# Sparse Random Hamiltonians Are Quantumly Easy

Chi-Fang Chen,[1,2,*] Alexander M. Dalzell,[2,3] Mario Berta,[4,5] Fernando G. S. L. Brandão,[2,1] and Joel A. Tropp[6]

[1]*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California, USA*
[2]*AWS Center for Quantum Computing, Pasadena, California, USA*
[3]*California Institute of Technology, Pasadena, California, USA*
[4]*Institute for Quantum Information, RWTH Aachen University, Aachen, Germany*
[5]*Department of Computing, Imperial College London, London, United Kingdom*
[6]*Department of Computing and Mathematical Sciences, Caltech, Pasadena, California, USA*

(Received 24 June 2023; revised 11 December 2023; accepted 14 December 2023; published 9 February 2024)

A candidate application for quantum computers is to simulate the low-temperature properties of quantum systems. For this task, there is a well-studied quantum algorithm that performs quantum phase estimation on an initial trial state that has a non-negligible overlap with a low-energy state. However, it is notoriously hard to give theoretical guarantees that such a trial state can be prepared efficiently. Moreover, the heuristic proposals that are currently available, such as with adiabatic state preparation, appear insufficient in practical cases. This paper shows that, for most random sparse Hamiltonians, the maximally mixed state is a sufficiently good trial state, and phase estimation efficiently prepares states with energy arbitrarily close to the ground energy. Furthermore, any low-energy state must have non-negligible quantum circuit complexity, suggesting that low-energy states are classically nontrivial and phase estimation is the optimal method for preparing such states (up to polynomial factors). These statements hold for two models of random Hamiltonians: (i) a sum of random signed Pauli strings and (ii) a random signed $d$-sparse Hamiltonian. The main technical argument is based on some new results in nonasymptotic random matrix theory. In particular, a refined concentration bound for the spectral density is required to obtain complexity guarantees for these random Hamiltonians.

DOI: 10.1103/PhysRevX.14.011014

Subject Areas: Condensed Matter Physics, Quantum Information, Statistical Physics

## I. INTRODUCTION

What are the strengths of quantum computers? The earliest (and still most compelling) candidates are factoring [1] and simulation of quantum systems [2–4]. While Shor's celebrated quantum algorithm for factoring [1] settles the quantum complexity of factoring, the complexity of quantum simulation at low energies has not been resolved. Indeed, we know from complexity theory that the ground-energy problem for general local Hamiltonians is QMA-hard in the worst case [5], so we anticipate that preparing ground states is generally intractable, even for quantum computers. This worst-case hardness persists for systems with additional physical constraints, including nearest-neighbor interaction in 1D [6] and translation invariance [7]. Although these results send a pessimistic signal, they merely indicate that any *proof* that the ground-energy problem is quantumly easy must further constrain the class of Hamiltonians, or else it can apply only for typical instances. Indeed, one may construct random families of Hamiltonians (Sec. I A) in the hope that their average-case complexity might be more favorable than the worst case.

Aside from complexity theory, the problem of preparing low-energy states arises in efforts to apply quantum computers to computational chemistry (for example, see Ref. [8]) and to condensed matter physics. To prepare a state of sufficiently low energy on a quantum computer, which can be used, e.g., for understanding chemical reaction pathways, a proposed quantum algorithm simply runs phase estimation on an initial trial state [9–12]. This method is efficient if the initial state has a non-negligible overlap with a low-energy state. Although the phase estimation part of the algorithm is well understood, we have an incomplete understanding of the time required to prepare a good initial state. In fact, recent numerical tests [13] suggest that, for some chemical systems, easily preparable initial states may have exponentially small

*Corresponding author: achifchen@gmail.com

(in system size) overlap. Moreover, preparing states with good overlap using the adiabatic algorithm may take exponential time, significantly impacting the end-to-end performance of the proposed quantum algorithm.

The search for tasks that are easy for quantum computers, in quantum chemistry or otherwise, is often implicitly a quest for quantum advantage: Quantum computers can be particularly helpful if the task is also classically hard. Unfortunately, proving classical hardness is challenging, and many once-promising candidates for classically hard problems have now been dequantized. For example, under certain classical access models, recent progress eliminates exponential quantum advantage in low-rank linear algebra tasks [14–17]. Still, hope remains that the Hamiltonian low-energy problem could provide quantum advantage [18]. With these thoughts in mind, the guiding question of this work is the following: Is there a classically nontrivial Hamiltonian whose low-energy states are provably easy to prepare?

Our work argues in the affirmative. In particular, we will show that the textbook phase estimation method (discussed above) works well for preparing low-energy states of typical random sparse Hamiltonians. Meanwhile, the low-energy states must have a large quantum circuit complexity, so they are plausibly nontrivial for classical computers to simulate.

The paper is organized as follows. First, we review relevant classes of Hamiltonians (Sec. I A) before presenting the main Hamiltonian model and the main results (Sec. II). Our proof strategy (Sec. III) exploits tools from nonasymptotic random matrix theory; Sec. IV contains further details and context. Last, we discuss the *classical* complexity of the low-energy problem, and we lay out future research directions in the search for quantum advantage (Sec. V). The remaining part of the work begins with proofs for the comparison principle (Appendix A), including the moments and the resolvent. We instantiate the nonasymptotic properties of GUE in Appendix B. The comparison results and GUE properties altogether allow us to calculate the properties of the Pauli string ensemble (Appendix C). In Appendix D, we prove the circuit-size lower bounds for the Pauli string ensemble, whose argument is independent of the comparison principle. Appendix E contains the missing proofs. Appendix F contains an argument for why interpolation methods do not immediately exploit higher matching moments.

### A. Related models

Before we give a statement of our main results, let us discuss how some familiar models fall short of answering our question. [19] We focus on random ensembles where quantitative statements are available.

(i) *Few-body Pauli models* [21]. As a natural generalization of the classical spin glasses (e.g., the Sherrington-Kirkpatrick (SK) model [22]), one replaces classical (commuting) constraints with noncommuting Pauli operators. An example is the ensemble of Hamiltonians given by

$$\boldsymbol{H} \coloneqq \sum_{i>j} g_{ij}^x \boldsymbol{\sigma}_i^x \boldsymbol{\sigma}_j^x + g_{ij}^y \boldsymbol{\sigma}_i^y \boldsymbol{\sigma}_j^y + g_{ij}^z \boldsymbol{\sigma}_i^z \boldsymbol{\sigma}_j^z, \quad \text{where } g_{ij}^x, g_{ij}^y, g_{ij}^z \sim \text{IID Gaussians}, \tag{1}$$

and $\boldsymbol{\sigma}_i^x, \boldsymbol{\sigma}_i^y, \boldsymbol{\sigma}_i^z$ denote the Pauli operators on qubit $i$. Heuristically, this model exhibits spin glass behavior at low temperatures [23], which suggests that finding low-energy states could be hard, even for quantum computers. [24] In the high-temperature regime, this model becomes classically easy: There exists an efficient algorithm that outputs a product state approximating the operator norm of the Hamiltonian $\|\boldsymbol{H}\|$ to a constant ratio [25]. We do not know whether there is a temperature range where the state remains quantumly easy but classically hard, nor do we know how to tackle this question.

(ii) *Sachdev-Ye-Kitaev (SYK) models* [26,27] *with fermionic degrees of freedom.* As an example, consider the four-body Hamiltonian

$$\boldsymbol{H}_{\text{SYK}} \coloneqq \sum_{i<j<k<\ell} g_{ijk\ell} \chi_i \chi_j \chi_k \chi_\ell \quad \text{where } \chi_i \chi_j + \chi_j \chi_i = 2\delta_{ij} \quad \text{and} \quad g_{ijk\ell} \sim \text{IID Gaussians}.$$

Nonrigorous arguments rooted in physics suggest that this model remains chaotic (instead of a spin glass) at very low temperatures [23,28,29]. If true, this is a strong hint that the SYK model is quantumly easy, and it gives a partial solution to our problem. [30] Unfortunately, it is challenging to sharpen the physics arguments into actual proofs. The only rigorous statement known to us is the recent work of Hastings and O'Donnell [33], which showed that a low-energy state with a *constant* ratio approximation of ground energy could be prepared by an efficient quantum algorithm. An extension of this result to arbitrarily low energies would also serve our question. Currently, we do not know of any analytical method suitable for the low-temperature regime of the SYK model. On the other hand, we are also unaware of any conclusive mathematical evidence that the low-temperature regime is classically hard; [34] in fact, finding a low-energy state with a constant ground-energy approximation for the *sparse* SYK model is known to be classically easy [35].

(iii) *Wigner's Gaussian unitary ensemble (GUE)* [36]. If we insist on provable models at low temperatures, we may consider the GUE with dimension $N = 2^n$, written in the Pauli string basis:

$$\boldsymbol{H}_{\mathrm{GUE}} := \sum_{\boldsymbol{\sigma}\in P} \frac{g_{\boldsymbol{\sigma}}}{N}\boldsymbol{\sigma} \quad \text{where } P := \{\mathbf{I}, \boldsymbol{\sigma}^x, \boldsymbol{\sigma}^y, \boldsymbol{\sigma}^z\}^{\otimes n}$$
$$\text{and} \quad g_{\boldsymbol{\sigma}} \sim \text{IID Gaussians.}$$

This nonlocal and nonsparse Hamiltonian seems unphysical, but it nevertheless served as a mathematical model for heavy nuclei [37]. Together with other random matrix ensembles (see Ref. [38] for a textbook introduction), the GUE provides a useful model for strongly interacting systems and for quantum information problems, thanks to its well-established properties [39]. As the Hamiltonian itself has exponentially many degrees of freedom, by a counting argument, the computational complexity (of low-energy state preparation or Hamiltonian evolution) is exponential $e^{\Omega(n)}$ with high probability for quantum computers; this is, at most, polynomially faster than exact diagonalization.

The ensemble we study in this work shares the properties of the few-body Pauli and SYK models, as it is sparse and instances can be efficiently specified. At the same time, like the GUE ensemble, we can accurately approximate the minimal energy and the density of states. Moreover, we show that polynomial-size quantum circuits are necessary and sufficient to generate low-energy states, up to arbitrarily good approximation ratios of the ground-state energy. The main downside of our ensemble is that, like the GUE ensemble, it is nonlocal, so it does not closely resemble the Hamiltonians that readily appear in nature.

## II. MAIN RESULTS

Now, let us present the model for which we will establish average-case quantum complexity for low-energy states. Consider an independent sum of *a few* random Pauli strings with random sign coefficients:

$$\boldsymbol{H}_{\mathrm{PS}} := \sum_{j=1}^{m} \frac{r_j}{\sqrt{m}}\boldsymbol{\sigma}_j \quad \text{where } \boldsymbol{\sigma} j \overset{\mathrm{IID}}{\sim} \{\mathbf{I}, \boldsymbol{\sigma}^x, \boldsymbol{\sigma}^y, \boldsymbol{\sigma}^z\}^{\otimes n}$$
$$\text{and} \quad r_j \overset{\mathrm{IID}}{\sim} \mathrm{UNIF}\{+1, -1\}. \tag{2}$$

The parameter $m$ will be polynomial in the number of qubits $n$, rather than exponential as in the GUE model.

Our main technical results show that its low-energy states enjoy two-sided bounds on circuit complexity.

*Theorem II.1.* (Low-energy states have low complexity). For any accuracy $\epsilon \geq 2^{-n/c_1}$, let $\boldsymbol{H}_{\mathrm{PS}}$ be drawn from the Pauli string ensemble (2) with

$$m = \left\lfloor c_2 \frac{n^5}{\epsilon^4} \right\rfloor$$

terms. Then, the following statement holds with probability of at least $1 - e^{-c_3 n^{1/3}}$ over a random draw $\boldsymbol{H}_{\mathrm{PS}}$ from the Pauli string ensemble. We can prepare a low-energy state $\boldsymbol{\rho}$ such that

$$\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) \leq \mathrm{Tr}[\boldsymbol{\rho}\boldsymbol{H}_{\mathrm{PS}}] \leq (1-\epsilon)\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) \tag{3}$$

using a circuit of size $G = \mathrm{Poly}(n, \epsilon^{-1})$. The quantities $c_1$, $c_2$, and $c_3$ are absolute constants, and $\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})$ denotes the smallest eigenvalue of $\boldsymbol{H}_{\mathrm{PS}}$ (which is typically negative).

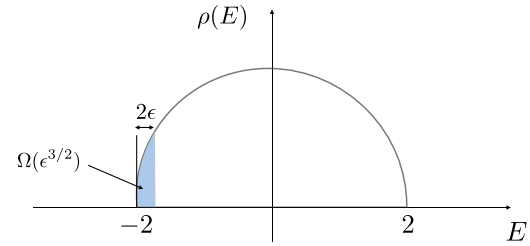


FIG. 1. Abundance of low-energy states. The contour illustrates the density $\rho$ versus the energy level $E$ for a semicircular distribution, which is (in the large dimension limit) the spectral distribution for the GUE. The semicircle spectral density implies the abundance of states near the ground energy. Performing phase estimation over the maximally mixed state gives a state with low energy $-2 + 2\epsilon$ with a decent probability $\Omega(\epsilon^{3/2})$.

See Appendix C 2 for the proof of Theorem II.1. Here, we elaborate on the interesting complexity theoretic aspects of this problem:

(i) *Arbitrarily good approximation of the ground energy.* For any polynomially small error $\epsilon \sim \mathrm{Poly}(n^{-1})$, there is a polynomially large choice $m = \mathrm{Poly}(n)$ for which a state with a $(1-\epsilon)$-ratio approximation of the ground energy (Fig. 1) can be prepared efficiently at gate complexity $G = \mathrm{Poly}(n)$. We hypothesize that the order of quantifiers can be exchanged, which would imply, for large enough $m = \mathrm{Poly}(n)$, that the low-energy states remain easy for any $\epsilon = \mathrm{Poly}(n^{-1})$. For further discussions, see point 2 in Sec. V.

(ii) *Phase estimation works.* As we will show, the quantum algorithm that produces the low-energy state $\boldsymbol{\rho}$ is very simple. Performing phase estimation over the maximally mixed state has a decent chance, at least $\Omega(\epsilon^{3/2})$, of returning a low-energy state obeying Eq. (3). A higher success probability is achieved via repeating the phase estimation step. The Hamiltonian simulation costs at most $\mathrm{Poly}(m, \epsilon^{-1})$ gates using off-the-shelf quantum simulation algorithms (e.g., Trotter [3] or qDRIFT [40]).

(iii) *End-to-end complexity.* This Hamiltonian problem is oracle-free and input-state-free, giving a complete picture. Further, the model description is entirely classical, and an instance can be generated using only $m(2n+1)$bits of randomness.

(iv) *Average case.* The statement holds with high probability over the Hamiltonian ensemble. Indeed, this model can produce an arbitrary local Hamiltonian in

the worst case, and we have no control over those instances.

(v) *Nonlocal, noncommuting Hamiltonians.* Most Pauli strings $\{\mathbf{I}, \boldsymbol{\sigma}^x, \boldsymbol{\sigma}^y, \boldsymbol{\sigma}^z\}^{\otimes n}$ act nontrivially on $\Theta(n)$ sites, and thus the Hamiltonian is nonlocal. The Hamiltonian is highly noncommutative since random Pauli strings anticommute with each other with probability $\frac{1}{2}$. Intuitively, the Pauli string ensemble is closer to a random matrix than to a local Hamiltonian.

(vi) *Sparse matrices.* From a linear algebra perspective, this model is a sparse, high-rank matrix (which has not been dequantized; see Sec. V). In general, a sparse matrix may not admit a simple Pauli decomposition; nevertheless, the same result extends to signed random $d$-sparse matrices (see Sec. IV B 1). However, the quantum easiness then requires access to a block encoding [41] of the Hamiltonian.

On the other hand, we argue that this problem is "very quantum" by proving a lower bound on the complexity of preparing low-energy states. As a disclaimer, we do not *prove* classical hardness for state preparation (see Sec. V), which is an intriguing open problem that we leave for future work.

*Theorem II.2.* (Small circuit gives bad energy) Fix a circuit architecture with $G$ two-qubit gates (e.g., 1D brickwork layout) with the initial state $|0\rangle$, and consider the family of all reachable states $\mathrm{Circ}(G)$. For any $\epsilon_1 \geq 0$, suppose $m \leq \epsilon_1^2 \times 2^n$. Then, with high probability over the random draw of the instance $\boldsymbol{H}_{\mathrm{PS}}$ from the Pauli string ensemble (2),

$$G = \tilde{o}(\epsilon_1 \sqrt{m}) \quad \text{implies} \quad \inf_{|\phi\rangle \in \mathrm{Circ}(G)} \langle \phi | \boldsymbol{H}_{\mathrm{PS}} | \phi \rangle \geq \epsilon_1 \times \mathbb{E}\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}).$$

Namely, all possible states $|\psi\rangle \in \mathrm{Circ}(G)$ parametrized by the circuit architecture fail to produce any low-energy state. The notation $\tilde{o}(\cdot)$ suppresses $\log(m)$ prefactors.

See Appendix D for the proof of Theorem II.2. In other words, we very often need a large circuit $G = \tilde{\Omega}(\epsilon_1 \sqrt{m})$ to describe the low-energy states; they are very entangled and far from product states. [42] Further, our circuit-size lower bound uses a direct counting argument, and it suggests the circuit should change over different random instances. Nevertheless, Theorem II.1 states the complementary result: An appropriate instance-dependent state can be prepared efficiently using the *simplest* quantum algorithms (Hamiltonian simulation and phase estimation).

The main caveat for our model is that it is nonlocal, unlike most physical Hamiltonians, and our argument is not immediately applicable to local Hamiltonians. Indeed, the spectral properties of the two types of models are different. As we will show, the Pauli string ensemble has a (compact) semicircular spectrum, while local Hamiltonians tend to have a tail in the spectrum. [45] Performing phase estimation with the maximally mixed state would not be able to access the low-energy states far in the low probability tail. Of course, we hope our results ultimately inspire a better understanding of preparing the low-energy states of local Hamiltonians. For further discussions, see point 3 in Sec. V.

Regardless, from a linear algebra and algorithm perspective, random sparse matrices are natural models to study. We emphasize that the main goal is to give a transparent toy model showcasing what quantum computers are good at, especially given recent developments in dequantization.

## III. PROOF IDEAS

Given a general, strongly interacting Hamiltonian, it seems daunting to control its behavior. However, we can make an exception in the case of certain random matrix ensembles where the matrices have predictable spectral properties. For example, it is well known that GUE matrices (see Appendix B) have a definite maximal eigenvalue and a semicircular spectral density $\rho(x)$:

$$\|\boldsymbol{H}_{\mathrm{GUE}}\| \approx 2 \quad \text{and} \quad \rho(x) \approx \frac{\sqrt{4 - x^2}}{2\pi}$$

(up to negligible fluctuation and with high probability).

See Fig. 1. Indeed, this fact alone hints that the low-energy states have a non-negligible density, independent of the system size:

$$\int_{-2}^{-2+2\epsilon} \rho(x) dx = \Theta(\epsilon^{3/2}).$$

In terms of complexity, directly running phase estimation on the maximally mixed state returns a low-energy state with decent probability: $\Omega(\epsilon^{3/2})$. The core of our argument is that the spectrum of the Pauli string ensemble (2) looks a lot like the spectrum of a GUE matrix (Fig. 2). As a consequence, it is also "easy" to find the low-energy states of the Pauli string ensemble.

How can we prove that the Pauli string ensemble also has a semicircular spectral density? The entire argument then boils down to a universality principle: The Pauli string ensemble, at moderately large $m$, mimics "smooth" properties of the GUE ensemble, including the maximum eigenvalue and the coarse-grained spectral density.

The mathematical argument is based on techniques from nonasymptotic random matrix theory (Sec. IV). Several novel results are required to address some of the particular challenges that arise in the quantum information problem.

