISN’T THIS JUST SCHUR’S LEMMA?” REPRESENTATIONS AND PROBABILITY THEORY ON GROUPS	19
2.1	Representations	19
2.2	Representation theory of the unitary group	21
2.3	The Haar measure and Weingarten calculus	23
2.4	The many faces of unitary t-designs	30
2.5	Random walks, spectral gaps and approximate designs	34
3	QUANTUM PSEUDORANDOMNESS AND GROWTH OF QUANTUM CIRCUIT COMPLEXITY	39
3.1	Unitary designs with a system-size independent number of non-Clifford unitaries	39
3.2	Improved design depths for random quantum circuits	89
3.3	Linear growth of quantum circuit complexity	133
4	QUANTUM ADVANTAGE EXPERIMENTS	155
4.1	Closing gaps of a quantum advantage with short-time Hamiltonian dynamics	155
5	COMPUTATIONAL COMPLEXITY AND RANDOM TENSOR NETWORKS	181
5.1	Contracting projected entangled pair states is average-case hard	181
5.2	Emergent statistical mechanics and disordered random matrix product states	192
6	SUMMARY AND OPEN PROBLEMS	207
6.1	The robust Brown-Susskind conjecture, designs in linear depth and the spectral gap of random quantum circuits	207
6.2	Approximate average-case hardness of output probabilities	208
6.3	Anticoncentration of boson sampling	208
	Acknowledgments	209
A	ADDITIONAL RESULTS	225
A.1	Efficient quantum pseudorandomness from quantum simulators	225
A.2	Incompressibility of parameterized quantum circuits	227

A.3	Baby steps towards anticoncentration of boson sampling	228
A.4	Is there an analogue of Gurvit's algorithm for tensor networks?	231

## PUBLICATIONS OF THE AUTHOR

---

The first-author publications featuring in this thesis are:

- [HHEG20] *Contracting projected entangled pair states is average-case hard*, J. Haferkamp, D. Hangleiter, J. Eisert, M. Gluza, *Physical Review Research*, **2**, 013010, (2020). [10.1103/PhysRevResearch.2.013010](https://doi.org/10.1103/PhysRevResearch.2.013010)

The author is the main contributor of this project. In particular, he is responsible for the proof of the main theorem and contributed to the presentation.

- [HHB<sup>+</sup>20] *Closing gaps of a quantum advantage with short-time Hamiltonian dynamics*, J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, J. Bermejo-Vega, *Physical Review Letters*, **125**, 250501, (2020). [10.1103/PhysRevLett.125.250501](https://doi.org/10.1103/PhysRevLett.125.250501)

The author is the main contributor of this project. He is in large parts responsible for the proofs of both main theorems and contributed to the presentation.

- [HBRE21] *Emergent statistical mechanics from properties of disordered random matrix product states*, J. Haferkamp, C. Bertoni, I. Roth, J. Eisert, *Physical Review X Quantum*, **2**, 040308, (2021). [10.1103/PRXQuantum.2.040308](https://doi.org/10.1103/PRXQuantum.2.040308)

The author is the main contributor of this project. In particular, he proposed the problem and is in large parts responsible for the manuscript, the graphics and developed most of the mathematical proofs.

- [HHJ21] *Improved spectral gaps for random quantum circuits: large local dimensions and all-to-all interactions*, J. Haferkamp, N. Hunter-Jones, *Physical Review A*, **104**, (2) 022417, (2021). [10.1103/PhysRevA.104.022417](https://doi.org/10.1103/PhysRevA.104.022417)

The author was the main contributor of this project. In particular, he developed most of the mathematical proofs and contributed to the presentation.

- [HFK<sup>+</sup>22] *Linear growth of quantum circuit complexity*, J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, N. Yunger Halpern, *Nature Physics*, **18**, 528-532 (2021) [10.1038/s41567-022-01539-6](https://doi.org/10.1038/s41567-022-01539-6). Presented as a contributed talk at QIP 2022.

The author is the main contributor of this project. In particular, he developed the mathematical proofs and contributed to the presentation.

- [HMMH<sup>+</sup>23] *Efficient unitary designs with a system-size independent number of non-Clifford gates*, J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, I. Roth, *Communications in Mathematical Physics* **397**, 995-1041 (2023) [10.1007/s00220-022-04507-6](https://doi.org/10.1007/s00220-022-04507-6). Presented as a contributed talk at QIP 2021 and TQC 2021.

The author is the main contributor of this project. In particular, he proposed the problem and the proof strategy of the main theorem is a generalization of the case  $t = 4$  which was worked out by the author. The author was involved in all the steps necessary for this generalization.

- [Haf22] *Random quantum circuits are unitary  $t$ -designs in depth  $O(\pi t^{5+o(1)})$* , J. Haferkamp, *Quantum* **6**, 795 (2022). [10.22331/q-2022-09-08-795](https://arxiv.org/abs/10.22331/q-2022-09-08-795)

The author is the sole contributor of this project.

The following publications are not included in this thesis.

- [GHS23] *Davydov-Yetter cohomology, comonads and Ocneanu rigidity*, A. Gainutdinov, J. Haferkamp, C. Schweigert, *Advances in Mathematics* **414**, 108853 (2023).
- [BGH<sup>+</sup>20] *Dynamical structure factors of dynamical quantum simulators*, M. L. Baez, M. Goihl, J. Haferkamp, J. Bermejo-Vega, M. Gluza, J. Eisert, *Proceedings of the National Academy of Sciences*, **117**,(42) , 26123-26134, (2021).
- [HIN<sup>+</sup>21] *Learnability of the output distributions of local quantum circuits*, M. Hinsche, M. Ioannou, A. Nietner, J. Haferkamp, Y. Quek, D. Hangleiter, J.-P. Seifert, J. Eisert, R. Sweke, *arXiv:2110.05517*.
- [YKH<sup>+</sup>22] *Resource theory of quantum uncomplexity*, N. Younger Halpern, N. B. T. Kothakonda, J. Haferkamp, A. Munson, J. Eisert, P. Faist, *Physical Review A* **106**, 062417 (2022).
- [HBV21] *Equivalence of contextuality and Wigner function negativity in continuous-variable quantum optics*, J. Haferkamp, J. Bermejo-Vega, *arXiv:2112.14788*, submitted to *Physical Review A*.
- [HIN<sup>+</sup>22] *A single T-gate makes distribution learning hard*, M. Hinsche, M. Ioannou, A. Nietner, J. Haferkamp, Y. Quek, D. Hangleiter, J.-P. Seifert, J. Eisert, R. Sweke, *arXiv:2207.03140*, submitted to *Physical Review Letters*. Best Poster Award at QIP 2023.

## INTRODUCTION

---

Determining macroscopic properties of materials is of key importance for the understanding of conductance and thermodynamics of solid-state materials [AS10, NB09], designing new sensors and devising novel quantum technologies [ABB<sup>+</sup>18] and inferring nuclear processes in stars or the early universe [Web05, ABB<sup>+</sup>05]. Often, however, it is not possible to find degrees of freedom enabling a concise description of a given system. In such a case there is usually no other way than to calculate numerically observables of interest from a microscopic description [GKKR96, GML<sup>+</sup>11, MSH<sup>+</sup>12, Scho5, Sch11, AL18, DDo6].

This being said, it is an unfortunate truth in life that many problems are hard to solve. This includes physics and, in particular, the derivation of macroscopic properties from a microscopic description. In general, the number of parameters necessary to fully characterize an arbitrary state of a physical system grows exponentially in the number of microscopic particles involved. This phenomenon is often referred to as the *curse of dimensionality*. Nevertheless, only specific and, thus, a small fraction of systems satisfy nature's constraints such as the locality of interactions. Therefore, the number of parameters alone cannot provide a satisfactory explanation for practical challenges and, in principle, these specific systems might allow for a more compressed description, e.g. by exploiting a low-entanglement structure via a tensor network ansatz. There are even examples of dynamics generating highly correlated quantum systems which can still be simulated efficiently. This includes states generated by so-called matchgate circuits [Valo2] and Clifford circuits [Got98]. Both circuit families admit a structure that can be exploited by a simulation algorithm: Matchgate circuits can be transformed into the dynamics of free fermions [Bra05] and are thus interaction-free from this point of view; Clifford circuits have restrictive algebraic properties. These examples are in contrast to many examples of quantum systems that can presently be build in laboratories such as Rydberg atoms [BSK<sup>+</sup>17], superconducting qubits [CW08] and trapped ions [BR12]. Even the controlled dynamics of ultracold bosonic atoms in optical lattices [TCF<sup>+</sup>12] quickly become intractable by classical simulation algorithms. This leads to the first guiding question of this thesis: *How complex are general interacting quantum systems?*

**Computational complexity and physics.** In order to provide at least partial answers to this general question, we need to show non-existence of such compressed descriptions for certain physical processes or systems. In general, ruling out algorithms or computational shortcuts is a highly non-trivial problem and a defining question for the field of computational complexity theory [AB09]. Arguably, the most prominent example of the notorious problem to rule out efficient solutions is the P vs NP conjecture. P vs NP asks, colloquially speaking, whether any computational problem that admits efficient proofs can also be solved efficiently to begin with [Aar16, AB09]. It is widely conjectured that this is not the case. That is, some so-called NP-hard problems cannot be solved efficiently <sup>1</sup>. Consequently, the hardness of explicit

---

<sup>1</sup> Formally in a time that is polynomial in the input size.

computational problems is often established by reducing the problem to the P vs NP conjecture or a variant of it.

The P vs NP conjecture provides us with beautiful examples of how theoretical computer science can relate to physical phenomena. A direct link between computational complexity theory and physics is the idea that nature itself behaves like a (quantum) computer. This is called the (quantum) *extended Church-Turing thesis* [BV93, VSD86]. While seemingly abstract, this hypothesis can be directly related to the behavior of spin glasses. These are highly disordered systems, typically modeled by Ising-type interactions, that can take very long times to cool down. It turns out that this behavior is a necessary consequence of the Church-Turing thesis: It was proven by Barahona [Bar82] that computing the ground state energy of a general Ising-type Hamiltonian is NP-hard. This means that a quick cooling for all such systems would equip nature with an in-built NP solver: Simply build your instance of the Ising ground state energy problem in a lab, cool it down and measure the energy! Even seemingly mundane phenomena, such as optimal configurations of soap bubbles, are subject to this line of reasoning [Aar, Aar05].

**Quantum vs. classical complexity.** While the example of spin glasses showcases how computational complexity can make statements about physics, its hardness is of a purely classical nature: Once found, the ground state can be easily stored and processed on a classical computer. What is more, the fact that nature does not perform these tasks quickly indicates that quantum dynamics will not be capable of solving this problem either. So are there tasks that nature performs via quantum computations that cannot be efficiently simulated? Indeed, there is strong evidence in form of Shor’s factoring algorithm [Sho99a] that the computational power of a full quantum computer is separate from classical computation.

Unfortunately, testing algorithms such as Shor’s factoring method on a large quantum computer might not be technologically feasible for a long time. A more recent approach to experimentally separate the computational power of quantum from classical dynamics are so-called “*quantum supremacy*” experiments. These are tasks that can be shown to be unsimulable conditional on certain complexity theoretic assumptions, such as the infinity of the polynomial hierarchy (a generalization of the P vs NP conjecture). In particular, they promise to be feasible long before fully fledged fault tolerant quantum computers can be expected. Indeed, Google prominently claimed quantum supremacy from random quantum circuit sampling in 2019 [AAB<sup>+</sup>19]. As anticipated, new classical algorithms emerged in the meantime and raised the bar for quantum supremacy [PCZ21].

A key aspect, common in all such proposals, is the idea of randomized experiments: The particular instance of the computational task that is performed by the experiment is drawn at random. This procedure guarantees that the dynamic is as devoid of exploitable structure as possible. The unsimulability of these sampling schemes would therefore demonstrate the *computational complexity of randomness in random quantum systems*.

**Complexity and local ensembles of states.** These sampling tasks are good candidates to separate quantum from classical computational power, but they provide little insight into how inherently complex the quantum states generated in these protocols are. For this we need quantifiers such as the (quantum) circuit complexity. This notion allows to measure the minimal number of operations required to prepare a given state. Finding lower bounds for explicit families of states is a notoriously difficult question. It is all the more surprising that often such bounds can be proven to hold *with high probability* for an ensemble of states. Finding such probabilistic

bounds is a key objective towards progress on the general complexity of interacting quantum systems.

Indeed, the *probabilistic method* [AS16] can often imply the existence (and even the abundance) of objects that lack an explicit construction. In this thesis, we will use the probabilistic method to establish lower bounds on the (quantum circuit) complexity of states under chaotic time evolution. Indeed, the probabilistic method has so far found several applications in quantum information theory. A most useful ensemble of states and unitaries is the uniform measure invariant under application of a unitary transformation [DF17]. For example, a seminal result by Page [Pag93] in quantum information theory is that the overwhelming majority of bipartite states are nearly maximally entangled. However, statements of the kind “most states have property X” clearly depend on the ensemble that specifies what is meant by “most”. In particular, in the context of complexity and quantum many-body physics, we are often interested in ensembles of quantum systems that respect the locality of natural interactions and need to understand how their properties compare to those of uniformly random states. This makes complexity a strong motivation to understand *randomness in random (locally) interacting quantum systems*. We will find that the applications of ensembles of states with locality restrictions go beyond complexity.

In this thesis we will make multiple advances at the intersection of randomness and quantum complexity theory. This line of research allows to make progress on fundamental questions about the computability of nature and at the same time delineate the limitations of emergent quantum technologies. In the following, we will relate the general questions and concepts outlined above to specific projects and assign them a place in this thesis.

## 1.1 EXTENDED OUTLINE.

In **Chapter 2**, we review ubiquitous aspects of group theory and representation theory. Multiple results in quantum information theory are based on the integration over groups. We develop most of these techniques from a simple statement in representation theory called Schur’s lemma. In discussions, when confronted with a seemingly complicated integration formula, one would often hear a puzzled: “But isn’t that just somehow Schur’s lemma?”. The answer to this is yes and in this review chapter we show how this works in general. Thereby, we formalize many of the notions we briefly introduce in the following, such as random walks, approximate designs, pseudorandomness and spectral gaps.

In **Chapter 3** we then present the first results of this thesis. To apply randomized arguments to the complexity of interacting quantum systems, we need to understand how toy models of locally random dynamics, so-called *random quantum circuits*, relate to uniform randomness over all states. Random quantum circuits are a well-established concept in the theory of quantum computing and quantum information theory. The abundance of applications, ranging from protocols for the processing of quantum information, quantum system characterization to even black hole physics, is partially due to the quick generation of *quantum pseudorandomness*. A key notion here is that of *unitary t-designs*. These are probability distributions on the unitary group which are indistinguishable from the uniform probability measure -- the *Haar measure* -- if only tested against polynomials of degree at most  $t$ . The Haar measure is the unique probability distribution on the unitary group that is invariant under multiplication with any unitary. In a seminal paper, Brandão, Harrow and Horodecki prove that quantum circuits on  $n$

qubits with randomly drawn gates approximate unitary  $t$ -designs after  $O(nt^{10.5})$  random gates for all  $t \geq O(2^{0.4n})$  [BaHH16].

Local random quantum circuits require a fully fledged fault tolerant quantum computer. While this is likely necessary for higher moments, small values of  $t$  require much less experimental resources. Indeed, the Clifford group is not only a 3-design but also efficiently simulable on a classical computer [Got98, AGo4]. Moreover, Clifford unitaries have desirable properties for fault tolerant implementation [BK05, Got97, CS96, Ste96, err]. At the same time, it fails *gracefully* [ZKGG16] to be a unitary 4-design. To generate higher designs, some non-Clifford resources are required. The objective of Section 3.1 is to quantify the non-Clifford resources that are sufficient to lift the Clifford group to an approximate  $t$ -design for small values of  $t$ :  $n \geq t^2$ . Surprisingly, we find that it suffices to inject  $O(t^4)$  single qubit non-Clifford gates into a random Clifford circuit to approximate unitary designs up to (additive) errors. Strikingly, this number is independent of the system-size and, therefore, the density of non-Cliffords is allowed to tend to zero in the thermodynamic limit.

The generation of approximate unitary  $t$ -designs has direct consequences for the *growth of quantum circuit complexity* in random quantum circuits. Effecting a unitary or a state (possibly up to some error) requires a minimal number of elementary 2-local unitaries. This minimal number is called the *quantum circuit complexity*. It was proven in Ref. [BCHJ<sup>+</sup>21] that a unitary drawn from a unitary  $t$ -design has circuit complexity at least  $\Omega(t)$  with overwhelming probability. This can be viewed as a partial derandomization of the fact that most unitaries are exponentially complex [Pre98], a quantum analogue of a result by Shannon [Sha49] for boolean functions. Combined with the Brandão, Harrow, Horodecki bound, the lower bound on unitary  $t$ -designs implies a bound of  $\Omega(T^{1/10.5})$  for the circuit complexity of states prepared by a random quantum circuit of depth  $T$ , *even for exponentially deep circuits*.

The exponential time scale, while practically inaccessible, is key to a conjecture in the context of the AdS/CFT correspondence and, in particular, for a proposal to resolve the wormhole growth paradox [Sus18, Sus16]. Here, random quantum circuits serve as a model for the holographic dual of a black holes dynamic. The Brown-Susskind conjecture states that with high probability the quantum circuit complexity of random quantum circuits grows at a linear rate for exponentially deep circuits [BS18a]. Roughly, this would mean that most local quantum circuits are essentially incompressible until their complexity saturates. Therefore, the complexity lower bound can be viewed as a significant step towards the Brown-Susskind conjecture in that it establishes a sublinear but *algebraic* bound on the complexity.

In the first publication of Section 3.2, we improve the design depth of local random quantum circuits with nearest neighbor interactions to  $O(nt^{5+o(1)})$  for  $t \leq O(2^{0.5n})$ . This implies a growth of complexity of  $O(T^{1/5+o(1)})$ , which constitutes the strongest known bound for error-robust notions of quantum circuit complexity. As an auxiliary result, we prove a strong convergence result for the Clifford interleaved circuits introduced in Section 3.1.

The generation of unitary  $t$ -designs at a depth  $O(nt)$  would imply the full Brown-Susskind conjecture [BCHJ<sup>+</sup>21] and was conjectured in Ref. [BaHH16, BHH16]. Evidence for this scaling was obtained in [HJ19], by considering random quantum circuits on large qudits, i.e. acting on  $(\mathbb{C}^q)^{\otimes n}$  in the limit of large local dimensions. One contribution of the second publication presented in Section 3.2 is that this abstract limit can be reduced to the condition  $t^2 \leq q$ . In the same publication, we also prove strong bounds on random quantum circuits without geometric locality restriction or all-to-all interactions. Moreover, we combine new numerical results with



analytical arguments to obtain the strongest known bound on the convergence to designs for  $t = 2, 3, 4, 5$ .

In Section 3.3 we change perspective and focus on the complexity of *exact* implementation of states. This far more restrictive notion allows us to prove that the complexity in random quantum circuit is at least linear in the depth of the random quantum circuit with probability 1. This rigorously proves a variant of the Brown-Susskind conjecture. The key idea is not to use unitary  $t$ -designs, but to exploit that the set of unitaries implementable in a given circuit architecture is a semialgebraic set [BCR13].

In **Chapter 4** of this thesis is concerned with the efficient simulability of specific computational problems related to quantum physics. In recent years, the primitive of sampling-based *quantum advantage experiments* or “quantum supremacy experiments” emerged. Examples include the boson sampling proposal [AA13], random quantum circuit sampling [AAB<sup>+</sup>19, BFNV19] and commuting quantum circuits [BMS17, BMS16a]. The objective of these schemes is not to solve some practical tasks. Instead, we only require the quantum device to exponentially outperform the best classical attempts to simulate them. As explained above, this alone would be a significant achievement, refuting the Church-Turing thesis. The typical models of quantum computing come with no mathematical guarantees of a speed-up. In fact, a polynomial algorithm for factoring integers would have no immediate implications for problems in classical computer science <sup>2</sup>, such as P vs NP.

While less powerful than access to a full quantum computer, the “supremacy” of sampling-based quantum advantage schemes can be related to classical computational assumptions. These arguments come with “loopholes” or “gaps” and much of the theoretical literature is focused on closing the gaps. In particular, the *average-case complexity* of approximating the output probabilities up to very small errors is an open problem for each of these schemes. Related thereto is a property called *anticoncentration* that roughly asserts that for most instances the output distribution is not too peaked. This anticoncentration property can be proven for some but not all schemes. Another loophole is the *verification* of the protocol: How do we know that the quantum device performed the computation we want it to? A recent quantum advantage scheme [BVHS<sup>+</sup>18] is based on a constant time evolution under an elementary Ising-type Hamiltonian and can be seen as a *quantum advantage experiment for quantum simulators*. A particular advantage of this scheme is that it comes with an efficient verification protocol (up to partial trust in the measurements) [HKSE17].

In **Chapter 4**, we close the anticoncentration gap for this protocol by proving the 2-design property for a random quantum circuit effected by measurements. Moreover, we show that a very accurate evaluation of the output distribution is indeed average-case complex by adapting an argument based on polynomial interpolation developed by Lipton for the permanent [Lip91, BFNV19].

**Chapter 5**, we focus on one of the key techniques for the simulation of quantum systems: tensor networks. The locality of interaction of Hamiltonians is related to a low-entanglement structure of states that can be prepared of ground states of such systems. Arguably, these states constitute the physical corner of Hilbert space. Tensor network states naturally captures low entanglement by modeling the correlation with local tensors with restricted dimensions. This ansatz has proven unreasonably successful [Orú14, BC17].

---

<sup>2</sup> Of course, it would break some hearts [pri].

The subclass of matrix product states are among the most ubiquitous tools in the simulation of condensed matter and in the theory of quantum information. Comfortably, this comes with a matching mathematical guarantee [Vid03] that not only their storage, but also their contraction can be efficiently computed on a classical computer. This allows to efficiently extract physical quantities such as expectation values of local observables from a matrix product approximation of a state. This situation drastically changes in 2D. Not only are practical challenges abundant, but it can also be proven that a very accurate contraction of a 2D tensor network representation (a so-called *projected entangled pair state* (PEPS)) is #P-hard [SWVC07]. The latter includes the class NP and it is therefore unlikely that an efficient algorithm for this task can exist. However, this only makes a statement about the *worst-case complexity*. That is, it does not rule out the existence of good heuristic algorithms that work fast and accurately for most systems. In the first half of **Chapter 5**, we prove a worst-to-average case reduction for the contraction of PEPS. These results imply that no heuristic algorithm can solve a constant fraction of the instances of the contraction problem with very high accuracy.

In the second half of **Chapter 5**, we study general properties of a naturally local ensemble of matrix product states. We employ a mapping of the resulting combinatorial problem to the partition functions of auxiliary statistical mechanics models. This allows us to prove equilibration of evolution under general Hamiltonians starting with “most MPS”, a local principle of maximal entropy and the extensivity of the second Rényi entropy of disjointed subsystems all with overwhelming probability.

Last, in **Appendix A** we present and prove results not contained in the publications that form this thesis. Some of these results are complementary to published results such as the generation of quantum pseudorandomness from random measurements in cluster states. Others approach independent problems, such as providing non-trivial upper bounds on moments of the permanent.

## ‘‘ISN’T THIS JUST SCHUR’S LEMMA?’’ REPRESENTATIONS AND PROBABILITY THEORY ON GROUPS

---

### 2.1 REPRESENTATIONS

In this thesis we assume familiarity with basic algebraic definitions such as topological groups, Lie-groups, rings, fields and (associative) algebras as well as the concepts of quantum information theory. The following chapter is intended as a quantum information theorist’s guide for representation theory and some of its applications. We do not follow any particular textbook, but for more details and depth we refer to Ref. [FH].

Groups are among the most ubiquitous objects in mathematics. In particular, they are closely related to the concept of symmetries. This thesis is concerned with quantum theory and in quantum theory most objects of interest can be located in vector spaces and more precisely in Hilbert spaces. If a group is supposed to describe a class of symmetries of a physical system it can only do so if the abstract group structure can be translated to operations on a Hilbert space. If such a map respects the group structure, it is called a *representation*:

**Definition 1.** *A (unitary) representation of a group  $G$  is a tuple  $(\rho, V)$ , where  $V$  is a Hilbert space and  $\rho : G \rightarrow \mathcal{U}(V)$  is a continuous map from  $G$  to the group of unitary operations on  $V$  such that  $\rho(g \circ h) = \rho(g)\rho(h)$  for all  $g, h \in G$  and  $\rho(e) = \mathbb{1}_V$ , where  $e \in G$  denotes the identity element of  $G$ . A subrepresentation is a vector space  $W \subset V$  such that  $\rho(U)W \subset W$  for all  $U$ .*

A representation of an algebra is defined similarly. We will often abuse notation and refer to the vector space  $V$  as the representation.

It can be shown from this definition that  $\rho$  also respects inverses:  $\rho(g^{-1}) = \rho(g)^\dagger$ .

**Example 1.** *We will meet the following examples multiple times again in this thesis.*

- *For every group  $G$  there is always the 1-dimensional trivial representation  $V_{\text{triv}}$  on which the group acts as  $\rho(g)v = v$  for  $g \in G$  and  $v \in V_{\text{triv}}$ .*
- *The traceless part of the adjoint representation of any matrix group  $G \subset \text{GL}(d) : V = \{M \in \mathbb{C}^{d \times d}, \text{Tr}[M] = 0\}$  and  $\rho(g)(M) := gMg^{-1}$ .*
- *Tensor products: For any two representation  $(V, \rho_V)$  and  $(W, \rho_W)$  of a group  $G$ , the vector space  $V \otimes W$  is also a representation with the representation  $\rho_{V \otimes W}(g)(v \otimes w) = \rho_V(g)v \otimes \rho_W(g)w$ . Most often, we will consider the case, where  $V = W$  and  $\rho_V = \rho_W$  or even higher tensor powers defined by iterating this definition.*

For many naturally occurring groups, representations can be decomposed into a class of *irreducible* representations, the ‘‘atoms’’ of representation theory:

**Definition 2.** *We call a representation  $(\rho, V)$  of a group  $G$  irreducible if there is no subrepresentation of  $\rho$  except  $\{0\}$  and  $V$ .*

In this thesis, we will only ever see groups and algebras that are semisimple:

**Definition 3.** A group is called *semisimple* if all representations over  $\mathbb{C}$  decompose as a direct sum of irreducible representations.

**Example 2.**

- The trivial representation is irreducible for every semisimple group.
- For the special unitary group, the adjoint representation decomposes into two irreps: The trivial irrep and the traceless part. We will show later as an application of the relation between commutants and representations that this traceless part is indeed an irreducible representation.

Intertwiners are linear maps that respect the structure of a representation:

**Definition 4.** Let  $(V, \rho)$  and  $(W, \sigma)$  be two representations of a group  $G$ , a linear map  $T : V \rightarrow W$  is called a *homomorphism* or *intertwiner* if it commutes with the group action, i.e.  $\sigma(g)T|v\rangle = T\rho(g)|v\rangle$ . If  $T$  is invertible, it is called an *isomorphism*.

A corner stone and one of the most useful statements in representation theory and in this thesis is *Schur's lemma*.<sup>1</sup>

**Theorem 1** (Schur's lemma). Let  $(V, \rho)$  and  $(W, \sigma)$  be irreducible representations of a semisimple group  $G$  and  $T : V \rightarrow W$  an intertwiner. If  $(V, \rho)$  and  $(W, \sigma)$  are not isomorphic, then  $T = 0$ . If we have  $(W, \sigma) = (V, \rho)$ , then  $T$  is a multiple of the identity  $\mathbb{1}_V$  on  $V$ .

This immediately implies that isomorphisms between irreducible representations are unique up to scalar factors. Indeed, in this thesis, we will only encounter intertwiner that are unitaries on the representation spaces, which implies uniqueness of isomorphisms up to a phase.

A useful structural tool in representation theory is the *commutant theorem*. Next, we will show that it follows directly from Schur's lemma. For this, consider the *commutant* of a set  $S \subset L(V)$  of linear maps acting on a vector space  $V$ :

$$\text{comm}(S) := \{T \in L(V), [L, T] = 0 \text{ for all } L \in S\}. \quad (2.1)$$

Moreover, for a representation  $(V, \rho)$  we set

$$\text{comm}(\rho) := \text{comm}(\{\rho(g), g \in G\}). \quad (2.2)$$

**Theorem 2** (Commutant theorem). Let  $(V, \rho)$  be a representation of a semisimple group  $G$ . Then,  $A := \text{comm}(\rho)$  is a semisimple algebra acting on  $V$  with irreducible representations  $W_\lambda$ , which can be labeled by all isomorphism classes  $\lambda$  represented by irreps  $(V_\lambda, \rho_\lambda)$  of  $G$  appearing in the decomposition of  $V$ . Then,  $V$  decomposes as follows:

$$V \simeq \bigoplus_{\lambda} V_{\lambda} \otimes W_{\lambda}, \quad (2.3)$$

where  $G$  acts on  $V$  as  $\sum_{\lambda} \rho_{\lambda}(g) \otimes \mathbb{1}_{W_{\lambda}}$  for all  $g \in G$  and  $A$  acts on  $V$  as  $\sum_{\lambda} \mathbb{1}_{V_{\lambda}} \otimes L$  for all  $L \in A$ .

<sup>1</sup> In general, Schur's lemma characterizes intertwiners between irreducible representations as elements of a division algebra. In the special case of representations over the complex numbers, we have the stronger implication presented here.

*Proof.* By definition of the irreps  $V_\lambda$ ,  $V$  decomposes as

$$V \simeq \sum_{\lambda} V_{\lambda} \otimes M_{\lambda} = \sum_{\lambda} V_{\lambda}^{\oplus \dim M_{\lambda}} \quad (2.4)$$

for some multiplicity spaces  $M_{\lambda}$ .<sup>2</sup> In the following we will use the bracket notation from quantum mechanics, i.e. we denote vectors as  $|\psi\rangle$  and dual vectors by  $\langle\psi|$ . Pick any ONB  $\{|k, \lambda\rangle\}$  of the spaces  $M_{\lambda}$ . Denote by  $P_{k, \lambda}$  the orthogonal projector onto the space  $V_{\lambda} \otimes |k, \lambda\rangle$  and let  $T : V \rightarrow V$  be an element of  $A$ . The latter is equivalent to  $T$  being an intertwiner from  $(V, \rho)$  to  $(V, \rho)$ . Consider the restricted operators  $P_{k, \lambda'} T P_{k', \lambda}$ . As the representation  $\rho(g)$  block diagonalizes into the irreducible representations, the restricted operators are intertwiners as well. In particular, we can directly invoke Schur's lemma multiple times and find that

$$\begin{aligned} T &= \sum_{k, \lambda} P_{k, \lambda} T \sum_{k', \lambda'} P_{k', \lambda'} \\ &= \sum_{k, \lambda, k', \lambda'} P_{k, \lambda} T P_{k', \lambda'} \\ &= \sum_{k, k', \lambda} P_{k, \lambda} T P_{k', \lambda} \\ &= \sum_{\lambda} \sum_{k, k', \lambda} (P_{V_{\lambda}} \otimes \langle k, \lambda |) T (P_{V_{\lambda}} \otimes |k', \lambda\rangle) \otimes |k, \lambda\rangle \langle k', \lambda| \\ &= \sum_{\lambda} \sum_{k, k', \lambda} L_{\lambda, k, k'} \mathbb{1}_{V_{\lambda}} \otimes |k, \lambda\rangle \langle k', \lambda| \\ &= \sum_{\lambda} \mathbb{1}_{V_{\lambda}} \otimes L_{\lambda}, \end{aligned} \quad (2.5)$$

for some linear operators  $L_{\lambda} : M_{\lambda} \rightarrow M_{\lambda}$ . Moreover, all operators of this form are in the commutant. Hence, we have

$$A = \bigoplus_{\lambda} \mathbb{1}_{V_{\lambda}} \otimes L(M_{\lambda}) \simeq \bigoplus_{\lambda} L(M_{\lambda}) \quad (2.6)$$

as an algebra. It follows from this characterization that the spaces  $M_{\lambda}$  are irreducible representations of  $A$ .  $\square$

## 2.2 REPRESENTATION THEORY OF THE UNITARY GROUP

In this section we focus on the representation theory of the unitary group. We have seen in the last section how the commutant algebra can be helped to characterize the representation theory of a group. Here, we will apply this general principle to the *diagonal representation* or *tensor power representation*  $\delta_t : \mathrm{SU}(d) \rightarrow \mathrm{U}((\mathbb{C}^d)^{\otimes t})$  of  $\mathrm{SU}(d)$ :

$$\delta_t(\mathbf{U}) := \mathbf{U}^{\otimes t}. \quad (2.7)$$

As it turns out, we also have to study the representation theory of the symmetric group  $S_t$ , the group of all bijections on  $[t]$ . The reason for this is the celebrated *Schur-Weyl duality*. For this statement, consider the representation  $r : S_t \rightarrow \mathrm{U}((\mathbb{C}^d)^{\otimes t})$  of the symmetric group defined by

$$r(\pi)|i_1, \dots, i_t\rangle = |i_{\pi(1)}, \dots, i_{\pi(t)}\rangle. \quad (2.8)$$

<sup>2</sup> The spaces  $V_{\lambda} \otimes M_{\lambda}$  are called isotypic components.

**Theorem 3.** *The following holds for the commutant of the diagonal action of  $SU(d)$ :*

$$\text{comm}(\delta_t) = \text{span}\{\tau(\pi), \pi \in S_t\}. \quad (2.9)$$

The special case  $t = 2$  already implies the statement made in Example 2 that the traceless part of the adjoint representation is irreducible:

**Proposition 1.** *The adjoint action  $\text{Ad} : SU(d) \rightarrow L(\mathbb{C}^{d \times d})$ ,  $\text{Ad}_U(M) := UMU^\dagger$  decomposes into two irreps:  $\{\alpha \mathbb{1}_d, \alpha \in \mathbb{C}\}$  and  $\{M \in \mathbb{C}^{d \times d}, \text{Tr}(M) = 0\}$ .*

*Proof.* First, recall that  $\mathbb{C}^{d \times d}$  is a Hilbert space with the Hilbert-Schmidt product. Therefore, we have  $\{M \in \mathbb{C}^{d \times d}, \text{Tr}(M) = 0\} = \{\alpha \mathbb{1}_d, \alpha \in \mathbb{C}\}^\perp$ . We have immediately that  $\{\alpha \mathbb{1}_d, \alpha \in \mathbb{C}\}$  is isomorphic to the trivial irrep. Therefore,  $\mathbb{C}^{d \times d}$  decomposes into at least two irreps. It can be easily verified that under the vectorization isomorphism, the adjoint representation is equivalent to the action  $\delta_{1,1} : U \mapsto U \otimes \bar{U}$ . It can be checked that the partial transpose  $\Gamma$  is an isomorphism of vector spaces  $\Gamma : \text{comm}(\delta_2) \rightarrow \text{comm}(\delta_{1,1})$ <sup>3</sup>. In particular,

$$\dim \text{comm}(\delta_{1,1}) = \dim \text{comm}(\delta_2) = \dim \text{span}\{\mathbb{1}, \mathbb{F}\} = 2.$$

Therefore, the commutant of the adjoint action can contain at most 2 irreducible representations up to isomorphisms and consequently, by Theorem 2, so does the adjoint representation.  $\square$

Similarly, we find:

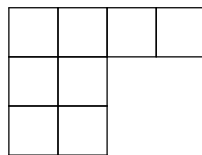
**Corollary 1.** *The symmetric subspace  $S^t(\mathbb{C}^d)$  is an irreducible representation of  $SU(d)$ .*

*Proof.* Observe that  $S^t(\mathbb{C}^d)$  is defined as the invariant subspace under the action  $r$  of  $S_t$ . In other words, it is the isotypical component in  $(\mathbb{C}^d)^{\otimes t}$  of the trivial representation of  $S_t$ . As the trivial irrep is 1-dimensional, the commutant theorem ensures that this isotypical component is an irrep of  $SU(d)$ .  $\square$

Even though less prominent in this thesis, essentially the same proof strategy works for the antisymmetric subspace  $A^t(\mathbb{C}^d)$ . There are more elementary and insightful proofs of the irreducibility of the symmetric subspace [Har13].

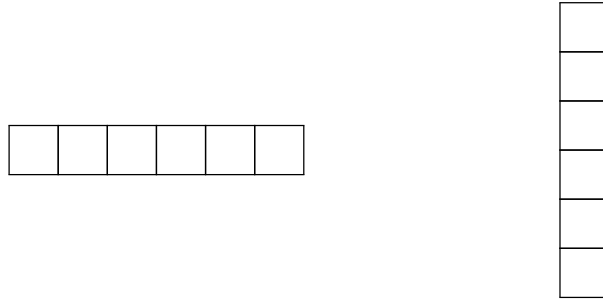
Irreps of the symmetric group can be uniquely labeled (up to isomorphisms) by *partitions*. That is, tuples of numbers  $\lambda = (\lambda_1, \dots, \lambda_m)$  of positive integers ordered from left to right by largest to smallest such that  $\lambda_1 + \dots + \lambda_m = t$ .

It is common to represent partitions by combinatorial objects called Young diagrams. Here, one places  $\lambda_1$  boxes horizontally in a first row,  $\lambda_2$  boxes in a second row etc. For example the partition  $8 = 4 + 2 + 2$  is represented by the Young tableaux



<sup>3</sup> However, the partial transpose is not an algebra isomorphism. In general, the two commutants are not isomorphic as algebras.

From every such Young diagram, one can construct an operator on the group algebra by labeling the boxes by numbers from 1 to  $t$  and then symmetrizing over the columns and antisymmetrizing over the rows. The result, the Young symmetrizer, will have (multiple copies of) a unique irrep as its image. We will not get into the details of this construction, but only take away from this the intuition that the longer the Young diagram horizontally, the more antisymmetric and the taller the diagram vertically, the more symmetric the corresponding irrep is. The extreme cases are the partitions  $(t)$  and  $(1, \dots, 1)$  corresponding to the trivial and the sign representation, respectively. As Young diagrams, these are represented by



Via the commutant theorem, these correspond to the symmetric subspace and the antisymmetric subspace as representations of  $SU(d)$ . This gives us the opportunity to state the following formulation of the Schur-Weyl duality:

**Corollary 2.** *Let  $\Lambda_{t,d}$  be the set of all Young diagrams with at most  $d$  boxes per row and  $t$  boxes overall. Then, we have the following decomposition of the  $t$ -th tensor power representation of  $SU(d)$  acting on  $(\mathbb{C}^d)^{\otimes t}$ :*

$$(\mathbb{C}^d)^{\otimes t} \simeq \bigoplus_{\lambda \in \Lambda_{t,d}} W_{d,\lambda} \otimes S_\lambda. \quad (2.10)$$

The spaces  $W_{d,\lambda}$  are often called Weyl modules and the  $S_\lambda$  are called Specht modules [FH].

It turns out that each irrep of  $SU(d)$  is contained in some tensor power of the standard representation. In particular, we can label the irreps of  $SU(d)$  also by Young diagrams. However, these labelings are only unique if we consider every column with  $d$  boxes to be trivial. It is therefore a standard convention to assign the empty Young diagram to the trivial representation of  $SU(d)$ . Considering the Schur-Weyl duality, the Young diagram labeling the standard representation corresponds to the trivial irrep of  $S_1$ , which is isomorphic to the trivial group. This is the Young diagram which consists of a single box. This immediately yields that the standard representation is irreducible

Highest weights [FH] are an equivalent way of labeling the irreducible representations by the "highest" irrep in the induced representation of the diagonal subgroup. Also compare Appendix B in the publication presented in Section 3.1. This latter approach of highest weights works more generally for compact Lie-groups, where one considers induced representations of a maximal torus [FH].

### 2.3 THE HAAR MEASURE AND WEINGARTEN CALCULUS

A central object in this thesis is the *Haar measure*. That is a generalization of the uniform probability distribution on an interval  $I = [0, b]$  to more complicated objects such as the set

of states or the unitary group. It turns out that the uniform measure on  $I$  can alternatively be defined as the unique measure that is *invariant* under addition modulo  $b$ . In particular, we have the defining condition

$$\frac{1}{b} \int_I f(x + y \pmod{b}) dx = \frac{1}{b} \int_I f(x) dx \quad (2.11)$$

for any measurable function  $f : I \rightarrow \mathbb{C}$  and any  $y \in I$ . Further, it can be easily seen that the operation  $x + y \pmod{b}$  is actually the group action inherited from  $U(1) \cong S^1 \cong I / \sim$ , where the equivalence relation  $\sim$  identifies the two end points. Translating the uniform measure via these isomorphisms from  $I / \sim$  to a measure on  $U(1)$  denoted by  $\mu_H$ , we can rewrite this defining condition on  $U(1)$  as

$$\int_{U(1)} f(UV) d\mu_H(U) = \int_{U(1)} f(U) d\mu_H(U). \quad (2.12)$$

This means,  $\mu_H$ , the uniform, or Haar measure on  $U(1)$  is the unique left invariant measure.

The following theorem provides the existence of a unique measure that generalizes the uniform measure on an interval.

**Theorem 4.** *Let  $G$  be a compact group. There is a unique left invariant probability measure on  $G$  that we denote by  $\mu_H$  and call the Haar measure.*

As the (normalized) Haar measure is a probability measure, we will often denote the integrals over it as expectation values:

$$\mathbb{E}_{U \sim \mu_H} f(U) := \int_G f(U) d\mu_H(U). \quad (2.13)$$

Analogously, a right invariant measure exists. For all groups of interest in this thesis, the left and right invariant measure will coincide, a property called unimodularity. The introduction (and explicit constructions in terms of random  $2 \times 2$  matrices) of the Haar measure on the classical Lie groups  $SU(d)$  and  $SO(d)$  go back to Hurwitz [Hur63, DF17].

Moreover, we can similarly define a unique invariant measure on manifolds that  $G$  acts on transitively. The main example of this is the state space  $\mathcal{S}_d = \{|\psi\rangle, \langle\psi|\psi\rangle = 1\}$ <sup>4</sup>. We denote by  $\mu_H$  the measure on  $\mathcal{S}_d$  defined by

$$\int f(U|\psi\rangle) d\mu_H(U) =: \int f(|\psi\rangle) d\mu_H(|\psi\rangle) \quad (2.14)$$

for any measurable function  $f$  and unitary  $U$ .

The following lemma provides the key link between the Haar measure and representation theory:

**Lemma 1** (Projector formula). *Let  $G$  be a compact group with Haar measure  $\mu_H$  and  $(V, \rho)$  a representation of  $G$ . Moreover, denote by  $P_{\text{triv}}$  the orthogonal projector onto the trivial isotypical ensemble of  $\rho$ . Then,*

$$P_{\text{triv}} = \int \rho(U) d\mu_H(U). \quad (2.15)$$

<sup>4</sup> Technically, the phase is unphysical information and thus we could and perhaps should reduce this to the projective space  $\mathbb{C}P$ .



Moreover, it holds that

$$\left( \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) \right) \left( \int \rho(\mathbf{U}) d\nu(\mathbf{U}) \right) = \left( \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) \right) \quad (2.16)$$

for all probability measures  $\nu$  on  $S\mathbf{U}(d)$ .

*Proof.* First, we can verify that the right hand side of Eq. (2.15) is a projector:

$$\begin{aligned} \left( \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) \right)^2 &= \int \int \rho(\mathbf{U}) \rho(\mathbf{V}) d\mu_{\mathbf{H}}(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{V}) \\ &= \int \int \rho(\mathbf{UV}) d\mu_{\mathbf{H}}(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{V}) \\ &= \int \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{V}) \\ &= \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}). \end{aligned} \quad (2.17)$$

Moreover, as  $\mu_{\mathbf{H}}$  is symmetric, we find that the RHS is also hermitian. Next, we need to characterize the image of the RHS. First, if  $|\psi\rangle$  is in the trivial isotypical component (or invariant subspace), then

$$\int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) |\psi\rangle = \int \rho(\mathbf{U}) |\psi\rangle d\mu_{\mathbf{H}}(\mathbf{U}) = \int |\psi\rangle d\mu_{\mathbf{H}}(\mathbf{U}) = |\psi\rangle. \quad (2.18)$$

For the other direction, let  $|\psi\rangle$  be in the image of  $\int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U})$ . Then,

$$\begin{aligned} \rho(\mathbf{V}) |\psi\rangle &= \rho(\mathbf{V}) \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) |\psi\rangle \\ &= \int \rho(\mathbf{VU}) d\mu_{\mathbf{H}}(\mathbf{U}) |\psi\rangle \\ &= \int \rho(\mathbf{U}) d\mu_{\mathbf{H}}(\mathbf{U}) |\psi\rangle \\ &= |\psi\rangle, \end{aligned} \quad (2.19)$$

which completes the proof. The second claim follows from a straightforward generalization of 2.17.  $\square$

An immediate consequence of this general principle is average over an adjoint of a representation:

**Corollary 3.**

$$\int_{\mathbf{H}} \rho(\mathbf{U}) \mathbf{A} \rho(\mathbf{U})^\dagger d\mu_{\mathbf{H}}(\mathbf{U}) = \mathbf{P}_{\text{comm}(\rho)}, \quad (2.20)$$

where  $\mathbf{P}_{\text{comm}(\rho)}$  denotes the orthogonal projector onto  $\text{comm}(\rho)$  with respect to the Hilber-Schmidt product on  $L(\mathbf{V})$ .

*Proof.* The proof is immediate from the fact that  $\text{Ad}^\rho$  defined by  $\text{Ad}_{\mathbf{U}}^\rho(\mathbf{A}) := \rho(\mathbf{U}) \mathbf{A} \rho(\mathbf{U})^\dagger$  defines a representation on  $L(\mathbf{V})$  and  $\text{Ad}_{\mathbf{U}}^\rho(\mathbf{A}) = \mathbf{A}$  is

$$\rho(\mathbf{U}) \mathbf{A} \rho(\mathbf{U})^\dagger = \mathbf{A} \iff \rho(\mathbf{U}) \mathbf{A} = \mathbf{A} \rho(\mathbf{U}).$$

$\square$

We now give two simple applications of these formulas. The first one is prototypical for many calculations in this thesis.

For these examples we to introduce the Schatten  $p$ -norms. Throughout this thesis, we will denote the Schatten  $p$ -norms of a matrix  $A \in \mathbb{C}^{D \times D}$  by

$$\|A\|_p := \text{Tr}[|A|^p]^{\frac{1}{p}}, \quad (2.21)$$

where  $|A| := \sqrt{AA^\dagger}$ . The case  $p = 1$  is also called the trace norm and is not to be confused with the  $l^1$ -norm of matrix as a vector in  $\mathbb{C}^{D^2}$ . The case  $p = \infty$  corresponds to the operator norm of  $A$  viewed as an element of  $L(\mathbb{C}^D)$ , the space of linear maps on  $\mathbb{C}^D$  and  $p = 2$  coincides with the Frobenius norm. For the latter case, we will mostly use the notation  $\|A\|_F$ .

**Application** On average, over the uniform measure on a  $d$ -dimensional state space, the probability of detecting an arbitrary fixed state  $|\phi\rangle$  is  $1/d$ . Indeed, we can compute

$$\begin{aligned} \mathbb{E}_{|\psi\rangle \sim \mu_H} |\langle \phi | \psi \rangle|^2 &= \mathbb{E} \text{Tr}[|\phi\rangle\langle\phi| |\psi\rangle\langle\psi|] \\ &= \text{Tr}[|\phi\rangle\langle\phi| \mathbb{E} |\psi\rangle\langle\psi|] \\ &= \text{Tr} \left[ |\phi\rangle\langle\phi| \mathbb{E} \int \mathbf{U} |\psi\rangle\langle\psi| \mathbf{U}^\dagger d\mu_H(\mathbf{U}) \right] \\ &\stackrel{\dagger}{=} \text{Tr} \left[ |\phi\rangle\langle\phi| \mathbb{E} \frac{1}{d} \mathbb{1}_d \text{Tr}(|\psi\rangle\langle\psi| \mathbb{1}_d) \right] \\ &= \frac{1}{d} \text{Tr}[|\phi\rangle\langle\phi|] \\ &= \frac{1}{d}. \end{aligned} \quad (2.22)$$

Here,  $\dagger$  follows from Corollary 3 by observing that the standard representation  $\rho(\mathbf{U}) = \mathbf{U}$  is irreducible and therefore the commutant of  $\mathbf{U}$  is the identity by Schur's lemma. Hence,

$$\int \mathbf{U} A \mathbf{U}^\dagger d\mu_H(\mathbf{U}) = P_{\text{span}(\mathbb{1}_d)}(A) = \frac{1}{\|\mathbb{1}_d\|_F^2} \mathbb{1}_d \text{Tr}(A \mathbb{1}_d) = \frac{1}{d} \mathbb{1}_d \text{Tr}(A).$$

The second application is a formula for averages over copies of a uniformly random state. We used a special case of this for a single copy in the previous application:

**Application** The following formula is key in many calculations involving random states:

$$\int (|\psi\rangle\langle\psi|)^{\otimes t} d\mu_H(\psi) = \frac{P_{\text{sym},t}}{\binom{d+t-1}{t}}, \quad (2.23)$$

where  $P_{\text{sym},t}$  denotes the orthogonal projector onto  $S^t(\mathbb{C}^d)$ . This follows from the fact that  $|\psi\rangle^{\otimes t} \in S^t(\mathbb{C}^d)$ . In particular, this space is an irreducible representation under the action  $\mathbf{U}^{\otimes t}$ . Then,

$$\int (|\psi\rangle\langle\psi|)^{\otimes t} d\mu_H(\psi) = \int \mathbf{U}^{\otimes t} (|\psi_0\rangle\langle\psi_0|)^{\otimes t} (\mathbf{U}^{\otimes t})^\dagger d\mu_H(\psi) \quad (2.24)$$

for some fixed  $|\psi_0\rangle$ . This projects  $(|\psi_0\rangle\langle\psi_0|)^{\otimes t}$  onto the commutant of the action  $\mathbf{U}^{\otimes t}$  by Corollary 3. However, by Schur's lemma, this commutant is a multiple of the orthogonal projector onto this irrep. Hence,

$$\int (|\psi\rangle\langle\psi|)^{\otimes t} d\mu_H(\psi) = \text{Tr} \left[ (|\psi_0\rangle\langle\psi_0|)^{\otimes t} \frac{P_{\text{sym},t}}{\|P_{\text{sym},t}\|_F} \right] \frac{P_{\text{sym},t}}{\|P_{\text{sym},t}\|_F} = \frac{P_{\text{sym},t}}{\|P_{\text{sym},t}\|_F^2}, \quad (2.25)$$

which yields the claimed result by observing that

$$\|P_{\text{sym},t}\|_{\mathbb{F}}^2 = \text{Tr}(P_{\text{sym},t}) = \dim S^t(\mathbb{C}^d) = \binom{d+t-1}{t},$$

where the last equation is the standard formula for the dimension of the symmetric subspace [Har13].

Another close connection between the Haar measure and representation theory arises from the fact that we can define the space  $L^2(G)$  of functions  $f : G \rightarrow \mathbb{C}$  that are square integrable over the Haar measure, with the usual inner product. This space comes with a representation called the *regular representation*:

$$\rho_{\text{reg}}(\mathbf{U}) : f \mapsto (\mathbf{V} \mapsto f(\mathbf{U}^{-1}\mathbf{V})). \quad (2.26)$$

A key result in representation theory is the *Peter-Weyl theorem* that roughly states that all irreducible representations appear in the decomposition of  $(L^2(G), \rho_{\text{reg}})$ .

**Theorem 5** (Peter-Weyl). *Let  $G$  be a compact group with Haar measure  $\mu_H$ . All irreducible representations of  $G$  are finite dimensional. Moreover, the set  $\Lambda_G$  of all irreducible representations  $V_\lambda$  up to isomorphisms is countable and we have the following decomposition*

$$L^2(G) \simeq \bigoplus_{\lambda \in \Lambda_G} V_\lambda^{\oplus \dim V_\lambda} \simeq \bigoplus_{\lambda \in \Lambda_G} L(V_\lambda). \quad (2.27)$$

Lastly, we introduce Weingarten calculus as a consequence of the projector formula derived in this section. We introduce the notation

$$|\psi_\pi\rangle := d^{-t/2} \text{vec}(r(\pi)), \quad (2.28)$$

where  $\text{vec} : \mathbb{C}^{D \times D} \rightarrow (\mathbb{C}^D)^{\otimes 2}$  defined by

$$\text{vec}(A) := \mathbb{1} \otimes A|\Omega\rangle, \quad (2.29)$$

for  $A \in \mathbb{C}^{D \times D}$  where  $|\Omega\rangle = \sum_i |ii\rangle$  denotes the (unnormalized) maximally entangled state for an ONB  $\{|i\rangle\}$  of  $\mathbb{C}^D$ . It can be seen that this map is alternatively defined by  $\text{vec}(|i\rangle\langle j|) = |i, j\rangle$ .

We know from Corollary 3 that  $M(\mu_H, t)$  is the orthogonal projector onto the span of the vectorizations  $|\psi_\pi\rangle$  of  $r(\pi)$ . Picking any orthonormal basis  $\{|\pi\rangle\}$  for this space, we can hence write the projector  $M(\mu_H, t)$  as a sum of the rank 1 projectors onto the basis states. Notice that, while  $\{|\psi_\pi\rangle\}$  constitutes a basis, the permutations are not mutually orthogonal. Nevertheless, we can expand each  $|\pi\rangle$  as a linear combination of the vectors  $|\psi_\pi\rangle$ . This yields the expansion

$$M(\mu_H, t) = \sum_{\pi \in S_t} |\pi\rangle\langle\pi| = \sum_{\pi, \sigma} (d^t W_d(\pi, \sigma)) |\psi_\pi\rangle\langle\psi_\sigma|. \quad (2.30)$$

The unique coefficients  $W_d(\pi, \sigma)$  are called Weingarten function. The factor  $d^t$  comes from the fact that we normalize the state  $|\psi_\pi\rangle$ .

To characterize these more precisely, we define the synthesis operator

$$\mathbf{V} = \sum_{\pi \in S_t} |\psi_\pi\rangle\langle\pi| \quad (2.31)$$

and the *Gram matrix*  $G$  by

$$G_{\pi\sigma} = (V^\dagger V)_{\pi\sigma} = \langle \psi_\pi | \psi_\sigma \rangle. \quad (2.32)$$

More precisely, we find

$$G_{\pi\sigma} = d^{-t} d^{\#\text{cycles}(\pi\sigma^{-1})}. \quad (2.33)$$

On the full space  $(\mathbb{C}^d)^{\otimes t}$  the matrices  $G$  and  $V$  are not invertible, but we can define the pseudoinverses  $G^+$  and  $V^+$ . That is, we only invert the matrices on their images and leave the kernel invariant. With these definitions, we can expand

$$\begin{aligned} \sum_{\pi \in S_t} |\pi\rangle \langle \pi| &= VV^+(V^+)^\dagger V^\dagger \\ &= V(V^\dagger V)^+ V^\dagger \\ &= VG^+V^\dagger \\ &= \sum_{\pi, \sigma} |\psi_\pi\rangle \langle \pi| G^+ |\sigma\rangle \langle \psi_\sigma| \\ &= \sum_{\pi, \sigma} (G^+)_{\pi, \sigma} |\psi_\pi\rangle \langle \psi_\sigma|. \end{aligned} \quad (2.34)$$

Comparing this to the expansion (2.30) characterizes the Weingarten functions as the matrix entries of the inverted Gram matrix:

$$W_d(\sigma, \pi) = d^{-t} \langle \pi | G^+ | \sigma \rangle \quad (2.35)$$

**Example 3.** We now put this formula to the test and derive a general integration formula for quadratic functions. For  $t = 2$ , we have the Gram matrix

$$G = \begin{pmatrix} \langle \psi_{\mathbb{1}} | \psi_{\mathbb{1}} \rangle & \langle \psi_{\mathbb{F}} | \psi_{\mathbb{1}} \rangle \\ \langle \psi_{\mathbb{1}} | \psi_{\mathbb{F}} \rangle & \langle \psi_{\mathbb{F}} | \psi_{\mathbb{F}} \rangle \end{pmatrix} = \frac{1}{d^2} \begin{pmatrix} \text{Tr}[\mathbb{1}] & \text{Tr}[\mathbb{F}] \\ \text{Tr}[\mathbb{F}] & \text{Tr}[\mathbb{1}] \end{pmatrix} = \frac{1}{d} \begin{pmatrix} d & 1 \\ 1 & d \end{pmatrix}. \quad (2.36)$$

We can verify that

$$G^{-1} = \frac{d^2}{d^2 - 1} \begin{pmatrix} 1 & -\frac{1}{d} \\ -\frac{1}{d} & 1 \end{pmatrix}. \quad (2.37)$$

Plugging this into (2.30) immediately yields the general formula

$$\int \mathbf{u}^{\otimes 2} A (\mathbf{u}^{\otimes 2})^\dagger = \frac{1}{d^2 - 1} \left( \mathbb{1} \text{Tr}(A) - \frac{1}{d} \mathbb{F} \text{Tr}(A) - \frac{1}{d} \mathbb{1} \text{Tr}(A\mathbb{F}) + \mathbb{F} \text{Tr}(A\mathbb{F}) \right). \quad (2.38)$$

Notice that we never had to construct an explicit ONB  $\{|\psi\rangle\}$  for this. This formula is ubiquitous in quantum information theory as expressions such as the LHS of (2.38) are common in the study of states that are invariant under local operations. In this context, the LHS of (2.38) is often referred to as a *twirl*. More generally, Eq. (2.35) can be used to exactly compute the  $t$ -fold moment operator. Indeed, strong bounds for the asymptotic behaviour of Weingarten functions for large dimensions are known [CM09, CM17]. However, there is no simple and explicit formula for higher moments and we are therefore incentivized to exploit as much symmetry as possible for higher moment calculations before we resort to Weingarten calculus. To exploit symmetries, a general formula by Collins and Sniady [CS06] can provide guidance: It expresses the integral

$\int \mathbf{U}^{\otimes t} \mathbf{A} (\mathbf{U}^\dagger)^{\otimes t}$  in terms of the *frame operator* applied to  $\mathbf{A}$  times a weighted sum of projectors onto the isotypic components of  $(\mathbb{C}^d)^{\otimes t}$ . Moreover, Ref. [CS06] also provides similar formulas for the integration over the orthogonal and symplectic group.

The formula (2.38) is a key ingredient for Ref. [HBRE21], which is presented in Chaper 5. In that chapter, we use the formula in a graphical calculus based on tensor network diagrams<sup>5</sup>:

$$\mathbb{E}_{\mathbf{U} \sim \mu_H} \left[ \begin{array}{c} \overline{\mathbf{U}} \\ \mathbf{U} \\ \overline{\mathbf{U}} \\ \mathbf{U} \end{array} \right] = \frac{1}{q^2 - 1} \left[ \begin{array}{c} ) \\ ( - \frac{1}{q} \\ ( - \frac{1}{q} \\ ( + \\ ( \end{array} \right]. \quad (2.39)$$

A diagrammatic approach to Weingarten calculus goes back to Ref. [BB96] (see also [NS21]). We refer to Ref. [BC17] for an extensive introduction to tensor network diagrams.

We can define the *frame operator* by

$$\mathbf{Q} = \mathbf{V} \mathbf{V}^\dagger = \sum_{\pi} |\psi_{\pi}\rangle \langle \psi_{\pi}| \geq 0. \quad (2.40)$$

We find that  $\sqrt{\mathbf{Q}^+} |\psi_{\pi}\rangle$  defines an orthonormal basis. This is because

$$|\psi_{\pi}\rangle = \mathbf{Q} \mathbf{Q}^+ |\psi_{\pi}\rangle = \sum_{\sigma} |\psi_{\sigma}\rangle \langle \psi_{\sigma} | \mathbf{Q}^+ |\psi_{\pi}\rangle \quad (2.41)$$

and the vectors  $|\psi_{\pi}\rangle$  are linearly independent. This general procedure is called symmetric orthogonalization (see e.g. Ref. [GNW21] for a similar application). Moreover, notice that the same procedure works for every group, where we instead consider the Gram matrix over a basis of the respective commutant.

If the permutations are approximately orthogonal, then we find that the frame operator is approximately the projector onto the span of permutations. Remarkably, the frame operator can also be interpreted as the moment operator of a (non-unitary) ensemble of matrices. Indeed, Wick's theorem [Wic50] shows that the  $t$ -th moment operator of matrices with i.i.d Gaussian entries is precisely the frame operator. It can be shown that for small submatrices the correlation between the entries can be neglected and that these submatrices are indeed approximately Gaussian [AA13]. This is conjectured even for  $t \times t$  submatrices with  $t = o(\sqrt{d})$  [Jiao6]. We conclude that the approximate orthogonality of permutation matrices is directly related to how Gaussian the Haar measure is up to  $t$  moments. As low-degree monomials in the entries of unitaries only see correlations in submatrices, we would expect the expectation values to behave effectively Gaussian. Indeed, this behavior can be made precise and is the key insight in many results on the generation of unitary designs. More precisely, it was shown in Ref. [BaHH16] that

$$\|\mathbf{M}(\mu_H, t) - \mathbf{Q}\|_{\infty} \leq \frac{t^2}{d}. \quad (2.42)$$

For higher moments, the correlations between the entries become increasingly important and we can therefore not expect that the scaling of this inequality can be greatly improved.

<sup>5</sup> Here, we use tensor network notation on the RHS. That is, strings between plaquettes represent contraction of indices, and plaquettes in parallel represent tensor products of operators. This notation will be used in Chapter 5.

Purely random bit strings are a ubiquitous design primitive for algorithms. There is reason, however, to view randomness as a precious resource and minimize the number of random bits required. First, randomness does not come perfectly uniform in nature. Whether there is true randomness (e.g. from quantum mechanics) or not, natural chaotic ensembles come with structure and biases. Second, the need for randomness in the design of algorithms is a puzzling phenomenon and many computer scientists believe that large classes of algorithms can be “derandomized” [AB09].

Consequently, it is desirable to construct *pseudorandom* probability distributions that mimic uniform randomness for practical purposes but require less random bits. These ensembles would be required to be *indistinguishable* from the uniform measure for an agent that has limited resources. Different limitations lead to different notions of pseudorandomness. For example, a standard notion of pseudorandom bit strings is *computational pseudorandomness*. These are ensembles, that cannot be distinguished from the uniform measure by any computation in polynomial time. Often, these guarantees come from conjectures in theoretical computer science.

In a more information theoretical setting, the agent can only read out  $t$  bits from the string. If any such  $t$  substrings are uniformly (or approximately uniformly) distributed, we speak of  $t$ -wise independence. Clearly, these ensembles can require far less random bits than the uniform distribution. Indeed, a single random bit suffices to generate a 1-wise independent ensemble: consider the bitstring  $01010101\dots$  and, with probability  $\frac{1}{2}$ , shift every bit to the right (or left) with periodic boundary conditions.

There are, in particular, powerful approximate constructions of  $t$ -wise independent permutations [BH08, Gow96, HMMR05]. More precisely, it can be shown that random classical reversible circuits of at most  $O(n^2 t^3)$  gates suffice to generate approximately  $t$ -wise independent ensembles.

A quantum version of pseudorandomness in general and  $t$ -wise independence in particular is naturally defined for probability distributions on state space or the unitary group. Intuitively, these are probability measures that are indistinguishable from the Haar measure if tested only against sufficiently simple functions. In particular, we define the notion of a *unitary  $t$ -design* [DCEL09, GAE07, AE07] as a probability measure  $\nu$  on the (special) unitary group that yields the same expectation values as the Haar measure for polynomials of degree at most  $t$ . In the following, we show that this definition has multiple, seemingly unrelated, characterizations.

We call a monomial  $f(x, y)$  balanced<sup>6</sup>, if the degree of each monomial in  $f$  in  $x$  is the same as the degree in  $y$ .

**Theorem 6.** *Let  $\nu$  be a probability measure on  $SU(d)$ . The following properties are equivalent.*

1.  $\mathbb{E}_{\mathcal{U} \sim \nu} f(\mathcal{U}, \overline{\mathcal{U}}) = \mathbb{E}_{\mathcal{U} \sim \mu_H} f(\mathcal{U}, \overline{\mathcal{U}})$  for all balanced polynomials  $f$  in the entries of  $\mathcal{U}$  and  $\overline{\mathcal{U}}$  of degree at most  $t$ .
2. Denote by  $\Phi_\nu^{(t)}$  the moment operator defined by

$$\Phi_\nu^{(t)}(\mathcal{A}) := \int \mathcal{U}^{\otimes t} \mathcal{A} (\mathcal{U}^{\otimes t})^\dagger d\nu(\mathcal{U}) \quad (2.43)$$

<sup>6</sup> Not to be confused with the notion of balanced boolean functions.

for every matrix  $A \in \mathbb{C}^{d^t \times d^t}$ . For  $\nu$  it holds that

$$\Phi_\nu^{(t)} = \Phi_{\mu_H}^{(t)}. \quad (2.44)$$

3. For the frame potential defined by

$$\mathcal{F}_\nu^{(t)} := \int \int |\text{Tr}(\mathbf{U}\mathbf{V}^\dagger)|^{2t} d\nu(\mathbf{U}) d\nu(\mathbf{V}) \quad (2.45)$$

it holds that

$$\mathcal{F}_\nu^{(t)} = \dim \text{span}(\mathbf{r}(\pi), \pi \in S_t). \quad (2.46)$$

*Proof.* 1.  $\implies$  2. This follows directly from the fact that for all  $|\psi\rangle$  and  $|\phi\rangle$ , we have

$$\langle \phi | \Phi_\nu^{(t)}(A) | \psi \rangle = \int \langle \phi | \mathbf{U}^{\otimes t} A (\mathbf{U}^{\otimes t})^\dagger | \psi \rangle d\nu(\mathbf{U}), \quad (2.47)$$

which is an expectation value over a balanced polynomial.

2.  $\implies$  1. First, notice that  $\Phi_\nu^{(t)} = \Phi_{\mu_H}^{(t)}$  implies  $\Phi_\nu^{(t-1)}(A) = \Phi_{\mu_H}^{(t-1)}$  by taking the partial trace over the  $t$  copies of  $\mathbb{C}^d$ . Moreover, by linearity of the expectation value we can restrict to balanced monomials but every balanced monomial can be written as

$$\langle i_1 | \otimes \dots \otimes \langle i_t | \mathbf{U}^{\otimes t} | j_1 \rangle \otimes \dots \otimes | j_t \rangle \langle i'_1 | \otimes \dots \otimes \langle i'_t | (\mathbf{U}^{\otimes t})^\dagger | j'_1 \rangle \otimes \dots \otimes | j'_t \rangle.$$

2.  $\implies$  3. It follows from Corollary 3 and the Schur-Weyl duality that  $\Phi_{\mu_H}^{(t)}$  is the projector onto the span of permutations. We introduce the short notation  $\mathbf{U}^{\otimes t, t} := \mathbf{U}^{\otimes t} \otimes \bar{\mathbf{U}}^{\otimes t}$ . We can reformulate

$$\begin{aligned} \mathcal{F}_\nu^{(t)} &= \int \int |\text{Tr}(\mathbf{U}\mathbf{V}^\dagger)|^{2t} d\nu(\mathbf{U}) d\nu(\mathbf{V}) \\ &= \int \int \text{Tr}(\mathbf{U}^{\otimes t, t} (\mathbf{V}^\dagger)^{\otimes t, t}) d\nu(\mathbf{U}) d\nu(\mathbf{V}) \\ &= \text{Tr} \left[ \int \mathbf{U}^{\otimes t, t} d\nu(\mathbf{U}) \left( \int \mathbf{U}^{\otimes t, t} d\nu(\mathbf{U}) \right)^\dagger \right] \\ &\stackrel{2.}{=} \text{Tr} \left[ \int \mathbf{U}^{\otimes t, t} d\mu_H(\mathbf{U}) \left( \int \mathbf{U}^{\otimes t, t} d\mu_H(\mathbf{U}) \right)^\dagger \right] \\ &= \text{Tr} \left[ \text{vec}(\Phi_\nu^{(t)}) \text{vec}(\Phi_\nu^{(t)})^\dagger \right] \\ &= \text{Tr} \left[ \text{vec}(\Phi_{\mu_H}^{(t)}) \text{vec}(\Phi_{\mu_H}^{(t)})^\dagger \right] \\ &= \text{Tr} \left[ \text{vec}(\Phi_{\mu_H}^{(t)}) \right] \\ &= \dim \text{span}(\mathbf{r}(\pi), \pi \in S_t). \end{aligned} \quad (2.48)$$

3.  $\implies$  1. We introduce the notation

$$\mathcal{M}(\nu, t) := \text{vec}(\Phi_\nu^{(t)}) = \int \mathbf{U}^{\otimes t, t} d\nu(\mathbf{U}). \quad (2.49)$$

Then, we can compute

$$\begin{aligned}
& \|M(\nu, t) - M(\mu_H, t)\|_F^2 \\
&= \text{Tr}[M(\nu, t)M(\nu, t)^\dagger] - \text{Tr}[M(\mu_H, t)M(\nu, t)^\dagger] - \text{Tr}[M(\nu, t)M(\mu_H, t)^\dagger] + \text{Tr}[M(\mu_H, t)M(\mu_H, t)^\dagger] \\
&\stackrel{\dagger}{=} \text{Tr}[M(\nu, t)M(\nu, t)^\dagger] - \text{Tr}[M(\mu_H, t)M(\mu_H, t)^\dagger] \\
&= \mathcal{F}_\nu^{(t)} - \mathcal{F}_{\mu_H}^{(t)},
\end{aligned} \tag{2.50}$$

where we have used Lemma 1 in  $\dagger$ . Eq. (2.50) immediately implies that  $\mathcal{F}_\nu^{(t)} \geq \mathcal{F}_{\mu_H}^{(t)}$ . Moreover, as  $\|\bullet\|_F$  is a norm, equality is satisfied if and only if  $M(\nu, t) = M(\mu_H, t)$ .  $\square$

Notice that for  $t \leq d$ , it was proven in [DS94] that

$$\dim \text{span}(r(\pi), \pi \in S_t) = t!, \tag{2.51}$$

or in other words, all permutations of tensor factors are linearly independent. Indeed, in the regime  $t \geq d$ , this relation does not hold anymore but the frame potential can nevertheless be characterized directly. In the general case, it was proven in Ref. [Rai98] that  $\mathcal{F}^{(t)}$  equals the number of permutations of length  $t$  with no increasing subsequence of length greater than  $d$ .

**Definition 5.** We call a measure that satisfies the conditions in Theorem 6 a unitary  $t$ -design.

In the special case that  $\nu$  is the Haar measure on a (Lie-)subgroup  $G$  of  $SU(d)$ , the  $t$ -design property is further equivalent to representation theoretical properties of  $G$ . A group that defines a unitary  $t$ -design is sometimes called a unitary  $t$ -group [BNRT18].

**Theorem 7.** Let  $\nu$  be the Haar measure on a Lie-subgroup of  $SU(d)$ . Then, the  $t$ -design property is equivalent to the following conditions:

1.  $\text{comm}(U^{\otimes t}, U \in G) = \text{comm}(U^{\otimes t}, U \in SU(d))$ .
2. The  $t$ -th tensor power representation of  $G$  has the same decomposition into irreps as  $SU(d)$ .
3. Every irrep in the  $t$ -th tensor power representation of  $SU(d)$  is also an irrep under the induced action of  $G$ .

*Proof.* It follows from Corollary 3 that

$$\Phi_\nu^{(t)} = P_{\text{comm}(U^{\otimes t}, U \in G)}. \tag{2.52}$$

It immediately follows that 1. is equivalent to condition 2. in Theorem 6.

2.  $\iff$  3. Consider the decomposition of  $t$ -th tensor power representation into irreducible representations  $V_\gamma$  of  $G$ :

$$(\mathbb{C}^d)^{\otimes t} \simeq \bigoplus_{\gamma} V_\gamma \otimes M_\gamma. \tag{2.53}$$

This equals the Schur-Weyl decomposition if and only if all irreps in the  $t$ -th tensor power representation of  $SU(d)$  remain irreducible under the action of  $G$ .



2.  $\implies$  1. For this equivalence, we can invoke the commutant theorem (Theorem 2), which characterizes the commutant of  $\{\mathbf{U}^{\otimes t}, \mathbf{U} \in \mathbf{G}\}$  as

$$\text{comm}\{\mathbf{U}^{\otimes t}, \mathbf{U} \in \mathbf{G}\} = \bigoplus_{\gamma} \mathbf{L}(M_{\gamma}). \quad (2.54)$$

Therefore, if the decomposition (2.53) equals the Schur-Weyl decomposition, then  $\text{comm}(\mathbf{U}^{\otimes t}, \mathbf{U} \in \mathbf{G}) = \text{comm}(\mathbf{U}^{\otimes t}, \mathbf{U} \in \text{SU}(d))$ .

1.  $\implies$  3. We have

$$\dim \text{comm}\{\mathbf{U}^{\otimes t}, \mathbf{U} \in \mathbf{G}\} = \sum_{\gamma} (\dim M_{\gamma})^2. \quad (2.55)$$

Assume an irrep  $W_{\lambda}$  of  $\text{SU}(d)$  in the Schur-Weyl decomposition decomposes non-trivially into irreps. Each of these now come with a multiplicity space at least as large as  $S_{\lambda}$ . Therefore, we find that  $\sum_{\gamma} (\dim M_{\gamma})^2$  is strictly larger than  $\sum_{\lambda} (\dim S_{\lambda})^2$ . Consequently, the two commutants have the same dimension (and are therefore equal), then no irrep in the  $t$ -th tensor power representation decomposes under the action of  $\mathbf{G}$ .  $\square$

In summary, unitary  $t$ -designs can be uniquely characterized by statistical properties, algebraic properties or representation theoretical properties. As such, they provide a natural link between all these fields. Switching between these different perspectives is going to provide an indispensable tool for the problems tackled in this thesis.

The key example of such a group in quantum information theory is the *Clifford group*. Let  $\mathcal{P}_n$  denote the Pauli group generated by all tensor products of  $n$  Pauli operators:

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.56)$$

The  $n$ -qubit Clifford group  $\text{Cl}(n)$  is the unitary normalizer of the Pauli group  $\mathcal{P}_n$ :

$$\text{Cl}(n) = \left\{ \mathbf{U} \in \mathbf{U}(2^n, \mathbf{Q}[i]) \mid \mathbf{U}\mathcal{P}_n\mathbf{U}^{\dagger} \subset \mathcal{P}_n \right\}. \quad (2.57)$$

Here, we restrict to elements with entries in  $\mathbf{Q}[i]$  as otherwise the group would become continuous due arbitrary phases. The Clifford group has multiple remarkable properties:

- It is a discrete group with

$$2^{n^2+2n} \prod_{j=1}^n (4^j - 1) \quad (2.58)$$

elements (see e.g. [GNW21]).

- It forms a unitary 3-design [Web15, Zhu17].
- Every element has a polynomially upper bounded circuit complexity [AG04, BM21].
- Moreover, computations performed by applications of Clifford unitaries can be efficiently classically simulated by the Gottesmann-Knill theorem [Got97, Got98].

- Clifford operations can be comparably easily protected from errors and hence serve as the foundation of most error-correcting codes [Got97].

The Clifford group however fails “gracefully” to be a unitary 4-design [ZKGG16]. More precisely, while the fourth tensor power representation of the unitary group decomposes under the induced action of the Clifford group, the number of irreducible representations remains constant in the system-size. Relatedly, the commutant of the fourth tensor power representation of the Clifford group is generated by the permutations plus a single additional generator: The projector

$$Q := 2^{-2n} \sum_{P \in \mathcal{P}_n} P^{\otimes 4}. \quad (2.59)$$

While the Clifford commutant’s dimension starts to grow rapidly with  $t$ , it remains true for higher moments that its dimension is independent of the system-size. In fact, the commutant of the Clifford group is characterized in Ref. [GNW21]. This observation is going to be the key ingredient in Chapter 3.

## 2.5 RANDOM WALKS, SPECTRAL GAPS AND APPROXIMATE DESIGNS

So far, we have seen that unitary  $t$ -designs satisfy strong constraints in particular, if the probability measure in question is the uniform measure on a subgroup of  $SU(d)$ . Given these strong properties, it is perhaps not surprising that the existence of unitary  $t$ -groups is heavily limited [BNRT18, GT05]. More precisely, there is no family of 4-groups for all number of qubits. More strikingly, the multiqubit Clifford group is actually the unique family of 3-designs that exists for all number of qudits. Similarly, every family of 2-designs in a prime power dimension is a subgroup of the multiqubit Clifford group and except for these there is only a finite list of sporadic designs. Of course, for many applications, we do not require a group structure at all and, in fact, it can be abstractly proven that finitely supported spherical  $t$ -designs always exist [BRV13] for every  $t$ . While explicit constructions of unitary  $t$ -designs are known [BNOZ22], the individual states in their support require exponentially deep quantum circuits. In fact, the circuits in these constructions to generate a unitary 4-design are of a depth that is sufficient to approximate the Haar measure. On the other hand, we usually do not require exact  $t$ -designs for practical purposes. Indeed, the majority of this thesis is concerned with probability distributions that only approximately satisfy the  $t$ -design property. Consequently, we will have to deal with moment operators that are not orthogonal projectors. Luckily, we are not the first to encounter this problem.

In general, the Fourier transform of a measure  $\nu$  on a compact Lie group  $G$  is the (in general infinite dimensional) operator

$$F_G(\nu) = \bigoplus_{\lambda \in \Lambda_G} \int \rho_\lambda d\nu(U) : \bigoplus_{\lambda \in \Lambda_G} V_\lambda \rightarrow \bigoplus_{\lambda \in \Lambda_G} V_\lambda. \quad (2.60)$$

For the special case of a torus  $T^d = S^1 \times \dots \times S^1$ , the irreducible representations are all 1-dimensional and uniquely labeled by tuples  $\lambda = (i_1, \dots, i_d)$  of integers. The corresponding representations are

$$\rho_\lambda((e^{ix_1}, \dots, e^{ix_d}))|v\rangle = e^{i\langle \lambda, x \rangle} |v\rangle. \quad (2.61)$$

Therefore,  $F_{T^d}(\nu)$  is technically an assignment of  $\lambda$  to the value  $\int e^{i\langle \lambda, x \rangle} d\nu(x)$ , which is the classical Fourier transform on the torus.

However, over a noncommutative group such as  $SU(d)$  for  $d \geq 2$ , the irreps  $V_\lambda$  are not all 1-dimensional. Therefore, it can in general be an intricate problem to characterize the spectrum of the Fourier transform for a given measure. For many applications, we are in need of precisely such a spectral analysis because this provides a way of measuring how random a distribution  $\nu$  is. More precisely, if all eigenvalues (except the trivial component) are sufficiently bounded away from 1, then we consider the distribution to be evenly distributed. Notice that the highest eigenvalue is always 1, realized on the trivial component and the Fourier transform of the Haar measure is by Lemma 1 precisely this component. We can quantify the randomness of a distribution by the following spectral gap:

$$\Delta(\nu) := 1 - \max_{\lambda \in \Lambda \setminus \{\text{triv}\}} \left\| \int \rho_\lambda(\mathbf{U}) d\mu(\mathbf{U}) \right\|_\infty = 1 - \|F(\nu) - F(\mu_H)\|_\infty. \quad (2.62)$$

We would expect a very uniformly random ensemble to give expectation values close to those of the completely uniform measure. A non-vanishing spectral gap does not guarantee closeness in expectation value. However, the gap can be exponentially amplified by an operation called *convolution*. Convolution is defined as an operation on probability measures over group. For this consider two probability measure  $\nu_1$  and  $\nu_2$ . The convolution  $\nu_1 * \nu_2$  is defined by drawing a random group element from the distribution  $\nu_1$  and another independently from  $\nu_2$  and then concatenating these two group elements. In measure theoretic terms the convolution is the pushforward measure of the product measure  $\nu_1 \times \nu_2$  under the group multiplication. The process of convoluting a measure with itself repeatedly is called a *random walk* and denoted by  $\nu^{*k} = \nu * \dots * \nu$ . In the following we denote  $F_{SU(d)}$  by  $F$ . We find that

$$\begin{aligned} F(\nu^{*k}) &= \bigoplus_{\lambda \in \Lambda_G} \int \rho_\lambda d\nu^{*k}(\mathbf{U}) \\ &= \bigoplus_{\lambda \in \Lambda_G} \int \dots \int \rho(\mathbf{U}_1 \dots \mathbf{U}_k) d\nu(\mathbf{U}_1) \dots d\nu(\mathbf{U}_k) \\ &= \bigoplus_{\lambda \in \Lambda_G} \int \dots \int \rho(\mathbf{U}_1) \dots \rho(\mathbf{U}_k) d\nu(\mathbf{U}_1) \dots d\nu(\mathbf{U}_k) \\ &= \left( \bigoplus_{\lambda \in \Lambda_G} \int \rho_\lambda d\nu(\mathbf{U}) \right)^k \\ &= F(\nu)^k. \end{aligned} \quad (2.63)$$

And as a consequence

$$\begin{aligned} \Delta(\nu^{*k}) &= 1 - \|F(\nu^{*k}) - F(\mu_H)\|_\infty = 1 - \|(F(\nu) - F(\mu_H))^k\|_\infty \\ &\geq 1 - \|F(\nu^{*k}) - F(\mu_H)\|_\infty^k = 1 - (1 - \Delta(\nu))^k. \end{aligned} \quad (2.64)$$

Moreover, notice that we have equality whenever  $F(\nu)$  is a normal operator and in particular, when it is hermitian. Therefore, for symmetric measure  $\nu$ , the gap amplification is strict.

As an example, consider a finite group  $H$  and a set of generators  $T$  of  $H$ . Then, let  $\nu_T$  be the uniform distribution over the elements of  $T$ . The spectral gap  $\Delta(\nu_T)$  characterizes if the Cayley

graph of  $H$  with respect to  $T$  is expanding. More concretely, a famous problem in the theory of random walks concerns the symmetric group  $S_{52}$ : the group of ways to arrange a deck of 52 cards. How many shuffles does it take to mix the deck completely? The group contains a daunting number of  $52!$  elements. Nevertheless, it was proven in Refs. [AD86, BD92, DGK83] that after 7 steps of a random walk that “shuffles” the deck of card, the total variation to the uniform distribution is below  $\frac{1}{2}$ . More generally, a cut-off phenomenon emerges [Dia96]: the distance to the uniform distributions remains almost constantly close to 1 until a threshold is reached. After this threshold, the distance decays exponentially. We will find similar cut-off behaviors in the generation of unitary designs.