Our first main result states that the trace polynomial moments of the Pauli string ensemble almost coincide with the corresponding moments of the GUE. In particular, by choosing a large enough moment, we can also compare the spectral norms of the two matrices. Throughout this work, we consider the normalized $p$-norms

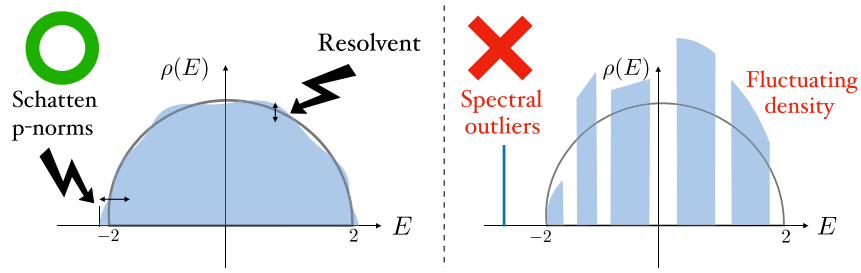


FIG. 2. Right panel: what could have gone wrong. For a generic matrix, the phase estimation strategy will not find a low-energy state if there are spectral outliers or if the spectral density gets too small near the ground energy. Left panel: almost a semicircle. For the Pauli string ensemble, we control the ground energy by Schatten $p$-norms and control the spectral density by the resolvent. Both values are comparable to the GUE, which has a favorable semicircular spectrum. Therefore, enough states remain near the ground energy, and thus phase estimation efficiently finds them.

$$\|\boldsymbol{O}\|_p := (\overline{\mathrm{Tr}}|\boldsymbol{O}|^p)^{1/p} \quad \text{and} \quad |||\boldsymbol{O}|||_p := (\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{O}|^p)^{1/p}$$
$$\text{where } \overline{\mathrm{Tr}} := \frac{1}{\mathrm{Tr}[\mathbf{I}]}\mathrm{Tr}. \tag{4}$$

We often denote $N := \mathrm{Tr}[\mathbf{I}] = 2^n$. We have the understanding that $|||\boldsymbol{O}|||_\infty := \text{ess sup}\,\|\boldsymbol{O}\|$.

*Theorem III.1.* ($p$-norms and operator norm) Let $p \in 2\mathbb{N}$ be an even natural number. The random Pauli string ensemble (2) satisfies the norm bound

$$\left| |||\boldsymbol{H}_{\mathrm{PS}}|||_p - |||\boldsymbol{H}_{\mathrm{GUE}}|||_p \right| \lesssim \left(\frac{p^{3/4}}{m^{1/4}} + \frac{p}{\sqrt{m}}\right)\left(1 + \frac{p^{3/4}}{2^{n/2}}\right). \tag{5}$$

The symbol $\lesssim$ suppresses constant factors. Furthermore, for $0 \le \epsilon \le 1/2$ and $m \le 2^{2n}$, there exist constants $c_1, c_2 > 0$ where

$$m \ge c_1 \frac{n^3}{\epsilon^4} \quad \text{ensures} \quad \Pr\left(\|\boldsymbol{H}_{\mathrm{PS}}\| \ge 2(1+\epsilon)\right) \le \exp(-c_2 n).$$

See Appendix C 1 for the proof of Theorem III.1. For a fixed moment $p$ that may depend on the number $n$ of sites, the right-hand side of Eq. (5) decays with the number $m$ of terms in the Hamiltonian. Applying Markov's inequality for $p = \Omega(\log(N)) = \Omega(n)$ and choosing $m = \mathrm{Poly}(n)$, we obtain a tail bound for the spectral norm.

Comparing the spectral densities of the two ensembles requires a more difficult argument. Ideally, we are interested in projectors to Hamiltonian eigenstates $|\phi\rangle\langle\phi|$. However, exact eigenstate projectors are tricky to handle. Instead, we consider the resolvent, which probes the "coarse-grained" energy projector at energies $\omega \pm \mathcal{O}(\eta)$. We define

$$\boldsymbol{R}_{\omega,\eta}(\boldsymbol{H}) := \frac{1}{\boldsymbol{H} - \omega + \mathrm{i}\eta} = \mathrm{i}\int_0^\infty \mathrm{e}^{\mathrm{i}(\boldsymbol{H}-\omega)t - \eta t}dt$$
$$\text{as a proxy for} \quad \frac{1}{\eta}\sum_{|\phi\rangle} \mathbb{1}\{|E(\phi) - \omega| \le \eta\} \times |\phi\rangle\langle\phi|, \tag{6}$$

where $\mathbb{1}$ is the indicator function. We often suppress parameter dependencies by writing $\boldsymbol{R} := \boldsymbol{R}_{\omega,\eta}(\boldsymbol{H})$. For intuition, the resolvent is diagonal in the Hamiltonian basis and spikes at energy $\omega$ with width $\mathcal{O}(\eta)$. See Fig. 3.

However, if we are especially interested in the states near certain energy $\omega$, the resolvent is not localized enough because the filter $E \mapsto 1/|E - \omega|$ decays too slowly as a function of the energy. [47] Instead, we can take the trace of resolvent powers so that the tail decays at the faster rate of about $|E - \omega|^{-p}$. In other words,

$$\frac{\eta^p}{N}\mathrm{Tr}|\boldsymbol{R}|^p = \frac{1}{N}\sum_E \frac{\eta^p}{|E - \omega + \mathrm{i}\eta|^p} \quad \text{as a proxy for}$$
$$\rho(\omega) \times \frac{\eta}{\sqrt{p}},$$

where the energy $\eta/\sqrt{p}$ is roughly the window where the weight $|\boldsymbol{R}|^p\eta^p$ remains large, $\Omega(1)$.

*Theorem III.2.* (Comparing the resolvent moments). Let $p \in 2\mathbb{N}$ be an even natural number. The resolvent (6) of the random Pauli string ensemble (2), written $\boldsymbol{R}_{\mathrm{PS}}$, can be compared with the resolvent $\boldsymbol{R}_{\mathrm{GUE}}$ of the GUE:

$$\left| |||\boldsymbol{R}_{\mathrm{PS}}|||_p - |||\boldsymbol{R}_{\mathrm{GUE}}|||_p \right| \lesssim \left(\frac{p^4}{\eta^5 m^2} + \frac{p^3}{\eta^5 m}\right)\left(1 + \frac{p^3}{2^{2n}}\right). \tag{7}$$

The symbol $\lesssim$ suppresses absolute constants.

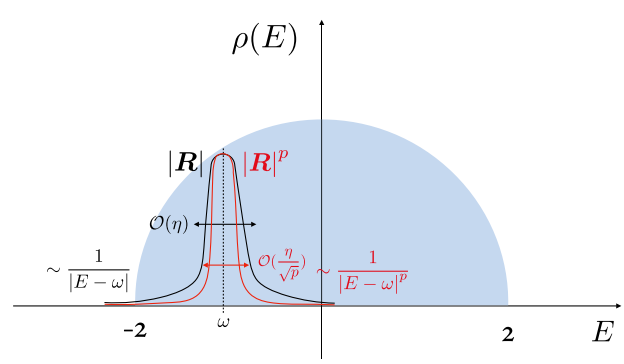


FIG. 3. Probing the spectrum by resolvents. The resolvent (black curve) centered at energy $\omega$ with resolution parameter $\eta$ filters out energies distant from $\omega$. Taking powers of the resolvent (red curve) focuses the filter on a narrower region around $\omega$.

See Appendix A 2 for the proof of Theorem III.2 in a more general setting. For moderately large $m$ (depending on the distance $\eta$ from the real line and the power $p$), the formula (7) controls the expected spectral density, filtered by the resolvent:

$$\mathbb{E}\mathrm{Tr}|\boldsymbol{R}_{\mathrm{PS}}|^p \approx \mathbb{E}\mathrm{Tr}|\boldsymbol{R}_{\mathrm{GUE}}|^p.$$

Since we want to make a statement that holds with high probability over realizations of the Pauli string ensemble, we also need to prove that the quantity $\mathrm{Tr}|\boldsymbol{R}|^p$ is concentrated near its expectation $\mathbb{E}\mathrm{Tr}|\boldsymbol{R}|^p$ (i.e., the spectral density does not fluctuate too much); see Theorem A.1. Last, since individual resolvents probe the local density, we may probe the *integrated* spectral density by placing consecutive resolvents. The abundance of low-energy states then implies that phase estimation has a decent chance of succeeding.

## IV. NEW RESULTS IN NONASYMPTOTIC RANDOM MATRIX THEORY

Our results for the Pauli string ensemble fall into the category of nonasymptotic universality laws for random matrices. This section provides some context for these results, as well as some details about the argument.

Asymptotic universality laws are among the celebrated classical achievements of random matrix theory (RMT). For example, Wigner showed that the semicircle law is the limiting spectral distribution of a (standardized) symmetric matrix with independent identically distributed (IID) Rademacher entries above the diagonal. The universality law for the Wigner matrix states that the detailed distribution of the entries does not affect the limiting spectral distribution, provided the first four moments are bounded. Subsequently, researchers obtained nonasymptotic comparisons between the spectrum of a Wigner-type matrix and the semicircle distribution. For surveys, see the monographs [48,49].

Our approach depends on a nonasymptotic comparison between the spectrum of the Pauli string ensemble (2) and a GUE matrix, whose spectral distribution approximately follows a semicircle law. This type of result does not fall within the scope of classical universality laws because the Pauli string ensemble barely has any randomness, let alone independent entries. To implement our program, we first observe that the low-order moments of the Pauli string ensemble match the low-order moments of a GUE matrix:

$$\mathbb{E}[\boldsymbol{H}_{\mathrm{PS}}] = \mathbb{E}[\boldsymbol{H}_{\mathrm{GUE}}] = 0;$$
$$\mathbb{E}[\boldsymbol{H}_{\mathrm{PS}} \otimes \boldsymbol{H}_{\mathrm{PS}}] = \mathbb{E}[\boldsymbol{H}_{\mathrm{GUE}} \otimes \boldsymbol{H}_{\mathrm{GUE}}] = \frac{1}{N^2}\sum_{\boldsymbol{\sigma}\in P}\boldsymbol{\sigma}\otimes\boldsymbol{\sigma};$$
$$\mathbb{E}[\boldsymbol{H}_{\mathrm{PS}} \otimes \boldsymbol{H}_{\mathrm{PS}} \otimes \boldsymbol{H}_{\mathrm{PS}}] = \mathbb{E}[\boldsymbol{H}_{\mathrm{GUE}} \otimes \boldsymbol{H}_{\mathrm{GUE}} \otimes \boldsymbol{H}_{\mathrm{GUE}}] = 0.$$

For a smooth statistic $f$ of the random matrices, we can take advantage of this coincidence by means of Lindeberg's exchange principle. Each of the random matrix models can be expressed as a sum of IID random matrices, and we can interpolate between the two models by swapping one summand at a time. At each step, we can control the change between the two models by expanding $f$ as a Taylor series to expose the polynomial moments. The terms in these expansions cancel through the third order, leaving a fourth-order error. Our argument is quite different from recent applications [50,51] of the Lindeberg principle in RMT.

In more detail, we consider two random Hermitian matrices $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$ that can be written as sums of independent, centered random matrices (all of the same dimension):

$$\boldsymbol{H} = \sum_{i=1}^{m}\boldsymbol{A}_i \quad \text{and} \quad \tilde{\boldsymbol{H}} = \sum_{i=1}^{m}\tilde{\boldsymbol{A}}_i. \tag{8}$$

Although less familiar than the classical random matrix ensembles, the independent sum model is much more flexible and has a wide scope of applicability; see Ref. [52] for examples. Suppose that the low-order polynomial moments of the summands match. In other words,

$$\mathbb{E}\boldsymbol{A}_i = \boldsymbol{0} \quad \text{and} \quad \mathbb{E}\boldsymbol{A}_i^{\otimes k} = \mathbb{E}\tilde{\boldsymbol{A}}_i^{\otimes k} \quad \text{for } k = 1,\ldots,t \quad \text{and} \quad i = 1,\ldots,m. \tag{9}$$

For example, the first three moments of a random Pauli string match the first three moments of a GUE matrix. More generally, constructive models in quantum information theory can match an arbitrary number $t$ of moments, similar to the case of a unitary $t$-design. Our work shows how to compare the spectral properties of models with many matching moments.

Our first universality result compares the trace polynomial moments of the two random matrices. These results allow us to control the spectral norm of the random matrices.

*Theorem IV.1.* (Universality for moments). Consider two families $(\boldsymbol{A}_i)$ and $(\tilde{\boldsymbol{A}}_i)$ of independent, random, Hermitian matrices whose moments match Eq. (9) up to order $t \geq 2$, and introduce the sums $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$ as in Eq. (8). Define the statistics

$$L_{p,k} := \left[\sum_{i=1}^{m}\left(|||\boldsymbol{A}_i|||_p^k + |||\tilde{\boldsymbol{A}}_i|||_p^k\right)\right]^{1/k} \quad \text{for } k \geq 1 \quad \text{and} \quad p \geq 2;$$
$$\sigma^2 := \sum_{i=1}^{m}\left(\|\mathbb{E}\boldsymbol{A}_i^2\| + \|\mathbb{E}\tilde{\boldsymbol{A}}_i^2\|\right).$$

Then, for each even natural number $p \in 2\mathbb{N}$, we have the bounds

$$\left| |||\boldsymbol{H}|||_p - |||\tilde{\boldsymbol{H}}|||_p \right| \leq \begin{cases} 2p^{t/(t+1)} \times L_{p,t+1} + 2p \times L_{p,p} \\ 2p^{t/(t+1)} \times (\sigma^2 L_{\infty,\infty}^{t-1})^{1/(t+1)} + 2p \times L_{p,p}. \end{cases} \tag{10}$$

The $p$-norm is defined in Eq. (4).

The proof of Theorem IV.1 appears in Appendix A 1. Theorem III.1 follows when we instantiate this result for the Pauli string ensemble (2) and the GUE.

We can obtain simpler versions of this result if we pass to the uniform bound $L_{p,\infty} = \max_i\{|||\boldsymbol{A}_i|||_p, |||\tilde{\boldsymbol{A}}_i|||_p\}$ on the summands. For example,

$$\left| |||\boldsymbol{H}|||_p - |||\tilde{\boldsymbol{H}}|||_p \right| \lesssim (1 + (m/p))^{1/(t+1)} \times pL_{p,\infty}.$$

Here, the symbol $\lesssim$ suppresses absolute constants only. Heuristically, we should consider $p \ll m$, so there are reductions in the error from matching more moments (i.e., increasing $t$).

Our second universality result provides a comparison for powers of the resolvents of independent sums. Define

$$\boldsymbol{R} := (\boldsymbol{H} - \omega + \mathrm{i}\eta)^{-1} \quad \text{and} \quad \tilde{\boldsymbol{R}} := (\tilde{\boldsymbol{H}} - \omega + \mathrm{i}\eta)^{-1} \quad \text{where } \omega \in \mathbb{R} \quad \text{and} \quad \eta > 0. \tag{11}$$

The random matrices $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$ are defined in Eq. (8).

*Theorem IV.2.* (Universality for resolvent moments). Instate the assumptions and notation of Theorem IV.1. For each even natural number $p \in 2\mathbb{N}$, the polynomial moments of the resolvent (11) are related by

$$\left| |||\boldsymbol{R}|||_p - |||\tilde{\boldsymbol{R}}|||_p \right| \lesssim \frac{1 + (m/p)}{\eta} \times \left( \frac{pL_{3p,\infty}}{\eta} \right)^{t+1}.$$

The symbol $\lesssim$ suppresses constants depending only on $t$.

See Appendix A 2 for the proof of Theorem IV.2. We obtain Theorem III.2 by instantiating the result for the Pauli string ensemble and the GUE.

The resolvent moment comparison (Theorem IV.2) is not sufficient to guarantee that a random realization $\boldsymbol{H}_{\mathrm{PS}}$ of the Pauli string ensemble places significant density on the low-energy states. To achieve this goal, we must also show that $|||\boldsymbol{R}_{\mathrm{PS}}|||_p$ is concentrated near its expected value. This claim requires a separate argument (Theorem A.1). The results on the concentration of trace moments of the resolvent are new.

### A. Related work

The field of RMT has historically focused on asymptotic limit laws for the spectral density of matrices from the classical ensembles (Wigner, Wishart, Jacobi, etc.). In this setting, there has also been a significant amount of research on rates of convergence, and some of these results can be interpreted as nonasymptotic universality laws. For example, see Bai and Silverstein (Chap. 8 in Ref. [48]).

In the last few years, researchers have recognized that the scope of the universality phenomenon extends well beyond the classical matrix ensembles. In particular, we have started to develop a deeper understanding of the independent sum model. Tropp obtained the first general result of this type [53]. His theory covers a sum of independent Gaussian random matrices, and it provides conditions under which the polynomial moments approximate the moments of the semicircle distribution. Building on Tropp's work, Bandeira *et al.* [54] developed a method for comparing a sum of independent Gaussian random matrices with a free probability model, which can capture a wider range of spectral distributions. With some effort, the techniques from these two papers can likely be applied to the Gaussian variant of the Pauli string ensemble (2) to obtain results similar to our main theorems.

The most immediate precedent for our work is a recent preprint by Brailovskaya and van Handel [55]. Their paper compares an independent sum $\boldsymbol{H}$ of random matrices with an independent sum $\boldsymbol{G}$ of Gaussian random matrices, where corresponding summands share the same mean and covariance:

$$\boldsymbol{H} = \sum_{i=1}^{m} \boldsymbol{A}_i \quad \text{and} \quad \boldsymbol{G} = \sum_{i=1}^{m} \tilde{\boldsymbol{A}}_i \quad \text{where } \mathbb{E}[\boldsymbol{A}_i] = \mathbb{E}[\tilde{\boldsymbol{A}}_i] \quad \text{and} \quad \mathbb{E}[\boldsymbol{A}_i \otimes \boldsymbol{A}_i] = \mathbb{E}[\tilde{\boldsymbol{A}}_i \otimes \tilde{\boldsymbol{A}}_i]. \tag{12}$$

The main result of the paper [55] provides conditions to guarantee that the two random matrix models have similar polynomial moments and polynomial resolvent moments.

*Theorem IV.3.* (Universality of moments and resolvent moments [55] ). Consider two random matrix models as in Eq. (12). Define the statistics

$$v := \left\| \sum_{i=1}^{m} \mathbb{E}\boldsymbol{A}_i^2 \right\| \quad \text{and} \quad L_\infty := \max_i \|\boldsymbol{A}_i\|.$$

Then, for every even natural number $p \in 2\mathbb{N}$, the polynomial moments and resolvents satisfy the bounds

$$\left| |||\boldsymbol{H}|||_p - |||\boldsymbol{G}|||_p \right| \lesssim (p^2 v L_\infty)^{1/3} + pL_\infty;$$
$$\left| |||\boldsymbol{R}_{\boldsymbol{H}}|||_p - |||\boldsymbol{R}_{\boldsymbol{G}}|||_p \right| \lesssim \frac{p^2 v L_\infty + p^3 L_\infty^3}{\eta^4}. \tag{13}$$

The $p$-norm is defined in Eq. (4), and the symbol $\lesssim$ suppresses absolute constants.

The proof of this result uses a version of Stein's method, inspired by Ref. [56]. The basic technique is to interpolate smoothly between the two random matrix models, preserving the second moments along the interpolation path. To control the derivative of a spectral function along the path, the authors use a cumulant expansion along with bounds on the higher derivatives of the function.

It is fruitful to compare the bounds (IV.1) and (IV.3). The variance parameter $v$ in Theorem IV.3 is never larger than the variance parameter $\sigma^2$ in Theorem IV.1 because the norm is inside the sum in $v$. The two quantities $\sigma^2$ and $v$ coincide for IID sums, but they can differ by a factor as large as the ambient dimension $N$, in general. The differences between the tail parameters ($L_{p,p}$, $L_{p,\infty}$, and $L_\infty$) are not an essential feature of the analysis; we have stated the simplest versions of the results, rather than the optimal versions.

On the other hand, the approach in Theorem IV.3 cannot provide more refined comparisons for random matrix models that match beyond the second moment (except perhaps when the third moments are identically zero). There are intrinsic reasons that continuous interpolation does not seem to extend beyond second moments (Appendix F). In contrast, the method based on Lindeberg exchange gracefully handles matching moments of any order.

As we will see (Sec. IV B), there are some natural settings where higher-order moments coincide. The resulting higher-order error bounds improve over the second-order bounds. In the setting of quantum information, we often need to take the moment parameter $p \sim \log N \sim n$, so this improvement is significant.

In addition, our argument is conceptually and technically simpler than the approach based on Stein's method and cumulant expansions. As a consequence, it may be easier to extend to other settings, and it may have a different scope of application. Altogether, our work contributes to the emerging toolkit for nonasymptotic RMT.

### B. Further examples

Our universality results apply to many different families of random matrix models, including examples that may not resemble the Gaussian models that are central to the comparison in Ref. [55]. For quantum computing applications, these families could potentially capture realistic sparse matrices better than random Pauli string sums. However, they generally require access to an additional block encoding, which we do not discuss in this work.

#### *1. Comparing sparse matrices with GUE*

In addition to the random Pauli string ensemble, we can describe another family of sparse random matrices that also matches the low moments of GUE. Therefore, the universality results (Theorems IV.1 and IV.2) show that these models nearly follow a semicircular distribution.

*Definition IV.1.* (Permutations with complex signs). A random complex signed permutation matrix is the product of a uniformly random permutation matrix $\boldsymbol{P}$ and a diagonal matrix $\boldsymbol{D}$ with complex signs:

$$\boldsymbol{Q} := \boldsymbol{D}\boldsymbol{P} \quad \text{where } \boldsymbol{D}_{ab} = \delta_{ab}\frac{r_a + \mathrm{i}r'_a}{\sqrt{2}} \quad \text{and} \quad r_a, r'_a \overset{\text{i.i.d.}}{\sim} \{1, -1\}.$$

*Proposition IV.1.* (Complex signed permutations). Consider random matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ that take the form

$$\boldsymbol{A} := \frac{\boldsymbol{Q} + \boldsymbol{Q}^\dagger}{\sqrt{2}} \quad \text{where } \boldsymbol{Q} \text{ is a complex signed permutation,}$$
$$\tilde{\boldsymbol{A}} \sim \boldsymbol{H}_{\mathrm{GUE}}.$$

For these models, the first three moments match:

$$\mathbb{E}[\boldsymbol{A}^{\otimes k}] = \mathbb{E}[\tilde{\boldsymbol{A}}^{\otimes k}] \quad \text{for each } k = 1, 2, 3.$$

See Appendix E 1 for the calculation. One may also consider random, real, signed permutations, [57] which match the first three moments of the Gaussian orthogonal ensemble (GOE).

#### *2. Higher-moment matching*

Even though higher-moment matching examples are less common, we can describe several pairs of models that match up to arbitrarily high moments. The first example considers conjugating a fixed matrix by random unitaries:

$$\tilde{\boldsymbol{A}}_i := \frac{1}{\sqrt{m}}\tilde{\boldsymbol{U}}_i\boldsymbol{\sigma}\tilde{\boldsymbol{U}}_i^\dagger \quad \text{where } \tilde{\boldsymbol{U}}_i \sim \text{Haar};$$
$$\boldsymbol{A}_i := \frac{1}{\sqrt{m}}\boldsymbol{U}_i\boldsymbol{\sigma}\boldsymbol{U}_i^\dagger \quad \text{where } \boldsymbol{U}_i \sim \text{unitary } t\text{-design.}$$

Indeed, if we take the unitaries $\boldsymbol{U}_i$ to be the Clifford circuits (exact 3-design) and $\boldsymbol{\sigma}$ to be a fixed Pauli string, we nearly obtain the Pauli string ensemble (up to the identity element that cannot be produced by conjugation). However, beyond Clifford circuits, we do not know other examples where the matrices $\boldsymbol{A}_i$ remain sparse.

If we insist on sparse matrices, another example is

$$\tilde{\boldsymbol{A}}_i := \frac{\tilde{\boldsymbol{Q}}_i + \tilde{\boldsymbol{Q}}_i^\dagger}{\sqrt{2m}} \quad \text{where } \tilde{\boldsymbol{Q}}_i \sim \text{IID}$$
complex signed permutations;
$$\boldsymbol{A}_i := \frac{\boldsymbol{Q}_i + \boldsymbol{Q}_i^\dagger}{\sqrt{2m}} \quad \text{where } \boldsymbol{Q}_i \sim \text{IID}$$
$t$-wise independent complex signed permutations.

In this context, $t$-wise independent permutation is exactly the $t$th moment matching condition $\mathbb{E}\boldsymbol{Q}_i^{\otimes t} = \mathbb{E}\tilde{\boldsymbol{Q}}_i^{\otimes t}$. Exact and *approximate* constructions for both $t$-designs [58,59] and $t$-wise independent permutations [60] are available in the literature. We leave a careful analysis of the approximate case, where very few random bits are needed, for future work.

## V. COMMENTS ON DEQUANTIZATION AND QUANTUM ADVANTAGE

In this section, we comment on the *classical* complexity for the low-energy problem. The flavor differs from local Hamiltonian problems because our model is highly nonlocal and has a semicircular spectrum.

(1) How far does dequantization go? As we mentioned, recent developments in dequantization show that many linear algebra tasks can be efficiently solved assuming certain classical access to a quantum state. In particular, existing results consider low-rank matrices for various tasks [14,16,17,61] or high-rank matrices but with constant accuracy [18].

In the setting of Theorem II.1, we provide an efficient classical witness for the optimum if the accuracy $\epsilon > 0$ is an arbitrarily fixed *constant* (with polynomially large $m$). The idea is a simple polynomial approximation. However, the cost of manipulating the witness is $\exp(\tilde{\Omega}(1/\sqrt{\epsilon}))$, which scales poorly with the constant $\epsilon$.

*Proposition V.1.* (Efficient classical witness at arbitrary constant accuracy). For any $\epsilon$ and large enough $m = \Omega(\mathrm{Poly}(n, \epsilon^{-1}))$, there is a degree $d = \mathcal{O}(1/\sqrt{\epsilon})$ polynomial $p_d(x)$ such that the associated ansatz state has low energy,

$$\boldsymbol{\rho} \propto p_d(\boldsymbol{H})^2 \quad \text{such that } \mathrm{Tr}[\boldsymbol{\rho}\boldsymbol{H}] \leq (1-\epsilon)\lambda_{\min}(\boldsymbol{H}).$$

Further, this can be efficiently verified classically in runtime

$$\mathcal{O}((dm)^{2d} nd).$$

*Proof of Proposition V.1.* Using power-series approximation (i.e., Taylor expansion) for the Gibbs state gives a suboptimal degree $d = \mathcal{O}(1/\epsilon)$. A better degree $d = \mathcal{O}(1/\sqrt{\epsilon})$ can be achieved using a Chebyshev-polynomial approximation of the Gibbs state. The verification algorithm simply evaluates all the $(dm)^{2d}$ terms in the ansatz state $\boldsymbol{\rho}$. Each of the $(dm)^{2d}$ terms requires $2d$ multiplications of Pauli strings, each with cost $\mathcal{O}(n)$. ∎

This indicates that eigenstates "far from the ground state" have polynomial classical complexity, which is reminiscent of the cost of dequantization methods [18] in the context of ground-energy estimation given good trial states. Still, the above classical polynomial witness gets stuck at a constant approximation ratio, while the quantum algorithm has no problem moving to better and better accuracy [62].

(2) What is a decision problem? To discuss quantum advantage, ideally one needs a problem with classical inputs and outputs, especially a decision problem. A candidate problem is to compute an approximation to the ground energy of our model. However, since our problem has randomness, we expect the spectrum to be concentrated around the semicircle. If the spectrum were exactly the semicircle, a classical algorithm could simply output the deterministic value. Therefore, the classical hardness, if it exists, must originate from the instance-to-instance *fluctuation* of the spectrum away from the semicircle density, and that is why we need the accuracy $\epsilon = 1/\mathrm{Poly}(n)$ to be small while the number of terms $m = \mathrm{Poly}(n)$ is not too large (otherwise the fluctuation becomes too small and predictable).

Acknowledging the above, a candidate problem for quantum advantage is deciding the density of states to a high precision. It also converts to a binary decision problem by setting a threshold.

*Question V.0.1.* (Task: Deciding the density of states). Given a Hamiltonian sampled from the Pauli string ensemble and a small parameter $\epsilon$, output the number of states at a small energy interval,

$$[-\delta, \delta] \subset [-2, 2] \quad \text{up to a multiplicative error } \epsilon.$$

Is it classically hard for some $m = \mathrm{Poly}(n)$, $\epsilon = \mathrm{Poly}(n)^{-1}$, $\delta = \mathrm{Poly}(n)^{-1}$?

There is a quantum algorithm that succeeds with gate complexity $\mathrm{Poly}(\epsilon^{-1}, \delta^{-1}, m)$: Our concentration argument for a low-energy density of states [Eq. (C3) in the proof of Theorem II.1] also implies that for each $\delta$, there is a polynomially large $m = \mathrm{Poly}(\delta^{-1}, n)$ such that the local density $[-\delta, \delta]$ is at least half of that of the semicircle. Therefore, phase estimation samples from this interval with $\Omega(\delta^{-1})$ success probability. Repeated trials [63] give a high-confidence estimate of the density of states to an error $\epsilon$ with $\mathrm{Poly}(1/\epsilon)$ algorithmic cost. Why consider the problem of approximating the density of states and not approximating the ground-state energy? Currently, we do not have control over the spectrum very close to the extreme eigenvalues; for fixed $m$, $n$, our current results do not rule out the possibility of a small spectral gap $\Omega(m^{-1/4}n^{5/4})$; while we believe the spectral gap is exponentially small, the proof will require further developments in nonasymptotic random matrix theory.

Proving classical hardness for the density-of-states problem, e.g., by using a problem already known to be hard, is more elusive. A general proof might be unrealistic since it would give a computational quantum advantage for an oracle-free average-case decision problem, something for which no other examples are known. Still, it would be interesting to provide arguments for it. A concrete step is to prove that the spectrum has a large enough instance-to-instance fluctuation away from the semicircle distribution such that the classical algorithm cannot succeed simply by always outputting the average value. We believe this to be true, and it would be interesting to test it numerically. However, a proof would require further developments in nonasymptotic random matrix theory.

(3) Quantum chaos and quantum advantage. Our work fits into the broader question of whether quantum chaos could be a source of quantumly easy problems and perhaps a quantum computational advantage. As we mentioned, Hastings and O'Donnell [33] made concrete progress on the SYK model, a prominent toy model of quantum chaos, by providing a low-energy state where Gaussian states are known to fail. Their results would serve the question at hand even better if the classical hardness argument can be improved or if the Hamiltonian remains provably easy near the ground state. The latter seems plausible on physical grounds as it remains "chaotic" near the ground energy. Indeed, if one were to formally assume quantum chaos in terms of the eigenstate thermalization hypothesis (ETH), one may prove that preparing low-energy states is quantumly easy because Gibbs sampling at low temperatures is efficient on a quantum computer [31].

Our work made progress in capturing quantum chaos and its consequences by studying random matrix models where nonasymptotic treatment of spectral properties is possible even near the ground energy. Still, we acknowledge that our model is nonlocal and perhaps deviates from local Hamiltonian problems in some aspects: The quantum easiness stems from the semicircular spectrum and does not directly explain why the low-energy problem of chaotic local Hamiltonians (whose spectral density has a tail instead) should also be easy. Nevertheless, we expect the following findings to extrapolate to local chaotic Hamiltonians: Random matrix behavior can emerge from very few bits of randomness, and the spectrum is smooth and free of outliers (Fig. 2).

Still, there is a wealth of quantum chaos phenomenology that requires formal treatment for quantum advantage implications. One direction is to show ETH (e.g., for the SYK models), which roughly means that nearby energy eigenstates are well connected to each other. We believe this can be formalized for the GUE, which should also extend to our Pauli string ensemble by the universality principle. Another direction is to reduce the locality of the Pauli string ensemble. In fact, our circuit complexity lower-bound argument remains nontrivial even when the locality $k$ of each Hamiltonian term is reduced from $k = \Theta(n)$ to $k = \log(n)$, which at least gives a hint of classical hardness.

## ACKNOWLEDGMENTS

We thank Thomas Vidick, John Preskill, Anand Natarajan, John Wright, Gil Refael, Mehdi Soleimanifar, András Gilyén, Sam McArdle, William Kretschmer, Michael Kastoryano, Hsin-Yuan (Robert) Huang, Leo Zhou, Nicola Pancotti, and Robbie King for helpful discussions. C. F. C. is supported by the Eddlemen Fellowship and AWS Center for Quantum Computing summer intern program. During this project, M. B. was additionally affiliated with the AWS Center for Quantum Computing, Pasadena, USA. J. A. T. was supported in part by ONR BRC Grant No. N00014-18-1-2363 and NSF FRG Grant No. 1952777. M. B. is supported by the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant No. EP/W032643/1.

## APPENDIX A: CALCULATIONS FOR THE LINDEBERG PRINCIPLE

In this appendix, we apply a version of the Lindeberg exchange principle for the $p$th moments and the resolvent moments. The main assumption we use is that two sums of independent matrices share the same lower-order moments. The main technical argument is readily illustrated in the moment calculation. The resolvent calculation is more involved because the resolvent is nonconvex. We also have to establish concentration for a random realization of the resolvent moment around its expected value.

### 1. Moments

We recapitulate the statement for moments.

*Theorem IV.1.* (Universality for moments). Consider two families $(\boldsymbol{A}_i)$ and $(\tilde{\boldsymbol{A}}_i)$ of independent, random, Hermitian matrices whose moments match Eq. (9) up to order $t \geq 2$, and introduce the sums $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$ as in Eq. (8). Define the statistics

$$L_{p,k} := \left[\sum_{i=1}^{m} (|||\boldsymbol{A}_i|||_p^k + |||\tilde{\boldsymbol{A}}_i|||_p^k)\right]^{1/k} \quad \text{for } k \geq 1 \quad \text{and} \quad p \geq 2;$$

$$\sigma^2 := \sum_{i=1}^{m} (\|\mathbb{E}\boldsymbol{A}_i^2\| + \|\mathbb{E}\tilde{\boldsymbol{A}}_i^2\|).$$

Then, for each even natural number $p \in 2\mathbb{N}$, we have the bounds

$$\left|\|\!|\boldsymbol{H}|\!\|_p - \|\!|\tilde{\boldsymbol{H}}|\!\|_p\right| \le \begin{cases} 2p^{t/(t+1)} \times L_{p,t+1} + 2p \times L_{p,p} \\ 2p^{t/(t+1)} \times (\sigma^2 L_{\infty,\infty}^{t-1})^{1/(t+1)} + 2p \times L_{p,p}. \end{cases}$$

The $p$-norm is defined in Eq. (4).

Our proof of Theorem IV.1 is based on the Lindeberg exchange principle. Roughly, we interpolate between the two sums by replacing one argument at each step. Since the low moments of the summands match, each replacement only changes the $p$-norm slightly, with an error on the order $(t+1)$. The calculation is straightforward, but it implicitly exploits noncommutativity properties of the random matrices in the moment matching. The error is a noncommutative polynomial of matrices, which we treat by a strict application of Hölder's inequality, entirely ignoring noncommutativity. Once we have replaced all the summands, we tie the estimates together using a self-bounding argument. To execute this step, we must solve a difference equation by passing to a continuous differential equation.

Lindeberg's method has recently been applied to RMT in Refs. [50,51,64]. Some of the other ideas and methods in this argument are from Refs. [53–55,65].

The basic argument relies on a general form of Hölder's inequality for Schatten norms. As we will see, we can introduce more refined moment inequalities to obtain some improvements.

*Fact A.1.* (Multivariate Hölder for random matrices). For any family $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ of square random matrices, possibly statistically dependent, the product satisfies the trace inequality

$$\mathbb{E}\overline{\mathrm{Tr}}\left|\prod_{i=1}^k \boldsymbol{X}_i\right| = \left\|\!\left|\prod_{i=1}^k \boldsymbol{X}_i\right|\!\right\|_1 \le \prod_{i=1}^k \|\!|\boldsymbol{X}_i|\!\|_{p_i} \quad \text{whenever} \quad \sum_{i=1}^k \frac{1}{p_i} = 1 \quad \text{and} \quad p_i \ge 0.$$

*Proof sketch.* In the deterministic setting, with two matrices, the result appears in Corollary 4.2.6 in Ref. [66]. Use induction to extend the bound to more than two deterministic matrices. To incorporate the normalized trace, note that the weighted geometric mean $(p_1, \ldots, p_\ell) \mapsto \prod_{i=1}^\ell a_i^{p_i}$ is homogeneous for fixed $a_j \ge 0$. To incorporate the expectation, recall that the weighted geometric mean is concave, and invoke Jensen's inequality. ∎

*Proof of Theorem IV.1: Two matching moments.* To illustrate the concept behind the argument, we carefully establish the first bound for the $t = 2$ case. Afterward, we describe the modifications required to extend the bound to $t > 2$ and to introduce the variance parameter $\sigma^2$.

Fix an even natural number $p \in 2\mathbb{N}$. Our goal is to compare the $p$th moment $\|\!|\cdot|\!\|_p^p$ of the two independent sums $\boldsymbol{S} = \sum_{i=1}^m \boldsymbol{A}_i$ and $\tilde{\boldsymbol{S}} = \sum_{i=1}^m \tilde{\boldsymbol{A}}_i$. The main idea is to update one summand at a time from $\boldsymbol{A}_j$ to $\tilde{\boldsymbol{A}}_j$, controlling the change in the $p$-norm at each step.

In detail, for each index $j = 0, 1, 2, \ldots, m$, we can define the hybrid matrix

$$\boldsymbol{S}_j := \sum_{i=1}^j \tilde{\boldsymbol{A}}_i + \sum_{i=j+1}^m \boldsymbol{A}_i \quad \text{where } \boldsymbol{S}_0 = \boldsymbol{S} \quad \text{and} \quad \boldsymbol{S}_m = \tilde{\boldsymbol{S}}.$$

Express the difference between the $p$th moments of $\boldsymbol{S}$ and $\tilde{\boldsymbol{S}}$ as a telescoping sum:

$$\mathbb{E}\overline{\mathrm{Tr}}\left(\sum_{i=1}^m \tilde{\boldsymbol{A}}_i\right)^p - \mathbb{E}\overline{\mathrm{Tr}}\left(\sum_{i=1}^m \boldsymbol{A}_i\right)^p = \sum_{j=1}^m (\mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}_j^p - \mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}_{j-1}^p). \tag{A1}$$

For even $p$, we can express the $p$-norm [67] in terms of a trace power: $\|\!|\boldsymbol{S}|\!\|_p^p = \mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}^p$. To bound the telescoping sum, we first analyze the single update error and then solve a recursion.

*Step I: Single update error.* Fix an index $j = 1, \ldots, m$. Let us give a bound for the change in the $p$-norm when we update $\boldsymbol{A}_j$ to $\tilde{\boldsymbol{A}}_j$. Define the unchanged part of the sum $\boldsymbol{S}_j$ by

$$\boldsymbol{S}_- := \boldsymbol{S}_j - \tilde{\boldsymbol{A}}_j = \boldsymbol{S}_{j-1} - \boldsymbol{A}_j.$$

We can control the change in the polynomial moment by performing a Taylor expansion of the polynomial moment at the unchanged part $\boldsymbol{S}_-$.

When expanding powers of a sum of matrices, keep in mind the scalar binomial expansion:

$$(x+y)^p = \sum_{k=0}^p \binom{p}{k} x^{p-k} y^k.$$

For matrices, the expansion takes the form

$$\begin{aligned}(\boldsymbol{S}_- + \boldsymbol{A}_j)^p &= \sum_{\text{words}} \boldsymbol{S}_- \cdots \boldsymbol{A}_j \boldsymbol{S}_- \cdots \boldsymbol{A}_j \boldsymbol{S}_- \cdots \boldsymbol{A}_j \boldsymbol{S}_- \cdots \\ &= \sum_{k=0}^p \boldsymbol{M}_k \quad \text{where } \left|\mathbb{E}\overline{\mathrm{Tr}}\left|\boldsymbol{M}_k\right|\right| \\ &\le \binom{p}{k} \|\!|\boldsymbol{S}_-|\!\|_p^{p-k} \|\!|\boldsymbol{A}_j|\!\|_p^k. \end{aligned} \tag{A2}$$

Likewise, $(\boldsymbol{S}_- + \tilde{\boldsymbol{A}}_j)^p = \sum_{i=1}^m \tilde{\boldsymbol{M}}_i$ with analogous bounds for the summands.

The bound in Eq. (A2) follows from Hölder's inequality (Fact A.1), applied for all possible relative positions of $\boldsymbol{S}_-$ and $\boldsymbol{A}_j$ with all parameters $p_i = p$. Note that this general bound ignores the noncommutativity of the matrices. The binomial coefficients $\binom{p}{k}$ are exactly the number of relative

positions of the $k$ appearances of $\boldsymbol{A}_j$ among the $p-k$ appearances of $\boldsymbol{S}_-$.