Another example of a random walk is Kac’s random walk on  $SO(d)$  [Kac47]. Here, each step of the walk consists of drawing a random  $2 \times 2$  submatrix from  $SO(2)$  and embed it into  $SO(d)$ . This walk was motivated from the theory of Brownian motions. An obvious analogue can be defined for  $SU(d)$ . Bounding the convergence of this process was a key objective of the theory of random walks [DSC00]. The full spectrum of the walk (and therefore the gap) was finally computed in [Mas03, CCL03]. We make use of the techniques developed here in Ref. [HHJ21].

A probability distribution  $\nu$  with spectral gap  $\Delta(\nu) > 0$  is called a  $(\infty, \Delta(\nu))$ -tensor product expander. The above example motivates why we should think of a gapped probability distribution as an expander, but where does the word “tensor” come from? To see this, we need to recall the definition of a unitary  $t$ -design in terms of moment operators:

$$M(\nu, t) = \int U^{\otimes t, t} d\nu(U) = \int U^{\otimes t, t} d\mu_H(U) = M(\mu_H, t). \quad (2.65)$$

The integral over a group action is unitarily equivalent to the integral over an isomorphic group action and block decomposes into the irreps of the representation. Moreover, it turns out that the representation we average over in the moment operators contains all irreps. This is a consequence of Chevalley’s generalization of Burnside’s theorem to Lie groups [Che16].

**Theorem 8.** *For a compact Lie group  $G$ , let  $\rho$  be a faithful (i.e. injective) irreducible representation. Then, every irrep of  $G$  is contained in  $\rho^{\otimes n} \otimes \bar{\rho}^{\otimes m}$  for some  $n, m \geq 0$ .*

Observe that the representation defined by  $U \mapsto U^{\otimes t, t}$  is isomorphic to  $\text{Ad}^{\otimes t}$ .  $\text{Ad}$  is irreducible if restricted to act on traceless matrices. Moreover, we find that it is faithful: consider  $U, V \in SU(d)$  such that  $U \otimes \bar{U} \neq V \otimes \bar{V}$ . Then,

$$0 < \|U \otimes \bar{U} - V \otimes \bar{V}\|_F^2 = 2d \left( 1 - \frac{1}{d} |\text{Tr}[UV^\dagger]|^2 \right). \quad (2.66)$$

But  $\frac{1}{d} |\text{Tr}[UV^\dagger]|^2$  is 1 if and only if  $U = V$  by Cauchy-Schwarz. Moreover,  $\text{Ad}$  is self-dual ( $\overline{\text{Ad}} \cong \text{Ad}$ ), so Burnside’s theorem implies that every irrep of  $SU(d)$  is contained in some tensor power of the adjoint action by Theorem 8. Therefore, we find

$$\Delta(\nu) = 1 - \sup_t \|M(\nu, t) - M(\mu_H, t)\|_\infty. \quad (2.67)$$

Often it suffices to consider only  $t$  moments for an application and hence we can define the more refined notion of a  $(t, \delta)$ -tensor product expander [HH08], which is a probability distribution  $\nu$  that satisfies

$$\delta = 1 - \|M(\nu, t) - M(\mu_H, t)\|_\infty. \quad (2.68)$$

In this sense, the spectral gap of the moment operator as defined in (2.68) measures how close a measure is to being a unitary  $t$ -design. As for the spectral gap of the Fourier transform, the spectral gap of moment operators is amplified exponentially by convolution powers. This often allows to translate a large spectral gap to more practical or operational notions of approximate unitary designs.

Before we present these notions of approximate designs, we introduce a third operator containing the same spectrum as the Fourier transform and the moment operators. Recall the definition of the left regular representation in Eq. (2.26) and consider the operators

$$\mathbb{T}_v : L^2(\mathrm{SU}(d)) \rightarrow L^2(\mathrm{SU}(d)), \quad \mathbb{T}_v := \int_{\mathrm{SU}(d)} \rho_{\mathrm{reg}}(\mathbb{U}) d\nu(\mathbb{U}). \quad (2.69)$$

These are sometimes called *Hecke operators* [LPS86, LPS87, BG08, BG12]. Hence, to translate the mathematical literature to a language closer to quantum information theory, it is often sufficient to realize that the operator  $\mathbb{T}_v$  block diagonalizes into irreducible representations and by the Peter-Weyl theorem (Theorem 5), all irreps are contained in  $L^2(\mathrm{SU}(d))$ .

We introduce the diamond norm [Wat18] of a superoperator  $\Phi : L(\mathbb{C}^D) \rightarrow L(\mathbb{C}^D)$  by

$$\|\Phi\|_{\diamond} := \max_{X, \|X\|_1 \leq 1} \|(\Phi \otimes \mathbb{1}_D)X\|_1. \quad (2.70)$$

In other words, the diamond norm is the stabilized version of the induced 1-norm:

$$\|\Phi\|_{1 \rightarrow 1} := \max_{X, \|X\|_1 \leq 1} \|\Phi X\|_1. \quad (2.71)$$

With this definition, we can introduce operationally and practically motivated definitions of approximate designs:

**Definition 6.** We call a probability distribution on  $\mathrm{SU}(d)$   $a(n)$

- additive  $\varepsilon$ -approximate unitary design if

$$\left\| \Phi_v^{(t)} - \Phi_{\mu_H}^{(t)} \right\|_{\diamond} \leq \varepsilon. \quad (2.72)$$

- strong additive  $\varepsilon$ -approximate unitary design if

$$\left\| \Phi_v^{(t)} - \Phi_{\mu_H}^{(t)} \right\|_{\diamond} \leq \frac{\varepsilon}{d^t}. \quad (2.73)$$

- relative  $\varepsilon$ -approximate design if

$$(1 - \varepsilon)\Phi_v^{(t)} \preceq \Phi_{\mu_H}^{(t)} \preceq (1 + \varepsilon)\Phi_v^{(t)}, \quad (2.74)$$

where  $A \succcurlyeq B$  if and only if  $B - A$  is completely positive for channels  $A$  and  $B$ .

An additive approximate design has low distinguishability from the Haar measure using the moment operator as a channel. More precisely, the 1-norm distance between two states  $\rho$  and  $\sigma$  quantifies how well  $\rho$  and  $\sigma$  can be distinguished via a single 2 outcome measurement. The induced 1-norm difference of two channels hence measures how well the two channels can be distinguished by applying them to some input state and then measuring the outcome. The

diamond norm additionally accounts for inputs entangled with ancillary systems and therefore quantifies the operational single-shot distinguishability of channels. Now consider the following scenario: You are given an unknown random source of unitaries and you want to know how evenly distributed this source is. An obvious operational approach would be to sample the  $t$ -th moment operator from this source and compare it to the  $t$ -th moment operator for the Haar measure. If the source is guaranteed to be distributed according to an additive approximate design, we know that the two channels are difficult to distinguish and the source is thus pseudo-random in an operational sense. That is not to say that additive approximate designs with constant  $\varepsilon$  do not have applications. Indeed, we present such an application in the appendix of the publication presented in Section 3.1.

The strong additive designs and relative designs, while technically inequivalent, tend to have the same properties. More precisely, both tend to have the same properties as exact designs for all practical purposes in quantum information theory. Indeed, a prototypical application in quantum information would involve concentration results for monomials in entries  $\langle i|U|0\rangle$  for a random unitary  $U$ . We have seen already that  $t$ -th moments of this random variable over the Haar measure scale like  $t!/d^t$ , where we usually have  $d = 2^n$  for a system-size of  $n$ . Consequently, we find that the strong additive and relative designs reproduce the scaling of Haar averages in  $n$  and therefore imply similar concentration results. For an example of this phenomenon, we refer to the bounds on circuit complexity from higher moments of overlaps of states in the proof of Proposition 2. Here, we require a scaling of  $O(d^{-t})$  to lift the statement to approximate designs. In this context, we also need to define approximate state designs:

**Definition 7.** We call a probability distribution  $\nu$  on  $\mathcal{S}_d$

- an additive  $\varepsilon$ -approximate state  $t$ -design if

$$\left\| \int (|\psi\rangle\langle\psi|)^{\otimes t} d\nu(\psi) - \int (|\psi\rangle\langle\psi|)^{\otimes t} d\mu_H(\psi) \right\|_1 \leq \varepsilon. \quad (2.75)$$

- a strong additive  $\varepsilon$ -approximate state  $t$ -design if

$$\left\| \int (|\psi\rangle\langle\psi|)^{\otimes t} d\nu(\psi) - \int (|\psi\rangle\langle\psi|)^{\otimes t} d\mu_H(\psi) \right\|_1 \leq \frac{\varepsilon}{d^t}. \quad (2.76)$$

Consider also Ref. [Low10], which compares more approximate notions of unitary designs.

Finally, *Random quantum circuits* are random walks on  $SU(q^n)$ . They can be viewed as an experimental protocol to generate pseudorandomness but also as a (toy) model for time evolution of chaotic systems. Arguably, it is the simplest model of dynamics that captures the concept of local interactions without imposing any particular further structure. In each step of this walk, one draws a random position (e.g. a random pair of nearest neighbors) and applies a Haar random unitary from  $SU(q^2)$  to this pair of qudits. It can be proven [BaHH16, Haf22] that this random walk is indeed gapped. Moreover, for moments satisfying  $t \leq e^{\Omega(n)}$  it was shown in a seminal paper [BaHH16] that these random quantum circuits indeed form tensor product expanders after a depth polynomial in the system-size and the moment.

In the next chapter we will define yet another random walk we call  $K$ -interleaved Clifford circuits, which quickly generates approximate  $t$ -designs for small values of  $t$ .

## QUANTUM PSEUDORANDOMNESS AND GROWTH OF QUANTUM CIRCUIT COMPLEXITY

---

### 3.1 UNITARY DESIGNS WITH A SYSTEM-SIZE INDEPENDENT NUMBER OF NON-CLIFFORD UNITARIES

Exact unitary 3-designs can be implemented by drawing a random unitary from the multi-qubit Clifford group. Clifford unitaries are well-behaved in a number of ways: They can be efficiently simulated classically and are comparably easy to implement fault-tolerantly. In fact, the difference to non-Clifford gates is so stark that modern resource theories of quantum computing treat Clifford gates as a free resource.

However, recent mathematical results show that higher designs can not be obtained exactly without implementing full Haar-randomness [BNRT18, GT05]. In seminal work [BaHH16], Brandao, Harrow and Horodecki have proven that  $t$ -designs can, nevertheless, be implemented approximately with local random quantum circuits of depth  $O(n^2 t^{10.5})$ . More recently, this depth was improved to  $O(n^2 t^{5+o(1)})$  in Ref. [Haf22], using the random walk introduced in this chapter. The full implementation of local random quantum circuits is daunting especially for near-term technology as it requires a fully fledged quantum computer. Moreover, the resulting circuits will not be simulable in polynomial time even for small fixed  $t$ . This is in stark contrast to the case of  $t = 3$ . This state of affairs suggest that -- in some ways costly -- non-Clifford gates have to be inserted into a random Clifford circuit in order to uplift unitary 3-designs to approximate higher-order unitary designs. This leads us to the central question underlying this work: *How many non-Clifford gates are required to generate an approximate unitary  $t$ -design?*

In this work, we show that surprisingly, the number on non-Clifford gates that need to be inserted into a random Clifford circuit to generate an additive approximate unitary  $t$ -design is independent of the system-size and the polynomial in  $t$ , provided that  $t$  is not too large:  $n \geq O(t^2)$ .

Let  $\nu$  be a probability measure on the unitary group  $U(d)$ . In this section, we work with (*additive*)  $\varepsilon$ -approximate  $t$ -designs. Recall that these are by Definition 6 distributions that satisfy

$$\left\| \Phi_{\nu}^{(t)} - \Phi_{\mu_H}^{(t)} \right\|_{\diamond} \leq \varepsilon, \quad (3.1)$$

where  $\|\cdot\|_{\diamond}$  is the diamond norm.

We consider uniformly drawn Clifford unitaries interleaved with a random single qubit gate drawn from  $\{K, K^{\dagger}, \mathbb{1}\}$ , where  $K$  is an arbitrary but fixed non-Clifford gate as illustrated in the figure. Our main result about these circuits is the following theorem:

**Theorem 9** (Unitary designs with few non-Clifford gates). *Let  $K \in U(2)$  be a non-Clifford unitary. Then, a  $K$ -interleaved Clifford circuit of depth  $O(t^4 \log^2(t) \log(1/\varepsilon))$  acting on  $n = O(t^2)$  qubits is an additive  $\varepsilon$ -approximate  $t$ -design.*

The number of non-Clifford gates in this result is system-size independent. Our result has multiple consequences: First, combined with the asymptotically optimal decomposition of Clifford unitaries [AGo4] into 2-local gates, our construction features an overall number of gates of  $O(n^2 t^4 \log^2(t) / \log(n))$ . This is an improved scaling compared to Ref. [BaHH16] in both  $t$  and  $n$  (for small  $t$ ). Second, Theorem 9 ensures the existence of families of additive  $\varepsilon$ -approximate unitary  $\log^{1/4}(n)$ -designs which are efficiently classically simulable Ref. [BBC<sup>+</sup>19].

We can also show that  $K$ -interleaved Clifford circuits of depth  $O(t^4 n)$  generate relative and strong additive designs (Definition 6) in the regime  $n = O(t^2)$ . This is not system-size independent anymore, but involves only a square root of the number of non-Clifford gates in other constructions [BaHH16, Haf22, HM18].

In order to make contact with a circuit constructions with random local gates, we in our work additionally identify rigorous bounds on the convergence of random walks of local Clifford generators to the moments of the uniform distribution on the Clifford group.

**Theorem 10** (Local random Clifford designs). *Let  $n \geq 12t$ , then a local random Clifford circuit of depth  $O(n^2 \log^{-2}(t) t^9 \log(1/\varepsilon))$  constitutes an  $\varepsilon$ -approximate  $t$ -design with respect to the uniform distribution on the Clifford group.*

This result is of independent interest and significantly improves the previously indicated scaling of  $O(n^8)$  [DLT02]. Together with Theorem 9 it provides a construction for unitary  $t$ -designs with a system-size-independent number of non-Clifford gates in terms of a circuit only consisting of random local gates.

A key tool we use is a variant of the Schur-Weyl duality for the Clifford group [GNW21]. This characterizes the commutant of the  $t$ -th diagonal action of the Clifford group in terms of a concept from symplectic geometry: stochastic Lagrangian subspaces  $T \in \Sigma_{t,t}$  of the vector space  $\mathbb{F}_2^{2t}$ .

We prove various auxiliary results about these Lagrangian subspaces. This includes a quantitative bound on the Gram-Schmidt orthogonalization of the resulting basis, which achieved by careful combinatorial bounds based on the statistics of cycles in random permutations. A most involved auxiliary results is a bound on the overlap of representations of Lagrangian subspaces with the commutant of the unitary group.

**Lemma 2.** *For all  $t$  and for all  $T \in \Sigma_{t,t} \setminus S_t$ , we have*

$$(Q_T | P_{\text{Haar}} | Q_T) \leq \frac{7}{8}, \quad (3.2)$$

where  $Q_T$  is the basis vector of the  $t$ -th commutant of the Clifford group corresponding to  $T$  and  $P_{\text{Haar}} = \Delta_t(\mu_{\text{Haar}})$  is the  $t$ -th moment operator of the single-qubit unitary group  $U(2)$ .

This is proven using a geometrical argument involving finite phase space methods. It is moreover essentially optimal, as we find examples that saturate a lower bound of  $7/10$ . This result can also be seen as a sanity check for the constant spectral gap conjecture for local random quantum circuits (compare the outlook of Ref. [Haf22]). We combine the before-mentioned bounds with deep results from harmonic analysis about spectral gaps of Hecke operators restricted to irreducible representations of Lie groups due to P. Varjú [Var13].



**Open Access** The following article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



# Efficient Unitary Designs with a System-Size Independent Number of Non-Clifford Gates

J. Haferkamp<sup>1</sup>, F. Montealegre-Mora<sup>2</sup>, M. Heinrich<sup>2,3</sup>, J. Eisert<sup>1</sup>, D. Gross<sup>2</sup>,  
I. Roth<sup>1,4</sup>

<sup>1</sup> Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin, Germany  
E-mail: [jonas.haferkamp@fu-berlin.de](mailto:jonas.haferkamp@fu-berlin.de)

<sup>2</sup> Institute for Theoretical Physics, University of Cologne, Cologne, Germany

<sup>3</sup> Present address: Quantum Technology Research Group, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>4</sup> Present address: Quantum Research Centre, Technology Innovation Institute, Abu Dhabi, UAE

Received: 14 June 2020 / Accepted: 16 August 2022  
Published online: 12 November 2022 – © The Author(s) 2022

**Abstract:** Many quantum information protocols require the implementation of random unitaries. Because it takes exponential resources to produce Haar-random unitaries drawn from the full  $n$ -qubit group, one often resorts to  $t$ -designs. Unitary  $t$ -designs mimic the Haar-measure up to  $t$ -th moments. It is known that Clifford operations can implement at most 3-designs. In this work, we quantify the non-Clifford resources required to break this barrier. We find that it suffices to inject  $O(t^4 \log^2(t) \log(1/\varepsilon))$  many non-Clifford gates into a polynomial-depth random Clifford circuit to obtain an  $\varepsilon$ -approximate  $t$ -design. Strikingly, the number of non-Clifford gates required is independent of the system size – asymptotically, the density of non-Clifford gates is allowed to tend to zero. We also derive novel bounds on the convergence time of random Clifford circuits to the  $t$ -th moment of the uniform distribution on the Clifford group. Our proofs exploit a recently developed variant of Schur-Weyl duality for the Clifford group, as well as bounds on restricted spectral gaps of averaging operators.

Random vectors and unitaries are ubiquitous in protocols and arguments of quantum information and many-body physics. In quantum information, a paradigmatic example is the *randomized benchmarking protocol* [1–3], which aims to characterize the error rate of quantum gates. There, random unitaries are used to average potentially complex errors into a single, easy to measure error rate. In many-body physics, random unitaries are used e.g. to model the dynamics that are thought to describe the mixing process that quantum information undergoes when absorbed into, and evaporated from, a black hole [4]. In these and related cases, one is faced with the issue that unitaries drawn uniformly from the full many-body group are *unphysical* in the sense that, with overwhelming probability, they cannot be implemented efficiently. The notion of a *unitary  $t$ -design* captures an efficiently realizable version of uniform randomness [5–7]. More specifically, a probability measure on the unitary group is a  $t$ -design if it matches the uniform Haar measure up to  $t$ -th moments.

Applications abound. The randomness provided by designs is used to foil attackers in quantum cryptography protocols [8–10]. It guards against worst case behavior in various

quantum [10–16] and classical [17] estimation problems. Designs allow for an efficient implementation of *decoupling* procedures, a primitive in quantum Shannon theory [18]. In quantum complexity, unitary designs are used as models for generic instances of time evolution that display a quantum computational speed-up [19,20]. Unitary designs are now standard tools for the quantitative study of toy models in high energy physics, quantum gravity, and quantum thermodynamics [4,21–23].

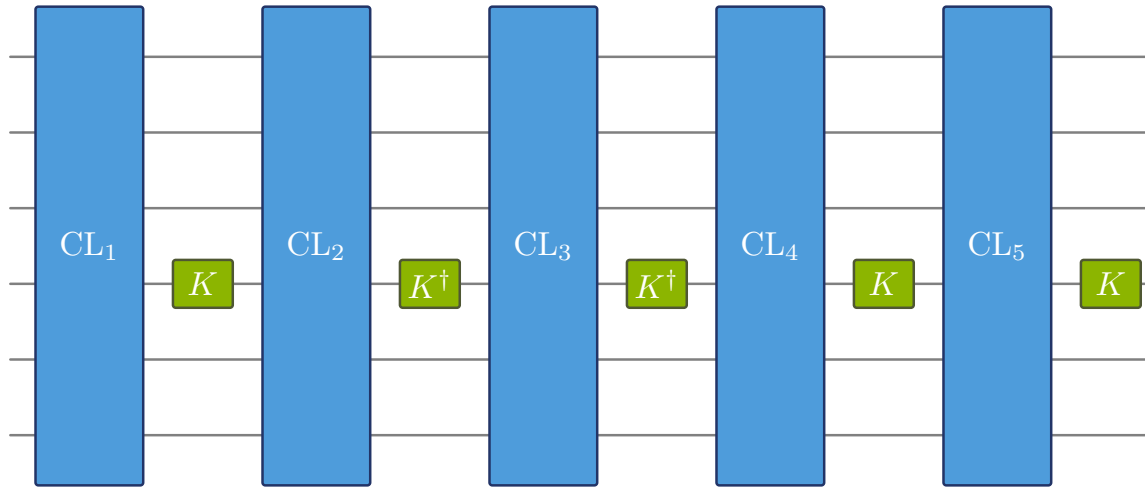
The multitude of applications motivates the search for efficient constructions of unitary  $t$ -designs [24–28]. In particular, Brandao, Harrow and Horodecki [24] show that local random circuits on  $n$  qubits with  $O(n^2 t^{10})$  many gates give rise to an approximate  $t$ -design. In practice, it is often desirable to find more structured implementations. Designs consisting of *Clifford operations* would be particularly attractive from various points of view: (i) Because the Clifford unitaries form a finite group, elements can be represented exactly using a small number ( $O(n^2)$ ) of bits. (ii) The Gottesman-Knill Theorem ensures that there are efficient classical algorithms for simulating Clifford circuits. (iii) Most importantly, in *fault-tolerant architectures* [29,30], Clifford unitaries tend to have comparatively simple realizations, while the robust implementation of general gates (e.g. via *magic-state distillation*) carries a significant overhead. The difference is so stark that in this context, Clifford operations are often considered to be a free resource, and the complexity of a circuit is measured solely in terms of the number of non-Clifford gates [31,32].

The Clifford group is known to form a unitary  $t$ -design for  $t = 2$  [9] and  $t = 3$  [33–35], but fails to have this property for  $t > 3$  [33–37]. In fact, the Clifford group is singled out among the finite subgroups of the unitary group by being a 3-design [38]. Moreover, Refs. [38,39] together imply that *any* local gate set that generates an exact unitary design of order  $t > 3$  must necessarily be universal, c.f. the discussion in Sect. 5. Hence, any efficient design construction for  $t > 3$  can only be approximate, and the Clifford group seems to be a distinguished starting point.

This leads us to the central question underlying this work: *How many non-Clifford gates are required to generate an approximate unitary  $t$ -design?* A direct application of the random circuit model of Ref. [24] yields an estimate of  $O(n^2 t^{10})$  non-Clifford operations. In this paper we show that a polynomial-sized random Clifford circuit, together with a *system size-independent* number of  $O(t^4 \log^2(t))$  non-Clifford gates – a “homeopathic dose” – is already sufficient.

We establish this main result for two different circuit models (Fig. 1). In Sect. 1.1, we consider alternating unitaries drawn uniformly from the Clifford group with a non-Clifford gate. This gives rise to an efficient quantum circuit, as there are classical algorithms for sampling uniformly from the Clifford group, and for producing an efficient gate decomposition of the resulting operation [40]. A somewhat simpler model is analyzed in Sect. 1.2. There, we assume that the Clifford layers are circuits consisting of gates drawn from a local Clifford gate set. These circuits will only approximate the uniform measure on the Clifford group. Theorem 2, which might be of independent interest, gives novel bounds on the convergence rate.

The key to this scaling lies in the structure of the commutant of the  $t$ -th tensor power of the Clifford group, described by a variant of Schur-Weyl duality developed in a sequence of recent works [36,41–43]. There, it has been shown that the dimension of this commutant – which measures the failure of the Clifford group to be a  $t$ -design from a representation theoretical perspective – is independent of the system size. Refs. [36,42] have used this insight to provide a construction for exact *spherical*  $t$ -designs that consist of a system size-independent number of Clifford orbits. It has been left as an



**Fig. 1.**  $K$ -interleaved Clifford circuits: We consider a model where random Clifford operations are alternated with a non-Clifford gate  $K$  or its inverse  $K^\dagger$

open problem whether these ideas can be generalized from spherical designs to the more complex notion of unitary designs, and whether the construction can be made efficient [42]. The present work resolves this question in the affirmative.

Finally, we note that in Ref. [44], it has been observed numerically that adding a single  $T$  gate to a random Clifford circuit has dramatic effects on the entanglement spectrum. A relation to  $t$ -designs was suspected. Our result provides a rigorous understanding of this observation.

## 1. Results

*1.1. Approximate  $t$ -designs with few non-Clifford gates.* To state our results precisely, we need to formalize the relevant notion of approximation, as well as the circuit model used. Let  $\nu$  be a probability measure on the unitary group  $U(d)$ . The measure  $\nu$  gives rise to a quantum channel

$$\mathbf{M}_t(\nu)(\rho) := \int_{U(d)} U^{\otimes t} \rho (U^\dagger)^{\otimes t} d\nu(U), \quad (1)$$

which applies  $U^{\otimes t}$ , with  $U$  chosen according to  $\nu$ . We will refer to  $\mathbf{M}_t(\nu)$  as the  $t$ -th moment operator associated with  $\nu$ . Following Ref. [27], we quantify the degree to which a measure approximates a  $t$ -design by the diamond norm distance of its moment operator to the moment operator of the Haar measure  $\mu_H$  on  $U(d)$ .

**Definition 1** (*Approximate unitary design*). Let  $\nu$  be a distribution on  $U(d)$ . Then  $\nu$  is an (additive)  $\varepsilon$ -approximate  $t$ -design if

$$\|\mathbf{M}_t(\nu) - \mathbf{M}_t(\mu_H)\|_\diamond \leq \varepsilon. \quad (2)$$

Denote the uniform measure on the multiqubit Clifford group  $\text{Cl}(2^n)$  by  $\mu_{\text{Cl}}$ , and let  $K$  be some fixed single-qubit non-Clifford gate. The circuit model we are considering (Fig. 1) interleaves Clifford unitaries drawn from  $\mu_{\text{Cl}}$ , with random gates from

$\{K, K^\dagger, \mathbb{1}\}$  acting on an arbitrary qubit.<sup>1</sup> Note that the concatenation of two unitaries drawn from measures  $\nu_1$  and  $\nu_2$  is described by the convolution  $\nu_1 * \nu_2$  of the respective measures. We thus arrive at this formal definition of the circuit model:

**Definition 2** (*K-interleaved Clifford circuits*). Let  $K \in U(2)$ . Consider the probability measure  $\xi_K$  that draws uniformly from the set  $\{K \otimes \mathbb{1}_{2^{n-1}}, K^\dagger \otimes \mathbb{1}_{2^{n-1}}, \mathbb{1}_{2^n}\}$ . A *K-interleaved Clifford circuit of depth k* is the random circuit acting on  $n$  qubits described by the probability distribution

$$\sigma_k := \underbrace{\mu_{\text{Cl}} * \xi_K * \cdots * \mu_{\text{Cl}} * \xi_K}_{k \text{ times}}. \quad (3)$$

For convenience, we work with the logarithm of base 2:  $\log(x) := \log_2(x)$ . We are now equipped to state the main result of this work in the form of a theorem:

**Theorem 1** (Unitary designs with few non-Clifford gates). *Let  $K \in U(2)$  be a non-Clifford unitary. There are constants  $C_1(K), C_2(K)$  such that for any  $k \geq C_1(K) \log^2(t)$  ( $t^4 + t \log(1/\varepsilon)$ ), a  $K$ -interleaved Clifford circuit with depth  $k$  acting on  $n$  qubits is an additive  $\varepsilon$ -approximate  $t$ -design for all  $n \geq C_2(K)t^2$ .*

We give the proofs of this theorem in Sect. 3. In Theorem 1, we consider uniformly drawn multiqubit Clifford unitaries. This can be achieved with  $O(n^3)$  classical random bits [40] and then implemented with  $O(n^2/\log(n))$  gates [45]. Combined with these results, Theorem 1 implies an overall gate count of  $O(n^2/\log(n)t^4 \log^2(t))$  improving the scaling compared to Ref. [24] in the dependence on both  $t$  and  $n$ . In this sense, our construction can be seen as a classical-quantum hybrid construction of unitary designs: The scaling is significantly improved by outsourcing as many tasks as possible to a classical computer. A construction in which all parts of the random unitary are local random circuits is considered in Corollary 2.

For designs generated from general random local circuits, numerical results suggest that convergence is much faster in practice than indicated by the proven bounds [46]. We expect that a similar effect occurs here, and that in fact very shallow  $K$ -interleaved Clifford circuits are sufficient to approximate  $t$ -designs. This intuition is supported by the numerical results of Ref. [44], which show that even a single  $T$ -gate has dramatic effects on the entanglement spectrum of a quantum circuit.

It is moreover noteworthy that circuits with few  $T$ -gates can be efficiently simulated [47–51]. The scaling of these algorithms is polynomial in the depth of the circuit, but exponential in the number of  $T$ -gates. Combined with our result, this implies that for fixed additive errors  $\varepsilon$ , there are families of  $\varepsilon$ -approximate unitary  $O(\log(n))$ -designs simulable in quasi-polynomial time. For the general random quantum circuit model, it is conjectured that a depth of order  $O(nt)$  suffices to approximate  $t$ -designs [24, 52]. If such a linear scaling is sufficient in our model, the quasi-polynomial time estimate for classical simulations would improve to polynomial.

For the proof of Theorem 1 we need to analyse the connection between the  $t$ -th moment operator of the Haar measure and the commutant of the diagonal action of the Clifford group. The latter was proven to be spanned by representations of so-called *stochastic*

<sup>1</sup> We use the set  $\{K, K^\dagger, \mathbb{1}\}$  instead of just  $\{K\}$  for technical reasons: Making the set closed under the adjoint causes the moment operator to be Hermitian. The identity is included to ensure that the concatenation of two random elements has a non-vanishing probability of producing a non-Clifford gate—a property that will slightly simplify the proof. Of course, in a physical realization, identity gates and the following Clifford operation are redundant and need not be implemented.

*Lagrangian sub-spaces* in Ref. [42]. In particular, we prove almost tight bounds on the overlap of the Haar operator with these basis vectors in Lemma 13 that might be of independent interest. This will allow us to invoke a powerful theorem by Varjú [53] on restricted spectral gaps of probability distributions on compact Lie groups to show that non-Clifford unitaries have a strong impact on representations of Lagrangian sub-spaces that are not also permutations. We combine this insight with a careful combinatorial argument about the Gram-Schmidt orthogonalization of the basis corresponding to stochastic Lagrangian sub-spaces to bound the difference to a unitary  $t$ -design in diamond norm.

Moreover, the bound for Theorem 1 allows us to prove a corollary about the stronger notion of *relative approximate designs*:

**Definition 3** (*Relative  $\varepsilon$ -approximate  $t$ -design*). We call a probability  $\nu$  a relative  $\varepsilon$ -approximate  $t$ -design if

$$(1 - \varepsilon)M_t(\nu) \preceq M_t(\mu_H) \preceq (1 + \varepsilon)M_t(\nu), \quad (4)$$

where  $A \preceq B$  if and only if  $B - A$  is completely positive.

**Corollary 1** ( *$K$ -interleaved Clifford circuits as relative  $\varepsilon$ -approximate  $t$ -designs*). There are constants  $C'_1(K), C'_2(K)$  such that a  $K$ -interleaved Clifford circuit is a relative  $\varepsilon$ -approximate  $t$ -design in depth  $k \geq C'_1(K) \log^2(t)(2nt + \log(1/\varepsilon))$  for all  $n \geq C'_2(K)t^2$ .

Hence, if we drop the system-size independence, we can achieve a scaling of  $O(nt)$  at least until  $t \sim \sqrt{n}$ .

While we believe the setting of  $K$ -interleaved Clifford circuits to be the more relevant case, the same method of proof works for *Haar*-interleaved Clifford circuits. Here, we draw not from the gate set  $\{K_i, K_i^\dagger, \mathbb{1}\}$ , but instead Haar-randomly from  $U(2)$ . The advantage is that we obtain explicit constants for the depth, while the depth in the  $K$ -interleaved setting has to depend on a constant (as  $K$  might be arbitrarily close to the identity).

**Proposition 1** (*Haar-interleaved Clifford circuits as additive  $\varepsilon$ -approximate  $t$ -designs*). For  $k \geq 36(33t^4 + 3t \log(1/\varepsilon))$ , *Haar-interleaved Clifford circuits with depth  $k$  form an additive  $\varepsilon$ -approximate  $t$ -design for all  $n \geq 32t^2 + 7$ .*

Similarly, variants of Corollary 1 for Haar-interleaved Clifford circuits can be obtained, here also without the  $\log^2(t)$  dependence. Finally, we discuss an application to higher Rényi entropies in “Appendix D”.

*1.2. Local random Clifford circuits for Clifford and unitary designs.* The circuits considered in the previous section require one to find the gate decomposition of a random Clifford operation. In this section, we analyze the case where the Clifford layers are circuits consisting of gates drawn from a local set of generators.

As a first step, we establish that a 2-local random Clifford circuit on  $n$  qubits of depth  $O(n^2 t^9 \log^{-2}(t) \log(1/\varepsilon))$  constitutes a relative  $\varepsilon$ -approximate Clifford  $t$ -design, i.e., reproduces the moment operator of the Clifford group up to the  $t$ -th order with a relative error of  $\varepsilon$ . We consider local random Clifford circuits that consist of 2-local quantum gates from a finite set  $G$  with is closed under taking the inverse and generates  $\text{Cl}(4)$ . We refer to such a set as a *closed, generating set*. A canonical example for such a closed, generating set is  $\{H \otimes \mathbb{1}, S \otimes \mathbb{1}, S^3 \otimes \mathbb{1}, \text{CX}\}$  where  $H$  is the Hadamard gate,  $S$  is the phase gate and CX is the cNOT-gate [54]. Such a set  $G$  induces a set of multi-qubit

Clifford unitaries  $\hat{G} \subset \text{Cl}(n)$  by acting on any pair of adjacent qubits on a line, where we adopt periodic boundary conditions. We then define the corresponding random Clifford circuits.

**Definition 4** (*Local random Clifford circuit*). Let  $G \subset \text{Cl}(4)$  be a closed, generating set containing the identity. Define the probability measure  $\sigma_G$  as the measure having uniform support on  $\hat{G} \subset \text{Cl}(n)$  acting on  $n$  qubits. A *local random Clifford circuit* of depth  $m$  is the random circuits described by the probability measure  $\sigma_G^{*m}$ .

For technical reasons, we again assume that the identity is part of the generating set. This assumption can be avoided but simplifies the argumentation in the following. As for the Definition 2 of  $K$ -interleaved Clifford circuits before, any upper bound on the depth of local random Clifford circuits with identity is a bound for those without.

Our result on local random Clifford circuits even holds for a stronger notion for approximations of designs, namely relative approximate designs. Write  $A \preceq B$  if  $B - A$  is positive semi-definite.

**Definition 5** (*Relative approximate Clifford  $t$ -designs*). Let  $\nu$  be a probability measure on  $\text{Cl}(2^n)$ . Then,  $\nu$  is a *relative  $\varepsilon$ -approximate Clifford  $t$ -design* if

$$(1 - \varepsilon)M_t(\mu_{\text{Cl}}) \preceq M_t(\nu) \preceq (1 + \varepsilon)M_t(\mu_{\text{Cl}}). \quad (5)$$

With this definition, our result reads as follows.

**Theorem 2** (*Local random Clifford designs*). Let  $n \geq 12t$ , then a local random Clifford circuit of depth  $O(n \log^{-2}(t)t^8(2nt + \log(1/\varepsilon)))$  constitutes a relative  $\varepsilon$ -approximate Clifford  $t$ -design.

The proof of the theorem is given in Sect. 4. This result is a significant improvement over the scaling of  $O(n^8)$ , which is implicit in Ref. [9].

We can combine this result with the bounds obtained in Sect. 3. To this end, consider a random circuit that  $k$ -times alternatingly applies a local random Clifford circuit of depth  $m$ , and a unitary drawn from the probability measure  $\xi_K$ . The corresponding probability measure is

$$\sigma_{k,m} := \underbrace{\sigma_G^{*m} * \xi_K * \dots * \sigma_G^{*m} * \xi_K}_{k \text{ times}}. \quad (6)$$

For these local random circuits we establish the following result:

**Corollary 2** (*Local random unitary design*). Let  $K \in U(2)$  be a non-Clifford gate and let  $G \subset \text{Cl}(4)$  be a closed, generating set. There are constants  $C_1''(K, G)$ ,  $C_2''(K)$ ,  $C_3''(K)$  such that whenever

$m \geq C_1''(K, G)n \log^{-2}(t)t^8(2nt + \log(1/\varepsilon))$  and  $k \geq C_2''(K) \log^2(t)(t^4 + t \log(1/\varepsilon))$ , the local random circuit  $\sigma_{k,m}$ , defined in (6), is an  $\varepsilon$ -approximate unitary  $t$ -design for all  $n \geq C_3''(K)t^2$ .

The complete argument for the corollary is given at the end of Sect. 4. After introducing technical preliminaries in Sect. 2, the remainder of the paper, Sect. 3 and Sect. 4, is devoted to the proofs of Theorem 1, Theorem 2 and the Corollary 2. Finally, in Sect. 5 we elaborate on and formalize as Proposition 3 the observation that there exists no non-universal gate set generating exact 4-designs for arbitrary system size. This observation is an immediate consequence of the classification of finite unitary  $t$ -groups and a criterion for the universality of finite gate sets [38, 39, 55].

## 2. Technical Preliminaries

*2.1. Operators and superoperators.* Given a (finite-dimensional) Hilbert space  $\mathcal{H}$ , we denote with  $L(\mathcal{H})$  the space of linear operators on  $\mathcal{H}$  with involution  $\dagger$  mapping an operator to its adjoint with respect to the inner product on  $\mathcal{H}$ .  $L(\mathcal{H})$  naturally inherits a Hermitian inner product, the *Hilbert-Schmidt inner product*

$$(A|B) := \text{Tr}(A^\dagger B), \quad \forall A, B \in L(\mathcal{H}). \quad (7)$$

As this definition already suggests, we will use “operator kets and bras” whenever we think it simplifies the notation. Concretely, we write  $|B) = B$  and denote with  $(A|$  the linear form on  $L(\mathcal{H})$  given by

$$(A| : B \longmapsto (A|B). \quad (8)$$

Following common terminology in quantum information theory, we call linear maps  $\phi : L(\mathcal{H}) \rightarrow L(\mathcal{H})$  on operators “superoperators”. We use  $\phi^\dagger$  to denote the adjoint map with respect to the Hilbert-Schmidt inner product. Note that with the above notation,  $\phi = |A)(B|$  defines a rank one superoperator with  $\phi^\dagger = |B)(A|$ . Moreover, we will denote by the superoperator  $\text{Ad}_A := A \cdot A^{-1}$  the *adjoint action* of an invertible operator  $A \in \text{GL}(\mathcal{H})$  on  $L(\mathcal{H})$ . For notational reasons, we sometimes write  $\text{Ad}(A)$  instead of  $\text{Ad}_A$ .

We consistently reserve the notation  $\|\cdot\|_p$  for the Schatten  $p$ -norms

$$\|A\|_p := \text{Tr}(|A|^p)^{1/p} = \|\sigma(A)\|_{\ell_p}, \quad (9)$$

where  $\sigma(A)$  is the vector of singular values of  $A$ . In particular, we use the *trace norm*  $p = 1$ , the *Frobenius* or *Hilbert-Schmidt norm*  $p = 2$  and the *spectral norm*  $p = \infty$ . Clearly, these norms can be defined for both operators and superoperators and we will use the same symbol in both cases. For the latter, however, there is also a family of induced operator norms

$$\|\phi\|_{p \rightarrow q} := \sup_{\|X\|_p \leq 1} \|\phi(X)\|_q. \quad (10)$$

Note that  $\|\cdot\|_{2 \rightarrow 2} \equiv \|\cdot\|_\infty$ . Finally, we are interested in “stabilized” versions of these induced norms, in particular the *diamond norm*

$$\|\phi\|_\diamond := \sup_{d \in \mathbb{N}} \|\phi \otimes \text{id}_{L(\mathbb{C}^d)}\|_{1 \rightarrow 1} = \|\phi \otimes \text{id}_{L(\mathcal{H})}\|_{1 \rightarrow 1}. \quad (11)$$

The following norm inequality will be useful [56]

$$\|\phi\|_\diamond \leq (\dim \mathcal{H})^2 \|\phi\|_\infty, \quad \|\phi\|_\infty \leq \sqrt{\dim \mathcal{H}} \|\phi\|_\diamond. \quad (12)$$



**2.2. Commutant of the diagonal representation of the Clifford group.** In this section, we review some of the machinery developed in Ref. [42]. Recall that the  $n$ -qubit *Clifford group*  $\text{Cl}(n)$  is defined as the unitary normalizer of the Pauli group  $\mathcal{P}_n$  as

$$\text{Cl}(n) = \left\{ U \in U(2^n, \mathbb{Q}[i]) \mid U\mathcal{P}_n U^\dagger \subset \mathcal{P}_n \right\}. \quad (13)$$

Here, we followed the convention to restrict the matrix entries to rational complex numbers. This avoids the unnecessary complications from an infinite center  $U(1)$  yielding a finite group with minimal center  $Z(\text{Cl}(n)) = Z(\mathcal{P}_n) \simeq \mathbb{Z}_4$ . The Clifford group can equivalently be defined in a less conceptual but more constructive manner: It is the subgroup of  $U(2^n)$  generated by CX, the controlled not gate, the Hadamard gate  $H$  and the phase gate  $S$ .

For this work, the  $t$ -th diagonal representation of the Clifford group, defined as

$$\tau^{(t)} : \text{Cl}(n) \longrightarrow U(2^{nt}), \quad U \longmapsto U^{\otimes t}, \quad (14)$$

will be of major importance. It acts naturally on the Hilbert space  $((\mathbb{C}^2)^{\otimes n})^{\otimes t}$  which can be seen as  $t$  copies of an  $n$ -qubit system. However, it will turn out that the operators commuting with this representation naturally factorize with respect to a different tensor structure on this Hilbert space, namely  $((\mathbb{C}^2)^{\otimes t})^{\otimes n} \simeq ((\mathbb{C}^2)^{\otimes n})^{\otimes t}$ . Because of the different exponents, it should be clear from the context which tensor structure is meant. We will make ubiquitous use of the description of the commutant of the diagonal representation in terms of *stochastic Lagrangian sub-spaces* [42]:

**Definition 6** (*Stochastic Lagrangian sub-spaces*). Consider the quadratic form  $q : \mathbb{Z}_2^{2t} \rightarrow \mathbb{Z}_4$  defined as  $q(x, y) := x \cdot x - y \cdot y \pmod{4}$ . The set  $\Sigma_{t,t}$  denotes the set of all sub-spaces  $T \subseteq \mathbb{Z}_2^{2t}$  being subject to the following properties:

1.  $T$  is totally  $q$ -isotropic:  $x \cdot x = y \cdot y \pmod{4}$  for all  $(x, y) \in T$ .
2.  $T$  has dimension  $t$  (the maximum dimension compatible with total isotropicity).
3.  $T$  is *stochastic*:  $(1, \dots, 1) \in T$ .

We call elements in  $\Sigma_{t,t}$  *stochastic Lagrangian sub-spaces*. We have

$$|\Sigma_{t,t}| = \prod_{k=0}^{t-2} (2^k + 1) \leq 2^{\frac{1}{2}(t^2+5t)}. \quad (15)$$

With this notion, we can now state the following key theorem from Ref. [42].

**Theorem 3** ([42]). *If  $n \geq t - 1$ , then the commutant  $\tau^{(t)}(\text{Cl}(n))'$  of the  $t$ -th diagonal representation of the Clifford group is spanned by the linearly independent operators  $r(T)^{\otimes n}$ , where  $T \in \Sigma_{t,t}$  and*

$$r(T) := \sum_{(x,y) \in T} |x\rangle\langle y|. \quad (16)$$

Since the representation in question is fixed throughout this paper, we will simplify the notation from now on and write  $\text{Cl}(n)' \equiv \tau^{(t)}(\text{Cl}(n))'$ . To make use of a more sophisticated characterization of the elements  $r(T)$  developed in Ref. [42, Section 4], we need the following definitions.

**Definition 7** (*Stochastic orthogonal group*). Consider the quadratic form  $q : \mathbb{Z}_2^t \rightarrow \mathbb{Z}_4$  defined as  $q(x) := x \cdot x \pmod{4}$ . The *stochastic orthogonal group*  $O_t$  is defined as the group of  $t \times t$  matrices  $O$  with entries in  $\mathbb{Z}_2$  such that  $q(Ox) = q(x)$  for all  $x \in \mathbb{Z}_2^t$ .

The subspace  $T_O := \{(Ox, x), x \in \mathbb{Z}_2^t\}$  is a stochastic Lagrangian subspace. Moreover, the operator  $r(O) := r(T_O)$  is unitary. We will therefore canonically embed the orthogonal stochastic group  $O_t \subset \Sigma_{t,t}$ . Notice that the permutation group on  $t$  objects, referred to as  $S_t$ , may be embedded into  $O_t$  by acting on the standard basis of  $\mathbb{Z}_2^t$ . Together with  $O_t$ , the following definition can be used to fully characterize the set of stochastic Lagrangian sub-spaces,  $\Sigma_{t,t}$ .

**Definition 8** (*Defect sub-spaces*). A defect subspace is a subspace  $N \subseteq \mathbb{Z}_2^t$  which is isotropic with respect to  $q$ , that is, that  $q(x) = 0$  for all  $x \in N$ .

The quadratic form  $q$  is what is known as a *generalized quadratic refinement* of the bi-linear form defined by the inner product  $(x, y) \mapsto x \cdot y \pmod{2}$  (see, e.g., Ref. [57, App. A] for a self-contained discussion). In the following, the ortho-complement  $N^\perp$  of a subspace  $N \subseteq \mathbb{Z}_2^t$  is taken with respect to the inner product modulo 2,

$$N^\perp = \{v \in \mathbb{Z}_2^t \mid v \cdot u = 0 \pmod{2}, \forall u \in N\}.$$

Notice that  $q(x) = 0$  implies that  $x \cdot \mathbf{1}_t = 0 \pmod{2}$ , where  $\mathbf{1}_t := (1, \dots, 1)^T$  is the all-ones vector. Thus, we do not need a separate clause requiring  $\mathbf{1}_t \in N^\perp$  in the definition of defect sub-spaces (compare Ref. [42, Def. 4.16]). Moreover, one may verify that  $2q(x) = 2x \cdot \mathbf{1}_t \pmod{4}$ . This implies, similarly, that if  $O$  preserves  $q$ , then  $O\mathbf{1}_t = \mathbf{1}_t$ . Borrowing the language of [42], all  $q$ -isometries are stochastic (compare the definition of the orthogonal stochastic group in that reference, [42, Def. 4.11]). The reason for these simplifications is that here we focus on the qubit case exclusively, while Ref. [42] works simultaneously for qubits and odd qudits. We use the names *stochastic orthogonal group* and *defect subspace* (rather than simply *q-isometry group* and *isotropic subspace*) to keep with the notation of that reference.

For any defect subspace  $N$ , it holds that  $N \subseteq N^\perp$  (and thus  $\dim N \leq t/2$ ). Because of this, defect sub-spaces  $N \subseteq \mathbb{Z}_2^t$  define *Calderbank-Shor-Sloane (CSS) codes*

$$\text{CSS}(N) := \{Z(p)X(q) \mid q, p \in N\}, \quad (17)$$

where the action of the multi-qubit Pauli operators is  $Z(p) |x\rangle := (-1)^{p \cdot x} |x\rangle$  and  $X(q) |x\rangle := |x + q\rangle$  for  $x \in \mathbb{Z}_2^t$ . The corresponding projector is given by

$$P_N := P_{\text{CSS}(N)} = \frac{1}{|N|^2} \sum_{q,p \in N} Z(p)X(q). \quad (18)$$

Since the order of the stabilizer group is  $2^{2 \dim N}$ ,  $P_N$  projects onto a  $2^{t-2 \dim N}$ -dimensional subspace of  $(\mathbb{C}^2)^{\otimes t}$ . For  $N = \{0\}$  we set  $P_{\text{CSS}(N)} := \mathbb{1}$ . We summarize the findings of Ref. [42, Sect. 4] in Thm. 4. We give a short proof to give an explicit relation between this theorem and the results of that work.

**Theorem 4** ([42]). Consider  $T \in \Sigma_{t,t}$ , then

$$r(T) = 2^{\dim N} r(O) P_{\text{CSS}(N)} = 2^{\dim N'} P_{\text{CSS}(N')} r(O') \quad (19)$$

for  $O, O' \in O_t$  and  $N, N'$  are unique defect sub-spaces with  $\dim N = \dim N'$ .

*Proof.* Recall from Ref. [42] that the code space range  $P_{\text{CSS}(N)}$  has an orthonormal basis of coset state vectors given by

$$\left\{ |N, [x]\rangle := \frac{1}{\sqrt{N}} \sum_{y \in N} |x + y\rangle \mid x \in N^\perp, [x] \in N^\perp/N \right\}.$$

One may compute that  $r(O) |N, [x]\rangle = |ON, [Ox]\rangle$ . This way,

$$r(O)P_{\text{CSS}(N)} = \sum_{[x] \in N^\perp/N} |ON, [Ox]\rangle \langle N, [x]|.$$

Comparing this equation to [42, Lem. 4.23] we see that the set  $\{2^{\dim N} r(O)P_{\text{CSS}(N)}\}_O$  is equal to the set of  $r(T)$  operators with right defect subspace given by  $N$ , i.e., with  $T_{RD} = N$  in the notation of that reference. This way, varying over  $N$  we obtain the full set  $\Sigma_{t,t}$ . The existence of a decomposition  $2^{\dim N} P_{\text{CSS}(N')} r(O')$  follows from the above by noting that  $r(O)P_{\text{CSS}(N)}r(O)^\dagger = P_{\text{CSS}(ON)}$ .  $\square$

**Lemma 1** (Norms of  $r(T)$ ). *Suppose  $r(T) = 2^{\dim N} r(O)P_N$  as in Theorem 4. Then it holds:*

$$\|r(T)\|_1 = 2^{t-\dim N}, \quad \|r(T)\|_2 = 2^{t/2}, \quad \|r(T)\|_\infty = 2^{\dim N}. \quad (20)$$

*Proof.* Since any Schatten  $p$ -norm is unitarily invariant, we have  $\|r(T)\|_p = 2^{\dim N} \|P_N\|_p$ . The statements follow from  $\text{rank } P_N = 2^{t-2\dim N}$ .  $\square$

In the following, we will often work with a normalized version of the  $r(T)$  operators which we define as

$$Q_T := \frac{r(T)}{\|r(T)\|_2} = 2^{-t/2} r(T). \quad (21)$$

### 3. Approximate Unitary $t$ -Designs

In this section, we give a bound on the number of non-Clifford gates needed to leverage the Clifford group to an approximate unitary  $t$ -design. This is made precise by the following two theorems which rely on two distinct proof strategies and come with different trade-offs.

**Theorem 1** (Unitary designs with few non-Clifford gates). *Let  $K \in U(2)$  be a non-Clifford unitary. There are constants  $C_1(K), C_2(K)$  such that for any  $k \geq C_1(K) \log^2(t)$  ( $t^4 + t \log(1/\varepsilon)$ ), a  $K$ -interleaved Clifford circuit with depth  $k$  acting on  $n$  qubits is an additive  $\varepsilon$ -approximate  $t$ -design for all  $n \geq C_2(K)t^2$ .*

Recall from Def. 2 that a  $K$ -interleaved Clifford circuit has an associated probability measure  $\sigma_K := (\mu_{\text{Cl}} * \xi_K)^{*k}$  where  $\xi_K$  is the measure which draws uniformly from  $\{K, K^\dagger, \mathbb{1}\}$  on the first qubit. Let us introduce the notation

$$\mathbf{R}(K) := \int_{U(2^n)} \text{Ad}_U^{\otimes t} d\xi_k(U) = \frac{1}{3} \left( \text{Ad}_K^{\otimes t} + \text{Ad}_{K^\dagger}^{\otimes t} + \text{id} \right) \otimes \text{id}_{n-1}. \quad (22)$$

Then, our goal is to bound the deviation of the moment operator

$$M_t(\sigma_k) = \int_{U(2^n)} \text{Ad}_U^{\otimes t} d\sigma_k(U) = \underbrace{M_t(\mu_{\text{Cl}})R(K) \dots M_t(\mu_{\text{Cl}})R(K)}_{k \text{ times}}, \quad (23)$$

from the Haar projector  $P_H \equiv M_t(\mu_H)$  in diamond norm. Using that  $P_H$  is invariant under left and right multiplication with unitaries, we have the identity

$$A^k - P_H = (A - P_H)^k, \quad (24)$$

for any mixed unitary channel  $A$ . Thus, we can rewrite the difference of moment operators as

$$M_t(\sigma_k) - P_H = [P_{\text{Cl}}R(K)]^k - P_H = [(P_{\text{Cl}} - P_H)R(K)]^k, \quad (25)$$

where we have introduced the shorthand notation  $P_{\text{Cl}} := M_t(\mu_{\text{Cl}})$ .

*Remark 1 (Non-vanishing probability of applying the identity).* We apply  $K, K^\dagger$  with equal probability in Theorem 1 such that  $R(K)$  is Hermitian. The non-vanishing probability of applying  $\mathbb{1}$ , i.e., of doing nothing, is necessary in the proof of Lemma 2, because we require the probability distribution  $\xi_K * \xi_K$  to have non-vanishing support on a non-Clifford gate. If  $\xi_K$  is the uniform measure on  $K$  and  $K^\dagger$ , then  $\xi_K * \xi_K$  has support on  $K^2, (K^\dagger)^2$  and  $\mathbb{1}$ . We can hence drop this assumption for gates that do not square to a Clifford gate. This is not the case for e.g. the  $T$ -gate.

Our proof strategy for Theorem 1 makes use of the following two lemmas which are proven in Sects. 6.1 and 6.2. The first lemma is key to the derivations in this section. It is based on a bound (Lemma 13) on the overlap of stochastic Lagrangian sub-spaces with the Haar projector and Theorem 5, a special case of a theorem about restricted spectral gaps of random walks on compact Lie groups due to Varjú [53].

**Lemma 2** (Overlap bound). *Let  $K$  be a single qubit gate which is not contained in the Clifford group. Then, there is a constant  $c(K) > 0$  such that*

$$\eta_{K,t} := \max_{\substack{T \in \Sigma_{t,t} - S_t \\ T' \in \Sigma_{t,t}}} \frac{1}{3} \left| (Q_T | \text{Ad}_K^{\otimes t} + \text{Ad}_{K^\dagger}^{\otimes t} + \text{id} | Q_{T'}) \right| \leq 1 - c(K) \log^{-2}(t). \quad (26)$$

The second lemma is of a more technical nature.

**Lemma 3** (Diamond norm bound). *Consider  $T_1, T_2 \in \Sigma_{t,t}$  and denote with  $N_1, N_2$  their respective defect spaces. Then, it holds that*

$$\| |Q_{T_1}\rangle\langle Q_{T_2}| \|_\diamond \leq 2^{\dim N_2 - \dim N_1}, \quad (27)$$

$$| \langle Q_{T_1} | Q_{T_2} \rangle | \leq 2^{-|\dim N_1 - \dim N_2|}. \quad (28)$$

The difficulty of using these results to bound the difference

$$M_t(\sigma_k) - P_H = [(P_{\text{Cl}} - P_H)R(K)]^k, \quad (29)$$

stems from the following reason: The range of the projector  $P_{\text{Cl}} - P_H$  is the ortho-complement of the space spanned by permutations  $Q_\pi^{\otimes n}$  for  $\pi \in S_t$  within the commutant of the Clifford group spanned by the operators  $Q_T^{\otimes n}$ . Although this is a conveniently

factorizing and well-studied basis, it is *non-orthogonal*. Thus, the projectors do not possess a natural expansion in this basis and we can not directly use the above bounds. However, we can write it explicitly in a suitable orthonormal basis of the commutant obtained by the Gram-Schmidt procedure from the basis  $\{Q_T^{\otimes n} \mid T \in \Sigma_{t,t}\}$ . We summarize the properties of this basis in the following lemma.

**Lemma 4** (Properties of the constructed basis). *Let  $\{T_j\}_{j=1}^{|\Sigma_{t,t}|}$  be an enumeration of the elements of  $\Sigma_{t,t}$  such that the first  $t!$  spaces  $T_j$  correspond to the elements of  $S_t$ . Then, the  $\{E_j\}$  constitutes an orthogonal (but not normalized) basis, where*

$$E_j := \sum_{i=1}^j A_{i,j} Q_{T_i}^{\otimes n} := \sum_{i=1}^j \left[ \sum_{\substack{\Pi \in S_j \\ \Pi(j)=i}} \text{sign}(\Pi) \prod_{l=1}^{j-1} (Q_{T_l} | Q_{T_{\Pi(l)}})^n \right] Q_{T_i}^{\otimes n}. \tag{30}$$

Denote by  $N_i$  the defect space of  $T_i$ . For  $n \geq \frac{1}{2}(t^2 + 5t)$ , we have

$$|A_{i,j}| \leq 2^{t^3+4t^2+6t-n} |\dim N_i - \dim N_j|, \quad \forall i, j, \tag{31}$$

$$|A_{i,j}| \leq 2^{2t^2+10t-n}, \quad \forall i \neq j. \tag{32}$$

Moreover, it holds that

$$1 - 2^{t^2+7t-n} \leq A_{j,j} \leq 1 + 2^{t^2+7t-n}. \tag{33}$$

We believe that the explicit bounds in Lemma 4 might be of independent interest in applications of the Schur-Weyl duality of the Clifford group. For the sake of readability, and as Theorem 1 holds up to an inexplicit constant, we will bound all polynomials in  $t$  by their leading order term in the following. Specifically, the bounds in Lemma 4 will be simplified by using the inequalities

$$t^3 + 4t^2 + 6t \leq 11t^3, \tag{34}$$

$$2t^2 + 10t \leq 12t^2 \leq 12t^3, \tag{35}$$

$$t^2 + 7t \leq 8t^2 \leq 8t^3 \tag{36}$$

which hold for all positive integers  $t$ .

*Proof of Theorem 1.* Notice that from (25), we have the expression

$$\|[P_{\text{Cl}}R(K)]^k - P_{\text{H}}\|_{\diamond} \tag{37}$$

$$= \left\| \left[ \left( \sum_{j=t!+1}^{|\Sigma_{t,t}|} \frac{1}{(E_j | E_j)} |E_j\rangle\langle E_j| \right) R(K) \right]^k \right\|_{\diamond} \tag{38}$$

$$= \left\| \sum_{j_1, \dots, j_m=t!+1}^{|\Sigma_{t,t}|} \prod_{l=1}^k \frac{1}{(E_{j_l} | E_{j_l})} |E_{j_1}\rangle\langle E_{j_1}| R(K) |E_{j_2}\rangle\langle E_{j_2}| \dots |E_{j_k}\rangle\langle E_{j_k}| R(K) \right\|_{\diamond} \tag{39}$$

$$\leq \sum_{j_1, \dots, j_k=t!+1}^{|\Sigma_{t,t}|} \prod_{l=1}^k \frac{1}{(E_{j_l} | E_{j_l})} \prod_{r=1}^{k-1} \left\| |E_{j_r}\rangle\langle E_{j_{r+1}}| \cdot \right\| \left\| |E_{j_1}\rangle\langle E_{j_k}| \right\|_{\diamond}. \tag{40}$$

We now bound each of the factors in each term above. First, we compute the squared norm of  $|E_j\rangle$ ,

$$(E_j|E_j) = \sum_{r,l=1}^j A_{r,j} A_{l,j} (Q_{T_r}|Q_{T_l})^n = A_{j,j}^2 + \sum_{k,l < j} A_{r,j} A_{l,j} (Q_{T_k}|Q_{T_l})^n. \quad (41)$$

Using Eqs. (32) and (33), we thus bound

$$\begin{aligned} (E_j|E_j) &\leq \left(1 + 2^{t^2+7t-n}\right)^2 + (j^2 - 1)4^{2t^2+10t-n} \\ &\leq \left(1 + 2^{t^2+7t-n}\right)^2 + |\Sigma_{t,t}|^2 4^{2t^2+10t-n} \\ &\leq 1 + 2^{31t^2-2n}, \end{aligned} \quad (42)$$

and in the same way

$$(E_j|E_j) \geq 1 - 2^{31t^2-2n}. \quad (43)$$

Now we use that  $n \geq 16t^2$ . Letting  $x := 2^{31t^2-2n} \in [0, \frac{1}{2}]$ , the inequalities  $1/(1-x) \leq 1+2x$  and  $1-2x \leq 1/(1+x)$  hold. This leads to

$$\frac{1}{(E_j|E_j)} = 1 + a_j \quad \text{with} \quad |a_j| \leq 2^{32t^2-2n}. \quad (44)$$

We now focus on the second factor,

$$|(E_i|R(K)|E_j)| \leq \sum_{r=1}^i \sum_{l=1}^j |A_{r,i} A_{l,j}| \cdot \left| (Q_{T_r}^{\otimes n} | R(K) | Q_{T_l}^{\otimes n}) \right|. \quad (45)$$

If for  $(Q_{T_r}|R(K)|Q_{T_l})$  one of the stochastic Lagrangian sub-spaces does not correspond to a permutation, Lemma 2 introduces a factor of  $\eta_{K,t}$ . If both correspond to a permutation, we redefine the factors in a way that leads to simpler expressions in the calculations used below. Namely, in this case we redefine  $A_{r,i}$  and  $A_{l,j}$  by multiplying it with 2. This is compensated by introducing a factor of  $\frac{1}{4}$  and letting

$$\bar{\eta}_{K,t} := \max \left\{ \frac{1}{4}, \eta_{K,t} \right\}. \quad (46)$$

We can do this as  $i$  and  $j$  do not correspond to permutations and hence  $A_{r,j}$  and  $A_{l,i}$  are exponentially suppressed, which remains true after rescaling by 2. In this case, moreover,  $r < t!+1 \leq i$  and  $l < t!+1 \leq j$ , so the factor  $|A_{r,i} A_{l,j}|$  will be exponentially suppressed according to (32) and so this redefinition will not affect the asymptotic scaling in  $n$ .

We provide two bounds for  $|(E_i|R(K)|E_j)|$  that will be used later on. We will use repeatedly that the diamond norm is multiplicative under the tensor product of superoperators [58, Thm. 3.49]. First, using (31), (33) and (28), we obtain

$$|(E_i | R(K) | E_j)| \leq \sum_{r=1}^i \sum_{l=1}^j |A_{r,i} A_{l,j}| \cdot \left| \left( Q_{T_r}^{\otimes n} \middle| R(K) \middle| Q_{T_l}^{\otimes n} \right) \right| \quad (47)$$

$$\leq \bar{\eta}_{K,t} (1 + 2^{8t^2-n}) \sum_{r=1}^i \sum_{l=1}^j 2^{24t^3-n|\dim N_r - \dim N_i| - n|\dim N_l - \dim N_j| - (n-1)|\dim N_l - \dim N_r|} \quad (48)$$

$$\leq \bar{\eta}_{K,t} (1 + 2^{8t^2-n}) |\Sigma_{t,t}|^2 2^{25t^3-n|\dim N_j - \dim N_i|} \quad (49)$$

$$\leq \bar{\eta}_{K,t} (1 + 2^{8t^2-n}) 2^{31t^3-n|\dim N_j - \dim N_i|}, \quad (50)$$

where we have used  $2^{|\dim N_l - \dim N_r|} \leq 2^t \leq 2^{t^3}$ , and the fact that for the rescaled  $A_{r,i}$ , the inequality (31) implies

$$A_{r,i} \leq 2^{11t^3 - |\dim N_r - \dim N_j| + 1} \leq 2^{12t^3 - |\dim N_r - \dim N_j|}$$

for all  $r, i$ . Moreover, we have used the triangle inequality,

$$\begin{aligned} & |\dim N_r - \dim N_i| + |-\dim N_l + \dim N_j| + |\dim N_l - \dim N_r| \\ & \geq |\dim N_r - \dim N_i - \dim N_l + \dim N_j + \dim N_l - \dim N_r| \\ & = |\dim N_j - \dim N_i|, \end{aligned} \quad (51)$$

in the inequality (49).

The second bound follows from Eqs. (32) and (33), and we consider two cases. If  $i \neq j$ , then

$$\begin{aligned} |(E_i | R(K) | E_j)| & \leq \sum_{r=1}^i \sum_{l=1}^j |A_{r,i} A_{l,j}| \cdot \left| \left( Q_{T_r}^{\otimes n} \middle| R(K) \middle| Q_{T_l}^{\otimes n} \right) \right| \\ & \leq \bar{\eta}_{K,t} (1 + 2^{8t^2-n}) |\Sigma_{t,t}|^2 2^{19t^2-n} \\ & \leq \bar{\eta}_{K,t} (1 + 2^{8t^2-n}) 2^{25t^2-n}. \end{aligned} \quad (52)$$

Otherwise,

$$|(E_i | R(K) | E_i)| \leq \sum_{r=1}^i \sum_{l=1}^i |A_{r,i} A_{l,i}| \cdot \left| \left( Q_{T_r}^{\otimes n} \middle| R(K) \middle| Q_{T_l}^{\otimes n} \right) \right| \quad (53)$$

$$\leq \bar{\eta}_{K,t} \left( |A_{i,i}|^2 + (i^2 - 1) 2^{12t^2-n} \right) \quad (54)$$

$$\leq \bar{\eta}_{K,t} \left( (1 + 2^{8t^2-n})^2 + (1 + 2^{8t^2-n}) 2^{16t^2-n} \right) \quad (55)$$

$$\leq \bar{\eta}_{K,t} (1 + 2^{16t^2-n})^3. \quad (56)$$

In inequality (54), we have bounded the term  $r = l = i$  using (33), and each of the other terms using (32). Moreover, in the inequalities (55) and (56) we use that  $i \leq |\Sigma_{t,t}|$ , and

$$1 + 2^{8t^2-n} \leq (1 + 2^{8t^2-n})^2 \leq (1 + 2^{16t^2-n})^2.$$

Lastly, we obtain from (31) and (27)

$$\| |E_i\rangle\langle E_j| \|_{\diamond} \leq \sum_{r=1}^i \sum_{l=1}^j |A_{r,i} A_{l,j}| \cdot \left\| \left| Q_{T_r}^{\otimes n} \right\rangle \left( Q_{T_l}^{\otimes n} \right) \right\|_{\diamond} \quad (57)$$

$$\leq |\Sigma_{t,t}| 2^{24t^3-n} |\dim N_r - \dim N_i| -n |\dim N_l - \dim N_j| +n (\dim N_l - \dim N_r) \quad (58)$$

$$\leq 2^{30t^3+n(\dim N_j - \dim N_i)}. \quad (59)$$

We now start piecing these expressions together to bound (40). Eqs. (59) and (44) give

$$\begin{aligned} & \| [P_{\text{Cl}} \mathbf{R}(K)]^k - P_{\text{H}} \|_{\diamond} \\ & \leq \left( 1 + 2^{32t^2-2n} \right)^k \sum_{j_1, \dots, j_k=t+1}^{|\Sigma_{t,t}|} 2^{30t^3+n(\dim N_{j_k} - \dim N_{j_1})} \prod_{r=1}^{k-1} | (E_{j_r} | R(K) | E_{j_{r+1}}) |. \end{aligned} \quad (60)$$

To bound (60), we will bunch together the contribution of all terms whose sequence  $\{j_1, \dots, j_k\}$  contains  $l$  changes. Moreover, we will treat differently the cases  $l \leq \lfloor t/2 \rfloor$  and  $l > \lfloor t/2 \rfloor$ . In the former case, we use (50) to get

$$\prod_{r=1}^{k-1} | (E_{j_r} | R(K) | E_{j_{r+1}}) | \leq \bar{\eta}_{K,t}^{k-1} (1 + 2^{16t^2-n})^{3(k-1)} 2^{l31t^3-n|\dim N_{j_k} - \dim N_{j_1}|}. \quad (61)$$

In this case, the factor of  $2^{n(\dim N_{j_k} - \dim N_{j_1})}$  coming from (59) is cancelled by the last factor of  $2^{-n|\dim N_{j_k} - \dim N_{j_1}|}$ .

In the latter case, we turn to (52) instead to obtain

$$\prod_{r=1}^{k-1} | (E_{j_r} | R(K) | E_{j_{r+1}}) | \leq \bar{\eta}_{K,t}^{k-1} (1 + 2^{16t^2-n})^{3(k-1)} 2^{l25t^2-ln}.$$

Here, the exponential factor coming from (59) is cancelled by  $2^{-ln}$  since  $\dim N_{j_k} - \dim N_{j_1} \leq \lfloor t/2 \rfloor$ . Counting the instances of sequences with  $l$  changes, we may put these considerations together to bound

$$\begin{aligned} \| [P_{\text{Cl}} \mathbf{R}(K)]^k - P_{\text{H}} \|_{\diamond} & \leq \left( 1 + 2^{32t^2-2n} \right)^k \left( 1 + 2^{16t^2-n} \right)^{3(k-1)} \\ & \quad \bar{\eta}_{K,t}^{k-1} \left[ \sum_{l=0}^{\lfloor \frac{t}{2} \rfloor} \binom{k}{l} |\Sigma_{t,t}|^{l+1} 2^{l31t^3} \right. \\ & \quad \left. + \sum_{l=\lfloor \frac{t}{2} \rfloor+1}^k \binom{k}{l} |\Sigma_{t,t}|^{l+1} 2^{(l-\lfloor \frac{t}{2} \rfloor)(25t^2-n)} 2^{\lfloor \frac{t}{2} \rfloor 25t^2} \right] \\ & \leq \left( 1 + 2^{32t^2-2n} \right)^{4k} \bar{\eta}_{K,t}^{k-1} \left[ \frac{t}{2} \binom{k}{\lfloor \frac{t}{2} \rfloor} |\Sigma_{t,t}|^{\lfloor \frac{t}{2} \rfloor+1} 2^{\lfloor \frac{t}{2} \rfloor 31t^3} \right] \end{aligned}$$



$$\begin{aligned}
 & + \sum_{l=1}^{k-\lfloor \frac{t}{2} \rfloor} \binom{k}{l+\lfloor \frac{t}{2} \rfloor} |\Sigma_{t,t}|^{l+1+\lfloor \frac{t}{2} \rfloor} 2^{l(25t^2-n)} 2^{13t^3} \Big] \\
 & \stackrel{\ddagger}{\leq} \left(1 + 2^{32t^2-2n}\right)^{4k} \bar{\eta}_{K,t}^{k-1} \left[ 2^{32t^4+t \log(k)} \right. \\
 & \quad \left. + k^{\lfloor \frac{t}{2} \rfloor} |\Sigma_{t,t}|^{1+\lfloor \frac{t}{2} \rfloor} 2^{13t^3} \sum_{l=0}^k \binom{k}{l} |\Sigma_{t,t}|^l 2^{l(25t^2-n)} \right] \\
 & \leq \left(1 + 2^{32t^2-2n}\right)^{4k} \bar{\eta}_{K,t}^{k-1} \left[ 2^{32t^4+t \log(k)} \right. \\
 & \quad \left. + 2^{18t^3+\log(k)t} \left(1 + 2^{28t^2-n}\right)^k \right] \\
 & \leq \left(1 + 2^{32t^2-2n}\right)^{4k} \left(1 + 2^{28t^2-n}\right)^k \\
 & \quad 2^{t \log(k)} \bar{\eta}_{K,t}^{k-1} \left[ 2^{32t^4} + 2^{18t^3} \right],
 \end{aligned}$$

where we have used in  $\ddagger$  that

$$\begin{aligned}
 \binom{k}{l+\lfloor \frac{t}{2} \rfloor} & = \frac{(k)!}{(k-l-\lfloor \frac{t}{2} \rfloor)! (l+\lfloor \frac{t}{2} \rfloor)!} \\
 & \leq (k-l-\lfloor \frac{t}{2} \rfloor + 1) \dots (k-l) \frac{k!}{(k-l)!} \\
 & \leq k^{\lfloor \frac{t}{2} \rfloor} \binom{k}{l}.
 \end{aligned}$$

Finally, noting that  $2^{32t^4} + 2^{18t^3} \leq 2^{33t^4}$  for all positive integers  $t$ , we obtain the bound

$$\|M_t(\sigma_k) - P_H\|_{\diamond} \leq 2^{33t^4+t \log(k)} \left(1 + 2^{32t^2-n}\right)^{5k} \bar{\eta}_{K,t}^{k-1}, \tag{62}$$

where  $\bar{\eta}_{K,t}$  is bounded by Lemma 2. Taking the logarithm and using the inequality  $\log(1+x) \leq x$  repeatedly, this implies Theorem 1.  $\square$

With the above bound, we can also prove Corollary 1.

*Proof of Corollary 1.* Consider the self-adjoint superoperator  $A := P_{\text{CI}}R(K)P_{\text{CI}}$ . As  $P_{\text{CI}}$  is a projector, we have with Eq. (24)

$$(A - P_H)^k = A^k - P_H = [P_{\text{CI}}R(K)]^k - P_H = M_t(\sigma_k) - P_H. \tag{63}$$

Using norm inequality between operator and diamond norm Eq. (12) and the previous result Eq. (62), we find

$$\begin{aligned}
 \|A - P_H\|_{\infty}^k & = \|(A - P_H)^k\|_{\infty} \leq 2^{nt/2} \|M_t(\sigma_k) - P_H\|_{\diamond} \\
 & \leq 2^{33t^4+t \log(k)+nt/2} \left(1 + 2^{32t^2-n}\right)^{5k} \bar{\eta}_{K,t}^{k-1}. \tag{64}
 \end{aligned}$$

Taking the  $k$ -th square root of the expression above, we obtain a sequence of infinitely many bounds for  $\|A - P_H\|_\infty$  which converges as  $k \rightarrow \infty$ . That limit gives

$$\|A - P_H\|_\infty \leq \left(1 + 2^{32t^2 - n}\right)^5 \bar{\eta}_{K,t}. \quad (65)$$

Combined with Ref. [24, Lem. 4], Eq. (65) implies the result.  $\square$

The bound in Eq. (62) also suffices to prove Proposition 1:

*Proof of Proposition 1.* The proof follows exactly as the proof of Theorem 1, but with the factor  $7/8$  instead of  $\bar{\eta}_{K,t}$  (compare Lemma 13). Using  $\log_2(7/8) \leq -0.19$  the result can be checked.  $\square$

#### 4. Convergence to Higher Moments of the Clifford Group

In this section, we aim to prove:

**Theorem 2** (Local random Clifford designs). *Let  $n \geq 12t$ , then a local random Clifford circuit of depth  $O(n \log^{-2}(t)t^8(2nt + \log(1/\varepsilon)))$  constitutes a relative  $\varepsilon$ -approximate Clifford  $t$ -design.*

The proof of Theorem 2 follows a well-established strategy [24, 59] in a sequence of lemmas. For the sake of readability, the proofs of these lemmas have been moved to Sect. 6.4. Given a measure  $\nu$  on the Clifford group  $\text{Cl}(n)$ , recall that its  $t$ -th moment operator was defined as

$$\mathbf{M}_t(\nu) := \int_{\text{Cl}(2^n)} \text{Ad}_U^{\otimes t} d\nu(U).$$

The idea of the proof is that if  $\mathbf{M}_t(\nu)$  is close to the moment operator  $\mathbf{M}_t(\mu_{\text{Cl}}) \equiv P_{\text{Cl}}$  of the uniform (Haar) measure  $\mu_{\text{Cl}}$  on the Clifford group,  $\nu$  is an approximate Clifford design. However, we have seen that there are different notions of closeness. We define its deviation in (superoperator) *spectral norm* as

$$g_{\text{Cl}}(\nu, t) := \|\mathbf{M}_t(\nu) - \mathbf{M}_t(\mu_{\text{Cl}})\|_\infty.$$

Then, we prove the following lemma in Sect. 6.4.

**Lemma 5** (Relative  $\varepsilon 2^{2tn}$ -approximate Clifford  $t$ -designs). *Suppose that  $0 \leq \varepsilon < 1$  is such that  $g_{\text{Cl}}(\nu, t) \leq \varepsilon$ . Then,  $\nu$  is a relative  $\varepsilon 2^{2tn}$ -approximate Clifford  $t$ -design.*

Recall that we have defined the measure  $\sigma_G$  on the Clifford group  $\text{Cl}(n)$  in Def. 4 by randomly drawing from a 2-local Clifford gate set  $G$  and applying it to a random qubit  $i$ , or to a pair of adjacent qubits  $(i, i + 1)$ , respectively. For this measure, we show that it fulfills the assumptions of Lemma 5:

**Proposition 2** (Clifford expander bound). *Let  $\sigma_G$  be as in Def. 4 and  $n \geq 12t$ . Then,  $g_{\text{Cl}}(\sigma_G, t) \leq 1 - c(G)n^{-1} \log^2(t)t^{-8}$  for some constant  $c(G) > 0$ .*

We will prove Proposition 2 in the end of this section. From this, Theorem 2 follows as a direct consequence:

*Proof of Theorem 2.* First, note that  $g_{\text{Cl}}(\nu^{*k}, t) = g_{\text{Cl}}(\nu, t)^k$  for all probability measures  $\nu$  on the Clifford group. This can be easily verified using the observation

$$M_t(\mu_{\text{Cl}})M_t(\nu) = M_t(\nu)M_t(\mu_{\text{Cl}}) = M_t(\mu_{\text{Cl}}). \quad (66)$$

Hence, combining the bound given by Proposition 2 and Lemma 5, we find that the  $k$ -step random walk  $\sigma_G^{*k}$  is a  $\varepsilon$ -approximate Clifford  $t$ -design, if we choose  $k = O(n \log^{-2}(t)t^8(2nt + \log(1/\varepsilon)))$ .  $\square$

For the sake of readability, let us from now on drop the dependence on  $G$  and write  $\sigma \equiv \sigma_G$ . In order to prove Proposition 2, we use a reformulation of  $g(\sigma, t)$  based on the following observation. Since  $G$  is closed under taking inverses, the moment operator  $M_t(\sigma)$  is self-adjoint with respect to the Hilbert-Schmidt inner product. Due to  $\sigma$  being a probability measure, its largest eigenvalue is 1 with eigenspace corresponding to the operator subspace which is fixed by the adjoint action  $\text{Ad}(g^{\otimes t})$  of all generators [59]. Equivalently, this is the subspace of operators which commute with any generator  $g^{\otimes t}$ . However, any operator commuting with all generators also commutes with every element in the Clifford group  $\text{Cl}(n)$  and vice versa. Hence, this subspace is nothing but the Clifford commutant  $\text{Cl}(n)'$  with projector  $P_{\text{Cl}} := M_t(\mu_{\text{Cl}})$ . Thus, the spectral decomposition is

$$M_t(\sigma) = P_{\text{Cl}} + \sum_{r \geq 2} \lambda_r(M_t(\sigma)) \Pi_r, \quad (67)$$

where  $\lambda_r(X)$  denotes the  $r$ -th largest eigenvalue of a normal operator  $X$ . Hence, we find

$$g(\sigma, t) = \|M_t(\sigma) - P_{\text{Cl}}\|_{\infty} = \lambda_*(M_t(\sigma)) := \max\{\lambda_2(M_t(\sigma)), |\lambda_{\min}(M_t(\sigma))|\}, \quad (68)$$

where  $\lambda_{\min}(M_t(\sigma))$  is the smallest eigenvalues of  $M_t(\sigma)$ . We continue by arguing that it sufficient to consider the case when  $\lambda_*(M_t(\sigma)) = \lambda_2(M_t(\sigma)) > 0$ .

To this end, consider the linear operator  $T_{\sigma} : L^2(\text{Cl}(n)) \rightarrow L^2(\text{Cl}(n))$  given as

$$T_{\sigma} f(g) := \int f(h^{-1}g) d\sigma(h). \quad (69)$$

This is the (Hermitian) averaging operator with respect to  $\sigma$  on the group algebra  $L^2(\text{Cl}(n))$ . The largest eigenvalue of  $T_{\sigma}$  is  $\lambda_1(T_{\sigma}) = 1$  and its eigenspace corresponds to the trivial representation. By Ref. [60, Lem. 1], its smallest eigenvalue is lower bounded by

$$\lambda_{\min}(T_{\sigma}) \geq -1 + 2\sigma(\mathbb{1}) = -1 + \frac{2}{|G|}, \quad (70)$$

where  $\sigma(\mathbb{1}) \equiv \sigma(\{\mathbb{1}\}) = 1/|G|$  is the probability of drawing the identity. According to the Peter-Weyl theorem, the spectrum of  $M_t(\sigma)$  is exactly the spectrum of the restriction of  $T_{\sigma}$  to the irreducible representations that appear in the representation  $U \mapsto \text{Ad}_U^{\otimes t}$ . In particular, we find  $\lambda_{\min}(M_t(\sigma)) \geq -1 + \frac{2}{|G|}$ . Let us assume that  $\lambda_*(M_t(\sigma)) = |\lambda_{\min}(M_t(\sigma))|$ . Then,  $g(\sigma, t) \leq 1 - 2/|G| < 1$  and hence we can argue as in the proof of Thm. 2 to show that local random Clifford circuits form relative  $\varepsilon$ -approximate Clifford  $t$ -designs in depth  $O(2nt + \log(1/\varepsilon))$ .

Therefore, we consider the more relevant case when  $\lambda_*(M_t(\sigma)) = \lambda_2(M_t(\sigma)) > 0$  in the following, this is

$$g(\sigma, t) = \|M_t(\sigma) - P_{\text{Cl}}\|_{\infty} = \lambda_2(M_t(\sigma)). \quad (71)$$

Since  $M_t(\sigma)$  is self-adjoint, we can interpret it as an Hamiltonian on the Hilbert space  $L((\mathbb{C}^2)^{\otimes nt})$ . In this light, it will turn out to be useful to recast Eq. (71) as the spectral gap of a suitable family of *local Hamiltonians* with vanishing ground state energy:

$$H_{n,t} := n(\text{id} - M_t(\sigma)) = \sum_{i=1}^n h_{i,i+1}, \quad \text{with } h_{i,i+1} := \frac{1}{|G|} \sum_{g \in G} \left( \text{id} - \text{Ad}(g_{i,i+1}^{\otimes t}) \right). \quad (72)$$

Let us summarize these findings in the following lemmas.

**Lemma 6** (Spectral gap). *Let  $\sigma$  be as in Def. 4 and  $H_{n,t}$  the Hamiltonian from Eq. (72). It holds that*

$$g(\sigma, t) = 1 - \frac{\Delta(H_{n,t})}{n}. \quad (73)$$

**Lemma 7** (Ground spaces). *The Hamiltonians  $H_{n,t}$  are positive operators with ground state energy 0. The ground space is given by the Clifford commutant*

$$\text{Cl}(n)' = \text{span} \left\{ r(T)^{\otimes n} \mid T \in \Sigma_{t,t} \right\}, \quad (74)$$

where  $\Sigma_{t,t}$  is the set of stochastic Lagrangian sub-spaces of  $\mathbb{Z}_2^t \oplus \mathbb{Z}_2^t$ .

In the remainder of this section, we will prove the existence of a uniform lower bound on the spectral gap of  $H_{n,t}$ . In combination with Lemma 6 and Lemma 5 this will imply Theorem 2. While it is highly non-trivial to show spectral gaps in the thermodynamic limits, we can use the fact that  $H_{n,t}$  is *frustration-free* (compare Lemma 7). This allows us to apply the powerful *martingale method* pioneered by Nachtergaele [61].

**Lemma 8.** (Lower bound to spectral gap) *Let the Hamiltonian  $H_{n,t}$  be as in Eq. (72) and assume that  $n \geq 12t$ . Then,  $H_{n,t}$  has a spectral gap satisfying*

$$\Delta(H_{n,t}) \geq \frac{\Delta(H_{12t,t})}{48t}. \quad (75)$$

*Proof of Proposition 2.* We can now combine the bound in (75) with any lower bound on the spectral gap independent of  $t$ . To this end, we make again use of the averaging operator  $T_{\sigma} : L^2(\text{Cl}(n)) \rightarrow L^2(\text{Cl}(n))$  introduced in Eq. (69) before. By Ref. [60, Cor. 1] we have that

$$\lambda_2(T_{\sigma}) \leq 1 - \frac{\eta}{d^2}, \quad (76)$$

where  $\eta$  is the probability of the least probable generator (here  $1/|G|n$ ) and  $d$  is the diameter of the associated Cayley graph (given in Ref. [62] as  $d = O(n^3 / \log(n))$ ).

Since the representation  $U \mapsto \text{Ad}_U^{\otimes t}$  contains a trivial component, the second largest eigenvalue of  $M_t(\sigma)$  can be at most  $\lambda_2(T_{\sigma})$ . Thus,  $H_{n,t}$  has a gap of at least  $\eta/d^2$ . Finally, by Lemma 8 it follows that

$$\Delta(H_{n,t}) \geq \frac{\Delta(H_{12t,t})}{48t} \geq c(G)t^{-8} \log(t)^2, \quad (77)$$

for a constant  $c(G)$ . We note that the applicability of Ref. [60, Cor. 1] to random walks on the Clifford group has also been observed in Ref. [9].  $\square$

We can combine Theorem 2 and Theorem 1 to obtain the following corollary:

**Corollary 2** (Local random unitary design). *Let  $K \in U(2)$  be a non-Clifford gate and let  $G \subset \text{Cl}(4)$  be a closed, generating set. There are constants  $C_1''(K, G)$ ,  $C_2''(K)$ ,  $C_3''(K)$  such that whenever*

$$m \geq C_1''(K, G)n \log^{-2}(t)t^8 (2nt + \log(1/\varepsilon)) \quad \text{and} \quad k \geq C_2''(K) \log^2(t)(t^4 + t \log(1/\varepsilon)),$$

*the local random circuit  $\sigma_{k,m}$ , defined in (6), is an  $\varepsilon$ -approximate unitary  $t$ -design for all  $n \geq C_3''(K)t^2$ .*

*Proof.* Consider the superoperator

$$\mathbf{M}_t(\sigma_{k,m}) = \int_{U(2^n)} \text{Ad}(U^{\otimes t}) d\sigma_{k,m}(U) = \underbrace{\mathbf{M}_t(\sigma^{*m})\mathbf{R}(K) \dots \mathbf{M}_t(\sigma^{*m})\mathbf{R}(K)}_{k \text{ times}}, \quad (78)$$

where  $\sigma^{*m}$  denotes the probability measure of a depth  $m$  local random walk on the Clifford group (cp. Def. 4). We would like to bound the difference between the Haar random  $t$ -th moment operator  $\mathbf{M}_t(\mu_{\text{H}}) =: P_{\text{H}}$  and  $\mathbf{M}_t(\sigma_{k,m})$ . Notice the following standard properties of  $P_{\text{H}}$ :

$$P_{\text{H}}\mathbf{M}_t(\nu) = \mathbf{M}_t(\nu)P_{\text{H}} = P_{\text{H}}, \quad \text{and} \quad P_{\text{H}}^\dagger = P_{\text{H}}, \quad (79)$$

for any probability measure  $\nu$  on  $U(2^n)$ . In particular, we have that  $P_{\text{H}}$  is an orthogonal projector. As in the last section, we make use of the spectral decomposition in Eq. (67) to decompose  $\mathbf{M}_t(\sigma^{*k})$  as follows:

$$\begin{aligned} \mathbf{M}_t(\sigma_{k,m}) - P_{\text{H}} &= [\mathbf{M}_t(\sigma^{*m})\mathbf{R}(K)]^k - P_{\text{H}} \\ &= \left[ \left( P_{\text{Cl}} + \sum_{i \geq 2} \lambda_i^m \Pi_i \right) \mathbf{R}(K) \right]^k - P_{\text{H}}. \end{aligned} \quad (80)$$

Recall the shorthand notation  $P_{\text{Cl}} := \mathbf{M}_t(\mu_{\text{Cl}})$ . Using the triangle inequality and the inequality (12), this implies

$$\begin{aligned} \|\mathbf{M}_t(\sigma_{k,m}) - P_{\text{H}}\|_{\diamond} &\leq \|[P_{\text{Cl}}\mathbf{R}(K)]^k - P_{\text{H}}\|_{\diamond} + 2^{2tn} \sum_{l=1}^k \binom{k}{l} \lambda_2^{lm} \\ &\leq \|[P_{\text{Cl}}\mathbf{R}(K)]^k - P_{\text{H}}\|_{\diamond} + k2^{2tn+1} \lambda_2^m. \end{aligned} \quad (81)$$

Note that we bounded the second largest eigenvalue  $\lambda_2$  of  $\mathbf{M}_t(\sigma)$  in Proposition 2. We can now combine Proposition 2 with (62) to obtain:

$$\|\mathbf{M}_t(\sigma_{k,m}) - P_{\text{H}}\|_{\diamond} \leq k2^{2tn+1} \lambda_2^m + 2^{33t^4+t \log(k)} \left(1 + 2^{32t^2-n}\right)^{5k} \bar{\eta}_{K,t}^k. \quad (82)$$

□

## 5. Singling out the Clifford Group

There are a number of ways to motivate the construction of approximate unitary  $t$ -designs from random Clifford circuits. From a practical point of view, Clifford gates are often comparatively easy to implement, in particular in fault-tolerant architectures. In this section, we point out that Refs. [38,39] together imply that the Clifford groups are also mathematically distinguished. We formulate this observation as Proposition 3: The finite case follows from the recently obtained classification of finite unitary subgroups forming  $t$ -designs, so-called *unitary  $t$ -groups*, by [38] building on earlier results by [55]. The infinite case is a corollary of a theorem about universality of finitely generated subgroups by [39].

This section is independent from the rest of the paper and has the sole purpose of highlighting the results in Refs. [38,39,55] and explicitly formulate their combined implications for the generation of unitary  $t$ -designs. Moreover, it might serve as an intuitive justification for the usefulness and omnipresence of Clifford unitaries in random circuit constructions.

For any subgroup  $G \subseteq U(d)$ , we let

$$\overline{G} := \{\det(U^\dagger)U \mid U \in G\} \subseteq SU(d).$$

Notice that  $\overline{G}$  is a unitary  $t$ -design if and only if  $G$  is.

Proposition 3 refers to  $t$ -designs generated by *finite gate sets*, which we define now. The starting point is a Hilbert space  $(\mathbb{C}^q)^{\otimes r}$  for some  $r$ . A finite gate set is a finite subset

$$\mathcal{G} \subset SU((\mathbb{C}^q)^{\otimes r}).$$

We will denote by  $\mathcal{G}_n$  the subgroup of  $SU((\mathbb{C}^q)^{\otimes n})$  generated by elements of  $\mathcal{G}$  acting on any  $r$  tensor factors (here  $r \leq n$ ). The number  $q$  is called the *local dimension* of  $\mathcal{G}$ .

**Proposition 3** (Singling out the Clifford group [38,39,55]). *Let  $t \geq 2$ , and let  $\mathcal{G}$  be a finite gate set with local dimension  $q \geq 2$ . Assume that (1) either all  $\mathcal{G}_n$  are finite or they are all infinite, and (2) there is an  $n_0$  such that for all  $n \geq n_0$ ,  $\mathcal{G}_n$  is a unitary  $t$ -design.*

*Then, one of the following cases apply:*

- (i) *If  $t = 2$ , we have either  $q$  prime and  $\mathcal{G}_n$  is isomorphic to a subgroup of the Clifford group  $\overline{\text{Cl}}(q^n)$ , or  $\mathcal{G}_n$  is dense in  $SU(q^n)$ ,*
- (ii) *If  $t = 3$ , we have either  $q = 2$  and  $\mathcal{G}_n$  is isomorphic to the full Clifford group  $\overline{\text{Cl}}(2^n)$  or  $\mathcal{G}_n$  is dense in  $SU(q^n)$ ,*
- (iii) *If  $t \geq 4$  then  $\mathcal{G}_n$  is dense in  $SU(q^n)$ .*

Note that a finitely generated infinite subgroup of  $SU(d)$  is always dense in some compact Lie subgroup (cp. [39, Fact 2.6]). In particular, it inherits a Haar measure from this Lie subgroup which allows for a definition of unitary  $t$ -design.

*a. Finite case.* In the classification in Ref. [38], the non-existence of finite unitary  $t$ -groups was shown for  $t \geq 4$  (and dimension  $d > 2$ ). Already the case  $t = 3$  is very restrictive, since the authors arrive at the following result:

**Lemma 9** (Ref. [38, Thm. 4]). *Suppose  $d \geq 5$  and consider a finite subgroup  $H < SU(d)$  which is a unitary 3-design. Then,  $H$  is either one of finitely many exceptional cases or  $d = 2^n$  and  $H$  is isomorphic to the Clifford group  $\overline{\text{Cl}}(2^n)$ .*

This establishes the finite version of (ii), the  $t = 3$  case.

The classification of unitary 2-designs is however more involved, it includes certain irreducible representations of finite unitary and symplectic groups (compare [38, Thm. 3 Lie-type case]), and a finite set of exceptions. The exceptions can be ruled out in the same way as above.

The former, the Lie-type cases, happen in dimensions  $(3^n \pm 1)/2$  and  $(2^n + (-1)^n)/3$ . There is no  $q$  for which there exists an  $n_0$  such that for all  $n \geq n_0$  there exists an  $m \in \mathbb{N}$  satisfying either

$$q^n = (3^m \pm 1)/2 \quad \text{or} \quad q^n = (2^m + (-1)^m)/3.$$

Thus, the assumptions of Prop. 3 rule these out. This establishes the finite version of (i).

*b. Infinite case.* Define the commutant for a set  $S \subset \text{SU}(d)$  of the adjoint action as

$$\text{Comm}(\text{Ad}_S) := \left\{ L \in \text{End}(\mathbb{C}^{d \times d}) \mid [\text{Ad}_g, L] = 0 \quad \forall g \in S \right\}.$$

We show that the second case can be reduced to Cor. 3.5 from Ref. [39] applied to the simple Lie group  $\text{SU}(d)$ .

**Lemma 10** ([39, Cor. 3.5]). *Given a finite set  $G \subset \text{SU}(d)$  such that  $\mathcal{G} = \langle G \rangle$  is infinite. Then, the group  $\mathcal{G}$  is dense in  $\text{SU}(d)$  if and only if*

$$\text{Comm}(\text{Ad}_{\mathcal{G}}) \cap \text{End}(\mathfrak{su}(d)) = \{\lambda \text{id}_{\mathfrak{su}(d)} \mid \lambda \in \mathbb{R}\}. \tag{83}$$

Recall that a subgroup  $\mathcal{G} \subseteq U(d)$  is a unitary 2-group if and only if  $\text{Comm}(U \otimes U \mid U \in \mathcal{G}) = \text{Comm}(U \otimes U \mid U \in U(d)) = \text{span}(\mathbb{1}, \mathbb{F})$ , where  $\mathbb{F}$  denotes the flip of two tensor copies (see also App. A). Let us denote the partial transpose on the second system of a linear operator  $A \in L(\mathbb{C}^d \otimes \mathbb{C}^d)$  by  $A^\Gamma$ . Then, one can easily verify that  $\Gamma$  induces a vector space isomorphism between  $\text{Comm}(U \otimes U \mid U \in \mathcal{G})$  and  $\text{Comm}(U \otimes \bar{U} \mid U \in \mathcal{G})$ . The image of the basis  $\{\mathbb{1}, \mathbb{F}\}$  is readily computed as

$$\mathbb{1}^\Gamma = \mathbb{1}, \quad \mathbb{F}^\Gamma = d \mid \Omega \rangle \langle \Omega \mid, \tag{84}$$

where  $\mid \Omega \rangle = d^{-1/2} \sum_{i=1}^d \mid ii \rangle$  is the maximally entangled state vector. Next, we use that  $U \otimes \bar{U} = \text{mat}(\text{Ad}_U)$  is the matrix representation of  $\text{Ad}_U = U \cdot U^\dagger$  with respect to the basis  $E_{i,j} = \mid i \rangle \langle j \mid$  of  $L(\mathbb{C}^d)$ . Thus, we have  $\text{Comm}(\text{Ad}_{\mathcal{G}}) \simeq \text{Comm}(U \otimes \bar{U} \mid U \in \mathcal{G})$  as algebras. Pulling the above basis of  $\text{Comm}(U \otimes \bar{U} \mid U \in \mathcal{G})$  back to  $\text{Comm}(\text{Ad}_{\mathcal{G}})$ , we then find:

$$\text{mat}^{-1}(\mathbb{1}) = \text{id}_{L(\mathbb{C}^d)}, \quad \text{mat}^{-1}(\mid \Omega \rangle \langle \Omega \mid) = \text{Tr}(\bullet) \text{id}_{L(\mathbb{C}^d)}. \tag{85}$$

Hence, we have shown that any element in  $\text{Comm}(\text{Ad}_{\mathcal{G}})$  is a linear combination of these two maps. However, by restricting to  $\mathfrak{su}(d)$ , the second map becomes identically zero, thus we have

$$\text{Comm}(\text{Ad}_{\mathcal{G}}) \cap \text{End}(\mathfrak{su}(d)) = \{\lambda \text{id}_{\mathfrak{su}(d)} \mid \lambda \in \mathbb{R}\}. \tag{86}$$

By Lemma 10, this shows that any finitely generated infinite unitary 2-group  $\mathcal{G} \leq \text{SU}(d)$  is dense in  $\text{SU}(d)$ . Since any unitary  $t$ -group is in particular a 2-group, this is also true for any  $t > 2$ .

## 6. Proofs

*6.1. Proof of overlap lemmas.* In this section, we prove three technical lemmas which are needed throughout this paper. These lemmas give bounds on the overlaps of the operators  $Q_T^{\otimes n}$  and hence quantify how far this basis is from an orthonormal basis of the commutant of the Clifford tensor power representation, i.e., for range  $P_{\text{Cl}}$ .

**Lemma 3** (Diamond norm bound). *Consider  $T_1, T_2 \in \Sigma_{t,t}$  and denote with  $N_1, N_2$  their respective defect spaces. Then, it holds that*

$$\| |Q_{T_1}\rangle\langle Q_{T_2}| \|_{\diamond} \leq 2^{\dim N_2 - \dim N_1}, \quad (27)$$

$$| \langle Q_{T_1} | Q_{T_2} \rangle | \leq 2^{-|\dim N_1 - \dim N_2|}. \quad (28)$$

*Proof.* First, recall that  $Q_T := 2^{-t/2}r(T)$ . Then, we make use of the following elementary bound on the diamond norm of rank one superoperator  $|A\rangle\langle B|$ :

$$\begin{aligned} \| |A\rangle\langle B| \|_{\diamond} &= \sup_{\|X\|_1=1} \|A \otimes \text{Tr}_1(B \otimes \mathbb{1}X)\|_1 \\ &\stackrel{\dagger}{\leq} \|A\|_1 \sup_{\|X\|_1=1} \|B \otimes \mathbb{1}X\|_1 \\ &\stackrel{\ddagger}{=} \|A\|_1 \|B \otimes \mathbb{1}\|_{\infty} \\ &= \|A\|_1 \|B\|_{\infty}. \end{aligned} \quad (87)$$

Here, we have used in  $\dagger$  that the partial trace is a contraction w.r.t.  $\|\cdot\|_1$  and in  $\ddagger$  a version of the duality between trace and spectral norm [63]. Given stochastic Lagrangians  $T_1$  and  $T_2$  with defect spaces  $N_1$  and  $N_2$ , we thus find using Lem. 1:

$$\| |Q_{T_1}\rangle\langle Q_{T_2}| \|_{\diamond} \leq 2^{-t} \|r(T_1)\|_1 \|r(T_2)\|_{\infty} = 2^{\dim N_2 - \dim N_1}. \quad (88)$$

To prove 2., we use Ref. [42, Eq. (4.25)] and that the transpose does not change the dimension of the corresponding defect subspace. Moreover, we assume w.l.o.g. that  $\dim N_2 \geq \dim N_1$ . We have

$$| \langle Q_{T_1} | Q_{T_2} \rangle | = 2^{-t} | \text{Tr}[r(T_1)r(T_2)^T] | = 2^{-t+\dim(N_1 \cap N_2)} | \text{Tr}[r(T)] | \quad (89)$$

where  $r(T)$  is described by a stochastic orthogonal and a defect space  $N_1^{\perp} \cap N_2 + N_1$ . Hence, we obtain (together with Hölder's inequality):

$$| \langle Q_{T_1} | Q_{T_2} \rangle | \leq 2^{-t+\dim(N_1 \cap N_2)} 2^{t-\dim(N_1^{\perp} \cap N_2 + N_1)}. \quad (90)$$

Using  $N \subseteq N^{\perp}$  for all defect spaces and the general identity  $\dim(V + W) = \dim V + \dim W - \dim(V \cap W)$ , this yields

$$| \langle Q_{T_1} | Q_{T_2} \rangle | \leq 2^{\dim(N_1 \cap N_2) - \dim N_1} \leq 2^{\dim N_2 - \dim N_1}. \quad (91)$$

□

Next, we define a *frame operator* associated to the basis  $Q_T^{\otimes n}$ . If the basis was orthogonal, this frame operator would simply be the projector  $P_{\text{Cl}}$  onto the Clifford commutant.



**Definition 9** (*Clifford frame operator*). We define the Clifford frame operator of the basis  $Q_T^{\otimes n}$  as

$$S_{\text{Cl}} := \sum_{T \in \Sigma_{t,t}} |Q_T\rangle\langle Q_T|^{\otimes n}. \quad (92)$$

Hence, a quantifier for the orthogonality of the  $Q_T^{\otimes n}$  basis is the distance of  $S_{\text{Cl}}$  to the projector  $P_{\text{Cl}}$ . As we prove in Lem. 12, we have  $P_{\text{Cl}} \approx S_{\text{Cl}}$  in spectral norm and we will use this result later in the proof of Lem. 8. In order to show this, we first derive a result on the *sum of overlaps* in Lem. 11.

Interestingly,  $S_{\text{Cl}}$  is *not* close to  $P_{\text{Cl}}$  in diamond norm (see. Ch. 15 in Ref. [64]). To derive our main result, we instead construct an orthogonalized basis from the  $Q_T^{\otimes n}$ . Some properties of the orthogonalized basis are proven in Lem. 4, which also makes use of Lem. 11.

**Lemma 11** (Overlap of stochastic Lagrangian sub-spaces). *We have  $\langle Q_T | Q_{T'} \rangle \geq 0$  for all  $T, T' \in \Sigma_{t,t}$ . Moreover, for all  $T \in \Sigma_{t,t}$  the sum of overlaps is*

$$\sum_{T' \in \Sigma_{t,t}} \langle Q_T | Q_{T'} \rangle^n = (-2^{-n}; 2)_{t-1} \leq 1 + t2^{t-n}, \quad (93)$$

where  $(-2^{-n}; 2)_{t-1} = \prod_{r=0}^{t-2} (1 + 2^{r-n})$  and the last inequality holds for  $n + 2 \geq t + \log_2(t)$ .

*Proof.* Denote by  $\text{Stab}(n)$  the set of stabilizer states on  $n$  qubits. Since the operators  $r(T)$  are entry-wise non-negative, we have  $\langle Q_T | Q_{T'} \rangle = 2^{-t} \text{Tr}(r(T)^\dagger r(T')) \geq 0$ . Note that  $r(T)^\dagger = r(\tilde{T})$  for a suitable  $\tilde{T} \in \Sigma_{t,t}$  (cp. Thm. 4). We obtain

$$\begin{aligned} \sum_{T' \in \Sigma_{t,t}} \langle Q_T | Q_{T'} \rangle^n &= \frac{1}{2^{tn}} \sum_{T' \in \Sigma_{t,t}} \text{Tr} \left[ r(\tilde{T})^{\otimes n} r(T')^{\otimes n} \right] \\ &\stackrel{\dagger}{=} \frac{2^n \prod_{r=0}^{t-2} (2^r + 2^n)}{2^{tn}} \text{Tr} \left[ r(\tilde{T})^{\otimes n} \mathbb{E}_{s \in \text{Stab}(n)} (|s\rangle\langle s|^{\otimes t}) \right] \\ &= \frac{2^n \prod_{r=0}^{t-2} (2^r + 2^n)}{2^{tn}} \mathbb{E}_{s \in \text{Stab}(n)} \langle s^{\otimes t} | r(\tilde{T})^{\otimes n} | s^{\otimes t} \rangle \\ &\stackrel{\ddagger}{=} \frac{2^n \prod_{r=0}^{t-2} (2^r + 2^n)}{2^{tn}} \\ &= \prod_{r=0}^{t-2} (1 + 2^{r-n}) \\ &\leq \left( 1 + 2^{t-2-n} \right)^{t-1} \\ &\stackrel{*}{\leq} \exp \left( (t-1)2^{t-n-2} \right), \end{aligned} \quad (94)$$

where we have again used [42, Thm. 5.3] in  $\dagger$  and in  $\ddagger$  that  $\langle s^{\otimes t} | r(T)^{\otimes n} | s^{\otimes t} \rangle = 1$  for all  $T \in \Sigma_{t,t}$  and all  $s \in \text{Stab}(n)$  (compare Ref. [42, Eq. (4.10)]). Finally, in  $*$  we have used the ‘‘inverse Bernoulli inequality’’  $(1+x)^r \leq e^{rx}$  which holds for all  $x \in \mathbb{R}$  and  $r \geq 0$ . By assumption, the following holds

$$0 \geq t + \log_2(t) - n - 2 \quad \Rightarrow \quad 1 \geq t2^{t-n-2} \geq (t-1)2^{t-n-2}. \quad (95)$$

Thus, we can use the inequality  $e^x \leq 1 + 2x$  for  $0 \leq x \leq 1$  to obtain

$$\begin{aligned} \sum_{T' \in \Sigma_{t,t}} (Q_T | Q_{T'})^n &\leq 1 + (t-1)2^{t-n-1} \\ &\leq 1 + t2^{t-n}. \end{aligned} \quad (96)$$

□

**Lemma 12.** *Let  $S_{\text{Cl}}$  be the Clifford frame operator and  $\Gamma$  the corresponding Gram matrix, i. e.  $\Gamma_{T,T'} = (Q_T | Q_{T'})^n$ . Then the following holds*

$$\|S_{\text{Cl}} - P_{\text{Cl}}\|_{\infty} = \|\Gamma - \mathbb{1}\|_{\infty} \leq (-2^{-n}; 2)_{t-1} - 1 \leq t2^{t-n}, \quad (97)$$

where  $(-2^{-n}; 2)_{t-1} = \prod_{r=0}^{t-2} (1 + 2^{r-n})$  and the last inequality holds for  $n + 2 \geq t + \log_2(t)$ .

*Proof.* Define the *synthesis operator* of the frame as the map

$$V : \mathbb{C}^{|\Sigma_{t,t}|} \rightarrow \text{Cl}(n)', \quad V = \sum_{T \in \Sigma_{t,t}} |Q_T^{\otimes n}\rangle \langle e_T|, \quad (98)$$

where  $e_T$  is the standard basis of the domain. Then, we have clearly  $\Gamma = V^\dagger V$  and  $S_{\text{Cl}}|_{\text{Cl}(n)'} = VV^\dagger$ . Since  $S_{\text{Cl}}$  and  $P_{\text{Cl}}$  are both identically zero on  $(\text{Cl}(n)')^\perp$ , this part does not contribute to the spectral norm. From this it is clear that

$$\|S_{\text{Cl}} - P_{\text{Cl}}\|_{\infty} = \|\Gamma - \mathbb{1}\|_{\infty}. \quad (99)$$

Moreover, we can compute

$$\begin{aligned} \|\Gamma - \mathbb{1}\|_{\infty} &= \left\| \sum_T \sum_{T'} (Q_T | Q_{T'})^n |e_T\rangle \langle e_{T'}| \right\|_{\infty} \\ &\leq \max_T \sum_{T' \neq T} (Q_T | Q_{T'})^n \\ &= (-2^{-n}; 2)_{t-1} - 1, \end{aligned} \quad (100)$$

where we have used that the spectral norm of Hermitian operators is bounded by the max-column norm and inserted the exact result of Lemma 11 in the last step. Finally, said lemma provides the desired bound for  $n + 2 \geq t + \log_2 t$ . □

## 6.2. Proof of Lemmas for Theorem 1.

**Lemma 2** (Overlap bound). *Let  $K$  be a single qubit gate which is not contained in the Clifford group. Then, there is a constant  $c(K) > 0$  such that*

$$\eta_{K,t} := \max_{\substack{T \in \Sigma_{t,t} - S_t \\ T' \in \Sigma_{t,t}}} \frac{1}{3} \left| (Q_T | \text{Ad}_K^{\otimes t} + \text{Ad}_{K^\dagger}^{\otimes t} + \text{id} | Q_{T'}) \right| \leq 1 - c(K) \log^{-2}(t). \quad (26)$$

The proof of Lemma 2 is based on two results. The first states that the basis elements  $r(T)$  of the commutant of tensor powers of the Clifford group either belong to the commutant of the powers of the unitary group, or else are far away from it.

**Lemma 13** (Haar symmetrization). *For all  $t$  and for all  $T \in \Sigma_{t,t} \setminus S_t$ , it holds that*

$$(Q_T | P_H | Q_T) = 2^{-t} \|P_H[r(T)]\|_2^2 \leq \frac{7}{8}, \quad (101)$$

where  $Q_T$  is as in Eq. (21) and  $P_H = M_t(\mu_H)$  is the  $t$ -th moment operator of the single-qubit unitary group  $U(2)$ .

The proof is given in Sect. 6.3. In Appendix C, we show that the constant  $7/8$  cannot be improved below  $7/10$ , by exhibiting a  $T$  that attains this bound.

The second ingredient to Lemma 2 is a powerful theorem by Varjú [53]. Here, we specialize this theorem to the unitary group:

**Theorem 5** ([53, Thm. 6]. *Let  $\nu$  be a probability measure on  $U(d)$ . Consider the averaging operator  $T_\nu(\nu)$  on a irreducible representation  $\pi_\nu : U(d) \rightarrow \text{End}(W_\nu)$  parameterized by highest weight  $\nu \in \mathbb{Z}^d$ :*

$$T_\nu(\nu) := \int_{U(d)} \pi_\nu(U) d\nu(U). \quad (102)$$

Then there are numbers  $C(d) > 0$  and  $r_0 > 0$  such that

$$\Delta_r(\nu) := 1 - \max_{0 < |\nu| \leq r} \|T_\nu(\nu)\|_\infty \geq C(d) \Delta_{r_0}(\nu) \log^{-2}(r), \quad (103)$$

where  $|\nu|^2 = \sum_i \nu_i^2$ .

*Proof of Lemma 2.* Consider the probability measure  $\xi_K$  that draws uniformly from the set  $\{K, K^\dagger, \mathbb{1}\}$ . Moreover, define  $\nu_K$  on  $U(2)$  as the average of the uniform measure on  $\{H, S, S^3\}$  and  $\xi_K * \xi_K$ . Hence, the according moment operator is

$$\begin{aligned} M_t(\nu_K) &:= \frac{1}{6} (\text{Ad}_H^{\otimes t} + \text{Ad}_S^{\otimes t} + (\text{Ad}_S^3)^{\otimes t}) + \frac{1}{2} M_t(\xi_K * \xi_K) \\ &= \frac{1}{6} (\text{Ad}_H^{\otimes t} + \text{Ad}_S^{\otimes t} + (\text{Ad}_S^3)^{\otimes t}) + \frac{1}{2} M_t(\xi_K)^2. \end{aligned} \quad (104)$$

As the Clifford group augmented with any non-Clifford gate is universal [65, Thm. 6.5], so is the probability measure  $\nu_K$ .

It follows from the representation theory of the unitary group (see App. B) that the representation  $U \mapsto \text{Ad}_U^{\otimes t}$  does not contain irreducible representations  $W_\nu$  with highest weight of length  $|\nu| > \sqrt{2}t$ . Thus, we can decompose into these irreducible representations as follows:

$$\begin{aligned} \|M_t(\nu_K) - P_H\|_\infty &= \left\| \bigoplus_{|\nu| \leq \sqrt{2}t} (T_\nu(\nu_K) - T_\nu(\mu_H)) \otimes \text{id}_{m_\nu} \right\|_\infty \\ &\leq \left\| \bigoplus_{0 < |\nu| \leq \sqrt{2}t} T_\nu(\nu_K) \right\|_\infty \end{aligned}$$

$$\begin{aligned}
&= \max_{0 < |v| \leq \sqrt{2t}} \|T_v(v_K)\|_\infty \\
&= 1 - \Delta_{\sqrt{2t}}(v_K).
\end{aligned} \tag{105}$$

Here,  $m_v$  denotes the multiplicity of the irreducible representation  $W_v$  (possibly zero). In the second step we have used that  $P_H$  has only support on the trivial irreducible representation  $v = 0$ , where both  $P_H$  and  $M_t(v_K)$  act as identity and thus cancel. Hence, only non-trivial irreducible representations are contributing. To bound  $\Delta_{\sqrt{2t}}(v_K)$ , we can invoke Theorem 5 combined with the fact that for any universal probability measure the restricted gap is non-zero:  $\Delta_r(v_K) > 0$  for all  $r \geq 1$  (compare e.g. Ref. [27]). Hence, we obtain

$$\begin{aligned}
\Delta_{\sqrt{2t}}(v_K) &\geq C(2)\Delta_{r_0}(v_K)\log^{-2}(\sqrt{2t}) \\
&\geq \frac{1}{4}C(2)\Delta_{r_0}(v_K)\log^{-2}(t) =: c'(K)\log^{-2}(t) > 0,
\end{aligned} \tag{106}$$

where  $c(K) > 0$ . Therefore, we have

$$\|M_t(v_K) - P_H\|_\infty \leq 1 - \Delta_{\sqrt{2t}}(v_K) \leq 1 - c'(K)\log^{-2}(t) =: \kappa_{t,K}, \tag{107}$$

Furthermore, consider the operator

$$X_T := \frac{(\text{id} - P_H)Q_T}{\|(\text{id} - P_H)Q_T\|_2}. \tag{108}$$

We obtain

$$\begin{aligned}
\|M_t(v_K) - P_H\|_\infty &= \max_{\|X\|_2=1} |(X|M_t(v_K) - P_H|X)| \\
&\geq \frac{|(X_T|M_t(v_K) - P_H|X_T)|}{\|X_T\|_2^2} \\
&= \frac{|(Q_T|(\text{id} - P_H)M_t(v_K)(\text{id} - P_H)|Q_T)|}{(Q_T|(\text{id} - P_H)^2|Q_T)} \\
&= \frac{|(Q_T|M_t(v_K)|Q_T) - (Q_T|P_H|Q_T)|}{1 - (Q_T|P_H|Q_T)} \\
&\geq \frac{(Q_T|M_t(v_K)|Q_T) - (Q_T|P_H|Q_T)}{1 - (Q_T|P_H|Q_T)}.
\end{aligned} \tag{109}$$

In the fourth step, we again used the properties of the Haar projector as in Eq. (79). Combining this with (107) and Lemma 13 we obtain

$$(Q_T|M_t(v_K)|Q_T) \leq \kappa_{t,K} + (1 - \kappa_{t,K})(Q_T|P_H|Q_T) \leq 1 - \frac{1}{8}c'(K)\log^{-2}(t). \tag{110}$$

We can use that  $(Q_T|\text{Ad}_S^{\otimes t}|Q_T) = (Q_T|\text{Ad}_{S^3}^{\otimes t}|Q_T) = (Q_T|\text{Ad}_H^{\otimes t}|Q_T) = 1$  for all  $T \in \Sigma_{t,t}$  because  $Q_T = 2^{-t/2}r(T)$  commutes with the  $t$ -th diagonal action of the single-qubit Clifford group (compare [42, Lem. 4.5]). We immediately obtain

$$(Q_T|M_t(\xi_K)^2|Q_T) \leq 1 - \frac{1}{4}c'(K)\log^{-2}(t). \tag{111}$$

From the Cauchy-Schwarz inequality, we now get

$$\begin{aligned}
 |(Q_T | M_t(\xi_K) | Q_{T'})| &\leq \sqrt{(Q_T | M_t(\xi_K)^2 | Q_T)} \\
 &\leq \sqrt{1 - \frac{1}{4}c'(K) \log^{-2}(t)} \\
 &\leq 1 - \frac{1}{8}c'(K) \log^{-2}(t) \\
 &=: 1 - c(K) \log^{-2}(t), \tag{112}
 \end{aligned}$$

where we have used that  $c'(K) \log^{-2}(t) \leq \Delta_{\sqrt{2}t}(\nu_K) \leq 1$  such that we can use the inequality  $\sqrt{1-x} \leq 1 - x/2$  for  $x \leq 1$ . This shows the claimed statement.  $\square$

*Remark 2 (Quantum gates with algebraic entries).* If we restrict to gates  $K$  that have only algebraic entries, we can apply the result from Ref. [66] and save the additional overhead of  $\log^2(t)$  in the scaling. This applies to the  $T$ -gate and for essentially all gates that might be used in practical implementations. Here, we have chosen the more general approach.

*Remark 3 (Implications for quantum information processing).* Theorem 5 has miscellaneous implications for quantum information processing. E.g. we can immediately combine this bound with the local-to-global lemma in Ref. [23, Lem. 16] to extend Ref. [24, Cor. 7] to gate sets with non-algebraic entries at the cost of an additional overhead of  $\log^2(t)$  in the scaling. The bottleneck to loosen the invertibility assumption as well is the local-to-global lemma which only works for Hermitian moment operators (symmetric distributions). Work to lessen the assumption of invertibility has been done in Ref. [67]. Extending this would be an interesting application which we, however, do not pursue in this work.

**Lemma 4** (Properties of the constructed basis). *Let  $\{T_j\}_{j=1}^{|\Sigma_{t,t}|}$  be an enumeration of the elements of  $\Sigma_{t,t}$  such that the first  $t!$  spaces  $T_j$  correspond to the elements of  $S_t$ . Then, the  $\{E_j\}$  constitutes an orthogonal (but not normalized) basis, where*

$$E_j := \sum_{i=1}^j A_{i,j} Q_{T_i}^{\otimes n} := \sum_{i=1}^j \left[ \sum_{\substack{\Pi \in S_j \\ \Pi(j)=i}} \text{sign}(\Pi) \prod_{l=1}^{j-1} (Q_{T_l} | Q_{T_{\Pi(l)}})^n \right] Q_{T_i}^{\otimes n}. \tag{30}$$

Denote by  $N_i$  the defect space of  $T_i$ . For  $n \geq \frac{1}{2}(t^2 + 5t)$ , we have

$$|A_{i,j}| \leq 2^{t^3+4t^2+6t-n} |\dim N_i - \dim N_j|, \quad \forall i, j, \tag{31}$$

$$|A_{i,j}| \leq 2^{2t^2+10t-n}, \quad \forall i \neq j. \tag{32}$$

Moreover, it holds that

$$1 - 2^{t^2+7t-n} \leq A_{j,j} \leq 1 + 2^{t^2+7t-n}. \tag{33}$$

*Proof.* The form of (30) is up to a constant the determinant formulation of the Gram-Schmidt procedure. First, note that the number of permutations of  $n$  elements with no fixed points is known from Ref. [68] to be

$$D(n) = n! \sum_{r=0}^n \frac{(-1)^r}{r!} \leq 2 \frac{n!}{e} \tag{113}$$

for  $n \geq 1$ . Here,  $D$  stands for ‘‘derangement’’ as permutations without fixed points are sometimes called. Then, the number of permutations having exactly  $k$  fixed points is  $\binom{n}{k}$  many choices of  $k$  points times the number  $D(n - k)$  of deranged permutations on the remaining  $n - k$  objects:

$$p(n, k) := \binom{n}{k} D(n - k) \leq 2e^{-1} \frac{n!}{k!}. \tag{114}$$

The following estimate for certain sums involving  $p(n, k)$  will shortly become useful. Note that we have for any  $M, L \in \mathbb{N}$  and  $m \in \mathbb{R}$  such that  $2^m > M - L$  and  $M \geq L \geq 1$ :

$$\begin{aligned} \sum_{k=0}^{M-L} p(M, k) 2^{-m(M-k)} &\leq \frac{2}{e} \sum_{k=0}^{M-L} 2^{-mM} M! \frac{2^{mk}}{k!} \\ &\leq \frac{2}{e} 2^{-mM} (M - L + 1) M! \frac{2^{m(M-L)}}{(M - L)!} \\ &\leq M^{L+1} 2^{-mL}. \end{aligned} \tag{115}$$

Here, we have used in the second inequality that  $2^{mk}/k!$  is monotonically increasing for  $k \leq M - L < 2^m$  and a standard bound on binomial coefficients in the last step.

We start by bounding the diagonal coefficients  $A_{j,j}$ . The idea is to divide the set of permutations into sets of permutations with exactly  $k$  fixed points. For any such permutation, the product of overlaps collapses to only  $j - 1 - k$  non-trivial inner products. By assumption  $n \geq \frac{1}{2}(t^2 + 5t) \geq t + \log_2 t$ , thus we can be bound any of those using Lemma 11 as

$$(Q_T | Q_{T'})^n \leq t 2^{t-n}, \quad \text{for all } T \neq T'. \tag{116}$$

Note that the trivial permutation (corresponding to  $k = j - 1$  fixed points) contributes by exactly 1 to the sum. Thus, we find the following bound using Eq. (115) with  $M = j - 1$ ,  $L = 1$  and  $m = n - t - \log_2 t$ :

$$\begin{aligned} A_{j,j} = |A_{j,j}| &\leq \sum_{\pi \in S_{j-1}} \prod_{l=1}^{j-1} (Q_l | Q_{\pi(l)})^n \\ &\leq 1 + \sum_{k=0}^{j-2} p(j - 1, k) 2^{-(n-t-\log_2 t)(j-1-k)} \\ &\leq 1 + (j - 1)^2 2^{-n+t+\log_2 t} \\ &< 1 + 2^{t^2+7t-n}, \end{aligned} \tag{117}$$

where we have used Eq. (15) in the last step as  $j - 1 < j \leq |\Sigma_{t,t}| \leq 2^{\frac{1}{2}(t^2+5t)}$ . Using the reverse triangle inequality, we get a lower bound in the same way:

$$A_{j,j} = |A_{j,j}| \geq 1 - \left| \sum_{\pi \in S_{j-1} \setminus \text{id}} \text{sign}(\pi) \prod_{l=1}^{j-1} (\mathcal{Q}_l | \mathcal{Q}_{\pi(l)})^n \right| \geq 1 - 2^{t^2+7t-n}. \quad (118)$$

Next, we will bound the off-diagonal terms  $A_{i,j}$ . It is well known that every permutation  $\Pi \in S_j$  can be written as a product of disjoint cycles. Given a  $\Pi \in S_j$  with  $\Pi(j) = i$ , consider the cycle  $j \mapsto i \mapsto i_1 \mapsto i_2 \mapsto \dots \mapsto i_r \mapsto j$  in  $\Pi$ . Then, we have the bound

$$\begin{aligned} \prod_{l=1}^{j-1} (\mathcal{Q}_{T_l} | \mathcal{Q}_{T_{\Pi(l)}})^n &\leq (\mathcal{Q}_{T_i} | \mathcal{Q}_{T_{i_1}})^n \dots (\mathcal{Q}_{T_{i_r}} | \mathcal{Q}_{T_j})^n \\ &\leq 2^{-n(|\dim N_i - \dim N_{i_1}| + \dots + |\dim N_{i_r} - \dim N_j|)} \\ &\leq 2^{-n|\dim N_i - \dim N_j|}, \end{aligned} \quad (119)$$

where we have used Lemma 3, the triangle inequality and a telescope sum. We set  $L := |\dim N_i - \dim N_j|$  and split the sum over permutations into those with more than or equal to  $j - L$  many fixed points and those with less. In the first case, we use Eq. (119) to bound the overlaps, in the second case we use Eq. (115) as before. This yields the following bound

$$\begin{aligned} |A_{i,j}| &\leq \sum_{\substack{\Pi \in S_j \\ \Pi(j)=i}} \prod_{l=1}^{j-1} (\mathcal{Q}_{T_l} | \mathcal{Q}_{T_{\Pi(l)}})^n \\ &\leq \sum_{k=j-L}^{j-1} p(j,k) 2^{-nL} + \sum_{k=0}^{j-L-1} p(j,k) 2^{-(n-t-\log_2 t)(j-1-k)} \\ &\leq \frac{2}{e} \sum_{k=j-L}^{j-1} \frac{j!}{k!} 2^{-nL} + 2^{n-t-\log_2 t} j^{L+2} 2^{-(n-t-\log_2 t)(L+1)} \\ &\leq L \frac{j!}{(j-L)!} 2^{-nL} + j^{L+2} 2^{-(n-t-\log_2 t)L} \\ &\leq L j^L 2^{-nL} + j^{L+2} 2^{-(n-t-\log_2 t)L} \\ &\leq L |\Sigma_{t,t}|^{L+2} 2^{-(n-t-\log_2 t)L} \\ &\leq 2^{\log_2 L} 2^{\frac{1}{2}(t^2+5t)(L+2)} 2^{(t+\log_2 t-n)L} \\ &= 2^{t^2+5t} 2^{\left(\frac{1}{2}t^2 + \frac{5}{2}t + t + \log_2 t - n\right)L} \\ &\leq 2^{\frac{1}{4}t^3 + \frac{11}{4}t^2 + 5t + (\frac{1}{2}+1)\log_2 t - nL} \\ &\leq 2^{t^3+4t^2+6t-n|\dim N_i - \dim N_j|}, \end{aligned} \quad (120)$$

where we have used again  $j \leq |\Sigma_{t,t}|$  and  $L \leq t/2$ .

Note that we can alternatively bound  $A_{i,j}$  for  $i \neq j$  using that the identity is not an allowed permutation, i. e. only permutations with less than  $j - 2$  fixed points can appear. With Eqs. (115) and (116), we get the following inequality

$$\begin{aligned}
 |A_{i,j}| &\leq \sum_{k=0}^{j-2} p(j, k) 2^{-(n-t-\log_2 t)(j-1-k)} \\
 &\leq j^3 2^{-(n-t-\log_2 t)} \\
 &\leq 2^{\frac{3}{2}t^2 + \frac{15}{2}t + \log_2 t - n} \\
 &\leq 2^{2t^2 + 10t - n}.
 \end{aligned}
 \tag{121}$$

□

### 6.3. Proof of Haar symmetrization Lemma 13.

**Lemma 13** (Haar symmetrization). *For all  $t$  and for all  $T \in \Sigma_{t,t} \setminus S_t$ , it holds that*

$$(Q_T | P_H | Q_T) = 2^{-t} \|P_H[r(T)]\|_2^2 \leq \frac{7}{8},
 \tag{101}$$

where  $Q_T$  is as in Eq. (21) and  $P_H = M_t(\mu_H)$  is the  $t$ -th moment operator of the single-qubit unitary group  $U(2)$ .

For an analysis of the tightness of the bound, see ‘‘Appendix C’’. Recall that

$$P_H[A] := \int_{U(2)} U^{\otimes t} A (U^\dagger)^{\otimes t} d\mu_H(U).
 \tag{122}$$

Let  $P_D$  be the Haar averaging operator, restricted to the diagonal unitaries. As it averages over a subgroup,  $P_D$  is a projection with range a super-set of  $P_H$ . By applying  $P_D$  to  $r(T)$ , we can turn the statement (101) from one involving *Hilbert space* geometry to one about the *discrete* geometry of stochastic Lagrangians. Indeed,

$$\begin{aligned}
 2^{-t} \|P_H[r(T)]\|_2^2 &= 2^{-t} \|P_H[P_D[r(T)]]\|_2^2 \\
 &\leq 2^{-t} \|P_D[r(T)]\|_2^2 \\
 &= 2^{-t} (r(T), P_D[r(T)]) \\
 &= 2^{-t} \sum_{(x,y) \in T} \sum_{(x',y') \in T} (|x\rangle\langle y|, P_D[|x'\rangle\langle y'|]) \\
 &= 2^{-t} \sum_{(x,y) \in T} \sum_{(x',y') \in T} (|x\rangle\langle y|, \int_0^{2\pi} e^{i2\phi(h(x')-h(y'))} |x'\rangle\langle y'| d\phi) \\
 &= 2^{-t} |\{(x, y) \in T \mid h(x) = h(y)\}| \\
 &= \Pr_{(x,y)}[h(x) = h(y)],
 \end{aligned}$$

i.e., the overlap is upper-bounded by the probability that a uniformly sampled element  $(x, y)$  of  $T$  has components of equal Hamming weight.



We will bound the probability in slightly different ways for spaces  $T$  with trivial (i.e., zero-dimensional) and non-trivial defect spaces.

*a. Case I: trivial defect sub-spaces* In this case,  $T = \{(Oy, y) \mid y \in \mathbb{F}_2^t\}$  for some orthogonal stochastic matrix  $O$ . The next proposition treats a slightly more general situation.

**Proposition 4** (Hamming bound). *Let  $O \in \text{GL}(\mathbb{F}_2^t)$ . Assume  $O$  has a column of Hamming weight  $r$ . Then the probability that  $O$  preserves the Hamming weight of a vector  $y$  chosen uniformly at random from  $\mathbb{F}_2^t$  satisfies the bound*

$$\Pr_y[h(Oy) = h(y)] \leq \frac{1}{2} + \begin{cases} 2^{-(r+1)} \binom{r+1}{(r+1)/2} & r \text{ odd} \\ 0 & r \text{ even.} \end{cases} \tag{123}$$

The bound in Eq. (123) decreases monotonically in  $r$ . Orthogonal stochastic matrices  $O$  satisfy  $r \equiv 1 \pmod{4}$ , so the smallest non-trivial  $r$  that can appear is  $r = 5$ , for which the bound gives .81.

The proof idea is as follows: For each  $y \in \mathbb{F}_2^t$ , the two vectors  $y, y + e_1$  differ in Hamming weight by  $\pm 1$ . But, if  $h(e_1) \neq 1$ , then  $h(Oy) - h(O(y + e_1))$  tends not to be  $\pm 1$ . In such cases,  $O$  does not preserve weights for *both*  $y$  and  $y + e_1$ . Applying this observation to randomly chosen vectors, we can show the existence of many vectors for which  $O$  changes the Hamming weight.

*Proof of Proposition 4.* Assume without loss of generality that the first  $r$  entries of  $Oe_1$  are 1, and the remaining  $t - r$  entries are 0.

Let  $y$  be a uniformly distributed random vector on  $\mathbb{F}_2^t$ , notice that also  $Oy$ , and  $O(y + e_1)$  are uniformly distributed. Using the union bound, we find that

$$\begin{aligned} \Pr[h(Oy) = h(y)] &= 1 - \Pr[h(Oy) \neq h(y)] \\ &= 1 - \frac{1}{2}(\Pr[h(Oy) \neq h(y)] + \Pr[h(Oy + Oe_1) \neq h(y + e_1)]) \\ &\leq 1 - \frac{1}{2}\Pr[h(Oy) \neq h(y) \vee h(Oy + Oe_1) \neq h(y + e_1)] \\ &= \frac{1}{2} + \frac{1}{2}\Pr[h(Oy) = h(y) \wedge h(Oy + Oe_1) = h(y + e_1)] \\ &\leq \frac{1}{2} + \frac{1}{2}\Pr[h(Oy) - h(Oy + Oe_1) = \pm 1]. \end{aligned}$$

We would like to compute  $\Pr[h(Oy) - h(O(y + e_1)) = \pm 1]$ . The vector  $O(y + e_1) = O(y) + O(e_1)$  arises from  $O(y)$  by flipping the first  $r$  components. This operation changes the Hamming weight by  $\pm 1$  if and only if the number of ones in the first  $r$  components of  $O(y)$  equals  $(r \pm 1)/2$ . For even  $r$ , this condition cannot be met, and correspondingly  $\Pr[h(Oy) - h(O(y + e_1)) = \pm 1] = 0$ .

In case of odd  $r$ , this probability becomes

$$\begin{aligned} \Pr[h(Oy) - h(O(y + e_1)) = \pm 1] &= 2^{-r} \binom{r}{(r-1)/2} + 2^{-r} \binom{r}{(r+1)/2} \\ &= 2^{-r} \binom{r+1}{(r+1)/2}. \end{aligned} \tag{124}$$

□

*b. Case II: non-trivial defect sub-spaces* We now turn to Lagrangians  $T$  with a non-trivial defect subspace.

**Proposition 5** (Defect Hamming bound). *Let  $\{0\} \neq N \subset \mathbb{F}_2^t$  be isotropic. There exists an  $n \in N$  such that if  $x$  is chosen uniformly at random from  $N^\perp$ , then*

$$\Pr_{x \in N^\perp}[h(x) = h(x + n)] \leq \frac{3}{4}.$$

*What is more, let  $T$  be a stochastic Lagrangian with non-trivial defect sub-spaces. Then, for an element  $(x, y)$  drawn uniformly from  $T$ , we have*

$$\Pr_{(x,y) \in T}[h(x) = h(y)] \leq \frac{7}{8}.$$

*Proof.* Let  $d = \dim N$ . Consider a  $t \times d$  column-generator matrix  $\Gamma$  for  $N$ . Permuting coordinates of  $\mathbb{F}_2^t$  and adopting a suitable basis, there is no loss of generality in assuming that  $\Gamma$  is of the form

$$\Gamma = \begin{pmatrix} G \\ \mathbb{1}_d \end{pmatrix}, \quad G \in \mathbb{F}_2^{(t-d) \times d}.$$

Note that

$$\gamma = (\mathbb{1}_{t-d}, G)$$

is a row-generator matrix for  $N^\perp$ . Indeed, the row-span has dimension  $t - d$  and the matrices fulfill

$$\gamma \Gamma = G + G = 0,$$

i.e., the inner product between any column of  $\Gamma$  and any row of  $\gamma$  vanishes. It follows that elements  $n \in N$ ,  $x \in N^\perp$  are exactly the vectors of respective form

$$n = \left( \underbrace{G\tilde{n}}_{t-d}, \underbrace{\tilde{n}}_d \right), \quad \tilde{n} \in \mathbb{F}_2^d; \quad x = \left( \underbrace{\tilde{x}}_{t-d}, \underbrace{G^T \tilde{x}}_d \right), \quad \tilde{x} \in \mathbb{F}_2^{t-d}.$$

In particular, if  $x$  is drawn uniformly from  $N^\perp$ , then the first  $t - d$  components are uniformly distributed in  $\mathbb{F}_2^{t-d}$ . For now, we restrict to the case where  $G$  has a column, say the first, with  $r \neq 1$  non-zero entries. We then choose  $n = (Ge_1, e_1)$  and argue as in Eq. (124) to obtain

$$\Pr_{x \in N^\perp}[h(x) = h(x + n)] \leq \sup_{1 \neq r \text{ odd}} 2^{-r} \binom{r+1}{(r+1)/2} = \frac{3}{4} \quad (\text{attained for } r = 3). \quad (125)$$

We are left with the case where all columns of  $G$  have Hamming weight 1. (If  $N$  is a defect subspace, then Def. 6.1 implies that every column of  $\Gamma$  has Hamming weight at least 4. We treat the present case merely for completeness). As  $N$  is isotropic, the columns of  $\Gamma$  have mutual inner product equal to 0:

$$\Gamma^T \Gamma = 0 \quad \Leftrightarrow \quad G^T G = -\mathbb{1} = \mathbb{1} \pmod{2}.$$

It follows that all columns have to be mutually orthogonal standard basis vectors  $e_i \in \mathbb{F}_2^{t-d}$ . Thus, by permutating the first  $t - d$  coordinates of  $\mathbb{F}_2^t$ , we can assume that  $G$  is of the form

$$G = \begin{pmatrix} \mathbb{1}_d \\ 0 \end{pmatrix}, \quad \Rightarrow \quad N = \{(\tilde{n} \oplus 0_{t-2d}, \tilde{n}) \mid \tilde{n} \in \mathbb{F}_2^d\}, \quad N^\perp = \{(\tilde{x}, \tilde{x}|_d) \mid \tilde{x} \in \mathbb{F}_2^{t-d}\},$$

where  $\tilde{x}|_d$  denotes the restriction of  $\tilde{x}$  to the first  $d$  components. Adding  $n := (e_1 \oplus 0, e_1)$  to  $x = (\tilde{x}, \tilde{x}|_d)$ , the Hamming weight of the two parts change both by  $\pm 1$ , giving  $h(x+n) = h(x) \pm 2$ . Thus, we have  $\Pr[h(x) = h(x+n)] = 0$ .

We have proven the first advertised claim. It implies the second one, as argued next. Let  $N$  be the left defect subspace of  $T$ . By Ref. [42, Prop. 4.17], we find the following.

- The restriction  $\{x \mid (x, y) \in T \text{ for some } y\}$  equals  $N^\perp$ .
- The stochastic Lagrangian  $T$  contains  $N \oplus 0$ .

Assume that  $(x, y)$  is distributed uniformly in  $T$ . By the first cited fact,  $x$  is distributed uniformly in  $N^\perp$ . By the second fact,  $(x+n, y)$  follows the same distribution as  $(x, y)$ , for each  $n \in N$ . Thus, repeating the argument in the proof of Proposition 4, we find that for any fixed  $n \in N$ :

$$\begin{aligned} \Pr[h(x) = h(y)] &= 1 - \Pr[h(x) \neq h(y)] \\ &\leq 1 - \frac{1}{2} \Pr[h(x) \neq h(y) \vee h(x+n) \neq h(y)] \\ &\leq \frac{1}{2} + \frac{1}{2} \Pr[h(x) = h(x+n)] \leq \frac{7}{8}. \end{aligned}$$

□

#### 6.4. Proof of Lemmas for Theorem 2.

**Lemma 5** (Relative  $\varepsilon 2^{2tn}$ -approximate Clifford  $t$ -designs). *Suppose that  $0 \leq \varepsilon < 1$  is such that  $g_{\text{Cl}}(\nu, t) \leq \varepsilon$ . Then,  $\nu$  is a relative  $\varepsilon 2^{2tn}$ -approximate Clifford  $t$ -design.*

*Proof.* This follows similar to Ref. [24, Lem. 4& Lem. 30]. Denote by  $|\Omega_{2^n}\rangle$  the maximally entangled state vector on  $\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n}$ . The condition in (5) is equivalent to

$$(1 - \varepsilon)\rho_{\text{Cl}} \leq \rho_\nu \leq (1 + \varepsilon)\rho_{\text{Cl}}, \quad (126)$$

as an operator inequality, where

$$\rho_\nu := (\Delta_\nu \otimes \mathbb{1})(|\Omega_{2^n}\rangle\langle\Omega_{2^n}|)^{\otimes t} \quad \text{and} \quad \rho_{\text{Cl}} := \rho_{\mu_{\text{Cl}}}. \quad (127)$$

We have a decomposition of  $(\mathbb{C}^{2^n})^{\otimes t}$  into irreducible representations of the Clifford group:

$$(\mathbb{C}^{2^n})^{\otimes t} \cong \bigoplus_{\gamma} C_\gamma \otimes L_\gamma, \quad (128)$$

where  $\{C_\gamma\}$  is the set of all equivalence classes of irreducible representations of  $\text{Cl}(n)$  that appear in the  $t$ -th order diagonal representation, and  $L_\gamma$  are the corresponding multiplicity spaces (which by the double commutant theorem are irreducible representations of the commutant algebra –we have chosen  $L$  for Lagrangian). This implies that

$$|\Omega_{2^n}\rangle^{\otimes t} \cong \sum_{\gamma} \sqrt{\frac{\dim L_\gamma \dim C_\gamma}{2^{nt}}} |\gamma, \gamma\rangle \otimes |\Omega_{C_\gamma}\rangle \otimes |\Omega_{L_\gamma}\rangle, \quad (129)$$

where  $|\Omega_{L_\gamma}\rangle$  and  $|\Omega_{C_\gamma}\rangle$  denote maximally entangling state vectors on two copies of  $L_\gamma$  and  $C_\gamma$ , respectively. Indeed, observe that  $|\Omega_{2^n}\rangle^{\otimes t} = 2^{-nt/2} \text{vec}(\mathbb{1})$  and that the identity restricted to sub-spaces is just the identity on these sub-spaces. The prefactors then follow from normalizing the vectorized identity operators on the direct summands.

Since  $\text{Cl}(n)$  acts via multiplication on the spaces  $C_\lambda$ , this implies that

$$\begin{aligned} \rho_{\text{Cl}} &= \int_{\text{Cl}(n)} (U \otimes \mathbb{1})^{\otimes t} (|\Omega_{2^n}\rangle \langle \Omega_{2^n}|)^{\otimes t} (U^\dagger \otimes \mathbb{1})^{\otimes t} d\mu_{\text{Cl}}(U) \\ &\cong \sum_{\gamma} \frac{\dim L_\gamma \dim C_\gamma}{2^{nt}} (|\gamma\rangle \langle \gamma|)^{\otimes 2} \otimes \left( \frac{\mathbb{1}_{C_\gamma}}{\dim C_\gamma} \right)^{\otimes 2} \otimes |\Omega_{L_\gamma}\rangle \langle \Omega_{L_\gamma}|, \end{aligned} \quad (130)$$

where the second line follows from Schur's lemma and the fact that  $\int U^{\otimes t} \bullet (U^\dagger)^{\otimes t}$  is trace preserving. The support of this operator is on the *symmetric subspace*  $\vee^t(\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n})$  [24, Lem 30.1]. The minimal eigenvalue of this operator restricted to the symmetric subspace is

$$\min_{\gamma} \frac{\dim L_\gamma}{2^{nt} \dim C_\gamma}, \quad (131)$$

which we now lower bound. Let  $\gamma^*$  denote the optimizer. By Schur-Weyl duality, the diagonal action of  $U(2^n)$  on  $(\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n})^{\otimes t}$  decomposes as  $\bigoplus_{\lambda} U_\lambda \otimes S_\lambda$  where as usual  $U_\lambda$  are Weyl modules and  $S_\lambda$  are Specht modules. Restricting this action to the Clifford group, the  $U_\lambda$  further decompose into irreducible representations

$$U_\lambda \simeq \bigoplus_{\gamma \in I_\lambda} C_\gamma \otimes \mathbb{C}^{d_{\lambda, \gamma}},$$

where  $I_\lambda$  is the spectrum of  $U_\lambda$  as a Clifford representation. Let  $\Lambda_0$  be the set of all  $\lambda$  such that  $\gamma^* \in I_\lambda$ , then as a Clifford representation

$$(\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n})^{\otimes t} \simeq C_{\gamma^*} \otimes \left( \bigoplus_{\lambda \in \Lambda_0} S_\lambda \otimes \mathbb{C}^{d_{\lambda, \gamma^*}} \right) \oplus (\text{other irreducible representations}). \quad (132)$$

Thus, as a vector space, we have

$$L_{\gamma^*} = \bigoplus_{\lambda \in \Lambda_0} S_\lambda \otimes \mathbb{C}^{d_{\lambda, \gamma^*}}. \quad (133)$$

In particular, for any  $\lambda \in \Lambda_0$  we have that  $\dim C_{\gamma^*} \leq \dim U_\lambda$  and  $\dim L_{\gamma^*} \geq \dim S_\lambda$ . Thus we get the following bound for the minimal eigenvalue:

$$\frac{\dim L_{\gamma^*}}{2^{nt} \dim C_{\gamma^*}} \geq \min_{\lambda \in \text{Part}(t, 2^n)} \frac{\dim S_\lambda}{2^{nt} \dim U_\lambda} \geq 2^{-2nt}. \quad (134)$$

The rest of the proof follows as in Ref. [24, Lem. 4], mutatis mutandis.  $\square$

In order to prove Lemma 8 we make use of the following result by [61] and Lemma 11 bounding certain sums of overlaps of the operators  $r(T)$ .

**Lemma 14** (Nachtergaele [61, Thm. 3]). *Let  $H_{[p,q]}$  for  $[p, q] \subset [n] = \{1, \dots, n\}$  be a family of positive semi-definite Hamiltonians with support on  $(\mathbb{C}^2)^{\otimes(q-p+1)} \subset (\mathbb{C}^2)^{\otimes n}$ . Assume there is a constant  $l \in \mathbb{N}$ , such that the following conditions hold:*

1. *There is a constant  $d_l > 0$  for which the Hamiltonians satisfy*

$$0 \leq \sum_{q=l}^n H_{[q-l+1,q]} \leq d_l H_{[1,n]}. \quad (135)$$

2. *There are  $Q_l \in \mathbb{N}$  and  $\gamma_l > 0$  such that there is a local spectral gap:*

$$\Delta(H_{[q-l+1,q]}) \geq \gamma_l, \quad \forall q \geq Q_l. \quad (136)$$

3. *Denote the ground state projector of  $H_{[p,q]}$  by  $G_{[p,q]}$ . There exist  $\varepsilon_l < 1/\sqrt{l}$  such that*

$$\|G_{[q-l+2,q+1]}(G_{[1,q]} - G_{[1,q+1]})\|_\infty \leq \varepsilon_l, \quad \forall q \geq Q_l. \quad (137)$$

Then, it holds that

$$\Delta(H_{[1,n]}) \geq \frac{\gamma_l}{d_l} \left(1 - \varepsilon_l \sqrt{l}\right)^2. \quad (138)$$

While conditions 1) and 2) are merely translation-invariance with finite range of interactions and frustration-freeness in disguise, the third condition is highly non-trivial and involves knowledge of the ground-space structure. Usually, finding the ground space in a basis can be just as hard as computing the spectral gap in the first place. Fortunately, the ground space structure of the Hamiltonians  $H_{n,t}$  is determined by the representation theory of the Clifford group. With little additional work, we obtain the following lemma about the ground space structure of our Hamiltonians.

**Lemma 8** (Lower bound to spectral gap). *Let the Hamiltonian  $H_{n,t}$  be as in Eq. (72) and assume that  $n \geq 12t$ . Then,  $H_{n,t}$  has a spectral gap satisfying*

$$\Delta(H_{n,t}) \geq \frac{\Delta(H_{12t,t})}{48t}. \quad (75)$$

*Proof.* We make use of the Nachtergaele lemma. We have to verify the three conditions of Lemma 14. As already stated in Ref. [61], the first two conditions hold directly for translation-invariant local Hamiltonians as in our case.

1. The first condition immediately follows from the fact that we consider a translation-invariant 2-local Hamiltonian. It is fulfilled for any choice of  $l \geq 2$  and  $d_l = l - 1$ .
2. The second condition follows again for all  $l \geq 2$  and the choice  $Q_l = l$ , since  $H_{[q-l+1,q]}$  is a sum of positive semi-definite operators for all  $q \geq l$  with spectrum that does not depend on  $q$  due to translation-invariance. Thus, we can set

$$\gamma_l := \Delta(H_{[q-l+1,q]}) > 0. \quad (139)$$

3. The third condition requires a calculation and a non-trivial choice of  $l$ . We have to bound the quantity

$$R_{q,l} := \left\| G_{[q-l+2,q+1]} (G_{[1,q]} - G_{[1,q+1]}) \right\|_{\infty}, \quad (140)$$

for all  $q \geq Q_l = l$ . Here,  $G_{[p,q]}$  denotes the orthogonal projector onto the ground space of  $H_{[p,q]}$ . Note that this ground space is simply a suitable translation of the Clifford commutant  $\text{Cl}(k)'$  for  $k = q - p + 1$  as shown in Lemma 7. Recall that it comes with a non-orthogonal basis  $Q_T^{\otimes k}$ , where

$$Q_T := \frac{r(T)}{\|r(T)\|_2} = 2^{-t/2} r(T), \quad T \in \Sigma_{t,t}. \quad (141)$$

Moreover, the projector  $G_{[p,q]}$  is also simply a translation of the Clifford projector  $P_{\text{Cl}(k)}$  projecting onto  $\text{Cl}(k)'$ . From the discussion in Sect. 6.1, we know that the Clifford frame operator

$$S_{\text{Cl}(k)} := \sum_T |Q_T\rangle\langle Q_T|^{\otimes k}, \quad (142)$$

is a suitable approximation to  $P_{\text{Cl}(k)}$  when  $k$  is large enough. Concretely, we have by Lem. 12:

$$\left\| S_{\text{Cl}(k)} - P_{\text{Cl}(k)} \right\|_{\infty} \leq (-2^{-k}; 2)_{t-1} - 1. \quad (143)$$

Defining the shorthand notation  $s_t(k) = (-2^{-k}; 2)_{t-1}$ , we in particular get the bound

$$\left\| S_{\text{Cl}(k)} \right\|_{\infty} \leq \left\| S_{\text{Cl}(k)} - P_{\text{Cl}(k)} \right\|_{\infty} + \left\| P_{\text{Cl}(k)} \right\|_{\infty} \leq s_t(k), \quad (144)$$

Let us introduce the shorthand notation  $G_q := G_{[1,q]} \equiv P_{\text{Cl}(q)}$ ,  $S_q = S_{[1,q]} \equiv S_{\text{Cl}(q)}$ , and  $G_{q,l} := G_{[q-l+2,q+1]}$ ,  $S_{q,l} := S_{[q-l+2,q+1]}$  for translations of the Clifford projector and frame operator, respectively. Notice that  $G_q - G_{q+1}$  is an orthogonal projector as the support of  $G_{q+1}$  is by definition contained in that of  $G_q$ . Therefore, restricted to the support of  $G_q$ , the operator  $G_q - G_{q+1}$  projects onto the orthogonal complement of the support of  $G_{q+1}$ . Combining this fact with the above inequalities, we find

$$\begin{aligned} R_{q,l} &= \left\| G_{q,l} (G_q - G_{q+1}) \right\|_{\infty} \\ &\leq \left\| (G_{q,l} - S_{q,l})(G_q - G_{q+1}) \right\|_{\infty} + \left\| S_{q,l}(G_q - G_{q+1}) \right\|_{\infty} \\ &\leq s_t(l) - 1 + \left\| S_{q,l}(S_q - S_{q+1}) \right\|_{\infty} + \left\| S_{q,l}(G_q - S_q) \right\|_{\infty} \\ &\quad + \left\| S_{q,l}(G_{q+1} - S_{q+1}) \right\|_{\infty} \\ &\leq \left\| S_{q,l}(S_q - S_{q+1}) \right\|_{\infty} + s_t(l) - 1 + s_t(l) (s_t(q) + s_t(q+1) - 2) \\ &\stackrel{q \geq l}{\leq} \left\| S_{q,l}(S_q - S_{q+1}) \right\|_{\infty} + (s_t(l) - 1) (2s_t(l) + 1) \\ &= \left\| \sum_{T \in \Sigma_{t,t}} |Q_T\rangle\langle Q_T|^{\otimes(q-l+1)} \otimes Y_T \right\|_{\infty} + (s_t(l) - 1) (2s_t(l) + 1), \quad (145) \end{aligned}$$

where the operator  $Y_T$  can be straightforwardly computed as

$$Y_T := \sum_{T' \neq T} \left( (|Q_{T'}\rangle\langle Q_T|)^{l-1} |Q_{T'}\rangle\langle Q_T|^{\otimes(l-1)} \right) \otimes \left( |Q_{T'}\rangle\langle Q_{T'}| (\text{id} - |Q_T\rangle\langle Q_T|) \right). \tag{146}$$

Invoking the synthesis operators

$$V_k = \sum_T \left| Q_T^{\otimes k} \right\rangle\langle e_T | : \mathbb{C}^{|\Sigma_{t,t}|} \longrightarrow \text{Cl}(k)', \tag{147}$$

introduced in Lemma 12, one can bound the above norm as

$$\begin{aligned} \left\| \sum_T |Q_T\rangle\langle Q_T|^{\otimes(q-l+1)} \otimes Y_T \right\|_\infty &= \left\| \sum_T V_{q-l+1} |e_T\rangle\langle e_T| V_{q-l+1}^\dagger \otimes Y_T \right\|_\infty \\ &\leq \left\| V_{q-l+1} V_{q-l+1}^\dagger \right\|_\infty \left\| \sum_T |e_T\rangle\langle e_T| \otimes Y_T \right\|_\infty \\ &= \left\| S_{q-l+1} \right\|_\infty \max_T \|Y_T\|_\infty \\ &\leq s_t(q-l+1) (s_t(l-1) - 1). \end{aligned} \tag{148}$$

Thus, we arrive at

$$\begin{aligned} R_{q,l} &\leq s_t(q-l+1) (s_t(l-1) - 1) + (s_t(l) - 1) (2s_t(l) + 1) \\ &\leq s_t(1) (s_t(l-1) - 1) + (s_t(l) - 1) (2s_t(l) + 1). \end{aligned} \tag{149}$$

For  $l+1 \geq t + \log_2(t)$ , we can use Lemma 11 to get:

$$\begin{aligned} R_{q,l} &\leq t2^{t-l+1} (1 + t2^{t-1}) + t2^{t-l} (3 + t2^{t-l}) \\ &= t^2 2^{2t-l} \left( \frac{5}{t} 2^{-t} + 2^{-l} + 1 \right) \\ &\leq 4t^2 2^{2t-l}. \end{aligned} \tag{150}$$

Finally choose any  $l \geq 4t + 4 \log_2(t) + 6$ , then we find

$$l \leq \frac{4^{l-2t}}{64t^2} \Rightarrow R_{q,l} \leq 4t^2 2^{2t-l} \leq \frac{1}{2\sqrt{l}} < \frac{1}{\sqrt{l}}, \quad \forall q \geq l. \tag{151}$$

In particular, we can choose  $l = 12t$ ,  $\varepsilon_l = 1/2\sqrt{l}$  to get the desired bound in Lemma 14  $\forall q \geq l$ .

Hence, for the choices  $l = 12t$ ,  $d_l = l - 1$ ,  $Q_l = l$ ,  $\gamma_l = \Delta(H_{12t,t})$  and  $\varepsilon_l = 1/2\sqrt{l}$ , Lemma 14 gives the claimed bound on the spectral gap:

$$\Delta(H_{n,t}) \geq \frac{\gamma_l}{d_l} (1 - \varepsilon_l^2 \sqrt{l}) \geq \frac{\Delta(H_{12t,t})}{48t}. \tag{152}$$

□

## 7. Summary and Open Questions

We have found that a number of non-Clifford gates independent of the system size suffices to generate  $\varepsilon$ -approximate unitary  $t$ -designs. This is surprising, conceptually interesting and practically relevant: After all, it is the main objective in quantum gate synthesis to minimize the number of non-Clifford gates in a circuit implementation of a given unitary. There are multiple open questions and ways to continue this work:

- Similar to the result in Ref. [24], the scaling in  $n$  is near to optimal, the scaling in  $t$  can probably be improved.
- Another natural open question is whether the condition  $n = O(t^2)$  can be lifted. Notably, this is reminiscent to the situation discussed in Ref. [69], where the improved scaling can be proven only in the regime  $t = o(n^{\frac{1}{2}})$ . In this work, the condition  $n = O(t^2)$  is related to the approximate orthogonality of the Lagrangian subspace. We use this fact repeatedly and in different flavours, but we can only prove it in this regime. In fact, in Lemma 12 we use the same technique that has been used in Ref. [24] to prove approximate orthogonality of permutations in the regimes  $t \leq 2^{O(0.4n)}$ . However, the commutant of the Clifford group is far larger than the span of permutations and we suspect that this bound is tight. Nevertheless, we cannot rule out that similar results can be proven without exploiting approximate orthogonality. This likely requires a detailed understanding of the representation theory of the Clifford group.
- Our result holds for additive errors in the diamond norm. For relative errors, our bounds can be used to obtain a quadratic advantage in the number of non-Clifford gates in Corollary 1. This still allows the density of non-Clifford gates to go to zero in the thermodynamic limit, but is not system-size independent anymore. In fact, it has been proven in Ref. [70] that this scaling is optimal for relative errors. It would be interesting to delineate more precisely for which notions of approximations a system-size independent result holds.
- We strongly expect that the results can be generalized to qudits for arbitrary  $d$ , giving rise to analogous conclusions concerning an independence of the system size for additive errors in the diamond norm.

We hope the present work stimulates such endeavors.

*Acknowledgements.* We would like to thank Richard Kueng, Lorenz Mayer and Adam Sawicki for helpful discussions. Moreover, we would like to thank Nick Hunter-Jones for pointing out the application presented in “Appendix D”. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 - 390534769, the ARO under contract W911NF-14-1-0098 (Quantum Characterization, Verification, and Validation), and the DFG (SPP1798 CoSIP, project B01 of CRC 183). The Berlin group has been supported in this work by the DFG (SPP1798 CoSIP, projects B01 and A03 of CRC 183, FOR 2724 and EI 519/14-1), the Einstein Research Foundation (Einstein Research Unit on quantum devices) and the Templeton Foundation. This work has also received funding from the European Union’s Horizon2020 research and innovation programme under grant agreement No. 817482 (PASQuanS). The research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory



regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Declarations

**Data Availability Statement** No data was generated in this work.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Appendix A: Unitary $t$ -designs

In the following, we review the concept of a *unitary  $t$ -design* [5–7], giving different but equivalent definitions which prove to be useful in different contexts. They also serve as starting point to explore connections to other mathematical fields, e. g. representation theory. To this end, let us introduce some notation. Define  $\mu_{\text{H}}$  to be the (normalized) Haar measure on  $U(d)$  and let  $\text{Hom}_{(t,t)}(U(d))$  be the space of homogeneous polynomials of degree  $t$  in both the entries of  $U \in U(d)$  as well as  $\bar{U}$ .

**Definition 10** (*Unitary  $t$ -design*). A probability measure  $\nu$  on  $U(d)$  is called a *unitary  $t$ -design* if the following holds for all  $p \in \text{Hom}_{(t,t)}(U(d))$ :

$$\int_{U(d)} p(U) d\nu(U) = \int_{U(d)} p(U) d\mu_{\text{H}}(U). \quad (\text{A1})$$

A subset  $D \subseteq U(d)$  is called a *unitary  $t$ -design*, if it comes with a probability measure  $\nu_D$  which, continued trivially to  $U(d)$ , is a unitary  $t$ -design. In particular, if  $D$  is finite,  $\nu_D$  is usually taken to be the (normalized) counting measure.

It might not come as a surprise that Def. 10 has not to be checked for any polynomial. Since any homogeneous polynomial  $p \in \text{Hom}_{(t,t)}(U(d))$  can be linearized as

$$p(U) = \text{Tr}(AU^{\otimes t,t}), \quad U^{\otimes t,t} := U^{\otimes t} \otimes \bar{U}^{\otimes t}, \quad (\text{A2})$$

the defining Eq. (A1) becomes

$$M_t(\nu) := \int_{U(d)} U^{\otimes t,t} d\nu(U) = \int_{U(d)} U^{\otimes t,t} d\mu_{\text{H}}(U) =: M_t(\mu_{\text{H}}). \quad (\text{A3})$$

Thus  $\nu$  is a unitary  $t$ -design if and only if its moment operator  $M_t(\nu)$  agrees with the one of the Haar measure. Note that the operators  $U^{\otimes t,t}$  are the matrix representation of the  $t$ -diagonal adjoint action  $\text{Ad}(U^{\otimes t}) = U^{\otimes t} \bullet (U^\dagger)^{\otimes t}$  with respect to the standard basis  $|i\rangle\langle j|$  of  $L(\mathbb{C}^d)$ . Thus, this can be equivalently stated as equality of the twirls  $M_t(\nu) = M_t(\mu_{\text{H}})$  over the two measures.

A particularly fruitful theory of designs is possible in the case where the design  $(G, \nu)$  itself constitutes a (locally compact) subgroup  $G \subseteq U(d)$  and  $\nu$  is the normalized Haar measure on  $G$ . Following Ref. [38], we call these *unitary  $t$ -groups*. In this case, we see that Eq. (A3) implies that the trivial isotype of the representation  $G \ni g \mapsto \text{Ad}_g^{\otimes t}$  shall agree with the trivial isotype of  $U(d) \ni U \mapsto \text{Ad}_U^{\otimes t}$ . Since the trivial isotype exactly corresponds to the commutant of the respective diagonal representations  $\tau_t : U \mapsto U^{\otimes t}$ , this is equivalent to the statement that the commutant of the representation  $\tau_t$  agrees with the commutant of the restriction  $\tau_t|_G$ . However, this is the case if and only if  $\tau_t|_G$  decomposes into the same irreducible representations as  $\tau_t$ .

## Appendix B: Representations of the Unitary Group

The representation theory of the unitary group can be understood using the theory of highest weight for compact Lie groups, see, for example Refs. [71–73]. We present a short summary of the part relevant to us here. Let  $\rho$  be an irreducible representation of  $U(d)$ , and consider the restriction  $\rho|_{D(d)}$  to the diagonal subgroup  $D(d) \simeq (S^1)^{\times d}$  (which is a so-called *maximal torus* in  $U(d)$ ). In general, this is a reducible representation of  $D(d)$ . Since  $D(d)$  is Abelian,  $\rho|_{D(d)}$  decomposes into one-dimensional irreducible representations, i. e. characters of  $D(d) \simeq (S^1)^{\times d}$ . Those are of the form  $\chi_u(\theta) := e^{iu^T\theta}$  for some vector  $u \in \mathbb{Z}^d$ , and thus we find

$$\rho|_{D(d)} \simeq \bigoplus_{u \in \mathbb{Z}^d} \chi_u \otimes \mathbb{1}_{m_u}, \quad (\text{B1})$$

where  $m_u \in \mathbb{N}$  are multiplicities. The vectors  $u$  for which  $m_u \neq 0$  are called the *weights* of  $\rho$ . Introducing a lexicographical ordering of the weights, we call a weight  $u$  higher than the weight  $v$  if  $u > v$ . The *theorem of the highest weight* states that any irreducible representation  $\rho$  has a highest weight and that irreducible representations with the same highest weight are isomorphic. Thus, irreducible representations are unambiguously labeled by their highest weight. Next, let us consider the tensor product  $\pi_u \otimes \pi_v$  of two irreducible representations labeled by their highest weights  $u$  and  $v$ . One can easily check that the weights of irreducible representations in  $\pi_u \otimes \pi_v$  have to be sums of weights of  $\pi_u$  and  $\pi_v$ . In particular, the highest weight of all irreducible representations is at most  $u + v$ .

As a relevant example consider the (irreducible) defining representation  $\rho : U \mapsto U$  of  $U(2)$ . Its restriction to the diagonal subgroup  $S^1 \times S^1$  decomposes as

$$\rho|_{S^1 \times S^1} \simeq \chi_{e_1} \oplus \chi_{e_2},$$

with highest weight  $e_1 = (1, 0)$ . Using  $\bar{\chi}_u = \chi_{-u}$ , the highest weight of the complex conjugate representation  $\bar{\rho} : U \mapsto \bar{U}$  can be immediately determined as  $(0, -1)$ . Hence, the weights of  $\rho \otimes \bar{\rho}$  are  $\{(0, 0), (1, -1), (-1, 1)\}$ . Here,  $(0, 0)$  is the highest weight of the trivial irreducible representation and  $(1, -1)$  the highest weight of the adjoint irrep. Finally, all irreducible representations appearing in  $(\rho \otimes \bar{\rho})^{\otimes t}$  have weights  $w$  satisfying  $(-t, t) \leq w \leq (t, -t)$  and, in particular,

$$w = \sum_{i=1}^t u_i$$

where  $u_i \in \{(0, 0), (1, -1), (-1, 1)\}$ . It follows that the Euclidean norm of these weights is at most  $\sqrt{2}t$ .

## Appendix C: Converse Bounds for Estimates in Sect. 6.3

Here, we collect various tightness results that limit the degree by which the estimates in Sect. 6.3 can be improved. The bound in Proposition 4 is tight in many cases. Most interestingly, the anti-identity [42]

$$\bar{\mathbb{1}} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \in \mathcal{O}_t, \quad (\text{C1})$$

meets the bound if both

$$r = t - 1 \quad \text{and} \quad t/2 = (r + 1)/2 \quad \text{are odd.} \tag{C2}$$

Indeed, the anti-identity flips the components of the input if its parity is odd, and leaves the input invariant if the parity is even. The flipping step preserves the Hamming weight if and only if  $h(a) = t/2$ . Thus

$$\begin{aligned} \Pr[h(Oa) = h(a)] &= \Pr[h(a) \text{ even}] + \Pr[h(a) \text{ odd} \wedge h(a) = t/2] \\ &= \Pr[h(a) \text{ even}] + \Pr[h(a) = t/2] \quad \text{(using (C2))} \\ &= \frac{1}{2} + 2^{-t} \binom{t}{t/2} \\ &= \frac{1}{2} + 2^{-(r+1)} \binom{r+1}{(r+1)/2}. \end{aligned}$$

Likewise, both estimates in Proposition 5 are tight. The first bound is saturated for  $N = \{0, (1, 1, 1, 1)\}$ . Indeed,  $N^\perp$  is the space of all even-weight elements of  $\mathbb{F}_2^4$ . The only non-trivial element of  $N$  is  $(1, 1, 1, 1)$  and adding it to an even-weight vector changes its weight if and only if the vector is in  $N$  itself. But  $|N|/|N^\perp| = 1/4$ . In an exactly analogous way, the second bound is tight for the stochastic Lagrangian with left and right defect spaces equal to the same  $N$ . As detailed in Example 4.27 of Ref. [42], this stochastic Lagrangian is the one identified in Ref. [74] as the sole non-trivial one in case of  $t = 4$ .

In contrast, we do not know (but suspect) that we pay a price by restricting from the full Haar symmetrizer to the one over diagonal matrices in Eq. (123). For the two cases that saturate the bounds in Proposition 4 and Proposition 5, we can compute the full projection explicitly and show that at least there, Eq. (123) indeed fails to be tight.