To proceed, since the order parameter $t=2$, the first and second moments of the random matrices $\boldsymbol{A}_i$ and $\tilde{\boldsymbol{A}}_i$ match for each index $i$. As a consequence,

$$\begin{aligned}\mathbb{E}\overline{\mathrm{Tr}}[\boldsymbol{A}_i\boldsymbol{B}] &= \mathbb{E}\overline{\mathrm{Tr}}[\tilde{\boldsymbol{A}}_i\boldsymbol{B}];\\ \mathbb{E}\overline{\mathrm{Tr}}[\boldsymbol{A}_i\boldsymbol{B}\boldsymbol{A}_i\boldsymbol{C}] &= \mathbb{E}\overline{\mathrm{Tr}}[\tilde{\boldsymbol{A}}_i\boldsymbol{B}\tilde{\boldsymbol{A}}_i\boldsymbol{C}] \quad \text{for arbitrary } \boldsymbol{B},\boldsymbol{C}\\ &\qquad \text{independent of } \boldsymbol{A}_i,\tilde{\boldsymbol{A}}_i.\end{aligned}$$

Crucially, this implies that subtracting the expected moments completely *cancels* the first-order terms $\boldsymbol{M}_1$ and $\tilde{\boldsymbol{M}}_1$ and the second-order terms $\boldsymbol{M}_2$ and $\tilde{\boldsymbol{M}}_2$. Thus,

$$\begin{aligned}\left|\mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}_j^p - \mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}_{j-1}^p\right| &= \left|\mathbb{E}\overline{\mathrm{Tr}}(\boldsymbol{S}_- + \tilde{\boldsymbol{A}}_j)^p - \mathbb{E}\overline{\mathrm{Tr}}(\boldsymbol{S}_- + \boldsymbol{A}_j)^p\right|\\ &= \left|\sum_{k=3}^{p}(\boldsymbol{M}_k - \tilde{\boldsymbol{M}}_k)\right|\\ &\le \sum_{k=3}^{p} p^k |||\boldsymbol{S}_{j-1}|||_p^{p-k}|||\boldsymbol{A}_j|||_p^k + (\boldsymbol{A}_j \mapsto \tilde{\boldsymbol{A}}_j)\\ &\le \frac{1}{4}|||\boldsymbol{S}_{j-1}|||_p^{p-3}(2p|||\boldsymbol{A}_j|||_p)^3\\ &\quad + \frac{1}{4}(2p|||\boldsymbol{A}_j|||_p)^p + (\boldsymbol{A}_j \mapsto \tilde{\boldsymbol{A}}_j). \qquad \text{(A3)}\end{aligned}$$

The notation $(\boldsymbol{A}_j \mapsto \tilde{\boldsymbol{A}}_j)$ denotes a replica of the first term with $\boldsymbol{A}_j$ replaced by $\tilde{\boldsymbol{A}}_j$. To reach the second line, we collect the higher-order terms $\boldsymbol{M}_3,\ldots,\tilde{\boldsymbol{M}}_p$ and $\tilde{\boldsymbol{M}}_3,\ldots,\tilde{\boldsymbol{M}}_p$, we bound the binomial coefficient $\binom{p}{k} \le p^k$, and we use the convexity of the $p$-norm $|||\boldsymbol{S}_-|||_p = |||\boldsymbol{S}_- + \mathbb{E}[\boldsymbol{A}_j]|||_p \le |||\boldsymbol{S}_- + \boldsymbol{A}_j|||_p = |||\boldsymbol{S}_{j-1}|||_p$. The last inequality bounds the geometric series using the elementary numerical inequality

$$\sum_{k=3}^{p} x^k = \sum_{k=3}^{p} 2^{-k}(2x)^k \le \frac{1}{4}((2x)^3 + (2x)^p) \quad \text{for } x \ge 0.$$

We have successfully established a comparison between the quantities $|||\boldsymbol{S}_j|||_p^p$ and $|||\boldsymbol{S}_{j-1}|||_p^p$.

*Step II: Solving the recursion.* We have expressed the difference between the moments of $\boldsymbol{S}$ and $\tilde{\boldsymbol{S}}$ as a telescoping sum (A1) of moments of hybrid matrices $\boldsymbol{S}_j$. The first part of the argument yields a bound on the change in moments at each step in terms of a smaller moment of the hybrid matrices. We can use these results to develop coupled difference inequalities, which we must solve.

Define the scalar quantities

$$x_j := \mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}_j^p \quad \text{for } j = 0,1,2,\ldots,m.$$

The boundary values of the sequence $(x_j)$ are the moments of the original independent sums that we seek to compare: $x_m = \mathbb{E}\overline{\mathrm{Tr}}\tilde{\boldsymbol{S}}^p$ and $x_0 = \mathbb{E}\overline{\mathrm{Tr}}\boldsymbol{S}^p$. To bound the differences of this sequence, we introduce the notation

$$\begin{aligned}a_j &:= \frac{(2p)^3}{4}(|||\boldsymbol{A}_j|||_p^3 + |||\tilde{\boldsymbol{A}}_j|||_p^3) \quad \text{and}\\ b_j &:= \frac{(2p)^p}{4}(|||\boldsymbol{A}_j|||_p^p + |||\tilde{\boldsymbol{A}}_j|||_p^p) \quad \text{and} \quad a_0 := b_0 := 0.\end{aligned}$$

Using the inequality (A1) for the $j$th step of the exchange, we arrive at the coupled difference inequalities:

$$|x_j - x_{j-1}| \le a_j \times x_{j-1}^{(p-3)/p} + b_j \quad \text{for each } j = 1,\ldots,m. \qquad \text{(A4)}$$

Our task is to produce bounds for the terminal value $x_m$ in terms of the initial value $x_0$ and the coefficients $a_j$ and $b_j$.

To do so, we pass to a differential equation. The proof appears at the end of this section.

*Lemma A.1.* (From differences to derivatives). Define coefficient functions

$$a(s) := a_{\lceil s\rceil} \quad \text{and} \quad b(s) := b_{\lceil s\rceil} \quad \text{for } s \in [0,m].$$

For a fixed integer $1 \le k \le p$, consider the differential inequality

$$\begin{cases} x'(s) \ge a(s) \times x(s)^{(p-k)/p} + b(s) & s \in [0,m]\\ x(0) = x_0. \end{cases} \qquad \text{(A5)}$$

Then, each solution $x(s)$ to the differential inequality overestimates the solution $(x_j)$ to the coupled difference inequalities (A4) in the sense that

$$x(j) \ge x_j \ge 0 \quad \text{for each } j = 0,\ldots,m.$$

The following ansatz provides a solution to the differential inequality (A5). The proof of this lemma appears at the end of the section.

*Lemma A.2.* (Ansatz for differential inequality). Fix an integer $1 \le k \le p$. Consider the function

$$y(s) := \left[x(0)^{k/p} + \frac{k}{p}\int_0^s a(u)du + \left(\int_0^s b(u)du\right)^{k/p}\right]^{p/k} \quad \text{for } s \in [0,m].$$

Then, $y$ solves the differential inequality (A.1).

We are now prepared to solve the coupled difference inequalities. Instantiate Lemmas A.1 and A.2 with parameters $s = m$ and $k = 3$ to arrive at the one-sided inequality

$$x_m^{3/p} \le y(m)^{3/p} = x_0^{3/p} + \frac{3}{p}\sum_{j=1}^{m} a_j + \left(\sum_{j=1}^{m} b_j\right)^{3/p}.$$

This inequality provides an upper bound for the difference $x_m^{3/p} - x_0^{3/p}$, where we recall that $x_m$ and $x_0$ are the $p$th moments of the two independent sums. Taking the third root and bounding the $\ell_3$ norm by the $\ell_1$ norm, we also have the estimate

$$x_m^{1/p} - x_0^{1/p} \le \left(\frac{3}{p}\sum_{j=1}^m a_j\right)^{1/3} + \left(\sum_{j=1}^m b_j\right)^{1/p}.$$

This statement is slightly weaker, but it may be easier to interpret and apply.

We may repeat the same argument by switching the roles of $\boldsymbol{A}_j$ and $\tilde{\boldsymbol{A}}_j$, noting that coefficients $a_j$ and $b_j$ remain the same. This yields the desired two-sided estimate:

$$\left|\left|\left|\sum_{i=1}^m \boldsymbol{A}_i\right|\right|\right|_p^3 - \left|\left|\left|\sum_{i=1}^m \tilde{\boldsymbol{A}}_i\right|\right|\right|_p^3\right| = \left|x_m^{3/p} - x_0^{3/p}\right| \le \frac{3}{p}\sum_{j=1}^m a_j + \left(\sum_{j=1}^m b_j\right)^{3/p}.$$

Similarly,

$$\left|\left|\left|\sum_{i=1}^m \boldsymbol{A}_i\right|\right|\right|_p - \left|\left|\left|\sum_{i=1}^m \tilde{\boldsymbol{A}}_i\right|\right|\right|_p\right| = \left|x_m^{1/p} - x_0^{1/p}\right| \le \left(\frac{3}{p}\sum_{j=1}^m a_j\right)^{1/3} + \left(\sum_{j=1}^m b_j\right)^{1/p}.$$

Introduce the values of $a_j$ and $b_j$ and evaluate the numerical constants to complete the proof of the theorem for random matrix models with matching second moments ($t = 2$). ■

*Proof of Theorem IV.1: More matching moments.* Using an analogous argument, we can obtain related results comparing random matrix models where the moments match. Fix $t \ge 2$. Suppose that each pair $\boldsymbol{A}_i$ and $\tilde{\boldsymbol{A}}_i$ of summands has matching moments up to order $t$. In this case, the terms $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_t$ cancel with $\tilde{\boldsymbol{M}}_1, \ldots, \tilde{\boldsymbol{M}}_t$, so the error depends only on the higher-order terms $\boldsymbol{M}_k$ and $\tilde{\boldsymbol{M}}_k$ for $k \ge t+1$.

Pursuing this observation, we arrive at the bound

$$E := \left|\left|\left|\sum_{i=1}^m \boldsymbol{A}_i\right|\right|\right|_p - \left|\left|\left|\sum_{i=1}^m \tilde{\boldsymbol{A}}_i\right|\right|\right|_p\right| \le \left(\frac{t+1}{p}\sum_{j=1}^m a_j\right)^{1/(t+1)} + \left(\sum_{j=1}^m b_j\right)^{1/p},$$

where

$$a_j := \frac{(2p)^{t+1}}{2^t}\left(|||\boldsymbol{A}_j|||_p^{t+1} + |||\tilde{\boldsymbol{A}}_j|||_p^{t+1}\right) \quad \text{and}$$
$$b_j := \frac{(2p)^p}{2^t}\left(|||\boldsymbol{A}_j|||_p^p + |||\tilde{\boldsymbol{A}}_j|||_p^p\right).$$

Using the notation $L_{p,k}$ from the statement of the theorem, we reach the estimate

$$E \le (2(t+1))^{1/(t+1)} \times p^{t/(t+1)} \times L_{p,t+1} + 2^{1-t/p} \times p \times L_{p,p}.$$

For $2 \le t \le p$, each of the leading constants is bounded above by 2. This completes the argument. ■

*Proof of Theorem IV.1: Refined statistics.* Finally, we establish the result with more precise statistics of the random summands. To do so, we simply replace Hölder's inequality (Fact A.1) by a more refined moment inequality. We require the following statement, which specializes Prop. 4.1 in Ref. [55].

*Fact A.2.* (Trace inequality for random matrices). Let $\boldsymbol{A}$ and $\boldsymbol{Y}$ be random Hermitian matrices that are statistically independent. Consider a product with $k$ copies of $\boldsymbol{A}$ and $(p-k)$ copies of $\boldsymbol{Y}$ in any order, where $p \ge k \ge 2$. Then,

$$\mathbb{E}\overline{\mathrm{Tr}}[\boldsymbol{A}\boldsymbol{Y}^{p_1}\cdots\boldsymbol{A}\boldsymbol{Y}^{p_k}] \le [\|\mathbb{E}\boldsymbol{A}^2\| \times |||\boldsymbol{A}|||_\infty^{k-2}] \times |||\boldsymbol{Y}|||_p^{p-k}.$$

In this expression, $p_1 + \cdots + p_k = p - k$ is an integer partition.

We invoke this result with $\boldsymbol{Y} = \boldsymbol{S}_-$ and with $\boldsymbol{A} = \boldsymbol{A}_j$ or $\boldsymbol{A} = \tilde{\boldsymbol{A}}_j$. We can also take the minimum of this bound with the bound via Hölder's inequality to see that the tail parameter $L_{p,p}$ does not get worse. The rest of the proof is the same. ■

Finally, we complete the proofs of the two lemmas that were required in the argument.

*Proof of Lemma A.1.* By assumption, $x(0) \ge x_0 = 0$. For an induction, assume that $x(j-1) \ge x_{j-1}$ for an index $j \ge 1$. Then,

$$\begin{aligned} &x(j) - x(j-1) \\ &= \int_{j-1}^j x'(s)ds \ge \int_{j-1}^j \left(a(s) \times x(s)^{\frac{p-k}{p}} + b(s)\right)ds \\ &\ge \int_{j-1}^j \left(a(j) \times x(j-1)^{\frac{p-k}{p}} + b(j)\right)ds \ge x_j - x_{j-1}. \end{aligned}$$

To reach the second line, note that the function $x(s)$ is increasing because the right-hand side of the differential inequality is positive. Then, observe that the coefficients $a(s) = a(\lceil s \rceil)$ and $b(s) = b(\lceil s \rceil)$ are constant on the domain of integration. By induction, we obtain

$$x(j) \ge x_j \quad \text{for each } j = 0, \ldots, m.$$

This is the stated result. ■

*Proof of Lemma A.2.* To verify that the ansatz satisfies the differential inequality, first note the initial condition $y(0) = x(0)$. Using this fact, we take the derivative

$$\begin{aligned}\frac{dy(s)}{ds} &= \frac{p}{k}\left[y(0)^{k/p} + \frac{k}{p}\int_0^s a(u)du + \left(\int_0^s b(u)du\right)^{\frac{k}{p}}\right]^{(p/k)-1}\left(\frac{k}{p}a(s) + \left(\int_0^s b(u)du\right)^{(k/p)-1} \times \frac{k}{p}b(s)\right)\\ &\geq a(s)\left[y(0)^{k/p} + \frac{k}{p}\int_0^s a(u)du + \left(\int_0^s b(u)du\right)^{k/p}\right]^{(p-k)/k} + b(s)\\ &= a(s)y(s)^{(p-k)/p} + b(s).\end{aligned}$$

The inequality depends on the fact that $(k/p) - 1 \leq 0$ and

$$\left(\int_0^s b(u)du\right)^{(k/p)-1} \geq \left[y(0)^{k/p} + \frac{k}{p}\int_0^s a(u)du + \left(\int_0^s b(u)du\right)^{k/p}\right]^{1-(p/k)}.$$

This is a direct calculation using the fact that all the terms are positive. Therefore, the ansatz solves the differential inequality. ■

### 2. Resolvent

The comparison principle extends to other functions besides polynomial moments. In this section, we study moments of the resolvent:

$$\boldsymbol{R} := \frac{1}{\boldsymbol{H} - \omega + \mathrm{i}\eta} \quad \text{and} \quad \tilde{\boldsymbol{R}} := \frac{1}{\tilde{\boldsymbol{H}} - \omega + \mathrm{i}\eta}.$$

As usual, $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$ are defined in Eq. (8). The parameters $\omega \in \mathbb{R}$ and $\eta > 0$.

*Theorem IV.2.* (Universality for resolvent moments). Instate the assumptions and notation of Theorem IV.1. For each even natural number $p \in 2\mathbb{N}$, the polynomial moments of the resolvent (11) are related by

$$\left|\|\boldsymbol{R}\|_p - \|\tilde{\boldsymbol{R}}\|_p\right| \lesssim \frac{1 + (m/p)}{\eta} \times \left(\frac{pL_{3p,\infty}}{\eta}\right)^{t+1}.$$

The symbol $\lesssim$ suppresses constants depending only on $t$.

Whenever the right-hand side is small [$\ll (p\eta)^{-1}$], we may take the $p$th power to obtain the expected density of states (filtered by the resolvent) up to a *multiplicative* error. For our Pauli string ensemble, we can achieve this outcome because $L_{3p,\infty} = 1/\sqrt{m}$, and the number $m$ of summands is chosen sufficiently large.

As compared with the polynomial moments, universality for the resolvent involves some additional technical challenges. They stem from the fact that the resolvent has an infinite Taylor series, and it is a nonconvex function of the random matrix. To address the first concern, we follow Ref. [55] and truncate the Taylor series at a carefully chosen order. To that end, let us recall the statement of Taylor's theorem with an integral remainder.

*Fact A.3.* (Taylor with integral remainder). If the function $f : [0, 1] \to \mathbb{R}$ is $K$ times continuously differentiable, then

$$f(1) = \sum_{k=0}^{K-1} \frac{f^{(k)}(0)}{k!} + \frac{1}{(K-1)!}\left[\int_0^1 f^{(K)}(s)(1-s)^{K-1}ds\right].$$

The Taylor expansion of the resolvent has a rather involved expression. Fortunately, we merely need bounds for the higher-order terms.

*Proposition A.1.* (Expanding the resolvent). For Hermitian matrices $\boldsymbol{S}$ and $\boldsymbol{A}$ of the same order, consider the matrix $\boldsymbol{Z} = \boldsymbol{S} + \mathrm{i}\eta\mathbf{I}$, where $\eta \in \mathbb{R}$. Then, for each even natural number $p \in 2\mathbb{N}$,

$$\begin{aligned}\frac{1}{|\boldsymbol{Z} + \boldsymbol{A}|^p} &= \sum_{k=0}^{3p} \boldsymbol{M}_k \quad \text{where } \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{M}_k|\\ &\leq \frac{(4p)^k}{\eta^{4k/3}} \times |||\boldsymbol{Z}^{-1}|||_p^{p-k/3}|||\boldsymbol{A}|||_{3p}^k \quad \text{for } k = 0, \ldots, 3p.\end{aligned} \tag{A6}$$

The term $\boldsymbol{M}_k$ is a noncommutative polynomial of degree $k$ in the variable $\boldsymbol{A}$ and degree $p + k$ in the variables $\boldsymbol{Z}^{-1}$ and $\boldsymbol{Z}^{-\dagger}$, where $^{-\dagger}$ refers to the conjugate transpose of the inverse.

*Proof of Proposition A.1.* First, we expand

$$\begin{aligned}\frac{1}{|\boldsymbol{Z} + \boldsymbol{A}|^p} &= \frac{1}{(\boldsymbol{Z} + \boldsymbol{A})^{p/2}} \times \frac{1}{((\boldsymbol{Z} + \boldsymbol{A})^\dagger)^{p/2}}\\ &= \sum_{\ell_1=0}^{\infty}(-\boldsymbol{Z}^{-1}\boldsymbol{A})^{\ell_1}\boldsymbol{Z}^{-1} \cdots \sum_{\ell_{p/2}=0}^{\infty}(-\boldsymbol{Z}^{-1}\boldsymbol{A})^{\ell_{p/2}}\boldsymbol{Z}^{-1}\\ &\quad \times \sum_{\ell'_1=0}^{\infty}(-\boldsymbol{Z}^{-\dagger}\boldsymbol{A})^{\ell'_1}\boldsymbol{Z}^{-\dagger} \cdots \sum_{\ell'_{p/2}=0}^{\infty}(-\boldsymbol{Z}^{-\dagger}\boldsymbol{A})^{\ell'_{p/2}}\boldsymbol{Z}^{-\dagger}.\end{aligned} \tag{A7}$$

The first line depends on the fact that $\boldsymbol{Z} + \boldsymbol{A}$ and $\boldsymbol{Z}^\dagger + \boldsymbol{A}$ commute. The second line uses the expansion $(\boldsymbol{Z} + \boldsymbol{A})^{-1} = \sum_{\ell=0}^{\infty}(-\boldsymbol{Z}^{-1}\boldsymbol{A})^\ell\boldsymbol{Z}^{-1}$.