One can expand the anti-id  $\bar{\mathbb{1}}$  in terms of Pauli operators [42]

$$\bar{\mathbb{1}} = \frac{1}{2}(\mathbb{1}^{\otimes t} + X^{\otimes t} + Y^{\otimes t} + Z^{\otimes t}). \tag{C3}$$

Then

$$\begin{aligned} 2^{-t}(r(\bar{\mathbb{1}}), P_H[r(\bar{\mathbb{1}})]) &= 2^{-t} \int \text{Tr } r(\bar{\mathbb{1}}) U^{\otimes t} r(\bar{\mathbb{1}})^\dagger (U^\dagger)^{\otimes t} \, d\mu_H(U) \\ &= 2^{-t-2} \sum_{i,j=0}^3 \int \text{Tr } \sigma_i^{\otimes t} U^{\otimes t} \sigma_j^{\otimes t} (U^\dagger)^{\otimes t} \, d\mu_H(U) \\ &= 2^{-t-2} \sum_{i,j} \int \left( \text{Tr } \sigma_i U \sigma_j U^\dagger \right)^t \, d\mu_H(U) \\ &= 2^{-2} + 2^{-t-2} \sum_{i,j \neq 0} \int \left( \text{Tr } \sigma_i U \sigma_j U^\dagger \right)^t \, d\mu_H(U) \\ &= 2^{-2} + 2^{-2} 9 \frac{1}{4\pi} \int_{S^2} x_1^t \, dx \\ &= \frac{1}{4} + \frac{9}{4} \frac{1}{4\pi} \frac{4\pi}{1+t} = \frac{1}{4} \left( 1 + \frac{9}{t+1} \right), \end{aligned} \tag{C4}$$

where in (C4), we have interpreted the Haar integral over inner products of Paulis as an integral over the Bloch sphere and in the next line, used the formula from [75]. For  $t = 2$ , Eq. (C1) is just the swap operator (i.e., a permutation), and the formula gives 1, as it should. The smallest non-trivial case is  $t = 6$  [42], where we get roughly  $0.571 < 0.65$ .

Next, we consider the CSS code  $P_N$  for  $N = (1, 1, 1, 1)$ . We use the results in Sect. 3 of Ref. [74]. For a given partition  $\lambda$ , let  $W_\lambda$  be the associated Weyl module and  $S_\lambda$  the Schur module. As in Ref. [74], let  $W_\lambda^+ \subset W_\lambda$  be the subspace such that

$$(W_\lambda \otimes S_\lambda) \cap \text{range } P_N = W_\lambda^+ \otimes S_\lambda.$$

For the projection operators onto the various spaces, we write  $P_\lambda$  (Schur module),  $Q_\lambda$  (Weyl module), and  $Q_\lambda^+$  (the subspace defined above). Then [74]

$$P_N = \sum_{\lambda} Q_\lambda^+ \otimes P_\lambda.$$

By Schur's Lemma,

$$P_H[P_N] = \sum_{\lambda} c_\lambda Q_\lambda \otimes P_\lambda,$$

for suitable coefficients  $c_\lambda$ , which are seen to equal  $c_\lambda = D_\lambda^+/D_\lambda$  by the fact that Haar averaging preserves the trace. Hence, using Table 1 of Ref. [74] for  $d = 2$ ,

$$2^{-t+2 \dim N} (P_N, P_H[P_N]) = 2^{-2} \sum_{\lambda} \frac{d_\lambda (D_\lambda^+)^2}{D_\lambda} = \frac{7}{10} < \frac{7}{8}.$$

## Appendix D: Saturation of Higher Rényi-entropies in $K$ -interleaved Clifford Circuits

Consider the Rényi-entropies which are defined as

$$S_\alpha(\rho) := \frac{1}{1-\alpha} \log \text{Tr}[\rho^\alpha] \quad (\text{D1})$$

for  $\alpha > 0$ . For  $\alpha \searrow 1$  the standard von Neumann entropy is recovered. Here, we are interested in the entanglement properties of random state vectors  $|\psi\rangle$  on  $n$  qubits. We consider a bi-partition of the  $n$  qubits into a set  $A$  consisting of constantly many qubits  $n_A$  and a set  $B$  of  $n_B = n - n_A$  many qubits that constitutes the complement of  $A$ . To derive concentration bounds on these quantities over random ensembles of states, we study the ‘‘higher purities’’  $\text{Tr}[\rho^\alpha]$  for positive integer  $\alpha$  in more detail. First, we compute

the Haar average of this quantity. Let  $\pi_{\text{cyc}} \in S_\alpha$  be any full  $\alpha$ -cycle. We compute

$$\begin{aligned}
\mathbb{E}_{U \sim \mu_H} \text{Tr}[\rho_A^\alpha] &= \mathbb{E}_{U \sim \mu_H} \text{Tr} \left[ \text{Tr}_B [|\psi\rangle\langle\psi|]^\alpha \right] \\
&= \mathbb{E}_{U \sim \mu_H} \text{Tr} \left[ r(\pi_{\text{cyc}})_A \otimes \mathbb{1}_B (|\psi\rangle\langle\psi|)^{\otimes \alpha} \right] \\
&= \binom{2^n + \alpha - 1}{\alpha}^{-1} \text{Tr} \left[ r(\pi_{\text{cyc}})_A \otimes \mathbb{1}_B P_{\text{sym}, \alpha} \right] \\
&= \binom{2^n + \alpha - 1}{\alpha}^{-1} \alpha!^{-1} \sum_{\sigma \in S_\alpha} \text{Tr} \left[ r(\pi_{\text{cyc}} \circ \sigma)_A \otimes r(\sigma)_B \right] \\
&= \binom{2^n + \alpha - 1}{\alpha}^{-1} \alpha!^{-1} \sum_{\sigma \in S_\alpha} 2^{n_A \# \text{cyc}(\pi_{\text{cyc}} \circ \sigma)} 2^{n_B \# \text{cyc}(\sigma)} \\
&= \frac{1}{2^n (2^n + 1) \dots (2^n + \alpha - 1)} \sum_{\sigma \in S_\alpha} 2^{n_A \# \text{cyc}(\pi_{\text{cyc}} \circ \sigma)} 2^{n_B \# \text{cyc}(\sigma)} \\
&= \frac{2^{\alpha n_B} 2^{n_A}}{2^n (2^n + 1) \dots (2^n + \alpha - 1)} + O(2^{-n_B}) \\
&= 2^{-(\alpha-1)n_A} + O(2^{-n_B}),
\end{aligned} \tag{D2}$$

where  $O(2^{-n_B})$  depends on  $\alpha$ . Therefore, up to an exponentially small correction, the average higher purity is minimal.

Next, we compute the same average over an additive  $\varepsilon$ -approximate unitary  $t$ -design. Recall that this is a probability distribution  $\nu$  such that

$$\|M_t(\nu) - M_t(\mu_H)\|_\diamond \leq \varepsilon. \tag{D3}$$

By definition of the diamond norm, this also implies

$$\|M_t(\nu) - M_t(\mu_H)\|_{1 \rightarrow 1} \leq \varepsilon. \tag{D4}$$

From this, we obtain

$$\begin{aligned}
\mathbb{E}_{U \sim \nu} \text{Tr}[\rho_A^\alpha] &= \mathbb{E}_{U \sim \nu} \text{Tr} \left[ \text{Tr}_B [|\psi\rangle\langle\psi|]^\alpha \right] \\
&= \text{Tr} \left[ r(\pi_{\text{cyc}})_A \otimes \mathbb{1}_B \mathbb{E}_{U \sim \nu} (|\psi\rangle\langle\psi|)^{\otimes \alpha} \right] \\
&\leq \text{Tr} \left[ r(\pi_{\text{cyc}})_A \otimes \mathbb{1}_B \mathbb{E}_{U \sim \mu_H} (|\psi\rangle\langle\psi|)^{\otimes \alpha} \right] \\
&\quad + \left| \text{Tr} \left[ r(\pi_{\text{cyc}})_A \otimes \mathbb{1}_B (M_t(\nu) - M_t(\mu_H)) \left[ (|\psi_0\rangle\langle\psi_0|)^{\otimes \alpha} \right] \right| \right| \\
&\leq \text{Tr} \left[ r(\pi_{\text{cyc}})_A \otimes \mathbb{1}_B \mathbb{E}_{U \sim \mu_H} (|\psi\rangle\langle\psi|)^{\otimes \alpha} \right] \\
&\quad + \left\| (M_t(\nu) - M_t(\mu_H)) \left[ (|\psi\rangle\langle\psi|)^{\otimes \alpha} \right] \right\|_1 \\
&\leq 2^{-(\alpha-1)n_A} + O(2^{-n}) + \varepsilon.
\end{aligned} \tag{D5}$$

It suffices to insert  $C(K) \log^2(t)(t^4 + t \log(1/\varepsilon))$  non-Clifford gates into random Clifford circuits to generate an additive  $\varepsilon$ -approximate  $t$ -designs. Therefore, we can choose  $\varepsilon = 2^{-2(\alpha-1)n_A}$  and  $t = \alpha$  and find that a  $K$ -interleaved Clifford circuit with  $k = C(K) \log^2(\alpha)(\alpha^4 + 2(\alpha-1)n_A)$  satisfies

$$\begin{aligned}
\mathbb{E}_{U \sim \sigma^{*k}} \text{Tr}[\rho_A^\alpha] &\leq (1 - 2^{-(\alpha-1)n_A}) 2^{-(\alpha-1)n_A} + O(2^{-n}) \\
&\leq (1 - 2^{-(\alpha-1)n_A} - O(2^{-n})) 2^{-(\alpha-1)n_A}.
\end{aligned} \tag{D6}$$

Therefore, for every constant  $n_A$  and  $\alpha$ , there is a classically simulable ensemble of quantum circuits that generate essentially minimal higher purities on average.

## References

1. Emerson, J., Alicki, R., Życzkowski, K.: Scalable noise estimation with random unitary operators. *J. Opt. B* **7**, S347–S352 (2005)
2. Magesan, E., Gambetta, J.M., Emerson, J.: Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012)
3. Knill, E., Leibfried, D., Reichle, R., Britton, J., Blakestad, R.B., Jost, J.D., Langer, C., Ozeri, R., Seidelin, S., Wineland, D.J.: Randomized benchmarking of quantum gates. *Phys. Rev. A* **77**, 012307 (2008)
4. Hayden, P., Preskill, J.: Black holes as mirrors: quantum information in random subsystems. *JHEP* **0709**, 120 (2007)
5. Dankert, C., Cleve, R., Emerson, J., Livine, E.: Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 012304 (2009)
6. Dankert, C.: MSc thesis, University of Waterloo (2005), [arXiv:quant-ph/0512217](https://arxiv.org/abs/quant-ph/0512217)
7. Gross, D., Audenaert, K., Eisert, J.: Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.* **48**, 052104 (2007)
8. Ambainis, A., Bouda, J., Winter, A.: Nonmalleable encryption of quantum information. *J. Math. Phys.* **50**, 042106 (2009)
9. DiVincenzo, D.P., Leung, D.W., Terhal, B.M.: Quantum data hiding. *IEEE Trans. Inf. Theory* **48**, 3580–3599 (2002)
10. Matthews, W., Wehner, S., Winter, A.: Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Commun. Math. Phys.* **291**, 813–843 (2009)
11. Sen, P.: Random measurement bases, quantum state distinction and applications to the hidden subgroup problem, *IEEE Conference on Computational Complexity*, 274–287 (2006)
12. Hayashi, A., Hashimoto, T., Horibe, M.: Reexamination of optimal quantum state estimation of pure states. *Phys. Rev. A* **72**, 032325 (2005)
13. Scott, A.J.: Optimizing quantum process tomography with unitary 2-designs. *J. Phys. A* **41**, 055308 (2008). [arXiv:0711.1017](https://arxiv.org/abs/0711.1017)
14. Zhu, H., Englert, B.-G.: Quantum state tomography with fully symmetric measurements and product measurements. *Phys. Rev. A* **84**, 022327 (2011)
15. Roth, I., Kueng, R., Kimmel, S., Liu, Y.-K., Gross, D., Eisert, J., Kliesch, M.: Recovering quantum gates from few average gate fidelities. *Phys. Rev. Lett.* **121**, 170502 (2018)
16. Kueng, R., Zhu, H., Gross, D.: Distinguishing quantum states using Clifford orbits (2016), [arXiv:1609.08595](https://arxiv.org/abs/1609.08595)
17. Gross, D., Krahmer, F., Kueng, R.: A partial derandomization of PhaseLift using spherical designs. *J. Fourier Anal. Appl.* **21**, 229–266 (2015)
18. Szehr, O., Dupuis, F., Tomamichel, M., Renner, R.: Decoupling with unitary approximate two-designs. *New J. Phys.* **15**, 053022 (2013)
19. Brandao, F.G.S.L., Horodecki, M.: Exponential quantum speed-ups are generic. *Quant. Inf. Comp.* **13**, 0901 (2013)
20. Haferkamp, J., Faist, P., Kothakonda, N., B.T., Eisert, J., Younger Halpern, N.: Linear growth of quantum circuit complexity. *Nature Phys.* **18**, 528–532 (2022). <https://doi.org/10.1038/s41567-022-01539-6>
21. Roberts, D.A., Yoshida, B.: Chaos and complexity by design. *JHEP* **04**, 121 (2017)
22. Masanes, L., Roncaglia, A.J., Acín, A.: Complexity of energy eigenstates as a mechanism for equilibration. *Phys. Rev. E* **87**, 032137 (2013)
23. Onorati, E., Buerschaper, O., Kliesch, M., Brown, W., Werner, A.H., Eisert, J.: Mixing properties of stochastic quantum Hamiltonians. *Commun. Math. Phys.* **355**, 905–947 (2017)
24. Brandão, F.G.S.L., Harrow, A.W., Horodecki, M.: Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.* **346**, 397–434 (2016)
25. F. G. S. L. Brandão, A. W. Harrow, M. Horodecki, Efficient quantum pseudorandomness, *Phys. Rev. Lett.* **116** (2016b)
26. Cleve, R., Leung, D., Liu, L., Wang, C.: Near-linear constructions of exact unitary 2-designs. *Quant. Inf. Comp.* **16**, 0721–0756 (2015)
27. Harrow, A.W., Low, R.A.: Random quantum circuits are approximate 2-designs. *Commun. Math. Phys.* **291**, 257–302 (2009). [arXiv:0802.1919](https://arxiv.org/abs/0802.1919)
28. Hunter-Jones, N.: Unitary designs from statistical mechanics in random quantum circuits (2019), [arXiv:1905.12053](https://arxiv.org/abs/1905.12053)
29. Gottesman, D.: An introduction to quantum error correction and fault-tolerant quantum computation, [arXiv:0904.2557](https://arxiv.org/abs/0904.2557)
30. Campbell, E.T., Terhal, B.M., Vuillot, C.: Roads towards fault-tolerant universal quantum computation. *Nature* **549**, 172–179 (2017)
31. Veitch, V., Mousavian, A.H., Gottesman, D., Emerson, J.: The resource theory of stabilizer quantum computation. *New J. Phys.* **16**, 013009 (2014)

32. Howard, M., Campbell, E.: Application of a resource theory for magic states to fault-tolerant quantum computing. *Phys. Rev. Lett.* **118**, 090501 (2017)
33. Webb, Z.: The Clifford group forms a unitary 3-design (2015), [arXiv:1510.02769](https://arxiv.org/abs/1510.02769)
34. Zhu, H.: Multiqubit clifford groups are unitary 3-designs. *Phys. Rev. A* **96**, 062336 (2017)
35. Kueng, R., Gross, D.: Qubit stabilizer states are complex projective 3-designs (2015), [arXiv:1510.02767](https://arxiv.org/abs/1510.02767)
36. Zhu, H., Kueng, R., Grassl, M., Gross, D.: The Clifford group fails gracefully to be a unitary 4-design, [arXiv:1609.08172](https://arxiv.org/abs/1609.08172)
37. Helsen, J., Wallman, J.J., Wehner, S.: Representations of the multi-qubit Clifford group. *J. Math. Phys.* **59**, 072201 (2018)
38. Bannai, E., Navarro, G., Rizo, N., Tiep, P.H.: Unitary  $t$ -groups. *J. Math. Soc. Japan* **72**, 909–921 (2020). <https://doi.org/10.2969/jmsj/82228222>
39. Sawicki, A., Karnas, K.: Universality of single qudit gates. *Ann. Henri Poincaré* **18**, 3515–3552 (2017)
40. Koenig, R., Smolin, J.A.: How to efficiently select an arbitrary Clifford group element. *J. Math. Phys.* **55**, 122202 (2014). [arXiv: 1406.2170](https://arxiv.org/abs/1406.2170)
41. Nezami, S., Walter, M.: Multipartite entanglement in stabilizer tensor networks (2016), [arXiv:1608.02595](https://arxiv.org/abs/1608.02595)
42. Gross, D., Nezami, S., WalMain, J., Gamburd, A.: Schur-Weyl duality for the Clifford group with applications. *Invent. Math.* **171**, 83–121 (2008). <https://doi.org/10.1007/s00222-007-0072-z>
43. Montealegre-Mora, F., Gross, D.: Rank-deficient representations in howe duality over finite fields arise from quantum codes (2019), [arXiv:1906.07230](https://arxiv.org/abs/1906.07230)
44. Zhou, S., Yang, Z.-C., Hamma, A., Chamon, C.: Single  $t$  gate in a Clifford circuit drives transition to universal entanglement spectrum statistics (2019), [arXiv:1906.01079](https://arxiv.org/abs/1906.01079)
45. Aaronson, S., Gottesman, D.: Improved simulation of stabilizer circuits. *Phys. Rev. A* **70**, 052328 (2004)
46. Cwiklinski, P., Howodecki, M., Mozrzymas, M., Pankowski, L., Studzinski, M.: Local random quantum circuits are approximate polynomial-designs - numerical results. *J. Phys. A* **46**, 305301 (2013)
47. Bravyi, S., Browne, D., Calpin, P., Campbell, E., Gosset, D., Howard, M.: Simulation of quantum circuits by low-rank stabilizer decomposition. *Quantum* **3**, 181 (2019)
48. Pashayan, H., Wallman, J.J., Bartlett, S.D.: Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Phys. Rev. Lett.* **115**, 070501 (2015)
49. Heinrich, M., Gross, D.: Robustness of magic and symmetries of the stabiliser polytope. *Quantum* **3**, 132 (2019)
50. Bravyi, S., Gosset, D.: Improved classical simulation of quantum circuits dominated by Clifford gates. *Phys. Rev. Lett.* **116**, 250501 (2016). <https://doi.org/10.1103/PhysRevLett.116.250501>
51. Seddon, J., Regular, B., Pashayan, H., Ouyang, Y., Campbell, E.: Quantifying quantum speedups: improved classical simulation from tighter magic monotones (2020), [arXiv:2002.06181](https://arxiv.org/abs/2002.06181)
52. Brandao, F.G.S.L., Chemsyany, W., Hunter-Jones, N., Kueng, R., Preskill, J.: Models of quantum complexity growth. *PRX Quantum* **2**, 030316 (2021). <https://doi.org/10.1103/PRXQuantum.2.030316>
53. Varju, P.: Random walks in compact groups. *Doc. Math.* **18**, 1137–1175 (2013)
54. Nielsen, M. A., Chuang, I. L.: *Quantum computation and quantum information*, Cambridge Series on Information and the Natural Sciences ( Cambridge University Press, 2000)
55. Guralnick, R.M., Tiep, P.H.: Decompositions of small tensor powers and Larsen’s conjecture. *Represent. Theory* **9**, 138–208 (2005)
56. Low, R.A.: *Pseudo-randomness and Learning in Quantum Computation* (2010), [arXiv: 1006.5227](https://arxiv.org/abs/1006.5227)
57. Klaus, Stephan: *Brown-Kervaire invariants* (Shaker, 1995)
58. Watrous, J.: *The theory of quantum information* (Cambridge university press, 2018)
59. Brown, W.G., Viola, L.: Convergence rates for arbitrary statistical moments of random quantum circuits, *Phys. Rev. Lett.* **104**, 250501
60. Diaconis, P., Saloff-Coste, L.: Comparison techniques for random walk on finite groups. *Ann. Probab.* **21**, 2131–2156 (1993)
61. Nachtergaele, B.: The spectral gap for some spin chains with discrete symmetry breaking. *Commun. Math. Phys.* **175**, 565–606 (1996)
62. Aaronson, S., Gottesman, D.: Improved simulation of stabilizer circuits. *Phys. Rev. A* **70**, 052328 (2004)
63. Bhatia, B.: *Matrix analysis*, Springer Science & Business Media, **169** (2013)
64. Heinrich, M.: On stabiliser techniques and their application to simulation and certification of quantum devices, Ph.D. thesis, University of Cologne (2021), <https://kups.ub.uni-koeln.de/50465/>
65. Nebe, G., Rains, E.M., Sloane, N.J.A.: The invariants of the Clifford groups (2001), [arXiv:math/0001038v2](https://arxiv.org/abs/math/0001038v2)
66. Bourgain, J., Gamburd, A.: A spectral gap theorem in  $SU(d)$  (2011), [arXiv: 1108.6264](https://arxiv.org/abs/1108.6264)
67. Mezher, R., Ghalbouni, J., Dgheim, J., Markham, D.: Efficient approximate unitary  $t$ -designs from partially invertible universal sets and their application to quantum speedup, [arXiv:1905.01504](https://arxiv.org/abs/1905.01504) (2019)
68. de Montmort, P.R.: *Essay d’analyse sur les jex de hazard*, seconde edition, Jacque Quillau, Paris (1753)

69. Nakata, Y., Hirche, C., Koashi, M., Winter, A.: Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics, *Physical Review X* **7** (2017), <https://doi.org/10.1103/PhysRevX.7.021006>
70. Leone, L., Oliviero, S.F.E., Zhou, Y., Hamma, A.: Quantum chaos is quantum. *Quantum* **5**, 453 (2021)
71. Bröcker, T., Dieck, T.: Representations of compact Lie groups, Graduate Texts in Mathematics (Springer-Verlag), <https://www.springer.com/de/book/9783540136781>
72. Fulton, W., Harris, J.: Representation theory, edited by W. Fulton and J. Harris, Graduate Texts in Mathematics (Springer), [https://doi.org/10.1007/978-1-4612-0979-9\\_2](https://doi.org/10.1007/978-1-4612-0979-9_2)
73. Goodman, R., Wallach, N. R.: Symmetry, representations, and invariants, edited by R. Goodman and N. R. Wallach, Graduate Texts in Mathematics (Springer), [https://doi.org/10.1007/978-0-387-79852-3\\_1](https://doi.org/10.1007/978-0-387-79852-3_1)
74. Zhu, H., Kueng, R., Grassl, M., Gross, D.: The Clifford group fails gracefully to be a unitary 4-design (2016). [arXiv:1609.08172](https://arxiv.org/abs/1609.08172)
75. Folland, G.B.: How to integrate a polynomial over a sphere. *Am. Math. Mon.* **108**, 446–448 (2001)

Communicated by A. Childs



Random quantum circuits (RQCs) are central in both quantum information and quantum many-body physics. They are a simple model of local chaotic dynamics and have been used to study thermalization and the dynamical spread of entanglement in strongly-interacting quantum systems. In particular, they lack structure due to the local randomization but still capture the concept of local interactions.

We will see that many of their properties can be derived from the fact that they quickly generate quantum pseudorandomness in the sense of approximate unitary  $t$ -designs. In particular, they will eventually generate approximate unitary  $t$ -designs even for exponentially large  $t$ , in contrast to the bounds we found in the previous chapter.

In this section we study the convergence of random quantum circuits to unitary  $t$ -designs. In particular, we prove several results on the convergence to unitary designs based on improved spectral gaps of the moment operators.

The approach of bounding the spectral gap to establish convergence to unitary designs has been considered before in Refs. [BH10, BaHH16]. Specifically, Ref. [BaHH16] considered local RQCs in one dimension, i.e. consisting of geometrically local nearest-neighbor interactions between  $n$  qudits of local dimension  $q$ , and proved that the spectral gap is bounded below by  $\Omega(n^{-1}t^{-5-3.1/\log(q)})$ .

It has been conjectured that the actual scaling is  $\sim \text{poly}(n)^{-1}$ , thus independent of  $t$ . This result would have exciting implications. For instance, the depth at which unitary designs are generated would improve to a linear scaling in  $t$ . This would imply a robust version of the Brown-Susskind conjecture [BS18a] for random quantum circuits. We will elaborate on this conjecture in the next section, where we prove a variant of it.

Here, we make improvements to multiple existing bounds on the spectral gap for a number of different random circuit architectures. This section contains two publications. The first improves the design depth of 1D random quantum circuits over qubits from  $O(nt^{10.5})$  to  $O(nt^{5+o(1)})$ . Notice that this holds for all  $t$  that satisfy  $t \leq O(2^{n/2})$  in stark contrast to the bound  $t^2 \leq O(n)$  in Section 3.1.

Before we explain the contributions in this section, we formalize the connection between unitary  $t$ -designs and circuit complexity. Let  $\mathcal{G}$  denote a universal gate set of 2-local gates. Moreover, denote by  $M_{\mathcal{G},R}$  the set of all unitaries that can be realized by concatenation of at most  $R$  gates in  $\mathcal{G}$ .

**Definition 8.**

- For  $\delta \in (0, 1]$  a state  $|\psi\rangle$  has  $\delta$ -state complexity

$$\mathcal{C}_\delta(|\psi\rangle) := \min\{R, \exists V \in M_{\mathcal{G},R} \text{ with } \frac{1}{2}\|V|0^n\rangle\langle 0^n|V^\dagger - |\psi\rangle\langle\psi|\|_1 \leq \delta\}. \quad (3.3)$$

- For  $\delta \in (0, 1]$  a unitary  $U$  has  $\delta$ -unitary complexity

$$\mathcal{C}_\delta(U) := \min\{R, \exists V \in M_{\mathcal{G},R} \text{ with } \|V - U\|_\diamond \leq \delta\}, \quad (3.4)$$

where  $U$  is the channel defined by  $U(\rho) = U\rho U^\dagger$  and  $V$  likewise.

The power of unitary  $t$ -designs can be seen in a counting argument that can be found e.g. in Ref. [BCHJ<sup>+</sup>21]. First, we consider fully Haar random global unitaries. Indeed, almost all Haar random unitaries are close to maximally complex. The argument for this folklore result (see e.g. [Pre98]) goes roughly as follows: The Haar measure behaves locally like the Lebesgue measure on a  $\dim \text{SU}(2^n) = 4^n - 1$  dimensional space. Therefore, the volume of a ball with radius  $\varepsilon$  scales like  $O(\varepsilon^{4^n})$  -- doubly exponentially small in the system-size. A circuit with  $R$  gates implements at most  $|\mathcal{G}|^R$  many unitaries. However, this implies that at most a fraction of  $|\mathcal{G}|^R O(\varepsilon^{4^n})$  is  $\varepsilon$ -close to a unitary that can be implemented with a circuit with  $R$  gates. This can only be a constant fraction if  $R$  scales like  $4^n$  and almost all unitaries have at least exponential circuit complexity.

It turns out that this argument can be partially derandomized using unitary  $t$ -designs.

**Proposition 2.** *Let  $|\psi\rangle$  be drawn from a strong additive  $\varepsilon$ -approximate state  $t$ -design as defined in Definition 7 with  $\varepsilon = O(1)$ . Then,  $C_\delta(|\psi\rangle) \geq \Omega(t)$  with high probability.*

*Proof.* Notice first that

$$\frac{1}{2} \|\mathbb{V}|0^n\rangle\langle 0^n|\mathbb{V}^\dagger - |\psi\rangle\langle\psi|\|_1 = \sqrt{1 - |\langle\psi|\mathbb{V}|0^n\rangle|^2}. \quad (3.5)$$

Hence,

$$\frac{1}{2} \|\mathbb{V}|0^n\rangle\langle 0^n|\mathbb{V}^\dagger - |\psi\rangle\langle\psi|\|_1 \leq \delta \iff |\langle\psi|\mathbb{V}|0^n\rangle|^2 \geq (1 - \delta^2)^{\frac{1}{2}}. \quad (3.6)$$

We use first the union bound and then a Markov inequality to obtain

$$\begin{aligned} \Pr \left[ |\langle\psi|\mathbb{V}|0^n\rangle|^2 \geq (1 - \delta^2)^{\frac{1}{2}}, \mathbb{V} \in M_{R,\mathcal{G}} \right] &= \Pr \left[ \bigcup_{\mathbb{V} \in M_{R,\mathcal{G}}} |\langle\psi|\mathbb{V}|0^n\rangle|^2 \geq (1 - \delta^2)^{\frac{1}{2}} \right] \\ &\leq \sum_{\mathbb{V} \in M_{R,\mathcal{G}}} \Pr \left[ |\langle\psi|\mathbb{V}|0^n\rangle|^2 \geq (1 - \delta^2)^{\frac{1}{2}} \right] \\ &= \sum_{\mathbb{V} \in M_{R,\mathcal{G}}} \Pr \left[ |\langle\psi|\mathbb{V}|0^n\rangle|^{2t} \geq (1 - \delta^2)^{\frac{1}{2}t} \right] \\ &\leq \sum_{\mathbb{V} \in M_{R,\mathcal{G}}} \frac{\mathbb{E} |\langle\psi|\mathbb{V}|0^n\rangle|^{2t}}{(1 - \delta^2)^{\frac{1}{2}t}} \\ &\leq |\mathcal{G}|^R \frac{\mathbb{E}_{\psi \sim \mu_H} |\langle\psi|0^n\rangle|^{2t} + \varepsilon 2^{-nt}}{(1 - \delta^2)^{\frac{1}{2}t}} \\ &= |\mathcal{G}|^R \frac{\binom{2^n+t-1}{t}^{-1} + \varepsilon 2^{-nt}}{(1 - \delta^2)^{\frac{1}{2}t}}. \end{aligned} \quad (3.7)$$

In the last inequality, we use that

$$\begin{aligned} \mathbb{E} |\langle\psi|\mathbb{V}|0^n\rangle|^{2t} &= \mathbb{E} \left| \text{Tr} \left[ (|\psi\rangle\langle\psi|)^{\otimes t} (\mathbb{V}|0\rangle\langle 0|\mathbb{V}^\dagger)^{\otimes t} \right] \right| \\ &\leq \mathbb{E}_{\psi \sim \mu_H} |\langle\psi|\mathbb{V}|0^n\rangle|^{2t} + \left| \text{Tr} \left[ (\mathbb{E} (|\psi\rangle\langle\psi|)^{\otimes t} - \mathbb{E}_{\psi \sim \mu_H} (|\psi\rangle\langle\psi|)^{\otimes t}) (\mathbb{V}|0\rangle\langle 0|\mathbb{V}^\dagger)^{\otimes t} \right] \right| \\ &\leq \mathbb{E}_{\psi \sim \mu_H} |\langle\psi|\mathbb{V}|0^n\rangle|^{2t} + \left| \mathbb{E} (|\psi\rangle\langle\psi|)^{\otimes t} - \mathbb{E}_{\psi \sim \mu_H} (|\psi\rangle\langle\psi|)^{\otimes t} \right| \\ &\leq \mathbb{E}_{\psi \sim \mu_H} |\langle\psi|\mathbb{V}|0^n\rangle|^{2t} + \frac{\varepsilon}{2^{nt}} \end{aligned} \quad (3.8)$$

via Hölder’s inequality.

The proof of Proposition 2 now follows from the fact that by Stirling’s approximation, we have asymptotically

$$\binom{2^n + t - 1}{t} \sim \left(\frac{2^n}{t}\right)^t. \quad (3.9)$$

Therefore, for

$$R \log |\mathcal{G}| \leq O\left(t(n - \log(t) - \log\left((1 - \delta^2)^{-\frac{1}{2}}\right))\right),$$

the probability of accidentally drawing a low complexity state is small.  $\square$

Notice that we did not use of any structure of quantum circuits of a certain depth. In principle, the above argument applies to any set of exponentially upper bounded cardinality.

**Improving the Brandão-Harrow-Horodecki exponent.** In the first publication presented in this section, we improve the spectral gap of random quantum circuits on qubits from the bound obtained in Ref. [BaHH16] of  $\Omega(n^{-1}t^{-9.5})$  to  $\Omega(n^{-1}t^{-(4+o(1))})$ . This is achieved by improving the unconditional gap (independent of  $t$ ). In Ref. [BaHH16], this bound was obtained by showing convergence of random quantum circuits to the Haar measure in the Wasserstein distance in exponential depth. This convergence result is strong enough to be translated to a bound on the spectral gap. We show that a close sibling of the random walk introduced in Chapter 3 converges very quickly (in depth  $4^n$ ) to the Haar measure up to polynomial factors. Using modern tools like Gao’s quantum union bound [Gao15, AAV16], we can again translate this into an improved gap:

**Proposition 3.** *We have the explicit bound*

$$\|M(\nu_n, t) - M(\mu_H, t)\|_\infty \leq 1 - 120000^{-1} n^{-5} 2^{-2n}. \quad (3.10)$$

This, together with small improvements of the reductions in Ref. [BaHH16] yields the new bound.

In the second publication, we consider parameter regimes and models that are closely related to the random quantum circuit model in 1D considered in Ref. [BaHH16].

**Constant gap for large local dimensions.** For large enough local dimensions  $q$  the spectral gap scales inverse linearly in  $n$  and is independent of  $t$ . This was proven in [HJ19] for  $q \geq q_0$  for some  $q_0$  depending inexplicitly on the circuit size and the moment. We circumvent this uncontrolled approximation and strongly improve the dependence of  $q_0$  to  $O(t^2)$ :

**Theorem 11** (Spectral gaps for large  $q$ ). *Local random quantum circuits on  $n$  qudits with local dimension  $q$  have a spectral gap that can be bounded by  $1/8n$ , and hence independent of  $t$  for all  $q \geq 6t^2$  and  $t \geq 1$ .*

An immediate corollary of this result is that 1D RQCs form approximate designs in a depth that scales essentially optimally in  $n$  and  $t$ :

**Corollary 4** (Linear design growth). *1D random quantum circuits on  $n$  qudits of local dimension  $q$  form approximate unitary  $t$ -designs when the depth is  $O(nt \log(q))$  for all  $q \geq 6t^2$ .*

We prove this for two strong definitions of approximate design, in terms of the relative difference of channels as in [BaHH16], as well as in diamond norm with exponentially small error.

**Improved scaling for non-local random quantum circuits.** For quantum circuits with all-to-all interactions one would expect a faster mixing time and hence quicker convergence to unitary designs. Interestingly, the bounds that can be derived from [BaHH16, BH10] scale worse in  $n$  than for 1D circuits. In this publication we improve the  $n$  scaling in the all-to-all case and even improve on the scaling in  $t$  for larger (but fixed) values of the local dimension:

**Theorem 12** (Spectral gap for non-local circuits). *Let  $n \geq \max\{\lceil 2.03 \log_q(t) \rceil, 6000\}$ , then there is a constant  $c(q)$  such that the spectral gap is lower bounded by  $c(q)n^{-1} \log^{-1}(n) \log(t)t^{-\alpha(q)}$ , where  $\alpha(q)$  is a function of we define explicitly, with  $\lim_{q \rightarrow \infty} \alpha(q) = 4.06$ .*

**Exact result for the smallest non-trivial example.** By constructing an explicit basis of the orthocomplement of the permutations for the smallest non-trivial case  $n = 3$  and  $t = 2$ , we obtain an explicit formula for the second highest eigenvalue of the moment operator:

$$\lambda_2(M(\nu_{\text{RQC}, n=3}, t=2)) = \frac{1}{2} + \frac{q}{2(q^2 + 1)}. \quad (3.11)$$

This can be combined with finite-size criteria to improve the constants in the convergence of random quantum circuits to unitary 2-designs for all  $n$ .

**Improved spectral gaps from numerics.** By numerically constructing the local moment operator, and using sparse matrix methods for eigenvalue approximation, we numerically compute the spectral gaps for the first few moments. Again using finite-size criteria to bound spectral gaps for all  $n$ , we provide improved constants for the convergence of  $n$ -qubit RQCs to unitary  $t$ -designs, up to  $t = 5$ . Specifically, the constants are improved by factors of  $10^9$  as compared with those given in [BaHH16], which is of particular importance for near-term applications of random circuits.

To obtain these results we combine several techniques from quantum many-body physics as well as the theory of random walks on Lie groups. Such a combination was used in [BaHH16] by applying the Nachtergaele method [Nac96] for spectral gaps of frustration-free Hamiltonians to lower bound the spectral gap. Instead, we exploit so-called Knabe bounds on the spectral gap [Kna88] to prove Theorem 11. This allows us to use that the ground space of the Hamiltonians in question admits an approximately orthonormal product basis for the regime  $q \geq \Omega(t^2)$ .

The technically more involved result is Theorem 12. The key to this bound is a surprisingly simple recursion relation that relates the spectral gap of a random quantum circuit with all-to-all interactions to the spectral gap of a simpler auxiliary walk. More specifically, each step of the auxiliary walk picks a random qudit and applies a Haar random unitary from  $\mathcal{U}(q^{n-1})$  to the system consisting of all other qudits. We can show that this walk mixes quickly. A similar technique was used for the spectral gap of Kac's random walk [CCLo3, Maso3].

Denote the second highest eigenvalue of the moment operator for the non-local random quantum circuit on  $n$  qudits by  $\Delta_n$  and the second eigenvalue for the auxiliary circuit by  $\gamma_n$ . Then we prove the following lemma:

**Lemma 3** (Recursion relation for non-local gap). *For all  $n > 2$  it holds that*

$$\Delta_n \leq \gamma_n + \Delta_{n-1}(1 - \gamma_n). \quad (3.12)$$

We first solve the auxiliary spectral gap problem in the regime  $n \geq O(\log(t))$ , for which we obtain the desired result using the approximate orthogonality of the permutation operators, and then combine this with a general spectral gap bound independent of  $t$  but exponential in  $n$ .

The following Paper is published in Quantum under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Copyright remains with the original copyright holders such as the authors or their institutions.

# Random quantum circuits are approximate unitary $t$ -designs in depth $O(nt^{5+o(1)})$

Jonas Haferkamp

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany

The applications of random quantum circuits range from quantum computing and quantum many-body systems to the physics of black holes. Many of these applications are related to the generation of quantum pseudorandomness: Random quantum circuits are known to approximate unitary  $t$ -designs. Unitary  $t$ -designs are probability distributions that mimic Haar randomness up to  $t$ th moments. In a seminal paper, Brandão, Harrow and Horodecki prove that random quantum circuits on qubits in a brickwork architecture of depth  $O(nt^{10.5})$  are approximate unitary  $t$ -designs. In this work, we revisit this argument, which lower bounds the spectral gap of moment operators for local random quantum circuits by  $\Omega(n^{-1}t^{-9.5})$ . We improve this lower bound to  $\Omega(n^{-1}t^{-4-o(1)})$ , where the  $o(1)$  term goes to 0 as  $t \rightarrow \infty$ . A direct consequence of this scaling is that random quantum circuits generate approximate unitary  $t$ -designs in depth  $O(nt^{5+o(1)})$ . Our techniques involve Gao's quantum union bound and the unreasonable effectiveness of the Clifford group. As an auxiliary result, we prove fast convergence to the Haar measure for random Clifford unitaries interleaved with Haar random single qubit unitaries.

Random unitaries are a ubiquitous concept in quantum information theory and quantum many-body physics. This ranges from practical applications such as randomized benchmarking [21, 37, 34] to mixing in black holes [30]. However, the applicability of uniformly (Haar) random unitaries and states is limited by the fact that they require exponential resources as counting arguments show [33]. It is therefore desirable to consider less complex probability distributions over the unitary group that are sufficiently random for practical purposes. This leads naturally to the notion of unitary  $t$ -designs [17, 18, 23, 6]. These are probability distributions which mimic the Haar measure up to  $t$ th moments.

There is an ongoing effort to find efficient constructions of unitary  $t$ -designs [16, 29]. In a seminal result, Brandão, Harrow and Horodecki proved that local random quantum circuits are approximate unitary  $t$ -designs after the application of  $O(n^2t^{10.5})$  random gates for all  $t \leq O(2^{0.4n})$  [9]. This result has several consequences and many results [28, 46, 42, 38] on the statistical properties of random processes over the unitary group directly depend on the bound from Ref. [9].

Recently, unitary  $t$ -designs played a key role in lower bounding quantum circuit complexity [11, 10]. Here, Ref. [11] shows that strong notions of quantum circuit complexity of a unitary  $t$ -design can be lower bounded by  $\Omega(nt)$  with high probability. Combined with the design depth from Ref. [9], this roughly implies that the complexity of most quantum circuit of depth  $T$  can be lower bounded by  $\Omega(T^{1/10.5})$ . This line of research is inspired by the Brown-Susskind conjecture [13] in the context of the AdS/CFT correspondence [48]. The conjecture states that the quantum circuit complexity grows at a linear rate for an exponentially long time. Indeed, Ref. [11] implies such a linear lower bound in the limit of large local dimension [31] (it was later shown that a scaling of  $\sim 6t^2$  for the local dimension is sufficient [26]). More recently, a variant of the Brown-Susskind conjecture for random quantum circuits was proven for the special case of the exact circuit complexity [25]. However, for error-robust (and therefore more operational) notions of quantum circuit complexity and for random quantum circuits over qubits, the best lower bound for exponentially long times is  $\Omega(T^{1/10.5})$ .

Here, we want to improve the scaling of the design depth in  $t$ . Indeed, the scaling in  $n$  is already optimal (up to logarithmic factors). It was conjectured in Ref. [9] that the true scaling of the design depth might be linear in  $t$ , which would directly imply an error-robust version of the Brown-Susskind conjecture [11].

In this work, we dissect the bound obtained on the spectral gap of random quantum circuits in Ref. [9] and improve the lower bound on this spectral gap from  $\Omega(n^{-1}t^{-9.5})$  to  $\Omega(n^{-1}t^{-4-o(1)})$ .

The lion's share of the improvement is a new unconditional bound on the spectral gap. We achieve this bound by considering an auxiliary random walk that interleaves global uniformly random Clifford unitaries with Haar random unitaries on a single qubit. This walk was introduced in Ref. [27, 52] to analyze the number of non-Clifford unitaries required to approximate unitary  $t$ -designs for small values of  $t$  satisfying  $O(t^2) \leq n$ . Indeed, this random walk has desirable properties for the generation of unitary designs. It was proven that, while relative error approximations require  $\theta(n)$  many non-Clifford gates [35, 45], an additive constant approximation of unitary designs requires only a system-size independent amount of non-Clifford resources [27]. We show that the path coupling technique on the unitary group [44] provides a fast convergence to the Haar measure for this auxiliary random walk. As the Clifford group is a finite group with upper bounded circuit complexity, the comparison technique from Ref. [19] provides fast mixing bounds, which allows us to approximate the uniform measure on the Clifford group with a local random walk. We then translate these bounds to a spectral gap for local random quantum circuits by invoking Gao's quantum union bound [22, 7].

## 1 Preliminaries

A central object of this paper is the moment superoperator defined with respect to a probability distribution  $\nu$  on the unitary group  $U(d)$ , defined as

$$\Phi_\nu^{(t)}(A) := \int U^{\otimes t} A (U^\dagger)^{\otimes t} d\nu(U). \quad (1)$$

We denote the Haar-measure on the unitary group by  $\mu_H$ .

We make use of the following isomorphism called vectorization:  $\text{vec} : \mathbb{C}^{d \times d} \rightarrow \mathbb{C}^{d^2}$  with  $\text{vec}(|i\rangle\langle j|) = |i\rangle \otimes |j\rangle$ . This isomorphism extends to superoperators, mapping them to matrices:  $\text{vec}(T)\text{vec}(M) := \text{vec}(T(M))$  for all  $M \in \mathbb{C}^{d \times d}$  for a superoperator  $T$ .

We introduce the following notation for the second highest eigenvalue of the moment operator:

$$g(\nu, t) := \left\| M(\nu, t) - M(\mu_H, t) \right\|_{\infty}, \quad (2)$$

where the  $t$ -th moment operator of a probability distribution is defined as

$$M(\nu, t) := \text{vec}\left(\Phi_{\nu}^{(t)}\right) = \int U^{\otimes t} \otimes \bar{U}^{\otimes t} d\nu(U) \quad (3)$$

and  $\|\bullet\|_{\infty}$  denotes the Schatten  $\infty$ -norm. The *spectral gap* of the moment operator is  $1 - g(\nu, t)$ . Notice that we defined the spectral gap to be the difference between the first  $t!$  eigenvalues and the  $(t! + 1)$ st eigenvalue of the moment operator. That means that for non-universal probability distributions the gap can be 0. A probability distribution  $\nu$  is called a  $(\lambda, t)$ -*tensor product expander* if the spectral gap of the moment operator satisfies  $1 - g(\nu, t) \leq \lambda$ .

Moreover, we denote the Schatten 1-norm by  $\|\bullet\|_p$  and the Schatten 2-norm/Frobenius norm by  $\|\bullet\|_F$ . The stabilized induced (Schatten) 1-norm – called *diamond norm* – of superoperators will be denoted by  $\|\bullet\|_{\diamond}$  [5, 50]. Roughly, the distance between channels in diamond norm quantifies the “one-shot” distinguishability using entangled input states.

The second highest eigenvalue  $g(\nu, t)$  can be amplified. Specifically, the  $k$ -fold convolution of  $\nu$  has the property that

$$g(\nu^{*k}, t) \leq g(\nu, t)^k. \quad (4)$$

Upper bounds on the second highest eigenvalue can be used to imply an approximate version of unitary designs [9]. We define approximate designs in two (inequivalent) ways, with a relative error and with an exponentially small additive error. The relation to the spectral gap turns out to be the same.

**Definition 1** (Approximate unitary designs).

1. A probability distribution  $\nu$  on  $U(d)$  is an  $\varepsilon$ -approximate unitary  $t$ -design if the moment superoperator obeys

$$\left\| \Phi_{\nu}^{(t)} - \Phi_{\mu_H}^{(t)} \right\|_{\diamond} \leq \frac{\varepsilon}{d^t}. \quad (5)$$

2. A probability distribution  $\nu$  on  $U(d)$  is a relative  $\varepsilon$ -approximate  $t$ -design if

$$(1 - \varepsilon)\Phi_{\nu}^{(t)} \preceq \Phi_{\mu_H}^{(t)} \preceq (1 + \varepsilon)\Phi_{\nu}^{(t)}, \quad (6)$$

where here  $A \preceq B$  if and only if  $B - A$  is a completely positive map.

By applying [9, Lemma 4], as well as the fact that  $\|\Phi_{\nu}^{(t)} - \Phi_{\mu_H}^{(t)}\|_{\diamond} \leq d^t g(\nu, t)$  [36], we can state:



**Lemma 1.** *Let  $\nu$  be a probability distribution on  $U(d)$  such that  $g(\nu, t) \leq \varepsilon/d^{2t}$ . Then  $\nu$  is an  $\varepsilon$ -approximate unitary  $t$ -design and obeys both Eq. (5) and Eq. (6).*

In this paper we consider the following architectures of random quantum circuits comprised of 2-local unitary gates on a system of  $n$  qubits.

**Definition 2** (Random quantum circuits).

1. Local (1D) random quantum circuits: *Let  $\nu_n$  denote the probability distribution on  $U((\mathbb{C}^2)^{\otimes n})$  defined by first choosing a random pair of adjacent qubits and then applying a Haar random unitary  $U_{i,i+1}$  from  $U(4)$ . We assume periodic boundary conditions.*
2. Brickwork random quantum circuits: *Apply first a unitary  $U_{1,2} \otimes U_{3,4} \otimes \dots$  and then a unitary  $U_{2,3} \otimes U_{4,5} \otimes \dots$ , where all  $U_{i,i+1}$  are drawn Haar-randomly. For simplicity we assume in this case an even number of qubits. We denote this distribution by  $\nu_n^{\text{bw}}$ .*

A standard tool for converting results on the gap on local random quantum circuits into brickwork circuits is the detectability lemma [4, 7]. We will also make use of a strong converse called the *quantum union bound* [22, 7]. We present both together:

**Lemma 2** (Detectability lemma and union bound). *Let  $H = \sum_i Q_i$ , where  $Q_i$  are orthogonal projectors in an arbitrary order. Assume that each  $Q_i$  commutes with all but  $g$  of the projectors. Moreover, assume that  $H$  is frustration-free, i.e. a state  $|\psi\rangle$  with  $\langle\psi|H|\psi\rangle = 0$  exists. Then, for a state  $|\psi^\perp\rangle$  orthogonal to the ground space of  $H$ ,*

$$\sqrt{1 - 4\Delta(H)} \leq \left\| \prod_i (\mathbb{1} - Q_i) |\psi^\perp\rangle \right\|_2 \leq \sqrt{\frac{1}{\Delta(H)/g^2 + 1}}. \quad (7)$$

Finally, we introduce the Clifford group, which plays a crucial role in the proof of Theorem 1: The  $n$ -qubit Clifford group  $\text{Cl}(n)$  is the unitary normalizer of the Pauli group  $\mathcal{P}_n$ :

$$\text{Cl}(n) := \left\{ U \in U(2^n, \mathbb{Q}[i]) \mid U\mathcal{P}_n U^\dagger \subset \mathcal{P}_n \right\}. \quad (8)$$

Our interest stems from the fact that this group forms a unitary 2-design [51, 53], while every element has a polynomially upper bounded circuit complexity [2, 12].

## 2 Revisiting the spectral gap of random quantum circuits

Here we review the key steps in the proof technique in Ref. [9]. We introduce the notation  $P_H := M(\mu_H, t)$  and  $P_{H,m}$  to specify that we are talking about the Haar moment operator on  $m$  qubits. Ref. [9] introduced the frustration-free Hamiltonian

$$H_{n,t} := \sum_{i=1}^n (\mathbb{1} - \mathbb{1}_{i-1} \otimes P_{H,2} \otimes \mathbb{1}_{n-i-1}). \quad (9)$$

With this definition it holds that [9, Lem. 16]:

$$g(\nu_n, t) = 1 - \frac{\Delta(H_{n,t})}{n}, \quad (10)$$

where  $\Delta(H_{n,t})$  denote the spectral gap of  $H_{n,t}$ .

Using Nachtergaele's martingale method for spin systems, Ref. [9] proves that for all  $n$  and  $t$  such that  $n \geq \lceil 2.5 \log_2(4t) \rceil$  the following bound holds:

$$\Delta(H_{n,t}) \geq \frac{\Delta(H_{\lceil 2.5 \log_2(4t) \rceil, t})}{4^{\lceil 2.5 \log_2(4t) \rceil}}. \quad (11)$$

Remarkably, this reduction not only establishes a constant spectral gap for every fixed value of  $t$ , but also turns every bound that only depends exponentially on the system size into one that depends polynomially on  $t$ .

The reduction (11) to smaller system sizes is then combined with an spectral gap that holds for all values of  $t$ . We call such a gap *unconditional*:

$$\Delta(H_{n,t}) \geq \frac{1}{n(5e)^n}. \quad (12)$$

This bound is obtained from a convergence result: By applying the path coupling technique on the unitary group [44], the authors of Ref. [9] show that random quantum circuits converge to the Haar measure in the Wasserstein distance. This convergence result is strong enough to imply the bound (12) on the spectral gap.

How does the exponent of  $t$  for the design depth in Ref. [9] decompose exactly? Ignoring logarithmic factors, the exponent is approximately

$$10.41 \approx \underbrace{1}_1 + \left( \underbrace{2}_2 + \underbrace{0.5}_3 \right) \times \left( \underbrace{\log_2 5}_4 + \underbrace{\log_2(e)}_5 \right). \quad (13)$$

1. This contribution comes from the conversion of a spectral gap to the stronger notions of approximate designs in Definition 1. This contribution is therefore a necessary consequence of using spectral gaps to bound the design depth.
2. Ref. [9] establishes approximate orthogonality of permutations of  $t$  tensor factors of  $\mathbb{C}^D$  in the regime  $D \geq t^2$ . In the application of Nachtergaele's martingale method [41], this fact is repeatedly used and the second contribution can be directly linked to the square in the condition  $t^2 \leq D$ . More precisely, the following inequality holds

$$\left\| M(\mu_H, t) - 2^{-tn} \sum_{\pi \in S_t} \text{vec}(r(\pi)) \text{vec}(r(\pi))^\dagger \right\|_\infty \leq \frac{t^2}{D}, \quad (14)$$

where  $r$  denotes the representation of the symmetric group that permutes the  $t$  tensor factors. The second operator in Eq. (14) is called the frame operator of the permutations. It has the same eigenvalues as the Gram matrix of the permutations and is also the moment operator of random matrices with i.i.d. Gaussian entries by Wick's theorem. The approximate orthogonality of

the permutations is therefore equivalent to the moments of the Haar measure being approximately equal to those of Gaussian matrices with i.i.d. entries. Each monomial of degree  $t$  only contains information about some  $t \times t$  submatrix. Indeed, small submatrices  $t \sim o(\sqrt{D})$  of Haar random unitaries are conjectured to be close to an i.i.d. Gaussian ensemble in total variation distance [1, 32]. However, as the entries of Haar random matrices are correlated, large submatrices cannot be close to an i.i.d Gaussian ensemble. Consequently, we cannot expect to make the RHS of Eq. (14) scale better than  $t/D$ .

3. This contribution is a consequence of Nachtergaele’s bound. More precisely, the final bound of Ref. [9, Lem. 18] is

$$\Delta(H_{n,t}) \geq \frac{\Delta(H_{l,t})}{4l}, \quad (15)$$

for any  $l$  that satisfies the inequality

$$\frac{6t^2}{2^l} \leq \frac{1}{2}l^{-\frac{1}{2}}. \quad (16)$$

The choice  $l = 2.5\lceil \log_2(4t) \rceil$  can be seen to always satisfy (16). This bound gives a comparably simple expressions at the expense of a slightly worse than optimal exponent in  $t$ . However, as the objective of this work is to pedantically optimize the exponent of  $t$ , we use modern bounds on Lambert’s W function to obtain such a tight bound in  $l$ . We present the resulting bound below in Observation 1.

4. The last two contributions are a consequence of the unconditional spectral gap in Eq. (12). In general, contribution 4. comes from an exponential decay  $(q^2 + 1)^{-n}$  (for qubits  $q = 2$ ) of the unconditional bound on the gap. This part of the bound (12) can be proven for random “staircase circuits” of the form  $U_{1,2}U_{2,3} \dots U_{n-1,n}$ , with Haar random 2-local unitaries  $U_{i,i+1}$ . The same bound holds for permutations of the unitaries  $U_{i,i+1}$ . In this work, we observe that the application of the path coupling method in Ref. [9] uses the full Haar randomness of the first gate  $U_{1,2}$  only. For all other gates, we merely need the second moments of the Haar measure and we would hence obtain the same bound if the gates  $U_{2,3}, \dots, U_{n-1}$  are drawn uniformly from the Clifford group. The key idea in our improved bound is to go one step further and directly apply the full  $n$ -qubit Clifford group and only worry about locality once this convergence result is translated to spectral gaps.
5. Permutations of these staircase random walks appear with non-zero probability after  $n$  steps of local random quantum circuits. More precisely, these are realized with a probability of  $n!/n^n$ . This explains the appearance of Euler’s constant  $e$  as Stirling’s approximation of the factorial is  $n! \sim (n/e)^n$ . By applying Gao’s union bound directly to the moment operators of Clifford brickwork circuits approximating the uniform distribution on the Cliffords, we can circumvent this counting step entirely.

We observe that the bound (11) can be slightly improved. A detailed argument for the following can be found in Appendix A.

**Observation 1** (Improved gap reduction). *For all  $n \geq \lceil 2 \log_2(4t) + 1.5\sqrt{\log_2(4t)} \rceil$  the following bound holds:*

$$\Delta(H_{n,t}) \geq \frac{\Delta\left(H_{\lceil 2 \log_2(4t) + 1.5\sqrt{\log_2(4t)} \rceil, t}\right)}{4 \lceil 2 \log_2(4t) + 1.5\sqrt{\log_2(4t)} \rceil}. \quad (17)$$

## 3 Results

### 3.1 Tensor product expanders and unitary designs

Our first result is a bound on the spectral gap of the moment operator of local random quantum circuits that is independent of the moment  $t$ :

**Theorem 1** (Unconditional gap). *We have the following bound for all  $t \geq 1$ :*

$$g(\nu_n, t) \leq 1 - 120000^{-1} n^{-5} 2^{-2n} \quad (18)$$

This can be combined with the slightly improved reduction in Observation 1 to yield the following bounds on the tensor product expander properties of local random quantum circuits:

**Theorem 2** (Tensor product expanders). *We have the following bounds for  $n \geq \lceil 2 \log_2(4t) + 1.5\sqrt{\log_2(4t)} \rceil$ :*

- $g(\nu_n, t) \leq 1 - \left(Cn \ln^5(t) t^{4+3\frac{1}{\sqrt{\log_2(t)}}}\right)^{-1}$
- $g(\nu_n^{\text{bw}}, t) \leq 1 - \left(3C \ln^5(t) t^{4+3\frac{1}{\sqrt{\log_2(t)}}}\right)^{-1}$ ,

where the constant can be taken to be  $C = 10^{13}$ .

*Proof.* The bound on  $g(\nu_n, t)$  follows from plugging the unconditional bound

$$\Delta(H_{m,t}) \geq 120000^{-1} m^{-4} 2^{-2m} \quad (19)$$

from Theorem 1 into the reduction (17) in Observation 1. The bound on  $g(\nu_n^{\text{bw}})$  can be obtained from this by applying the detectability lemma (Lemma 2):

$$g(\nu_n^{\text{bw}}) \leq \sqrt{\frac{1}{\Delta(H_{n,t})/4 + 1}} \leq 1 - \left(3C \ln^5(t) t^{4+3\frac{1}{\sqrt{\log_2(t)}}}\right)^{-1}, \quad (20)$$

which is the second part of Theorem 2. □

By Lemma 1, this immediately implies the following corollary, which is the main application of Theorem 2:

**Corollary 1** (Unitary designs). *For  $n \geq \lceil 2 \log_2(4t) + 1.5\sqrt{\log_2(4t)} \rceil$ , the following bounds hold*

- *Local random quantum circuits are  $\varepsilon$ -approximate unitary  $t$ -designs after*

$$k \geq C n \ln^5(t) t^{4+3\frac{1}{\sqrt{\log_2(t)}}} (2nt + \log_2(1/\varepsilon)) \quad (21)$$

*steps.*

- *Brickwork quantum circuits are  $\varepsilon$ -approximate unitary  $t$ -designs after*

$$k \geq C n \ln^5(t) t^{4+3\frac{1}{\sqrt{\log_2(t)}}} (2nt + \log_2(1/\varepsilon)) \quad (22)$$

*steps.*

In comparison to the exponent in Eq. (13), the new exponent decomposes as follows.

$$5 + 3\frac{1}{\sqrt{\log_2(t)}} = \underbrace{1}_{1.} + \left( \underbrace{2}_{2.} + \underbrace{1.5\frac{1}{\sqrt{\log_2(t)}}}_{3.} \right) \times \underbrace{2}_{4.}, \quad (23)$$

where we again ignored the log-factors in (21). Here, contribution 1. and 2. appear for the same reason as 1. and 2. in Eq. (13). Contribution 3. is a direct consequence of Observation 1 and 4. follows from the improved bound on the unconditional gap in Theorem 1.

By applying an argument from Ref. [9], the same scaling can be implied for random quantum circuits drawn from a discrete invertible gate set with algebraic entries, using a powerful result by Bourgain and Gamburd [8]. This bound comes with an implicit constant depending on the gate set. Moreover, at the expense of additional polylogarithmic factors in  $t$ , the assumption of algebraic entries can be dropped as proven in Ref. [47] using on a theorem from Ref. [49]. Similarly, it was shown recently that the assumption of invertibility can be dropped at the expense of an additional factor of  $n$  [47].

In this paper we focus on random quantum circuits over qubits. This is because there are very concrete bounds available for the maximal circuit complexity of the multiqubit Clifford group [2, 12]. However, the same proof strategy works for random quantum circuits defined over every local dimension that is prime because in these dimensions the Clifford group forms a unitary 2-design. Therefore, all results in this paper can be proven for every prime power dimension with constants depending on the maximal circuit complexity of the multiqubit Clifford group.

## 3.2 Quantum circuit complexity

We apply our result to the growth of circuit complexity [11, 10]. Let  $\mathcal{G}$  denote a universal gate set of 2-local gates. Moreover, denote by  $M_{\mathcal{G},R}$  the set of all unitaries that can be realized by concatenation of at most  $R$  gates in  $\mathcal{G}$ .

**Definition 3** (Quantum circuit complexity). For  $\delta \in (0, 1]$  a

- state  $|\psi\rangle$  has  $\delta$ -state complexity

$$\mathcal{C}_\delta(|\psi\rangle) := \min\{R, \exists V \in M_{G,R} \text{ with } \frac{1}{2}\|V|0^n\rangle\langle 0^n|V^\dagger - |\psi\rangle\langle\psi|\|_1 \leq \delta\}. \quad (24)$$

- unitary  $U$  has  $\delta$ -unitary complexity

$$\mathcal{C}_\delta(U) := \min\{R, \exists V \in M_{G,R} \text{ with } \|\mathbf{V} - \mathbf{U}\|_\diamond \leq \delta\}, \quad (25)$$

where  $\mathbf{U}$  is the channel defined by  $\mathbf{U}(\rho) = U\rho U^\dagger$  and  $\mathbf{V}$  likewise.

We can use the following bound:

**Theorem 3** (Informal, Ref. [9]). Let  $\nu$  be a relative approximate unitary  $t$ -design for some  $t \leq O(2^{n/2})$ . Then, a unitary  $U$  drawn from  $\nu$  satisfies  $\mathcal{C}_\delta(U) \geq \Omega(nt)$  and  $\mathcal{C}_\delta(U|\psi) \geq \Omega(nt)$  with high probability.

Combining this theorem with Corollary 1 yields the following lower bounds on the circuit complexity:

**Corollary 2** (Growth of quantum circuit complexity). Let  $U$  be drawn from a random quantum circuit in brickwork architecture of depth  $T$ . Moreover, let  $\delta \in (0, 1)$  be constant in the system size. Then, with probability  $1 - e^{-\Omega(n)}$ , we have for the

- $\delta$ -state complexity

$$\mathcal{C}_\delta(U0^n) \geq \Omega\left(T^{1/(5+o(1))}\right)$$

- $\delta$ -unitary complexity

$$\mathcal{C}_\delta(U) \geq \Omega\left(T^{1/(5+o(1))}\right)$$

until  $T \geq \Omega(2^{n/(2+o(1))})$ .

For the error-robust notions of quantum circuit complexity defined above Corollary 2 provides the strongest known bounds for random quantum circuits over qubits of superpolynomial depth.

Ref. [11] in fact provides an even stronger result. Even an operational notion of complexity that quantifies the resources necessary to distinguish a unitary from the maximally depolarizing channel (the ‘‘most useless’’ channel) is lower bounded by  $\Omega(T^{1/(5+o(1))})$ .

## 4 Proof of Theorem 2 and Theorem 1

Denote by  $\mu_{H,1}$  the Haar measure on the subgroup  $U(2) \otimes \mathbb{1}_{n-1}$  and  $\mu_{\text{Cl}}$  the uniform measure on the Clifford group. We apply the path coupling technique [14] for the unitary group developed in Ref. [44] to the following auxiliary random walk that is defined by

$$\sigma^{*k} := (\mu_{\text{Cl}} * \mu_{H,1} * \mu_{\text{Cl}})^{*k}. \quad (26)$$

This walk was defined in [27] and shown to generate unitary  $t$ -designs in depth  $k = O(t^4)$  for ‘‘small’’  $t$ . More precisely, the result holds for the regime  $t^2 \leq O(n)$  and for a weaker definition of approximate unitary design.

The first lemma is an unconditional spectral gap on the random walk  $\sigma$ :

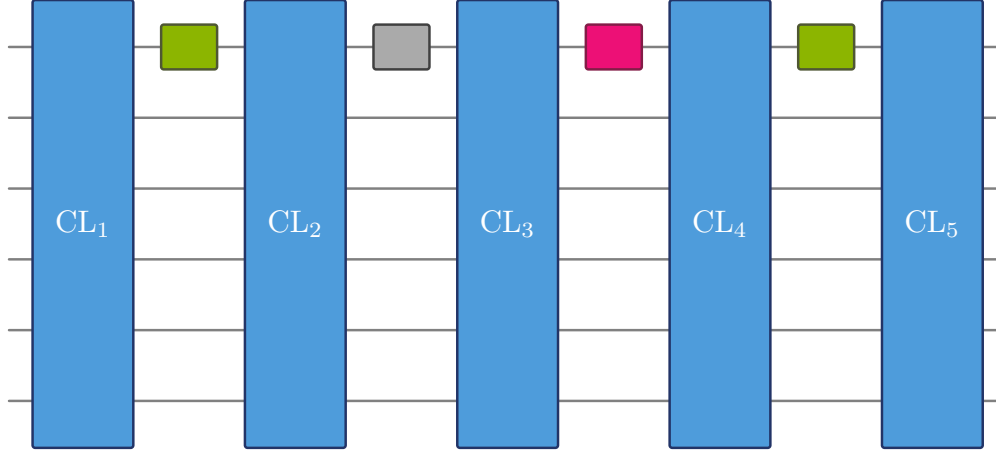


Figure 1: Four steps of the random walk  $\sigma$  on  $n = 6$  qubits.

**Lemma 3.** *We have for all  $t \geq 1$  and  $n \geq 1$  that*

$$g(\sigma, t) \leq 1 - \frac{3}{2} \frac{1}{2^{2n} - 1} \quad (27)$$

The path coupling technique provides a bound on the convergence to the Haar measure in the Wasserstein distance, which is based on the notion of a random coupling: A random variable  $(X, Y)$  is called a coupling for the distributions  $\nu_1$  and  $\nu_2$  if the marginal distributions for random variables  $X$  and  $Y$  are  $\nu_1$  and  $\nu_2$ , respectively. This allows us to define the Wasserstein distance:

$$W_{\mathfrak{g},p}(\nu_1, \nu_2) := \inf\{(\mathbb{E} \mathfrak{g}(X, Y)^p)^{1/p}, \quad (X, Y) \text{ is a random coupling}\}. \quad (28)$$

In Ref. [9] a convergence result is proven for for the distance  $W_{\text{Rie},2}$ . Here,  $\mathfrak{g}_{\text{Rie}}$  is the metric induced by the unique Riemannian metric that is invariant under left and right translations. For our purposes, however, it is sufficient to prove a convergence in the weaker distance  $W_{\text{Fro},2}$ . We remark that the stronger convergence result can be obtained from our analysis as in Ref. [9].

It was proven in Ref. [44] that infinitesimal contractions can be boosted to a global contraction of a random walk. More precisely, Theorem 2 in Ref. [44] contains the following lemma as a special case:

**Lemma 4.** *Suppose that*

$$\limsup_{X \rightarrow Y} \left\{ \frac{W_{\text{Fro},2}(\nu * \delta_X, \nu * \delta_Y)}{\|X - Y\|_F} : \|X - Y\|_F \leq \varepsilon \right\} \leq \eta, \quad (29)$$

for some probability distribution  $\nu$  over  $SU(d)$ . Then, for all probability measures  $\mu_1$  and  $\mu_2$ ,

$$W_{\text{Fro},2}(\nu * \mu_1, \nu * \mu_2) \leq \eta W_{\text{Fro},2}(\mu_1, \mu_2). \quad (30)$$

Using Lemma 4, we prove the following convergence result:

**Lemma 5.** *For every  $k \geq 1$ , we have the bound:*

$$W_{\text{Fro},2}(\sigma^{*k}, \mu_H) \leq \left(1 - \frac{3}{2^{2n} - 1}\right)^{k/2} \pi 2^{n/2}. \quad (31)$$

*Proof.* A single step of the random walk  $\sigma$  applied to a fixed unitary  $X \in U(2^n)$  by left multiplication yields:

$$X \rightarrow \{C'(U \otimes \mathbb{1}_{n-1})CX\}, \quad (32)$$

for  $U \in U(2)$  and  $C, C' \in \text{Cl}(n)$ . The unitary  $Y$  undergoes the same transformation. We can now introduce a family of random couplings for  $(\sigma * \delta_X, \sigma * \delta_Y)$ :

$$X' = C'(UV \otimes \mathbb{1}_{n-1})CX \quad Y' = C'(U \otimes \mathbb{1}_{n-1})CY \quad (33)$$

defined for a unitary  $V \in U(2)$  that is independent of  $U$  but can depend on  $X, Y$  and  $C$ .

We need to bound:

$$\mathbb{E}\|X' - Y'\|_{\mathbb{F}}^2 = \mathbb{E}\|C'(UV \otimes \mathbb{1}_{n-1})CX - C'(U \otimes \mathbb{1}_{n-1})CY\|_{\mathbb{F}}^2. \quad (34)$$

We can choose  $V$  to be minimal for all  $C, X$  and  $Y$ . Then, we find

$$\begin{aligned} \mathbb{E}\left(\min_V \|C'(UV \otimes \mathbb{1}_{n-1})CX - C'(U \otimes \mathbb{1}_{n-1})CY\|_{\mathbb{F}}^2\right) &= 2\left(\text{Tr}[\mathbb{1}_n] - \max_V \text{Tr}[V \otimes \mathbb{1}_{n-1} CXY^\dagger C^\dagger]\right) \\ &= 2\left(\text{Tr}[\mathbb{1}_n] - \mathbb{E}\|\text{Tr}_{[2,n]}(CXY^\dagger C^\dagger)\|_1\right), \end{aligned} \quad (35)$$

where we used the variational characterization of the Schatten 1-norm. We can choose  $X$  and  $Y$  to be infinitesimally close together and write

$$R := XY^\dagger = \exp(i\varepsilon H) = \mathbb{1}_n + i\varepsilon H - \frac{\varepsilon^2}{2}H^2 + O(\varepsilon^3), \quad (36)$$

with  $\|H\|_{\mathbb{F}} \leq 1$ . By Taylor expanding the 1-norm, this implies (as in Ref. [9]):

$$\begin{aligned} \mathbb{E}\left(\min_V \|C'(UV \otimes \mathbb{1}_{n-1})CX - C'(U \otimes \mathbb{1}_{n-1})CY\|_{\mathbb{F}}^2\right) \\ = \varepsilon^2\left(\text{Tr}(H^2) - \frac{1}{2^{n-1}}\mathbb{E}\left[\text{Tr}\left(\text{Tr}_{[2,n]}[CHC^\dagger]\right)^2\right]\right) + O(\varepsilon^3). \end{aligned} \quad (37)$$

Denote by  $\mathbb{F}$  the swap operator defined by  $\mathbb{F} : \mathbb{C}^d \otimes \mathbb{C}^d \rightarrow \mathbb{C}^d \otimes \mathbb{C}^d$ ,  $\mathbb{F}|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$ . The ‘‘replica trick’’

$$\text{Tr}[A^2] = \text{Tr}[A^{\otimes 2}\mathbb{F}] \quad (38)$$

for every matrix  $A$  can be applied to show

$$\begin{aligned} \mathbb{E}\left[\text{Tr}\left(\text{Tr}_{[2,n]}[CHC^\dagger]\right)^2\right] &= \text{Tr}\left[(CHC^\dagger)^{\otimes 2}\mathbb{F}_1 \otimes \mathbb{1}_{[2,n]}\right] \\ &= \text{Tr}\left[H^{\otimes 2}(C^\dagger)^{\otimes 2}(\mathbb{F}_1 \otimes \mathbb{1}_{[2,n]})C^{\otimes 2}\right]. \end{aligned} \quad (39)$$

We now use that the Clifford group is an exact unitary 2-design: the expectation value in the above equation is the same as that for Haar-random unitaries  $C$ . We



employ standard formulas [3, Lem. IV.3] for the twirl over the Haar measure:

$$\begin{aligned}
\mathbb{E}_{C \sim \mu_{\text{Cl}}} & \text{Tr}[H^{\otimes 2} (C^\dagger)^{\otimes 2} (\mathbb{F}_1 \otimes \mathbb{1}_{[2,n]}) C^{\otimes 2}] \\
&= \text{Tr} \left[ H^{\otimes 2} \left( \frac{2 + 2^{n-1}}{2^n + 1} \frac{1}{2} (\mathbb{1} + \mathbb{F}) - \frac{2 - 2^{n-1}}{2^n - 1} \frac{1}{2} (\mathbb{1} - \mathbb{F}) \right) \right] \\
&= \frac{1}{2} \left( \frac{2 + 2^{n-1}}{2^n + 1} + \frac{2 - 2^{n-1}}{2^n - 1} \right) \text{Tr}[H^{\otimes 2} \mathbb{F}] + \frac{1}{2} \left( \frac{2 + 2^{n-1}}{2^n + 1} - \frac{2 - 2^{n-1}}{2^n - 1} \right) \text{Tr}[H^{\otimes 2} \mathbb{1}] \\
&= \left( \frac{2 - 2^{-1}}{2^n - 2^{-n}} \right) \text{Tr}[H^2] + \left( \frac{2^{2n-1} - 2}{2^{2n} - 1} \right) \text{Tr}[H]^2 \\
&\geq \left( \frac{2 - 2^{-1}}{2^n - 2^{-n}} \right) \text{Tr}[H^2].
\end{aligned} \tag{40}$$

Consequently, we obtain

$$\mathbb{E} \left( \min_V \|C'(UV \otimes \mathbb{1}_{n-1})CX - C'(U \otimes \mathbb{1}_{n-1})CY\|_F^2 \right) \leq \left( 1 - \frac{3}{2^{2n} - 1} \right) \varepsilon^2 \text{Tr}(H^2) + O(\varepsilon^3). \tag{41}$$

From

$$\|X - Y\|_F^2 = \varepsilon^2 \text{Tr}(H^2) + O(\varepsilon^3), \tag{42}$$

we find

$$\mathbb{E} \|X' - Y'\|_F^2 \leq \left( 1 - \frac{3}{2^{2n} - 1} \right) \|X - Y\|_F^2 + O(\varepsilon^3). \tag{43}$$

Notice, that for  $X \neq Y$ , we can always choose  $\varepsilon$ , such that  $\text{Tr}(H^2) = \|H\|_F^2 = 1$ . Therefore, we find that

$$\begin{aligned}
\frac{W_{\text{Fro},2}(\sigma * \delta_X, \sigma * \delta_Y)}{\|X - Y\|_F} &\leq \frac{(\mathbb{E} \|X' - Y'\|_F^2)^{\frac{1}{2}}}{\|X - Y\|_F} \\
&\leq \frac{\sqrt{\left( 1 - \frac{3}{2^{2n} - 1} \right) \|X - Y\|_F^2 + O(\varepsilon^3)}}{\|X - Y\|_F} \\
&\leq \sqrt{\left( 1 - \frac{3}{2^{2n} - 1} \right)} + \sqrt{\frac{O(\varepsilon^3)}{\varepsilon^2 + O(\varepsilon^3)}} \\
&\leq \sqrt{\left( 1 - \frac{3}{2^{2n} - 1} \right)} + O(\sqrt{\varepsilon}),
\end{aligned} \tag{44}$$

with  $\varepsilon \rightarrow 0$  as  $X \rightarrow Y$ . Invoking Lemma 4  $k$  times with  $\eta := \sqrt{\left( 1 - \frac{3}{2^{2n} - 1} \right)}$  implies

$$\begin{aligned}
W_{\text{Fro},2}(\sigma^{*k}, \mu_H) &= W_{\text{Fro},2}(\sigma^{*k} * \delta_{\mathbb{1}}, \sigma * \mu_H) \\
&\leq \eta^k W_{\text{Fro},2}(\delta_{\mathbb{1}}, \mu_H) \\
&\leq \eta^k (\mathbb{E}_U \|\mathbb{1} - U\|_F^2)^{\frac{1}{2}} \\
&= \sqrt{2} 2^{n/2} \eta^k,
\end{aligned} \tag{45}$$

which is the statement of Lemma 5.  $\square$

We can now prove Lemma 3:

*Proof of Lemma 3.* We use the following formula [9], which holds for any probability distribution  $\nu$ :

$$g(\nu, t) \leq 2tW_{\text{Fro},1}(\nu, \mu_H) \quad (46)$$

Combining this with Lemma 5 yields:

$$\begin{aligned} g(\sigma^{*k}, t) &\leq 2tW_{\text{Fro},1}(\sigma^{*k}, \mu_H) \\ &\leq 2tW_{\text{Fro},2}(\sigma^{*k}, \mu_H) \\ &\leq \left(1 - \frac{3}{2^{2n} - 1}\right)^{k/2} 2\sqrt{2}t2^{n/2}, \end{aligned} \quad (47)$$

where the second inequality follows immediately from Jensen's inequality.  $\sigma$  is a symmetric measure, that is, it is invariant under taking inverses. Consequently, the moment operators are hermitian, which implies

$$g(\sigma^{*k}, t) = g(\sigma, t)^k. \quad (48)$$

Notice that the inequality  $g(\sigma^{*k}, t) \leq g(\sigma, t)^k$  holds for all measure but not equality. We plug (48) into (47), take the  $k$ -th square root and the limit  $k \rightarrow \infty$ . This yields

$$g(\sigma, t) \leq \left(1 - \frac{3}{2^{2n} - 1}\right)^{1/2} \leq 1 - \frac{1}{2} \frac{3}{2^{2n} - 1}. \quad (49)$$

□

Before we can prove Theorem 1, we need another auxiliary bound. More precisely, we show that a local random walk over Clifford generators quickly approximates the uniform measure on the Clifford group: Apply first a unitary  $U_{1,2} \otimes U_{3,4} \otimes \dots$  and then a unitary  $U_{2,3} \otimes U_{4,5} \otimes \dots$ , where all  $U_{i,i+1}$  are drawn uniformly from the Clifford group on 2 qubits. We denote this distribution by  $\nu_n^{\text{Cl,bw}}$

The Clifford group is a finite group with polynomially bounded circuit complexity [2, 12] and we can use this fact to apply the comparison technique [19]. This gives the following bound:

**Lemma 6.** *The following bound holds for all  $t \geq 1$ :*

$$\left\| M\left(\nu_n^{\text{Cl,bw}}, t\right) - M\left(\mu_{\text{Cl}}, t\right) \right\|_{\infty} \leq 1 - \frac{1}{2000n^3}. \quad (50)$$

*Proof.* Consider the averaging operator on the group algebra  $T_{\nu} : L^2(\text{Cl}_n) \rightarrow L^2(\text{Cl}_n)$  defined by

$$(T_{\nu}f)(g) := \int f(h^{-1}g)d\nu(h). \quad (51)$$

By [19], we have

$$\lambda_2(T_{\sigma}) \leq 1 - \frac{\eta}{d^2}, \quad (52)$$

where  $\eta$  is the probability of the least probable generator and  $d$  is the minimal number of generators necessary to generate any group element. Here,  $d = 9n$  by [12] and  $\eta = 1/|\text{Cl}_2|n = 1/24n$ . By the Peter-Weyl theorem, the entire spectrum of the moment operators is contained in the spectrum of the averaging operator and thus we obtain Eq. (50). □

We remark that the comparison technique was previously applied to the Clifford group in Ref. [20].

We can now put everything together.

*Proof of Theorem 1.* For simplicity, we use the following notation:

$$P_{\text{Cl},n} = M(\mu_{\text{Cl},n}, t). \quad (53)$$

We use

$$\left\| M(\nu_n^{\text{Cl,bw}}, t)^k - P_{\text{Cl},n} \right\|_{\infty} \leq \left\| M(\nu_n^{\text{Cl,bw}}, t) - P_{\text{Cl},n} \right\|_{\infty}^k \leq \left(1 - \frac{1}{2000n^3}\right)^k. \quad (54)$$

In particular, we have the approximation

$$\left\| M(\nu_n^{\text{Cl,bw}}, t)^{k_n} - P_{\text{Cl},n} \right\|_{\infty} \leq \frac{1}{2} \frac{1}{2^{2n} - 1} \quad (55)$$

for a number of layers  $k_n$  that we can choose as

$$k_n = 6000n^4. \quad (56)$$

Therefore, by applying the triangle inequality twice, we obtain

$$\begin{aligned} & \left\| M(\nu_n^{\text{Cl,bw}}, t)^{k_n} M(\mu_{H,1}, t) M(\nu_n^{\text{Cl,bw}}, t)^{k_n} - P_H \right\|_{\infty} \\ & \leq \left\| P_{\text{Cl},n} M(\mu_{H,1}, t) P_{\text{Cl},n} - P_H \right\|_{\infty} + \frac{1}{2^{2n} - 1} \\ & = \left\| M(\sigma, t) - P_H \right\|_{\infty} + \frac{1}{2^{2n} - 1} \\ & \leq 1 - \frac{1}{2} \frac{3}{2^{2n} - 1} + \frac{1}{2^{2n} - 1} \\ & \leq 1 - \frac{1}{2(2^{2n} - 1)}, \end{aligned} \quad (57)$$

where we used Lemma 3 in the third inequality.

$M(\nu_n^{\text{Cl,bw}}, t)^{k_n} M(\mu_{H,1}, t) M(\nu_n^{\text{Cl,bw}}, t)^{k_n}$  is a product of  $(2nk_n + 1)$  orthogonal projectors. In the following we denote these projectors as  $\mathbb{1} - Q_1, \dots, \mathbb{1} - Q_{2nk_n+1}$ . E.g. we set

$$Q_1 := \mathbb{1} - P_{\text{Cl},2} \otimes \mathbb{1}_{n-2} \quad (58)$$

and  $Q_{2nk_n+1} := \mathbb{1} - P_{H,1}$ . The order of these labels will not matter in the following. As in Lemma 2, we define a Hamiltonian  $\tilde{H} := \sum_{i=1}^{2nk_n+1} Q_i$ . We can now relate the bound (57) to the gap of the Hamiltonian  $\tilde{H}$  via the quantum union bound or

converse detectability lemma (Lemma 2):

$$\begin{aligned}
& \|M(\nu_n^{\text{Cl,bw}}, t)^{k_n} M(\mu_{H,1}, t) M(\nu_n^{\text{Cl,bw}}, t)^{k_n} - P_H\|_\infty \\
&= \max_{|\psi^\perp\rangle} \|M(\nu_n^{\text{Cl,bw}}, t)^{k_n} M(\mu_{H,1}, t) M(\nu_n^{\text{Cl,bw}}, t)^{k_n} |\psi^\perp\rangle\|_2 \\
&= \max_{|\psi^\perp\rangle} \left\| \prod_{i=1}^{2nk_n+1} (\mathbb{1} - Q_i) |\psi^\perp\rangle \right\|_2 \\
&\geq \sqrt{1 - 4\Delta(\tilde{H})} \\
&\geq 1 - 4\Delta(\tilde{H}).
\end{aligned} \tag{59}$$

Combining this with (57), yields:

$$\frac{1}{8} \frac{1}{2^{2n} - 1} \leq \Delta(\tilde{H}). \tag{60}$$

We still need to relate the gap of  $\tilde{H}$  to that of  $H_{n,t}$ . Using the operator inequalities  $P_{\text{Cl}} \geq P_H$  and  $P_{H,1} \geq P_{H,2}$ , we obtain

$$\begin{aligned}
\tilde{H} &= 2k_n \sum_{i=1}^n (\mathbb{1} - \mathbb{1}_{i-1} \otimes P_{\text{Cl},2} \otimes \mathbb{1}_{n-i-1}) + (\mathbb{1} - P_{H,1} \otimes \mathbb{1}_{n-1}) \\
&\leq 2k_n \sum_{i=1}^n (\mathbb{1} - \mathbb{1}_{i-1} \otimes P_{H,2} \otimes \mathbb{1}_{n-i-1}) + (\mathbb{1} - P_{H,1} \otimes \mathbb{1}_{n-1}) \\
&\leq (2k_n + 1) \sum_{i=1}^n (\mathbb{1} - \mathbb{1}_{i-1} \otimes P_{H,2} \otimes \mathbb{1}_{n-i-1}) \\
&= (2k_n + 1) H_{n,t}.
\end{aligned} \tag{61}$$

Moreover, notice that  $\tilde{H}$  and  $H_{n,t}$  have the same ground state space with ground state energy 0 as the gate set consisting of all 2-local Clifford unitaries plus single qubit unitaries are universal [43]. Indeed, it was proven in Ref. [29] that the eigenvalue 1 eigenspace of a moment operator for a probability distribution with universal support equals the image of the Haar random moment operator. Consequently, we find

$$\Delta(H_{n,t}) \geq \frac{1}{8(2k_n + 1)(2^{2n} - 1)} \geq 120000^{-1} n^{-4} 2^{-2n}. \tag{62}$$

Together with Eq. (10), this implies Theorem 1.  $\square$

## 5 Outlook: Gaps from representations of the Clifford group?

In this work, we improved the scaling of the design depth for local random quantum circuits over qubits. The lion's share of the improvement comes from the near-optimal convergence of an auxiliary random walk that interleaves global random Clifford unitaries and single qubit Haar random unitaries.

The key open question, posed in Ref. [9], is to either prove or disprove a subexponential unconditional gap. An unconditional gap that scales inverse polynomially

in  $n$  would imply a robust version of the Brown-Susskind conjecture for random quantum circuits. For approaches based on techniques from harmonic analysis, see Ref. [26]. The techniques presented in this paper open up another potential path towards this goal: It suffices to prove such a gap for the auxiliary walk defined in Eq. (26).

The commutant of tensor powers of the Clifford group was studied in Ref. [24, 39, 40] and it was proven in Ref. [27] that the action of a single qubit Haar average has a strong shrinking effect on a natural basis of the Clifford commutant labeled by Lagrangian subspaces of  $\mathbb{Z}_2^{2t}$ . Therefore, Lemma 13 in Ref. [27] can be viewed as a sanity check for the constant spectral gap conjecture.

To prove the generation of state designs in linear depth, we do not require a spectral gap of the entire moment operator. Indeed, it would suffice to show a spectral gap of the moment operator restricted to the endomorphisms of the symmetric subspace  $\text{End}(S^t[(\mathbb{C}^2)^{\otimes n}])$ :

$$\left\| (M(\nu_n, t) - P_H)|_{\text{vec}(\text{End}(S^t[(\mathbb{C}^2)^{\otimes n}]})} \right\|_{\infty} \leq 1 - \text{poly}^{-1}(n)? \quad (63)$$

To characterize the commutant of powers of the Clifford group on the endomorphisms of the symmetric subspace, one would need to understand how the symmetric subspace decomposes into irreducible representations of the Clifford group.

## 6 Acknowledgements

I am grateful to Nick Hunter-Jones, Richard Kueng and Felipe Montealegre-Mora for fruitful discussions. Moreover, I want to thank Jens Eisert for detailed comments on the manuscript. This work was funded by the foundational questions institute (FQXi).

## References

- [1] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 333–342, 2011. doi:10.1364/QIM.2014.QTh1A.2.
- [2] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004. doi:10.1103/PhysRevA.70.052328.
- [3] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: restructuring quantum information’s family tree. *Proc. R. Soc. A*, 465:2537, 2009. doi:10.1098/rspa.2009.0202.
- [4] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani. The Detectability Lemma and Quantum Gap Amplification. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC ’09, page 417, 2009. doi:10.1145/1536414.1536472.
- [5] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 20–30, 1998. doi:10.1145/276698.276708.

- [6] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *Computational Complexity, 2007. CCC '07. Twenty-Second Annual IEEE Conference on*, pages 129–140, June 2007. doi:10.1109/CCC.2007.26.
- [7] A. Anshu, I. Arad, and T. Vidick. Simple proof of the detectability lemma and spectral gap amplification. *Phys. Rev. B*, 93:205142, 2016. doi:10.1103/PhysRevB.93.205142.
- [8] J. Bourgain and A. Gamburd. A spectral gap theorem in  $su(d)$ . *Journal of the European Mathematical Society*, 14(5):1455–1511, 2012. doi:10.4171/JEMS/337.
- [9] 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. doi:10.1007/s00220-016-2706-8.
- [10] F. G. S. L. Brandao, A. W. Harrow, and M. Horodecki. Efficient quantum pseudorandomness. *Physical review letters*, 116(17):170502, 2016. doi:10.1103/PhysRevLett.116.170502.
- [11] Fernando GSL Brandão, Wissam Chemissany, Nicholas Hunter-Jones, Richard Kueng, and John Preskill. Models of quantum complexity growth. *PRX Quantum*, 2(3):030316, 2021. doi:10.1103/PRXQuantum.2.030316.
- [12] S. Bravyi and D. Maslov. Hadamard-free circuits expose the structure of the Clifford group. *IEEE Transactions on Information Theory*, 67(7):4546–4563, 2021. doi:10.1109/TIT.2021.3081415.
- [13] A. R. Brown and L. Susskind. Second law of quantum complexity. *Phys. Rev.*, D97:086015, 2018. doi:10.1103/PhysRevD.97.086015.
- [14] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, page 223, 1997. doi:10.1109/SFCS.1997.646111.
- [15] I. Chatzigeorgiou. Bounds on the Lambert function and their application to the outage analysis of user cooperation. *IEEE Communications Letters*, 17(8):1505–1508, 2013. doi:10.1109/LCOMM.2013.070113.130972.
- [16] R. Cleve, D. Leung, L. Liu, and C. Wang. Near-linear constructions of exact unitary 2-designs. *Quant. Inf. Comp.*, 16:0721–0756, 2015. doi:10.26421/QIC16.9-10-1.
- [17] C. Dankert. Efficient simulation of random quantum states and operators, 2005. doi:10.48550/arXiv.quant-ph/0512217.
- [18] 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. doi:10.1103/PhysRevA.80.012304.
- [19] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walk on finite groups. *The Annals of Probability*, pages 2131–2156, 1993. doi:10.1214/aoap/1177005359.

- [20] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal. Quantum data hiding. *IEEE, Trans. Inf Theory*, 48:3580–599, 2002. doi:[10.48550/arXiv.quant-ph/0103098](https://doi.org/10.48550/arXiv.quant-ph/0103098).
- [21] J. Emerson, R. Alicki, and K. Życzkowski. Scalable noise estimation with random unitary operators. *J. Opt. B: Quantum Semiclass. Opt.*, 7(10):S347, 2005. doi:[10.1088/1464-4266/7/10/021](https://doi.org/10.1088/1464-4266/7/10/021).
- [22] J. Gao. Quantum union bounds for sequential projective measurements. *Phys. Rev. A*, 92:052331, 2015. arXiv:[1410.5688](https://arxiv.org/abs/1410.5688), doi:[10.1103/PhysRevA.92.052331](https://doi.org/10.1103/PhysRevA.92.052331).
- [23] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: On the structure of unitary designs. *J. Math. Phys.*, 48:052104, 2007. doi:[10.1063/1.2716992](https://doi.org/10.1063/1.2716992).
- [24] D. Gross, S. Nezami, and M. Walter. Schur–Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations. *Communications in Mathematical Physics*, 385(3):1325–1393, 2021. doi:[10.1007/s00220-021-04118-7](https://doi.org/10.1007/s00220-021-04118-7).
- [25] J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, and N. Yunger Halpern. Linear growth of quantum circuit complexity. *Nature Physics*, 18:528–532, 2021. doi:[10.1038/s41567-022-01539-6](https://doi.org/10.1038/s41567-022-01539-6).
- [26] J. Haferkamp and N. Hunter-Jones. Improved spectral gaps for random quantum circuits: large local dimensions and all-to-all interactions. *Physical Review A*, 104(2):022417, 2021. doi:[10.1103/PhysRevA.104.022417](https://doi.org/10.1103/PhysRevA.104.022417).
- [27] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth. Quantum homeopathy works: Efficient unitary designs with a system-size independent number of non-Clifford gates. 2020. doi:[10.48550/arXiv.2002.09524](https://doi.org/10.48550/arXiv.2002.09524).
- [28] A. Harrow and S. Mehraban. Approximate unitary  $t$ -designs by short random quantum circuits using nearest-neighbor and long-range gates. *arXiv preprint arXiv:1809.06957*, 2018. doi:[10.48550/arXiv.1809.06957](https://doi.org/10.48550/arXiv.1809.06957).
- [29] A. W. Harrow and R. A. Low. Random quantum circuits are approximate 2-designs. *Communications in Mathematical Physics*, 291(1):257–302, 2009. doi:[10.1007/s00220-009-0873-6](https://doi.org/10.1007/s00220-009-0873-6).
- [30] P. Hayden and J. Preskill. Black holes as mirrors: Quantum information in random subsystems. *JHEP*, 09:120, 2007. doi:[10.1088/1126-6708/2007/09/120](https://doi.org/10.1088/1126-6708/2007/09/120).
- [31] N. Hunter-Jones. Unitary designs from statistical mechanics in random quantum circuits. 2019. arXiv:[1905.12053](https://arxiv.org/abs/1905.12053).
- [32] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *The Annals of Probability*, 34(4):1497–1529, 2006. doi:[10.1214/009117906000000205](https://doi.org/10.1214/009117906000000205).
- [33] E. Knill. Approximation by quantum circuits. *arXiv preprint*, 1995. doi:[10.48550/arXiv.quant-ph/9508006](https://doi.org/10.48550/arXiv.quant-ph/9508006).

- [34] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, 2008. doi:10.1103/PhysRevA.77.012307.
- [35] L. Leone, S. F. E. Oliviero, Y. Zhou, and A. Hama. Quantum chaos is quantum. *Quantum*, 5:453, 2021. doi:10.22331/q-2021-05-04-453.
- [36] R. A. Low. Pseudo-randomness and Learning in Quantum Computation. *arXiv preprint*, 2010. PhD Thesis, 2010. doi:10.48550/arXiv.1006.5227.
- [37] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, 2012. arXiv:1109.6887, doi:10.1103/PhysRevA.85.042311.
- [38] R. Mezher, J. Ghalbouni, J. Dgheim, and D. Markham. Efficient quantum pseudorandomness with simple graph states. *Physical Review A*, 97(2):022333, 2018. doi:10.1103/PhysRevA.97.022333.
- [39] F. Montealegre-Mora and D. Gross. Rank-deficient representations in the theta correspondence over finite fields arise from quantum codes. *Representation Theory of the American Mathematical Society*, 25(8):193–223, 2021. doi:10.1090/ert/563.
- [40] F. Montealegre-Mora and D. Gross. Duality theory for Clifford tensor powers. *arXiv preprint*, 2022. doi:10.48550/arXiv.2208.01688.
- [41] B. Nachtergaele. The spectral gap for some spin chains with discrete symmetry breaking. *Commun. Math. Phys.*, 175:565, 1996. doi:10.1007/BF02099509.
- [42] Y. Nakata, C. Hirche, M. Koashi, and A. Winter. Efficient quantum pseudorandomness with nearly time-independent hamiltonian dynamics. *Physical Review X*, 7(2):021006, 2017. doi:10.1103/PhysRevX.7.021006.
- [43] G. Nebe, E. M. Rains, and N. J. A Sloane. The invariants of the Clifford groups. *arXiv preprint*, 2001. doi:10.48550/arXiv.math/0001038.
- [44] R. I. Oliveira. On the convergence to equilibrium of Kac’s random walk on matrices. *Ann. Appl. Probab.*, 19:1200, 2009. doi:10.1214/08-AAP550.
- [45] S. F. E. Oliviero, L. Leone, and A. Hama. Transitions in entanglement complexity in random quantum circuits by measurements. *Physics Letters A*, 418:127721, 2021. doi:10.1016/j.physleta.2021.127721.
- [46] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert. Mixing properties of stochastic quantum Hamiltonians. *Communications in Mathematical Physics*, 355(3):905–947, 2017. doi:10.1007/s00220-017-2950-6.
- [47] M. Oszmaniec, A. Sawicki, and M. Horodecki. Epsilon-nets, unitary designs and random quantum circuits. *IEEE Transactions on Information Theory*, 2021. doi:10.1109/TIT.2021.3128110.
- [48] L. Susskind. Black Holes and Complexity Classes. *arXiv preprint*, 2018. doi:10.48550/arXiv.1802.02175.



- [49] P. P. Varjú. Random walks in compact groups. *Doc. Math.*, 18:1137–1175, 2013. doi:10.48550/arXiv.1209.1745.
- [50] J. Watrous. *The theory of quantum information*. Cambridge university press, 2018. doi:10.1017/9781316848142.
- [51] Z. Webb. The Clifford group forms a unitary 3-design. *Quantum Info. Comput.*, 16:1379, 2016. doi:10.5555/3179439.3179447.
- [52] S. Zhou, Z. Yang, A. Hamma, and C. Chamon. Single T gate in a Clifford circuit drives transition to universal entanglement spectrum statistics. *SciPost Physics*, 9(6):087, 2020.
- [53] H. Zhu. Multiqubit clifford groups are unitary 3-designs. *Phys. Rev. A*, 96:062336, 2017. doi:10.1103/PhysRevA.96.062336.

## A Proof of Observation 1

In this appendix we show how Observation 1 follows from bounds on Lambert’s W function. The improvement obtained in Observation 1 makes the difference between a design depth of  $O(nt^6)$  and a design depth of  $O(nt^{5+o(1)})$ .

*Proof.* Eq. (16) is implied by

$$l - \frac{1}{2 \ln(2)} \ln(l) \geq 2 \log_2(4t). \quad (64)$$

First, we consider the equation  $y - a \ln(y) = b$ . This is equivalent to

$$e^{-\frac{y}{a}} \left( -\frac{y}{a} \right) = -\frac{e^{-\frac{b}{a}}}{a}. \quad (65)$$

If the right hand side of Eq. (65) is larger than  $-1/e$ , this can be solved by

$$y = -a W_{-1} \left( -\frac{e^{-b/a}}{a} \right), \quad (66)$$

where  $W_{-1}$  denotes one of the branches [15] of Lambert’s W function (also called log product function). By definition, all branches of Lambert’s W function solve the equation

$$y e^y = x \quad (67)$$

for  $y$ .

We can now use the following bound for  $x > 0$  proven in Ref. [15]:

$$W_{-1}(e^{-x-1}) \geq -1 - (2x)^{\frac{1}{2}} - x. \quad (68)$$

With  $x = b/a + \ln(a) - 1$ ,  $b = 2 \log_2(4t)$  and  $a = (2 \ln(2))^{-1}$ , this implies

$$\begin{aligned} y &\leq -a \left( -1 - \sqrt{2 \left( \frac{b}{a} + \ln(a) - 1 \right)} - \frac{b}{a} - \ln(a) + 1 \right) \\ &\leq 2 \log_2(4t) + 1.5 \sqrt{\log_2(4t)}. \end{aligned} \quad (69)$$

The LHS of Eq. (64) is monotone in  $l$ . Hence, choosing

$$l = \left\lceil 2 \log_2(4t) + 1.5 \sqrt{\log_2(4t)} \right\rceil \quad (70)$$

implies Eq. (64) and thus, by Nachtergaele's bound, Eq. (17).  $\square$

Ref. [HHJ21] is not contained in the online version for copyright reasons.

In random quantum circuits the quantum circuit complexity serves as a proxy for studying quantum chaotic Hamiltonian dynamics, with profound implications for a broad range of quantum complex many-body systems; for instance, quantum complexity is increasingly understood as being essential to capture properties of black hole horizons. In particular, connections between gate complexity and holography in high-energy physics, in the context of the anti-de-Sitter-space/conformal-field-theory (AdS/CFT) correspondence have been prominently considered [Sus16, SS14, BRS<sup>+</sup>16, BS18b, BFV19]. The “complexity equals volume” conjecture [SS14] suggests that the correspondence’s boundary state has a complexity proportional to the volume behind the event horizon of a black hole in the bulk geometry. This conjecture was developed in the context of the wormhole growth paradox: The volume of a wormhole grows at a steady rate for an exponentially long time. However, if the volume corresponds to a local observable in the dual quantum theory, we would expect it to saturate quickly due to thermalization. As a resolution, Susskind and collaborators suggest that the volume could instead correspond to a measure of the inherent complexity of the dual state such as the circuit complexity. These connections have motivated studies of quantum complexity as a means of illuminating quantum many-body systems’ complex behaviors.

In related work, Brown and Susskind conjecture that quantum circuits’ complexity generically grows linearly for an exponentially long time [BS18a, Sus18]. The intuition behind this conjecture is that the space of all unitaries is vast and unless the circuits in question are exponentially deep, it should be more likely to expand into new territory by applying a gate. In particular, collisions between circuits should be rare and most quantum circuits are expected to not hit an accidental short-cut to the identity.

In this work, we prove the above notorious conjecture for a straightforward notion of quantum circuit complexity: the minimal number of gates necessary to implement a unitary exactly. We denote this quantity by  $\mathcal{C}(\mathbf{U})$  and  $\mathcal{C}(|\psi\rangle)$  for states. Informally, an architecture is simply a way to arrange the gates. In a random quantum circuit in a given architecture we draw all gates independently from the Haar measure on  $SU(4)$ . We call a segment of an architecture a *backwards light cone* if there is a site from which a perturbation can spread to the full system-size in the reversed circuit. We prove the following theorem:

**Theorem 13** (Linear growth of complexity). *Let  $\mathbf{U}$  denote a unitary implemented by a random quantum circuit in an architecture that contains  $T$  disjoint backwards light cones. The unitary’s circuit complexity is lower-bounded as*

$$\mathcal{C}(\mathbf{U}) \geq \frac{T}{9} - \frac{n}{3}, \quad (3.13)$$

*with unit probability, until the number of gates grows to  $T \geq 4^n - 1$ . The same bound holds for  $\mathcal{C}(\mathbf{U}|0^n)$ , until  $T \geq 2^{n+1} - 1$ .*

Effectively, this means that almost all quantum circuits have no significant “short-cuts” [Pir22]. Moreover, this result provides direct evidence for Brown-Susskind conjecture for chaotic Hamiltonian dynamics such as the SYK model and a sanity check for the “complexity equals volume” conjecture.

The theorem requires the architecture to be a concatenation of blocks that each have a light cone. The reason for this is that resulting quantum circuits need to be sufficiently scrambling for

the complexity to grow steadily. In fact, the generality of Theorem 13 allows us to show that in fact *most architectures* admit linear growth of complexity. That is, if the circuit architecture itself is drawn at random, a similar linear lower bound on the complexity holds with exponentially high probability. Arguably, the most studied architecture for random quantum circuits is the *brickwork architecture*, consisting of staggered layers of nearest-neighbor gates on an  $n$ -qubit chain, which yields the following bound:

$$C_u(\mathcal{U}|0^n\rangle) \geq \frac{\# \text{ of layers}}{18n} - \frac{n}{3}. \quad (3.14)$$

Moreover, we prove that there exists an error  $\varepsilon > 0$  such that a  $\varepsilon$ -approximate version of circuit complexity is subject to the same bound with a probability arbitrarily close to 1.

The proof of Theorem 13 is surprisingly short, given the established difficulty of lower-bounding the exact circuit complexity. In order to do so, we introduce new mathematical methods into quantum information theory, which we regard as a relevant contribution in its own right. Our proof strategy combines differential topology and elementary algebraic geometry with an inductive construction of Clifford circuits.

An idea central to the proof is to view the quantum circuits in an architecture  $A$  as a smooth map  $F^A$  from an algebraic set  $SU(4)^R$  to another algebraic set and manifold  $SU(2^n)$ . The *Tarski-Seidenberg theorem*, a powerful result in the theory of quantifier elimination, guarantees that the image  $\mathcal{U}(A) = F^A(SU(4)^R)$  is a *semi-algebraic set* and as such has a well defined integer dimension. Since the publication of Ref. [HFK<sup>+</sup>22] shorter proofs for variants of the main result were found in Ref. [Liz2]. Both proofs are based on the (semi-)algebraic properties of  $\mathcal{U}(A)$ .

We call  $\dim \mathcal{U}(A)$  the *accessible dimension* of the architecture  $A$ . Using the algebraic geometric properties of  $SU(4)^R$ , we can show that unitaries in the image of an architecture generating a low dimensional set are of probability 0 with respect to drawing random circuits from an architecture that generates a higher dimensional set.  $\dim \mathcal{U}(A)$  can be linearly upper bounded and all unitaries with a fixed number of gates  $R$  are necessarily in the image of the contraction map  $F^{A'}$  for  $A'$  having  $R$  gates. As a consequence, it suffices to prove a lower bound on the dimension  $\dim \mathcal{U}(A)$  in terms of the number of disjoint complete backward light cones. We obtain such a lower bound in the following proposition:

**Proposition 4** (Lower bound of accessible dimension). *Let  $A_T$  denote an architecture with  $T$  disjoint complete backward light cones. Assume that  $A_T$  consists of causal slices of  $\leq L$  gates each. The architecture's accessible dimension is lower-bounded as  $\dim \mathcal{U}(A_T) \geq T$ .*

We achieve this by showing that the dimension  $\dim \mathcal{U}(A_T)$  equals the maximal rank of the contraction map  $F^A$  due to the properties of the unitary group as a real algebraic set. In more detail, we use the irreducibility of the unitary group in the Zariski topology to conclude that the rank of  $F^A$  is maximal except for a subset of measure 0. This allows us to invoke the constant rank theorem from differential topology to show that locally  $F^A$  is equivalent to its own derivative.

As a consequence, it suffices to show that there exists a single tuple of gates such that  $F^{A_T}$  has a rank of at least  $T$  at this tuple. We proceed to construct such an example of a circuit for which the Jacobian has an image of dimension at least  $T$ . We construct a tuple of Clifford gates such that 2-local Pauli operators inserted into the resulting circuit yield at least  $T$  different Pauli matrices.

The exact circuit complexity seems to be more tame than robust versions in general. In fact, a recent paper [JW22] provides explicit states with exponential exact circuit complexity, a task far out of reach for boolean functions. The exact complexity cannot be directly related to the separation of complexity classes and thus does not necessarily suffer from the same obstructions. Notice that the circuit complexity for boolean functions is robust to errors as they form a discrete set, so there is no classical analogue of it. This opens the fascinating possibility to test quantum versions of notorious conjectures with a rigorously defined notion of circuit complexity.



OPEN

# Linear growth of quantum circuit complexity

Jonas Haferkamp<sup>1,2</sup>✉, Philippe Faist<sup>1</sup>, Naga B. T. Kothakonda<sup>1,3</sup>, Jens Eisert<sup>1,2</sup> and Nicole Yunger Halpern<sup>4,5,6,7,8</sup>

**The complexity of quantum states has become a key quantity of interest across various subfields of physics, from quantum computing to the theory of black holes. The evolution of generic quantum systems can be modelled by considering a collection of qubits subjected to sequences of random unitary gates. Here we investigate how the complexity of these random quantum circuits increases by considering how to construct a unitary operation from Haar-random two-qubit quantum gates. Implementing the unitary operation exactly requires a minimal number of gates—this is the operation's exact circuit complexity. We prove a conjecture that this complexity grows linearly, before saturating when the number of applied gates reaches a threshold that grows exponentially with the number of qubits. Our proof overcomes difficulties in establishing lower bounds for the exact circuit complexity by combining differential topology and elementary algebraic geometry with an inductive construction of Clifford circuits.**

Complexity is a pervasive concept at the intersection of computer science, quantum computing, quantum many-body systems and black hole physics. In general, complexity quantifies the resources required to implement a computation. For example, the complexity of a Boolean function can be defined as the minimal number of gates, chosen from a given gate set, necessary to evaluate the function. In quantum computing, the circuit model provides a natural measure of complexity for pure states and unitaries: a unitary transformation's quantum circuit complexity is the size, measured with the number of gates, of the smallest circuit that effects the unitary. Similarly, a pure state's quantum circuit complexity definable is the size of the smallest circuit that produces the state from a product state.

Quantum circuit complexity, by quantifying the minimal size of any circuit that implements a given unitary, is closely related to computational notions of complexity. The latter quantify the difficulty of solving a given computational task with a quantum computer and determine quantum complexity classes. Yet quantum circuit complexity can subtly differ from computational notions of quantum complexity: the computational notion depends on the difficulty of finding the circuit. In the following, we refer to quantum circuit complexity as 'quantum complexity' for convenience.

Quantum complexity has risen to prominence recently due to connections between gate complexity and holography in high-energy physics, in the context of the anti-de-Sitter space/conformal field theory (AdS/CFT) correspondence<sup>1–5</sup>. In the bulk theory, a wormhole's volume grows steadily for exponentially long times. By contrast, in boundary quantum theories, local observables tend to thermalize much more quickly. This contrast is known as the 'wormhole-growth paradox'<sup>1</sup>. It appears to contradict the AdS/CFT correspondence, which postulates a mapping of physical operators between the bulk theory and a quantum boundary theory. A resolution has been proposed in the 'complexity equals volume' conjecture: the wormhole's volume is conjectured to be dual not to a local quantum observable, but to the boundary state's quantum complexity<sup>2</sup>. Similarly, the 'complexity equals action' conjecture posits

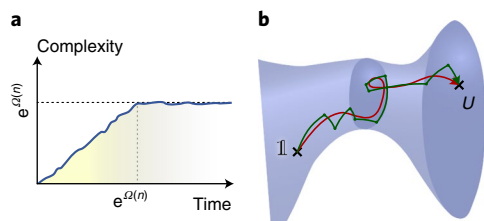
that a holographic state's complexity is dual to a certain space–time region's action<sup>6</sup>.

A counting argument reveals that the vast majority of unitaries have near-maximal complexities<sup>7</sup>. Yet lower-bounding the quantum complexity is a long-standing open problem in quantum information theory. The core difficulty is that the gates performed early in a circuit may partially cancel with gates performed later. One can rarely rule out the existence of a 'shortcut', a seemingly unrelated but smaller circuit that generates the same unitary. Consequently, quantum-gate-synthesis algorithms, which decompose a given unitary into gates, run for times exponential in the system size<sup>8</sup>. Approaches to lower-bounding unitaries' quantum complexities include Nielsen's geometric picture<sup>9–13</sup>.

A key question in the study of quantum complexity is the following. Consider constructing deeper and deeper circuits for an  $n$ -qubit system, by applying random two-qubit gates. At what rate does the circuit complexity increase? Brown and Susskind conjectured that the complexity of quantum circuits generically grows linearly for an exponentially long time<sup>4,14</sup>. Intuitively, the conjecture is that most circuits are fundamentally 'incompressible': no substantially shorter quantum circuit effects the same unitary. Quantum complexity, if it grows linearly with a generic circuit's depth, strongly supports the 'complexity equals volume' conjecture as a proposal to the wormhole-growth paradox<sup>1,2</sup>. The conjecture therefore implies that complexity growth is as generic as thermalization<sup>15,16</sup> and operator growth<sup>17,18</sup> (the spreading of an initially local operator's support in the Heisenberg picture). However, in contrast to easily measurable physical quantities, which thermalize rapidly, complexity grows for an exponentially long time. Brown and Susskind have supported their conjecture using Nielsen's geometric approach (Fig. 1b)<sup>9–12</sup>.

Brandão et al.<sup>19</sup> recently proved a key result about the growth of quantum complexity under random circuits. The authors leveraged the mathematical toolbox of  $t$ -designs, finite collections of unitaries that approximate completely random unitaries. A  $t$ -design is a probability distribution, over unitaries, whose first  $t$  moments equal the Haar measure's moments<sup>20–22</sup>. The Haar measure is the unique

<sup>1</sup>Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin, Germany. <sup>2</sup>Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany. <sup>3</sup>Institute for Theoretical Physics, University of Cologne, Cologne, Germany. <sup>4</sup>ITAMP, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA. <sup>5</sup>Department of Physics, Harvard University, Cambridge, MA, USA. <sup>6</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>7</sup>Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>8</sup>Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA. ✉e-mail: [jonas.haferkamp@fu-berlin.de](mailto:jonas.haferkamp@fu-berlin.de)



**Fig. 1 | The geometric approach to complexity provides a strong intuitive and physical basis for the complexity growth conjecture that we prove.**

**a.** The complexity has been conjectured to grow linearly under random quantum circuits until times exponential in the number  $n$  of qubits<sup>4</sup>. The blue region depicts part of the space of  $n$ -qubit unitaries. A unitary  $U$  has a complexity that we define as the minimal number of two-qubit gates necessary to effect  $U$  (green jagged path; each path segment represents a gate). Nielsen’s complexity<sup>9–12</sup>, involved in ref. 4, attributes a high metric cost to directions associated with nonlocal operators. In this geometry, the unitary’s complexity is the shortest path that connects  $I$  to  $U$  (red line). Nielsen’s geometry suggests the toolbox of differential geometry, avoiding circuits’ discreteness. The circuit complexity upper-bounds Nielsen’s complexity; opposite bounds hold for approximate circuit complexity<sup>12</sup>.

unitarily invariant probability measure over a compact group. Reference<sup>19</sup> proved that quantum complexity robustly grows polynomially in a random circuit’s size. The complexity’s growth was shown to be linear in the circuit’s size if the local Hilbert-space dimension is large.

We prove that the complexity of a random circuit grows linearly with time (with the number of gates applied). We consider unitaries constructed from quantum circuits composed of Haar-random two-qubit gates. The focus of our proof is the set of unitaries that can be generated with a fixed arrangement of gates. We show that this set’s dimension, which we call accessible dimension, serves as a good proxy for the quantum complexity of almost every unitary in the set. Our bound on the complexity holds for all random circuits described above, with probability 1. Instead of invoking unitary designs<sup>19</sup> or Nielsen’s geometric approach<sup>9–12</sup>, we employ elementary aspects of differential topology and algebraic geometry, combined with an inductive construction of Clifford circuits. The latter are circuits that transform Pauli strings to Pauli strings up to a phase<sup>23–27</sup>.

This work is organized as follows. First, we introduce the set-up and definitions. Second, we present the main result, the complexity’s exponentially long linear growth. Third, we present a high-level overview of the proof. The key mathematical steps follow, in the Methods. Two corollaries follow: an extension to random arrangements of gates and an extension to slightly imperfect gates. In the Discussion we compare our results with known results and explain our work’s implications for various subfields of quantum physics. Finally, we discuss the opportunities engendered by this work. In Supplementary Appendix A we review the elementary algebraic geometry required for the proof. Proof details are provided in Supplementary Appendix B. We elaborate on states’ complexities in Supplementary Appendix C. We prove two corollaries in Supplementary Appendices D and E. Finally, we compare notions of circuit complexity in Supplementary Appendix F.

**Preliminaries.** This work concerns a system of  $n$  qubits. For convenience, we assume that  $n$  is even. We simplify tensor-product notation as  $|0^k\rangle := |0\rangle^{\otimes k}$ , for  $k = 1, 2, \dots, n$ , and  $\mathbb{1}_k$  denotes the  $k$ -qubit identity operator. Let  $U_{j,k}$  denote a unitary gate that operates on qubits  $j$  and  $k$ . Such gates need not couple the qubits together and need not be geometrically local. An architecture is an arrangement of some fixed number  $R$  of gates (Fig. 2a).

**Definition 1. (Architecture)** An architecture is a directed acyclic graph that contains  $R \in \mathbb{Z}_{>0}$  vertices (gates). Two edges (qubits) enter each vertex, and two edges exit.

Figure 2b,c illustrates example architectures governed by our results.

- A brickwork is the architecture of any circuit formed as follows. Apply a string of two-qubit gates:  $U_{1,2} \otimes U_{3,4} \otimes \dots \otimes U_{n-1,n}$ . Then apply a staggered string of gates, as shown in Fig. 2b. Perform this pair of steps  $T$  times in total, using possibly different gates each time.
- A staircase is the architecture of any circuit formed as in Fig. 2c. Apply a stepwise string of two-qubit gates:  $U_{n,n-1} U_{n-2,n-1} \dots U_{2,1}$ . Repeat this process  $T$  times, using possibly different gates each time.

The total number of gates in the brickwork architecture, as in the staircase architecture, is  $R = (n - 1)T$ . Our results extend to more general architectures, for example, the architecture depicted in Fig. 2a and architectures of non-nearest-neighbour gates. Circuits of a given architecture can be formed randomly.

**Definition 2. (Random quantum circuit)** Let  $A$  denote an arbitrary architecture. A probability distribution can be induced over the architecture- $A$  circuits as follows: for each vertex in  $A$ , draw a gate Haar-randomly from  $SU(4)$ . Then contract the unitaries along the edges of  $A$ . Each circuit so constructed is called a random quantum circuit.

Implementing a unitary with the optimal gates, in the optimal architecture, concretizes the notion of complexity.

**Definition 3. (Exact circuit complexities)** Let  $U \in SU(2^n)$  denote an  $n$ -qubit unitary. The (exact) circuit complexity  $C_u(U)$  is the least number of two-qubit gates in any circuit that implements  $U$ . Similarly, let  $|\psi\rangle$  denote a pure quantum state vector. The (exact) state complexity  $C_{\text{state}}(|\psi\rangle)$  is the least number  $r$  of two-qubit gates  $U_1, U_2, \dots, U_r$ , arranged in any architecture, such that  $U_1 U_2 \dots U_r |0^n\rangle = |\psi\rangle$ .

We now define a backwards light cone, a concept that helps us focus on sufficiently connected circuits. Consider creating two vertical cuts in a circuit (dashed lines, Fig. 2). The gates between the cuts form a block. We say that a block contains a backwards light cone if some qubit  $t$  links to each other qubit  $t'$  via a directed path of gates (a path that may be unique to  $t'$ ). The backwards light cone consists of the gates in the paths.

**Main result, linear growth of complexity in random quantum circuits.** Our main result is a lower bound on the complexities of random unitaries and states. The bound holds with unit probability.

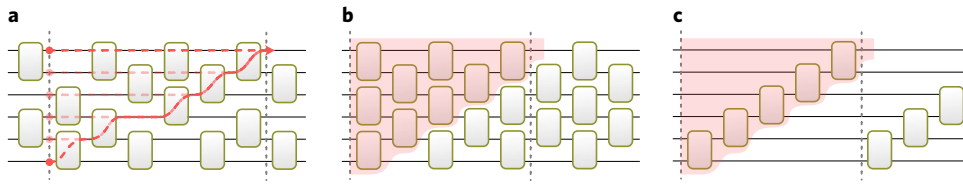
**Theorem 1. (Linear growth of complexity)** Let  $U$  denote a unitary implemented by a random quantum circuit in an architecture formed by concatenating  $T$  blocks of  $\leq L$  gates each, each block containing a backwards light cone. The unitary’s circuit complexity is lower-bounded as

$$C_u(U) \geq \frac{R}{9L} - \frac{n}{3}, \tag{1}$$

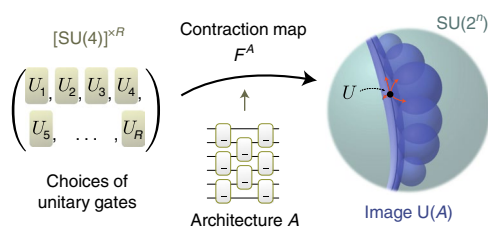
with unit probability, until the number of gates grows to  $T \geq 4^n - 1$ . The same bound holds for  $C_{\text{state}}(U|0^n)$ , until  $T \geq 2^{n+1} - 1$ .

The theorem governs all architectures that contain enough backwards light cones. The brickwork architecture forms a familiar special case. Let us choose for a brickwork’s blocks to contain  $2n$  of the columns in Fig. 2b. Each block contains  $L = n(n - 1)$  gates (in the absence of periodic boundary conditions), yielding the lower bound  $C_u(U) \geq \frac{R}{9n(n-1)} - \frac{n}{3}$ . Another familiar example is the staircase architecture. A staircase’s blocks can have the least  $L$  possible,  $n - 1$ , which yields the strongest bound.





**Fig. 2 | Our result relies on architectures and their backwards light cones. a**, An architecture specifies how  $R$  two-qubit gates are arranged in an  $n$ -qubit circuit. The gates need not be applied to neighbouring qubits, although they are depicted in this way for convenience. Our result involves blocks with the following property: the block contains a qubit reachable from each other qubit via a path (red dashed line), possibly unique to the latter qubit, that passes only through gates in the block. **b**, The brickwork architecture interlaces layers of gates on a one-dimensional (1D) chain. In a 1D architecture with geometrically local gates, such as the brickwork architecture, each block has a backwards light cone (light red region) that touches the qubit chain's edges. In the brickwork architecture, a minimal backwards-light-cone-containing block consists of  $\sim n^2$  gates. **c**, The staircase architecture, too, acts on a 1D qubit chain. The circuit consists of layers in which  $n - 1$  gates act on consecutive qubit pairs. A minimal backwards-light-cone-containing block consists of  $n - 1$  gates.



**Fig. 3 | The  $R$ -gate architecture  $A$  is associated with a contraction map  $F^A$ .**  $F^A$  maps a list of input gates (a point in  $[\text{SU}(4)]^{\times R}$ ) to an  $n$ -qubit unitary  $U$  in  $\text{SU}(2^n)$ . The unitary results from substituting the gates into the architecture.  $F^A$  has an image  $\mathcal{U}(A)$ , which consists of the unitaries implementable with the architecture.  $A$  has an accessible dimension,  $d_A$ , equal to the dimension of  $\mathcal{U}(A)$ . Our core technical result is that  $d_A$  grows linearly with  $R$ . To bridge this result to complexity, consider an arbitrary architecture  $A'$  formed from fewer gates than a constant fraction of  $R$ . Such an architecture's accessible dimension satisfies  $d_{A'} < d_A$ , as we show. Therefore, every unitary in  $\mathcal{U}(A)$  has a complexity linear in  $R$ , except for a measure-0 set. The proof relies on algebraic geometry. A key concept is the rank of  $F^A$  at a point. The rank counts the local degrees of freedom in the image (orange arrows).

**High-level overview of the proof of Theorem 1.** Consider fixing an  $R$ -gate architecture  $A$ , then choosing the gates in the architecture. The resulting circuit implements some  $n$ -qubit unitary. All the unitaries implementable with  $A$  form a set  $\mathcal{U}(A)$  (compare Fig. 3). Our proof relies on properties of  $\mathcal{U}(A)$ —namely, on the number of degrees of freedom in  $\mathcal{U}(A)$ . We define this number as the architecture's accessible dimension,  $d_A = \dim(\mathcal{U}(A))$  (Fig. 3). The following section contains a formal definition; here, we provide intuition. As the  $n$ -qubit unitaries form a space of dimension  $4^n$ ,  $d_A \in [0, 4^n]$ . The greater the  $d_A$ , the more space  $\mathcal{U}(A)$  fills in the set of  $n$ -qubit unitaries. Considering  $\mathcal{U}(A)$  circumvents the intractability of calculating a unitary's circuit complexity. To better understand the form of  $\mathcal{U}(A)$ , we study the set's dimension, which is the accessible dimension. Importantly, the accessible dimension enables us to compare the sets  $\mathcal{U}(A)$  generated by different architectures. Distinct accessible dimensions imply that the lower-dimensional set has measure zero in the higher-dimensional set. As a proxy for quantum complexity, the accessible dimension plays a role similar to  $t$ -designs in refs. 19,28. Our first technical result lower-bounds the sufficiently connected architecture's accessible dimension.

**Proposition 1.** (Lower bound on accessible dimension) *Let  $A_T$  denote an architecture formed by concatenating  $T$  blocks of  $\leq L$  gates each, each block containing a backwards light cone. The architecture's accessible dimension is lower-bounded as*

$$d_{A_T} \geq T \geq \frac{R}{L}. \quad (2)$$

We can upper-bound  $d_A$ , for an arbitrary architecture  $A$ , by counting parameters. To synopsise the argument in Supplementary Appendix B, 15 real parameters specify each two-qubit unitary. Each qubit shared by two unitaries makes three parameters redundant. Hence

$$d_A \leq 9R + 3n. \quad (3)$$

The accessible dimension reaches its maximal value,  $4^n$ , after a number of gates exponential in  $n$ . Similarly, the circuit complexity reaches its maximal value after exponentially many gates. This parallel suggests  $d_A$  as a proxy for the circuit complexity. The next section justifies the use of  $d_A$  as a proxy.

The proof of Theorem 1 revolves around the accessible dimension  $d_{A_T}$  of a certain  $R$ -gate architecture  $A_T$ . The main idea is as follows. Let  $R'$  be less than a linear fraction of  $R$ . More specifically, let  $9R' + 3n < T = R/L$ . For every  $R'$ -gate architecture  $A'$ ,  $d_{A'} < d_{A_T}$  holds by a combination of equations (2) and (3). Consequently, Supplementary Appendix B shows that  $\mathcal{U}(A')$  has zero probability in  $\mathcal{U}(A_T)$ , according to the measure in Definition 2. Therefore, almost every unitary  $U \in \mathcal{U}(A_T)$  has a complexity greater than the greatest possible  $R'$ . Inequality (1) follows.

## Discussion

We have proven a prominent physics conjecture proposed by Brown and Susskind for random quantum circuits<sup>4,14</sup>: a local random circuit's quantum complexity grows linearly in the number of gates until reaching a value exponential in the system size. To prove this conjecture, we have introduced a technique for bounding complexity. The proof rests on our connecting the quantum complexity to the accessible dimension, the dimension of the set of unitaries implementable with a given architecture (arrangement of gates). Our core technical contribution is a lower bound on the accessible dimension. The bound rests on techniques from differential topology and algebraic geometry.

Theorem 1 is a rigorous demonstration of the linear growth of random qubit circuits' complexities for exponentially long times. The bound holds until the complexity reaches  $C_u(U) = \Omega(4^n)$ —the scaling, up to polynomial factors, of the greatest complexity achievable by any  $n$ -qubit unitary<sup>29</sup>. One hurdle has stymied attempts to prove that the quantum complexity of local random circuits grows linearly: most physical properties (described with, for example, local observables or correlation functions) reach fixed values in times subexponential in the system size. One must progress beyond such properties to prove that the complexity grows linearly at

superpolynomial times. We overcome this hurdle by identifying the accessible dimension as a proxy for the complexity.

Theorem 1 complements another rigorous insight about complexity growth. In ref. <sup>19</sup>, the linear growth of complexity is proven in the limit of large local dimension  $q$  and for a strong notion of quantum circuit complexity, with help from ref. <sup>30</sup>. Furthermore, depth- $T$  random qubit circuits have complexities that scale as  $\mathcal{O}(T^{1/11})$  until  $T = \exp(\mathcal{O}(n))$  (refs. <sup>19,22</sup>). The complexity scales the same way for other types of random unitary evolution, such as a continuous-time evolution under a stochastically fluctuating Hamiltonian<sup>31</sup>. Finally, ref. <sup>19</sup> addresses bounds on convergence to unitary designs<sup>22,30–32</sup>, translating these bounds into results about circuit complexity. Theorem 1 is neither stronger nor weaker than the results of ref. <sup>19</sup>, which govern a more operational notion of complexity—how easily  $U|0^n\rangle\langle 0^n|U^\dagger$  can be distinguished from the maximally mixed state.

Our work is particularly relevant to the holographic context surrounding the Brown–Susskind conjecture. There, random quantum circuits are conjectured to serve as proxies for chaotic quantum dynamics generated by local time-independent Hamiltonians<sup>33</sup>. Reference <sup>34</sup> has introduced this conjecture into black hole physics, and ref. <sup>1</sup> has discussed the conjecture in the context of holography. A motivation for invoking random circuits is that random circuits can be analysed more easily than time-independent Hamiltonian dynamics. Time-independent Hamiltonian dynamics are believed to be mimicked also by time-fluctuating Hamiltonians<sup>31</sup> and by random ensembles of Hamiltonians. Furthermore, complexity participates in analogies with thermodynamics, such as a second law of quantum complexity<sup>4</sup>. Our techniques can be leveraged to construct an associated resource theory of complexity<sup>35</sup>.

In the context of holography, the complexities of thermofield double states have attracted recent interest<sup>1,36–38</sup>. Thermofield double states are pure bipartite quantum states for which each subsystem's reduced state is thermal. In the context of holography, thermofield double states are dual to eternal black holes in anti-de-Sitter space<sup>36</sup>. Such a black hole's geometry consists of two sides connected by a wormhole, or Einstein–Rosen bridge. The wormhole's volume grows for a time exponential in the number of degrees of freedom of the boundary theory<sup>1,4</sup>. As discussed above, random quantum circuits are expected to capture the (presumed Hamiltonian) dynamics behind the horizon. If they do, the growth of the wormhole's volume is conjectured to match the growth of the boundary state's complexity<sup>1,2,4</sup>; both are expected to reach a value exponentially large in the number of degrees of freedom. Our results govern the random circuit that serves as a proxy for the dynamics behind the horizon. That random circuit's complexity, our results show strikingly, indeed grows to exponentially large values. This conclusion reinforces the evidence that quantum circuit complexity is the right quantity with which to resolve the wormhole-growth paradox<sup>1</sup>.

## Outlook

Our main result governs exact circuit complexity. In Supplementary Corollary 2, we generalize the result to a slightly robust notion of circuit complexity. There, the complexity depends on our tolerance of the error in the implemented unitary. Yet, the error tolerance can be uncontrollably small. The main challenge in extending our results to approximate complexity is that the accessible dimension crudely characterizes the set of unitaries implementable with a given architecture. Consider attempting to enlarge this set to include all the  $n$ -qubit unitaries that lie close to the set in some norm. The enlarged set's dimension is  $4^n$ . The reason for this is that the enlargement happens in all directions of  $SU(2^n)$ . Therefore, our argument does not work as for the exact complexity. Extending our results to approximations therefore offers an opportunity for future work. Approximations may also illuminate random circuits as instruments for identifying quantum advantages<sup>39,40</sup>; they would show that a polynomial-size quantum circuit cannot be compressed

substantially while achieving a good approximation. These observations motivate an uplifting of the present work to robust notions of quantum circuit complexity allowing for implementation errors in the distinguishability of states or channels<sup>41</sup> (see, for example, ref. <sup>19</sup>). A possible uplifting might look as follows. Let  $A$  denote an  $R$ -gate architecture, and let  $A'$  denote an  $R'$ -gate architecture. Suppose that the accessible dimensions obey  $d_{A'} < d_A$ . A unitary implemented with  $A$  has no chance of occupying the set  $\mathcal{U}(A')$ , which has a smaller dimension than  $\mathcal{U}(A)$ . Consider enlarging  $\mathcal{U}(A')$  to include the unitaries that lie  $\epsilon$ -close, for some  $\epsilon > 0$ . If  $\mathcal{U}(A')$  is sufficiently smooth and well-behaved, we expect the enlarged set's volume, intersected with  $\mathcal{U}(A)$ , to scale as  $\sim \epsilon^{d_A - d_{A'}}$ . Furthermore, suppose that unitaries implemented with  $A$  are distributed sufficiently evenly in  $\mathcal{U}(A)$  (rather than being concentrated close to  $\mathcal{U}(A')$ ). All the unitaries in  $\mathcal{U}(A)$  except a small fraction  $\sim \epsilon^{d_A - d_{A'}}$  could not lie in  $\mathcal{U}(A')$ . We expect, therefore, that all the unitaries in  $\mathcal{U}(A)$  except a fraction  $\sim \epsilon^{d_A - d_{A'}}$  have  $\epsilon$ -approximate complexities greater than  $R'$ .

A related opportunity is a proof that Nielsen's geometric complexity measure grows linearly under random circuits. Such a proof probably requires a more refined characterization of  $\mathcal{U}(A)$  than its dimension. The quantum complexity in Theorem 1 does not lower-bound Nielsen's complexity. Hence our main results do not immediately imply a similar bound for Nielsen's complexity. However, proving the approximate circuit complexity's linear growth would suffice to lower-bound Nielsen's complexity because of the known inequalities between Nielsen's complexity and the circuit complexity (Fig. 1b; for example, ref. <sup>12</sup>).

We expect our machinery based on geometry<sup>42–47</sup> and properties of the Clifford<sup>27,48,49</sup> group to be applicable to random processes that more closely reflect a variety of systems that are studied in the many-body physics community. Examples include randomly fluctuating dynamics<sup>31</sup>, which implement random quantum circuits when Trotterized, and thermofield double states undergoing random ‘shocks’<sup>5,50,51</sup>. Additionally, hybrid circuits—random unitary circuits punctuated by intermediate measurements—have recently attracted much interest<sup>52,53</sup>, as the amount of entanglement present in such systems appears to undergo phase transitions induced by the rate at which they are measured. A generalization of the accessible dimension to such systems might reveal to what extent circuit complexity, as a measure of entanglement in deep dynamics, undergoes similar phase transitions. We hope that the present work, by innovating machinery for addressing complexity, stimulates further quantitative studies of holography, scrambling and chaotic quantum dynamics.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-022-01539-6>.

Received: 5 July 2021; Accepted: 4 February 2022;

Published online: 28 March 2022

## References

- Susskind, L. Computational complexity and black hole horizons. *Fortsch. Phys.* **64**, 24–43 (2016).
- Stanford, D. & Susskind, L. Complexity and shock wave geometries. *Phys. Rev. D.* **90**, 126007 (2014).
- Brown, A. R., Roberts, D. A., Susskind, L., Swingle, B. & Zhao, Y. Complexity, action and black holes. *Phys. Rev. D.* **93**, 086006 (2016).
- Brown, A. R. & Susskind, L. Second law of quantum complexity. *Phys. Rev. D.* **97**, 086015 (2018).
- Bouland, A., Fefferman, B. & Vazirani, U. Computational pseudorandomness, the wormhole growth paradox and constraints on the AdS/CFT duality. In *Proc. ITCS'20* (2020).

6. Brown, A. R., Roberts, D. A., Susskind, L., Swingle, B. & Zhao, Y. Holographic complexity equals bulk action? *Phys. Rev. Lett.* **116**, 191301 (2016).
7. Poulin, D., Qarry, A., Somma, R. & Verstraete, F. Quantum simulation of time-dependent Hamiltonians and the convenient illusion of Hilbert space. *Phys. Rev. Lett.* **106**, 170501 (2011).
8. Gosset, D., Kliuchnikov, V., Mosca, M. & Russo, V. An algorithm for the  $T$ -count. *Quant. Inf. Comput.* **14**, 1277–1301 (2014).
9. Nielsen, M. A. A geometric approach to quantum circuit lower bounds. Preprint at <https://arxiv.org/abs/quant-ph/0502070> (2005).
10. Nielsen, M. A., Dowling, M. R., Gu, M. & Doherty, A. C. Quantum computation as geometry. *Science* **311**, 1133–1135 (2006).
11. Nielsen, M. A., Dowling, M. R., Gu, M. & Doherty, A. C. Optimal control, geometry and quantum computing. *Phys. Rev. A* **73**, 062323 (2006).
12. Dowling, M. R. & Nielsen, M. A. The geometry of quantum computation. *Quant. Inf. Comput.* **8**, 861–899 (2008).
13. Eisert, J. Entangling power and quantum circuit complexity. *Phys. Rev. Lett.* **127**, 020501 (2021).
14. Susskind, L. Black holes and complexity classes. Preprint at <https://arxiv.org/abs/1802.02175> (2018).
15. Eisert, J., Friesdorf, M. & Gogolin, C. Quantum many-body systems out of equilibrium. *Nat. Phys.* **11**, 124–130 (2015).
16. Polkovnikov, A., Sengupta, K., Silva, A. & Vengalattore, M. Nonequilibrium dynamics of closed interacting quantum systems. *Rev. Mod. Phys.* **83**, 863–883 (2011).
17. Swingle, B. Unscrambling the physics of out-of-time-order correlators. *Nat. Phys.* **14**, 988–990 (2018).
18. Maldacena, J., Shenker, S. H. & Stanford, D. A bound on chaos. *J. High Energy Phys.* **1608**, 106 (2016).
19. Brandao, F. G. S. L., Chemsassy, W., Hunter-Jones, N., Kueng, R. & Preskill, J. Models of quantum complexity growth. *PRX Quantum* **2**, 030316 (2021).
20. Gross, D., Audenaert, K. M. R. & Eisert, J. Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.* **48**, 052104 (2007).
21. Dankert, C., Cleve, R., Emerson, J. & Livine, E. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 012304 (2009).
22. Brandão, F. G. S. L., Harrow, A. W. & Horodecki, M. Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.* **346**, 397–434 (2016).
23. Bolt, B., Room, T. G. & Wall, G. E. On the Clifford collineation, transform and similarity groups. I. *J. Austr. Math. Soc.* **2**, 60–79 (1961).
24. Bolt, B., Room, T. G. & Wall, G. E. On the Clifford collineation, transform and similarity groups. II. *J. Austr. Math. Soc.* **2**, 80–96 (1961).
25. Calderbank, A. R., Rains, E. M., Shor, P. W. & Sloane, N. J. A. Quantum error correction and orthogonal geometry. *Phys. Rev. Lett.* **78**, 405 (1997).
26. Calderbank, A. R., Rains, E. M., Shor, P. M. & Sloane, N. J. A. Quantum error correction via codes over  $GF(4)$ . *IEEE Trans. Inf. Theory* **44**, 1369–1387 (1998).
27. Gottesman, D. Stabilizer codes and quantum error correction. Preprint <https://arxiv.org/abs/quant-ph/9705052> (1997).
28. Brandão, F. G. S. L., Harrow, A. W. & Horodecki, M. Efficient quantum pseudorandomness. *Phys. Rev. Lett.* **116**, 170502 (2016).
29. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
30. Hunter-Jones, N. Unitary designs from statistical mechanics in random quantum circuits. Preprint at <https://arxiv.org/abs/1905.12053> (2019).
31. Onorati, E. et al. Mixing properties of stochastic quantum Hamiltonians. *Commun. Math. Phys.* **355**, 905 (2017).
32. Nakata, Y., Hirche, C., Koashi, M. & Winter, A. Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics. *Phys. Rev. X* **7**, 021006 (2017).
33. Nahum, A., Vijay, S. & Haah, J. Operator spreading in random unitary circuits. *Phys. Rev. X* **8**, 021014 (2018).
34. Hayden, P. & Preskill, J. Black holes as mirrors: quantum information in random subsystems. *J. High Energy Phys.* **0709**, 120 (2007).
35. Yunger Halpern, N. et al. Resource theory of quantum uncomplexity. Preprint at <https://arxiv.org/abs/2110.11371> (2021).
36. Maldacena, J. M. Eternal black holes in anti-de Sitter. *J. High Energy Phys.* **04**, 021 (2003).
37. Chapman, S. et al. Complexity and entanglement for thermofield double states. *SciPost Phys.* **6**, 034 (2019).
38. Susskind, L. Entanglement is not enough. Preprint at <https://arxiv.org/abs/1411.0690> (2014).
39. Neill, C. et al. A blueprint for demonstrating quantum supremacy with superconducting qubits. Preprint at <https://arxiv.org/abs/1709.06678> (2017).
40. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
41. Watrous, J. Semidefinite programs for completely bounded norms. *Theory Comput* **5**, 217–238 (2009).
42. Bochnak, J., Coste, M. & Roy, M.-F. *Real Algebraic Geometry* Vol. 36 (Springer, 2013).
43. Hall, B. *Lie Groups, Lie Algebras and Representations: an Elementary Introduction* Vol. 222 (Springer, 2015).
44. Lee, J. M. in *Introduction to Smooth Manifolds* (eds Axler, S. & Ribet, K.) 1–31, 2nd edn (Springer, 2013).
45. Milne, J. S. *Algebraic Groups: the Theory of Group Schemes of Finite Type over a Field* Vol. 170 (Cambridge Univ. Press, 2017).
46. Sard, A. Hausdorff measure of critical images on Banach manifolds. *Am. J. Math.* **87**, 158–174 (1965).
47. Khaneja, N. & Glaser, S. Cartan decomposition of  $SU(2^n)$ , constructive controllability of spin systems and universal quantum computing. Preprint at <https://arxiv.org/abs/quant-ph/0010100> (2000).
48. Cleve, R., Leung, D., Liu, L. & Wang, C. Near-linear constructions of exact unitary 2-designs. *Quant. Inf. Comput.* **16**, 0721–0756 (2016).
49. Aaronson, S. & Gottesman, D. Improved simulation of stabilizer circuits. *Phys. Rev. A* **70**, 052328 (2004).
50. Shenker, S. H. & Stanford, D. Black holes and the butterfly effect. *J. High Energy Phys.* **2014**, 67 (2014).
51. Shenker, S. H. & Stanford, D. Multiple shocks. *J. High Energy Phys.* **2014**, 1–20 (2014).
52. Li, Y., Chen, X. & Fisher, M. P. A. Measurement-driven entanglement transition in hybrid quantum circuits. *Phys. Rev. B* **100**, 134306 (2019).
53. Skinner, B., Ruhman, J. & Nahum, A. Measurement-induced phase transitions in the dynamics of entanglement. *Phys. Rev. X* **9**, 031009 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Having overviewed the proof at a high level, here we fill in the key mathematics. Three points need clarifying. First, we must rigorously define the accessible dimension, or the dimension of  $\mathcal{U}(A)$ , which is not a manifold. Second, we must prove Proposition 1. Finally, we must elucidate steps in the proof of Theorem 1. We address these points using the toolbox of algebraic geometry. We associate with every  $R$ -gate architecture  $A$  a contraction map  $F^A: \text{SU}(4)^{\times R} \rightarrow \text{SU}(2^n)$ . This function maps a list of gates to an  $n$ -qubit unitary. The unitary results from substituting the gates into the architecture  $A$  (Fig. 3). The map contracts every edge (qubit) shared by two vertices (gates) in  $A$ .

The image of  $F^A$  is the set  $\mathcal{U}(A)$  of unitaries implementable with the architecture  $A$ .  $\mathcal{U}(A)$  is a semialgebraic set, consisting of the solutions to a finite set of polynomial equations and inequalities over the real numbers (Supplementary Appendix A provides a review). That  $\mathcal{U}(A)$  is a semialgebraic set follows from the Tarski–Seidenberg principle, a deep result in semialgebraic geometry (Supplementary Appendix A). A semialgebraic set’s dimension quantifies the degrees of freedom needed to describe the set locally. More precisely, a semialgebraic set decomposes into manifolds. The greatest dimension of any such manifold equals the semialgebraic set’s dimension. The dimension of  $\mathcal{U}(A)$  is the architecture  $A$ ’s accessible dimension. More restricted than a semialgebraic set is an algebraic set, which consists of the solutions to a finite set of polynomial equations.

Just as the contraction map’s image will prove useful, so will the map’s rank, defined as follows. Let  $x = (U_1, U_2, \dots, U_R) \in \text{SU}(4)^{\times R}$  denote an input into  $F^A$ , such that the  $U_j$  denote two-qubit gates. The map’s rank at  $x$  is the rank of a matrix that approximates  $F^A$  linearly around  $x$  (the rank of the map’s Jacobian at  $x$ ). The rank is low at  $x$  if perturbing  $x$  can influence the  $n$ -qubit unitary only along few directions in  $\text{SU}(2^n)$ .

Crucially, we prove that  $F^A$  has the same rank throughout the domain, except on a measure-zero set, where  $F^A$  has a lesser rank. The greater, ‘dominating’ rank is the dimension of  $\mathcal{U}(A)$ . To formalize this result, let  $E_r$  denote the locus of points at which  $F^A$  has a rank of  $r \geq 0$ . Let  $E_{<r} = \bigcup_{r' < r} E_{r'}$  denote the set of points where  $F^A$  has a lesser rank. Let  $r_{\max}$  denote the maximum rank achieved by  $F^A$  at any point  $x$ . We prove the following lemma in Supplementary Appendix B, using the dimension theory of real algebraic sets.

**Lemma 1.** (Low-rank locus) *The low-rank locus  $E_{<r_{\max}}$  is an algebraic set of measure 0 and so is closed (in the Lie-group topology). Equivalently,  $E_{r_{\max}}$  is an open set of measure 1. Consequently,  $d_A = r_{\max}$ .*

Lemma 1 guarantees that the contraction map’s rank equals the accessible dimension  $d_A$  almost everywhere in  $\mathcal{U}(A)$ .

We now turn to the proof of Proposition 1. The rank  $r$  of  $F^A$  at each point  $x$  lower-bounds  $r_{\max}$ , by definition. Consider an architecture  $A_T$  of  $T$  blocks, each containing a backwards light cone. We identify an  $x$  at which  $r$  is lower-bounded by a quantity that grows linearly with  $R$  (the number of gates in the architecture  $A_T$ ). We demonstrate the point’s existence by constructing circuits from Clifford gates.

Consider a choice  $x = (U_1, U_2, \dots, U_R) =: (U_j)_j$  of unitary gates. Perturbing a  $U_j$  amounts to appending an infinitesimal unitary:  $U_j \mapsto \tilde{U}_j = e^{i\epsilon H} U_j$ . The  $H$  denotes a two-qubit Hermitian operator and  $\epsilon \in \mathbb{R}$ .  $H$  can be written as a linear combination of two-qubit Pauli strings  $S_k$ . (An  $n$ -qubit Pauli string is a tensor product of  $n$  single-site operators, each of which is a Pauli operator  $[X, Y$  or  $Z]$  or the identity,  $\mathbb{1}_1$ . The  $4^n$   $n$ -qubit Pauli strings form a basis for the space of  $n$ -qubit Hermitian operators.) Consider perturbing each gate  $U_j$  using a combination of all 15 nontrivial two-qubit Pauli strings (Supplementary Fig. 4a):  $x = (U_j)_j \mapsto \tilde{x} = (\exp(i \sum_{k=1}^{15} \epsilon_{j,k} S_k) U_j)_j$ , wherein  $\epsilon_{j,k} \in \mathbb{R}$ . The perturbation  $x \mapsto \tilde{x}$  causes a perturbation  $U = F^{A_T}(x) \mapsto \tilde{U} = F^{A_T}(\tilde{x})$  of the image under  $F^{A_T}$ . The latter perturbation is, to first order,  $\partial_{\epsilon_{j,k}} \tilde{U}|_{\epsilon_{j,k}=0}$ . This derivative can be expressed as the original circuit with the Pauli string  $S_k$  inserted immediately after the gate  $U_j$  (Supplementary Fig. 4b).

The rank of  $F^{A_T}$  at  $x$  is the number of parameters  $\epsilon_{j,k}$  needed to parameterize a general perturbation of  $U = F^{A_T}(x)$  within the image set  $\mathcal{U}(A_T)$ . To lower-bound the rank of  $F^{A_T}$  at a point  $x$ , we need only show that  $\geq r$  parameters  $\epsilon_{j,k}$  perturb  $F^{A_T}(x)$  in independent directions. To do so, we express the derivative as

$$\partial_{\epsilon_{j,k}} F^{A_T}(\tilde{x})|_{\epsilon_{j,k}=0} = K_{j,k} F^{A_T}(x), \tag{4}$$

where  $K_{j,k}$  denotes a Hermitian operator (Supplementary Fig. 4c).  $K_{j,k}$  results from conjugating  $S_k$ , the Pauli string inserted into the circuit after gate  $U_j$ , with the later gates. The physical significance of  $K_{j,k}$  follows from perturbing the gate  $U_j$  in the direction  $S_k$  by an infinitesimal amount  $\epsilon_{j,k}$ . The image  $F^{A_T}(x)$  is consequently perturbed, in  $\text{SU}(2^n)$ , in the direction  $K_{j,k}$ .

We choose for the gates  $U_j$  to be Clifford operators. The Clifford operators are the operators that map the Pauli strings to the Pauli strings, to within a phase, via conjugation. For every Clifford operator  $C$  and Pauli operator  $P$ ,  $CPC^\dagger$  equals a phase times a Pauli string by definition of the Clifford group. As a result, the operators  $K_{j,k}$  are Pauli strings (up to a phase). Two Pauli strings are linearly independent if and only if they differ. For Clifford circuits, therefore, we can easily verify whether perturbations of  $x$  cause independent perturbation directions in  $\text{SU}(2^n)$ : we need only show that the resulting operators  $K_{j,k}$  are distinct.

We apply that fact to prove Proposition 1, using the following observation. Consider any Pauli string  $P$  and any backwards-light-cone-containing block of any architecture. We can insert Clifford gates into the block such that two operations are equivalent: (1) operating on the input qubits with  $P$  before the extended block and (2) operating with the extended block, then with a one-qubit  $Z$ . Supplementary Fig. 4d depicts the equivalence, which follows from the structure of backwards light cones. We can iteratively construct a Clifford unitary that reduces the Pauli string’s weight until producing a single-qubit operator. See Supplementary Appendix B for details.

We now prove Proposition 1 by recursion. Consider an  $R'$ -gate architecture  $A_{T'}$  formed from  $T' < 4n - 1$  blocks, each containing a backwards light cone and each of  $\leq L$  gates. Assume that there exists a list  $x'$  of Clifford gates, which can be slotted into  $A_{T'}$ , such that  $F^{A_{T'}}(x')$  has a rank  $\geq T'$  at  $x'$ . Consider appending a backwards-light-cone-containing block to  $A_{T'}$ . The resulting architecture corresponds to a contraction map whose rank is  $\geq T' + 1$ .

By assumption, we can perturb  $x'$  such that its image,  $F^{A_{T'}}(x')$ , is perturbed in  $\geq T'$  independent directions in  $\text{SU}(2^n)$ . These directions can be represented by Pauli operators  $K'_{j,m,k_m}$ , wherein  $m = 1, 2, \dots, T'$ , by equation (4). Let  $P$  denote any Pauli operator absent from  $\{K'_{j,m,k_m}\}$ . We can append to  $A_{T'}$  a backwards-light-cone-containing block, forming an architecture  $A_{T'+1}$  of  $T' + 1$  backwards light cones. We design the new block from Clifford gates such that two operations are equivalent: (1) applying  $P$  to the input qubits before the extended blocks and (2) applying the extended block, then a single-site  $Z$ . We denote by  $x''$  the list of gates in  $x'$  augmented with the gates in the extended block. Conjugating the  $K'_{j,m,k_m}$  with the new block yields operators  $K'_{j,m,k_m}$ , for  $m = 1, 2, \dots, T'$ . They represent the directions in which the image  $F^{A_{T'+1}}(x'')$  is perturbed by the original perturbations of  $A_{T'}$ . The  $K'_{j,m,k_m}$  are still linearly independent Pauli operators. Also, the  $K'_{j,m,k_m}$  and the single-site  $Z$  form an independent set, because  $P$  is not in  $\{K'_{j,m,k_m}\}$ . Meanwhile, the single-site  $Z$  is a direction in which the last block’s final gate can be perturbed. The operators  $K_{j,m,k_m}$  augmented with the single-site  $Z$ , therefore span  $T' + 1$  independent directions along which  $F^{A_{T'+1}}(x'')$  can be perturbed. Therefore,  $T' + 1$  lower-bounds the rank of  $F^{A_{T'+1}}$ .

We apply the above argument recursively, starting from an architecture that contains no gates. The following result emerges: consider any architecture  $A_T$  that consists of  $T$  backwards-light-cone-containing blocks. At some point  $x$ , the map  $F^{A_T}$  has a rank lower-bounded by  $T$ . Lemma 1 ensures that the same bound applies to  $d_{A_T}$ .

To conclude the proof of Theorem 1, we address an architecture  $A'$  whose accessible dimension satisfies  $d_{A'} < d_{A_T}$ . Consider sampling a random circuit with the architecture  $A_T$ . We must show that the circuit has a zero probability of implementing a unitary in  $\mathcal{U}(A')$ . To prove this claim, we invoke the constant-rank theorem: consider any map whose rank is constant locally—in any open neighbourhood of any point in the domain. In that neighbourhood, the map is equivalent to a projector, up to a diffeomorphism. We can apply the constant-rank theorem to the contraction map:  $F^{A_T}$  has a constant rank throughout  $E_{r_{\max}}$ , by Lemma 1. Therefore,  $F^{A_T}$  acts locally as a projector throughout  $E_{r_{\max}}$ —and so throughout  $\text{SU}(4)^{\times R}$ , except on a measure-0 region, by Lemma 1. Consider mapping an image back, through a projector, to a pre-image. Suppose that the image forms a subset of dimension lower than the whole range’s dimension. The backward mapping just adds degrees of freedom to the image. Therefore, the pre-image locally has a dimension less than the domain’s dimension. Hence the pre-image is of measure 0 in the domain. We use the unitary group’s compactness to elevate this local statement to the global statement in Theorem 1.

Data availability

No data or code have been generated in this work.

Acknowledgements

We thank A. Harrow and R. Küng for discussions and P. Varjú for introducing us to the algebraic geometrical methods used in this Article. N.Y.H. thanks S. Chapman, M. Walter and the other organizers of the 2020 Lorentz Center workshop ‘Complexity: from quantum information to black holes’ for inspiration. This work has been funded by the DFG (EI 519/14-1 to J.E. and J.H.; CRC 183 to J.E.), for which this is an inter-node Berlin-Cologne project, and FOR 2724 (to J.E., P.F.), by the Einstein Research Foundation, the FQXi (‘Information as fuel’) (to J.E., J.H.) and by an NSF grant for the Institute for Theoretical Atomic, Molecular and Optical Physics at Harvard University and the Smithsonian Astrophysical Observatory (to N.Y.H.). Administrative support was provided by the MIT CTP (N.Y.H.).

Author contributions

J.H. developed the basic proof technique. J.H., P.F., N.B.T.K., J.E. and N.Y.H. wrote the manuscript and contributed to the results.

Funding

Open access funding provided by Freie Universität Berlin

Competing interests

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41567-022-01539-6>.

**Correspondence and requests for materials** should be addressed to Jonas Haferkamp.

**Peer review information** *Nature Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Supplementary information**

---

**Linear growth of quantum circuit complexity**

---

In the format provided by the authors and unedited

# Supplementary Material for “Linear growth of quantum circuit complexity”

J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert and N. Yunger Halpern

## Appendix A: Algebraic and semialgebraic sets

For convenience, we review elementary aspects of algebraic geometry over the real numbers. We apply these properties in the proof of Theorem 1. Ref. [45] contains a more comprehensive treatment.

**Definition 4** (Algebraic set). *A subset  $V \subseteq \mathbb{R}^m$  is called an algebraic set, or an algebraic variety, if, for a set of polynomials  $\{f_j\}_j$ ,*

$$V = \{x \in \mathbb{R}^m \mid f_j(x) = 0\}. \quad (\text{A1})$$

*A subset  $V' \subseteq V$  is called an algebraic subset if  $V'$  is an algebraic set. We call a subset  $W \subseteq \mathbb{R}^m$  a semialgebraic set if, for sets  $\{f_j\}_j$  and  $\{g_k\}_k$  of polynomials,*

$$W = \{x \in \mathbb{R}^m \mid f_j(x) = 0, g_k(x) \leq 0\}. \quad (\text{A2})$$

A natural topology on algebraic sets is the Zariski topology.

**Definition 5** (Zariski topology). *Let  $V$  denote an algebraic set. The Zariski topology is the unique topology whose closed sets are the algebraic subsets of  $V$ .*

A traditional definition of “dimension” for algebraic sets involves irreducible sets.

**Definition 6** (Irreducible set). *Let  $X$  denote a topological space.  $X$  is called irreducible if it is not the union of two proper closed subsets.*

**Definition 7** (Dimension of algebraic sets). *Let  $V$  be an algebraic set that is irreducible with respect to the Zariski topology. The dimension of  $V$  is the maximal length  $d$  of any chain  $V_0 \subset V_1 \subset \dots \subset V_d$  of distinct nonempty irreducible algebraic subsets of  $V$ .*

The relevant algebraic sets in the proof of Theorem 1 are  $\text{SU}(4)^{\times R}$  and  $\text{SU}(2^n)$ . Our interest in semialgebraic sets stems from the following principle. In the following, we refer to a function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  as a *polynomial map* if its entries are polynomials in the entries of its input.

**Theorem 2** (Tarski-Seidenberg principle). *Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a polynomial map. If  $W$  is a semialgebraic set, so is  $F(W)$ .*

The Tarski-Seidenberg principle applies to the map that contracts sets of quantum gates. This application is important for us, because it provides a natural notion of dimension for the contraction map’s image.

All semialgebraic sets (and hence all algebraic sets) decompose into smooth manifolds.

**Theorem 3** (Stratification of semialgebraic sets). *If  $W$  is a semialgebraic set, then  $W = \bigcup_{j=1}^N M_j$ , wherein each  $M_j$  denotes a smooth manifold. If  $W$  is an algebraic set of dimension  $d$  in the sense of Definition 7, then  $\max_j \{\dim(M_j)\} = d$ .*

This  $\max_j \{\dim(M_j)\}$  does not depend on the decomposition chosen. This independence motivates the following definition:

**Definition 8** (Dimension of semialgebraic sets). *Let  $W$  denote a semialgebraic set, such that  $W = \bigcup_{j=1}^N M_j$ , wherein each  $M_j$  denotes a manifold. The greatest dimension of any manifold,  $\max_j \{\dim(M_j)\}$ , is the semialgebraic set’s dimension.*

This definition generalizes Definition 7, due to Theorem 3. One more fact about semialgebraic sets’ dimensions will prove useful:

**Lemma 2** (Dimension of an image). *Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a polynomial map. If  $W$  is a dimension- $d$  semialgebraic set,  $F(W)$  is of dimension  $\leq d$ .*

The bound follows from combining the results of Ref. [45, Prop. 2.8.7] with the results of Ref. [45, Prop. 2.8.6]. (Ref. [45] invokes a semialgebraic mapping, which encompasses polynomial maps.)

## Appendix B: Proof of the main theorem and lemmata

In this appendix, we prove Lemma 3, Lemma 4, and the main theorem. The proofs rely on the topics reviewed in Appendix A, as well as the following notation and concepts. In differential geometry, the rank of  $F^A$  at the point  $x = (U_1, U_2, \dots, U_R)$  is defined as the rank of the derivative  $D_x F^A$ . Mapping lists of gates to unitaries,  $F$  is a complicated object. We can more easily characterize a map from real numbers to real numbers. Related is a map from Hermitian operators to Hermitian operators: An  $n$ -qubit state evolves under a Hamiltonian represented by a  $2^n \times 2^n$  Hermitian operator, which has  $(2^n)^2 = 4^n$  real parameters. Therefore, for convenience, we shift focus from unitaries to their Hermitian generators. We construct a map whose domain is the algebra  $\mathfrak{su}(4)^{\times R} \simeq \mathbb{R}^{15R}$  that generates  $SU(4)^{\times R}$ . The range is the set of  $n$ -qubit Hermitian operators,  $\mathfrak{su}(2^n) \simeq \mathbb{R}^{4^n}$ . We construct such a map from three steps, depicted by the dashed lines in Fig. 1.

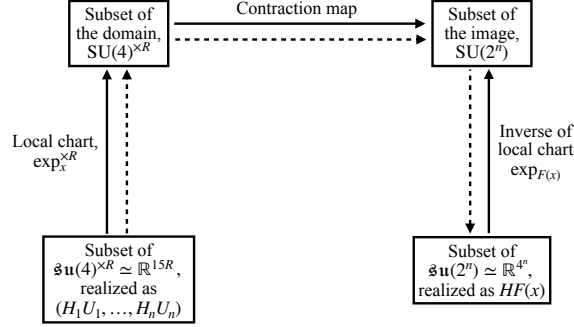


Figure 1. Three-part map used in the proof of Lemma 3.  $H_j$  denotes the  $j^{\text{th}}$  two-qubit Hermitian operator,  $U_j$  denotes the  $j^{\text{th}}$  two-qubit unitary, and  $H$  denotes an  $n$ -qubit Hermitian operator.

The first step is a chart, a diffeomorphism that maps one manifold to another invertibly. Our chart acts on the algebra  $\mathfrak{su}(4)^{\times R}$  that generates  $SU(4)^{\times R}$ . To define the chart, we parameterize an element  $H$  of the  $j^{\text{th}}$  copy of  $\mathfrak{su}(4)$ :

$$H = \sum_{\substack{\alpha, \beta \in \{\mathbb{1}, X, Y, Z\} \\ (\alpha, \beta) \neq (\mathbb{1}, \mathbb{1})}} \lambda_{j, \alpha, \beta} \alpha \otimes \beta, \quad (\text{B1})$$

wherein  $\lambda_{j, \alpha, \beta} \in \mathbb{R}$ . For each point  $x = (U_1, U_2, \dots, U_R) \in SU(4)^{\times R}$ , we define the local exponential chart  $\exp_x^{\times R} : \mathfrak{su}(4)^{\times R} \rightarrow SU(4)^{\times R}$  as  $\exp_x^{\times R}(H_1, \dots, H_R) := (e^{iH_1} U_1, \dots, e^{iH_R} U_R)$ , and we define the analogous  $\exp_U : \mathfrak{su}(2^n) \rightarrow SU(2^n)$  as  $\exp_U(H) := e^{iH} U$ . These charts are standard for matrix Lie groups. Both are locally invertible in small neighbourhoods around  $x$  and  $U$ , by a standard result in Lie-group theory [46]. The three-part map, represented by the dashed lines in Fig. 1, has the form  $\exp_{F^A(x)}^{-1} \circ F^A \circ \exp_x^{\times R}$ .

We now characterize the map's derivative, to characterize the derivative of  $F^A$ , to characterize the rank of  $F^A$ . Denote by  $D_0$  the derivative evaluated where the Hermitian operators are set to zero, such that each chart reduces to the identity operation. The image of  $D_0(\exp_{F^A(x)}^{-1} \circ F^A \circ \exp_x^{\times R})$  is spanned by the operators

$$\partial_{\lambda_{j, A, B}} \left( \exp_{F^A(x)}^{-1} \circ F^A \circ \exp_x^{\times R} \right) \Big|_0. \quad (\text{B2})$$

These operators have the form

$$U_R \dots U_{j+1} P U_j \dots U_1, \quad (\text{B3})$$

wherein  $P$  denotes a two-qubit Pauli operator. We apply the setting above to prove the following restatement of Lemma 1 in the methods section.

**Lemma 3 (Low-rank locus).** *The low-rank locus  $E_{<r_{\max}}$  is an algebraic set of measure 0 and so is closed (in the Lie-group topology). Equivalently,  $E_{r_{\max}}$  is an open set of measure 1. Consequently,  $d_A = r_{\max}$ .*

*Proof.* Consider representing an operator (B3) as a matrix relative to an arbitrary tensor-product basis. To identify the matrix's form, we imagine representing the unitaries in  $SU(4)^{\times R}$  as matrices relative to the corresponding tensor-product basis for  $\mathbb{C}^2 \otimes \mathbb{C}^2$ . Combining the unitary matrices' elements polynomially yields the matrix elements of (B3).



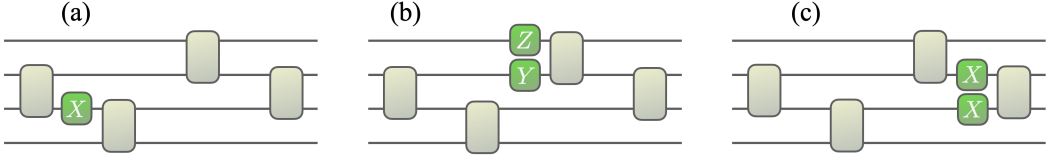


Figure 2. Examples of partial derivatives  $\partial_{\lambda_j, \alpha, \beta} (\exp_{F^A(x)}^{-1} \circ F^A \circ \exp_x^{\times 4})|_{\lambda_j, \alpha, \beta=0}$  that span the image of  $D_0(\exp_{F^A(x)} \circ F^A \circ \exp_x^{\times 4})$ .

$D_x F^A$  has the same rank as  $D_0(\exp_{F^A(x)}^{-1} \circ F^A \circ \exp_x^{\times R})$ , because  $\exp_x^{\times R}$  and  $\exp_{F^A(x)}$  are local charts [47]. Recall that  $E_{<r_{\max}}$  denotes the locus of points, in  $SU(4)^{\times R}$ , where  $F^A$  has a rank  $< r_{\max}$ . Equivalently, by the invertible-matrix theorem,  $E_{<r_{\max}}$  consists of the points where certain minors of  $D_0(\exp^{-1} \circ F^A \circ \exp_x^{\times R})$ —the determinants of certain collections of  $r_{\max} \times r_{\max}$  matrix elements—vanish. The determinants' vanishing implies a set of equations polynomial in the matrix elements of  $D_0(\exp^{-1} \circ F^A \circ \exp_x^{\times R})$ —and so, by the last paragraph, polynomial in the entries of matrices in  $SU(4)^{\times R}$ .  $SU(4)^{\times R}$  is a real algebraic set, being the set of operators that satisfy the polynomial equations equivalent to  $UU^\dagger = \mathbb{1}$  and  $\det U = 1$ . Thus, by Definition 4, the points of rank  $< r$  form an algebraic subset of  $SU(4)^{\times R}$ .

We can now invoke properties of algebraic subsets, reviewed in Appendix A. First, we prove that  $SU(4)^{\times R}$  is irreducible in the Zariski topology. The Zariski topology of  $SU(4)^{\times R}$  is coarser than the topology inherited from  $(\mathbb{C}^{4 \times 4})^{\times R}$ , identified with  $\mathbb{R}^{32R}$ . As  $SU(4)^{\times R}$  is connected in the finer topology, so is  $SU(4)^{\times R}$  connected in the Zariski topology. This connectedness implies that  $SU(4)^{\times R}$  is irreducible, as  $SU(4)^{\times R}$  is an algebraic group [48, Summary 1.36]. Being irreducible,  $SU(4)^{\times R}$  has a dimension à la Definition 7. If the low-rank locus  $E_{<r_{\max}}$  is not all of  $SU(4)^{\times R}$ , then it is, by Definition 7, a lower-dimensional algebraic subset. Every dimension- $N$  algebraic subset decomposes into a collection of submanifolds, each of which has dimension  $\leq N$  [45, Prop. 9.1.8]. As a proper submanifold has measure 0,  $E_{<r_{\max}}$  has measure 0. As an algebraic subset,  $E_{<r_{\max}}$  is closed in the Lie-group topology.