Next, to the matrix $\boldsymbol{M}_k$, we add all terms with total power $k$ on the matrix $\boldsymbol{A}$ and total power $p + k$ on the matrix $\boldsymbol{Z}^{-1}$ or $\boldsymbol{Z}^{-\dagger}$. For $0 \leq k < 3p$, there are $\binom{p+k-1}{k} \leq (4p)^k$ such

terms. Then, we apply Hölder's inequality (Fact A.1) to each term contributing to $\boldsymbol{M}_k$. This step yields

$$\mathbb{E}\mathrm{Tr}|\boldsymbol{M}_k| \le (4p)^k \times |||\boldsymbol{Z}^{-1}|||_{q_1}^{p+k} |||\boldsymbol{A}|||_{q_2}^k \quad \text{for } q_1 = \frac{3p}{3p-k} \quad \text{and} \quad q_2 = \frac{3p}{k}.$$

To bring the bounds into the same form, note that $|||\boldsymbol{Z}^{-1}|||_{q_1} \le |||\boldsymbol{Z}^{-1}|||_p$ for $q_1 \le p$ and $|||\boldsymbol{A}_q||| \le |||\boldsymbol{A}|||_{3p}$ for $q_2 \le 3p$. To treat additional powers of $\boldsymbol{Z}^{-1}$, use a uniform bound for the resolvent: $|||\boldsymbol{Z}^{-1}|||_{q_1}^{4k/3} \le \|\boldsymbol{Z}^{-1}\|^{4k/3} \le \eta^{-4k/3}$. We reach the desired bound for terms of order $0 \le k < 3p$.

For the remainder term in the Taylor expansion ($K = 3p$), we may compute the $K$th derivative using Eq. (A7) and invoke the same method to obtain a bound:

$$\left|\frac{d^{3p}}{ds^{3p}} \mathbb{E}\overline{\mathrm{Tr}} \frac{1}{|\boldsymbol{Z} + s\boldsymbol{A}|^p}\right| \le (3p)! \times (4p)^{3p} \times \frac{|||\boldsymbol{A}|||_{3p}^{3p}}{\eta^{3p}} \quad \text{for } s \in [0,1].$$

We have applied the uniform bound $\|(\boldsymbol{Z} + s\boldsymbol{A})^{-1}\| \le \eta^{-1}$. Introduce the last display equation into the integral remainder term in the Taylor expansion (A.3). We reach the required estimate for $K = 3p$. ■

*Proof of Theorem IV.2.* To obtain a comparison of the resolvents, we apply Lindeberg's method again. For clarity of argument, we will assume that there are $t = 2$ matching moments; the general case is similar. Define hybrid matrices and their resolvents:

$$\boldsymbol{H}_j := \sum_{i=1}^{j} \tilde{\boldsymbol{A}}_i + \sum_{j+1}^{m} \boldsymbol{A}_i \quad \text{and}$$

$$\boldsymbol{R}_j := \frac{1}{\boldsymbol{H}_j - \omega + \mathrm{i}\eta} \quad \text{for } j = 0, \ldots, m.$$

Consider the telescoping sum

$$\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p = \sum_{j=1}^{m} (\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p).$$

We must bound each of the terms in the telescope.

*Step I: Single update error.* Fix an index $j = 1, \ldots, m$. The $j$th update replaces the summand $\boldsymbol{A}_j$ with $\tilde{\boldsymbol{A}}_j$. Define the unchanged part of the matrix and its resolvent:

$$\boldsymbol{H}_- := \boldsymbol{H}_j - \tilde{\boldsymbol{A}}_j = \boldsymbol{H}_{j-1} - \boldsymbol{A}_j \quad \text{and} \quad \boldsymbol{R}_- := (\boldsymbol{H}_- - \omega + \mathrm{i}\eta)^{-1}.$$

Since the moments of $\boldsymbol{A}_j$ and $\tilde{\boldsymbol{A}}_j$ match up to second order and these matrices are independent of $\boldsymbol{H}_-$, the terms $\boldsymbol{M}_0$, $\boldsymbol{M}_1$, $\boldsymbol{M}_2$ cancel the terms $\tilde{\boldsymbol{M}}_0, \tilde{\boldsymbol{M}}_1, \tilde{\boldsymbol{M}}_2$ in the Taylor expansion of the resolvent powers (Proposition A.1). Thus,

$$\begin{aligned}
|\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p| &= |\mathbb{E}\overline{\mathrm{Tr}}|(\boldsymbol{H}_- - \omega + \mathrm{i}\eta) + \boldsymbol{A}_j|^{-p} - \mathbb{E}\overline{\mathrm{Tr}}|(\boldsymbol{H}_- - \omega + \mathrm{i}\eta) + \tilde{\boldsymbol{A}}_j|^{-p}| \\
&\le \frac{1}{4} |||\boldsymbol{R}_-|||_p^{p-1} \frac{(8p)^3 |||\tilde{\boldsymbol{A}}_j|||_{3p}^3}{\eta^4} + \frac{1}{4} \frac{(8p)^{3p} |||\tilde{\boldsymbol{A}}_j|||_{3p}^{3p}}{\eta^{4p}} + (\tilde{\boldsymbol{A}}_j \mapsto \boldsymbol{A}_j) \\
&\le \frac{1}{2} |||\boldsymbol{R}_-|||_p^{p-1} \frac{(8pc_j)^3}{\eta^4} + \frac{1}{2} \frac{(8pc_j)^{3p}}{\eta^{4p}} \quad (\text{setting } c_j := \max\{|||\boldsymbol{A}_j|||_{3p}, |||\tilde{\boldsymbol{A}}_j|||_{3p}\}) \\
&\le |||\boldsymbol{R}_{j-1}|||_p^{p-1} \frac{(8pc_j)^3}{\eta^4} + \frac{3}{2} \frac{(8pc_j)^{3p}}{\eta^{4p}}.
\end{aligned} \tag{A8}$$

The first inequality bounds the geometric series of the error terms (A6):

$$\sum_{k=3}^{3p} x^k \le \frac{1}{4}((2x)^3 + (2x)^{3p}) \quad \text{for } x := \frac{4p}{\eta^{4/3}} \times |||\boldsymbol{R}_-|||_p^{-1/3} |||\tilde{\boldsymbol{A}}_j|||_{3p}.$$

The second inequality combines the bounds for the two different summands $\boldsymbol{A}_j$ and $\tilde{\boldsymbol{A}}_j$.

The third inequality requires some comment. By another Taylor expansion, we may control the moments of $\boldsymbol{R}_-$ using the moments of $\boldsymbol{R}_{j-1}$:

$$\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p \le \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p \left(\sum_{k=0}^{3p} y^k\right) \quad \text{where } y := \frac{4p}{\eta^{4/3}} \cdot |||\boldsymbol{R}_{j-1}|||_p^{-1/3} |||\boldsymbol{A}_j|||_{3p}.$$

Bounding the geometric series as $\sum_{k=0}^{3p} y^k \le 2(1 + ((2y)^{3p})^{p/(p-1)})$ and noting that $|||\boldsymbol{A}_j|||_{3p} \le c_j$, we find that

$$\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p \le 2\left(\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p + \frac{(8pc_j)^{3p}}{\eta^{4p}}\right).$$

Lastly, raise both sides to the $(p-1)/p$ power, and use the numerical inequality $(a+b)^{(p-1)/p} \le a^{(p-1)/p} + b^{(p-1)/p}$ for $a, b \ge 0$ to reach Eq. (A8). A similar bound holds when $t \ge 2$ moments match.

*Step II: Solving the recursion.* The recursion is similar to the proof of Theorem IV.1. We will present this argument for a general choice of $t \ge 2$. First, introduce the scalar variables

$$x_j \coloneqq \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p \quad \text{for } j = 0, \ldots, m.$$

Define the coefficients $a_0 \coloneqq b_0 \coloneqq 0$ and

$$a_j \coloneqq \frac{1}{2^{t-2}} \frac{(2(t+2)pc_j)^{t+1}}{\eta^{t+2}} \quad \text{and}$$
$$b_j \coloneqq \frac{3}{2^{t-1}} \frac{(2(t+2)pc_j)^{(t+1)p}}{\eta^{(t+2)p}}.$$

The updates (A2) can then be written as a scalar recursion,

$$|x_j - x_{j-1}| \le a_j \cdot x_{j-1}^{(p-1)/p} + b_j \quad \text{for each } j = 1, \ldots, m.$$

Repeating the same arguments as before (via Lemmas A.1 and A.2), we obtain control on the endpoints of the sequence:

$$\begin{aligned} |x_m^{1/p} - x_0^{1/p}| &\le \frac{1}{p}\sum_{j=1}^m a_j + \left(\sum_{j=1}^m b_j\right)^{1/p} \\ &\le 8m \frac{p^t((t+2)L_{3p,\infty})^{t+1}}{\eta^{t+2}} \\ &\quad + \left(\frac{3}{2}\right)^{1/p} m^{1/p} \frac{(2(t+2)pL_{3p,\infty})^{t+1}}{2^{(t-2)/p}\eta^{t+2}} \\ &\le (2^{t+2} + 10m/p) \frac{((t+2)pL_{3p,\infty})^{t+1}}{\eta^{t+2}}. \end{aligned}$$

The second inequality uses the uniform bound $c_j = \max\{|||\tilde{\boldsymbol{A}}_j|||_{3p}, |||\boldsymbol{A}_j|||_{3p}\} \le L_{3p,\infty}$. To reach the last line, note that $(3/2)^{1/p} \le 2$, drop the denominator $2^{(t-2)/p} \ge 1$, and apply Young's inequality to determine that $m^{1/p} \le 1 + m/p$. ∎

### 3. Concentration for resolvent trace

In this section, we study the $q$th moments of the resolvent trace, which, by Markov's inequality, gives the concentration of the local density of states needed for Theorem II.1. The concentration fundamentally differs from the calculation for the expectation and does not explicitly refer to an ideal random matrix ensemble (e.g., the GUE). It suffices to introduce an independent copy by the convexity of the $q$-norm,

$$|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q \le |\overline{\mathrm{Tr}}|\boldsymbol{R}|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}'|^p|_q \quad \text{where}$$
$$|x|_q \coloneqq (\mathbb{E}|x|^q)^{1/q},$$

which allows us to utilize a powerful concentration inequality for martingales. The estimate depends on an expected moment $|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q$, which we bound independently in Appendix A 3 a.

*Theorem A.1.* (Concentration for resolvent trace). For independent centered matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$, consider identical copies $\boldsymbol{A}'_j$ of $\boldsymbol{A}_j$. Then, the resolvent trace is concentrated as

$$\begin{aligned} |\overline{\mathrm{Tr}}|\boldsymbol{R}|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}'|^p|_q \lesssim |\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q \Bigg( &\frac{\sqrt{q}p^2}{\eta^2}\sqrt{\sum_{j=1}^m |||\boldsymbol{A}_j|||_\infty^4} \\ &+ \frac{\sqrt{q}p}{\eta^2}\sqrt{\sum_{j=1}^m \sigma_*(\boldsymbol{A}_j)^2} \\ &+ \frac{qp}{\eta}\left(\sum_{j=1}^m |||\boldsymbol{A}_j|||_\infty^q\right)^{1/q}\Bigg), \end{aligned}$$

where

$$\sigma_*(\boldsymbol{A})^2 \coloneqq \sup_{|||u\rangle\|=|||v\rangle\|=1} \mathbb{E}_{\boldsymbol{A}}|\langle u|\boldsymbol{A}|v\rangle|^2.$$

Crucially, the estimate depends on a variance-like quantity $\sigma_*^2(\boldsymbol{A})$ that more fully reflects the randomness of the random matrix $\boldsymbol{A}$. For the Pauli string ensemble, this quantity is significantly smaller than the ordinary matrix variance: $\sigma_*^2(\boldsymbol{A}) = (mN)^{-1} \ll m^{-1} = \mathbb{E}|||\boldsymbol{A}_j|||^2$. The quantity $\sigma_*^2(\boldsymbol{A})$ arises from the following bound.

*Fact A.4.* Consider a random matrix $\boldsymbol{A}$ and a fixed matrix $\boldsymbol{B}$ with compatible dimensions. Then,

$$\mathbb{E}_{\boldsymbol{A}}|\mathrm{Tr}[\boldsymbol{A}\boldsymbol{B}]|^2 \le \sigma_*^2(\boldsymbol{A}) \cdot \mathrm{Tr}[|\boldsymbol{B}|]^2.$$

*Proof of Fact A.4.* Consider the singular value decomposition $\boldsymbol{B} = \sum_j |v_j\rangle s_j \langle u_j|$. Then,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{A}}\mathrm{Tr}[\boldsymbol{A}\boldsymbol{B}]\mathrm{Tr}[\boldsymbol{A}^\dagger\boldsymbol{B}^\dagger] &= \mathbb{E}_{\boldsymbol{A}}\sum_j \langle u_j|\boldsymbol{A}|v_j\rangle s_j \sum_j \langle v_j|\boldsymbol{A}^\dagger|u_j\rangle s_j \\ &\le \sup_{j,i} \mathbb{E}_{\boldsymbol{A}}|\langle u_j|\boldsymbol{A}|v_j\rangle||\langle u_i|\boldsymbol{A}|v_i\rangle| \\ &\quad \cdot \mathrm{Tr}[|\boldsymbol{B}|]\mathrm{Tr}[|\boldsymbol{B}|] \\ &\le \sup_{|||u\rangle\|=|||v\rangle\|=1} \mathbb{E}_{\boldsymbol{A}}|\langle u|\boldsymbol{A}|v\rangle|^2 \\ &\quad \cdot \mathrm{Tr}[|\boldsymbol{B}|]\mathrm{Tr}[|\boldsymbol{B}|]. \end{aligned}$$

The first inequality pushes the expectation inside the sum, and it applies Hölder's inequality to the sum. The second inequality is the Cauchy-Schwarz one. This is the desired result. ∎

In addition, the proof of Theorem A.1 employs a refined scalar martingale inequality as follows. (For an introduction to martingales, see Ref. [68].)

*Theorem A.2.* (Ref. [69]). For the scalar martingales difference sequence $d_j$ (i.e., $\mathbb{E}_{j-1}d_j := \mathbb{E}[d_j|d_{j-1},\ldots,d_1]=0$), we have that

$$\left|\max_{k\le n}\left|\sum_{j=1}^{k} d_j\right|\right|_q \lesssim \left(\sqrt{q}\left|\sum_{j=1}^{n}\mathbb{E}_{j-1}d_j^* d_j\right|_{\frac{q}{2}}^{1/2} + q\left|\max_{1\le j\le n}|d|_j\right|_q\right).$$

Significantly, the conditional expectation $\mathbb{E}_{j-1}$ appears inside the norm, which then allows us to exploit the second-moment properties of $\boldsymbol{A}_i$ via Fact A.4. Otherwise, applying a crude martingale inequality, such as uniform smoothness, gives a looser bound in terms of $\mathbb{E}\|\boldsymbol{A}_j\|^2$ instead of $\sigma_*^2(\boldsymbol{A})$. The weaker bound does not properly reflect the randomness in $\boldsymbol{A}_j$.

*Proof of Theorem A.1.* As usual, we write the telescoping sum

$$\overline{\mathrm{Tr}}|\boldsymbol{R}'|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}|^p = \sum_{j=1}^{m}(\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p) \quad \text{with} \quad \boldsymbol{R}_j := \frac{1}{\boldsymbol{H}_j - \omega + \mathrm{i}\eta} \quad \text{and} \quad \boldsymbol{H}_j := \sum_{i=1}^{j}\boldsymbol{A}'_i + \sum_{j+1}^{m}\boldsymbol{A}_i.$$

By construction, the updates compose a martingale difference sequence:

$$\mathbb{E}[\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p|\boldsymbol{A}'_{j-1},\boldsymbol{A}_{j-1},\ldots,\boldsymbol{A}'_1,\boldsymbol{A}_1] = 0. \tag{A9}$$

This point is evident because we can swap the random variables $\boldsymbol{A}_j$ and $\boldsymbol{A}'_j$ without changing the distribution. To analyze the martingale, we expand the difference using the algebraic identity for a difference of powers:

$$\begin{aligned}
\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p &= \overline{\mathrm{Tr}}[\boldsymbol{R}_j^\dagger\boldsymbol{R}_j|\boldsymbol{R}_j|^{p-2} - \boldsymbol{R}_{j-1}^\dagger\boldsymbol{R}_{j-1}|\boldsymbol{R}_{j-1}|^{p-2}]\\
&= \overline{\mathrm{Tr}}[(\boldsymbol{R}_j^\dagger - \boldsymbol{R}_{j-1}^\dagger)\cdot\boldsymbol{R}_{j-1}|\boldsymbol{R}_{j-1}|^{p-2}] + \cdots\\
&= \overline{\mathrm{Tr}}[\boldsymbol{R}_j^\dagger(\boldsymbol{H}_{j-1} - \boldsymbol{H}_j)\boldsymbol{R}_{j-1}^\dagger\cdot\boldsymbol{R}_{j-1}|\boldsymbol{R}_{j-1}|^{p-2}] + \cdots\\
&= \overline{\mathrm{Tr}}[\boldsymbol{R}_j^\dagger(\boldsymbol{A}_j - \boldsymbol{A}'_j)|\boldsymbol{R}_{j-1}|^p] + \cdots\\
&= a_j - b_j + c_j.
\end{aligned}$$

The second equality is a telescoping sum. For the moment, we have suppressed other telescoping terms, such as $\overline{\mathrm{Tr}}[\boldsymbol{R}_j^\dagger(\boldsymbol{R}_j - \boldsymbol{R}_{j-1})\cdot|\boldsymbol{R}_{j-1}|^{p-2}]$. The third equality uses the matrix identity $\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} = \boldsymbol{B}^{-1}(\boldsymbol{B} - \boldsymbol{A})\boldsymbol{A}^{-1}$. The last line regroups into three types of terms:

$$\begin{aligned}
a_j &:= \overline{\mathrm{Tr}}[\boldsymbol{R}_j^\dagger\boldsymbol{A}_j|\boldsymbol{R}_j|^p] + \overline{\mathrm{Tr}}[|\boldsymbol{R}|_j^2\boldsymbol{A}_j\boldsymbol{R}_j|\boldsymbol{R}_j|^{p-2}] + \cdots =: \sum_{\substack{r+s=p+1,\\ r,s\ge1}}\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j\boldsymbol{R}_j^{(s)}],\\
b_j &:= \overline{\mathrm{Tr}}[\boldsymbol{R}_{j-1}^\dagger\boldsymbol{A}'_j|\boldsymbol{R}_{j-1}|^p] + \overline{\mathrm{Tr}}[|\boldsymbol{R}|_{j-1}^2\boldsymbol{A}'_j\boldsymbol{R}_{j-1}|\boldsymbol{R}_{j-1}|^{p-2}] + \cdots =: \sum_{\substack{r+s=p+1,\\ r,s\ge1}}\overline{\mathrm{Tr}}[\boldsymbol{R}_{j-1}^{(r)}\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}],\\
c_j &:= \overline{\mathrm{Tr}}[\boldsymbol{R}_j^\dagger\boldsymbol{A}_j(|\boldsymbol{R}_{j-1}|^p - |\boldsymbol{R}_j|^p)] - \overline{\mathrm{Tr}}[(\boldsymbol{R}_j^\dagger - \boldsymbol{R}_{j-1}^\dagger)\boldsymbol{A}'_j|\boldsymbol{R}_{j-1}|^p] + \cdots\\
&=: \sum_{\substack{r+s=p+1,\\ r,s\ge1}}\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j(\boldsymbol{R}_{j-1}^{(s)} - \boldsymbol{R}_j^{(s)})] - \overline{\mathrm{Tr}}[(\boldsymbol{R}_j^{(r)} - \boldsymbol{R}_{j-1}^{(r)})\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}],
\end{aligned}$$

using the identities for each $r$ and $s$,

$$\begin{aligned}
\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j\boldsymbol{R}_{j-1}^{(s)}] &= \overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j\boldsymbol{R}_j^{(s)}] + \overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j(\boldsymbol{R}_{j-1}^{(s)} - \boldsymbol{R}_j^{(s)})],\\
\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}] &= \overline{\mathrm{Tr}}[\boldsymbol{R}_{j-1}^{(r)}\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}] + \overline{\mathrm{Tr}}[(\boldsymbol{R}_j^{(r)} - \boldsymbol{R}_{j-1}^{(r)})\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}].
\end{aligned}$$

To simplify the expression above, we defined

$$\boldsymbol{R}_j^{(r)} := \begin{cases} \boldsymbol{R}_j^\dagger|\boldsymbol{R}_j|^{r-1} \text{ or } \boldsymbol{R}_j|\boldsymbol{R}_j|^{r-1} & \text{if } r \text{ is odd}\\ |\boldsymbol{R}|_j^r & \text{if } r \text{ is even.}\end{cases}$$

When $r$ is odd, we overload the same symbol for the above two possible types of expressions, both of which fit into the same argument since we eventually take norms everywhere. Note that $|\boldsymbol{R}|_j^{r-1}\boldsymbol{R}_j^\dagger = \boldsymbol{R}_j^\dagger|\boldsymbol{R}|_j^{r-1}$ since they can be simultaneously diagonalized in the eigenbasis of $\boldsymbol{H}_j$.

It remains to study the concentration for each sequence $a_j, b_j, c_j$. The main observation is that each sequence forms a martingale difference sequence,

$$\mathbb{E}[a_j, b_j, c_j | \boldsymbol{A}'_{j-1}, \boldsymbol{A}_{j-1}, \ldots, \boldsymbol{A}'_1, \boldsymbol{A}_1] = 0.$$

Indeed, the sequence $a_j$ gives a martingale difference sequence since $\boldsymbol{A}_j$ is independent of $\boldsymbol{R}_j$ (similarly for the sequence $b_j$). Though complicated, the sequence $c_j$ also gives a martingale difference sequence since $c_j = (\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p) - a_j - b_j$ and Eq. (A9). Intuitively, the sequences $a_j$, $b_j$ have concentration controlled by the second moments, which fully exploits randomness in the random matrix $\boldsymbol{A}_j$ via Fact A.4. The sequence $c_j$ is higher order in $\boldsymbol{A}_j$ and $\boldsymbol{A}'_j$, and we simply use crude inequalities.

By uniform smoothness, the martingale difference sequence $c_j$ satisfies

$$\left|\sum_{j=1}^m c_j\right|_q^2 \le \sum_{j=1}^m (q-1)|c_j|_q^2.$$

We evaluate the individual $q$-norms,

$$\begin{aligned}
|c_j|_q &\le \sum_{\substack{r+s=p+1,\\ r,s\ge 1}} \left(|\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j(\boldsymbol{R}_{j-1}^{(s)} - \boldsymbol{R}_j^{(s)})]|_q + |\overline{\mathrm{Tr}}[(\boldsymbol{R}_j^{(r)} - \boldsymbol{R}_{j-1}^{(r)})\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}]|_q\right)\\
&\le \sum_{\substack{r+s+t=p+2,\\ r,s,t\ge 1}} \left(|\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}\boldsymbol{A}_j\boldsymbol{R}_j^{(s)}(\boldsymbol{A}'_j - \boldsymbol{A}_j)\boldsymbol{R}_{j-1}^{(t)}]|_q + |\overline{\mathrm{Tr}}[\boldsymbol{R}_j^{(r)}(\boldsymbol{A}_j - \boldsymbol{A}'_j)\boldsymbol{R}_{j-1}^{(s)}\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(t)}]|_q\right)\\
&\le \sum_{\substack{r+s+t=p+2,\\ r,s,t\ge 1}} \Big(|\|\boldsymbol{A}_j\|\|\boldsymbol{A}'_j - \boldsymbol{A}_j\|\|\boldsymbol{R}_j\|_q^{r/q}\|\boldsymbol{R}_j\|_q^{s/q}\|\boldsymbol{R}_{j-1}\|_q^{t/q}|_q\\
&\quad + |\|\boldsymbol{A}'_j\|\|\boldsymbol{A}'_j - \boldsymbol{A}_j\|\|\boldsymbol{R}_j\|_q^{r/q}\|\boldsymbol{R}_{j-1}\|_q^{s/q}\|\boldsymbol{R}_{j-1}\|_q^{t/q}|_q\Big)\\
&\le 4\cdot\frac{p(p+1)}{2}\frac{|\|\boldsymbol{A}_j\||_\infty^2}{\eta^2}|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q.
\end{aligned}$$

The second inequality further expands the resolvent difference. The third inequality uses Hölder's inequality for the trace. The last inequality uses Hölder's inequality, counts the combination of $r$, $s$, $t$ by $\binom{p+1}{2}$, assumes that $\boldsymbol{A}$ and $\boldsymbol{A}'$ have the same distributions, and assumes that $\boldsymbol{R}_j$, $\boldsymbol{R}_{j-1}$ have the same distribution as $\boldsymbol{R}$.

For the sequence $b_j$, we apply the scalar martingale inequality (Theorem A.2). We calculate the predictable quadratic variation

$$\begin{aligned}
\left|\sum_{j=1}^m \mathbb{E}_{j-1} b_j^* b_j\right|_{\frac{q}{2}} &\le \sum_{\substack{r+s=p+1,\\ r,s\ge 1}} \left|\sum_{j=1}^m \overline{\mathrm{Tr}}[\boldsymbol{R}_{j-1}^{(r)}\boldsymbol{A}'_j\boldsymbol{R}_{j-1}^{(s)}]^2\right|_q\\
&\le p^2\left|\sum_{j=1}^m \sup_{\||u\rangle\|=\||v\rangle\|=1} \mathbb{E}_{\boldsymbol{A}_j}|\langle u|\boldsymbol{A}_j|v\rangle|^2\cdot\overline{\mathrm{Tr}}[|\boldsymbol{R}_{j-1}|^{p+1}]^2\right|_{\frac{q}{2}}\\
&\le \frac{p^2}{\eta^2}|\overline{\mathrm{Tr}}[|\boldsymbol{R}|^p]|_q^2\cdot\sum_{j=1}^m \sigma_*(\boldsymbol{A}_j)^2
\end{aligned}$$

using Fact A.4 and the uniform bound on resolvent $\|\boldsymbol{R}\| \le \eta^{-1}$. We calculate the maximum by

$$|\max_{1\le j\le m}|b_j||_q \le \left(\sum_{j=1}^m |b_j|_q^q\right)^{1/q} \le p\,\frac{(\sum_{j=1}^m |\|\boldsymbol{A}_j\||_\infty^q)^{1/q}}{\eta}|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q.$$

The bound for $a_j$ is completely analogous. Combine the above estimates to obtain the desired result. ■

### a. Expected moments

To make use of Theorem A.1, we also need to estimate the expected moments via the comparison argument.

*Theorem A.3.* (Expected resolvent moments). For independent centered matrices $\boldsymbol{A}_i$, suppose the moments match that of idealized matrices $\tilde{\boldsymbol{A}}_i$,

$$\mathbb{E}\boldsymbol{A}_i = 0 \quad \text{and} \quad \mathbb{E}\boldsymbol{A}_i^{\otimes k} = \mathbb{E}\tilde{\boldsymbol{A}}_i^{\otimes k} \quad \text{for each } k = 1, \ldots, t \quad \text{and} \quad i = 1, \ldots, m.$$

Then,

$$||\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q^{1/p} - |\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q^{1/p}| \lesssim \frac{1 + m/pq}{\eta}\left(\frac{pqL}{\eta}\right)^{t+1} \quad \text{where } L := \max_i(|||\boldsymbol{A}_i|||_{3pq}, |||\tilde{\boldsymbol{A}}_i|||_{3pq}).$$

The symbol $\lesssim$ suppresses constants depending only on $t$.

*Proof.* We begin with a telescoping sum,

$$\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}|^p = \sum_{j=1}^{m}(\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}_{j-1}|^p) =: d_j \quad \text{with} \quad \boldsymbol{R}_j := \frac{1}{\boldsymbol{H}_j - \omega + \mathrm{i}\eta} \quad \text{and} \quad \boldsymbol{H}_j := \sum_{i=1}^{j}\tilde{\boldsymbol{A}}_i + \sum_{j+1}^{m}\boldsymbol{A}_i.$$

We move on to control the moments of trace $|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q^q$, which uses a similar argument as Theorems IV.1 and IV.2. We present the calculation for $t = 2$, but the general case is analogous.

*Step I: Single update error.* We again start with the telescoping sum

$$\mathbb{E}(\overline{\mathrm{Tr}}|\boldsymbol{R}|^p)^q - \mathbb{E}(\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p)^q = \sum_{j=0}^{m}\mathbb{E}(\overline{\mathrm{Tr}}|\boldsymbol{R}|_j^p)^q - \mathbb{E}(\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|_{j-1}^p)^q.$$

The Taylor expansions satisfy the bound from Fact A.1:

$$\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p = \overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p + \sum_{k=1}^{3p}\overline{\mathrm{Tr}}\,\boldsymbol{M}_k \quad \text{where } |\overline{\mathrm{Tr}}\boldsymbol{M}_k| \le (4p)^k\|\boldsymbol{R}_-\|_p^{p-k/3}\frac{\|\boldsymbol{A}_j\|_{3p}^k}{\eta^{4k/3}},$$

$$(\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p)^q = (\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p)^q + \sum_{k=1}^{3qp}f_k \quad \text{where } |f_k| \le (4qp)^k(\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p)^q\left(\|\boldsymbol{R}_-\|_p^{-1/3}\frac{\|\boldsymbol{A}_j\|_{3p}}{\eta^{4/3}}\right)^k.$$

The first inequality is analogous to the calculation (A6). Recall that the $p$-norms are normalized as $\|\boldsymbol{O}\|_p = (\overline{\mathrm{Tr}}|\boldsymbol{O}|^p)^{1/p}$. The second inequality proceeds with an additional $q^k$ factor. We then bound the expected increments by canceling the first- and second-order terms $f_1$ and $f_2$,

$$\begin{aligned}
|\mathbb{E}(\overline{\mathrm{Tr}}|\boldsymbol{R}|_j^p)^q - \mathbb{E}(\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|_{j-1}^p)^q| &\le \mathbb{E}\sum_{k=3}^{3qp}(4qp)^k|\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p|^q\left(\|\boldsymbol{R}_-\|_p^{-1/3}\frac{\|\boldsymbol{A}_j\|_{3p}}{\eta^{4/3}}\right)^k + (\boldsymbol{A}_j \mapsto \tilde{\boldsymbol{A}}_j)\\
&\le \frac{1}{4}|\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p|_q^{q-1/p}\left(\frac{8qp|||\boldsymbol{A}_j|||_{3pq}}{\eta^{4/3}}\right)^3 + \frac{1}{4}\left(\frac{8qp|||\boldsymbol{A}_j|||_{3pq}}{\eta^{4/3}}\right)^{3qp} + (\boldsymbol{A}_j \mapsto \tilde{\boldsymbol{A}}_j)\\
&\le \frac{1}{2}|\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p|_q^{q-1/p}\left(\frac{8qp\ell_j}{\eta^{4/3}}\right)^3 + \frac{1}{2}\left(\frac{8qp\ell_j}{\eta^{4/3}}\right)^{3qp}\\
&\le |\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p|_q^{q-1/p}\left(\frac{8qp\ell_j}{\eta^{4/3}}\right)^3 + \frac{3}{2}\left(\frac{8qp\ell_j}{\eta^{4/3}}\right)^{3qp}.
\end{aligned}$$

The second inequality uses Hölder's inequality with respect to the expectation, sums the geometric series, and uses the convenient estimate $|\|\boldsymbol{A}_j\|_{3p}|_{3pq} \le |||\boldsymbol{A}_j|||_{3pq}$. The third inequality sets $\ell_j := \max(|||\tilde{\boldsymbol{A}}_j|||_{3pq}, |||\boldsymbol{A}_j|||_{3pq})$. The last inequality establishes a self-bounding argument by

$$\mathbb{E}(\overline{\mathrm{Tr}}|\boldsymbol{R}_-|^p)^q \le 2\left(\mathbb{E}(\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p)^q + \left(\frac{8qp|||\boldsymbol{A}_j|||_{3pq}}{\eta^{4/3}}\right)^{3qp}\right)$$

in a similar vein as Eq. (A8).

*Step II: Solving the recursion.* Again, we simplify the recursion by defining scalar variables (and we also consider matching moments up to order $t$),

$$x_j := \mathbb{E}(\overline{\mathrm{Tr}}|\boldsymbol{R}_j|^p)^q, \qquad a_j := \frac{1}{2^{t-2}}\frac{(2(t+2)qp\ell_j)^{t+1}}{\eta^{t+2}},$$
$$\text{and} \quad b_j := \frac{3}{2^{t-1}}\frac{(2(t+2)qp\ell_j)^{(t+1)qp}}{\eta^{(t+2)qp}}$$

and $a_0 := b_0 := 0$. The updates in Eq. (A8) can then be written as a scalar recursion

$$|x_j - x_{j-1}| \le a_j \cdot x_{j-1}^{\frac{qp-1}{qp}} + b_j \quad \text{for each } j = 1, \ldots, m,$$

which, in fact, takes the exact same form as Theorem IV.2 up to $p \to qp$. Regardless, we write down the remaining calculation for completeness. As before, the arguments (Lemmas A.1 and A.2) give the bound for the endpoints,

$$\begin{aligned}|x_m^{1/pq} - x_0^{1/pq}| &\le \frac{1}{qp}\sum_{j=1}^m a_j + \left(\sum_{j=1}^m b_j\right)^{\frac{1}{qp}} \\ &\le 8m\frac{(qp)^t((t+2)L)^{t+1}}{\eta^{t+2}} \\ &\quad + \left(\frac{3}{2}\right)^{1/qp}(1+m/qp)\frac{(2(t+2)qpL)^{t+1}}{2^{(t-1)/qp}\eta^{t+2}} \\ &\le (2^{t+2} + 10m/qp)\frac{((t+2)qpL)^{t+1}}{\eta^{t+2}}.\end{aligned}$$

The second inequality uses the uniform bound $\ell_j = \max(|||\tilde{\boldsymbol{A}}_j|||_{3p}, |||\boldsymbol{A}_j|||_{3p}) \le L$. The last inequality drops the denominator $2^{(t-1)/qp} \ge 1$ and uses Young's inequality for $m^{1/qp} \le 1 + m/qp$. This is the second desired result. ■

## APPENDIX B: PROPERTIES OF THE GUE

In this section, we instantiate the properties of GUE matrices in the nonasymptotic regime. For our purposes, most of the quantities highly concentrate [up $\mathcal{O}(1/\mathrm{Poly}(N))$ deviations] and can be practically regarded as constants.

*Definition B.1.* (GUE ensemble). The N-by-N Gaussian unitary ensemble is a family of complex Hermitian random matrices specified by

$$\boldsymbol{H}_{ij} = \frac{g_{ij} + \mathrm{i}g'_{ij}}{\sqrt{2N}} \quad \text{if } j > i,$$
$$\boldsymbol{H}_{ii} = \frac{g_{ii}}{\sqrt{N}},$$

where $g_{ii}, g_{ij}, g'_{ij}$ are independent standard Gaussians.

### 1. *p*th moments

First, we given explicit bounds for the moments of the GUE ensemble. This kind of result is a consequence of classical explicit formulas for the moments of the GUE; for example, see Lemma 3.3.1 in Ref. [70]. It also follows from more recent work on nonasymptotic random matrix theory, such as Ref. [53]. For our purposes, it is convenient to derive the statement as a consequence of the main results (Theorem 2.7 in Ref. [54]), applied to a GUE matrix.

*Theorem B.1.* (Moment bounds). For even $p$ and a random GUE matrix with dimension $N$,

$$|||\boldsymbol{H}_{\mathrm{GUE}}|||_p \le 2\left(1 + \frac{(p/2)^{3/4}}{\sqrt{N}}\right).$$

The number 2 is exactly the maximal eigenvalue of the semicircular distribution with unit variance.

*Proof.* We can compare the following random matrices:

$$\begin{aligned}\boldsymbol{H}_{\mathrm{GUE}} = \boldsymbol{X} &= \sum_{j\ge i}\left(g_{ij}\frac{|i\rangle\langle j| + |j\rangle\langle i|}{\sqrt{2N}} + g'_{ij}\frac{\mathrm{i}(|i\rangle\langle j| - |j\rangle\langle i|)}{\sqrt{2N}}\right) \\ &\sim \sum_k g_k \boldsymbol{A}_k \\ \boldsymbol{X}_{\mathrm{free}} &:= \sum_k \boldsymbol{A}_k \otimes s_k,\end{aligned}$$

where $(g_{ij}, g'_{ij}, g_k)$ are independent standard normal variables and $(s_k)$ is a free semicircular family. The result (Theorem 2.7 in Ref. [54]) states that

$$\begin{aligned}|||\boldsymbol{X}|||_p - \|\boldsymbol{X}_{\mathrm{free}}\|_p| &\le 2(p/2)^{3/4}\sqrt{\sup_{\mathrm{Tr}|\boldsymbol{M}|^2=1}\sum_k |\mathrm{Tr}[\boldsymbol{A}_k\boldsymbol{M}]|^2} \\ &= 2(p/2)^{3/4}\sqrt{\frac{1}{N}\sup_{\mathrm{Tr}|\boldsymbol{M}|^2=1}\sum_{i,j}|M_{ij}|^2} \\ &= \frac{2^{1/4}p^{3/4}}{\sqrt{N}}.\end{aligned}$$

Recall the unconditional bound $\|\boldsymbol{X}_{\mathrm{free}}\| \le 2\|\mathbb{E}\boldsymbol{X}^2\| = 2$ to conclude the proof (p. 208 in Ref. [71]). ■

### 2. Resolvent moments

We calculate the resolvent moments for GUE matrices. Recall that

$$\boldsymbol{R} := \frac{1}{\boldsymbol{H} - \omega + \mathrm{i}\eta}.$$

*Fact B.1.* (Consequence of Corollary 11.4 in Ref. [72]). For a random instance $\boldsymbol{H}_{\mathrm{GUE}}$ of the GUE with dimension $N$, define the empirical spectral density

$$\rho(E) := \frac{1}{N}\sum_{i=1}^{N}\delta(E-\lambda_i(\boldsymbol{H}_{\mathrm{GUE}})).$$

Then, there is an absolute constant $c$ such that we have

$$\sup_{E\in\mathbb{R}}\left|\int_{-\infty}^{E}(\rho(E')-\rho_{sc}(E'))dE'\right| \le \frac{c}{\sqrt{N}} \tag{B1}$$

with probability of at least $1-(1/N)$.

*Proof.* In the setting of Corollary 11.4 in Ref. [72]), set $D=1$ and $\epsilon=1/2$. The range $|E|\le 10$ extends to infinity since the semicircle density $\rho_{sc}$ is supported on $[-2,2]$ and the error must be decreasing for $|E|\ge 2$. ∎

*Corollary B.1.* (Resolvent moments for the GUE). There is an absolute constant $c$ such that for each $\omega$, $\eta$, we have

$$|\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_{\omega,\eta}(\boldsymbol{H}_{\mathrm{GUE}})|^p - S_{\omega,\eta,p}| \le \frac{c}{\eta^p\sqrt{N}}$$
$$\text{where } S_{\omega,\eta,p} := \int_{-2}^{2}\frac{\sqrt{4-x^2}}{2\pi}\frac{1}{|x-\omega+\mathrm{i}\eta|^p}dx.$$

*Proof.* Let $f(E)=|E-\omega+\mathrm{i}\eta|^{-p}$. Then,

$$\begin{aligned}
&|\mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}_{\mathrm{GUE}}|^p - S_p^p| \\
&\le \mathbb{E}\left|\int_{-\infty}^{\infty} f(E)(\rho_{\mathrm{GUE}}(E)-\rho_{sc}(E))dE\right| \\
&= \mathbb{E}\left|\int_{-\infty}^{\infty} f'(E)\int_{-\infty}^{E}(\rho_{\mathrm{GUE}}(E')-\rho_{sc}(E'))dE'dE\right| \\
&\le \left(1-\frac{1}{N}\right)\cdot\frac{c}{\sqrt{N}}\left(\int_{-\infty}^{\infty}|f'(E)|dE\right) + \frac{1}{N}\cdot 2\max_{E\in\mathbb{R}}|f(E)| \\
&\le \frac{2c}{\eta^p\sqrt{N}} + \frac{2}{\eta^p N}.
\end{aligned}$$

The first inequality uses integration by parts and the boundary value $\int_{-\infty}^{\infty}(\rho_{\mathrm{GUE}}(E')-\rho_{sc}(E'))dE'=0$. The third line uses Fact B.1 to handle the high probability event (B1). To reach the last line, we compute the integral using the fact that the resolvent power $f(E)=|E-\omega+\mathrm{i}\eta|^{-p}$, increasing for $E<\omega$ and decreasing for $E>\omega$, so the integral equals $2f(\omega)=2\eta^{-p}$. To bound the maximum, note that $0\le f(E)\le\eta^{-p}$. Finally, increase the constant $c$ as needed to combine the terms. ∎

*Proposition B.1.* (GUE: Concentration for resolvent moments). For a matrix with jointly Gaussian entries

$$\boldsymbol{H} := \sum_i g_i\boldsymbol{A}_i \quad \text{where } g_i\sim\mathcal{N}(0,1)$$

and even $p$, the spectral density (probed by resolvent powers) is concentrated as

$$|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q \lesssim \sqrt{q}\,\frac{p}{\eta^{p+1}}\sigma_*$$
$$\text{where } \sigma_* := \sqrt{\sup_{\|\boldsymbol{w}\|=\|\boldsymbol{v}\|=1}\sum_i|\langle\boldsymbol{v}|\boldsymbol{A}_i|\boldsymbol{w}\rangle|^2}.$$

For a GUE matrix, we have $\sigma_*=N^{-1/2}$.