Finally, we prove that  $d_A = r_{\max}$ . In a small open neighborhood  $V$  of a point  $x \in E_{r_{\max}}$ , the contraction map's rank is constant, by Lemma 3. By the constant-rank theorem [47, Thm 5.13], therefore,  $F^{AT}$  acts locally as a projector throughout  $E_{r_{\max}}$ —and so throughout  $SU(4)^{\times R}$  (except on a region of measure 0, by Lemma 3). The projector has a rank, like  $F^{AT}$ , of  $r_{\max}$ . A rank- $r_{\max}$  projector has an image that is a dimension- $r_{\max}$  manifold. Hence  $r_{\max} \leq d_A$ . The other direction,  $d_A \leq r_{\max}$ , follows directly from Sard's theorem [49]. Let  $X_r$  denote the set of points where  $F^A$  is rank- $r$ . As  $F^A$  is a smooth map, Sard's theorem ensures that  $r$  upper-bounds the Hausdorff dimension of the image  $F^A(X_r)$ . As  $F^A(SU(4))$  is a semialgebraic set, it stratifies into manifolds, by Theorem 3. Therefore, the Hausdorff dimension coincides with the semialgebraic set's dimension.  $\square$

Lemma 3, combined with the following lemma, implies Proposition 1.

**Lemma 4** (Existence of a high-rank point). *Let  $T \in \mathbb{Z}_{>0}$  denote any nonnegative integer. Consider any architecture  $A_T$  formed from  $T$   $L$ -gate, backwards-light-cone-containing blocks. The map  $F^{AT}$  has the greatest rank possible,  $r_{\max} \geq T$ .*

*Proof.* Without loss of generality, we assume that all  $T$  blocks have identical architectures. This assumption will simplify the notation below. We can lift the assumption by complicating the notation.

Consider an arbitrary point  $x = (U_1, U_2, \dots, U_R) \in SU(4)^{\times R}$ . For all  $x$ , the contraction map  $F^{AT}$  has a derivative characterized, in the proof of Lemma 3, with local charts  $\exp_{F^{AT}(x)}$  and  $\exp_x^{\times R}$ . The number of gates in  $A_T$  is  $R \leq TL$ . The map  $F^{AT}$  has an image spanned by the partial derivatives  $\partial_{\lambda_j, \alpha, \beta} (\exp_{F^{AT}(x)}^{-1} \circ F^{AT} \circ \exp_x^{\times R})|_{\lambda_j, \alpha, \beta=0}$ . Each partial derivative has the form

$$U_R U_{R-1} \dots U_{j+1} (\alpha \otimes \beta) U_j U_{j-2} \dots U_1 \quad (\text{B4})$$

(Fig. 2).  $\alpha$  and  $\beta$  denote Pauli operators; each acts nontrivially on just one of the two qubits on which  $U_j$  acts nontrivially. We implicitly pad operators with identities wherever necessary, such that the operators act on the appropriate Hilbert space.

We aim to lower-bound the greatest possible rank,  $r_{\max}$ , of the map  $F^{AT}$ . To do so, we construct a point

$$x_T = \left( \underbrace{C_1^{(1)}, \dots, C_1^{(L)}}_{L \text{ gates}}, \dots, \underbrace{C_T^{(1)}, \dots, C_T^{(L)}}_{L \text{ gates}} \right) \in SU(4)^{\times R}. \quad (\text{B5})$$

We will choose for the  $C_j^{(i)}$ 's to be Clifford gates. A gate's subscript,  $j$ , labels the blocks to which the gate belongs. The superscript,  $i$ , labels the gate's position within the block. The gates constitute a block as  $C_j^{(L)} C_j^{(L-1)} \dots C_j^{(1)} =: C_j$ . Our

construction of  $C_j$  relies on a property of an arbitrary Pauli operator  $Q_j$ : We can choose the Clifford gates  $C_j^{(i)}$  such that block  $C_j$  maps  $Q_j$  to a  $Z$  on qubit  $t$ :  $C_j Q_j C_j^\dagger = Z_t \equiv \mathbb{1}^{\otimes(t-1)} \otimes Z \otimes \mathbb{1}^{\otimes(n-t)}$ . We now show how the existence of such a Clifford unitary  $C_j$  implies Lemma 4. Afterward, we show to construct  $C_j$ .

Let us choose the Pauli strings  $Q_j$  that guide our construction of the Clifford block  $C_j$ . We choose the  $Q_j$ 's inductively over  $T$  such that  $\{(C_T C_{T-1} \dots C_j) Q_j (C_{j-1} C_{j-2} \dots C_1)\}_{1 \leq j \leq T}$  is linearly independent. We start with an arbitrary Pauli string  $Q_1$ . The form of  $Q_1$  guides our construction of  $C_1$ . Second, we choose for  $Q_2$  to be an arbitrary Pauli string  $\neq C_1 Q_1 C_1^\dagger$ .  $Q_2$  guides our construction of  $C_2$ . Third, we choose for  $Q_3$  to be an arbitrary Pauli string outside  $\text{span}\{C_1 C_2 Q_1 C_2^\dagger C_1^\dagger, C_2 Q_2 C_2^\dagger\}$ . This  $Q_3$  guides our construction of  $C_3$ . After  $T$  steps, we have constructed all the  $Q_j$ 's and  $C_j$ 's. If  $T < 4^n - 1$ , enough Pauli strings exist that, at each step, a Pauli string lies outside the relevant span.

The operators  $(C_T C_{T-1} \dots C_j) Q_j (C_{j-1} C_{j-2} \dots C_1)$ , for  $j \in [1, T]$ , are in the image of  $D_0(\exp_{F^{AT}}^{-1} \circ F^{AT} \circ \exp_{x_T}^{\times(R)})$ :

$$\begin{aligned} \partial_{\lambda_{jL, z_1, z}} \left( \exp_{F^{AT}}^{-1} \circ F^{AT} \circ \exp_{x_T}^{\times(R)} \right) \Big|_0 &= (C_T C_{T-1} \dots C_{j+1}) (\mathbb{1}_{t-1} \otimes Z_t \otimes \mathbb{1}_{n-t}) (C_j C_{j-1} \dots C_1) \\ &= (C_T C_{T-1} \dots C_j) Q_j (C_{j-1} C_{j-2} \dots C_1). \end{aligned} \quad (\text{B6})$$

We have assumed, without loss of generality, that each block's final gate acts on qubit  $t$ . For all  $j \in [1, T]$ , the operators  $(C_T C_{T-1} \dots C_j) Q_j (C_{j-1} C_{j-2} \dots C_1)$  are in the image of  $D_0(\exp_{F^{AT}}^{-1} \circ F^{AT} \circ \exp_{x_T}^{\times(R)})$  and are linearly independent. Therefore, the rank of  $F^{AT}$  at the point  $x_T$  is  $\geq T$ .

In the remainder of this proof, we provide the missing link: We show that, for every Pauli string  $P$ , we can construct a backwards-light-cone-containing block that implements a Clifford unitary  $C = C^{(L)} C^{(L-1)} \dots C^{(1)}$  such that  $CPC^\dagger = Z_t$ . We drop subscripts because subscripts index blocks and this prescription underlies all blocks. By definition, each block contains a qubit  $t$  to which each other qubit  $t'$  connects via gates in the block. The path from a given qubit  $t'$  depends on  $t'$ , and multiple paths may connect a  $t'$  to  $t$ . Also, one path may connect  $t$  to multiple qubits. We choose an arbitrary complete set of paths (which connect all the other qubits to  $t$ ) that satisfies the merging property described below. To introduce the merging property, we denote by  $m$  the number of paths in the set. Let  $p \in [1, m]$  index the paths. Path  $p$  contacts the qubits in the order  $i_{p,1} \mapsto i_{p,2} \mapsto \dots \mapsto i_{p,l_p} = t$ , reaching  $l_p \in [1, L+1]$  qubits. We choose the paths such that they merge whenever they cross: If  $i_{p,j} = i_{p',j'}$ , then  $i_{p,j+k} = i_{p',j'+k}$  for all  $k \in \{1, 2, \dots, l_p - j = l_{p'} - j'\}$ . We choose for all the gates outside these paths to be identities. Next, we choose the nontrivial gates in terms of an arbitrary Pauli string.

Let  $P = \bigotimes_{j=1}^n P_j$  denote an arbitrary nontrivial  $n$ -qubit Pauli string. Some Clifford unitary  $C$  maps  $P$  to a Pauli string that acts nontrivially on just one qubit, which we choose to be  $t$ . Indeed, consider conjugating an arbitrary  $n$ -qubit Pauli operator  $P$  with a uniformly random Clifford operator  $C$ . The result,  $C^\dagger P C$ , is a uniformly random  $n$ -qubit Pauli operator [50]. Therefore, for every initial Pauli operator  $P$  and every final Pauli operator, some Clifford operator  $C$  maps one to the other.

We arbitrarily choose for the string's nontrivial single-qubit Pauli operator to be  $Z$ . Let  $i_{p,k_p}$  denote the first index  $j$  in  $i_{p,1} \mapsto i_{p,2} \mapsto \dots \mapsto i_{p,l_p}$  for which  $P_j \neq \mathbb{1}$ . By definition,  $P_{i_{p,k}} \otimes P_{i_{p,k+1}}$  is a nontrivial Pauli string. There exists a two-local Clifford gate  $C^{p,(0)}$  that transforms  $P_{i_{p,k}} \otimes P_{i_{p,k+1}}$  into a  $Z$  acting on qubit  $i_{p,k+1}$ :

$$C^{p,(0)} (P_{i_{p,k}} \otimes P_{i_{p,k+1}}) (C^{p,(0)})^\dagger = \mathbb{1}_{i_k} \otimes Z_{i_{k+1}}. \quad (\text{B7})$$

Operating with  $C^{p,(0)}$  (padded with  $\mathbb{1}$ 's) on the whole string  $P$  yields another Pauli string:

$$C^{p,(0)} P (C^{p,(0)})^\dagger = \bigotimes_{j=1}^n P_j^{p,(1)}. \quad (\text{B8})$$

Let  $i_{p,\ell}$  denote the first index  $j$  for which  $P_j^{p,(1)}$  is a nontrivial Pauli operator. Since

$$P_{i_{p,k+1}}^{(1)} \otimes P_{i_{p,k+2}}^{(1)} = Z_{i_{p,k+1}} \otimes P_{i_{p,k+2}}, \quad (\text{B9})$$

$i_{p,\ell} = i_{p,k+1}$ . There exists a two-local Clifford gate  $C^{p,(1)}$  that shifts the  $Z$  down the path:

$$C^{p,(1)} \left( P_{i_{p,k+1}}^{p,(1)} \otimes P_{i_{p,k+2}}^{p,(1)} \right) (C^{p,(1)})^\dagger = \mathbb{1}_1 \otimes Z_{i_{p,k+2}}. \quad (\text{B10})$$

We perform this process—of shifting the  $Z$  down the path and leaving an  $\mathbb{1}_1$  behind—for every path simultaneously. For example, if we begin with two equal-length paths,  $C^{(2)} = C^{p,(2)} C^{p',(2)}$ . This simultaneity is achievable until two paths merge. Whenever paths merge, we choose the next Clifford gate such that we proceed along the merged path. Every qubit is visited, and every path ends at qubit  $t$ . Therefore, we have constructed a circuit that implements a Clifford operation  $C$  such that  $CPC^\dagger = Z_t$ . Figure 4(d) depicts an example of this construction.  $\square$

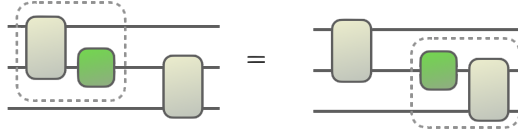


Figure 3. Along every contraction is a redundant copy of the gauge group  $SU(2)$ .

The foregoing proof has a surprising implication: A map's rank is somewhat divorced from a circuit's complexity. The rank of  $F^{A_T}$  at  $x_T$  is at least  $T$ , which could be a large number. Yet, the contracted unitary corresponding to this circuit is Clifford. Hence the extended circuit's complexity surpassed the original circuit's complexity only a little—by, at most,  $O(n^2/\log(n))$  [51].

Finally, we combine Lemmata 3 and 4 to prove Theorem 1:

**Theorem 1** (Linear growth of complexity). *Let  $U$  denote a unitary implemented by a random quantum circuit in an architecture formed by concatenating  $T$  blocks of  $\leq L$  gates each, each block containing a backwards light cone. The unitary's circuit complexity is lower-bounded as*

$$\mathcal{C}_u(U) \geq \frac{R}{9L} - \frac{n}{3}, \quad (1)$$

with unit probability, until the number of gates grows to  $T \geq 4^n - 1$ . The same bound holds for  $\mathcal{C}_{\text{state}}(U|0^n)$ , until  $T \geq 2^{n+1} - 1$ .

*Proof of Theorem 1.* We reuse the notation introduced in Lemmata 3 and 4. Examples include  $A_T$ , an arbitrary architecture that satisfies the assumptions in Lemma 4 and that consists of  $R \leq TL$  gates.  $F^{A_T}$  denotes the corresponding contraction map.  $E_{r_{\max}}$  denotes the locus of points at which  $F^{A_T}$  achieves its greatest rank,  $r_{\max}$ . In a small open neighborhood  $V$  of a point  $x \in E_{r_{\max}}$ , the contraction map's rank is constant, by Lemma 3. By the constant-rank theorem [47, Thm 5.13], therefore,  $F^{A_T}$  acts locally as a projector throughout  $E_{r_{\max}}$ —and so throughout  $SU(4)^{\times R}$  (except on a region of measure 0, by Lemma 3). The projector has a rank, like  $F^{A_T}$ , of  $r_{\max}$ . Therefore, in the open set  $V \subseteq SU(4)^{\times R}$ ,  $F^{A_T}$  is equivalent, up to a diffeomorphism, to the projection

$$(x_1, \dots, x_{\dim(SU(4)^{\times R})}) \mapsto (x_1, \dots, x_{r_{\max}}, \underbrace{0, \dots, 0}_{\dim(SU(2^n)) - r_{\max}}). \quad (\text{B11})$$

For simplicity of notation, we identify  $V$  with its image under the local diffeomorphism (we do not distinguish  $V$  from its image notationally).

The open subset  $V$  contains, itself, an open subset that decomposes as a product:  $V_1 \times V_2 \subseteq V$ , such that  $x \in V_1 \times V_2$  and, as suggested by Eq. (B11),

$$V_1 \subseteq \mathbb{R}^{r_{\max}}, \quad \text{and} \quad V_2 \subseteq \mathbb{R}^{\dim(SU(4)^{\times R}) - r_{\max}}. \quad (\text{B12})$$

(Again to simplify notation, we are equating the local sets  $V_{j=1,2}$  with their images, under local charts, in  $\mathbb{R}^m$ , for  $m \in \mathbb{Z}_{>0}$ .) From now on,  $V_1 \times V_2$  is the open subset of interest. The contraction map's equivalence to a projector, in  $V_1 \times V_2$ , will help us compare high-depth circuits with low-depth circuits: Consider a circuit whose contraction map takes some local neighborhood to an image of some dimension. How does the dimension differ between high-depth circuits and low-depth circuits? We start by upper-bounding the dimension for low-depth circuits.

We have been discussing an  $R$ -gate architecture  $A_T$ . Consider any smaller architecture  $A'$  of  $R' < R$  gates.  $A'$  is encoded in a contraction map  $F^{A'}$  whose domain is  $SU(4)^{\times R'}$ . As explained in the proof of Lemma 3,  $F^{A'}$  is a polynomial map. Therefore,  $F^{A'}$  has a property prescribed by the Tarski-Seidenberg principle [45] (Theorem 2): The image  $F^{A'}(SU(4)^{\times R'})$  is a semialgebraic set of dimension  $\leq \dim(SU(4)^{\times R'}) = R' \dim(SU(4)) = 15R'$ .

We can strengthen this bound: Consider contracting two gates that share a qubit. The shared qubit may undergo a one-qubit gate specified by three parameters (one parameter per one-qubit Pauli). The one-qubit gate can serve as part of the first two-qubit gate or as part of the second two-qubit gate; which does not affect the contraction. Hence the contraction contains 3 fewer parameters than expected. In other words, the contraction has a redundant copy of the gauge group  $SU(2)$ : Every unitary  $U \in SU(4)$  decomposes as  $(U_1 \otimes U_2)K(V_1 \otimes V_2)$ , wherein  $K = e^{i(aZ \otimes Z + bY \otimes Y + cX \otimes X)}$  and the  $U_j$  and the  $V_j$  denote single-qubit unitaries [52]. Let us classify the shared qubit as an input of the second two-qubit gate. A two-qubit gate in a circuit's bulk accepts two input qubits outputted by earlier gates. So we might expect an  $R'$ -gate circuit to have  $\dim(F^{A'}(SU(4)^{\times R'})) \leq 15R' - 2 \times 3R' = 9R'$ . But the first  $n/2$  gates [the leftmost vertical line of gates in Fig. 2(a)] receive

their input qubits from no earlier gates. So we must restore  $3 \times 2$  parameters for each of the  $n/2$  initial gates, or restore  $3n$  parameters total:

$$\dim(F^{A'}(\text{SU}(4)^{\times R'})) \leq 9R' + 3n. \quad (\text{B13})$$

We have upper-bounded the dimension for low-depth circuits.

Technically, this bound on the dimension does not follow from Lemma 2 as it does for the bound  $\dim F^{A'}(\text{SU}(4)^{\times R'}) \leq 15R'$ . The reason is, the quotient space  $\text{SU}(4)^{\times R'} / \text{SU}(2)^{\times (2R'-n)}$  is not necessarily semialgebraic. This difficulty can be resolved via Sard's theorem [49], which asserts, as a special case, that the Hausdorff dimension of a smooth map's image is bounded by its domain's dimension. A semialgebraic set's dimension is the greatest dimension in its stratification and so agrees with the Hausdorff dimension.

We now lower-bound the corresponding dimension for high-depth circuits. We can do so by lower-bounding the greatest possible rank,  $r_{\max}$ , of a high-depth architecture's contraction map,  $F^{A_T}$ : In an open neighborhood of  $x \in \text{SU}(4)^{\times R}$ ,  $F^{A_T}$  is equivalent to a projector, which has some rank. The neighborhood's image, under the projector, is a manifold. The manifold's dimension equals the projector's rank. Therefore, we bound the rank to bound the dimension.

Augmenting an architecture with  $T$  ( $\leq L$ )-gate blocks increases the contraction map's greatest possible rank,  $r_{\max}$ , by  $\geq T - 1$ . Therefore, for an architecture- $A_T$  circuit of  $R \leq TL$  gates, we have constructed a point of rank  $T \geq R/L$ . Therefore,

$$r_{\max} \geq R/L. \quad (\text{B14})$$

We have lower-bounded the dimension of the image of a high-depth architecture's contraction map [the rank in Ineq. (B14)] and have upper-bounded the analogous dimension for a low-depth architecture [Ineq. (B13)]. The high-depth-architecture dimension upper-bounds the low-depth-architecture dimension,

$$\dim(F^{A'}(\text{SU}(4)^{\times R'})) < r_{\max}, \quad (\text{B15})$$

if

$$9R' + 3n < r_{\max}, \quad (\text{B16})$$

by Ineq. (B13). Furthermore, by Ineqs. (B13) and (B14), Ineq. (B15) holds if  $9R' + 3n < R/L$ , or  $R' < \frac{R}{9L} - \frac{n}{3}$ .

$$R' < \frac{R}{9L} - \frac{n}{3}. \quad (\text{B17})$$

holds. We have upper-bounded the short circuit's gate count in terms of the deep circuit's gate count.

Let us show that, if Ineq. (B17) holds, the points in  $\text{SU}(4)^{\times R}$  resulting in unitaries that can be prepared with short circuits form a set of measure 0. We will begin with a point  $x \in E_{r_{\max}}$ ; apply the short-architecture contraction map  $F^{A'}$ ; and follow with the deep-architecture contraction map's inverse,  $(F^{A_T})^{-1}$ . The result takes up little space in  $\text{SU}(4)^{\times R}$ , we will see.

To make this argument rigorous, we recall the small open neighborhood  $V_1 \times V_2$  of  $x \in E_{r_{\max}}$ . In  $V_1 \times V_2$ ,  $F^{A'}(\text{SU}(4)^{\times R'})$  has the preimage, under  $F^{A_T}$ , of

$$(F^{A_T}|_{V_1 \times V_2})^{-1} \left( F^{A'}(\text{SU}(4)^{\times R'}) \right) \simeq \left[ F^{A'}(\text{SU}(4)^{\times R'}) \cap V_1 \right] \times V_2. \quad (\text{B18})$$

The  $\simeq$  represents our identification of the map  $F^A$  with its representation in local charts. By the proof of Lemma 3,  $F^{A'}(\text{SU}(4)^{\times R'})$  is a semialgebraic set. Therefore, by Theorem 3,  $F^{A'}(\text{SU}(4)^{\times R'})$  is a union of smooth manifolds. Each manifold is of dimension  $\leq 9R' + 3n$ , by Theorem 3 and Ineq. (B13). By Eq. (B12),  $V_2$  is of dimension  $\dim(\text{SU}(4)^{\times R}) - r_{\max}$ . Therefore,  $[F^{A'}(\text{SU}(4)^{\times R'}) \cap V_1] \times V_2$  consists of manifolds of dimension  $\leq 9R' + 3n + \dim(\text{SU}(4)^{\times R}) - r_{\max}$ . Using Ineq. (B16), we can cancel the  $9R' + 3n$  with the  $-r_{\max}$ , at the cost of loosening the bound:  $[F^{A'}(\text{SU}(4)^{\times R'}) \cap V_1] \times V_2$  consists of manifolds of dimension  $< \dim(\text{SU}(4)^{\times R})$ . As a collection of manifolds of submaximal dimension, the points in  $\text{SU}(4)^{\times R}$  that contract to unitaries implementable by short circuits (satisfying (B17)), restricted to a small open neighborhood  $V_1$ , form a set of measure 0.

Let us extend this conclusion about  $n$ -qubit unitaries—about images of maps  $F^{A'}$ —to a conclusion about preimages—about lists of gates. By Lemma 3,  $E_{r_{\max}}$  is of measure 1. Therefore, for every  $\varepsilon > 0$ , there exists a compact subset  $K \subseteq E_{r_{\max}}$  of measure  $1 - \varepsilon$ . Since  $K$  is compact, for any cover of  $K$  by open subsets, a finite subcover exists. The foregoing paragraph shows that, restricted to each open set in this finite subcover, the preimage of the unitaries reached by lower-depth circuits is of measure 0. Therefore, the preimage of the  $R'$ -gate, architecture- $A'$  circuits is of measure  $\leq \varepsilon$ . Since  $\varepsilon > 0$  is arbitrary, the preimage is of measure 0. The foregoing argument holds for each architecture  $A'$  of  $R'$  gates. Hence each preimage forms a set of measure 0. The total measure is subadditive. So the union of the preimages, over all architectures with  $\leq R'$  gates, is of measure 0. We have proven the circuit-complexity claim posited in Theorem 1. The state-complexity claim follows from tweaks to the proof (Appendix C).  $\square$

### Appendix C: Proof of the linear growth of state complexity

At the end of Appendix B, we proven part of Theorem 1—that circuit complexity grows linearly with the number of gates. Here, we prove rest of the theorem—that state complexity grows linearly. We need only tweak the proof presented in Appendix B.

Consider instead of the contraction map  $F^{A_T}$ , the map that contracts a list of gates, forming an architecture- $A_T$  circuit, and applies the circuit to  $|0^n\rangle$ , to get

$$G^{A_T} : \text{SU}(4)^{\times R} \rightarrow \mathcal{S}^{2 \times 2^n - 1} \subseteq \mathbb{C}^{2^n}. \quad (\text{C1})$$

The argument works the same as in Appendix B, with one exception: The derivative  $D_x G^{A_T}$  has an image that does not contain  $4^n - 1$  nontrivial linearly independent Pauli operators. Rather, the image contains the computational basis  $\{|i^\kappa x\rangle\}_{x \in \{0,1\}^n, \kappa \in \{0,1\}}$  formed by applying tensor products of  $Z, X$  and  $Y$  to  $|0^n\rangle$ . (We denote the imaginary number  $\sqrt{-1}$  by  $i$ .) The proof of Lemma 3 ports over without modification, as  $G^{A_T}$  is a polynomial map between algebraic sets.

The proof of Lemma 4 changes slightly. We must prove the existence of a point  $x \in \text{SU}(4)^{\times R}$  at which  $G^{A_T}$  has a rank at least linear in the circuit depth. The only difference in the proof is, we must choose the operators  $Q_j$  inductively such that the states  $(C_T C_{T-1} \dots C_j) Q_j (C_{j-1} C_{j-2} \dots C_1) |0^n\rangle$  are linearly independent. Such a choice is possible if  $T < 2 \times 2^n - 1$ , the number of real parameters in a pure  $n$ -qubit state vector.

### Appendix D: Randomized architectures

From Theorem 1 follows a bound on the complexity of a doubly random circuit: Not only the gates, but also the gates' positions, are drawn randomly. This model features in Ref. [22]. Our proof focuses on nearest-neighbor gates, but other models (such as all-to-all interactions) yield similar results.

**Corollary 1** (Randomized architectures). *Consider drawing an  $n$ -qubit unitary  $U$  according to the following probability distribution: Choose a qubit  $j$  uniformly randomly. Apply a Haar-random two-qubit gate to qubits  $j$  and  $j + 1$ . Perform this process  $R$  times. With high probability, the unitary implemented has a high complexity: For all  $\alpha \in [0, 1)$ ,*

$$\Pr \left( \mathcal{C}_u(U) \geq \alpha \frac{R}{9n(n-1)^2} - \frac{n}{3} \right) \geq 1 - \frac{1}{1-\alpha} (n-1)e^{-n}. \quad (\text{D1})$$

*Proof.* The proof relies on the following strategy: We consider constructing blocks randomly to form a circuit. If the blocks contain enough gates, we show, many of the blocks contain backwards light cones. This result enables us to apply Theorem 1 to bound the circuit's complexity.

Consider drawing  $L$  gates' positions uniformly randomly. For each gate, the probability of drawing position  $(j, j + 1)$  is  $1/(n - 1)$ . The probability that no gates act at position  $(j, j + 1)$  is  $(1 - 1/(n - 1))^L$ . Let us choose for each block to contain  $L = n(n - 1)^2$  gates. Define a binary random variable  $I_j$  as follows: If one of the gates drawn during steps  $(j - 1)n(n - 1), (j - 1)n(n - 1) + 1, \dots, jn(n - 1)$  acts at  $(j, j + 1)$ , then  $I_j = 1$ . Otherwise,  $I_j = 0$ . With high probability, gates act at all positions:

$$p := \Pr \left( \bigwedge_{j=1}^{n-1} (I_j = 1) \right) = \left( 1 - \left( 1 - \frac{1}{n-1} \right)^{n(n-1)^2} \right)^{n-1} \geq (1 - e^{-n})^{n-1} \geq 1 - (n-1)e^{-n}. \quad (\text{D2})$$

We have invoked the inverse Bernoulli inequality and the Bernoulli inequality. We will use this inequality to characterize blocks that contain backwards light cones.

Consider drawing  $T$   $L$ -gate blocks randomly, as described in the corollary. Denote by  $X$  the number of blocks in which at least one position is bereft of gates: For some  $j$ ,  $I_j = 0$ . With high probability,  $X$  is small: For all  $a \in (0, T]$ ,

$$\Pr(X \geq a) \leq \frac{T(1-p)}{a} \leq T(n-1)e^{-L/n}/a, \quad (\text{D3})$$

by Markov's inequality. Let us choose for the threshold to be  $a = (1 - \alpha)T$ . With overwhelming probability,  $\alpha T$  blocks satisfy  $\bigwedge_j (I_j = 1)$  and so contain gates that act at all positions  $(j, j + 1)$  in increasing order. Therefore, these blocks contain a staircase architecture and so contain backwards light cones. Therefore, a slight variation on Theorem 1 governs the  $\alpha T \times L = \alpha R$  gates that form the blocks. Strictly speaking, Theorem 1 governs only consecutive backwards-light-cone-containing blocks. In contrast, extra gates may separate the blocks here. However, the extra gates can only increase the contraction map's image.

Therefore, the additional  $(1 - \alpha)TL$  gates cannot decrease the accessible dimension  $d_{A_T}$ . Therefore, the bound from Theorem 1 holds. With probability  $\geq 1 - \frac{1}{1-\alpha}(n-1)e^{-n}$  over the choice of architecture,

$$\mathcal{C}_u(U) \geq \frac{R - (1 - \alpha)R}{9n^2(n-1)} - \frac{n}{3}, \quad (\text{D4})$$

with probability one over the choice of gates. This bound is equivalent to Ineq. (D1).  $\square$

### Appendix E: Linear growth of slightly robust circuit complexity

Corollary 2 extends Theorem 1 to accommodate errors in the target unitary's implementation. We prove Corollary 2 by drawing on the proof of Theorem 1 and reusing notation therein.

**Corollary 2** (Slightly robust circuit complexity). *Let  $U$  denote the  $n$ -qubit unitary implemented by any random quantum circuit in any architecture  $A_T$  that satisfies the assumptions in Theorem 1. Let  $U'$  denote the  $n$ -qubit unitary implemented by any circuit of  $R' \leq R/(9L) - n/3$  gates. For every  $\delta \in (0, 1]$ , there exists an  $\varepsilon := \varepsilon(A_T, \delta) > 0$  such that the Frobenius distance  $d_F(U, U') \geq \varepsilon$ , with probability  $1 - \delta$ , unless  $R/L > 4^n - 1$ .*

*Proof of Corollary 2.* The proof of Theorem 1 can be modified to show that, for every  $\delta > 0$ , there exists an open set  $B \subseteq \text{SU}(2^n)$  that contains  $F^{A'}(\text{SU}(4)^{\times R'})$ , such that the preimage  $(F^{A_T})^{-1}(B)$  is small—of measure  $\leq \delta$ . The modification is as follows. For every  $\delta' > 0$ , there exists a measure- $(1 - \delta')$  compact subset  $K$  of  $E_{r_{\max}}$ . As  $K$  is compact, there exists a finite cover of  $K$  that has the following properties:  $K$  is in the union  $\cup_j V^j$  of subsets  $V^j$ . On the  $V^j$ , the contraction map  $F^{A_T}$  is equivalent to a projector, up to a local diffeomorphism. As in the proof of Theorem 1, we can assume, without loss of generality, that  $V^j = V_1^j \times V_2^j$ . The  $V_1^j$  and  $V_2^j$  are defined analogously to the  $V_1$  and  $V_2$  in the proof of Theorem 1. For each  $V^j$ , there exists an open neighborhood  $W^j$  of  $F^{A'}(\text{SU}(4)^{\times R'}) \cap V_1^j$  such that  $W^j$  has an arbitrarily small measure  $\delta_j'' > 0$ . Therefore,  $B := \cup_j W^j$  has a preimage of measure  $\leq \delta' + \sum_j \delta_j'' = \delta$ . Each of the summands, though positive, can be arbitrarily small.

The Frobenius norm induces a metric  $d_F$  on  $\text{SU}(4)^{\times R}$ . In terms of  $d_F$ , we define the function

$$d_F(\cdot, F^{A_T}(\text{SU}(4)^{\times R}) \setminus B) : F^{A'}(\text{SU}(4)^{\times R'}) \rightarrow \mathbb{R}_{\geq 0}. \quad (\text{E1})$$

This function is continuous, and  $F^{A'}(\text{SU}(4)^{\times R'})$  is compact. Therefore, the function achieves its infimum at a point  $x_{\min} \in F^{A'}(\text{SU}(4)^{\times R'})$ . Therefore, the minimal distance to  $F^{A_T}(\text{SU}(4)^{\times R}) \setminus B$  is  $d_F(x_{\min}, F(\text{SU}(4)^{\times R}) \setminus B)$ . Since  $B$  is open,  $F^A(\text{SU}(4)^{\times R}) \setminus B$  is closed and so compact. By the same argument,

$$\varepsilon(A_T, \delta) := d_F(x_{\min}, F(\text{SU}(4)^{\times R}) \setminus B) = \inf_{y \in F(\text{SU}(4)^{\times R}) \setminus B} \{d_F(x_{\min}, y)\} = d_F(x_{\min}, y_{\min}) > 0. \quad (\text{E2})$$

We have identified an  $\varepsilon > 0$  that satisfies Corollary 2.  $\square$

### Appendix F: Notions of circuit complexity

As circuit complexity is a widely popular concept, there is a zoo of quantities that measure it. We prove our main theorem for the straightforward definition of exact circuit implementation—the most straightforward notion of a circuit complexity—and for a version of approximate circuit complexity (Corollary 2) with an uncontrollably small error. In this appendix, we briefly mention other notions of complexity, partially to review other notions and partially to place the main text's findings in a wider context. Let  $U \in \text{SU}(2^n)$  denote a unitary. Ref. [10] discusses notions of approximate circuit complexity.

**Definition 9** (Approximate circuit complexity). *The approximate circuit complexity  $\mathcal{C}_u(U, \eta)$  is the least number of 2-local gates, arranged in any architecture, that implements  $U$  up to an error  $\eta > 0$  in operator norm  $\|\cdot\|$ .*

This definition is similar in mindset to the above (slightly) robust definition of a circuit complexity. For every pair  $U, U' \in \text{SU}(2^n)$  of circuits, the Frobenius distance between them satisfies

$$\frac{1}{2^n} d_F(U, U') \leq \|U - U'\| \leq d_F(U, U'). \quad (\text{F1})$$

A widely used proxy for quantum circuit complexity—one that is increasingly seen as a complexity measure in its own right—is Nielsen's geometric approach to circuit and state complexity [9, 10, 12]. This approach applies geometric reasoning to

circuit complexity and led to many intuitive insights, including Brown and Susskind's conjectures about the circuit complexity's behavior under random evolution. To connect to cost functions as considered in Nielsen's framework, consider 1-local and 2-local Hamiltonian terms  $H_1, H_2, \dots, H_m$  in the Lie algebra  $\mathfrak{su}(2^n)$  of traceless Hermitian matrices, normalized as  $\|H_j\| = 1$  for  $j = 1, 2, \dots, m$ . Consider generating a given unitary, by means of a control system, following Schrödinger's equation:

$$\frac{d}{dt}U(t) = -iH(t)U(t), \text{ wherein } H(t) = \sum_{j=1}^m h_j(t)H_j. \quad (\text{F2})$$

The control function  $[0, \tau] \rightarrow \mathbb{R}^m$  is defined as  $t \mapsto (h_1(t), \dots, h_m(t))$  and satisfies  $U(0) = \mathbb{1}$ . That is, a quantum circuit results from time-dependent control. In practice, not all of  $\mathbb{R}^m$  reflects meaningful control parameters; merely a control region  $\mathcal{R} \subset \mathbb{R}^m$  does. With each parameterized curve is associated a cost function  $c : \mathcal{R} \rightarrow \mathbb{R}$ , so that the entire cost of a unitary  $U \in \text{SU}(2^n)$  becomes

$$C(U) := \inf_{T, t \mapsto H(t)} \int_0^\tau dt c(H(t)). \quad (\text{F3})$$

We take the infimum over all time intervals  $[0, \tau]$  and over all control functions  $t \mapsto H(t)$  such that the control parameters are in  $\mathcal{R}$  for all  $t \in [0, \tau]$  and such that  $U(\tau) = U$ . Several cost functions are meaningful and have been discussed in the literature. A common choice is

$$c_p(H(t)) = \left( \sum_{j=1}^m h_j(t)^p \right)^{1/p}. \quad (\text{F4})$$

In particular,  $c_2$  gives rise to a sub-Riemannian metric. For the resulting cost  $C_2(U)$ , Ref. [11] establishes a connection between the approximate circuit complexity and the cost: Any bound on the approximate circuit complexity, with an approximation error bounded from below independently of the system size, immediately implies a lower bound on the cost.

**Theorem 4** (Approximate circuit complexity and cost [11]). *For every integer  $n$ , every  $U \in \text{SU}(2^n)$  and every  $\eta > 0$ ,*

$$\mathcal{C}_u(U, \eta) \leq c \frac{C_2(U)^3 n^6}{\eta^2}. \quad (\text{F5})$$

The quantity on the right-hand side can, in turn, be upper-bounded:  $C_2(U) \leq C_1(U)$ . This  $C_1$  has a simple interpretation in terms of a weighted gate complexity [13].

**Definition 10** (Weighted circuit complexities). *Let  $U \in \text{SU}(2^n)$  denote a unitary. The weighted circuit complexity  $\mathcal{C}_w(U)$  equals the sum of the weights of 2-local gates, arranged in any architecture, that implement  $U$ , wherein each gate  $U_j$  is weighted by its strength  $W(U_j)$ , defined through*

$$W(U) := \inf \{ \|h\| : U = e^{ih} \}. \quad (\text{F6})$$

The weighted circuit complexity  $\mathcal{C}_w(U)$  turns out to equal the cost  $C_1(U)$  for any given unitary. We can grasp this result by Trotter-approximating the time-dependent parameterized curve in the definition of  $C_1(U)$ .

**Lemma 5** (Weighted circuit complexity and cost). *If  $n$  denotes an integer and  $U \in \text{SU}(2^n)$ , then*

$$\mathcal{C}_w(U) = C_1(U). \quad (\text{F7})$$

Therefore, the weighted circuit complexity grows like the cost  $C_1$ . By implication, the circuit complexity's growth will be reflected by a notion of circuit complexity that weighs the quantum gates according to their strengths. Again, once the main text's approximate circuit complexity is established with an  $n$ -independent approximation error, one finds bounds on the weighted circuit complexity, as well.

The last important notion of circuit complexity that has arisen in the recent literature is that of Ref. [19]. Denote by  $\mathcal{G}_a \subset \text{SU}(2^{2n})$  the set of  $2n$ -qubit unitary circuits comprised of  $\leq a$  elementary quantum gates, wherein the first  $n$  qubits form the actual system and the next  $n$  qubits form a memory. Let  $\mathcal{M}_b$  denote the class of all two-outcome measurements, defined on  $2n$  qubits, that require quantum circuits whose implementation requires  $\leq b$  elementary quantum gates. Define

$$\beta(r, U) := \text{maximize } \left| \text{tr} \left( M \{ [U \otimes \mathbb{1}] |\phi\rangle\langle\phi| [U \otimes \mathbb{1}]^\dagger - [\mathbb{1}/2^n \otimes \text{tr}_1(|\phi\rangle\langle\phi|)] \} \right) \right|, \quad (\text{F8})$$

$$\text{subject to } M \in \mathcal{M}_b, |\phi\rangle = V|0^{2n}\rangle, V \in \mathcal{G}_a, r = a + b. \quad (\text{F9})$$

In terms of this quantity, Ref. [19] defined strong unitary complexity.

**Definition 11** (Strong unitary complexity [19]). *Let  $r \in \mathbb{R}$  and  $\delta \in (0, 1)$ . A unitary  $U \in SU(2^n)$  has strong unitary complexity  $\leq r$  if*

$$\beta(r, U) \geq 1 - \frac{1}{2^{2n}} - \delta, \quad (\text{F10})$$

denoted by  $\tilde{\mathcal{C}}(U, \delta) \geq r$ .

While seemingly technically involved, the definition is operational. The definition is also more stringent and demanding than more-traditional definitions of approximate circuit complexity. To concretize this statement, we denote the diamond norm by  $\|\cdot\|_\diamond$  [53].

**Lemma 6** (Implications of strong unitary complexity [19]). *Suppose that  $U \in U(2^n)$  obeys  $\tilde{\mathcal{C}}(U, \delta) \geq r + 1$  for some  $\delta \in (0, 1)$ ,  $r \in \mathbb{R}$ , arbitrary measurement procedures that include the Bell measurement. Then*

$$\min_{c_u(V) \leq r} \frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_\diamond > \sqrt{\delta}. \quad (\text{F11})$$

That is, it is impossible to accurately approximate  $U$  with circuits  $V$  of  $< r$  elementary quantum gates.

$\mathcal{U}$  and  $\mathcal{V}$  denote the unitary quantum channels defined by  $\mathcal{U}(\rho) = U\rho U^\dagger$  and  $\mathcal{V}(\rho) = V\rho V^\dagger$ . The diamond norm between them is

$$\begin{aligned} \frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_\diamond &= \frac{1}{2} \sup_{\rho} \|(U \otimes \mathbb{1})\rho(U \otimes \mathbb{1})^\dagger - (V \otimes \mathbb{1})\rho(V \otimes \mathbb{1})^\dagger\|_1 \\ &\leq \frac{1}{2} \sup_{\rho} \|[(U - V) \otimes \mathbb{1}]\rho(U \otimes \mathbb{1})^\dagger\|_1 + \frac{1}{2} \sup_{\rho} \|(U \otimes \mathbb{1})\rho[(U - V) \otimes \mathbb{1}]^\dagger\|_1. \end{aligned} \quad (\text{F12})$$

We have added and subtracted a term and have used the triangle inequality. Therefore,

$$\frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_\diamond \leq \frac{1}{2} \|U - V\|_\infty \left( \sup_{\rho} \|\rho(U \otimes \mathbb{1})^\dagger\|_1 + \sup_{\rho} \|(V \otimes \mathbb{1})\rho\|_1 \right) \leq \|U - V\|_\infty, \quad (\text{F13})$$

as the operator norm is a weakly unitarily invariant norm. Therefore,  $\tilde{\mathcal{C}}(U, \delta) \geq r + 1$  implies that  $\mathcal{C}(U, \delta) \geq r$ . That is, the strong unitary complexity of Ref. [19] is tighter than approximate circuit complexity. A topic of future work will be the exploration of the growth of approximate notions of complexity with an approximation error independent of the system size.

## Appendix G: Figures for the methods section



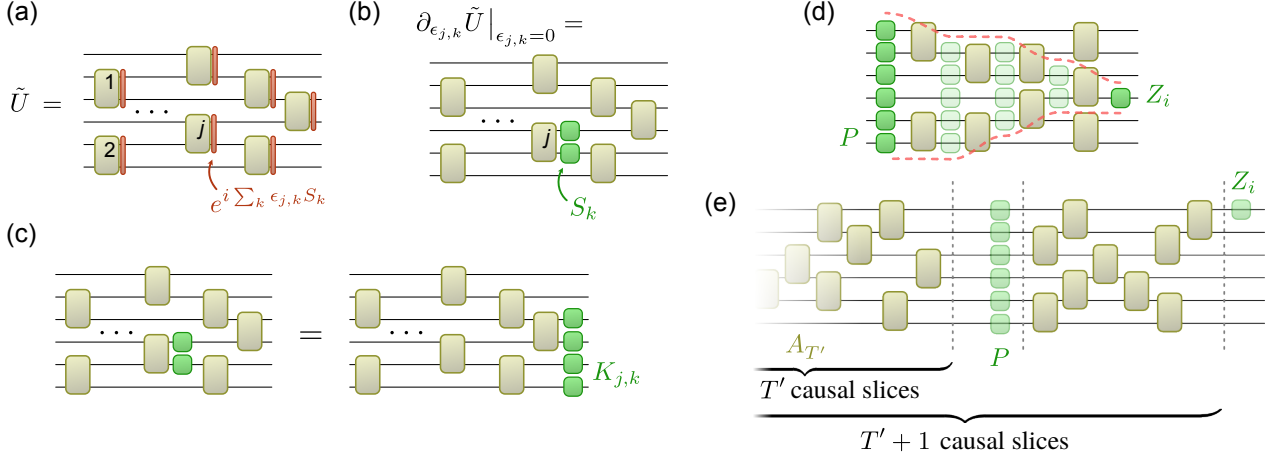


Figure 4. Our core technical result is a lower bound on the accessible dimension. **(a)** Each gate  $U_j$  is perturbed with a unitary  $e^{i\epsilon_{j,k} S_k}$ , generated by a 2-qubit Pauli operator  $S_k$  and parameterized with an infinitesimal  $\epsilon_{j,k} \in \mathbb{R}$ . Perturbing the gate perturbs the  $n$ -qubit unitary, turning  $U$  into  $\tilde{U} \approx U$ . **(b)** A key quantity is the derivative of  $\tilde{U}$  with respect to a parameter  $\epsilon_{j,k}$ , evaluated at  $U$ . Taking this derivative is equivalent to inserting the Pauli string  $S_k$  immediately after the gate  $U_j$ . **(c)** The derivative depicted in panel (b) is equivalent to following the circuit with a Hermitian operator  $K_{j,k}$  [Eq. (1)]. The operator  $K_{j,k}$  results from conjugating  $S_k$  with the gates after  $U_j$ . If the circuit consists of Clifford gates, then  $K_{j,k}$  is a Pauli string, since Clifford gates map the Pauli strings to Pauli strings. Therefore, a perturbation of  $U_j$  in the direction of  $S_k$  results in a perturbation of the resulting unitary  $U$  in the direction of  $K_{j,k}$  in  $SU(2^n)$ . **(d)** The following is true of every backwards-light-cone-containing block and every Pauli string  $P$  (leftmost green squares): The block's gates can be chosen to be Cliffords that map  $P$  to a single-site  $Z$ . The Clifford gates first map  $P$  to a Pauli string that acts nontrivially on fewer qubits (pale green squares), then to a Pauli string on fewer qubits, and so on until the Pauli string dwindles to one qubit (rightmost green square). **(e)** Our lower bound is proven by recursion as explained in the main text.

## QUANTUM ADVANTAGE EXPERIMENTS

---

### 4.1 CLOSING GAPS OF A QUANTUM ADVANTAGE WITH SHORT-TIME HAMILTONIAN DYNAMICS

In principle, errors in quantum computers can be corrected faster than they appear. This result, known as the “threshold theorem”, was shown already in the 1990s [ABO97]. With it comes the prospect of quantum algorithms that efficiently solve classical problems for which no polynomial classical algorithm is known [Sho99b]. Indeed, it is widely believed that such an efficient classical algorithm does not exist and that quantum computers are therefore exponentially separated from classical computers. This would violate the extended Church-Turing thesis, which claims that any model of computation that can be build in nature can be simulated (with a polynomial overhead) on a classical computer [VSD86,BV93]. An experimental realization of such a violation would be a major milestone in the field of computing and is referred to as a *quantum advantage* or “*quantum supremacy*” [Pre13,HE22].

In order to conclusively demonstrate the superior computational power of quantum devices we must hold ourselves to a particularly high standard of evidence. While several examples of quantum devices that outperform certain classical algorithms have been reported [TCF<sup>+</sup>12, BFH<sup>+</sup>15, CHZ<sup>+</sup>16, BSK<sup>+</sup>17, ZPH<sup>+</sup>17], to have high confidence that these devices are providing bona-fide quantum speedups, we must give evidence that *no classical algorithm* will ever be able to solve this problem efficiently.

However, the realization of fault-tolerant devices capable of outperforming classical computers for practical problems appears to be far out of reach for current technology [RWS<sup>+</sup>17, CKM19, GE21]. Recently, a paradigm of sampling based quantum advantage schemes emerged. This appears to be a particularly natural task for quantum computers as all measurements are intrinsically random. We know from Bell’s theorem that the correlations between measurement outcomes can, in general, not be generated with a classical (realistic) local hidden variable model. We expect that, similarly, classical computers cannot generate the correlations in general quantum measurements efficiently and refer to Ref. [HE22] for the discussion of quantum advantages as a computational analogue of Bell’s inequality.

Multiple candidates for such random quantum sampling schemes emerged such as commuting quantum circuits [BJS11,BMS16a,BMS17], boson sampling [AA13] and universal random circuit sampling [BIS<sup>+</sup>18,AAB<sup>+</sup>19]. In fact, in a technological breakthrough, a successful experiment was prominently reported by Google Inc. [AAB<sup>+</sup>19], demonstrating a verified advantage over classical algorithms of that time.

However, in light of recent advances in classical simulations of Google’s experiment [PCZ21], it is possible that the race between quantum experiments and classical simulation will only be decided in a regime where even the verification algorithms fail. This is because all verification procedures for Google’s random circuit sampling experiment also require exponentially long times. Without verification there is no way of knowing whether the experiment was successful.

In the near future, only imperfect and small universally programmable quantum devices are becoming available in laboratories around the world [BIS<sup>+</sup>18, WHL<sup>+</sup>17], however. At the same time, large-scale quantum simulators that outperform known classical algorithms are already available today in platforms such as cold atoms in optical lattices [BFH<sup>+</sup>15, CHZ<sup>+</sup>16, BDN<sub>12</sub>] or ion traps [ZPH<sup>+</sup>17]. However, rigorous results on the hardness of those simulations are very sparse and what is more the available hardness results rely on unproven albeit plausible conjectures beyond standard complexity-theoretic assumptions [HE22].

Quantum advantage schemes based on quantum simulators that involve the constant-time evolution of translation-invariant Ising Hamiltonians [GWD<sub>17</sub>, BVHS<sup>+</sup>18, HKSE<sub>17</sub>] have been proposed. These schemes come with an efficient and rigorous certification protocol that only requires partial trust in single-spin measurements. One of the simplest realizations can be performed in three steps:

1. **Preparation.** Arrange  $N$  qubits on a rectangular lattice, draw  $N$  random angles  $\beta_i \in (0, 2\pi]$  and prepare the state (up to normalization)

$$|\psi_\beta\rangle = \bigotimes_{i=1}^N (|0\rangle + e^{i\beta_i}|1\rangle). \quad (4.1)$$

2. **Coupling.** Couple along the edges of the lattice via time evolution under an Ising Hamiltonian

$$H := \sum_{i,j \in E} J_{i,j} Z_i Z_j - \sum_{i \in V} h_i Z_i. \quad (4.2)$$

3. **Measurement.** Measure each qubit in the  $X$  basis.

While only little entanglement is generated, the resulting probability distribution can be related to highly correlated distributions. This is because the protocol is equivalent to measuring a cluster state in a random basis and therefore implements a randomized measurement based quantum computation. From a practical perspective, the design of such simple translationally invariant schemes is highly desirable in that they are very close to what can be achieved in scalable present-day experimental architectures such as cold atoms in optical lattices [BDN<sub>12</sub>].

The central open problem in the complexity theoretic argument for hardness of all such sampling schemes revolves around its robustness to noise [AA<sub>13</sub>, BMS<sub>16b</sub>]. Proving this key conjecture called *approximate average-case hardness* has remained elusive for all known practical schemes that are amenable to the Stockmeyer proof strategy. Aaronson and Arkhipov have also observed, however, that evidence for approximate average-case hardness can be provided using certain properties of the sampled distribution: First, *exact average-case hardness* constitutes a necessary criterion for the approximate version thereof. Second, the so-called *anti-concentration* property reduces the notion of approximation that is necessary for the hardness proof to a more plausible one that involves only relative errors. One of the key challenges in the field, therefore, is to close these conceptual, mathematical, and complexity-theoretic “loopholes” for schemes that are feasible on large-scale quantum simulators.

In this work, we close these loopholes simultaneously for the arguably simplest quantum simulation architecture on a square lattice of Ref. [BVHS<sup>+</sup>18] -- thus bringing it up to the highest

standard to date in terms of evidence for computational intractability. First, we prove *anti-concentration* for this model. In fact, our main contribution is to establish an even stronger property than anti-concentration, namely, that the effective circuits generated by the architectures mimic Haar-randomness up to second moments -- surprisingly -- already on square ( $n \times \mathcal{O}(n)$ ) lattices. In precise terms, we prove that these circuits form an approximate 2-design.

**Theorem 14.** *Consider the architectures of quantum simulation with local rotation angles chosen uniformly from  $[0, 2\pi)$  on an  $n \times m$  lattice with  $m \in \mathcal{O}(n + \log(1/\varepsilon))$ . When measuring the first  $m - 1$  columns in the  $X$ -basis, the effective unitary acting on the last column forms a relative  $\varepsilon$ -approximate unitary 2-design.*

In fact, the emergence of relatively  $\varepsilon$ -approximate 2-designs is a much more powerful result than mere anti-concentration. First, observe that the 2-design property directly implies anti-concentration [HBVSE18, BFK18, MB17, HM18]. Second, we note that generating the moments of the Haar measure is considered even stronger evidence for hardness of classical simulation than mere anti-concentration [BIS<sup>+</sup>18]. But already rigorously establishing anti-concentration is a difficult endeavor, in our case particularly so due to the low -- in fact constant -- depth of the circuits. Indeed, for the prominent case of random circuit sampling anti-concentration holds at depth  $\mathcal{O}(\sqrt{n})$  on a  $\sqrt{n} \times \sqrt{n}$  2D grid [HM18], and at depth  $\mathcal{O}(n)$  in 1D [BHH16, HBVSE18], but it is not expected for constant depth [BIS<sup>+</sup>18, BMS17]. With our work, we prove anti-concentration at constant depth, but in a very different model of random circuits. Our result implies the first non-trivial anti-concentration bound for constant-time, translation-invariant dynamics on a square lattice, going significantly beyond direct measurement-based embeddings [MSM17, HBVSE18, GWD17, MGDM18, MGDM19].

As our second main contribution, we prove average-case hardness for *exactly evaluating* the output probabilities of the architectures. To do so, we extend a result of [BFNV19] showing exact average-case hardness of universal circuit sampling [BIS<sup>+</sup>18] to the translation invariant case. Informally, we obtain the following result:

**Theorem 15** (Average-case hardness). *It is #P-hard to exactly compute any  $3/4 + 1/\text{poly}(N)$  fraction of the output probabilities of the architectures of quantum simulation.*

Our work also demonstrates that these average-case hardness methods are applicable to many other sampling architectures, such as continuous forms of IQP circuits [BMS16b] and other measurement-based schemes [GWD17, MSM17, MGDM19]. Our results have significant implications for quantum simulation, giving the strongest complexity theoretic evidence to date that “simple” constant-time Hamiltonian evolutions on the square lattice [BVHS<sup>+</sup>18] cannot be classically simulated.

Ref. [HHB<sup>+</sup>19] is not contained in the online version for copyright reasons.

---

COMPUTATIONAL COMPLEXITY AND RANDOM TENSOR NETWORKS

---

## 5.1 CONTRACTING PROJECTED ENTANGLED PAIR STATES IS AVERAGE-CASE HARD

Storing a general state of the system on a computer is impossible and hence one seeks for efficient variational families of states. Tensor networks are a prime example of such an ansatz class [FNW92,VMCo8,Sch11,Orú14,BC17,ECP10]. Despite their spectacular success in one-dimension [VGRCo4,GKSS05,VC06,MWNM07,TCF<sup>+</sup>12,FKE<sup>+</sup>13,PTBO10,BPM12,KBM12,Pro11,RSB<sup>+</sup>13] as so-called matrix-product states [PGVWC07,VC06,VMCo8], the most natural tensor network ansatz in two-dimensions, called projected entangled pair states (PEPS) [VWPGCo6], turned out to be burdened by a peculiar difficulty: even to calculate the normalization of PEPS is computationally intractable as has been shown by [SWVCo7]. More precisely, the normalization or evaluation of a local expectation value within the PEPS ansatz class is a computational task which is complete for the complexity class #P.

In this section, we provide complexity theoretical evidence against strong heuristic algorithms for generic PEPS. This extends the worst-case #P-hardness result [SWVCo7] to the average-case and is an even more challenging obstruction to overcome. More precisely, we show that determining the normalization of as well as expectation values in a PEPS is #P-hard for a constant fraction of the instances.

It is known that PEPS contraction algorithms often work well in practice for reasonable condensed-matter systems [LCB14,RTP<sup>+</sup>17] which at first sight may seem at odds with the results presented here and in Ref. [SWVCo7]. However, many important problems have additional structure that may render the PEPS contraction feasible. Specifically, it was proven in Ref. [SBE17] that local normalized expectation values of injective PEPS with *uniformly gapped parent Hamiltonian* can be evaluated in quasi-polynomial time, i.e., faster than conjectured by the exponential-time hypothesis.

We define PEPS on a family of graphs  $G = (V, E)$  with  $|V| = N$  vertices. Every vertex  $v$  stands for a local spin system of *bond dimension*  $D$ . In the projective construction of PEPS one thinks of every edge  $e \in E$  as a maximally entangled state  $\sum_{i=1}^D |i\rangle|i\rangle$  in a virtual  $D$ -dimensional spin systems. A specific PEPS is described by operators  $P^{[v]} : \mathbb{C}^D \otimes \dots \otimes \mathbb{C}^D \rightarrow \mathbb{C}^d$ . It is defined as the state vector in  $\mathcal{H}$  resulting from the application of all  $P^{[v]}$  for all  $v \in V$ . Note that by this the obtained PEPS is not necessarily normalized.

While the polynomial description is a clear advantage of PEPS, it remains notoriously difficult to contract PEPS. This is needed for obtaining physical quantities of interest like expectation values. Specifically, we consider the following computational task, which is an essential ingredient in PEPS contraction algorithms. It is one of the key insights in Ref. [SWVCo7] that this problem is in fact #P-complete for the case that  $G$  is a square lattice.

We can formalize this task as follows:

**Problem 1** (PEPS-contraction).

**Input.** A graph  $G$  and corresponding finite PEPS-data  $(P^{[v]})_v$  describing an unnormalized state  $|\psi\rangle$  and with bond dimension  $D = \text{poly}(N)$ .

**Output.**  $\langle\psi|\psi\rangle$ .

We find that PEPS-contraction is hard in the same sense as canonical combinatorial problems [AA13, Lip91]: Both problems admit random self-reducibility. A problem is *randomly self-reducible* if the evaluation of any instance  $x$  can be reduced to the evaluation of random instances  $y_1, \dots, y_k$  with a bounded probability independent of the input. The seminal result by [Lip91] proves this property for the permanent.

While worst-case results are ubiquitous in computer science, average-case hardness is rarely established rigorously and most known examples of average-case complexity results concern #P-complete problems. Moreover, more often than not, the proof is a variant of Lipton's technique. An exception are lattice problems starting with the work of Ajtai [Ajt96].

In the proof of our main result, we show random self-reducibility for PEPS contraction:

**Theorem 16** (Average case hardness of PEPS contraction). *Suppose there exists an algorithm  $\mathcal{O}$  that solves Problem 1 for square lattices in polynomial time with probability  $\frac{3}{4} + \frac{1}{\text{poly}(N)}$  when instances are drawn from  $\mathcal{P}$ . Then, there exists a randomized algorithm  $\mathcal{O}'$  that solves any instance of Problem 1 in polynomial time with exponentially high probability  $1 - 2^{-\text{poly}(N)}$ .*

To do so, we consider a generic PEPS in the sense that all entries of the tensor  $P^{[v]}$  are drawn independently at random from the finite precision approximation of the normal distribution centered around 0 and with standard deviation  $\sigma$ . The main result is then that any algorithm that solves the PEPS-contraction for around  $\frac{3}{4}$  of the instances can be lifted to a randomized algorithm that solves PEPS-contraction in the worst-case and is, hence, very unlikely to exist. Indeed, our proof works analogously to Lipton's technique for the permanent and is intriguingly simple: We interpolate between any instance of the problem and a generic instance. We then prove that polynomially many queries to a good enough heuristic solver would suffice to learn the permanent as a function of the interpolation parameter. Then, evaluating in the original function yields a good estimate with high probability. As the original instance was arbitrary, we conclude random self-reducibility.

## Contracting projected entangled pair states is average-case hard

Jonas Haferkamp, Dominik Hangleiter, Jens Eisert, and Marek Gluza

*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*



(Received 31 October 2018; revised manuscript received 27 October 2019; published 3 January 2020)

An accurate calculation of the properties of quantum many-body systems is one of the most important yet intricate challenges of modern physics and computer science. In recent years, the tensor network ansatz has established itself as one of the most promising approaches enabling striking efficiency of simulating static properties of one-dimensional systems and abounding numerical applications in condensed matter theory. In higher dimensions, however, a connection to the field of computational complexity theory has shown that the accurate normalization of the two-dimensional tensor networks called projected entangled pair states (PEPS) is  $\#P$ -complete. Therefore an efficient algorithm for PEPS contraction would allow solving exceedingly difficult combinatorial counting problems, which is considered highly unlikely. Due to the importance of understanding two- and three-dimensional systems the question currently remains: Are the known constructions typical of states relevant for quantum many-body systems? In this work, we show that an accurate evaluation of normalization or expectation values of PEPS is as hard to compute for typical instances as for special configurations of highest computational hardness. We discuss the structural property of average-case hardness in relation to the current research on efficient algorithms attempting tensor network contraction, hinting at a wealth of possible further insights into the average-case hardness of important problems in quantum many-body theory.

DOI: [10.1103/PhysRevResearch.2.013010](https://doi.org/10.1103/PhysRevResearch.2.013010)

### I. INTRODUCTION

Determining the properties of quantum many-body systems is of paramount importance in our efforts to understand conductance and thermodynamics of solid-state materials [1,2], designing new sensors and devising novel quantum technologies [3], inferring nuclear processes in stars or the early universe [4,5]. However, oftentimes it is not possible to find degrees of freedom enabling a concise description of a given system in terms of an effective model featuring essentially no interactions. In such a case, there is usually no easy way out but to calculate numerically observables of interest from a Hamiltonian description [6–12]. Here, however, we face a particular challenge namely that the state space of quantum many-body systems demands a number of parameters that grows exponentially with the amount of constituents of the system. If so, even storing the state of the system on a computer becomes impossible and hence one seeks for efficient variational families of states. Tensor networks are a prime example of such an ansatz class [10,13–17]. Despite their spectacular success in one dimension [18–29] as so-called matrix-product states [14,20,30], the most natural tensor network ansatz in two-dimensions, called projected entangled pair states (PEPS) [31], turned out to be burdened by a peculiar difficulty: even to calculate the normalization

of PEPS is computationally intractable as has been shown in Ref. [32].

More precisely, the normalization or evaluation of a local expectation value within the PEPS ansatz class is a computational task which is complete for the complexity class  $\#P$ , i.e., is as hard as any other problem in this class [33–35]. Paradigmatic  $\#P$  problems consist in counting the solutions to decision problems which are complete for the class  $NP$ . Intuitively, counting the solutions to a hard problem can only be harder. Within the current state of knowledge in computer science the optimal runtime for  $NP$ -complete problems is unknown. However, it is widely conjectured that there are no algorithms with polynomial runtime solving any  $NP$ -complete problem. For the famous SAT-problem, there is even the *exponential-time hypothesis* [36], which conjectures that an exponential runtime is optimal for the problem.

Physically, one can invoke the Church-Turing-Deutsch principle [37] that interprets computations as physical processes.  $NP$  has been established to correspond to the cooling of spin glasses [38]. These materials are known to sometimes take an extremely long time to cool down. On the other hand, very many solid-state materials seem to cool down much faster. Indeed, insights in computer science suggest that the hardness of  $NP$ -complete problems lies in few tough instances with particularly rugged landscape. Phenomena like this are described in the framework of *average-case* complexity. While many  $NP$ -complete problems like 3-SAT are unlikely to be hard on average for uniform distributions [39], average-case hard problems are ubiquitous for the class  $\#P$ . Recently, first examples directly relevant to demonstrating computational separation between classical and quantum devices have been pointed out [40,41].

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



There are several approaches for a rigorous theory of average-case complexity. Arguably the most natural is *random self-reducibility*, an immediate consequence of which is that a machine powerful enough to solve, e.g., three quarters of the instances would allow solving all instances. Thus it becomes implausible to find heuristic algorithms that solve significant numbers of instances as the self-reducibility structure would imply efficiency even for those instances that are particularly hard.

In this work, we provide strong complexity theoretical indications that the latter is not the case for generic PEPS due to a random self-reducibility structure that we uncover. This extends the worst-case #P-hardness result [32] to the average case and is an even more challenging obstruction to overcome. Technically, we make an extensive use of the recent insightful work in Ref. [41], where average-case hardness has been established in the context of quantum circuits, and we also employ some of the results established in Ref. [40]. Our main result is the following theorem.

*Theorem 1 (Informal).* Contracting a subset making up a  $\frac{3}{4} + \frac{1}{\text{poly}(N)}$  fraction of the instances drawn from an entrywise Gaussian distribution is #P-hard.

We explain in Sec. IV how this can be generalized to many probability distributions that satisfy an autocorrelation property. In particular a similar statement would hold if the local tensors are drawn from a uniform distribution supported on a bounded region, i.e., the overall “shape” is not crucial as long as the distribution is not infinitely peaked or has unusually broad tails. Firstly, this rules out the possibility that the computational hardness could be hidden in particular instances that are intractable, as it says that one could use the algorithm  $\mathcal{O}$  to construct an algorithm  $\mathcal{O}'$  that is efficient for all inputs. Secondly, it is important to note that Theorem 1 requires exact computation [32] but a different variant of Theorem 1 shows the following. Approximation up to errors of the form  $2^{-\text{poly}(N)}$  for  $N$  the system size, is also intractable on average, however, under stronger requirements on the algorithm  $\mathcal{O}$ . Our choice of the probability distribution is similar to that of Sec. 9.1 of Ref. [40], where the evaluation of the so-called *permanent* is considered which is also a #P-complete computational problem. Note that the result holds for arbitrary graphs as well, though the statement is trivial in one dimension [42].

In certain special instances fast algorithms might still be feasible. For example it is known that matrix-product states admit a polynomial time deterministic contraction algorithm [42]. However, even in two dimensions, this can happen under strong physical assumptions forcing the problem to admit a local structure [43,44]. Additionally, for certain subclasses some heuristic algorithms [42–63] (see Refs. [45,64] for reviews) yield results of practical importance [65–74]. Our average-case hardness result, however, suggests that these approaches could break down even for relevant PEPS instances as otherwise difficult computational problems would admit (quasi-) polynomial algorithms.

Physically, for disordered systems, one would expect any accurate ground state approximation by a PEPS to inherit the randomness of the Hamiltonian [75]. Hence in this setting, we provide evidence of intractability. Oftentimes, however, further physical assumptions are justified: While these com-

pletely generic PEPS are relevant for the study of strongly disordered systems, in many practically meaningful settings (in particular in the study of topological order), the relevant PEPS are translation-invariant. Remarkably, a worst-to-average case reduction as described in this paper works just as well for translation-invariant systems but we are unaware of a hardness result in the worst case for such systems.

## II. DISCUSSION

Before we formalize the above in a rigorous setting, we discuss various aspects of this result.

### A. Translation invariance

In many physical applications, e.g., in solid state materials and specifically in systems admitting topological order, the system of interest is translation-invariant. Hence, the data specifying the PEPS efficiently should reflect this symmetry and one would naturally set all local tensors to be equal. In this case, we do not know the corresponding computational problem to be #P-hard, for example the #P-hard instances in Ref. [32] are not translation-invariant. However, our worst-to-average case reduction works just as well in this special case, simply by choosing  $(Q^{[v]})_v = (Q)_v$ , where  $Q$  is drawn from the Gaussian distribution  $\mathcal{N}_{\mathbb{C}}(0, \sigma)^{D^d}$ . The same argument and statement of the main theorem goes through. This leaves us with two mutually exclusive options: If the translation-invariant problem is hard for a complexity class  $C$ , then it follows that the problem is  $C$ -hard on average in the sense of our main theorem. If the problem is merely in  $P$ , then it is enough to find a heuristic for about 3/4 of the inputs to find a full randomized algorithm. On the other hand, if  $C = \#P$ , then even the translation-invariant PEPS contraction problem would appear to be average-case intractable. We are unaware of random self-reducibility results for complexity classes other than #P. We thus expect a dichotomy: Either the translation-invariant problem is in  $P$  or it is #P-complete.

### B. Evaluation precision

As far as we know, it is state of the art in computer science to prove random self-reducibility structures for problems given the promise that  $\mathcal{O}$  works with at least exponential precision. In fact, we can improve our main theorem for this case too, at the cost of requiring  $\mathcal{O}$  to function with a probability of  $1 - \frac{1}{12N}$ , where  $N$  denotes the system size. The reason for this trade-off is that subtleties arise in the technical steps, where the Berlekamp-Welch algorithm has to be replaced with a noise-resistant method. However, in the bigger picture, it does not seem possible to extend the seminal idea of Lipton to  $\mathcal{O}$  working with lower precision. The reason is that the method crucially depends on the extrapolation of polynomials which is highly sensitive to noise. Related questions of precision relaxation are of interest in quantum information theory in the context of searching for quantum speed-ups. Here, certain precision relaxations are conjectured to be average-case hard as well [40,41].

### C. Expectation values

The computational problem is concerned with PEPS contractions. The quantity that one computes is the norm of

the respective PEPS. However, in most physical applications the quantities of interest are expectation values of a local observable  $\hat{A}$

$$\langle \hat{A} \rangle_\psi = \frac{\langle \psi | \hat{A} | \psi \rangle}{\langle \psi | \psi \rangle}, \quad (1)$$

where  $|\psi\rangle$  refers to the PEPS specified by local tensors. Notice that this problem and its unnormalized version have both been proven to be  $\#\mathbf{P}$ -complete in Ref. [32] as well. For any algorithm that uses PEPS normalization as an intermediate step our main theorem is directly of interest and reflects the fundamental structure of the problem at hand. In the general case we can prove a worst-to-average result for this quantity as well. It is easy to see that our discussion of PEPS contraction carries over to the discussion of unnormalized expectation values. We show that a close analog of Theorem 1 holds for this quantity as well. The normalized expectation value is slightly more subtle in the following sense: the analog of the function  $q$  is not a polynomial but a rational function  $q/p$  where the degrees of both polynomials  $q$  and  $p$  are bounded by  $2N$ . We can then use interpolation for rational functions on enough sampling points to obtain the respective coefficients.

#### D. Implications on practical tensor network algorithms

The results found here have interesting implications to the performance of PEPS contraction algorithms aimed at solving condensed-matter problems [10,14,15]. There are three insights that are important in this respect. Firstly, the results laid out here relate average-case to worst-case complexity. In that, they apply to any tensor network contraction algorithm as the structure of random self-reducibility shows that if a given method  $\mathcal{O}$  has trouble at less than a quarter of instances, these can in principle be treated with a small polynomial runtime overhead by our construction of the randomized algorithm  $\mathcal{O}'$  (and, for that matter, our results also pertain to algorithms in  $\mathbf{P}$ ). Secondly, it is known that PEPS contraction algorithms often work well in practice for reasonable condensed-matter systems [45,64] which may seem at first sight at odds with the results presented here and in Ref. [32]. For this, one has to acknowledge that many important problems have additional structure that may render the PEPS contraction feasible. Specifically, it has been proven in Ref. [44] that local normalized expectation values of injective PEPS with *uniformly gapped parent Hamiltonian* can be evaluated in quasipolynomial time, i.e., faster than conjectured by the exponential-time hypothesis. Following up on this observation, it seems conceivable that one can devise PEPS algorithms that provide ground states of systems in a trivial phase (possibly even with convergence proofs), by making use of techniques of quasiadiabatic evolution [76,77], applying short circuits to product states as ground states of trivial parents. Having said that, any such approach would require keeping track of ground states of families of Hamiltonians. Thirdly, in most practical algorithms used in practice, in contrast, some initial condition for the PEPS is chosen, which is iteratively refined via sweeps, until a good convergence to the ground state is encountered. In fact, in practice, the PEPS data are initially often chosen randomly, following a refinement in sweeps by iteratively minimizing the energy evaluated from

a local Hamiltonian. The results laid out here show that it is crucial to devise meaningful schemes making reasonable choices of these initial conditions. However, our average-case hardness results of PEPS contraction indicate that one should be particularly cautious when choosing such initial states.

### III. PROBLEM SETTING

We now come to the technical section of this paper. In this section, we describe the problem in a rigorous setting.

#### A. Projected entangled pair states

Here we recall the definition of PEPS [52] and review the computational problem from Ref. [32] concerning the contraction of PEPS. We consider a family of graphs  $G = (V, E)$  with  $|V| = N$ . Every vertex  $v$  stands for a local spin system described by a Hilbert space  $\mathcal{H}_v := \mathbb{C}^d$ . The physical Hilbert space is, thus,  $\mathcal{H} := \mathcal{H}_v^{\otimes N} = (\mathbb{C}^d)^{\otimes N}$ . In the projective construction of PEPS, one thinks of every edge  $e \in E$  as a maximally entangled state  $\sum_{i=1}^D |i\rangle|i\rangle$  in a virtual  $D$ -dimensional spin systems. A specific PEPS is then described by linear operators  $P^{[v]} : \mathbb{C}^D \otimes \dots \otimes \mathbb{C}^D \rightarrow \mathbb{C}^d$ , where the number of copies of  $\mathbb{C}^D$  is the number of adjacent edges for  $v$ . It is defined as the state vector in  $\mathcal{H}$  resulting from the application of all  $P^{[v]}$  for all  $v \in V$ . Note that by this the obtained PEPS is not necessarily normalized. The virtual dimension is assumed to satisfy  $D = \text{poly}(N)$  and is called *bond dimension*. In our discussion, it will be crucial to discriminate between the PEPS, which is a state vector in  $\mathcal{H}$ , and its specification  $(P^{[v]})_v$ . We will refer to the latter as *PEPS data*. A PEPS is called *translation-invariant* if the local tensors satisfy  $P^{[v]} = P^{[w]} = P$  for all  $v, w \in V$ . These states have already been proven to be immensely useful in condensed matter research but the full regime of applicability is still open. Here, we assume open boundary conditions but our results carry over to the periodic case too.

#### B. PEPS evaluation

PEPS are described by polynomial data only. However, the physical problem we want to tackle remains notoriously difficult in that contraction of PEPS is computationally hard. This is needed for obtaining physical quantities of interest like expectation values of local observables. Specifically, the following computational tasks are the essential ingredients of PEPS contraction algorithms:

*Problem 1 (PEPS-contraction). Input:* A graph  $G$  and corresponding finite PEPS data  $(P^{[v]})_v$  describing an unnormalized state  $|\psi\rangle$  and with bond dimension  $D = \text{poly}(N)$ .

*Output:*  $\langle \psi | \psi \rangle$ .

*Problem 2 (PEPS-contraction:UEV). Input:* The same input as in Problem 1 and additionally a local observable  $\hat{A}$ .

*Output:*  $\langle \psi | \hat{A} | \psi \rangle$ .

*Problem 3 (PEPS-contraction:NEV). Input:* The same input as in Problem 1 and additionally a local observable  $\hat{A}$ .

*Output:*  $\langle \psi | \hat{A} | \psi \rangle / \langle \psi | \psi \rangle$ .

It is one of the key insights in Ref. [32] that these problems are in fact  $\#\mathbf{P}$ -complete for the case that  $G$  is a square lattice. In the following, we recall the arguments leading to this observation. The construction uses measurement based quantum

computing [78–80]. Measurement based quantum computing based on cluster states performs a computation by initializing the cluster state on a square lattice and successively applying local sharp (projective) measurements to the local qubits. This is a universal model of a quantum computer and we can use it to encode any quantum circuit in a PEPS with polynomially bounded bond dimension. Notice first that the cluster state is a PEPS with bond dimension  $D = 2$ . However, the outcome of the quantum computation performed by the measurements depends on the random outcomes. This is dealt with by correcting the outcome with Pauli operators depending on the random outcomes. The PEPS encoding the quantum circuit is now obtained by applying an additional projector  $|a\rangle\langle a|$ , where  $a$  is the outcome that does not give rise to a nontrivial Pauli correction. Hardness follows from encoding the problem of counting solutions for a Boolean formula: Given a Boolean formula  $f$ , finding  $\#_1(f) := |\{x, f(x) = 1\}|$  is #P-complete.

We prove all results for two canonical choices: The first is to draw entry-wise from a uniform distribution centered around zero and truncated at some chosen threshold  $\sigma$ , which we will denote by  $\mathcal{U} = \mathcal{U}_{\mathbb{C}}(0, \sigma)$  and the product distribution by  $\mathcal{P}_1 := \mathcal{U}^{D^d dN}$ . Almost equivalently we could draw from a Gaussian distribution. We will denote this Gaussian distribution with  $\mathcal{P}_2 := \mathcal{G}^{D^d dN} := \mathcal{N}_{\mathbb{C}}(0, \sigma)^{D^d dN}$ . This is reminiscent to a discussion about the permanent with entries in the complex numbers in Sec. 9.1 of Ref. [40]. More precisely, we prove the following technical theorems.

*Theorem 2 (Worst-to-average reduction).* Suppose there exists a machine  $\mathcal{O}$  that solves Problem 1 or 2 within precision  $2^{-\text{poly}(N)}$  for square lattices in polynomial time with a probability of  $1 - \frac{1}{12N}$  over the instance drawn from  $\mathcal{P}_i$  for  $i = 1, 2$ . Then, there exists a machine  $\mathcal{O}'$  that solves any instance with precision  $2^{-\text{poly}(N)}$  of the respective problem in randomized polynomial time with exponentially high probability.

We will prove this theorem first, as it requires the most technical work. If we do not relax to exponential precision but require perfect arithmetical evaluation of the machine  $\mathcal{O}$ , we obtain a much stronger worst-to-average reduction:

*Theorem 3 (Stronger worst-to-average reduction).* Suppose it exists a machine  $\mathcal{O}$  that solves Problem 1 or 2 exactly for square lattices in polynomial time with a probability of  $\frac{3}{4} + \frac{1}{\text{poly}N}$  drawn from  $\mathcal{P}_i$ , with  $i = 1, 2$ . Then, there exists a machine that solves any instance of the respective problem in randomized polynomial time with exponentially high precision.

Finally, requiring perfect evaluation, we obtain a worst-to-average reduction for the normalized expectation value problem as well:

*Theorem 4 (Normalized expectation values).* Suppose it exists a machine  $\mathcal{O}$  that solves Problem 3 exactly for square lattices in polynomial time with a probability of  $\frac{3}{4} + \frac{1}{\text{poly}N}$  drawn from  $\mathcal{P}_i$  with  $i = 1, 2$ . Then there exists a machine that solves any instance of the respective problem in randomized polynomial time with exponentially high precision.

### C. Proof idea

There are several precise mathematical candidates for a definition of *average-case hardness*. We find that PEPS

contraction is average-case hard in the same sense as certain combinatorial problems [40,81]: They admit a property called random self-reducibility. A problem is *randomly self-reducible* if the evaluation of any instance  $x$  can be reduced to the evaluation of random instances  $y_1, \dots, y_k$  with a bounded probability independent of the input. We will sketch how this is done for the permanent and PEPS giving the essential proof idea, see Ref. [41] for a particularly clear exposition in the context of quantum circuits. The complete argument can be found in Sec. IV.

In a seminal result, Ref. [81] has proven random self-reducibility for the evaluation of the permanent, a function that takes as an input a square matrix and outputs a number. The permanent of an  $n \times n$  matrix  $A$  over a finite field is defined as the “determinant without signs”:

$$\text{perm}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^n A_{i, \sigma(i)}, \quad (2)$$

where  $S_n$  is the symmetric group. However, very unlike the determinant, the permanent turns out to yield a difficult combinatorial problem: Its evaluation has been proven to be #P-complete in Ref. [82]. The proof of random-self reducibility is rooted in the algebraic fact that the permanent defines a polynomial of degree  $n$  in the entries of its input matrix  $A$ . More precisely, the strategy is to take any (hard) instance  $A$  that we want to compute, draw a uniformly random matrix  $B$  and define

$$E(t) := A + tB, \quad (3)$$

for a parameter  $t$  in the finite field. Notice that  $E(t)$  is uniformly random for any  $t$  because  $B$  is, even though  $E(t)$  and  $E(t')$  are correlated. The permanent of these matrices is a polynomial  $q(t) := \text{perm}(E(t))$  of degree  $n$ . Even if the algorithm  $\mathcal{O}$  fails to accurately output  $\text{perm}(A)$  it will, by assumption, likely correctly evaluate  $q(t_i)$  for a choice of  $t_i$ . The idea is to infer  $q(0)$  from the values at  $\{t_i\}$  via polynomial interpolation. We will explain this step in more detail in the next paragraph for the setting of PEPS.

We sketch how the worst to average-case reduction works for PEPS contractions. For a detailed and formal proof we refer to Sec. IV. A major difference to Lipton’s result for the permanent is that we work over the complex numbers, for which there is no uniform distribution. Instead, we work with an entry-wise Gaussian distribution.

Intuitively, we scramble independently the individual tensors. Given a hard instance  $(P^{[v]})_v$ , we draw random PEPS data  $(Q^{[v]})_v$  and define

$$(R(t)^{[v]})_v := t(P^{[v]})_v + (1 - t)(Q^{[v]})_v. \quad (4)$$

Thus  $(R(0)^{[v]})_v = (Q^{[v]})_v$  and  $(R(1)^{[v]})_v = (P^{[v]})_v$ . Notice that PEPS data and PEPS since the above definition has nothing to do with the addition of the corresponding states. This choice of a scrambled operator is suitable for us because it allows us to deal with a subtlety arising from the fact that the PEPS data  $(R(t)^{[v]})_v$  is not Gauss-random even though  $(Q^{[v]})_v$  is. This is different to the setting of Ref. [81] but has been worked out for boson sampling [40], where it was shown that the difference is immaterial for small  $t$ . This carries over to our case as we discuss in Sec. IV.

#### IV. PROOFS

We can now provide rigorous proofs for Theorems 2–4.

##### A. Proof of Theorem 2

Before we turn to presenting the proof, we state a modification of Lemma 48 in Ref. [40]. Let us denote with  $\mathcal{N}_{\mathbb{C}}(\mu, \sigma)$  the normal distribution over the complex numbers with mean  $\mu$  and standard deviation  $\sigma$ . The lemma establishes that products of normal distributions with small mean are close to a product of the standard normal distribution with zero mean.

*Lemma 5 (Autocorrelation of Gaussian distributions).* For the distributions

$$\mathcal{D}_1 := \mathcal{N}_{\mathbb{R}}(0, (1 - \varepsilon)^2 \sigma)^M, \quad (5)$$

$$\mathcal{D}_2 := \prod_{i=1}^M \mathcal{N}_{\mathbb{R}}(v_i, \sigma) \quad (6)$$

with  $v \in \mathbb{C}^M$ , it holds that

$$\|\mathcal{D}_1 - \mathcal{N}_{\mathbb{R}}(0, \sigma)^M\| \leq 2M\varepsilon, \quad (7)$$

$$\|\mathcal{D}_2 - \mathcal{N}_{\mathbb{R}}(0, \sigma)^M\| \leq \frac{1}{\sigma} \|v\|_1, \quad (8)$$

where  $\|\cdot\|$  denotes the total variation distance and  $v \in \mathbb{C}^M$ . The same result holds if we substitute  $\mathcal{N}$  with  $\mathcal{U}$ .