*Proof of Proposition B.1.* This is a standard application of Gaussian concentration inequalities. We cannot find the particular function of interest elsewhere, so we include a derivation adapted from Ref. [55]. Some of the estimates could be loose, in general, but suffice for our purposes. Consider the function

$$f(\boldsymbol{x}) := \overline{\mathrm{Tr}}\left|\sum_i x_i\boldsymbol{A}_i - \omega + \mathrm{i}\eta\right|^{-p}.$$

We bound the Lipschitz constant

$$\begin{aligned}
|f(\boldsymbol{x})-f(\boldsymbol{y})| &= |\bar{\mathrm{T}}\mathrm{r}|\boldsymbol{R}_{\boldsymbol{x}}|^p - \bar{\mathrm{T}}\mathrm{r}|\boldsymbol{R}_{\boldsymbol{y}}|^p| \\
&= |\bar{\mathrm{T}}\mathrm{r}[\boldsymbol{R}_{\boldsymbol{x}}^\dagger\boldsymbol{R}_{\boldsymbol{x}}|\boldsymbol{R}_{\boldsymbol{x}}|^{p-2} - \boldsymbol{R}_{\boldsymbol{y}}^\dagger\boldsymbol{R}_{\boldsymbol{x}}|\boldsymbol{R}_{\boldsymbol{x}}|^{p-2}] + \cdots| \\
&\le |\bar{\mathrm{T}}\mathrm{r}[\boldsymbol{R}_{\boldsymbol{x}}^\dagger(\boldsymbol{H}_{\boldsymbol{y}}-\boldsymbol{H}_{\boldsymbol{x}})\boldsymbol{R}_{\boldsymbol{y}}^\dagger\boldsymbol{R}^x|\boldsymbol{R}_{\boldsymbol{x}}|^{p-2}] + \cdots| \\
&\le \frac{p}{\eta^{p+1}}\|\boldsymbol{H}_{\boldsymbol{x}}-\boldsymbol{H}_{\boldsymbol{y}}\|.
\end{aligned}$$

The second equality is a telescoping sum. The first inequality uses the identity $\boldsymbol{A}^{-1}-\boldsymbol{B}^{-1}=\boldsymbol{B}^{-1}(\boldsymbol{B}-\boldsymbol{A})\boldsymbol{A}^{-1}$. The last inequality uses the triangle inequality, the uniform bound that $\|\boldsymbol{R}\|\le\eta^{-1}$, and the coarse bound $\|\boldsymbol{A}\|_1\le\|\boldsymbol{A}\|$, which holds because we are using normalized Schatten norms. Lastly, we relate the operator norm to the Euclidean norm of the coefficients,

$$\begin{aligned}
\|\boldsymbol{H}_{\boldsymbol{x}}-\boldsymbol{H}_{\boldsymbol{y}}\| &= \left\|\sum_i(x_i-y_i)\boldsymbol{A}_i\right\| = \sup_{\boldsymbol{w},\boldsymbol{v}}\sum_i\langle\boldsymbol{v}|(x_i-y_i)\boldsymbol{A}_i|\boldsymbol{w}\rangle \\
&\le \left(\sup_{\boldsymbol{w},\boldsymbol{v}}\sum_i|\langle\boldsymbol{v}|\boldsymbol{A}_i|\boldsymbol{w}\rangle|^2\right)^{1/2}\cdot\|\boldsymbol{x}-\boldsymbol{y}\|_{\ell_2} \\
&= \sigma_*\cdot\|\boldsymbol{x}-\boldsymbol{y}\|_{\ell_2}.
\end{aligned}$$

The inequality is the Cauchy-Schwarz one. To conclude the proof, recall that an $L$-Lipschitz function of a standard Gaussian vector is $L^2$-sub-Gaussian (Theorem 5.6 in Ref. [73]). ∎

The preceding concentration argument also allows us to bound

$$|\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q \le \mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p + |\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q.$$

GUE matrices have strong concentration properties, so the right-hand side of the previous display is always dominated by the expectation term in our applications.

## APPENDIX C: PROPERTIES OF THE PAULI STRING ENSEMBLE

In this section, we compare the properties of the Pauli string ensemble with the GUE. Recall that

$$\boldsymbol{H}_{\mathrm{PS}} = \sum_{j=1}^{m} \boldsymbol{A}_j \quad \text{where } \boldsymbol{A}_j \overset{\mathrm{IID}}{\sim} \frac{1}{\sqrt{m}} \cdot \pm\{\mathbf{I}, \boldsymbol{\sigma}^x, \boldsymbol{\sigma}^y, \boldsymbol{\sigma}^z\}^{\otimes n}.$$

We will compare the Pauli string ensemble with the GUE Hamiltonian,

$$\boldsymbol{H}_{\mathrm{GUE}} = \sum_{j=1}^{m} \tilde{\boldsymbol{A}}_j \quad \text{where } \tilde{\boldsymbol{A}}_j \overset{\mathrm{IID}}{\sim} \frac{1}{\sqrt{m}} \boldsymbol{H}_{\mathrm{GUE}}.$$

Comparing the $p$th moments controls the spectral norm, while the resolvent moments control the spectral density.

### 1. Moments and the spectral norm

We use the $p$th moments to bound the spectral norm.

*Theorem III.1.* ($p$-norms and operator norm). Let $p \in 2\mathbb{N}$ be an even natural number. The random Pauli string ensemble (2) satisfies the norm bound

$$\left| |||\boldsymbol{H}_{\mathrm{PS}}|||_p - |||\boldsymbol{H}_{\mathrm{GUE}}|||_p \right| \lesssim \left( \frac{p^{3/4}}{m^{1/4}} + \frac{p}{\sqrt{m}} \right) \left( 1 + \frac{p^{3/4}}{2^{n/2}} \right).$$

The symbol $\lesssim$ suppresses constant factors. Furthermore, for $0 \le \epsilon \le 1/2$ and $m \le 2^{2n}$, there exist constants $c_1, c_2 > 0$ where

$$m \ge c_1 \frac{n^3}{\epsilon^4} \quad \text{ensures} \quad \Pr\left( \|\boldsymbol{H}_{\mathrm{PS}}\| \ge 2(1+\epsilon) \right) \le \exp(-c_2 n).$$

To obtain a smaller multiplicative error $\epsilon$, note that the parameters $p$ and $m$ only need to increase at a polynomial rate (as a function of the number $n$ of sites and the parameter $\epsilon$).

*Proof of Theorem III.1.* The $p$-norm estimates use Theorem IV.1. To obtain the desired tail bounds, recall the $p$-norm for GUE matrices,

$$|||\boldsymbol{H}_{\mathrm{GUE}}|||_p \le 2 \cdot \left( 1 + \frac{(p/2)^{3/4}}{\sqrt{N}} \right).$$

By Markov's inequality,

$$\begin{aligned}
&\Pr\left( \|\boldsymbol{H}_{\mathrm{PS}}\| \ge t \right) \\
&\le \frac{\mathbb{E}\|\boldsymbol{H}_{\mathrm{PS}}\|^p}{t^p} \le N \frac{|||\boldsymbol{H}_{\mathrm{PS}}|||_p^p}{t^p} \\
&\le \left( \frac{\mathrm{e}^{\log(N)/p}}{t} \left( 2 + \mathcal{O}\left( \frac{p^{3/4}}{\sqrt{N}} + \frac{p^{3/4}}{m^{1/4}} + \frac{p}{\sqrt{m}} \right) \right) \right)^p \\
&\le \left( \frac{1+\epsilon/2}{1+\epsilon} \right)^p \le \mathrm{e}^{-c_1 \log(N)/4}, \qquad \text{setting } t = 2(1+\epsilon).
\end{aligned}$$

The second inequality converts the operator norm to the $p$th moments by $\|\boldsymbol{H}\|^p \le \mathrm{Tr}\boldsymbol{H}^p$, and the third inequality keeps the leading-order terms via the notation $\mathcal{O}(\cdot)$. The third line uses $m \le N^2$ and chooses appropriate parameters $p = c\log(N)/\epsilon$ and $m = c_1[p^4/\log(N)]$ so that the numerator is bounded by $2(1+\epsilon/2)$. The last inequality uses the elementary estimate $[(1+\epsilon/2)/1+\epsilon] \le \mathrm{e}^{-\epsilon/4}$ for $\epsilon \le 1/2$. Note that $N = 2^n$ to obtain the desired result. ■

### 2. Abundance of low-energy states and success of phase estimation

In this section, we combine the bounds on the minimal eigenvalue (Theorem III.1) and the density of states (Theorem A.1) to estimate the number of states with approximately the ground energy. This immediately implies that applying phase estimation on the maximally mixed state returns a low-energy state as a witness with a non-negligible success probability.

*Theorem II.1.* (Low-energy states have low complexity). For any accuracy $\epsilon \ge 2^{-n/c_1}$, let $\boldsymbol{H}_{\mathrm{PS}}$ be drawn from the Pauli string ensemble (2) with

$$m = \left\lfloor c_2 \frac{n^5}{\epsilon^4} \right\rfloor$$

terms. Then, the following statement holds with probability of at least $1 - \mathrm{e}^{-c_3 n^{1/3}}$ over a random draw $\boldsymbol{H}_{\mathrm{PS}}$ from the Pauli string ensemble. We can prepare a low-energy state $\boldsymbol{\rho}$ such that

$$\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) \le \mathrm{Tr}[\boldsymbol{\rho}\boldsymbol{H}_{\mathrm{PS}}] \le (1-\epsilon)\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})$$

using a circuit of size $G = \mathrm{Poly}(n, \epsilon^{-1})$. The quantities $c_1$, $c_2$, and $c_3$ are absolute constants, and $\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})$ denotes the smallest eigenvalue of $\boldsymbol{H}_{\mathrm{PS}}$ (which is typically negative).

*Proof of Theorem II.1.* We are interested in controlling the integrated density of states in an energy window *just* above the ground state,

$$\frac{\#\{|E\rangle : E \le E_0\}}{N} \quad \text{where } E_0 = \lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) + \mathcal{O}(\epsilon).$$

This relative number of states guarantees the success probability that phase estimation returns a state $\boldsymbol{\rho}$ within this energy range (up to the phase estimation precision). The idea is to construct a proxy for the low-energy projector by consecutive local resolvents, which are essentially local energy subspace projectors (Fig. 4). Consider

$$\begin{aligned}
\eta^p \sum_{2 \le \ell \bar{\omega} \le E_0} |\boldsymbol{R}_{\ell\bar{\omega},\eta}|^p &=: \sum_E |E\rangle\langle E| q_{E_0}(E) \\
&=: \boldsymbol{Q}(E_0) \quad \text{as a proxy for the projector} \\
&\quad \times \sum_E |E\rangle\langle E| \mathbb{1}(E \le E_0)
\end{aligned}$$

at low energy $E_0 := -(1-\epsilon/3)\cdot 2$. The resolvents are spaced appropriately,

$$|\boldsymbol{R}_{\ell\bar{\omega},\eta}|^p := \frac{1}{|\boldsymbol{H}-\ell\bar{\omega}+i\eta|^p} \quad \text{for } \ell\in\mathbb{Z}, \qquad \bar{\omega} := \frac{\epsilon}{\sqrt{p}}\cdot 2,$$
$$\eta = \frac{\epsilon}{3}\cdot 2, \quad \text{and} \quad p = \lfloor c_1 \log(N)\rfloor.$$

We will calculate $\mathbb{E}\overline{\mathrm{Tr}}\,\tilde{\boldsymbol{Q}}(E_0)$ for GUE, show that $\overline{\mathrm{Tr}}\boldsymbol{Q}(E_0)$ for the Pauli string ensemble takes comparable values, and then extract the low-energy density of state.

*GUE values.* Recall the GUE resolvent values (Corollary B.1)

$$\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}_{\ell\bar{\omega},\eta}|^p \geq S_p^p/2 \gtrsim \frac{1}{\eta^p}\sqrt{2-(\ell-1)\bar{\omega}}\cdot\frac{\epsilon}{\sqrt{p}}$$

where $S_{\ell\bar{\omega},\eta,p}$ is an integral over the semicircle defined in Corollary B.1. The second inequality uses the GUE estimate (Corollary B.1) and imposes the simplifying constraint $N \geq \mathrm{const}/\epsilon^4$ such that $S_{\ell\bar{\omega},\eta,p}/2 \geq \mathrm{const}N^{-1/2}$ even near the spectral edge. The third inequality evaluates $S_{\ell\bar{\omega},\eta,p}$.

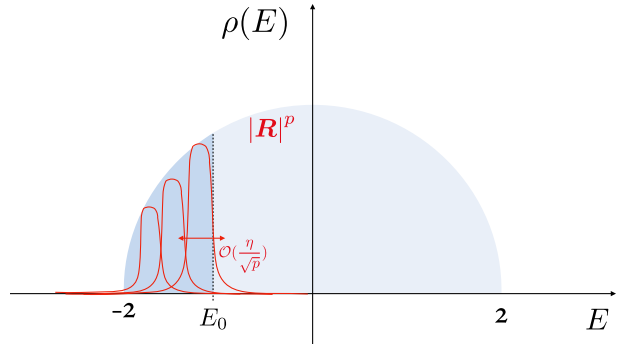


FIG. 4. Probing the low-energy states $E \leq E_0$ via consecutive resolvent powers.

Apply a crude Riemann sum over the semicircular density near the edge and drop constants to obtain bounds on the expected value,

$$\mathbb{E}\overline{\mathrm{Tr}}\,\tilde{\boldsymbol{Q}}(E_0) := \sum_{2\leq\ell\bar{\omega}\leq E_0}\eta^p\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}_{\ell\bar{\omega},\eta}|^p = \Omega(\epsilon\sqrt{\epsilon}) \quad \text{using} \quad \int\sqrt{x}dx = \frac{2}{3}x\sqrt{x}. \tag{C1}$$

*Pauli's string ensemble values.* Taking a crude union bound over the local resolvents, we ensure all of them are at least half of the GUE expectation with high probability,

$$\begin{aligned}\Pr\left(\overline{\mathrm{Tr}}\boldsymbol{Q}(E_0) \leq \frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}\tilde{\boldsymbol{Q}}(E_0)\right) &\leq \sum_{\ell\bar{\omega}\leq E_0}\Pr\left(\overline{\mathrm{Tr}}|\boldsymbol{R}_{\ell\bar{\omega},\eta}|^p \leq \frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}_{\ell\bar{\omega},\eta}|^p\right)\\ &\lesssim \sqrt{p}\mathrm{e}^{-\Omega(\log(N)^{1/3})} \ll 1 \quad \text{(Claim)}.\end{aligned} \tag{C2}$$

*Extracting the density of states.* Assuming the claim holds, it remains to extract the spectral density from the low-energy subspace proxy $\boldsymbol{Q}(E_0)$. We crudely estimate the function $q_{E_0}(E)$ by splitting the spectrum by a step function

$$\begin{aligned}\overline{\mathrm{Tr}}\boldsymbol{Q}(E_0) &= \frac{1}{N}\sum_E q_{E_0}(E)\mathbb{1}(E\leq E_0+\eta) + q_{E_0}(E)\mathbb{1}(E\geq E_0+\eta)\\ &\lesssim \frac{1}{N}\sum_E \mathbb{1}(E\leq E_0+\eta) + \mathrm{e}^{-\Omega(p)}\mathbb{1}(E\geq E_0+\eta).\end{aligned}$$

The second inequality uses that $q_{E_0}(E)\lesssim 1$ and that $q_{E_0}(E) = \mathrm{e}^{-\Omega(p)}$ for $E\geq E_0+\eta$. Rearrange to bound the number of low-energy states,

$$\begin{aligned}\frac{\#\{|E\rangle : E\leq -(1-2\epsilon/3)\cdot 2\}}{N} &= \sum_E\frac{\mathbb{1}(E\leq E_0+\eta)}{N} \gtrsim \eta^p\overline{\mathrm{Tr}}\boldsymbol{Q}(E_0) - \mathrm{e}^{-\Omega(p)}\\ &\gtrsim \epsilon\sqrt{\epsilon} - \mathrm{e}^{-\Omega(\log(N))} \quad [\text{with high probability (C2)}]\\ &\gtrsim \epsilon\sqrt{\epsilon} \quad [\text{setting } \log(N)\geq \mathrm{const}\log(1/\epsilon)].\end{aligned}$$

The second inequality uses concentration (C2) and plugs in the GUE value $\mathbb{E}\overline{\mathrm{Tr}}\,\tilde{\boldsymbol{Q}}(E_0)$ [Eq. (C1)], which is approximately the semicircle integral. The third inequality imposes additional constraints $\log(N)\geq\mathrm{const}\log(1/\epsilon)$, which can be combined with $N\geq\mathrm{const}/\epsilon^4$ by $N\geq\epsilon^{-c_1}$ for some constant $c_1$. Combine this with the tail bound (Theorem III.1) for the

operator norm $\|\boldsymbol{H}_{\mathrm{PS}}\|$ for $\epsilon' = \epsilon/6$ [using that $\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) \geq -\|\boldsymbol{H}_{\mathrm{PS}}\|$ and that $(1 - 2/3\epsilon/1 + \epsilon/6) \geq 1\text{–}5\epsilon/6$] to find that there are many low-energy states,

$$\frac{\#\{|E\rangle : E \leq (1 - 5\epsilon/6)\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})\}}{N} \geq \Omega(\epsilon^{3/2}) \quad \text{with high probability} \quad 1 - \mathrm{e}^{-\Omega(\log(N)^{1/3})},$$

drawing from the Hamiltonian ensemble. Consequently, performing phase estimation with energy resolution $\mathcal{O}(\epsilon)$ on the maximally mixed state prepares a low-energy state $\boldsymbol{\rho}$ such that

$$\mathrm{Tr}[\boldsymbol{\rho}\boldsymbol{H}_{\mathrm{PS}}] \leq (1 - 5\epsilon/6)\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) \quad \text{with success probability} \quad \Omega(\epsilon^{3/2}),$$

which costs $\mathrm{Poly}(m, 1/\epsilon)$ using any off-the-shelf quantum simulation algorithm such as Trotter [3], Qubitization [74], or qDrift [40] for Hamiltonian simulation within phase estimation. We may amplify the success probability to $1 - \epsilon/(6\sqrt{m})$ using $\mathcal{O}(\epsilon^{-3/2}\log(\sqrt{m}/\epsilon))$ repeats. When all repetitions fail, we output the maximally mixed state, which, even in the worst case, has energy upper bounded by $\sqrt{m}$. The resulting output state uses gate complexity

$$G = (\text{number of repeats}) \cdot (\text{QPE cost}) = \Omega\left(\mathrm{Poly}\left(m, \frac{1}{\epsilon}\right)\right),$$

and the energy of the output state [75] is at most $(1 - \epsilon)\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})$, which is the desired result.