*Proof of Lemma 5.* We prove the lemma for the Gaussian case. The uniform can be obtained similarly. We obtain with the triangle inequality for the total variation distance:

$$\|\mathcal{D}_1 - \mathcal{G}^M\| \leq M \|\mathcal{N}_{\mathbb{R}}(0, (1 - \varepsilon)^2 \sigma) - \mathcal{N}_{\mathbb{R}}(0, \sigma)\|. \quad (9)$$

With the relation between total variation distance and  $L^1$  norm, we obtain

$$\begin{aligned} \|\mathcal{D}_1 - \mathcal{G}^M\| &\leq \frac{M}{2} \int_{-\infty}^{\infty} \left| \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sqrt{2\pi\sigma(1-\varepsilon)}} e^{-\frac{x^2}{2\sigma^2(1-\varepsilon)}} \right| dx \\ &= \frac{M}{2\sqrt{2\pi\sigma(1-\varepsilon)}} \int_{-\infty}^{\infty} \left| (1-\varepsilon)e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{x^2}{2\sigma^2(1-\varepsilon)}} \right| dx \\ &\leq \frac{M\varepsilon}{2\sqrt{2\pi\sigma(1-\varepsilon)}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \\ &\quad + \frac{M}{2\sqrt{2\pi\sigma(1-\varepsilon)}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{x^2}{2\sigma^2(1-\varepsilon)}} dx \\ &= \frac{M\varepsilon}{2(1-\varepsilon)} + \frac{M}{2(1-\varepsilon)} - \frac{M}{2} = \frac{M\varepsilon}{1-\varepsilon} \leq 2M\varepsilon. \end{aligned} \quad (10)$$

The second inequality follows using again the triangle inequality:

$$\begin{aligned} \|\mathcal{D}_2 - \mathcal{G}^M\| &\leq \sum_{i=1}^M \|\mathcal{N}_{\mathbb{R}}(v_i, \sigma) - \mathcal{N}_{\mathbb{R}}(0, \sigma)\| \\ &= \sum_{i=1}^M \frac{1}{2} \int_{-\infty}^{\infty} \left| \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-v_i)^2}{2\sigma^2}} - \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} \right| dx \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^M \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} \left| e^{-\frac{(x-v_i)^2}{2\sigma^2}} - e^{-\frac{x^2}{2\sigma^2}} \right| dx \\ &\leq \sum_{i=1}^M \frac{|v_i|}{\sigma} = \frac{\|v\|_1}{\sigma}, \end{aligned} \quad (11)$$

where the last inequality follows from a straightforward calculation. ■

*Proof of Theorem 2.* For simplicity, we set  $\sigma = 1$ . Furthermore, we restrict to the case of Problem 1 as the proof for the case of Problem 2 is completely analogous. Consider Problem 1 and a hard instance defined by the data  $(P^{[v]})_v$ , e.g., the encoding of a Boolean function as was done in Ref. [32]. It suffices to consider a  $(P^{[v]})_v$  with all matrix entries being bounded by 1 as all instances constructed in Ref. [32] admit this form. Furthermore, we draw PEPS data from the standard Gaussian distribution entrywise, denoted as  $(Q^{[v]})_v \sim \mathcal{G}^{D^4 dN}$ . Analogously to Lipton [81], we define

$$(R(t)^{[v]})_v := t(P^{[v]})_v + (1-t)(Q^{[v]})_v. \quad (12)$$

Now, let  $|\psi(t)\rangle$  denote the PEPS corresponding to these data. In analogy to the discussion of the permanent, we define the function  $q(t) := \langle \psi(t) | \psi(t) \rangle$ . Notice that this function is a polynomial in  $t$  with degree  $r = 2N$ , which scales polynomially in the input length. Before we can apply Theorem 8, we have to deal with the fact that the  $(R(t)^{[v]})_v$  are not distributed according to the Gaussian distribution. We will need only very small  $t$  bounded by some  $\varepsilon > 0$ , such that the difference between the respective distributions is immaterial. Specifically, the  $(R(t)^{[v]})_v$  tensors are distributed according to

$$\mathcal{D} = \prod_{i=1}^{D^4 dN} \mathcal{N}_{\mathbb{C}}(t p_i, (1-t)^2). \quad (13)$$

Thus, from a triangle inequality and Lemma 5, we obtain

$$\|\mathcal{D} - \mathcal{G}^{D^4 dN}\| \leq (4D^4 dN + 2D^4 dN)\varepsilon = (6D^4 dN)\varepsilon \quad (14)$$

for  $|t| \leq \varepsilon$ , by identifying  $\mathbb{C}$  with  $\mathbb{R}^2$ . It will suffice to set

$$\varepsilon := \frac{\delta}{6D^4 dN} \quad (15)$$

and  $\delta := \frac{1}{12N}$ . This implies that for a small enough inverse polynomial  $\varepsilon$ , we can make the total variation distance polynomially small. Let  $\{t_i\}_{i \in [r+1]}$  be the set of  $r+1$  equidistant points in  $[0, \varepsilon]$ . We will now use the assumption from the theorem's statement that the machine  $\mathcal{O}$  works for a  $1 - \delta$  fraction of the instances drawn from  $\mathcal{G}^{D^4 dN}$ . Using (14), we obtain for the success probability of the machine evaluating at the points  $t_i$  accurately up to within precision  $2^{-\text{poly}N}$

$$\begin{aligned} \Pr[\mathcal{O}((R^{[v]})_v(t_i)) - q(t_i)] &\leq 2^{-\text{poly}N} \\ &\geq 1 - \delta - \|\mathcal{D} - \mathcal{G}^{D^4 dN}\| \\ &\geq 1 - 2\delta, \end{aligned} \quad (16)$$

where we used that the total variation distance is an upper bound on the difference in probability the two distributions could possibly assign to an event.

Finally, we obtain the probability of  $r + 1$  consecutive successful evaluations as

$$\begin{aligned} \Pr[\{|i \in [r + 1], |\mathcal{O}(t_i) - q(t_i)| \leq 2^{-\text{poly}N}\}| = r + 1] \\ = (1 - 2\delta)^{r+1} = \left(1 - \frac{1}{6N}\right)^{r+1} \\ \geq 1 - \frac{2N + 1}{6N} = \frac{2}{3} - \frac{1}{6N}, \end{aligned} \quad (17)$$

by Bernoulli's inequality. Here, we abbreviated  $\mathcal{O}((R^{[v]})_v(t_i))$  with  $\mathcal{O}(t_i)$ . Given the evaluation values at the  $t_i$ , we can solve for the coefficients and obtain a polynomial  $\tilde{q}$  which satisfies  $|\tilde{q}(t_i) - q(t_i)| \leq 2^{-\text{poly}N}$  for all  $t_i$  with high probability. The machine  $\mathcal{O}'$  then evaluates  $\tilde{q}(1)$ , which is an estimate for  $q(1) = \langle \psi | \psi \rangle$ .

To bound the error on this estimate we will use two powerful results: The first on noisy extrapolations and the second on noisy interpolations of polynomials. A version of the following lemma was proven in Ref. [83], see also Sec. 9.1 in Ref. [40].

*Lemma 6 (Paturi).* Let  $p : \mathbb{R} \rightarrow \mathbb{R}$  be a polynomial of degree  $r$  and suppose  $|p(x)| \leq \Delta$  for all  $x$  such that  $|x| \leq \varepsilon$ . Then,  $|p(1)| \leq \Delta e^{2r(1+1/\varepsilon)}$ .

The following result was proven in Rakhmanov [84].

*Theorem 7 (Rakhmanov).* Let  $E_k$  denote the set of  $k$  equidistant points in  $(-1, 1)$ . Then, for a polynomial  $p : \mathbb{R} \rightarrow \mathbb{R}$  with degree  $r$  such that  $|p(y)| \leq 1$  for all  $y \in E_k$ , it holds that

$$|p(x)| \leq C \log \left( \frac{\pi}{\arctan \left( \frac{k}{r} \sqrt{\mathcal{R}^2 - x^2} \right)} \right) \quad (18)$$

with

$$|x| \leq \mathcal{R} := \sqrt{1 - \frac{r^2}{k^2}}. \quad (19)$$

We will use the second result to bound the error between the points and then use the first result to bound the error on  $\tilde{q}(1)$ . For the proof, we shift the polynomial  $p$  such that the interval of interest is centered around the origin. Furthermore, we can straightforwardly implement that we work with a smaller interval. We obtain that

$$\mathcal{R} = \sqrt{1 - \frac{r^2}{(r+1)^2} \frac{\varepsilon}{2}} = \sqrt{\frac{4N+1}{(2N+1)^2} \frac{\varepsilon}{2}}. \quad (20)$$

Restricting to the strict subinterval  $[-\frac{\mathcal{R}}{2}, \frac{\mathcal{R}}{2}]$ , we can apply Theorem 7 and obtain the following bound for all  $t \in [-\frac{\mathcal{R}}{2}, \frac{\mathcal{R}}{2}]$ ,

$$\begin{aligned} |p(t)| &\leq 2^{-\text{poly}N} C \ln \left( \frac{\pi}{\arctan \left( \frac{k}{r} \sqrt{\mathcal{R}^2 - x^2} \right)} \right) \\ &\leq 2^{-\text{poly}N} C \ln \left( \frac{\pi}{\arctan(2\mathcal{R})} \right) \leq 2^{-\frac{1}{2}\text{poly}N}. \end{aligned} \quad (21)$$

Finally, we can apply Lemma 6. This yields the desired bound on the difference between the estimate  $\tilde{q}(1)$  and the actual value  $q(1)$ :

$$\begin{aligned} |\tilde{q}(1) - q(1)| &= |p(1)| \leq 2^{-\frac{1}{2}\text{poly}N + 4 \log_2(e)N(1+2/\mathcal{R})} \\ &= 2^{-\text{poly}N} \end{aligned} \quad (22)$$

for a sufficiently large  $\text{poly}$ . Finally, we remark that the success probability can be exponentially amplified by repeating the above procedure polynomially many times because of the Chernoff bound. ■

## B. Proof of Theorem 3

The superior bound in Theorem 3 follows from the fact that we can invoke the Berlekamp-Welch algorithm in the interpolation step. The latter is a provably correct algorithm for the interpolation of polynomials due to Ref. [85]. Compare also Bouland *et al.* [41].

*Theorem 8 (Berlekamp-Welch [85]).* Let  $q$  be a degree- $r$  polynomial over any field  $\mathbb{F}$ . Suppose we are given  $k$  pairs of elements  $\{(x_i, y_i)\}_{i=1, \dots, k}$  with all  $x_i$  distinct with the promise that  $y_i = q(x_i)$  for at least  $\max(r + 1, (k + r)/2)$  points. Then, one can recover  $q$  exactly in  $\text{poly}(k, r)$  deterministic time.

As explained in Sec. III C, we arrive at a polynomial  $q(t) = \langle \psi(t) | \psi(t) \rangle$  of degree  $r = 2N$ . Instead of  $r + 1$  queries to the machine  $\mathcal{O}$ , we query it  $k = \text{poly}(N)$  times. Berlekamp-Welch requires that at least  $\frac{k+r}{2}$  of obtained  $k$  data points are correct in order to reconstruct the polynomial. We furthermore assume that  $k > r$ . From Markov's inequality and the union bound, we obtain

$$\begin{aligned} \Pr \left[ |\{i, \mathcal{O}(t_i) = q(t_i)\}| \geq \frac{k+r}{2} \right] &\geq 1 - \frac{2\mathbb{E}}{k-r} \\ &\geq 1 - \frac{2\left(\frac{1}{4} - \frac{1}{\text{poly}N}\right)k}{k-r} = 1 - \frac{k}{2(k-r)} + \frac{2k}{\text{poly}(N)(k-r)} \\ &= \frac{1}{2} - \frac{r}{2(k-r)} + \frac{2k}{\text{poly}(N)(k-r)}, \end{aligned} \quad (23)$$

where we abbreviate the expectation value in question with  $\mathbb{E}$ . Thus, by choosing  $k$  polynomially large, we obtain an expression that is polynomially close to  $1/2$ . Again, by repeating the procedure a polynomial number of times and taking a majority vote we can amplify this probability exponentially. With this probability, the Berlekamp-Welch algorithm outputs  $q$  exactly and we can simply evaluate  $q(1)$  without having to worry about the error of extrapolation. It seems appropriate to point out that we are in fact not drawing data from the Gaussian distribution in this case but from a discrete analog of it. However, this does not change the details of our analysis.

## C. Proof of Theorem 4

We know that the function we are interested in can be described by the quotient of two polynomials of degree at most  $r = 2N$ . This leaves us with  $4N + 1$  unknown coefficients. There is an equivalent of the Berlekamp-Welch algorithm for rational functions [86]. Invoking this algorithm, the proof proceeds analogously to the proof of Theorem 3.

## V. EXPONENTIAL DEPENDENCE ON PEPS DATA

The argument in the main text emphasizes the demanding precision that is required when specifying the PEPS data. In this section, we stress that this is not merely done for complexity-theoretic reasons: A pair of states can be defined by very similar PEPS data, while their norms can be vastly

different. In fact, to specify the norm of a PEPS, one needs exponential precision in the PEPS data, as a moment of thought reveals. This is already true in one spatial dimension for matrix product states. Take  $D = 2$ ,  $d = 2$ , an a translation-invariant open boundary condition MPS, so that the vertex set  $V$  is that of  $N$  sites,  $E$  reflecting nearest neighbor interactions. The linear operators  $P^{[v]} = P$  are for all  $v$  defined by

$$P^{[v]} = \sum_{i=1,2} \sum_{\alpha,\beta=1}^D A[i]_{\alpha,\beta}|i\rangle\langle\alpha, \beta|, \quad (24)$$

where for the state vector  $|\psi\rangle$  we take

$$A[0] := \text{diag}(1, 0), \quad A[1] := \text{diag}(0, 1). \quad (25)$$

The boundary conditions are taken open, as in the main text, and fixed by vector  $|0\rangle$  and the respective dual. Obviously, this is a representation of the product  $|0, \dots, 0\rangle$  with norm  $\langle\psi|\psi\rangle = 1$ . For  $|\phi\rangle$ , we choose

$$B[0] := \text{diag}(1, 0), \quad B[1] := \text{diag}(\eta, 1), \quad (26)$$

with the same boundary conditions, for some  $\eta > 0$ . It is still straightforward to compute the norm, invoking the transfer operator

$$\mathbb{E} := B[0] \otimes B^*[0] + B[1] \otimes B^*[1] = \text{diag}(1 + \eta^2, \eta, \eta, 1). \quad (27)$$

This gives

$$\langle\phi|\phi\rangle = \langle 0|\mathbb{E}^N|0\rangle = (1 + \eta^2)^N. \quad (28)$$

Clearly, for the two states to feature norms that are the same up to a constant, an in  $N$  exponentially small  $\eta > 0$  is required. In fact, even for a bond dimension  $D = 1$  one could have come to a similar conclusion. However,  $|\psi\rangle$  and  $|\phi\rangle$  are even vastly different in their entanglement properties, the latter featuring an entanglement entropy of a symmetrically bisected chain that is extensive in  $N$ .

## VI. OUTLOOK

In this work, we presented the first average-case complexity result in the context of quantum many-body systems, specifically tensor network states. Our main result is structural, namely we prove that the hard instances of PEPS contraction make up a significant fraction of all instances. Physically, this means that contraction of PEPS with random tensors is likely to be computationally hard to accurately evaluate. Conceptually, we establish structural similarities to the evaluation of the permanent. Our results hold under the assumption of accurate or exponential precision. In Sec. V, we stress that also on physical grounds, to demand exponential precision is very much reasonable. However, in a physical context it is often sufficient to evaluate observables up to polynomial precision. The major *open problem* is thus to extend the presented analysis to this case. For PEPS contractions establishing such a result would have direct practical implications. Furthermore, we are not aware of any #P-completeness result for translation-invariant PEPS. Thus the general *open question* should be: what are the instances of PEPS for which known contraction methods have convergence guarantees? It is our hope that further research at the interface between computer science and quantum many-body physics will provide exciting insights to this question.

## ACKNOWLEDGMENTS

We thank Adam Bouland, Bill Fefferman, Juani Bermejo-Vega, Christopher Chubb, Ingo Roth, Alexander Nietner, Augustine Kshetrimayum, Frederik Hahn, and Radu Curticapean for valuable discussions and comments. In particular we are grateful to Adam Bouland and Bill Fefferman for pointing out to us the result of Rakhmanov [84]. This work was supported by the DFG (EI 519/14-1, EI 519/7-1, EI 519/15-1, and CRC 183 B1), the ERC (TAQ), and the Templeton Foundation.

- 
- [1] A. Altland and B. D. Simons, *Condensed Matter Field Theory* (Cambridge University Press, Cambridge, UK, 2010).
  - [2] Y. V. Nazarov and Y. M. Blanter, *Quantum Transport: Introduction to Nanoscience* (Cambridge University Press, Cambridge, UK, 2009).
  - [3] A. Acin, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm, The European quantum technologies roadmap, *New J. Phys.* **20**, 080201 (2018).
  - [4] F. Weber, Strange quark matter and compact stars, *Prog. Part. Nucl. Phys.* **54**, 193 (2005).
  - [5] I. Arsene, I. Bearden, D. Beavis, C. Besliu, B. Budick, H. Bøggild, C. Chasman, C. Christensen, P. Christiansen, J. Cibor *et al.*, Quark-gluon plasma and color glass condensate at RHIC? The perspective from the BRAHMS experiment, *Nucl. Phys. A* **757**, 1 (2005).
  - [6] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
  - [7] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, *Rev. Mod. Phys.* **83**, 349 (2011).
  - [8] W. Metzner, M. Salmhofer, C. Honerkamp, V. Meden, and K. Schönhammer, Functional renormalization group approach to correlated fermion systems, *Rev. Mod. Phys.* **84**, 299 (2012).
  - [9] U. Schollwöck, The density-matrix renormalization group, *Rev. Mod. Phys.* **77**, 259 (2005).
  - [10] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Phys.* **326**, 96 (2011).
  - [11] F. Alet and N. Laflorencie, Many-body localization: An introduction and selected topics, *C. R. Phys.* **19**, 498 (2018).
  - [12] T. DeGrand and C. DeTar, *Lattice Methods for Quantum Chromodynamics* (World Scientific, Singapore, 2006).
  - [13] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, *Commun. Math. Phys.* **144**, 443 (1992).
  - [14] F. Verstraete, V. Murg, and J. I. Cirac, Matrix product states, projected entangled pair states, and variational renormalization

- group methods for quantum spin systems, *Adv. Phys.* **57**, 143 (2008).
- [15] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, *Ann. Phys.* **349**, 117 (2014).
- [16] J. C. Bridgeman and C. T. Chubb, Hand-waving and interpretive dance: an introductory course on tensor networks, *J. Phys. A* **50**, 223001 (2017).
- [17] J. Eisert, M. Cramer, and M. B. Plenio, Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (2010).
- [18] F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac, Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems, *Phys. Rev. Lett.* **93**, 207204 (2004).
- [19] D. Gobert, C. Kollath, U. Schollwöck, and G. Schütz, Real-time dynamics in spin-1/2 chains with adaptive time-dependent density matrix renormalization group, *Phys. Rev. E* **71**, 036102 (2005).
- [20] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, *Phys. Rev. B* **73**, 094423 (2006).
- [21] S. R. Manmana, S. Wessel, R. M. Noack, and A. Muramatsu, Strongly Correlated Fermions After a Quantum Quench, *Phys. Rev. Lett.* **98**, 210405 (2007).
- [22] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas, *Nat. Phys.* **8**, 325 (2012).
- [23] T. Fukuhara, A. Kantian, M. Endres, M. Cheneau, P. Schauß, S. Hild, D. Bellem, U. Schollwöck, T. Giamarchi, C. Gross *et al.*, Quantum dynamics of a mobile spin impurity, *Nat. Phys.* **9**, 235 (2013).
- [24] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Entanglement spectrum of a topological phase in one dimension, *Phys. Rev. B* **81**, 064439 (2010).
- [25] J. H. Bardarson, F. Pollmann, and J. E. Moore, Unbounded Growth of Entanglement in Models of Many-Body Localization, *Phys. Rev. Lett.* **109**, 017202 (2012).
- [26] C. Karrasch, J. H. Bardarson, and J. E. Moore, Finite-Temperature Dynamical Density Matrix Renormalization Group and the Drude Weight of Spin-1/2 Chains, *Phys. Rev. Lett.* **108**, 227206 (2012).
- [27] T. Prosen, Open *XXZ* Spin Chain: Nonequilibrium Steady State and a Strict Bound on Ballistic Transport, *Phys. Rev. Lett.* **106**, 217206 (2011).
- [28] J. P. Ronzheimer, M. Schreiber, S. Braun, S. S. Hodgman, S. Langer, I. P. McCulloch, F. Heidrich-Meisner, I. Bloch, and U. Schneider, Expansion Dynamics of Interacting Bosons in Homogeneous Lattices in One and Two Dimensions, *Phys. Rev. Lett.* **110**, 205301 (2013).
- [29] I. Arad, Z. Landau, U. Vazirani, and T. Vidick, Rigorous RG algorithms and area laws for low energy eigenstates in 1D, *Commun. Math. Phys.* **356**, 65 (2017).
- [30] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, *Quantum Inf. Comput.* **7**, 401 (2007).
- [31] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, Criticality, the Area Law, and the Computational Power of Projected Entangled Pair States, *Phys. Rev. Lett.* **96**, 220601 (2006).
- [32] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, The Computational Complexity of PEPS, *Phys. Rev. Lett.* **98**, 140506 (2007).
- [33] S. Arora and B. Barak, *Computational Complexity: A Modern Approach* (Cambridge University Press, Cambridge, UK, 2009).
- [34] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2002).
- [35] S. Aaronson, *Quantum Computing Since Democritus* (Cambridge University Press, Cambridge, UK, 2013).
- [36] D. Lokshantov, D. Marx, and S. Saurabh, Lower bounds based on the exponential time hypothesis, *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **105**, 41 (2011).
- [37] D. Deutsch, Quantum theory, the Church–Turing principle and the universal quantum computer, *Proc. R. Soc. London A* **400**, 97 (1985).
- [38] F. Barahona, On the computational complexity of Ising spin glass models, *J. Phys. A* **15**, 3241 (1982).
- [39] J. Feigenbaum and L. Fortnow, On the random-self-reducibility of complete sets, in *Proceedings of the Sixth Annual Structure in Complexity Theory Conference* (IEEE, Piscataway, NJ, 1991).
- [40] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, *Theor. Comput. Sci.* **9**, 143 (2013).
- [41] A. Bouland, B. Feffermann, C. Nirkhe, and U. Vazirani, Quantum supremacy and the complexity of random circuit sampling, *Nat. Phys.* **15**, 159 (2019).
- [42] Y.-Y. Shi, L.-M. Duan, and G. Vidal, Classical simulation of quantum many-body systems with a tree tensor network, *Phys. Rev. A* **74**, 022320 (2006).
- [43] A. Anshu, I. Arad, and A. Jain, How local is the information in tensor networks of matrix product states or projected entangled pairs states, *Phys. Rev. B* **94**, 195143 (2016).
- [44] M. Schwarz, O. Buerschaper, and J. Eisert, Approximating local observables on projected entangled pair states, *Phys. Rev. A* **95**, 060102(R) (2017).
- [45] M. Lubasch, J. I. Cirac, and M.-C. Banuls, Unifying projected entangled pair states contractions, *New J. Phys.* **16**, 033014 (2014).
- [46] M. Levin and C. P. Nave, Tensor Renormalization Group Approach to Two-Dimensional Classical Lattice Models, *Phys. Rev. Lett.* **99**, 120601 (2007).
- [47] Z.-C. Gu and X.-G. Wen, Tensor-entanglement-filtering renormalization approach and symmetry-protected topological order, *Phys. Rev. B* **80**, 155131 (2009).
- [48] Z. Y. Xie, H. C. Jiang, Q. N. Chen, Z. Y. Weng, and T. Xiang, Second Renormalization of Tensor-Network States, *Phys. Rev. Lett.* **103**, 160601 (2009).
- [49] H. H. Zhao, Z. Y. Xie, Q. N. Chen, Z. C. Wei, J. W. Cai, and T. Xiang, Renormalization of tensor-network states, *Phys. Rev. B* **81**, 174411 (2010).
- [50] S. Yang, Z.-C. Gu, and X.-G. Wen, Loop Optimization for Tensor Network Renormalization, *Phys. Rev. Lett.* **118**, 110504 (2017).
- [51] G. Evenbly, Algorithms for tensor network renormalization, *Phys. Rev. B* **95**, 045117 (2017).
- [52] F. Verstraete and J. I. Cirac, Renormalization algorithms for quantum-many body systems in two and higher dimensions, [arXiv:cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066).


- [53] R. Orús and G. Vidal, Infinite time-evolving block decimation algorithm beyond unitary evolution, *Phys. Rev. B* **78**, 155117 (2008).
- [54] L. Vanderstraeten, J. Haegeman, P. Corboz, and F. Verstraete, Gradient methods for variational optimization of projected entangled-pair states, *Phys. Rev. B* **94**, 155123 (2016).
- [55] M. T. Fishman, L. Vanderstraeten, V. Zauner-Stauber, J. Haegeman, and F. Verstraete, Faster methods for contracting infinite 2D tensor networks, *Phys. Rev. B* **98**, 235148 (2018).
- [56] T. Nishino, Density matrix renormalization group method for 2D classical models, *J. Phys. Soc. Jpn.* **64**, 3598 (1995).
- [57] T. Nishino, K. Okunishi, and M. Kikuchi, Numerical renormalization group at criticality, *Phys. Lett. A* **213**, 69 (1996).
- [58] T. Nishino and K. Okunishi, Corner transfer matrix algorithm for classical renormalization group, *J. Phys. Soc. Jpn.* **66**, 3040 (1997).
- [59] R. Orús and G. Vidal, Simulation of two-dimensional quantum systems on an infinite lattice revisited: Corner transfer matrix for tensor contraction, *Phys. Rev. B* **80**, 094403 (2009).
- [60] R. Orús, Exploring corner transfer matrices and corner tensors for the classical simulation of quantum lattice systems, *Phys. Rev. B* **85**, 205117 (2012).
- [61] A. Kshetrimayum, M. Rizzi, J. Eisert, and R. Orús, A Tensor Network Annealing Algorithm for Two-Dimensional Thermal States, *Phys. Rev. Lett.* **122**, 070502 (2019).
- [62] I. Markov and Y. Shi, Simulating quantum computation by contracting tensor networks, *SIAM J. Comp.* **38**, 963 (2008).
- [63] I. Arad and Z. Landau, Quantum computation and the evaluation of tensor networks, *SIAM J. Comp.* **39**, 3089 (2010).
- [64] S.-J. Ran, E. Tarrico, C. Peng, X. Chen, L. Tagliacozzo, G. Su, and M. Lewenstein, *Tensor Network Contractions: Methods and Applications to Quantum Many-Body Systems*, Lecture Notes in Physics, Vol. 964 (Springer, 2000).
- [65] P. Corboz, R. Orús, B. Bauer, and G. Vidal, Simulation of strongly correlated fermions in two spatial dimensions with fermionic projected entangled-pair states, *Phys. Rev. B* **81**, 165104 (2010).
- [66] P. Corboz, S. R. White, G. Vidal, and M. Troyer, Stripes in the two-dimensional  $t$ - $J$  model with infinite projected entangled-pair states, *Phys. Rev. B* **84**, 041108(R) (2011).
- [67] I. Niesen and P. Corboz, A tensor network study of the complete ground state phase diagram of the spin-1 bilinear-biquadratic Heisenberg model on the square lattice, *SciPost Phys.* **3**, 030 (2017).
- [68] Y. H. Matsuda, N. Abe, S. Takeyama, H. Kageyama, P. Corboz, A. Honecker, S. R. Manmana, G. R. Foltin, K. P. Schmidt, and F. Mila, Magnetization of  $\text{SrCu}_2(\text{BO}_3)_2$  in Ultrahigh Magnetic Fields up to 118 T, *Phys. Rev. Lett.* **111**, 137204 (2013).
- [69] A. Kshetrimayum, T. Picot, R. Orús, and D. Poilblanc, Spin-1/2 kagome XXZ model in a field: Competition between lattice nematic and solid orders, *Phys. Rev. B* **94**, 235146 (2016).
- [70] S. Yan, D. A. Huse, and S. R. White, Spin-liquid ground state of the  $S = 1/2$  Kagome Heisenberg antiferromagnet, *Science* **332**, 1173 (2011).
- [71] D. Poilblanc, J. I. Cirac, and N. Schuch, Chiral topological spin liquids with projected entangled pair states, *Phys. Rev. B* **91**, 224431 (2015).
- [72] T. Picot, M. Ziegler, R. Orús, and D. Poilblanc, Spin- $S$  kagome quantum antiferromagnets in a field with tensor networks, *Phys. Rev. B* **93**, 060407(R) (2016).
- [73] H. J. Liao, Z. Y. Xie, J. Chen, Z. Y. Liu, H. D. Xie, R. Z. Huang, B. Normand, and T. Xiang, Gapless Spin-Liquid Ground State in the  $S = 1/2$  Kagome Antiferromagnet, *Phys. Rev. Lett.* **118**, 137202 (2017).
- [74] J.-Y. Chen, L. Vanderstraeten, S. Capponi, and D. Poilblanc, Non-Abelian chiral spin liquid in a quantum antiferromagnet revealed by an iPEPS study, *Phys. Rev. B* **98**, 184409 (2018).
- [75] T. B. Wahl, A. Pal, and S. H. Simon, Efficient Representation of Fully Many-Body Localized Systems Using Tensor Networks, *Phys. Rev. X* **7**, 021018 (2017).
- [76] S. Bachmann, S. Michalakis, B. Nachtergaele, and R. Sims, Automorphic equivalence within gapped phases of quantum lattice systems, *Commun. Math. Phys.* **309**, 835 (2012).
- [77] M. B. Hastings, Quasi-adiabatic continuation for disordered systems: Applications to correlations, Lieb-Schultz-Mattis, and Hall conductance, [arXiv:1001.5280](https://arxiv.org/abs/1001.5280).
- [78] R. Raussendorf and H. J. Briegel, A One-Way Quantum Computer, *Phys. Rev. Lett.* **86**, 5188 (2001).
- [79] R. Raussendorf, D. E. Browne, and H. J. Briegel, Measurement-based quantum computation with cluster states, *Phys. Rev. A* **68**, 022312 (2003).
- [80] D. Gross and J. Eisert, Novel Schemes for Measurement-Based Quantum Computation, *Phys. Rev. Lett.* **98**, 220503 (2007).
- [81] R. Lipton, *New Directions in Testing: Distributed Computing and Cryptography* (American Mathematical Society, Providence, RI, 1991).
- [82] L. G. Valiant, Quantum circuits that can be simulated classically in polynomial time, *SIAM Comput.* **31**, 1229 (1979).
- [83] R. Paturi, On the degree of polynomials that approximate symmetric Boolean functions, in *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 1992), pp. 468–474.
- [84] E. A. Rakhmanov, Bounds for polynomials with a unit discrete norm, *Ann. Math.* **165**, 55 (2007).
- [85] L. Welch and E. Berlekamp, Error correction for algebraic block codes, US Patent, 4,633,470 (1986).
- [86] R. Movassagh, Efficient unitary paths and quantum computational supremacy: A proof of average-case hardness of Random Circuit Sampling, [arXiv:1810.04681](https://arxiv.org/abs/1810.04681).



Random matrix theory has countless applications in condensed matter physics, allowing to predict a wide range of topics ranging from universal conductance fluctuations, weak localization or coherent back-scattering [ABDF11]. Closer to notions of quantum information theory, it has been shown that for uniformly drawn states, such ideas lead to a principle of maximum entropy [SR95, HLSWo4, HLWo6], an insight that has implications in the context of quantum computing [GFEo9]. However, from a physical perspective, such random states do not respect the local structure present in most naturally occurring systems. What is more common and natural, in contrast, are quantum states that emerge from ground states of gapped Hamiltonians. Indeed, in common experiments, good approximations of ground states of local Hamiltonian characterized by finite-ranged interactions can often be feasibly prepared by means of cooling procedures or by resorting to suitable loading procedures. A key question that arises, therefore, is to what extent one can expect such ground states to exhibit the same or similar properties as (Haar-random) uniformly chosen ones. Of particular interest is the apparent emergence of properties that are usually assumed true in statistical mechanics from the isolated Hamiltonian dynamics of generic quantum states, for instance the tendency to relax to an equilibrium state and to exhibit maximum entropy. To tackle these questions, we use the fact that ground states of one dimensional gapped local Hamiltonian can be approximated arbitrarily well by Matrix Product States (MPS), i.e. states of the form

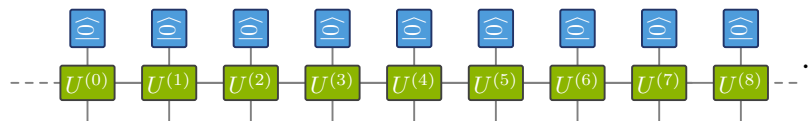
$$|\psi\rangle = \sum_{i_1, \dots, i_N} \text{Tr} \left[ A_{i_1}^{(1)} A_{i_2}^{(2)} \dots A_{i_n}^{(n)} \right] |i_1, \dots, i_N\rangle. \quad (5.1)$$

Here  $A_{i_j}^{(k)}$  is a  $D \times D$  matrix.  $D$  is referred to as the bond dimension. In graphical calculus (compare [BC17]):



$$\dots \text{---} \square \text{---} \square \text{---} \square \text{---} \square \text{---} \square \text{---} \square \text{---} \square \text{---} \square \text{---} \square \text{---} \dots \quad (5.2)$$

Many properties are known for random translationally invariant Matrix Product States including a principle of maximum entropy [GGJN18, CGGPG13] and extensivity of the Rényi 2-entropy with respect to equidistant disconnected subsystems [RW20]. In this work we focus on the properties of typical *disordered* Matrix Product States. More precisely, it is known that MPSs can be unitarily embedded [PGVWC07]: here, one equips each tensor with another leg and feeds in an arbitrary state vector  $|0\rangle \in \mathbb{C}^d$ , to get



$$\dots \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \begin{array}{c} \square \\ \square \end{array} \text{---} \dots \quad (5.3)$$

Each unitary  $U^{(0)}, \dots, U^{(n)} \in U(dD)$  can be seen as mapping one  $\mathbb{C}^d \otimes \mathbb{C}^D$  input system to another  $\mathbb{C}^d \otimes \mathbb{C}^D$  output system. A natural definition of typical (i.e. Haar random) disordered Matrix Product State is hence the following:

**Definition 9** (Random matrix product state). *A random matrix product state (RMPS) of local dimension  $d$ , system-size  $n$  and bond dimension  $D$  is a state defined by (5.3) with each unitary  $U^{(1)}, \dots, U^{(n)} \in U(dD)$  drawn i.i.d. randomly from the Haar measure. We denote the resulting measure as  $\mu_{d,n,D}$ .*

We first prove that states drawn from this measure equilibrate exponentially well with a probability exponentially close to 1 in the system-size  $n$  under unitary evolution with mild requirements on the Hamiltonian. More precisely, given an observable  $A$ , "equilibration" means that for the overwhelming majority of time (for example after a short equilibration time) the expectation value of  $A$  will take a value very close to its infinite time average, meaning that the quantity

$$\Delta A_\psi^\infty := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t |\langle \psi | A(t') | \psi \rangle - A_\psi^\infty|^2 dt' \quad (5.4)$$

measuring fluctuations from the infinite time average

$$A_\psi^\infty := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \langle \psi | A(t') | \psi \rangle dt' \quad (5.5)$$

should be small. We thus prove the following:

**Theorem 17** (Equilibration of RMPS). *Let  $H$  be a Hamiltonian with non-degenerate spectrum and non-degenerate spectral gap and  $A$  an observable evolving in time governed by  $H$ . Let  $|\psi\rangle$  be a RMPS drawn from  $\mu_{d,n,D}$  and  $|\psi'\rangle := |\psi\rangle / \sqrt{\langle \psi | \psi \rangle}$ . Then there are universal constants  $c_1, c_2$  such that for  $n$  sufficiently large the infinite time average  $\Delta A_\psi^\infty$ , (5.4) fulfills*

$$\Pr \left( \Delta A_\psi^\infty \leq e^{-c_1 \alpha(d,D)n} \right) \geq 1 - e^{-\Omega(n)}. \quad (5.6)$$

Our second result is once again motivated by the phenomenon of equilibration in quantum many-body physics and the endeavor to provide a rigorous foundation for it. It has been proven in Ref. [WGRE19] that systems equilibrate if their energy eigenstates have Rényi entropy that is extensive in the system-size  $n$ , which means that

$$S_2(\text{Tr}_A[|j\rangle\langle j|]) \geq g(j)n \quad (5.7)$$

for a sufficiently well-behaved function  $g$  and an *arbitrary* subsystem  $A$ . This property has been dubbed *entanglement ergodicity* [WGRE19]. Moreover, Ref. [RW20] shows that generic translation-invariant MPS have extensive Rényi entropy if one considers a bi-partition of the spin chain into the subsystem  $A$  that corresponds to every  $k$ th spin and the rest. That is, the entropy grows proportional to the boundary  $|\partial A|$ . Here, we prove such a result with explicit quantitative bounds for disordered RMPS with overwhelming probability.

**Theorem 18** (Extensivity of entanglement entropies). *Suppose  $n$  is divisible by  $k$  and  $A$  consists of every  $k$ -th qudit. Let  $\rho'_A$  be the normalized density matrix of a RMPS drawn from  $\mu_{d,n,D}$  reduced on  $A$ . Then,*

$$\Pr \left( S_2(\rho'_A) \geq \Omega\left(\frac{n}{k}\right) \right) \geq 1 - e^{-\Omega(n/k)}. \quad (5.8)$$

Finally, we show a principle of maximum entropy for typical disordered ground states of local gapped Hamiltonians. We show that the Rényi  $z$ -entropy of a small connected subchain of qudits close to maximal:

**Theorem 19** (Almost maximum entropy for reduced states). *Let  $A$  be a subset  $l$  of consecutive qudits, let  $\rho'_A$  be the normalized density matrix of a RMPS drawn from  $\mu_{d,n,D}$  reduced to  $A$ . Then for any real  $r$  and  $n \geq 2 \frac{\log D}{\log d} + l$  we have*

$$\Pr(\text{Tr}[\rho_A'^2] \geq \Omega(D^{-r}) + d^{-l}) \leq O\left(D^{-(2-r)}\right). \quad (5.9)$$

The proofs of the above results are based on a graphical calculus, as all of them are obtained via mappings from the combinatorial problem obtained from applying elementary Weingarten calculus (as introduced in Chapter 2). This mapping produces partition functions of a one-dimensional statistical mechanical model. In fact, the method we make use of is inspired by recent work on random quantum circuits [ZN19,HJ19]. We showcase how this mapping can be a powerful tool in a conceptually different context. We also make use of a generalization of the Cauchy-Schwarz inequality to tensor networks without self-contractions [KKEG19].


## Emergent Statistical Mechanics from Properties of Disordered Random Matrix Product States

Jonas Haferkamp,<sup>1,2,\*</sup> Christian Bertoni<sup>1</sup>, Ingo Roth<sup>1,3</sup> and Jens Eisert<sup>1,2</sup>

<sup>1</sup>*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin 14195, Germany*

<sup>2</sup>*Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin 14109, Germany*

<sup>3</sup>*Quantum Research Centre, Technology Innovation Institute (TII), Abu Dhabi*

 (Received 16 April 2021; revised 13 July 2021; accepted 24 August 2021; published 13 October 2021)

The study of generic properties of quantum states has led to an abundance of insightful results. A meaningful set of states that can be efficiently prepared in experiments are ground states of gapped local Hamiltonians, which are well approximated by matrix product states. In this work, we introduce a picture of generic states within the trivial phase of matter with respect to their nonequilibrium and entropic properties. We do so by rigorously exploring nontranslation-invariant matrix product states drawn from a local independent and identically distributed Haar measure. We arrive at these results by exploiting techniques for computing moments of random unitary matrices and by exploiting a mapping to partition functions of classical statistical models, a method that has led to valuable insights on local random quantum circuits. Specifically, we prove that such disordered random matrix product states equilibrate exponentially well with overwhelming probability under the time evolution of Hamiltonians featuring a nondegenerate spectrum. Moreover, we prove two results about the entanglement Rényi entropy: the entropy with respect to sufficiently disconnected subsystems is generically extensive in the system size, and for small connected systems, the entropy is almost maximal for sufficiently large bond dimensions.

DOI: [10.1103/PRXQuantum.2.040308](https://doi.org/10.1103/PRXQuantum.2.040308)

The application of random matrix theory to the study of interacting quantum many-body systems has proven to be a particularly fruitful endeavor, in fact, in various readings. This includes countless applications in condensed matter physics, allowing us to predict a wide range of topics ranging from universal conductance fluctuations, weak localization, or coherent backscattering [1]. In a mindset closer to notions of quantum information theory, it has been shown that, for uniformly drawn states, such ideas lead to a principle of maximum entropy [2–4], an insight that has implications in the context of quantum computing [5]. That said, from an operational perspective, quantum states that are uniformly random—in the sense that they are drawn from a global invariant measure—are not particularly natural in many contexts. From a physical perspective, such random states do not respect locality present in most naturally occurring systems. What is more common and natural, in contrast, are quantum states that emerge from ground states of gapped Hamiltonians. Indeed, in common experiments, often good approximations of ground

states of local Hamiltonian characterized by finite-ranged interactions can be feasibly prepared, by means of cooling procedures or by resorting to suitable loading procedures. A key question that arises, therefore, is to what extent one can expect such ground states to exhibit the same or similar properties as (Haar-random) uniformly chosen ones.

This question can be interpreted in several readings. A particularly important one from the perspective of out-of-equilibrium physics is, specifically, to what extent such states would eventually equilibrate in time under the evolution of general Hamiltonians [6–8]. Equilibration is an important concept in the foundations of quantum statistical mechanics and considerations of how apparent equilibrium states seem to emerge under closed system quantum dynamics. Equilibration refers to properties becoming apparently and effectively stationary, even though the entire system would remain to undergo unitary dynamics generated by some local Hamiltonian.

In this work, we consider such typical ground states from a fresh perspective. More precisely, we prove several concentration-type results for so-called *matrix product states* (MPSs), instances of tensor network states that approximate ground states of gapped one-dimensional quantum systems well. This class of states hence indeed captures those that can be obtained by cooling local Hamiltonians with a spectral gap. To pick such random states seems a most meaningful approach to respect natural restrictions of locality. Since one can readily see such

\*jonas.haferkamp@fu-berlin.de

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

states as being ground states of disordered parent Hamiltonians [9], they can be viewed as being typical representatives of one-dimensional quantum phases of matter. Indeed, while the significance of random ensembles of quantum states respecting locality has been appreciated early on [10], only recently, first steps have been taken towards a rigorous understanding of such ensembles, specifically concerning spectral properties and decays of correlations of generic MPSs [11,12].

Specifically, the powerful technical tool we bring in to this kind of study is a framework – seemingly out of context – related to mappings to partition functions of a one-dimensional statistical mechanical models. We also develop this picture further. Equipped with this technical tool, we prove with overwhelming probability that a randomly chosen MPS—drawn according to the independent and identically distributed (i.i.d.) Haar measure in the physical and the virtual dimensions of the tensor network state—equilibrates exponentially under the time evolution of Hamiltonians with nondegenerate spectrum. Moreover, as a second result, and also motivated by the equilibration of systems exhibiting *many-body localization* [13–15], we prove the extensivity of the Rényi-2 entropy with overwhelming probability with respect to equidistant disconnected subsystems. This result complements recent work obtained for translation-invariant MPSs [16]. A third result is an improved *principle of maximum entropy* for disordered random MPSs. More precisely, we show that the Rényi-2 entropy is almost maximal for small connected subsystems up to errors polynomially small in the bond dimension  $D$ . This again complements results for random translation-invariant MPSs [11,17]. Another complementary result is given in Refs. [18,19], where formulas for the asymptotic values of various quantum entropies are derived in the setting of ergodic Markov chains.

Again, this substantial progress in studying generic equilibrium and out-of-equilibrium states of matter are facilitated by a technical tool introduced to the context at hand. The proofs of the above results are based on a graphical calculus, as all of them are obtained via mappings to partition functions of a one-dimensional statistical mechanical model. These models can be obtained by an application of the *Weingarten calculus*. In fact, the method we make use of is inspired by recent work on *random quantum circuits*. Such random quantum circuits have recently prominently been discussed in the literature, both in the context of quantum computing where they are used to show a quantum advantage over classical algorithms [20,21], as well as in proxies for scrambling dynamics [22–29]. In that context, similar mappings have been exploited [26]. We showcase here, therefore, how this mapping can be a very much helpful tool in a conceptually very different context. We also make use of a generalization of the Cauchy-Schwarz inequality to tensor networks without self-contractions [30]. In the Appendix,

we argue that the low-entanglement structure of MPSs prevents them from having generic properties of uniform states in the sense that they will now form an approximate complex projective 2-design [31] in a meaningful sense. The upshot of this work is that, using a machinery of mappings to partition functions of a one-dimensional statistical mechanical model and the Weingarten calculus, a wealth of out-of-equilibrium and equilibrium properties of generic one-dimensional quantum phases of matter can be rigorously computed.

## I. SETTING

### A. Random matrix product states

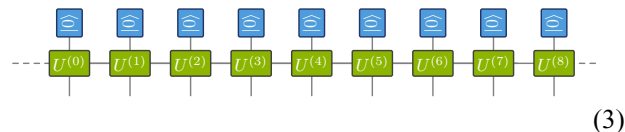
We start by stating the underlying model of random quantum states used throughout this work. A very reasonable model of random matrix product states already used in Ref. [10] is the following. Consider an arbitrary MPS vector [32,33] with periodic boundary conditions and  $n$  constituents of local dimension  $d$ . Such a state vector can be written as

$$|\psi\rangle = \sum_{i_1, \dots, i_n} \text{Tr}[A_{i_1}^{(1)} A_{i_2}^{(2)} \cdots A_{i_n}^{(n)}] |i_1, \dots, i_n\rangle. \quad (1)$$

Here, the  $A^{(i)}$  are complex valued  $D \times D$  matrices that specify the quantum state at hand, where  $D$  is referred to as the *bond dimension*, sometimes also called the tensor train rank. In a commonly used graphical calculus, this is represented as



We consider MPSs that can be unitarily embedded [34,35]. Here, one equips each tensor with another leg and feeds in an arbitrary state vector  $|0\rangle \in \mathbb{C}^d$  to get



Each unitary  $U^{(0)}, \dots, U^{(n)} \in U(dD)$  can be seen as mapping one  $\mathbb{C}^d \otimes \mathbb{C}^D$  input system to another  $\mathbb{C}^d \otimes \mathbb{C}^D$  output system. These can be normalized by choosing an appropriate normalization for the boundary vectors. We can construct MPSs with periodic boundary conditions analogously. This motivates a very natural probability measure.

**Definition 1** (Random matrix product state). *A random matrix product state (RMPS) of local dimension  $d$ , system size  $n$ , and bond dimension  $D$  is a state defined by Eq. (3)*

with each unitary  $U^{(1)}, \dots, U^{(n)} \in U(dD)$  drawn i.i.d. randomly from the Haar measure. We denote the resulting measure as  $\mu_{d,n,D}$ .

Note that this definition can be regarded as drawing the  $A^{(i)}$  tensor cores uniformly from the Stiefel manifold of isometries. This probability measure makes a lot of sense: it is a distribution over random disordered, nontranslation-invariant quantum states. Once again, each realization will have a *parent Hamiltonian* [9], a gapped local Hamiltonian for which the given state is an exact ground state. In this sense, the probability measure discussed here can equally be seen as a probability measure on disordered, random local Hamiltonians. It should be noted that this probability measure takes values on state vectors that are not exactly normalized. It is, however, easy to see that the values of  $\langle \psi | \psi \rangle$  are strongly concentrated around unity.

**Lemma 1** (Concentration around a unit norm). *It holds that*

$$\Pr(|\langle \psi | \psi \rangle - 1| \geq \varepsilon) \leq \varepsilon^{-2} d^{-n}. \quad (4)$$

This will be proven in Sec. V.

### B. Effective dimension and equilibration

We now turn to discussing concepts of equilibration in quantum many-body dynamics. Equilibration refers to the observation that, for an observable  $A$ , the expectation value of a time-evolving quantum many-body system will for the overwhelming times take the same values as the expectation value with respect to the infinite time average: the expectation value then looks “equilibrated” [7,36]. For a given arbitrary state vector  $|\psi\rangle$  that reflects the initial pure state at time  $T = 0$ , this expectation value of an observable  $A$  with respect to the infinite time average takes the form

$$A_{\psi}^{\infty} := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \langle \psi | A(t') | \psi \rangle dt'. \quad (5)$$

The fluctuations around this infinite time average are defined as [36]

$$\Delta A_{\psi}^{\infty} := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t |\langle \psi | A(t') | \psi \rangle - A_{\psi}^{\infty}|^2 dt'. \quad (6)$$

The reduced density matrices on a local region are determined by local observables. It hence suffices to consider the above definitions. In particular, if  $\Delta A_{\psi}^{\infty} \ll 1$  then the reduced density matrices on a small region cannot deviate from the infinite time average except for very brief time periods. In this sense a small value for  $\Delta A_{\psi}^{\infty}$  implies equilibration.

## II. EQUILIBRATION OF RANDOM MATRIX PRODUCT STATES

Equipped with these preparations, we are in the position to state our first main result. On the equilibration of RMPSs, following nonequilibrium dynamics, we prove the following theorem.

**Theorem 1** (Equilibration of RMPSs). *Let  $H$  be a Hamiltonian with nondegenerate spectrum and nondegenerate spectral gap, and let  $A$  be an observable evolving in time governed by  $H$ . Let  $|\psi\rangle$  be a RMPS drawn from  $\mu_{d,n,D}$  and  $|\psi'\rangle := |\psi\rangle / \sqrt{\langle \psi | \psi \rangle}$ . Then there are constants independent of the system parameters  $c_1, c_2$  such that, for sufficiently large  $n$ , the infinite time average  $\Delta A_{\psi}^{\infty}$ , Eq. (6), fulfills*

$$\Pr(\Delta A_{\psi}^{\infty} \leq e^{-c_1 \alpha(d,D)n}) \geq 1 - e^{-c_2 \alpha(d,D)n} \quad (7)$$

with

$$\alpha(d,D) = \log \left( \frac{d - 1/(dD^2)}{(1 + 1/D)[1 + 1/(dD)]} \right). \quad (8)$$

This shows that, under the time evolution of many Hamiltonians, almost all matrix product states equilibrate exponentially well. The proof of Theorem 1 implies a second result that we can informally state as follows. Assume that we draw Hamiltonians from an ensemble such that all marginal distributions for most eigenstates are distributed according to the RMPS measure introduced above. Then, for an arbitrary initial state, the system equilibrates exponentially well with overwhelming probability. This is motivated by the fact that the equilibration of systems exhibiting many-body localization still lacks a completely rigorous explanation. In particular, it is known that the energy eigenstates of many-body localized systems satisfy an *area law for the entanglement entropy* [14,37,38] and can be described by matrix product states [14]. It seems unrealistic that any natural ensemble of Hamiltonians has RMPSs as marginal eigenstates, but we believe that enough of this result might survive for more structured ensembles to prove equilibration of many-body localization systems.

### A. Proof techniques

In this section, we present many of the proof techniques relevant for this work. In order to obtain Theorem 1, we observe that Theorem 3 of Ref. [39] can be generalized to any distribution  $\nu$  on states provided that the effective dimension is large. The effective dimension measures how much overlap the initial state has with the energy eigenstates of the Hamiltonian. In particular, we have the following key result from Refs. [39–41] that we make use of.

**Lemma 2** ([39–41]). *Consider a Hamiltonian with nondegenerate spectrum and nondegenerate spectral gaps, i.e.,  $E_n - E_m = E_j - E_k$  if and only if  $n = j, m = k$ , where  $E_n$  labels the eigenvalues of the Hamiltonian. Then,*

$$\Delta A_\psi^\infty = O(1/D_{\text{eff}}) \tag{9}$$

with

$$1/D_{\text{eff}} := \sum_j |\langle \psi | j \rangle|^4, \tag{10}$$

where  $\{|j\rangle\}$  is the eigenbasis of the Hamiltonian  $H$ .

Theorem 1 follows immediately from Lemma 2 in combination with the following statement.

**Lemma 3** (Bound to effective dimension). *For all state vectors  $|\phi\rangle$ , we have*

$$\mathbb{E}_{\psi \sim \mu_{d,n,D}} |\langle \psi | \phi \rangle|^4 \leq 2 \frac{(1 + 1/D)^n [1 + 1/(dD)]^n}{(d^2 - 1/D^2)^n}. \tag{11}$$

We are trying to find an upper bound on

$$\mathbb{E} |\langle \psi | \phi \rangle|^4 = \langle \phi |^{\otimes 2} \mathbb{E} (|\psi\rangle \langle \psi|)^{\otimes 2} | \phi \rangle^{\otimes 2}. \tag{12}$$

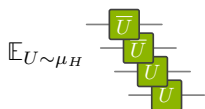
We make use of the Weingarten calculus [42,43], as elaborated upon here in the following statement.

**Lemma 4** (Weingarten calculus). *The  $t$ th moment operator of Haar-random unitaries is given by*

$$\mathbb{E}_{U \sim \mu_H} U^{\otimes t} \otimes \bar{U}^{\otimes t} = \sum_{\sigma, \pi \in S_t} \text{Wg}(\sigma^{-1}\pi, q) |\sigma\rangle \langle \pi|, \tag{13}$$

where  $|\sigma\rangle := (\mathbb{1} \otimes r(\sigma))|\Omega\rangle$  with  $r$  being the representation of the symmetric group  $S_t$  on  $(\mathbb{C}^q)^{\otimes t}$  that permutes the vectors in the tensor product, and  $|\Omega\rangle = \sum_{j=1}^q |j, j\rangle$  the maximally entangled state vector up to normalization.

The Weingarten calculus is a powerful tool in particular when suitably combined with Penrose tensor-network diagrams providing a graphical calculus; e.g., for  $t = 2$  and  $U(q)$ , Eq. (13) takes the form

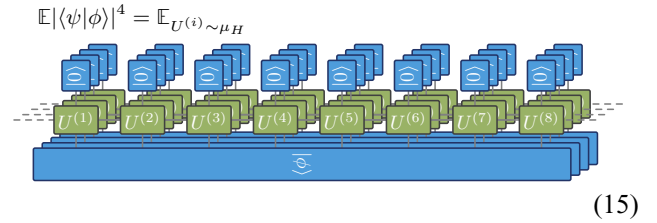


$$\mathbb{E}_{U \sim \mu_H}$$

$$= \frac{1}{q^2 - 1} \left[ \begin{array}{c} \left( \right) \left( -\frac{1}{q} \right) \left( -\frac{1}{q} \right) \left( + \right) \left( \right) \end{array} \right] \tag{14}$$

Graphically, we can express Eq. (12) as

$$\mathbb{E} |\langle \psi | \phi \rangle|^4 = \mathbb{E}_{U^{(i)} \sim \mu_H}$$



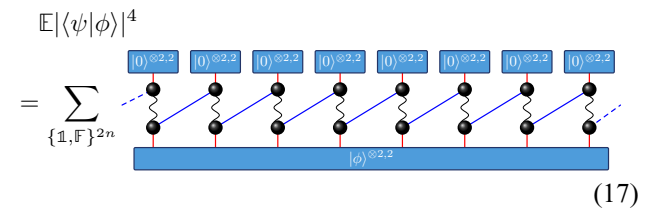
$$\tag{15}$$

By evaluating each  $\mathbb{E}_{U^{(i)}} U^{(i)} \otimes U^{(i)} \otimes \bar{U}^{(i)} \otimes \bar{U}^{(i)}$  individually according to Eq. (14), this can be reformulated as a partition function. Introducing the notation

$$|\psi\rangle^{\otimes 2,2} := |\psi\rangle^{\otimes 2} \otimes \overline{|\psi\rangle}^{\otimes 2}, \tag{16}$$

we obtain

$$\mathbb{E} |\langle \psi | \phi \rangle|^4 = \sum_{\{\mathbb{1}, \mathbb{F}\}^{2n}}$$



$$\tag{17}$$

Here, the black balls correspond to a choice of an element of  $S_2 = \{\mathbb{1}, \mathbb{F}\}$  with  $\mathbb{F}$  the swap permutation. The wiggly line corresponds to different weights for every pair of permutations  $(\pi, \sigma)$  with the corresponding value of the Weingarten function  $\text{Wg}(\pi^{-1}\sigma, q)$  according to Eq. (14) with  $q = dD$ . The red edges denote contractions over  $\mathbb{C}^d$  and the blue edges are contractions over  $\mathbb{C}^D$ . This is reminiscent of Ref. [26], where the frame potential of random quantum circuits is mapped to a partition function with local degrees of freedom corresponding to permutations. Note however that, for every permutation  $\pi \in S_2$ , we always have

$$\langle \pi | 0 \rangle^{\otimes 2,2} = |\langle 0 | 0 \rangle|^2 = 1. \tag{18}$$

Moreover, every summand contains a factor of the form  $\langle \phi |^{\otimes 2,2} \otimes_{i=1}^n |\sigma_i\rangle$ . We can bound this contribution using the following generalization of the Cauchy-Schwarz inequality [30].

**Lemma 5** (Cauchy-Schwarz inequality for tensor networks [30]). *Consider a tensor network  $(T, C)$  with  $J \geq 2$  tensors  $T = (\mathcal{t}^j)_{j \in \{1, \dots, J\}}$  such that no tensor in the contraction  $C$  self-contracts, i.e., no string connects a tensor with itself. Then,*

$$|C(T)| \leq \prod_{j=1}^J \|\mathcal{t}^j\|_F, \tag{19}$$

where  $\|\cdot\|_F$  is the Frobenius norm of the tensor  $\mathcal{t}^j$  viewed as a vector.

*Proof of Lemma 3.* As the tensor network contraction  $\langle \phi |^{\otimes 2,2} \bigotimes_{l=1}^n |\sigma_l\rangle$  does not contain self-contractions, this yields

$$\left| \langle \phi |^{\otimes 2,2} \bigotimes_{l=1}^n |\sigma_l\rangle \right| \leq \| |\phi\rangle \|_F^4 = 1. \quad (20)$$

Therefore, we can apply a triangle inequality to the sum in Eq. (17) to obtain the bound

$$\mathbb{E} |\langle \psi | \phi \rangle|^4 \leq \sum_{\{\mathbb{1}, \mathbb{F}\}^{2n}} \left| \begin{array}{c} \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \end{array} \right| \quad (21)$$

As contractions over  $\mathbb{C}^D$  we have  $\langle \mathbb{1} | \mathbb{1} \rangle = \langle \mathbb{F} | \mathbb{F} \rangle = D^2$  and

$$\langle \mathbb{1} | \mathbb{F} \rangle = \langle \mathbb{F} | \mathbb{1} \rangle = D. \quad (22)$$

This allows us to obtain a sufficient upper bound on  $1/D_{\text{eff}}$  via a combinatorial argument. Consider a sequence  $(\sigma_1 \pi_1), (\sigma_2 \pi_2), \dots, (\sigma_n \pi_n) \in \{\mathbb{1}, \mathbb{F}\}^{2n}$ . Here, the  $\sigma$  refer to the balls on top in Eq. (21) and  $\pi$  to the ones on the bottom. If  $\sigma_i = \pi_i$ , the total interaction between sites  $i$  and  $i + 1$  contributes with a term

$$\frac{D^2}{d^2 D^2 - 1}. \quad (23)$$

If  $\sigma_i \neq \pi_i$ , this contribution is divided by  $dD$ , and if  $\pi_i \neq \sigma_{i+1}$ , it is divided by  $D$ . Hence, we sum over all possible ways of choosing  $\sigma_i \neq \pi_i$  or  $\pi_i \neq \sigma_{i+1}$ , and divide by the corresponding factor, to get

$$\begin{aligned} \mathbb{E} |\langle \phi | \psi \rangle|^4 &\leq 2 \frac{D^{2n}}{(d^2 D^2 - 1)^n} \sum_{l=0}^n \sum_{r=0}^n \binom{n}{l} \binom{n}{r} D^{-l-r} d^{-r} \\ &= 2 \frac{(1 + 1/D)^n [1 + 1/(dD)]^n}{(d^2 - 1/D^2)^n}. \end{aligned} \quad (24)$$

This implies Lemma 3. ■

Theorem 1 now follows from applying Lemma 3 to Eq. (10) together with an application of Markov's inequality and Lemma 1.

*Proof of Theorem 1.* By Lemma 3 we have

$$\begin{aligned} \mathbb{E}(1/D_{\text{eff}}) &= \sum_j \mathbb{E}(|\langle \psi | j \rangle|^4) \\ &\leq 2 \frac{(1 + 1/D)^n [1 + 1/(dD)]^n}{[d - 1/(dD^2)]^n} \\ &= 2e^{-\alpha n} \end{aligned} \quad (25)$$

with  $\alpha = \alpha(d, D)$  as defined in Eq. (8). Picking two positive constants  $k_1, k_2 > 0$  with  $k_1 < 1/2$ , since  $\Delta A_{\psi}^{\infty} = O(1/D_{\text{eff}})$ , Markov's inequality yields

$$\Pr(\Delta A_{\psi}^{\infty} \leq e^{-k_1 \alpha n}) \geq 1 - e^{-k_2 \alpha n}. \quad (26)$$

Let  $N := \langle \psi | \psi \rangle$ , and let  $|\psi'\rangle = N^{-1/2} |\psi\rangle$  be the normalized state vector. We then have

$$\Delta A_{\psi'}^{\infty} = \frac{\Delta A_{\psi}^{\infty}}{N^2}. \quad (27)$$

Suppose that  $\Delta A_{\psi'}^{\infty} \leq e^{-k_1 \alpha n}$  and  $|N - 1| \leq e^{-k_1 \alpha n}$ . Then  $N^2 \geq (1 - e^{-k_1 \alpha n})^2$  and

$$\Delta A_{\psi'}^{\infty} \leq \frac{e^{-k_1 \alpha n}}{(1 - e^{-k_1 \alpha n})^2} \leq e^{-c_1 \alpha n} \quad (28)$$

for some constant  $c_1 > 0$  and the integer  $n$  being large enough. Then, by the union bound, we get

$$\begin{aligned} \Pr(\Delta A_{\psi'}^{\infty} \leq e^{-c_1 \alpha n}) &\geq 1 - \Pr(\Delta A_{\psi}^{\infty} \geq e^{-k_1 \alpha n} \text{ or } |N - 1| \geq e^{-k_1 \alpha n}) \\ &\geq \Pr(\Delta A_{\psi}^{\infty} \leq e^{-k_1 \alpha n}) - \Pr(|N - 1| \geq e^{-k_1 \alpha n}) \\ &\geq 1 - e^{-k_2 \alpha n} - \Pr(|N - 1| \geq e^{-k_1 \alpha n}) \\ &\geq 1 - e^{-k_2 \alpha n} - e^{-2k_1 \alpha n} d^{-n} \\ &\geq 1 - e^{-k_2 \alpha n} - e^{-(1-2k_1) \alpha n}, \end{aligned} \quad (29)$$

where we have used Lemma 1 and  $\alpha \leq \log d$ . Finally,

$$\begin{aligned} \Pr(\Delta A_{\psi'}^{\infty} \leq e^{-c_1 \alpha n}) &\geq 1 - e^{-k_2 \alpha n} - e^{-\alpha(1-2k_1)n} \\ &\geq 1 - e^{-c_2 \alpha n} \end{aligned} \quad (30)$$

for a constant  $c_2 > 0$  and sufficiently large  $n$ . ■

### III. EXTENSIVITY OF THE RÉNYI-2 ENTANGLEMENT ENTROPY

In this section, we turn to proving our second result. This is once again motivated by the phenomenon of equilibration in quantum many-body physics and the endeavor to provide a rigorous foundation for it. It has been proven



in Ref. [44] that systems equilibrate if their energy eigenstates have Rényi entropy that is extensive in the system size  $n$ , which means that

$$S_2(\text{Tr}_A[|j\rangle\langle j|]) \geq g(j)n \tag{31}$$

for a sufficiently well-behaved function  $g$  and *some* subsystem  $A$ . This property has been dubbed *entanglement ergodicity* [44]. Motivated by this insight, it has been proven in Ref. [16] that generic translation-invariant MPSs have extensive Rényi entropy if one considers a bipartition of the chain into the subsystem that corresponds to every  $k$ th site and the rest. That is, the entropy grows proportional to the boundary  $|\partial A|$ . Considering this particular partitioning is enough, since one partition of the system satisfying Eq. (31) suffices to consider the system entanglement ergodic, and to show the equilibration properties proven in Ref. [44]. Here, we prove such a result with explicit quantitative bounds for disordered RMPSs with overwhelming probability. Interestingly, the details of the correlation length of the state is of no concern for the bound presented. Even though the correlation length is expected to be small compared to the system size but nonzero (compare also the results of Ref. [12] in the translationally invariant case), we still arrive at an extensive bound of the respective entropy, independent of how the distance  $k$  between the sites precisely relates to the correlation length.

**Theorem 2** (Extensivity of entanglement entropies). *Suppose that  $n$  is divisible by a positive integer  $k$  and that  $A$  consists of what remains after tracing out every  $k$ th qudit. Let  $\rho'_A$  be the normalized density matrix of a RMPS drawn from  $\mu_{d,n,D}$  reduced on  $A$ . Then,*

$$\Pr\left[S_2(\rho'_A) \geq \Omega\left(\frac{n}{k}\right)\right] \geq 1 - e^{-\Omega(n/k)}. \tag{32}$$

*Proof.* The proof of Theorem 2 follows similar lines as that of Lemma 3. We first show that the purity ( $\text{Tr}[\rho_A^2]$ ) is almost minimal and then apply Markov's inequality. We have, for the expected purity of a subsystem,

$$\begin{aligned} \text{Tr}[\rho_A^2] &= \mathbb{E}_{\psi \sim \nu}(\text{Tr}[\rho_A^2]) \\ &= \mathbb{E}_{\psi \sim \nu}(\text{Tr}[\mathbb{F}_{A,\bar{A}} \rho_A^{\otimes 2}]) \\ &= \text{Tr}\{\mathbb{F}_{A,\bar{A}} \text{Tr}_{B,\bar{B}}[\mathbb{E}(|\psi\rangle\langle\psi|)^{\otimes 2}]\} \\ &= \text{Tr}[\mathbb{F}_{A,\bar{A}} \otimes \mathbb{1}_{B,\bar{B}} \mathbb{E}(|\psi\rangle\langle\psi|)^{\otimes 2}]. \end{aligned} \tag{33}$$

In the same graphical notation as in the proof of Lemma 3, this amounts to

$$\mathbb{E} \text{Tr}[\rho_A^2] = \sum_{\{\mathbb{1}, \mathbb{F}\}^{2n}} \text{Diagram} \tag{34}$$

here depicted for the subset  $A$  corresponding to every  $k = 3$  spin. Similar to the strategy laid out in Ref. [26], we sum over every lower ball to obtain a statistical model. For this, we define the following two interactions. These are

$$\text{Diagram} := \sum_{\mathbb{1}, \mathbb{F}} \text{Diagram}, \quad \text{Diagram} := \sum_{\mathbb{1}, \mathbb{F}} \text{Diagram} \tag{35}$$

Graphically, this yields the chain

$$\mathbb{E} \text{Tr}[\rho_A^2] = \sum_{\{\mathbb{1}, \mathbb{F}\}^n} \text{Diagram} \tag{36}$$

In order to proceed, we compute

$$\begin{aligned} \mathbb{F} \text{---} \square \text{---} \mathbb{F} &= 1, \\ \mathbb{1} \text{---} \square \text{---} \mathbb{1} &= \frac{dD^2 - d}{D^2 d^2 - 1} = \eta(d, D), \\ \mathbb{1} \text{---} \square \text{---} \mathbb{F} &= 0, \\ \mathbb{F} \text{---} \square \text{---} \mathbb{1} &= \frac{d^2 D - D}{D^2 d^2 - 1} = \eta(D, d) \end{aligned} \tag{37}$$

with the notation

$$\eta(x, y) := \frac{xy^2 - x}{x^2 y^2 - 1}. \tag{38}$$

For the green plaquettes, we simply switch the roles of  $\mathbb{1}$  and  $\mathbb{F}$ . As an intermediate step, let us compute the same but with  $j$  blue plaquettes. The sum over the black balls is implicit, and the following is obtained by simply counting how many switches between  $\mathbb{1}$  and  $\mathbb{F}$  are allowed:

$$\begin{aligned}
& \mathbb{F} \text{---} \boxed{\text{---}} \mathbb{F} = 1, \\
& \mathbb{1} \text{---} \boxed{\text{---}} \mathbb{1} = \eta(d, D)^j, \\
& \mathbb{1} \text{---} \boxed{\text{---}} \mathbb{1} = \mathbb{F} = 0, \\
& \mathbb{F} \text{---} \boxed{\text{---}} \mathbb{1} = \eta(D, d) \sum_{l=1}^j \eta(d, D)^{j-l} \\
& \quad = \eta(D, d) \frac{1 - \eta(d, D)^j}{1 - \eta(d, D)}
\end{aligned} \tag{39}$$

We can group the chain in Eq. (36) into blocks of  $k - 1$  spins consisting of all spins to the right of a green plaquette except the one before the next green plaquette. For  $k \geq 2$ , we have, using Eq. (39) with  $j = k - 1$  combined with Eq. (37) for the final green plaquette,

$$\begin{aligned}
& \mathbb{1} \text{---} \boxed{\text{---}} \mathbb{1} \\
& \quad = \eta(d, D)^{k-1} + \eta(D, d)^2 \frac{1 - \eta(d, D)^{k-1}}{1 - \eta(d, D)} \\
& \quad \leq \eta(d, D)^{k-1} + \eta(D, d), \\
& \mathbb{1} \text{---} \boxed{\text{---}} \mathbb{F} \\
& \quad = \eta(D, d) \\
& \quad \leq \eta(d, D)^{k-1} + \eta(D, d), \\
& \mathbb{F} \text{---} \boxed{\text{---}} \mathbb{1} \\
& \quad = \eta(d, D) \eta(D, d) \frac{1 - \eta(d, D)^{k-1}}{1 - \eta(d, D)} \\
& \quad \leq \eta(D, d) + \eta(d, D)^{k-1}, \\
& \mathbb{F} \text{---} \boxed{\text{---}} \mathbb{F} = \eta(d, D).
\end{aligned} \tag{40}$$

In this step, we have used the facts that  $\eta(x, y) \leq 1/x$  and, for  $x \geq 2$ ,

$$\frac{1}{1 - \eta(x, y)} \leq 2. \tag{41}$$

We sum over the number of spins with value  $\mathbb{1}$ . We make use of the fact that every spin with value  $\mathbb{1}$  contributes a factor of  $\eta(d, D)^{k-1} + \eta(D, d)$  (via the plaquette to its right), to find that

$$\begin{aligned}
\mathbb{E} \text{Tr}[\rho_A^2] & \leq \sum_{i=0}^{n/k} \binom{n/k}{i} [\eta(d, D)^{k-1} + \eta(D, d)]^i \eta(d, D)^{n/k-i} \\
& = [\eta(d, D)^{k-1} + \eta(D, d) + \eta(d, D)]^{n/k} \\
& = \eta(d, D)^{n/k} [1 + \eta(d, D)^{k-2} \\
& \quad + \eta(D, d) \eta(d, D)^{-1}]^{n/k}.
\end{aligned} \tag{42}$$

Hence, for  $d^{(1/2)} \leq D$  and large enough  $k$ , this expectation value becomes arbitrarily close to the minimal value  $d^{-n/k}$ . In the regime  $D^2 \leq d$ , the Rényi-2 entropy is bounded by the entanglement area law [38]

$$|\partial A| \log(D) = \frac{2n}{k} \log(D). \tag{43}$$

As in the proof of Theorem 1, the expression for the expectation value in Eq. (42) implies a concentration result via Markov's inequality and the union bound. This concentration inequality extends to the Rényi-2 entropy. ■

#### IV. MAXIMUM ENTROPY FOR SMALL CONNECTED SUBSYSTEMS

In this section, we show that small connected subsystems will, with high probability, feature a close to maximum entropy. A principle of maximum entropy for translation-invariant RMPs has been proved in Refs. [11,17].

**Theorem 3** (Almost maximum entropy for reduced states). *Let  $A$  be a subset of  $l$  consecutive qudits, and let  $\rho'_A$  be the normalized density matrix of a RMPs drawn from  $\mu_{d,n,D}$  reduced to  $A$ . Then, for any real  $r$ ,*

$$\Pr\{\text{Tr}[\rho_A'^2] \geq \Omega(D^{-r}) + d^{-l}\} \leq O(D^{-(2-r)}) \tag{44}$$

for

$$n \geq 2 \frac{\log D}{\log d} + l. \tag{45}$$

*Proof.* Let  $\rho_A$  be the unnormalized quantum state. In the disordered case we consider here, we obtain analogously to the previous section the expression

$$\begin{aligned}
& \mathbb{E} \text{Tr}[\rho_A^2] \\
& = \sum_{\{\mathbb{1}, \mathbb{F}\}^n} \dots \boxed{\text{---}} \dots
\end{aligned} \tag{46}$$

for the subset  $A$  consisting of  $l$  consecutive qudits. Summing over the four contributions in Eq. (39), we obtain

$$\begin{aligned}
\mathbb{E} \text{Tr}[\rho_A^2] & = \eta(d, D)^{n-l} + \eta(d, D)^l + 0 \\
& \quad + \eta(D, d)^2 \frac{1 - \eta(D, d)^l}{1 - \eta(D, d)} \frac{1 - \eta(D, d)^{n-l}}{1 - \eta(D, d)}.
\end{aligned} \tag{47}$$

By Eq. (41), for  $d, D \geq 2$ ,

$$\mathbb{E} \text{Tr}[\rho_A^2] \leq \frac{1}{d^l} + \frac{1}{d^{n-l}} + 4 \frac{1}{D^2} \tag{48}$$

holds. Let  $N = \text{Tr}[\rho]$  be the norm squared of the MPS. We have

$$\text{Tr}[\rho_A^2] \geq N^2 \frac{1}{d^l}, \quad (49)$$

and using [see Eq. (67) in the proof of Lemma 1]

$$\mathbb{E}(N^2) = 1 + \eta(d, D)^n \geq 1, \quad (50)$$

we find that

$$\begin{aligned} \mathbb{E}(\text{Tr}[\rho_A^2] - N^2 d^{-l}) &\leq 4 \frac{1}{D^2} + \frac{1}{d^{n-l}} + \frac{1}{d^{n+l}} \\ &\leq 6 \frac{1}{D^2}, \end{aligned} \quad (51)$$

where we have used

$$n \geq 2 \frac{\log D}{\log d} + l, \quad \text{i.e., } d^{l-n} \leq \frac{1}{D^2}. \quad (52)$$

Markov's inequality then yields

$$\Pr(\text{Tr}[\rho_A^2] - N^2 d^{-l} \geq D^{-r}) \leq \frac{1}{6} D^{-(2-r)}. \quad (53)$$

We now need to ensure that the normalization of the state does not worsen the bound. We employ the union bound in a similar manner as in the proof of Theorem 1. Let  $\rho'_A = \rho_A/N$  be the normalized state. Suppose that

$$\text{Tr}[\rho_A^2] - N^2 d^{-l} \leq D^{-r} \quad (54)$$

and  $|N - 1| \leq \epsilon$ . Then  $N^2 \geq (1 - \epsilon)^2$  and

$$\text{Tr}[\rho_A^2] - d^{-l} \leq \frac{D^{-r}}{(1 - \epsilon)^2}. \quad (55)$$

Pick  $\epsilon := t d^{-kn}$ , with  $k \leq 1 - r/4$  and  $t < 1$ . Then  $\epsilon \leq t d^{kl} D^{-2k}$  and

$$\text{Tr}[\rho_A^2] - d^{-l} \leq O(D^{-r}). \quad (56)$$

By the same union bound argument as in the proof of Theorem 1, we get

$$\begin{aligned} \Pr\{\text{Tr}[\rho'_A] - d^{-l} \leq O(D^{-r})\} \\ &\geq \Pr\{\text{Tr}[\rho_A] - N^2 d^{-l} \leq O(D^{-r})\} - \Pr(|N - 1| \leq d^{-kn}) \\ &\geq 1 - O(D^{-(2-r)}) - d^{-(1-2k)n} \\ &\geq 1 - O(D^{-(2-r)}), \end{aligned} \quad (57)$$

where we have used the fact that  $k \leq 1 - r/4$  and hence  $d^{(1-2k)n} \geq D^{2-r}$ . This concludes the proof. ■

The bound on the expectation value (47) is of the form

$$\mathbb{E} \text{Tr}[\rho_A^2] \leq d^{-l} + O(D^{-2}) + O(d^{l-n}). \quad (58)$$

In comparison, a bound for the expectation value of the form  $d^{-l} + O(D^{-1/10})$  for  $D \geq n^5$  has been obtained in Ref. [17] for random translation-invariant MPSs. For periodic boundary conditions, a similar result has been proven in Ref. [11]. Perhaps not surprisingly, our bound in the disordered case scales slightly better in  $D$ . Moreover, the bond dimension  $D$  is not required to grow with the system size  $n$ .

In Ref. [17], Levy's lemma has been employed in order to obtain an exponential concentration bound. This is not possible in our case, as the Lipschitz constant of the purity is necessarily lower bounded by  $O(D)$ . As a matter of fact, consider the MPS generated by choosing

$$U^{(k)} := \mathbb{1}_d \otimes \mathbb{1}_D \quad (59)$$

for all  $k$ , i.e.,  $\rho = D^2 |0\rangle\langle 0|^{\otimes n}$ . Then, for any  $A$ ,  $\text{Tr}[\rho_A^2] = D^2$ . Now pick a site  $k$  and a unitary

$$U := \mathbb{1}_d \otimes V \quad (60)$$

with  $\text{Tr}(V) = 0$ . For example, we can choose

$$V := \text{diag}(e^{2\pi i j/D})_{j=0, \dots, D-1}. \quad (61)$$

We construct the MPS  $\sigma$  with the identity on every site and  $U$  on site  $k$ . For any  $A$  containing  $k$ ,  $\text{Tr}[\sigma_A^2] = 0$ . The Lipschitz constant of the function

$$\begin{aligned} U(dD)^{\times n} &\rightarrow \mathbb{R}, \\ (U) &\mapsto \text{Tr}[\rho_A^{U^2}], \end{aligned} \quad (62)$$

where  $\rho^U$  is the MPS constructed with the unitaries and where we equip  $U(dD)^{\times n}$  with the product Frobenius norm, is then bounded by

$$L \geq \frac{|\text{Tr}[\rho_A^2] - \text{Tr}[\sigma_A^2]|}{\|\mathbb{1} - U\|_2} \geq \frac{D^2}{2D} = O(D), \quad (63)$$

where we have made use of the triangle inequality to further bound the denominator.

## V. CONCENTRATION AROUND THE UNIT NORM

In this section, we make use of the machinery of statistical mechanics mappings in order to show that the norm of the vectors  $|\psi\rangle$  is exponentially concentrated around unity, i.e., around being an actual quantum state.

**Lemma 6** (Concentration around a unit norm). *It holds that*

$$\Pr(|\langle \psi | \psi \rangle - 1| \geq \epsilon) \leq \epsilon^{-2} d^{-n}. \quad (64)$$

To prove this statement, first note that

$$\mathbb{E}\langle\psi|\psi\rangle^2 = \text{Tr}[\mathbb{F}(|\psi\rangle\langle\psi|)^{\otimes 2}]. \quad (65)$$

Moreover, using first-order Weingarten calculus, it is easy to see that  $\mathbb{E}\langle\psi|\psi\rangle = 1$ . Graphically, this corresponds to the spin chain depicted in Eq. (34) with only blue plaquettes as

$$\begin{aligned} & \mathbb{E}\langle\psi|\psi\rangle^2 \\ &= \sum_{\{1, \mathbb{F}\}^n} \dots \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \dots \end{aligned} \quad (66)$$

However, given the values of the plaquettes in Eq. (37), the only nonzero contributions are all spins  $\mathbb{1}$  or all spins  $\mathbb{F}$ . This yields

$$\mathbb{E}\langle\psi|\psi\rangle^2 = 1 + \left( \frac{dD^2 - d}{D^2d^2 - 1} \right)^n \leq 1 + d^{-n}. \quad (67)$$

Now from Markov's inequality we obtain

$$\text{Pr}[(\langle\psi|\psi\rangle - 1)^2 \geq \varepsilon] \leq \varepsilon^{-1} \mathbb{E}(1 - \langle\psi|\psi\rangle)^2 \leq \varepsilon^{-1} d^{-n}. \quad (68)$$

This exponential concentration allows us to prove all other concentration results without regarding the normalization. In fact, we combine every calculation with a union bound and the above concentration result such that the normalization does not change the statements. It is also noteworthy that all these concentration results follow from simple applications of Markov's inequality and do not require geometric methods such as Levy's lemma [45].

## VI. LOCAL EXPECTATION VALUES

Consider an observable  $O$  acting on  $(\mathbb{C}^d)^{\otimes l}$  for some small number of sites  $l$ . Here we consider the expectation values  $\langle\psi|O \otimes \mathbb{1}_{(\mathbb{C}^2)^{\otimes n-l}}|\psi\rangle$ . We assume without loss of generality that  $O$  is traceless. First, note that we have

$$\mathbb{E}\langle\psi|O \otimes \mathbb{1}_{(\mathbb{C}^2)^{\otimes n-l}}|\psi\rangle = \text{Tr}(O) = 0. \quad (69)$$

We want to show that the expectation values are concentrated around 0. One possible way to do this is to exploit the results of Sec. IV to argue that the local density matrix is close in norm with high probability to the maximally mixed state, as done in Ref. [17], and conclude by Hölder's inequality that, for the expectation value expressed in terms of the reduced density matrix  $\rho$ , the

following holds:

$$|\text{Tr}[O\rho]| = |\text{Tr}[O(\rho - \mathbb{1}/d^l)]| \leq \|\rho - \mathbb{1}/d^l\|_\infty \|O\|_1. \quad (70)$$

Here  $\|O\|_1$  is a constant in  $n$  and  $D$ . Nevertheless, we want to showcase how our method can be applied directly to this problem.

To showcase the flexibility of our approach, we provide a direct bound on the second moment

$$\mathbb{E}\langle\psi|O \otimes \mathbb{1}_{(\mathbb{C}^2)^{\otimes n-l}}|\psi\rangle^2 = \mathbb{E}\text{Tr}[(O \otimes \mathbb{1}_{(\mathbb{C}^2)^{\otimes n-l}}|\psi\rangle\langle\psi|)^{\otimes 2}]. \quad (71)$$

We consider the case  $l = 1$  for simplicity. Graphically, we have

$$\begin{aligned} & \mathbb{E}\langle\psi|O|\psi\rangle^2 \\ &= \dots \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \square \text{---} \bullet \text{---} \dots \end{aligned} \quad (72)$$

where  $|O\rangle := [(\mathbb{1} \otimes O)|\Omega\rangle]^{\otimes 2}$ . We obtain the interactions

$$\begin{aligned} \mathbb{1} \rightsquigarrow \square &= -\frac{\text{Tr}[O^2]}{D^2d^3 - d}, \\ \mathbb{F} \rightsquigarrow \square &= \frac{\text{Tr}[O^2]D^2}{D^2d^2 - 1}, \\ \mathbb{1} \rightsquigarrow \square &= -\frac{\text{Tr}[O^2]D}{D^2d^3 - d}, \\ \mathbb{F} \rightsquigarrow \square &= \frac{\text{Tr}[O^2]D}{D^2d^2 - 1} \end{aligned} \quad (73)$$

from this. Simply by ignoring the one negative contribution (all spins  $\mathbb{1}$ ), we obtain

$$\begin{aligned} & \mathbb{E}\langle\psi|O \otimes \mathbb{1}_{(\mathbb{C}^2)^{\otimes n-l}}|\psi\rangle^2 \\ & \leq \text{Tr}[O^2] \left[ \frac{D}{D^2d^2 - 1} \eta(D, d) \sum_{i=0}^{n-2} \eta(d, D)^i \right. \\ & \quad \left. + \frac{D^2}{D^2d^2 - 1} \left( \frac{d^2D - D}{D^2d^2 - D} \right)^{n-2} \right] \\ & \leq \text{Tr}[O^2] \left( D^{-2} \frac{1 - d^{-n+1}}{1 - d^{-1}} + D^{-n+1} \right) \\ & \leq 2D^{-2} \text{Tr}[O^2] \end{aligned} \quad (74)$$

for  $n \geq 2$  as a combinatorial bound. We can now achieve a concentration result for local expectation values from Markov's inequality. Previously, typicality of expectation values was argued for translation-invariant MPSs in

Ref. [46], where the authors argued an exponential concentration if  $D$  grows faster than  $\Omega(n^2)$ . However, their proof is based on the concentration of measure phenomenon and the bound they obtained on the Lipschitz constant involves the assumption that the second highest eigenvalue of the transfer operator is small enough for essentially all instances. While a numerical assessment indicates that this is the case for all practical purposes, it is difficult to obtain explicit rigorous bounds for the dependence on  $n$  this way. Our simple bound from second moments only scales as  $D^{-2}$  (as opposed to  $e^{-\Omega(D)}$ ), but it is independent of  $n$  and does not require any further assumptions.

### VII. OUTLOOK

In this work, we have systematically explored RMPSs where the individual tensors have been chosen independently according to the Haar measure. Exploiting a mapping to a one-dimensional statistical mechanics model, we have been in the position to compute expectation values of various quantities for such disordered RMPSs. We have derived concentration results for the effective dimension, implying equilibration under the time evolution of generic Hamiltonians, and concentration results for the Rényi-2 entropy and the expectation values of local observables.

While we have put properties of such families of quantum states into the focus of our analysis, it should be clear that by means of the parent Hamiltonian concept mentioned above, we could have discussed ensembles of local Hamiltonians. For translational-invariant RMPSs [12], this picture is particularly transparent.

An obvious further problem is to consider quantities that require higher moments, such as higher Rényi entropies  $S_\alpha$  with  $2 \leq \alpha \in \mathbb{Z}$ . This would result in more complicated statistical models with  $\alpha!$  many local degrees of freedom. Another open question concerns higher-dimensional systems. It would be interesting to apply a similar analysis to projected entangled pair states.

For translational-invariant RMPSs, the expected correlation length has been proven in Refs. [11]. The family of states defined in this fashion gives rise to generic representatives of the *trivial phase of matter* [9] with unit probability. If a state has symmetries, i.e., if the state vector satisfies  $U_g^{\otimes n}|\psi\rangle = e^{i\theta_g}|\psi\rangle$  for some real phase  $\theta_g$ , and where  $g \mapsto U_g$  is a linear unitary representation of a symmetry group  $G$ , then this symmetry is (for a suitable phase gauge) reflected on the virtual level of the MPS as a projective unitary representation of group  $G$ , satisfying

$$V_g V_h = e^{i\omega(g,h)} V_{gh} \tag{75}$$

for a real phase  $(g, h) \mapsto \omega(g, h)$  [47]. Different *phases of matter* respecting these symmetries in symmetry-protected topological order are now captured by equivalence classes, called cohomology classes, they again forming a group,

the second cohomology group  $H_2[G, U(1)]$  of  $G$  over  $U(1)$  [9,48,49]. That said, it now makes sense to think of RMPSs that respect a physical symmetry and think of *typical symmetry-protected topological (SPT) phases of matter*. Here, the Haar measure is chosen in each of the blocks of a direct sum on the virtual level, respecting the projective unitary representation of group  $G$ . In this sense, one can speak of common representatives of SPT phases, a line of thought that will be elaborated upon elsewhere.

More broadly put, this work can be seen as a contribution to a bigger program concerned with understanding generic phases of quantum matter by means of *random tensor networks*. Indeed, properties of random tensor networks can often be easier computed than those of tensor networks in which the entries are specifically chosen. Randomness hence serves as a computational tool, a line of thought that can be dated back to Ref. [3] and further.

In Ref. [50], such a line of thought has already been applied to identify properties of holographic tensor networks, where a desirable property to be a so-called perfect tensor turns out to be approximately satisfied with high probability. Building upon this insight, one can compute properties of the resulting boundary state. More ambitiously still, it makes a lot of sense to think of holographic random tensor network models in which the tensors have further structure, e.g., to be match-gate tensor networks [51]. Similarly, questions of properties of higher-dimensional cubic tensor networks arise along similar lines. It is the hope that the present work can provide insights and a powerful machinery to address such further questions when exploring typical instances of phases of matter.

### ACKNOWLEDGMENTS

We would like to warmly thank Alexander Altland, Alex Goëßmann, Dominik Hangleiter, Nick Hunter-Jones, Richard Kueng, Amin Thainat, and Carolin Wille for fruitful discussions. We would like to thank the DFG (CRC 183, project A03, FOR 2724, EI 519/15-1, EI 519/17-1) and the FQXi for support.

### APPENDIX

A natural follow up question is whether RMPSs have features similar to generic Haar-random states. Naively, an argument as above might be used to show that RMPSs form approximate spherical 2-designs [31]. The difference of moment operators in the 2-norm can be bounded using the frame potential as

$$\left\| \mathbb{E}_{\psi \sim \nu} (|\psi\rangle\langle\psi|)^{\otimes 2} - \frac{2P_{\text{sym}}}{d^n(d^n + 1)} \right\|_F^2 = \mathcal{F}_{2,\nu} - \frac{2}{d^n(d^n - 1)}, \tag{A1}$$

and the frame potential is given by

$$\mathcal{F}_{2,\nu} := \mathbb{E}_{\psi, \phi \sim \nu} |\langle \psi | \phi \rangle|^4, \quad (\text{A2})$$

which is reminiscent of the expression in Eq. (11). After all, random product states constitute an exact projective 1-design.

However, RMPSSs with polynomially bounded bond dimensions have low-entanglement structure by definition. Since the Schmidt rank along any bipartition is bounded by  $D$ , the purity is bounded below by  $1/D$ , but the average over this quantity is exponentially small for an approximate 2-design. In more detail, consider a probability measure such that

$$\left\| \mathbb{E}_{\psi \sim \nu} (|\psi\rangle\langle\psi|)^{\otimes 2} - \frac{2P_{\text{sym}}}{d^n(d^n+1)} \right\|_F \leq \varepsilon. \quad (\text{A3})$$

With this notion we obtain a straightforward bound on the entanglement purity over a bipartition of the spin chain into subsets  $A$  and  $B$  of equal size  $n/2$ . With the norm inequality  $\|\cdot\|_1 \leq \sqrt{\dim \mathcal{H}} \|\cdot\|_F$  [see, e.g., Eq. (1.2.6) of Ref. [52]], we obtain

$$\begin{aligned} \mathbb{E}_{\psi \sim \nu} (\text{Tr}[\rho_A^2]) &= \mathbb{E}_{\psi \sim \nu} (\text{Tr}[\mathbb{F}_{A,\bar{A}} \rho_A^{\otimes 2}]) \\ &= \text{Tr}\{\mathbb{F}_{A,\bar{A}} \text{Tr}_{B,\bar{B}}[\mathbb{E}_{\psi \sim \nu} (|\psi\rangle\langle\psi|)^{\otimes 2}]\}. \end{aligned} \quad (\text{A4})$$

Therefore,

$$\begin{aligned} \mathbb{E}_{\psi \sim \nu} (\text{Tr}[\rho_A^2]) &\leq \frac{2}{d^n(d^n+1)} \text{Tr}\{\mathbb{F}_{A,\bar{A}} \text{Tr}_{B,\bar{B}}[P_{\text{sym}}]\} + d^n \varepsilon \\ &= \frac{1}{d^n(d^n+1)} \text{Tr}\{\mathbb{F}_{A,\bar{A}} \text{Tr}_{B,\bar{B}}[\mathbb{F}_{AB,\bar{A}\bar{B}} + \mathbb{1}_{AB,\bar{A}\bar{B}}]\} + d^n \varepsilon \\ &= \frac{2d^{3n/2}}{d^n(d^n+1)} + d^n \varepsilon. \end{aligned} \quad (\text{A5})$$

As a small expectation value implies the existence of instances with small values, this leads to a contradiction for

$$\varepsilon \leq \frac{d^{-n}}{D}. \quad (\text{A6})$$

This argument rules out a ‘‘cutoff’’ phenomenon, where the error is exponentially suppressed in the bond dimension  $D$  only after some polynomial threshold has been surpassed.

---

[1] G. Akemann, J. Baik, P. Di Francesco, editors., *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, Oxford, 2015).

- [2] P. Hayden, D. Leung, P. W. Shor, and A. Winter, Randomizing quantum states: Constructions and applications, *Comm. Math. Phys.* **250**, 371 (2004).
- [3] P. Hayden, D. W. Leung, and A. Winter, Aspects of generic entanglement, *Comm. Math. Phys.* **265**, 95 (2006).
- [4] J. Sánchez-Ruiz, Simple proof of page’s conjecture on the average entropy of a subsystem, *Phys. Rev. E* **52**, 5653 (1995).
- [5] 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).
- [6] J. Eisert, M. Friesdorf, and C. Gogolin, Quantum many-body systems out of equilibrium, *Nat. Phys.* **11**, 124 (2015).
- [7] C. Gogolin and J. Eisert, Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems, *Rep. Prog. Phys.* **79**, 56001 (2016).
- [8] A. Polkovnikov, K. Sengupta, A. Silva, and M. Venkatachalan, Non-equilibrium dynamics of closed interacting quantum systems, *Rev. Mod. Phys.* **83**, 863 (2011).
- [9] N. Schuch, D. Perez-Garcia, and I. Cirac, Classifying quantum phases using matrix product states and projected entangled pair states, *Phys. Rev. B* **84**, 165139 (2011).
- [10] S. Garnerone, T. R. de Oliveira, S. Haas, and P. Zanardi, Statistical properties of random matrix product states, *Phys. Rev. A* **82**, 052312 (2010).
- [11] C. E. Gonzales-Guillen, M. Junge, and I. Nechita, On the spectral gap of random quantum channels. [arXiv:1811.08847](https://arxiv.org/abs/1811.08847).
- [12] C. Lancien and D. Pérez-García, Correlation length in random MPS and PEPS. [arXiv:1906.11682](https://arxiv.org/abs/1906.11682).
- [13] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Colloquium: Many-body localization, thermalization, and entanglement, *Rev. Mod. Phys.* **91**, 021001 (2019).
- [14] M. Friesdorf, A. H. Werner, W. Brown, V. B. Scholz, and J. Eisert, Many-Body Localisation Implies That Eigenvectors are Matrix-Product States, *Phys. Rev. Lett.* **114**, 170505 (2015).
- [15] D. A. Huse, R. Nandkishore, and V. Oganesyan, Phenomenology of fully many-body-localized systems, *Phys. Rev. B* **90**, 174202 (2014).
- [16] A. Rolandi and H. Wilming, Extensive Rényi entropies in matrix product states. [arXiv:2008.11764](https://arxiv.org/abs/2008.11764), (2020).
- [17] B. Collins, C. E. González-Guillén, and D. Pérez-García, Matrix product states, random matrix theory and the principle of maximum entropy. [arXiv:1201.6324](https://arxiv.org/abs/1201.6324), (2012).
- [18] R. Movassagh and J. Schenker, An ergodic theorem for homogeneously distributed quantum channels with applications to matrix product states. [arXiv:1909.11769](https://arxiv.org/abs/1909.11769), (2019).
- [19] R. Movassagh and J. Schenker, Theory of ergodic quantum processes. [arXiv:2004.14397](https://arxiv.org/abs/2004.14397), (2020).
- [20] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [21] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, R. Barends, B. Burkett, Y. Chen, and Z. Chen, A blueprint for demonstrating quantum supremacy with superconducting qubits, *Science* **360**, 195 (2017).
- [22] 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).

- [23] F. G. S. L. Brandao, W. Chemsyany, N. Hunter-Jones, R. Kueng, and J. Preskill, Models of quantum complexity growth. [arXiv:1912.04297](#).
- [24] A. Chandran and C. R. Laumann, Semiclassical limit for the many-body localization transition, *Phys. Rev. B* **92**, 024301 (2015).
- [25] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth, Quantum homeopathy works: Efficient unitary designs with a system-size independent number of non-clifford gates. [arXiv:2002.09524](#).
- [26] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits. [arXiv:1905.12053](#), (2019).
- [27] A. Nahum, S. Vijay, and J. Haah, Operator Spreading in Random Unitary Circuits, *Phys. Rev. X* **8**, 021014 (2018).
- [28] C. Sünderhauf, D. Pérez-García, D. A. Huse, N. Schuch, and J. I. Cirac, Localization with random time-periodic quantum circuits, *Phys. Rev. B* **98**, 134204 (2018).
- [29] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, *Phys. Rev. B* **99**, 174205 (2019).
- [30] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, Guaranteed recovery of quantum processes from few measurements, *Quantum* **3**, 171 (2019).
- [31] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, *J. Math. Phys.* **48**, 052104 (2007).
- [32] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, *Commun. Math. Phys.* **144**, 443 (1992).
- [33] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, *Quantum Inf. Comput.* **7**, 401 (2007).
- [34] D. Gross and J. Eisert, Quantum computational webs, *Phys. Rev. A* **82**, 040303(R) (2010).
- [35] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, *Quant. Inf. Comp.* **5** and **6**, 401 (2007).
- [36] N. Linden, S. Popescu, A. J. Short, and A. Winter, Quantum mechanical evolution towards thermal equilibrium, *Phys. Rev. E* **79**, 061103 (Jun 2009).
- [37] B. Bauer and C. Nayak, Area laws in a many-body localised state and its implications for topological order, *J. Stat. Mech.* **2013**, P09005 (2013).
- [38] J. Eisert, M. Cramer, and M. B. Plenio, Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (2010).
- [39] Y. Huang and A. W. Harrow, Instability of localization in translation-invariant systems, [arXiv:1907.13392](#) (2019).
- [40] P. Reimann, Foundation of Statistical Mechanics under Experimentally Realistic Conditions, *Phys. Rev. Lett.* **101**, 190403 (2008).
- [41] H. Tasaki, From Quantum Dynamics to the Canonical Distribution: General Picture and a Rigorous Example, *Phys. Rev. Lett.* **80**, 1373 (1998).
- [42] P. W. Brouwer and C. W. J. Beenakker, Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems, *J. Math. Phys.* **37**, 4904 (1996).
- [43] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, *Comm. Math. Phys.* **264**, 773 (2006).
- [44] H. Wilming, M. Goihl, I. Roth, and J. Eisert, Entanglement-Ergodic Quantum Systems Equilibrate Exponentially Well, *Phys. Rev. Lett.* **123**, 200604 (2019).
- [45] M. Ledoux, *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., (2001).
- [46] S. Garnerone, T. R. de Oliveira, and P. Zanardi, Typicality in random matrix product states, *Phys. Rev. A* **81**, 032336 (2010).
- [47] M. Sanz, M. M. Wolf, D. Pérez-García, and J. I. Cirac, Matrix product states: Symmetries and two-body hamiltonians, *Phys. Rev. A* **79**, 042308 (2009).
- [48] X. Chen, Z.-C. Gu, and X.-G. Wen, Classification of gapped symmetric phases in one-dimensional spin systems, *Phys. Rev. B* **83**, 035107 (2011).
- [49] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Entanglement spectrum of a topological phase in one dimension, *Phys. Rev. B* **81**, 064439 (2010).
- [50] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, Holographic duality from random tensor networks, *J. High En. Phys.* **2016**, 9 (2016).
- [51] A. Jahn, M. Gluza, F. Pastawski, and J. Eisert, Holography and criticality in matchgate tensor networks, *Sci. Adv.* **5**, eaaw0092 (2019).
- [52] R. A. Low, Pseudo-randomness and learning in quantum computation. [arXiv:1006.5227](#), (2010).

## SUMMARY AND OPEN PROBLEMS

---

In this thesis, we made advances at the intersection of quantum many-body physics, complexity theory and probability theory. We saw how to generate increasingly evenly distributed ensembles of states and unitaries from randomizing over quantum circuits and provided several strong bounds that quantify this phenomenon. We also used randomness to imply strong bounds on the growth of quantum circuit complexity. These results provide powerful examples of how randomness can be used as a proof technique for complexity. On the other hand, we also saw how complexity can arise from randomizing over the instances of a computational problem. In the practically motivated task of contracting tensor networks, we ruled out strong heuristic algorithms by turning any such algorithm into an efficient randomized solver for all tensor networks. Lastly, we explored the physical properties of generic matrix product states, thereby studying how local ensembles of states relate to properties of uniformly random states.

In the following we comment on open questions that we made partial progress on.

### 6.1 THE ROBUST BROWN-SUSSKIND CONJECTURE, DESIGNS IN LINEAR DEPTH AND THE SPECTRAL GAP OF RANDOM QUANTUM CIRCUITS

Large parts of this thesis are motivated by the Brown-Susskind conjecture for random quantum circuits, namely the intuition that complexity should grow linearly for almost all circuits for an exponentially long time. This conjecture can be formulated for essentially any sensible notion of circuit complexity. A crucial result in this thesis is that the conjecture holds for the exact implementation complexity. A more operational notion should allow for an implementation error, e.g. in the 1-norm of states quantifying the single shot distinguishability [BCHJ<sup>+</sup>21]. The result of Section 3.3 can even be shown to imply such a linear growth of complexity for an implementation error that is  $> 0$  but depends uncontrollably on the system-size. The problem with this is that we ideally would like the error to be constant in the system-size. To make any progress on the scaling in the system-size it seems like we need to use information about the curvature of the sets of states generated by circuits of a given depth.

A purely combinatorial approach based on unitary  $t$ -designs [BCHJ<sup>+</sup>21] works perfectly fine for operational/robust notions of circuit complexity. The best bound on the growth rate of robust complexity was based on this approach in Section 3.2 and scales like  $\Omega(\text{depth}^{1/(5+o(1))})$ .

Therefore, two seemingly different natural approaches to the robust Brown-Susskind conjecture fail for different reasons. However, while based on different mathematics, both approaches are versions of counting arguments. More precisely, they can both be viewed as partial derandomizations of counting arguments: a volume based counting argument presented in Section 3.2 is derandomized by unitary  $t$ -designs and a counting argument based on the dimension of the unitary group/state space can be derandomized by dimensions of sets of unitaries with a fixed circuit complexity in Section 3.3. In both approaches, we use that random quantum circuits are “expressive” and generate increasing amounts of randomness. This motivates to search for a



way to quantify the randomness of ensembles of states and unitaries that is weaker than the convergence for higher moments but stronger than the dimension of the ensemble’s support.

## 6.2 APPROXIMATE AVERAGE-CASE HARDNESS OF OUTPUT PROBABILITIES

To this day, there is no (oracle free) unconditional argument that relates superpolynomial speed-ups of (realistic) quantum devices to the separation of classical complexity classes such as P vs NP. On the other hand, an efficient sampler for quantum supremacy schemes up to multiplicative errors implies the collapse of the polynomial hierarchy. The problem here is that even a fully-fledged quantum computer will never be capable of such a simulation. This is because multiplicative errors allow for no error at all for output probabilities that are 0. A much more realistic notion of approximation is additive in total variation distance/  $\ell^1$  norm. Reducing hardness of sampling up to constant errors in total variation distance rests on one key conjecture: approximate average-case hardness.

In various settings, one can show worst-to-average case reductions based on polynomial interpolation, as developed by Lipton for the permanent [Lip91]. As it was explained in Chapter 4, this can be adapted to output probabilities of sampling schemes with continuous degrees of freedom. Usually, these reductions yield average-case complexity for approximation up to errors of the form  $2^{-\text{poly}(n)}$ . In comparison, for random quantum circuits, we would need average-case complexity for approximations up to errors of  $O(2^{-n})$ . Recently, the robustness of worst-to-average case reductions for random quantum circuits was improved to  $O(2^{-O(n)})$  [Kro22, BFL22, KMM22]

Unfortunately, recent work establishes strong limitations on worst-to-average case reductions based on gate-wise interpolations for errors that scale as  $2^n$  [BFL22, DFG<sup>+</sup>21]. In particular, every proof of average-case complexity for this level of robustness necessarily depends on the depth of the random quantum circuit beyond the worst-case hardness.

## 6.3 ANTICONCENTRATION OF BOSON SAMPLING

While well-established for most circuit-based sampling schemes, the anticoncentration conjecture for boson sampling remains wide open. While numerical results suggest that the distribution of the permanent is as flat as the distribution of the determinant [AA13], the complexity of the permanent is mirrored in the mathematical difficulty to characterize its distribution: It turns out that the usual approach of computing second moments in combination with the Paley-Zygmund inequality is not sufficient to establish the anticoncentration conjecture in this case.

A natural path forward is to compute higher moments and use them to characterize the permanent distribution. Unfortunately, this seems to require a very precise computation of the higher moments, which turns out to be a highly non-trivial combinatorial problem [Nez21]. In particular, we would need upper bounds that are far stronger than the ones obtained in Appendix A. Moreover, it is observed in Ref. [Nez21] that the permanent distribution for Gaussian matrices might not even be fully characterized by its integer moments due to possible heavy tails of the distribution.

## ACKNOWLEDGMENTS

---

First and foremost, I want to thank Jens Eisert for countless discussions, encouragements, unwaivering support and for creating the unique environment in Dahlem that allowed me to thrive.

I also want to thank Juani Bermejo-Vega for her patience with me when I was new to this field and for introducing me to some of the topics closest to my heart.

Research is not conducted in a vacuum and I was lucky enough to collaborate with many bright and kind researchers from all over the world. I want to thank Laura Baez, Juani Bermejo-Vega, Christian Bertoni, Adam Bouland, Ellen Derbyshire, Jens Eisert, Philippe Faist, Bill Fefferman, Azat Gainutdinov, Marek Gluza, Marcel Goihl, David Gross, Dominik Hangleiter, Markus Heinrich, Marcel Hinsche, Nick Hunter-Jones, Marios Ioannou, Yifan Jia, Naga Kothakonda, Richard Kueng, Felipe Montealegre-Mora, Anthony Munson, Alexander Nietner, Hakob Pashayan, Yihui Quek, Ingo Roth, Christoph Schweigert, Jean-Pierre Seifert, Ryan Sweke, Jadwiga Wilkens and Nicole Yunger Halpern for the shared joy of discovering something new, their hard work and creativity.

I want to thank Sevag Gharibian, David Gross, Richard Kueng for inviting me to the inspiring research visits that I have learned so much from.

I want to thank Jens Eisert, Markus Heinrich, Felipe Montealegre-Mora and Vicky Schneider for comments on this manuscript.

I am grateful for all the coffee breaks, chats, nights out, pasta, beers and games of chess I had with my colleagues and friends over the years. All of you made this experience far more interesting.

Nicht fehlen darf meine Familie: Karina, Jan, Celine, Edda and Wilhelm Haferkamp, Lotte and Alfred Decker, Urte, Gev and Jele Drücke, Lutz and Christine Neugebauer dafür, dass ich mich immer bei euch zuhause fühle. Ohne euch und eure Unterstützung wäre ich nicht wo ich bin.

Last, I want to thank Vicky Schneider, without whom all of this would be impossible, for her love, for shared indian food and for being my closest companion.



## BIBLIOGRAPHY

---

- [AA13] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. *Th. Comp.*, 9:143--252, 2013.
- [AAB<sup>+</sup>19] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505--510, 2019.
- [AALV09] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani. The detectability lemma and quantum gap amplification. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 417--426, 2009.
- [Aar] S. Aaronson. What google won't find. <https://scottaaronson.blog/?p=266>.
- [Aar05] S. Aaronson. Guest column: NP-complete problems and physical reality. *ACM Sigact News*, 36(1):30--52, 2005.
- [Aar11] S. Aaronson. A linear-optical proof that the permanent is #P-hard. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2136):3393--3405, 2011.
- [Aar16] S. Aaronson.  $P \neq NP$ ? *Open problems in mathematics*, 2016.
- [AAV16] A. Anshu, I. Arad, and T. Vidick. Simple proof of the detectability lemma and spectral gap amplification. *Physical Review B*, 93(20):205142, 2016.
- [AB09] S. Arora and B. Barak. *Computational Complexity A Modern Approach*. Cambridge University Press, Cambridge, 2009. OCLC: 840446601.
- [ABB<sup>+</sup>05] I Arsene, IG Bearden, D Beavis, C Besliu, B Budick, H Bøggild, C Chasman, CH Christensen, P Christiansen, J Cibor, et al. Quark--gluon plasma and color glass condensate at rhic? the perspective from the brahms experiment. *Nuclear Physics A*, 757(1-2):1--27, 2005.
- [ABB<sup>+</sup>18] A. Acin, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm. The European quantum technologies roadmap. *New J. Phys.*, 20:080201, 2018.
- [ABDF11] G. Akemann, J. Baik, and P. Di Francesco. *The Oxford handbook of random matrix theory*. Oxford University Press, 2011.
- [ABO97] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 176--188, 1997.

- [AD86] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333--348, 1986.
- [AE07] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 129--140. IEEE, 2007.
- [AG04] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, 2004.
- [AH12] S. Aaronson and T. Hance. Generalizing and derandomizing Gurvits's approximation algorithm for the permanent. *arXiv:1212.0025*, 2012.
- [Ajt96] M. Ajtai. Generating hard instances of lattice problems. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 99--108, 1996.
- [AL18] F. Alet and N. Laflorencie. Many-body localization: an introduction and selected topics. *Comptes Rendus Physique*, 2018.
- [AS10] A. Altland and B. D. Simons. *Condensed matter field theory*. Cambridge University Press, 2010.
- [AS16] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 2016.
- [BaHH16] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.*, 346:397--434, 2016.
- [Bar82] F. Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.
- [BB96] P. W. Brouwer and C. W. J. Beenakker. Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems. *Journal of Mathematical Physics*, 37(10):4904--4934, 1996.
- [BB01] J.-L. Brylinski and R. Brylinski. Universal quantum gates. 2001. *arXiv:quant-ph/0108062*.
- [BBC<sup>+</sup>19] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard. Simulation of quantum circuits by low-rank stabilizer decomposition. *Quantum*, 3:181, 2019.
- [BC17] J. C. Bridgeman and C. T. Chubb. Hand-waving and interpretive dance: an introductory course on tensor networks. *Journal of physics A: Mathematical and theoretical*, 50(22):223001, 2017.
- [BCHJ<sup>+</sup>21] F. G. S. L. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill. Models of quantum complexity growth. *PRX Quantum*, 2(3):030316, 2021.
- [BCR13] J. Bochnak, M. Coste, and M.-F. Roy. *Real algebraic geometry*, volume 36. Springer Science & Business Media, 2013.

- [BD92] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, pages 294--313, 1992.
- [BDN12] I. Bloch, J. Dalibard, and S. Nascimbene. Quantum simulations with ultracold quantum gases. *Nat. Phys.*, 8:267, 2012.
- [BFH<sup>+</sup>15] S. Braun, M. Friesdorf, J. S. Hodgman, M. Schreiber, J. P. Ronzheimer, A. Riera, M. del Rey, I. Bloch, J. Eisert, and U. Schneider. Emergence of coherence and the dynamics of quantum phase transitions. *PNAS*, 112:3641--3646, 2015.
- [BFK18] A. Bouland, J. F. Fitzsimons, and D. E. Koh. Complexity classification of conjugated Clifford circuits. In *Proc. 33rd Comp. Compl. Conf.*, page 21. Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik, 2018.
- [BFL22] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu. Noise and the frontier of quantum supremacy. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1308--1317. IEEE, 2022.
- [BFNV19] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani. Quantum supremacy and the complexity of random circuit sampling. *Nat. Phys.*, 15:159--163, March 2019.
- [BFV19] A. Bouland, B. Fefferman, and U. Vazirani. Computational pseudorandomness, the wormhole growth paradox, and constraints on the AdS/CFT duality. *arXiv:1910.14646*, 2019.
- [BGo8] J. Bourgain and A. Gamburd. On the spectral gap for finitely-generated subgroups of  $SU(2)$ . *Invent. math.*, 171(1):83--121, January 2008.
- [BG12] J. Bourgain and A. Gamburd. A Spectral Gap Theorem in  $SU(d)$ . *Journal of the European Mathematical Society*, 14(5):1455--1511, 2012.
- [BGH<sup>+</sup>20] M. L. Baez, M. Goihl, J. Haferkamp, J. Bermejo-Vega, M. Gluza, and J. Eisert. Dynamical structure factors of dynamical quantum simulators. *PNAS*, 117(42):26123--26134, 2020.
- [BH08] A. Brodsky and S. Hoory. Simple permutations mix even better. *Random Structures & Algorithms*, 32(3):274--289, 2008.
- [BH10] F. G. S. L. Brandão and M. Horodecki. Exponential quantum speed-ups are generic. *arXiv:1010.3654*, 2010.
- [BHH16] F. G. S. L. Brandao, A. W. Harrow, and M. Horodecki. Efficient quantum pseudorandomness. *Phys. Rev. Lett.*, 116, 2016.
- [BIS<sup>+</sup>18] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, J. M. Martinis, and H. Neven. Characterizing quantum supremacy in near-term devices. *Nature Phys.*, 14:595--600, 2018.
- [BJS11] M. J. Bremner, R. Jozsa, and D. J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2126):459--472, 2011.

- [BK05] S. Bravyi and A. Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Phys. Rev. A*, 71(2):022316, 2005.
- [BLSF19] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.
- [BM21] S. Bravyi and D. Maslov. Hadamard-free circuits expose the structure of the Clifford group. *IEEE Transactions on Information Theory*, 67(7):4546--4563, 2021.
- [BMS16a] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.*, 117:080501, 2016.
- [BMS16b] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.*, 117:080501, 2016.
- [BMS17] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum*, 1:8, 2017.
- [BNOZ22] E. Bannai, Y. Nakata, T. Okuda, and D. Zhao. Explicit construction of exact unitary designs. *Advances in Mathematics*, 405:108457, 2022.
- [BNRT18] E. Bannai, G. Navarro, N. Rizo, and P. H. Tiep. Unitary t-groups. *arXiv:1810.02507*, 2018.
- [BPM12] J. H. Bardarson, F. Pollmann, and J. E. Moore. Unbounded growth of entanglement in models of many-body localization. *Phys. Rev. Lett.*, 109(1):017202, 2012.
- [BR12] R. Blatt and C. F. Roos. Quantum simulations with trapped ions. *Nat. Phys.*, 8:277, 2012.
- [Bra05] S. Bravyi. Lagrangian representation for fermionic linear optics. *Quantum Inf. and Comp.*, 5:216, 2005.
- [BRS<sup>+</sup>16] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao. Complexity, action, and black holes. *Phys. Rev. D*, 93:086006, 2016.
- [BRV13] A. Bondarenko, D. Radchenko, and M. Viazovska. Optimal asymptotic bounds for spherical designs. *Annals of mathematics*, pages 443--452, 2013.
- [BS18a] A. R. Brown and L. Susskind. Second law of quantum complexity. *Physical Review D*, 97(8):086015, 2018.
- [BS18b] A. R. Brown and L. Susskind. Second law of quantum complexity. *Phys. Rev. D*, 97:086015, 2018.
- [BSK<sup>+</sup>17] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletic, and M. D. Lukin. Probing many-body dynamics on a 51-atom quantum simulator. *Nature*, 551(7682):579--584, November 2017.

- [BV93] E. Bernstein and U. Vazirani. Quantum complexity theory. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 11--20, 1993.
- [BVHS<sup>+</sup>18] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert. Architectures for quantum simulation showing a quantum speedup. *Phys. Rev. X*, 8:021010, 2018.
- [CCL03] E. A. Carlen, M. C. Carvalho, and M. Loss. Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta mathematica*, 191(1):1--54, 2003.
- [CGGPG13] B. Collins, C. E. Gonzalez-Guillen, and D. Perez-Garcia. Matrix product states, random matrix theory and the principle of maximum entropy. *Commun. Math. Phys.*, 320:663--677, 2013.
- [Che16] C. Chevalley. *Theory of Lie Groups (PMS-8), Volume 8*. Princeton University Press, 2016.
- [CHZ<sup>+</sup>16] J. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross. Exploring the many-body localization transition in two dimensions. *Science*, 352(6293):1547--1552, June 2016.
- [CKM19] E. Campbell, A. Khurana, and A. Montanaro. Applying quantum algorithms to constraint satisfaction problems. *Quantum*, 3:167, 2019.
- [CM09] B. Collins and S. Matsumoto. On some properties of orthogonal Weingarten functions. *Journal of Mathematical Physics*, 50(11):113516, 2009.
- [CM17] B. Collins and S. Matsumoto. Weingarten calculus via orthogonality relations: new applications. *arXiv:1701.04493*, 2017.
- [CS96] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54(2):1098, 1996.
- [CS06] B. Collins and P. Sniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Commun. Math. Phys.*, 264(3):773--795, 2006.
- [CW08] J. Clarke and F. K. Wilhelm. Superconducting quantum bits. *Nature*, 453(7198):1031--1042, 2008.
- [DCEL09] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, 2009.
- [DD06] T. DeGrand and C. DeTar. *Lattice methods for quantum chromodynamics*. World Scientific, 2006.
- [DF17] P. Diaconis and P. J. Forrester. Hurwitz and the origins of random matrix theory in mathematics. *Random Matrices: Theory and Applications*, 6(01):1730001, 2017.
- [DFG<sup>+</sup>21] A. Deshpande, B. Fefferman, A. V. Gorshkov, M. J. Gullans, P. Niroula, and O. Shtanko. Tight bounds on the convergence of noisy random circuits to uniform. *arXiv preprint arXiv:2112.00716*, 2021.



- [DGK83] P. Diaconis, R. L. Graham, and W. M. Kantor. The mathematics of perfect shuffles. *Advances in applied mathematics*, 4(2):175--196, 1983.
- [Dia96] P. Diaconis. The cutoff phenomenon in finite Markov chains. *PNAS*, 93(4):1659--1664, 1996.
- [DLTo2] D. P DiVincenzo, D. W. Leung, and B. M. Terhal. Quantum data hiding. *IEEE, Trans. Inf Theory*, 48:3580--599, 2002.
- [DS94] P. Diaconis and M. Shahshahani. On the eigenvalues of random matrices. *Journal of Applied Probability*, 31(A):49--62, 1994.
- [DSCoo] P. Diaconis and L. Saloff-Coste. Bounds for Kac's master equation. *Commun. Math. Phys.*, 209(3):729--755, 2000.
- [ECP10] J. Eisert, M. Cramer, and M. B. Plenio. Area laws for the entanglement entropy. *Rev. Mod. Phys.*, 82:277, 2010.
- [err] Error Correction Zoo. <https://errorcorrectionzoo.org/c/surface>.
- [FH] W. Fulton and J. Harris. *Representation theory*. Graduate Texts in Mathematics. Springer.
- [FKE<sup>+</sup>13] T. Fukuhara, A. Kantian, M. Endres, M. Cheneau, P. Schauß, S. Hild, D. Bellem, U. Schollwöck, T. Giamarchi, C. Gross, et al. Quantum dynamics of a mobile spin impurity. *Nat. Phys.*, 9(4):235, 2013.
- [FNW92] M. Fannes, B. Nachtergaele, and R. F. Werner. Finitely correlated states on quantum spin chains. *Commun. Math. Phys.*, 144(3):443--490, 1992.
- [GAE07] D. Gross, K. M. R. Audenaert, and J. Eisert. Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.*, 48:052104, 2007.
- [Gao15] J. Gao. Quantum union bounds for sequential projective measurements. *Phys. Rev. A*, 92(5):052331, 2015.
- [GE21] C. Gidney and M. Ekerå. How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits. *Quantum*, 5:433, 2021.
- [GFE09] D. Gross, S. T Flammia, and J. Eisert. Most quantum states are too entangled to be useful as computational resources. *Physical review letters*, 102(19):190501, 2009.
- [GGJN18] C. E. González-Guillén, M. Junge, and I. Nechita. On the spectral gap of random quantum channels. *arXiv preprint arXiv:1811.08847*, 2018.
- [GHS23] A. Gainutdinov, J. Haferkamp, and C. Schweigert. Davydov-Yetter cohomology, comonads and Ocneanu rigidity, 2023.
- [GKKR96] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.*, 68(1):13, 1996.

- [GKSS05] D. Gobert, C. Kollath, U. Schollwöck, and G. Schütz. Real-time dynamics in spin-1/2 chains with adaptive time-dependent density matrix renormalization group. *Phys. Rev. E*, 71(3):036102, 2005.
- [Gly10] D. G. Glynn. The permanent of a square matrix. *European Journal of Combinatorics*, 31(7):1887--1891, 2010.
- [GML<sup>+</sup>11] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner. Continuous-time monte carlo methods for quantum impurity models. *Rev. Mod. Phys.*, 83(2):349, 2011.
- [GNW21] D. Gross, S. Nezami, and M. Walter. Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem and de Finetti representations. *Commun. Math. Phys.*, 385(3):1325--1393, 2021.
- [Got97] D. Gottesman. Stabilizer codes and quantum error correction. *quant-ph/9705052*, 1997.
- [Got98] D. Gottesman. The Heisenberg representation of quantum computers. *quant-ph/9807006*, 1998.
- [Gow96] W.T. Gowers. An almost  $m$ -wise independent random permutation of the cube. *Combinatorics, Probability and Computing*, 5(2):119--130, 1996.
- [GT05] R. M. Guralnick and P. H. Tiep. Decompositions of small tensor powers and Larsen’s conjecture. *Represent. Theory*, 9:138--208, 2005.
- [Guro5] L. Gurvits. On the complexity of mixed discriminants and related problems. In *International Symposium on Mathematical Foundations of Computer Science*, pages 447--458. Springer, 2005.
- [GWD17] X. Gao, S.-T. Wang, and L.-M. Duan. Quantum supremacy for simulating a translation-invariant ising spin model. *Phys. Rev. Lett.*, 118:040502, 2017.
- [Haf22] J. Haferkamp. Random quantum circuits are unitary  $t$ -designs in depth  $O(nt^{5+o(1)})$ . *Quantum*, 6:795, 2022.
- [Har13] A. W. Harrow. The church of the symmetric subspace. *arXiv:1308.6595*, 2013.
- [HBK21] T. Haug, K. Bharti, and M. S. Kim. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum*, 2(4):040309, 2021.
- [HBRE21] J. Haferkamp, C. Bertoni, I. Roth, and J. Eisert. Emergent statistical mechanics from properties of disordered random matrix product states. *Phys. Rev. X Q.*, 2:040308, 2021.
- [HBV21] J. Haferkamp and J. Bermejo-Vega. Equivalence of contextuality and Wigner function negativity in continuous-variable quantum optics. *arXiv:2112.14788*, 2021.
- [HBVSE18] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert. Anticoncentration theorems for schemes showing a quantum speedup. *Quantum*, 2:65, May 2018.

- [HE22] D. Hangleiter and J. Eisert. Computational advantage of quantum random sampling. *arXiv preprint arXiv:2206.04079*, 2022.
- [HFK<sup>+</sup>22] J. Haferkamp, P. Faist, N. T. B. Kothakonda, J. Eisert, and N. Yunger Halpern. Linear growth of quantum circuit complexity. *Nat. Phys.*, 18:528--532, 2022.
- [HH08] M. B. Hastings and A. W. Harrow. Classical and quantum tensor product expanders. *arXiv preprint arXiv:0804.0011*, 2008.
- [HHB<sup>+</sup>19] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega. Closing gaps of a quantum advantage with short-time hamiltonian dynamics. *arXiv:1908.08069*, 2019.
- [HHB<sup>+</sup>20] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega. Closing gaps of a quantum advantage with short-time Hamiltonian dynamics. *Phys. Rev. Lett.*, 125:250501, 2020.
- [HHEG20] J. Haferkamp, D. Hangleiter, J. Eisert, and M. Gluza. Contracting projected entangled pair states is average-case hard. *Phys. Rev. Res.*, 2:013010, 2020. *arXiv:1810.00738*.
- [HHJ21] J. Haferkamp and N. Hunter-Jones. Improved spectral gaps for random quantum circuits: large local dimensions and all-to-all interactions. *Phys. Rev. A*, 104:022417, 2021.
- [HIN<sup>+</sup>21] M. Hinsche, M. Ioannou, A. Nietner, J. Haferkamp, Y. Quek, D. Hangleiter, J.-P. Seifert, J. Eisert, and R. Sweke. Learnability of the output distributions of local quantum circuits. *arXiv:2110.05517*, 2021.
- [HIN<sup>+</sup>22] M. Hinsche, M. Ioannou, A. Nietner, J. Haferkamp, Y. Quek, D. Hangleiter, J.-P. Seifert, J. Eisert, and R. Sweke. A single T-gate makes distribution learning hard. *arXiv:2207.03140*, 2022.
- [HJ19] N. Hunter-Jones. Unitary designs from statistical mechanics in random quantum circuits. *arXiv:1905.12053*, 2019.
- [HKSE17] D. Hangleiter, M. Kliesch, M. Schwarz, and J. Eisert. Direct certification of a class of quantum simulations. *Quantum Sci. Technol.*, 2(1):015004, 2017.
- [HLSW04] P. Hayden, D. Leung, P. W. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Commun. Math. Phys.*, 250(2):371--391, 2004.
- [HLW06] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. *Commun. Math. Phys.*, 265(1):95--117, 2006.
- [HM18] A. Harrow and S. Mehraban. Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates. *arXiv:1809.06957*, 2018.
- [HMMH<sup>+</sup>23] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth. Efficient unitary designs with a system-size independent number of non-Clifford gates. *Commun. Math. Phys.*, 397:995--1041, 2023.

- [HMMR05] S. Hoory, A. Magen, S. Myers, and C. Rackoff. Simple permutations mix well. *Theoretical computer science*, 348(2-3):251--261, 2005.
- [Hur63] A. Hurwitz. Ueber die Erzeugung der Invarianten durch Integration. In *Mathematische Werke*, pages 546--564. Springer, 1963.
- [Jia06] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *The Annals of Probability*, 34(4):1497--1529, 2006.
- [JW22] Y. Jia and M. M. Wolf. Hay from the haystack: explicit examples of exponential quantum circuit complexity. *arXiv:2205.06977*, 2022.
- [Kac47] M. Kac. Random walk and the theory of Brownian motion. *The American Mathematical monthlly*, 54(7P1):369--391, 1947.
- [KBM12] C. Karrasch, J. H. Bardarson, and J. E. Moore. Finite-temperature dynamical density matrix renormalization group and the drude weight of spin-1/2 chains. *Phys. Rev. Lett.*, 108(22):227206, 2012.
- [KKEG19] M. Kliesch, R. Kueng, J. Eisert, and D. Gross. Guaranteed recovery of quantum processes from few measurements. *Quantum*, 3:171, 2019.
- [KKT20] V. Kocharovsky, V. Kocharovsky, and S. Tarasov. Unification of the nature's complexities via a matrix permanent—critical phenomena, fractals, quantum computing, #P-complexity. *Entropy*, 22(3):322, 2020.
- [KMM22] Y. Kondo, R. Mori, and R. Movassagh. Quantum supremacy and hardness of estimating output probabilities of quantum circuits. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1296--1307. IEEE, 2022.
- [Kna88] S. Knabe. Energy gaps and elementary excitations for certain vbs-quantum antiferromagnets. *Journal of statistical physics*, 52(3-4):627--638, 1988.
- [Kro22] H. Krovi. Average-case hardness of estimating probabilities of random quantum circuits with a linear scaling in the error exponent. *arXiv:2206.05642*, 2022.
- [LCB14] M. Lubasch, J. I. Cirac, and M.-C. Banuls. Unifying projected entangled pair states contractions. *New J. Phys.*, 16:033014, 2014.
- [Li22] Z. Li. Short proofs of linear growth of quantum circuit complexity. *arXiv:2205.05668*, 2022.
- [Lip91] R. Lipton. New directions in testing. *Distributed computing and cryptography*, pages 191--202, 1991.
- [Low10] R. A. Low. Pseudo-randomness and Learning in Quantum Computation. *arXiv:1006.5227*, 2010.
- [LPS86] A. Lubotzky, R. Phillips, and P. Sarnak. Hecke operators and distributing points on the sphere I. *Communications on Pure and Applied Mathematics*, 39(S1):S149--S186, 1986.

- [LPS87] A. Lubotzky, R. Phillips, and P. Sarnak. Hecke operators and distributing points on  $S_2$ . II. *Communications on Pure and Applied Mathematics*, 40(4):401--420, 1987.
- [Mas03] D. K. Maslen. The eigenvalues of Kac's master equation. *Mathematische Zeitschrift*, 243(2):291--331, 2003.
- [MB17] R. L. Mann and M. J. Bremner. On the Complexity of Random Quantum Computations and the Jones Polynomial. *arXiv:1711.00686*, 2017.
- [MGDM18] R. Mezher, J. Ghalbouni, J. Dgheim, and D. Markham. Efficient quantum pseudorandomness with simple graph states. *Phys. Rev. A*, 97:022333, 2018.
- [MGDM19] R. Mezher, J. Ghalbouni, J. Dgheim, and D. Markham. Efficient approximate unitary t-designs from partially invertible universal sets and their application to quantum speedup. *1905.01504*, 2019.
- [MSH<sup>+</sup>12] W. Metzner, M. Salmhofer, C. Honerkamp, V. Meden, and K. Schönhammer. Functional renormalization group approach to correlated fermion systems. *Rev. Mod. Phys.*, 84(1):299, 2012.
- [MSM17] J. Miller, S. Sanders, and A. Miyake. Quantum supremacy in constant-time measurement-based computation: A unified architecture for sampling and verification. *Phys. Rev. A*, 96:062320, 2017.
- [MWNMo7] S. R. Manmana, S. Wessel, R. M. Noack, and A. Muramatsu. Strongly correlated fermions after a quantum quench. *Phys. Rev. Lett.*, 98(21):210405, 2007.
- [Nac96] B. Nachtergaele. The spectral gap for some spin chains with discrete symmetry breaking. *Commun. Math. Phys.*, 175(3):565--606, 1996.
- [NB09] Y. V. Nazarov and Y. M. Blanter. *Quantum transport: introduction to nanoscience*. Cambridge University Press, 2009.
- [Nez21] S. Nezami. Permanent of random matrices from representation theory: moments, numerics, concentration, and comments on hardness of boson-sampling. *arXiv:2104.06423*, 2021.
- [NS21] I. Nechita and S. Singh. A graphical calculus for integration over random diagonal unitary matrices. *Linear Algebra and its Applications*, 613:46--86, 2021.
- [OBK<sup>+</sup>17] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert. Mixing Properties of Stochastic Quantum Hamiltonians. *Commun. Math. Phys.*, 355(3):905--947, November 2017.
- [Orú14] R. Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117--158, 2014.
- [OSH21] M. Oszmaniec, A. Sawicki, and M. Horodecki. Epsilon-nets, unitary designs and random quantum circuits. *IEEE Transactions on Information Theory*, 2021.
- [Pag93] D. N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71(9):1291, 1993.

- [PCZ21] F. Pan, K. Chen, and P. Zhang. Solving the sampling problem of the sycamore quantum supremacy circuits. *arXiv:2111.03011*, 2021.
- [PGVWC07] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Inf. Comput.*, 7:401, 2007.
- [Pir22] L. Piroli. Random circuits have no shortcuts. *Nat. Phys.*, 18(5):482--483, 2022.
- [Pre98] J. Preskill. Lecture notes for physics 229: Quantum information and computation. *California Institute of Technology*, 16(1):1--8, 1998.
- [Pre13] J. Preskill. Quantum computing and the entanglement frontier. *Bulletin of the American Physical Society*, 58, 2013.
- [pri] private communication.
- [Pro11] T. Prosen. Open xxz spin chain: Nonequilibrium steady state and a strict bound on ballistic transport. *Phys. Rev. Lett.*, 106:217206, May 2011.
- [PTBO10] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa. Entanglement spectrum of a topological phase in one dimension. *Phys. Rev. B*, 81:064439, Feb 2010.
- [Rai98] E. M. Rains. Increasing subsequences and the classical groups. *Electronic Journal of Combinatorics*, 5:Art--No, 1998.
- [Rau05] R. Raussendorf. Quantum computation via translation-invariant operations on a chain of qubits. *Physical Review A*, 72(5), November 2005. *arXiv: quant-ph/0505122*.
- [RSB<sup>+</sup>13] J. P. Ronzheimer, M. Schreiber, S. Braun, S. S. Hodgman, S. Langer, I. P. McCulloch, F. Heidrich-Meisner, I. Bloch, and U. Schneider. Expansion dynamics of interacting bosons in homogeneous lattices in one and two dimensions. *Phys. Rev. Lett.*, 110(20):205301, 2013.
- [RTP<sup>+</sup>17] S.-J. Ran, E. Tirrito, C. Peng, X. Chen, G. Su, and M. Lewenstein. Review of tensor network contraction approaches. *arXiv:1708.09213*, 2017.
- [RW20] A. Rolandi and H. Wilming. Extensive Rényi entropies in matrix product states. *arXiv:2008.11764*, 2020.
- [RWS<sup>+</sup>17] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer. Elucidating reaction mechanisms on quantum computers. *Proc. Natl. Ac. Sc.*, 114(29):7555--7560, 2017.
- [Rys63] H. J. Ryser. *Combinatorial mathematics*, volume 14. American Mathematical Soc., 1963.
- [SBE17] M. Schwarz, O. Buerschaper, and J. Eisert. Approximating local observables on projected entangled pair states. *Phys. Rev. A*, 95:060102(R), 2017.
- [Scho5] U. Schollwöck. The density-matrix renormalization group. *Rev. Mod. Phys.*, 77(1):259, 2005.

- [Sch11] U. Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Ann. Phys.*, 326(1):96--192, 2011.
- [Sha49] C. E. Shannon. The synthesis of two-terminal switching circuits. *The Bell System Technical Journal*, 28(1):59--98, 1949.
- [Sho99a] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303--332, 1999.
- [Sho99b] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Statist. Comput.*, 41(2):303--332, 1999.
- [SR95] J. Sánchez-Ruiz. Simple proof of Page's conjecture on the average entropy of a subsystem. *Physical Review E*, 52(5):5653, 1995.
- [SS14] D. Stanford and L. Susskind. Complexity and shock wave geometries. *Phys. Rev. D*, 90:126007, 2014.
- [Ste96] A. M. Steane. Error correcting codes in quantum theory. *Phys. Rev. Lett.*, 77(5):793, 1996.
- [Sus16] L. Susskind. Computational complexity and black hole horizons. *Fort. Phys.*, 64:24-43, 2016.
- [Sus18] L. Susskind. Black holes and complexity classes. *arXiv:1802.02175*, 2018.
- [SVW08] D. M. Schlingemann, H. Vogts, and R. F. Werner. On the structure of clifford quantum cellular automata. *Journal of Mathematical Physics*, 49(11):112104, 2008.
- [SWVC07] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac. Computational complexity of projected entangled pair states. *Phys. Rev. Lett.*, 98:140506, 2007.
- [TCF<sup>+</sup>12] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch. Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional bose gas. *Nat. Phys.*, 8(4):325, 2012.
- [Val79] L. G. Valiant. The complexity of computing the permanent. *Theoretical computer science*, 8(2):189--201, 1979.
- [Val02] L. G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comp.*, 31:1229--1254, 2002.
- [Var13] P. Varjú. Random walks in compact groups. *Documenta Mathematica*, 18:1137--1175, 2013.
- [VC06] F. Verstraete and J. I. Cirac. Matrix product states represent ground states faithfully. *Phys. Rev. B*, 73(9):094423, 2006.
- [VGRC04] F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac. Matrix product density operators: Simulation of finite-temperature and dissipative systems. *Phys. Rev. Lett.*, 93(20):207204, 2004.

- [Vido3] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91(14):147902, 2003.
- [VMCo8] F. Verstraete, V. Murg, and J. I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Adv. Phys.*, 57(2):143--224, 2008.
- [VSD86] A. Vergis, K. Steiglitz, and B. Dickinson. The complexity of analog computation. *Mathematics and computers in simulation*, 28(2):91--113, 1986.
- [VWPGCo6] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac. Criticality, the area law, and the computational power of projected entangled pair states. *Phys. Rev. Lett.*, 96(22):220601, 2006.
- [Wat18] J. Watrous. *The theory of quantum information*. Cambridge University Press, 2018.
- [Web05] F. Weber. Strange quark matter and compact stars. *Progress in Particle and Nuclear Physics*, 54(1):193--288, 2005.
- [Web15] Z. Webb. The Clifford group forms a unitary 3-design. 2015. arXiv:1510.02769.
- [WGRE19] H. Wilming, M. Goihl, I. Roth, and J. Eisert. Entanglement-ergodic quantum systems equilibrate exponentially well. *Phys. Rev. Lett.*, 123(20):200604, 2019.
- [WHL<sup>+</sup>17] H. Wang, Y. He, Y.-H. Li, Z.-E. Su, B. Li, H.-L. Huang, X. Ding, M.-C. Chen, C. Liu, J. Qin, et al. High-efficiency multiphoton boson sampling. *Nature Photonics*, 11(6):361, 2017.
- [Wic50] G.-C. Wick. The evaluation of the collision matrix. *Phys. Rev.*, 80(2):268, 1950.
- [YKH<sup>+</sup>22] N. Yunger Halpern, N. B. T. Kothakonda, J. Haferkamp, A. Munson, J. Eisert, and P. Faist. Resource theory of quantum uncomplexity. *Phys. Rev. A*, 106:062417, 2022.
- [Zhu17] H. Zhu. Multiqubit Clifford groups are unitary 3-designs. *Phys. Rev. A*, 96:062336, Dec 2017.
- [ZKGG16] H. Zhu, R. Kueng, M. Grassl, and D. Gross. The Clifford group fails gracefully to be a unitary 4-design. 2016. arXiv: 1609.08172.
- [ZN19] T. Zhou and A. Nahum. Emergent statistical mechanics of entanglement in random unitary circuits. *Physical Review B*, 99(17):174205, 2019.
- [ZPH<sup>+</sup>17] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature*, 551:601--604, 2017.

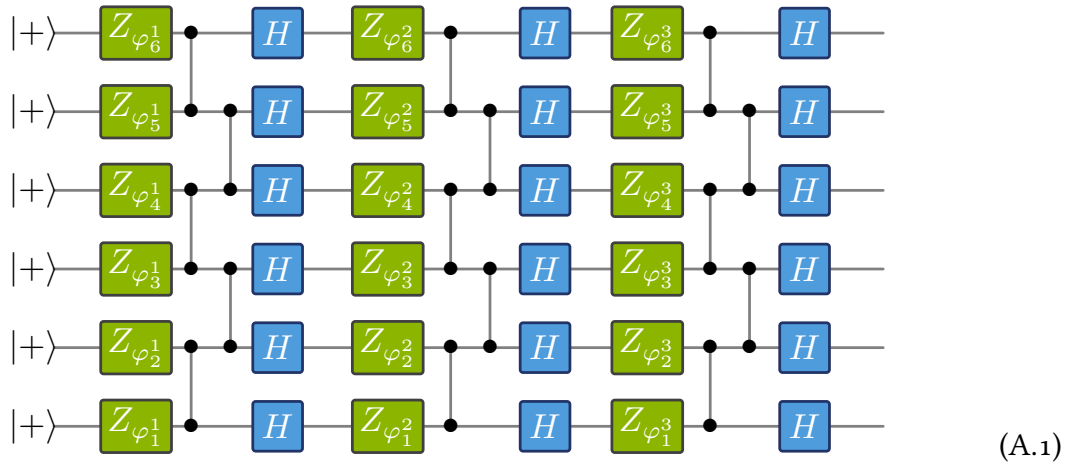




## ADDITIONAL RESULTS

### A.1 EFFICIENT QUANTUM PSEUDORANDOMNESS FROM QUANTUM SIMULATORS

Recall from Chapter 4 that measuring the 2D cluster states in uniformly random X-Y plane bases results in the following effective circuit acting on the last column [HHB<sup>+</sup>20, BVHS<sup>+</sup>18]



Here, global entangling Clifford unitaries

$$E := \left( \prod_{i=n}^1 H_i \right) \left( \prod_{i=1}^{n-1} CZ_{i,i+1} \right), \quad (\text{A.2})$$

are interleaved with a tower of single qubit Z rotations:  $\prod_{i=1}^n e^{i\beta_i^1 Z_i}$ .

**Theorem 20.** *The circuit in (A.1) forms an  $\varepsilon$ -approximate unitary  $t$ -design in depth*

$$m = O(nt(t^{4+o(1)}n + \log(1/\varepsilon))).$$

Notice that each layer consists of  $O(n)$  many gates, so the overall gate count scales as  $O(n^3)$  in the system-size. This is in contrast to the result in [HHB<sup>+</sup>20], where the number of required gates scales as  $O(n^2)$  but only for  $t = 2$ . It was already proven in [Rau05] that this constitutes a universal model of quantum computing and in fact, any unitary  $U \in U(2^n)$  can be approximated by a long enough circuit of the above form. This insight is based on the properties of the unitary  $E$  viewed as a Clifford cellular automaton [SVW08]. We can commute the rotation  $e^{i\beta_i^1 Z_i}$  through the circuit and obtain more complicated and, in particular, entanglement generating unitaries. The effect can be observed by the action of  $E$  on  $Z_i := \mathbb{1}_{[1,\dots,i-1]} \otimes Z \otimes \mathbb{1}_{[i+1,n]}$  with  $i$  somewhere in the bulk (i.e.  $i$  is far enough away from 1 and  $n$ ):

$$EZ_i E^\dagger = X_i, \quad E^2 Z_i (E^\dagger)^2 = X_{i-1} Z_i X_{i+1}, \quad E^3 Z_i (E^\dagger)^3 = X_{i-2} Z_{i-1} X_i Z_{i+1} X_{i+2}. \quad (\text{A.3})$$

This expansion carries on until one of the non-trivial Pauli operators is one of the boundary qubits 1 or  $n$ . Then, the process turns into a shift. This is best seen by starting with the operator  $Z_1$ :

$$EZ_1E^\dagger = X_1, \quad E^2Z_1(E^\dagger)^2 = Z_1X_2 \quad E^3Z_1(E^\dagger)^3 = Z_2X_3. \quad (\text{A.4})$$

It is this process in particular that we will exploit to prove the generation of unitary  $t$ -designs.

We consider a random quantum circuit of the form A.1 on  $n$  qubits with  $n$  layers and denote the corresponding probability measure on the unitary group by  $\nu$ . The corresponding moment operator is of the following form

$$M_t(\nu) = \left( E^{\otimes t,t} \prod_{i=1}^n P_t^{Z_i} \right)^n, \quad (\text{A.5})$$

where we define for any Pauli word  $p$

$$P_t^p := \frac{1}{2\pi} \int_0^{2\pi} (e^{i\varphi p})^{\otimes t,t} d\varphi. \quad (\text{A.6})$$

We successively commute all operators  $e^{\beta_i^j}$  from their position in the circuit to the right of the entangling unitaries  $E$ . This is equivalent to reformulating the moment operator as

$$M_t(\nu) = (E^n)^{\otimes t,t} P_t^{p_1} \dots P_t^{p_{n^2}}, \quad (\text{A.7})$$

with suitable Pauli words  $p_i$ . Since  $E^{\otimes t,t}$  is unitary, we have that

$$\|M_t(\nu) - M_t(\mu_H)\|_\infty = \|P_t^{p_1} \dots P_t^{p_{n^2}} - M_t(\mu_H)\|_\infty. \quad (\text{A.8})$$

The latter is the spectral gap of a product of projectors and we can apply the detectability lemma [AAV16,AALV09] to obtain the bound:

$$\|P_t^{p_1} \dots P_t^{p_{n^2}} - M_t(\mu_H)\|_\infty \leq \frac{1}{\sqrt{\frac{\Delta(H_{n,t})}{9} + 1}} \quad (\text{A.9})$$

for the Hamiltonian

$$H_{n,t} = \sum_{i=1}^{n^2} \mathbb{1} - P_t^{p_i}. \quad (\text{A.10})$$

Each of these summands is a positive operator. Moreover,  $H_{n,t}$  is frustration-free, i.e. it has global ground space with eigenvalue 0. We know from (A.4) that the Pauli words  $Z_i \otimes X_{i+1}$  for all  $1 \leq i \leq n-1$  and  $Z_i, X_i$  for  $1 \leq i \leq n$  among the Pauli words  $p_1, \dots, p_{n^2}$ . This implies the operator inequality

$$\begin{aligned} H_{n,t} &\geq \sum_{j=1}^n (\mathbb{1} - P_t^{Z_j}) + \sum_{j=1}^n (\mathbb{1} - P_t^{X_j}) + \sum_{j=1}^{n-1} (\mathbb{1} - P_t^{Z_j X_{j+1}}) \\ &\geq \frac{1}{2} \left( \sum_{i=1}^{n-1} h_{i,i+1} \right), \end{aligned}$$

where

$$h_{i,i+1} := 5\mathbb{1} - p_t^{Z_i} - p_t^{X_i} - p_t^{Z_{i+1}} - p_t^{X_{i+1}} - p_t^{Z_i X_{i+1}}. \quad (\text{A.11})$$

In particular, up to normalization (by a factor of  $1/5$ ) the summands  $h_{i,i+1}$  are moment operators corresponding to the distribution  $\sigma$  of picking with equal probability a gate  $e^{i\varphi_i^X X_i}, e^{i\varphi_i^Z Z}, \dots$  with equal probability. Since the gates  $e^{i\varphi_i^X X_i}, e^{i\varphi_i^Z Z}$  are universal on  $\text{SU}(2)$  and  $e^{i\varphi_i^{ZX} Z_i X_{i+1}}$  is entangling for almost all values of  $\varphi_i^{ZX} \in [0, 2\pi)$ , it follows from the result in [BB01] that this is a universal gate set and hence  $\sigma$  a universal distributions. For universal distributions it was proven in Refs [Var13, OSH21, HMMH<sup>+</sup>23]) that

$$\|M_t(\sigma) - M_t(\mu_H)\|_\infty \leq 1 - c' \log^{-2}(t) \quad (\text{A.12})$$

for a universal constant  $c' > 0$ . This local spectral gap can be elevated to a global spectral gap [BaHH16, OBK<sup>+</sup>17]:

$$\Delta(H_{n,t}) \geq (1 - \|M_t(\sigma) - M_t(\mu_H)\|_\infty) \Delta(H_{n,t}^H), \quad (\text{A.13})$$

where

$$H_{n,t}^H = \sum_{i=1}^{n-1} \mathbb{1} - M_t(\mu_{H,i,i+1}) \quad (\text{A.14})$$

is the local Hamiltonian corresponding to a locally Haar-random walk (compare [BaHH16]). We can combine this bound with the following bound from [Haf22]:

$$\Delta(H_{n,t}^H) \geq c_2 \times t^{-(4+o(1))}. \quad (\text{A.15})$$

Overall, we obtain a bound of

$$\Delta(H_{n,t}) \geq c_3 \log^{-2}(t) t^{-(4+o(1))}. \quad (\text{A.16})$$

Plugging this into (A.9) yields

$$\|M_t(\nu) - M_t(\mu_H)\|_\infty \leq 1 - c \log^{-2}(t) t^{-(4+o(1))}. \quad (\text{A.17})$$

Theorem 20 now follows from a standard argument [BaHH16] (that can also be found in Section 3.2).

## A.2 INCOMPRESSIBILITY OF PARAMETERIZED QUANTUM CIRCUITS

The effective circuits in (A.1) can also be prepared on a quantum computer. As such, they are an example of parameterized quantum circuits [BLSF19]. These are candidates for emergent quantum technologies like quantum machine learning or variational quantum optimizer. In this context, a key concept is that of overparameterization [HBK21]: How many of the parameters are redundant for the preparation of states? The parameters in the case of (A.1) are the angles  $\varphi_j^i$ .

Here, we observe that the techniques from Section 3.3 can be used to show that at least for certain parameterized circuits of linear depth, there is no overparameterization in this sense! The set of unitaries generated has a dimension precisely equal to the number of parameters:

**Proposition 5.** Denote by  $\mathcal{U}_T^{\text{par}}$  the set of unitaries that can be generated by circuits introduced in Eq. (A.1) of depth  $T \leq n$ . Then,  $\mathcal{U}_T^{\text{par}}$  is a semialgebraic set with

$$\dim \mathcal{U}_T^{\text{par}} = nT. \quad (\text{A.18})$$

*Proof.* First, notice that  $\mathcal{U}_T^{\text{par}}$  is the image of the contraction map  $\text{F}^{\text{par}} : \mathcal{U}(1)^{\times nT} \rightarrow \mathcal{U}(2^n)$  that maps  $e^{\theta i} \mapsto \begin{pmatrix} e^{\theta i} & 0 \\ 0 & e^{-\theta i} \end{pmatrix}$  and then contracts the circuits (A.1) with the above single-qubit diagonal matrices. This is a polynomial map applied to an algebraic set and hence, by the Tarski-Seidenberg principle [BCR13], the image is a semialgebraic set. For the dimension formula, we observe that again, by the same techniques as in Section 3.3, the maximally attainable rank among all choices of parameters equals  $\dim \mathcal{U}^{\text{par}}$ . To prove the second claim of Proposition 5 we only need to find  $nT$  linearly independent matrices in the image of the Jacobian. It turns out that it suffices to choose  $\varphi_j^i = 0$  for all angles. The partial derivatives are then of the form

$$E^{j-1} Z_i E^{T-j+1} = E^T P_{i,j}. \quad (\text{A.19})$$

From the pattern established in the previous section, it is clear that each tuple  $(i, j) \in [1, n] \times [1, T]$  corresponds to  $nT$  different Pauli operators.  $\square$

The same result is likely true for even deeper circuits. As  $E^n = -\mathbb{1}$ , this would require us to choose non-trivial angles  $\varphi_j^i$ .

### A.3 BABY STEPS TOWARDS ANTICONCENTRATION OF BOSON SAMPLING

The permanent is one of the most studied matrix polynomials, second perhaps only to the determinant. In contrast to the determinant the permanent is notoriously difficult to compute [Val79, Aar11, AA13]. In particular, the #P-hardness of approximating the permanent is related to various other hardness results [KKT20] and is at the heart of the quantum advantage paradigm of boson sampling [AA13]. Specifically in the context of quantum advantages, one is often concerned with the permanent of random matrices such as submatrices of Haar random unitary matrices. A key conjecture for quantum advantages is the anticoncentration conjecture. While the full conjecture remains out of reach, a recent preprint by S. Nezami [Nez21] makes substantial progress on the permanent of random matrices. In particular, this new work provides formulas for various special cases, a powerful expansion in terms of irreducible representation and lower bounds for the higher moments of the permanent distribution.

The permanent can be interpreted as the probability amplitude of a state subject to non-interacting bosonic dynamics [AA13, Aar11]. In this section, we show that this quantum mechanical point of view allows for a simple derivation of this lower bounds. Moreover, along similar lines, we find a simple method to obtain non-trivial upper bounds on the moments. We show how this general method can yield a bound of  $O(t^{-\frac{1}{2}k})$  on the  $t^{\text{th}}$  moment of a  $k \times k$  minor.

We work over the unitary group  $\mathcal{U}(d)$  on  $d$  modes. Each element  $U \in \mathcal{U}(d)$  acts as a passive Gaussian transformation on the bosonic (symmetric) subspace of  $n$  particles with  $d$  modes  $S^n(\mathbb{C}^d) \subset (\mathbb{C}^d)^{\otimes n}$  as  $\mathcal{U}^{\otimes n}$ .

The key to the observation in this section is the following standard formulation of the permanent as a probability amplitude in quantum mechanics [AA13, Aar11]. Let  $M$  be the upper-left  $k \times k$  minor of a unitary matrix  $U \in U(d)$  with  $k \leq d$ . Then,

$$\text{perm}(M) = \langle \psi | U^{\otimes k} | \psi \rangle, \quad \text{where } |\psi\rangle := k!^{\frac{1}{2}} P_{\text{sym}} |1, \dots, k\rangle. \quad (\text{A.20})$$

For the moments of the permanent this yields:

$$\mathbb{E}_U |\text{perm}(M)|^{2t} = \mathbb{E}_U |\langle \psi |^{\otimes t} U^{\otimes kt} | \psi \rangle^{\otimes t}|^2 = \mathbb{E}_U \langle \psi |^{\otimes t} U^{\otimes kt} | \psi \rangle^{\otimes t} \langle \psi |^{\otimes t} (U^\dagger)^{\otimes kt} | \psi \rangle^{\otimes t}. \quad (\text{A.21})$$

We see that the unitaries  $U^{\otimes kt}$  form a subgroup of  $U(d^{kt})$ . The following elementary observation captures the intuition that overlaps are smaller in larger spaces on average:

**Observation 1.** *Let  $G \subseteq H \subseteq U(\mathcal{H})$  be Lie subgroups of the unitary group acting on a Hilbert space  $\mathcal{H}$ . Then, for  $\phi \in \mathcal{H}$ , we have*

$$\mathbb{E}_{U \sim \mu_G} |\langle \phi | U | \phi \rangle|^2 \geq \mathbb{E}_{U \sim \mu_H} |\langle \phi | U | \phi \rangle|^2, \quad (\text{A.22})$$

where  $\mu_{G,H}$  denote the uniform measure on  $G$  and  $H$  respectively.

*Proof.* We present two short proofs of this statement. The first is the following calculation. Here,  $\|\bullet\|_F$  denotes the Frobenius norm:

$$\begin{aligned} 0 &\leq \left\| \mathbb{E}_{U \sim \mu_H} U | \phi \rangle \langle \phi | U^\dagger - \mathbb{E}_{V \sim \mu_G} V | \phi \rangle \langle \phi | V^\dagger \right\|_F^2 \\ &= \text{Tr} \left[ \left( \mathbb{E}_{U \sim \mu_H} U | \phi \rangle \langle \phi | U^\dagger - \mathbb{E}_{V \sim \mu_G} V | \phi \rangle \langle \phi | V^\dagger \right)^2 \right] \\ &= \text{Tr} \left[ \mathbb{E}_{U, U' \sim \mu_H} U | \phi \rangle \langle \phi | U^\dagger U' | \phi \rangle \langle \phi | U'^\dagger - 2 \text{Tr} \left[ \mathbb{E}_{U \sim \mu_H} \mathbb{E}_{V \sim \mu_G} U | \phi \rangle \langle \phi | U^\dagger V | \phi \rangle \langle \phi | V^\dagger \right] \right. \\ &\quad \left. + \text{Tr} \left[ \mathbb{E}_{V, V' \sim \mu_G} V | \phi \rangle \langle \phi | V^\dagger V' | \phi \rangle \langle \phi | V'^\dagger \right] \right] \\ &= \mathbb{E}_{U \sim \mu_G} |\langle \phi | U | \phi \rangle|^2 - 2 \mathbb{E}_{U \sim \mu_H} |\langle \phi | U | \phi \rangle|^2 + \mathbb{E}_{U \sim \mu_H} |\langle \phi | U | \phi \rangle|^2 \\ &= \mathbb{E}_{U \sim \mu_G} |\langle \phi | U | \phi \rangle|^2 - \mathbb{E}_{U \sim \mu_H} |\langle \phi | U | \phi \rangle|^2, \end{aligned} \quad (\text{A.23})$$

where we have used repeatedly the left and right invariance of the measures  $\mu_G$  and  $\mu_H$ .

Alternatively, Eq. (A.22) follows directly from representation theory. Recall Lemma 1, which states that for every representation  $\rho$  of a group  $H$ ,

$$\mathbb{E}_{U \sim \mu_H} \rho(U) = P_{\text{triv}}, \quad (\text{A.24})$$

where  $P_{\text{triv}}$  is the orthogonal projector onto the trivial isotypic component of  $\rho$ . As every trivial representation is also trivial under the induced representation of a subgroup  $G$ , we immediately obtain the following operator inequality of projectors:

$$\mathbb{E}_{U \sim \mu_H} \rho(U) \leq \mathbb{E}_{U \sim \mu_G} \rho(U). \quad (\text{A.25})$$

This immediately implies Eq. (A.22) by choosing  $\rho(U) = U \otimes \bar{U}$ .  $\square$

By choosing the correct subgroups, this yields a short proof of the following result from [Nez21, Theorem 5.1]:

**Theorem 21.**

$$\mathbb{E}_{\mathbf{U}} |\text{perm}(\mathbf{M})|^{2t} \geq \binom{\binom{k+d-1}{d} + t - 1}{t}^{-1}. \quad (\text{A.26})$$

*Proof.* Observe that  $|\psi\rangle \in S^t[S^d(\mathbb{C}^d)]$ . We can choose  $G = \{\mathbf{U}^{\otimes kt}, \mathbf{U} \in \mathbf{U}(d)\}$  forming a subgroup of  $H = \{\mathbf{U}^{\otimes t}, \mathbf{U} \in \mathbf{U}(S^k(\mathbb{C}^d))\}$ . Hence,

$$\begin{aligned} \mathbb{E}_{\mathbf{U}} |\text{perm}(\mathbf{M})|^{2t} &\geq \mathbb{E}_{\mathbf{U} \sim \mu_H} \langle \psi |^{\otimes t} \mathbf{U}^{\otimes t} | \psi \rangle^{\otimes t} \langle \psi |^{\otimes t} (\mathbf{U}^\dagger)^{\otimes t} | \psi \rangle^{\otimes t} \\ &= \frac{\langle \psi |^{\otimes t} \mathbf{P}_{\text{sym}, \binom{k+d-1}{d}, t} | \psi \rangle^{\otimes t}}{\dim S^t(S^k(\mathbb{C}^d))} \\ &= \frac{1}{\dim S^t(S^k(\mathbb{C}^d))}, \end{aligned} \quad (\text{A.27})$$

which implies the result.  $\square$

Theorem 21 follows in Ref. [Nez21] as a consequence of a very general expansion formula. In the remainder of this section we apply Observation 1 to obtain *upper bounds* on the permanent. This is possible by choosing  $H = \{\mathbf{U}^{\otimes kt}, \mathbf{U} \in \mathbf{U}(d)\}$  and  $G$  to be a small enough subgroup such that the expectation values over it can be explicitly evaluated.

**Theorem 22** (Upper bound on moments of the permanent). *Let  $\mathbf{M}$  be a  $k \times k$  minor of a Haar random unitary  $\mathbf{U} \in \mathbf{U}(d)$ . Then,*

$$\mathbb{E}_{\mathbf{U}} |\text{perm}(\mathbf{M})|^{2t} \leq (\pi t)^{-\frac{1}{2} \lfloor d/2 \rfloor} \quad (\text{A.28})$$

*If additionally  $t \leq d$ , then we have the stronger inequality*

$$\mathbb{E}_{\mathbf{U}} |\text{perm}(\mathbf{M})|^{2t} \leq \min \left[ (\pi t)^{-\frac{1}{2} \lfloor k/2 \rfloor}, (\pi k)^{-\frac{1}{2} \lfloor t/2 \rfloor} \right]. \quad (\text{A.29})$$

*Proof.* We first consider the case  $k = d = 2$ . We obtain an interesting bound from considering diagonal matrices in the Hadamard eigenbasis. This is the following group:

$$G' = D_H := \{HXH, X = \text{diag}(x_1, x_2) \in \mathbf{U}(2)\}. \quad (\text{A.30})$$

A general element of this group has the form:

$$HXH = \frac{1}{2} \begin{pmatrix} x_1 + x_2 & x_1 - x_2 \\ x_1 - x_2 & x_1 + x_2 \end{pmatrix}. \quad (\text{A.31})$$

Hence we find

$$\text{perm}(HXH) = \frac{1}{2} (x_1^2 + x_2^2). \quad (\text{A.32})$$

This yields:

$$\mathbb{E} |\text{perm}(HXH)|^{2t} = \mathbb{E} \frac{1}{4^t} (x_1^2 + x_2^2)^t (\bar{x}_1^2 + \bar{x}_2^2)^t = \frac{1}{4^t} \sum_{j=0}^t \binom{t}{j}^2 = \frac{1}{4^t} \binom{2t}{t} \leq \frac{1}{\sqrt{\pi t}}, \quad (\text{A.33})$$

by Stirling's approximation.

Now consider a larger value of  $d$ . We can always choose a subgroup of  $G' \subset \mathbb{U}(d)$  by considering block diagonal matrices with  $k$  independent copies of  $\mathbb{U}(2)$  on the diagonal:

$$\begin{pmatrix} [\mathbb{U}(2)] & \cdots & \\ & \ddots & \\ & & \cdots & [\mathbb{U}(2)] \end{pmatrix}$$

The maximal number of such copies that fit into a  $k \times k$  minor is  $\lfloor k/2 \rfloor$ . As the permanent factorizes for block diagonal matrices, we immediately obtain the bound:

$$\mathbb{E}_{\mathbb{U}} |\text{perm}(M)|^{2t} \leq (\pi t)^{-\frac{1}{2} \lfloor k/2 \rfloor}. \quad (\text{A.34})$$

The stronger form (A.29) follows from the moment-size duality derived in [Nez21].  $\square$

A natural generalization of the above example for general  $d$  is to consider the discrete Fourier transform  $F_d$ , where  $F_2 = H$  and the group

$$G' = D_{F_d} := \left\{ F_d X F_d^\dagger, X = \text{diag}(x_1, \dots, x_d) \in \mathbb{U}(d) \right\}. \quad (\text{A.35})$$

The case  $d = 3$  yields

$$\text{perm}(F_3 X F_3^\dagger) = \frac{1}{9} \left[ 2(x_1^3 + x_2^3 + x_3^3) + 3x_1 x_2 x_3 \right]. \quad (\text{A.36})$$

This generalizes to an interesting family of polynomials. The bound (A.28) can likely be improved by choosing larger blocks and the moments of the permanent of random diagonal matrices in the Fourier basis of these blocks. Even better bounds can be expected from the latter problem in arbitrary dimensions without using block diagonal embeddings. Unfortunately, this might be a difficult combinatorial problem: The matrices in question are known as *circulant matrices*. Expanding the permanent of these as a polynomial in the eigenvalues can be linked to #P-hard problems [KKT20].

#### A.4 IS THERE AN ANALOGUE OF GURVIT'S ALGORITHM FOR TENSOR NETWORKS?

The analogy between the permanent and tensor networks goes beyond the #P average-case hardness. Both functions, the tensor network contraction and the permanent, are low-degree polynomials in the entries of the input tensors. In this section, we show further that we can define an analogue of formulas for the permanent as expectation values over binary random variables [Rys63, Gly10, AH12]. Such formulas immediately give rise to natural sampling algorithms [Gur05, AH12].

**Definition 10.** A tensor network  $T$  is a graph  $G = (V, E)$  together with a tensor  $T^v \in \bigotimes_{e \in N(v)} \mathbb{C}^{D_e}$  for every vertex  $v$ , where  $N(v) \subseteq E$  denotes the set of adjacent edges to  $v$ .

Contraction over all edges defines the number  $C(T)$ . We define the max-1-norm of a tensor network:



**Definition 11.** The max-1-norm of a tensor network  $T$  is defined by

$$\|T\|_{m,1} := \max_{v \in V} \|T^v\|_1, \quad (\text{A.37})$$

where  $\|T^v\|_1$  is the 1-norm of  $T^v$  as a vector.

**Theorem 23.** There is a randomized approximation scheme that, given a tensor network  $T$  outputs an estimate of  $C(T)$  that is within an additive error of  $\varepsilon \|T\|_{m,1}^{|V|}$  with high probability and runtime in

$$\mathcal{O}\left(|V|, \max_{e \in E} D_e, \varepsilon^{-2}\right).$$

For every edge  $e \in E$ , we draw a random string  $x^e \in \{-1, 1\}^{D_e}$ . Then, we define a smaller tensor network  $\tilde{T}(v, x)$  for every vertex  $v$ : The graph of this network is the vertex  $v$  together with the edges  $e \in N(v)$  that now connect  $v$  to new, artificial vertices  $v^e$ . We locate the tensor  $T^v$  at  $v$  and  $x^e$  at  $v^e$ . Next we define the estimators

$$\text{Est}(x) := \prod_{v \in V} C(\tilde{T}(x, v)). \quad (\text{A.38})$$

None of these equals the actual tensor network contraction, but we find the following analogue to Ryser’s formula for the permanent: The contraction of a tensor networks  $T$  can be expressed as

$$C(T) = \mathbb{E}_x(\text{Est}(x)). \quad (\text{A.39})$$

This follows from the fact that:

$$\mathbb{E}_x x_i^e x_j^e = \delta_{ij}. \quad (\text{A.40})$$

The algorithm to estimate  $C(T)$  is thus simply to randomly draw  $m = \mathcal{O}(\varepsilon^{-2})$  many  $x_1, \dots, x_m$  and then output the mean:

$$M := \frac{\text{Est}(x_1) + \dots + \text{Est}(x_m)}{m} \quad (\text{A.41})$$

The bound in Theorem 23 follows from a standard application of the Chernoff bound. This immediately implies  $|C(T)| \leq \|T\|_{m,1}^{|V|}$ . Hence, we have  $\lim_{n \rightarrow \infty} C(T_n) = 0$  for all families of tensor networks  $T_n$  with sufficiently (but constantly) small entries. However, a stronger inequality is available for tensor networks that don’t have “self-loops” in form of a generalization of the Cauchy-Schwarz inequality. More precisely, it was proven in Ref. [KKEG19] that for every tensor network  $T$  in which every edge has two distinct vertices, the contraction is upper bounded by the product of the vector 2-norms of the tensors and, in fact, we used this inequality in Section 5.2.

Therefore, while we can define an analogue to Gurvit’s algorithm for tensor networks, the runtime guarantees are only polynomial (for reasonable errors) in a regime where stronger bounds exist already and simply outputting 0 is a better estimator. Unfortunately, even in the regime of exponential runtimes, the 1-norm usually gives very weak guarantees. We can compare this to the analogous algorithm for the permanent [Gur05, AH12], which merely depends on the operator norm of the input tensor/matrix.