*Proof of Claim.* It remains to prove the claim (C2); this is where we invoke concentration for the resolvent (Theorem A.1). To reiterate, for each $\boldsymbol{R} = \boldsymbol{R}_{\ell\bar{\omega},\eta}$, we want to show

$$\text{(WTS)} \quad \Pr\left(\overline{\mathrm{Tr}}|\boldsymbol{R}_{\ell\omega,\eta}|^p \leq \frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}_{\ell\omega,\eta}|^p\right) \leq \mathrm{e}^{-\Omega(\log(N)^{1/3})} \quad \text{for each } \ell, \bar{\omega} \tag{C3}$$

for parameters $\eta = (\epsilon/3)\cdot 2$, $p = \lfloor c_1 \log(N)\rfloor$, and

$$q = \theta(\log(N)),$$
$$m = \Omega\left(\frac{\log(N)^5}{\epsilon^4}\right).$$

For each $\ell, \bar{\omega}$, we use the shorthand $\boldsymbol{R} = \boldsymbol{R}_{\ell\bar{\omega},\eta}$ and rearrange

$$\begin{aligned}\Pr\left(\overline{\mathrm{Tr}}|\boldsymbol{R}|^p \leq \frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p\right) &\leq \Pr\left(|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p| \geq \frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - |\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|\right)\\ &\leq \left(\frac{|\overline{\mathrm{Tr}}[|\boldsymbol{R}|^p - |\boldsymbol{R}'|^p]|_q}{\frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - |\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|}\right)^q.\end{aligned}$$

The last inequality is Markov's. We proceed in bounding the denominator

$$\begin{aligned}\frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - |\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{R}|^p| &= \frac{1}{2}|||\tilde{\boldsymbol{R}}|||_p^p\left(1 - 2\left|1 - \frac{|||\boldsymbol{R}|||_p^p}{|||\tilde{\boldsymbol{R}}|||_p^p}\right|\right)\\ &\geq \frac{1}{2}|||\tilde{\boldsymbol{R}}|||_p^p(1 - \mathcal{O}(\mathrm{e}^{\mathcal{O}(1/\log(N))} - 1)).\end{aligned}$$

The inequality uses $|1 - (1 + x)^p| \leq |1 - \mathrm{e}^{|x|p}|$ for

$$x := \left|\frac{|||\boldsymbol{R}|||_p}{|||\tilde{\boldsymbol{R}}|||_p} - 1\right| = \frac{|\,|||\boldsymbol{R}|||_p - |||\tilde{\boldsymbol{R}}|||_p|}{|||\tilde{\boldsymbol{R}}|||_p} \lesssim \frac{p^3 + p^4/m}{m\eta^4}\left(1 + \frac{p^{3/4}}{N^{1/2}}\right) = \mathcal{O}\left(\frac{1}{p\log(N)}\right),$$

which compares resolvent moments (Theorem IV.1), $|||\boldsymbol{H}_{\mathrm{GUE}}|||_p \leq 2\cdot[1 + (p/2)^{3/4}/\sqrt{N}]$, $|||\tilde{\boldsymbol{R}}|||_p \geq 1/2\eta$, and the values of the parameters $q$, $p$, $m$, $\eta$.

For the numerator, we evaluate Theorem A.1,

$$\begin{aligned}|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p - \overline{\mathrm{Tr}}|\boldsymbol{R}'|^p|_q &\le |\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q\left(\frac{\sqrt{q}p^2}{m\eta^2}+\frac{\sqrt{q}p}{\eta\sqrt{N}}+\frac{qpm^{1/q-1}}{\eta}\right)\\ &\le \left(|\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q^{1/p}+\mathcal{O}\left(\frac{p^3q^3+p^4q^4/m}{m\eta^5}\left(1+\frac{(2qp)^{3/4}}{\sqrt{N}}\right)^4\right)\right)^p\cdot\left(\frac{\sqrt{q}p^2}{m\eta^2}+\frac{\sqrt{q}p}{\eta\sqrt{N}}+\frac{qpm^{1/q-1}}{\eta}\right)\\ &\lesssim \left(\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p+\mathcal{O}\left(\frac{\log(N)^{1+1/6}}{\eta^{p+1}\sqrt{N}}\right)\right)\frac{\epsilon^2}{\log(N)^{3-1/6}}.\end{aligned}$$

The first inequality evaluates the second-moment quantity

$$\sigma_*(\boldsymbol{A}_i)^2=\frac{1}{m}\sup_{\||u\rangle\|=\||v\rangle\|=1}\mathbb{E}_{\boldsymbol{A}_i}|\langle u|\boldsymbol{A}_i|v\rangle|^2=\frac{1}{m}\sup_{\||u\rangle\|=\||v\rangle\|=1}\langle u|\overline{\mathrm{Tr}}[|v\rangle\langle v|]|u\rangle=\frac{1}{mN}.$$

The second inequality compares $|\overline{\mathrm{Tr}}|\boldsymbol{R}|^p|_q^{1/p}$ with $|\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q^{1/p}$ (Theorem A.3). The last inequality plugs in the values of $q$, $p$, $\eta$ in terms of $N$, $\epsilon$ and uses concentration for the Gaussian resolvent $|\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q \le \mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p + |\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p - \mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p|_q \le \mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p + \mathcal{O}(\sqrt{q}p/\eta^{p+1}\sqrt{N})$ (Proposition B.1). We obtain the desired result (C3),

$$\begin{aligned}\Pr\left(\overline{\mathrm{Tr}}|\boldsymbol{R}|^p\le\frac{1}{2}\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p\right)&\le\left(\frac{(\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p+\mathcal{O}(\frac{\log(N)^{1+1/6}}{\eta^{p+1}\sqrt{N}}))\frac{\epsilon^2}{\log(N)^{3-1/6}}}{\frac{1}{2}|||\tilde{\boldsymbol{R}}|||_p^p(1-\mathcal{O}(\mathrm{e}^{\mathcal{O}(1/\log(N))}-1)).}\right)^q\le \mathrm{e}^{-\Omega(\log(N)^{1/3})}\\ &\le \mathrm{e}^{-c_3n^{1/3}}.\end{aligned}$$

The second inequality uses $\mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}|^p = \mathbb{E}\overline{\mathrm{Tr}}|\tilde{\boldsymbol{R}}_{\ell\overline{\omega},\eta}|^p \gtrsim \sqrt{\epsilon/\sqrt{p}}^3 \gtrsim \sqrt{\epsilon/\sqrt{\log(N)}}^3$ so that the base is smaller than 1 for large enough $N$. The last inequality introduces an explicit constant. This concludes the proof of Claim (C2). ■

## APPENDIX D: CIRCUIT-SIZE LOWER BOUNDS FOR LOW-ENERGY STATES

The GUE is unitarily invariant; thus, any subspace of low-energy eigenvectors will be Haar random. Consequently, preparing a low-energy state of the GUE necessarily requires large circuit complexity. Does enough of this randomness carry over to sparse Hamiltonian ensembles such that a similar statement can be made? In this section, we prove that, with high probability over the ensemble, any circuit that prepares a low-energy state necessarily has a large circuit size. To show Theorem II.2, we split this into the following two statements. We first calculate the expected norm.

*Proposition D.1.* For the Pauli string ensemble given in Eq. (2), $\mathbb{E}\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}}) \le -1/2$.

*Proof.* By symmetry of the ensemble,

$$\begin{aligned}2\mathbb{E}|\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})| &= \mathbb{E}|\lambda_{\min}(\boldsymbol{H}_{\mathrm{PS}})| + \mathbb{E}|\lambda_{\max}(\boldsymbol{H}_{\mathrm{PS}})|\\ &\ge \mathbb{E}\|\boldsymbol{H}_{\mathrm{PS}}\| \ge \mathbb{E}\|\boldsymbol{H}_{\mathrm{PS}}\|_2 = 1.\end{aligned}$$

The last inequality holds for the normalized Schatten $p$-norms, defined in Eq. (4). ■

*Lemma D.1.* (Small circuit fails to give low-energy states). Fix a circuit architecture of two-qubit gates with size $G$ with the initial state $|0\rangle$ and consider the family of all reachable states $\mathrm{Circ}(G)$. Suppose $m \le \epsilon^2 N^2$; then, there are constants $c_1$, $c_2$ such that

$$\begin{aligned}&G \le c_1\epsilon\sqrt{m}\log^{-1}(m) \quad \text{implies}\\ &\qquad \Pr\left[\inf_{|\psi\rangle\in\mathrm{Circ}(G)}\langle\psi|\boldsymbol{H}_{\mathrm{PS}}|\psi\rangle \le -\epsilon\right] \le \exp\left(-c_2\epsilon\sqrt{m}\right).\end{aligned}$$

*Proof.* The proof uses concentration inequality for any nonrandom state and bootstraps for an epsilon net by the union bound. Consider the random variable associated with a fixed pure state,

$$\begin{aligned}\langle\boldsymbol{H}_{\mathrm{PS}}\rangle_\psi = \sum_{i=1}^m a_i \quad &\text{where } \langle\boldsymbol{H}_{\mathrm{PS}}\rangle_\psi := \langle\psi|\boldsymbol{H}_{\mathrm{PS}}|\psi\rangle\\ \text{and} \quad a_i &:= \frac{1}{\sqrt{m}}\langle\boldsymbol{A}_i\rangle_\psi.\end{aligned}$$

We use the symmetry of the Pauli string ensemble $\boldsymbol{H}_{\mathrm{PS}} \sim -\boldsymbol{H}_{\mathrm{PS}}$ to consider the more intuitive maximization problem. We also drop the subscript $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{PS}}$ for simplicity.

*Variance of a Pauli string.* First, we calculate the variance of a random Pauli string with arbitrary fixed input

$$\mathbb{E}_{\boldsymbol{A}_i}|\langle \boldsymbol{A}_i\rangle_\psi|^2 = \mathbb{E}_{\boldsymbol{A}_i}\langle\psi|\boldsymbol{A}_i|\psi\rangle\langle\psi|\boldsymbol{A}_i|\psi\rangle = \frac{1}{N}\langle\psi|\mathrm{Tr}[|\psi\rangle\langle\psi|]|\psi\rangle = \frac{1}{N} \ll 1.$$

The second equality evaluates the second moment of the Pauli string ensemble. Notice that the variance is exponentially smaller than the maximal value of the random variable. In other words, a fixed input state is very unlikely to "align" with the random Pauli string $\boldsymbol{A}_i$. Intuitively, the random Pauli strings are very *noncommutative* and thus cannot be simultaneously diagonalized on a preferred basis.

*Variance of the total energy.* From the variance of the individual terms, we may obtain a tail bound for the sum via Bernstein's inequality,

$$\Pr[\langle \boldsymbol{H}\rangle_\psi \geq t] \leq \exp\left(\frac{-t^2/2}{v + Lt/3}\right) \quad \text{where } v \coloneqq \sum_i^m \mathbb{E}[a_i^2] = \frac{1}{N} \quad \text{and} \quad a_i \leq L \coloneqq 1/\sqrt{m}. \tag{D1}$$

In other words, any deterministic input state (that does not correlate with the Hamiltonian) is very likely to have small energies.

*Union bound over an epsilon net.* By a union bound, good concentration implies that the energies must be *simultaneously* small for a large family of deterministic input states, specifically, the input states drawn from an epsilon net over a small circuit. For a circuit consisting of $G$ gates, there exists an

$$\left(\frac{\epsilon}{2\sqrt{m}}\right)\text{net}\{|\psi_i\rangle\} \quad \text{for } \mathrm{Circ}(G) \quad \text{with cardinality} \quad \#\{|\psi_i\rangle\} \leq \exp\left(\mathcal{O}(G\log(G\sqrt{m}/\epsilon))\right),$$

which is justified as follows. Any circuit with $G$ two-qubit gates is equivalently given by a product of fixed CNOT gates interspersed with $KG$ single-parameter single-qubit rotation gates by certain angles, with $K = \mathcal{O}(1)$. If we cast a $(\epsilon/(2\sqrt{m}KG))$ net over the interval $[0, 2\pi]$ for each of these $KG$ rotation angles, the set of circuits we generate will form an $\epsilon/(2\sqrt{m})$ net over states in $\mathrm{Circ}(G)$, and the cardinality of the set is at $(4\pi\sqrt{m}KG/\epsilon)^{KG}$. One of the elements of this net is guaranteed to approximate the state $|\psi\rangle \in \mathrm{Circ}(G)$ that achieves the supremum of $\langle \boldsymbol{H}\rangle_\psi$ up to error $\epsilon/(2\sqrt{m})$, and since $\|\boldsymbol{H}\| \leq \sqrt{m}$ holds, we have that $\sup_{|\psi\rangle}\langle \boldsymbol{H}\rangle_\psi \leq \max_i\langle \boldsymbol{H}\rangle_{\psi_i} + \epsilon/2$. Therefore, we have reduced the supremum over the state on a size-$G$ circuit to the maximum over the $(\epsilon/(2\sqrt{m}))$ net, where the union bound applies [Eq. (D1)],

$$\begin{aligned}
&\Pr\left[\sup_{|\psi\rangle\in\mathrm{Circ}(G)}\langle \boldsymbol{H}\rangle_\psi \geq \epsilon\right] \\
&\leq \Pr\left[\max_i\langle \boldsymbol{H}\rangle_{\psi_i} \geq \frac{\epsilon}{2}\right] \\
&\leq \#\{|\psi_i\rangle\}\cdot\exp\left(\frac{-\epsilon^2/8}{1/N + \epsilon/6\sqrt{m}}\right) \\
&\leq \#\{|\psi_i\rangle\}\cdot\exp\left(-\min(8\epsilon\sqrt{m}/8, \epsilon^2 N/8)\right).
\end{aligned}$$

Therefore, there exist constants $c_1$, $c_2$ such that

$$G \leq c_1\min(\epsilon\sqrt{m}, \epsilon^2 N)\cdot\log^{-1}(m) \quad \text{implies} \quad \Pr\left[\sup_{|\psi\rangle\in\mathrm{Circ}(G)}\langle \boldsymbol{H}\rangle_\psi \geq \epsilon\right] \leq \exp\left(-c_2\min(\epsilon\sqrt{m}, \epsilon^2 N)\right).$$

Plug in the assumption that $m \leq \epsilon^2 N^2$ to obtain the desired result. ■

We suspect the true circuit complexity to be $\Omega(m)$, but the current union bound argument can only give $\tilde{\Omega}(\sqrt{m})$. The concentration inequality needs to handle the event when the same Pauli string occurs $\Omega(\sqrt{m})$ times.

Still, we obtain a growing circuit-size lower bound $\Omega(\sqrt{m})$ by an elementary argument. In retrospect, it crucially depends on the noncommutativity of Pauli strings: The variance is suppressed by dimension. In contrast, the argument only gives $\Omega(n)$ circuit-size lower bounds (which is useless) for random, complete, $k$-local Hamiltonians for fixed $k$. Concretely, let $P_k$ be the set of Pauli strings of weight $k$ and consider the ensemble $\boldsymbol{H} = \sum_{\boldsymbol{\sigma}\in P_k} r_{\boldsymbol{\sigma}}\boldsymbol{\sigma}$, where $r_{\boldsymbol{\sigma}}$ are uniform random signs. Then, as in the proof of Lemma D.1, define $a_{\boldsymbol{\sigma}} = r_{\boldsymbol{\sigma}}\langle \boldsymbol{H}_{\boldsymbol{\sigma}}\rangle_\psi$, and compute [viewing $k = \mathcal{O}(1)$]

$$v = \sum_{\boldsymbol{\sigma}\in P_k}\mathbb{E}[a_{\boldsymbol{\sigma}}^2] = \Theta(|P_k|) = \Theta(n^k) \quad \text{and} \quad a_{\boldsymbol{\sigma}} \leq L = 1.$$

Here, the variance is much larger, and the optimum is roughly $\|\boldsymbol{H}\| = \Theta(\sqrt{vn})$. Plugging this into the union bound yields

$$\Pr\left[\sup_i\langle \boldsymbol{H}\rangle_{\psi_i} \geq \frac{\epsilon}{2}\sqrt{vn}\right] \leq \#\{|\psi_i\rangle\}\cdot\exp\left(\Theta\left(\frac{-\epsilon^2 vn/8}{v + \epsilon\sqrt{vn}/6}\right)\right) \sim \#\{|\psi_i\rangle\}\mathrm{e}^{-\Theta(\epsilon^2 n)}.$$

The union bound only supports size-$\mathcal{O}(n)$ circuits, roughly the circuit size of product states.

## APPENDIX E: MISSING PROOFS

In this appendix, we collect missing proofs.

### 1. Proof of Fact IV.1

*Proof.* The first and third moments vanish for both sets of matrices $\boldsymbol{A}_i$ and $\tilde{\boldsymbol{A}}_i$. We calculate the second moment

$$\begin{aligned}\mathbb{E}[\boldsymbol{A}_i \otimes \boldsymbol{A}_i] &= \frac{1}{2m}\mathbb{E}(\boldsymbol{DP} + \boldsymbol{P}^\dagger \boldsymbol{D}^\dagger) \otimes (\boldsymbol{DP} + \boldsymbol{P}^\dagger \boldsymbol{D}^\dagger).\\ &= \frac{1}{2m}\mathbb{E}\boldsymbol{DP} \otimes \boldsymbol{P}^\dagger \boldsymbol{D}^\dagger + \boldsymbol{P}^\dagger \boldsymbol{D}^\dagger \otimes \boldsymbol{DP}\\ &= \mathbb{E}[\boldsymbol{H}_{\mathrm{GUE}} \otimes \boldsymbol{H}_{\mathrm{GUE}}].\end{aligned}$$

The first equality uses $\mathbb{E}[\boldsymbol{D} \otimes \boldsymbol{D}^\dagger] = 0$. The second inequality is

$$\begin{aligned}\mathbb{E}[\boldsymbol{DP} \otimes \boldsymbol{P}^\dagger \boldsymbol{D}^\dagger] &= \sum_i \mathbb{E}|i\rangle\langle i|\boldsymbol{P} \otimes \boldsymbol{P}^\dagger|i\rangle\langle i|\\ &= \frac{1}{N}\sum_i \sum_j |i\rangle\langle j| \otimes |j\rangle\langle i|,\end{aligned}$$

which is proportional to the GUE value. This is the desired result. ■

## APPENDIX F: DIFFICULTY CANCELING HIGHER MOMENTS VIA INTERPOLATION

In this appendix, we give a heuristic reason why it seems difficult for interpolation-based methods to utilize higher-moment matching. Again, consider a set of independent matrices $\boldsymbol{A}_i$ whose low moments match those of some idealized matrices $\tilde{\boldsymbol{A}}_i$,

$$\mathbb{E}\boldsymbol{A}_i = 0 \quad \text{and} \quad \mathbb{E}\boldsymbol{A}_i^{\otimes k} = \mathbb{E}\tilde{\boldsymbol{A}}_i^{\otimes k} \quad \text{for each } k = 1, \ldots, t \quad \text{and} \quad i = 1, \ldots, m.$$

Suppose we construct an interpolation path

$$\boldsymbol{S}(t) \coloneqq f(t)\sum_i \boldsymbol{A}_i + g(t)\sum_i \tilde{\boldsymbol{A}}_i. \tag{F1}$$

Then, consider the expected $p$th moment and expand in powers of $\boldsymbol{A}_1$,

$$\begin{aligned}\mathbb{E}\boldsymbol{S}(t)^p = {}& \mathbb{E}(f(t)\boldsymbol{A}_1 + g(t)\tilde{\boldsymbol{A}}_1)\cdots(f(t)\boldsymbol{A}_1 + g(t)\tilde{\boldsymbol{A}}_1)\cdots\\ &+ \mathbb{E}(f(t)\boldsymbol{A}_1 + g(t)\tilde{\boldsymbol{A}}_1)\cdots(f(t)\boldsymbol{A}_1 + g(t)\tilde{\boldsymbol{A}}_1)\cdots\\ &\times (f(t)\boldsymbol{A}_1 + g(t)\tilde{\boldsymbol{A}}_1) + \cdots.\end{aligned}$$

Suppose the second and third moments do not vanish. If we want the time derivative to vanish, then we generally need

$$\begin{aligned}f(t)^2\mathbb{E}\boldsymbol{A}_1 \otimes \boldsymbol{A}_1 + g(t)^2\mathbb{E}\tilde{\boldsymbol{A}}_1 \otimes \tilde{\boldsymbol{A}}_1 = \mathrm{const} &\Longrightarrow f(t)^2 + g(t)^2 = 1,\\ f(t)^3\mathbb{E}\boldsymbol{A}_1 \otimes \boldsymbol{A}_1 \otimes \boldsymbol{A}_1 + g(t)^3\mathbb{E}\tilde{\boldsymbol{A}}_1 \otimes \tilde{\boldsymbol{A}}_1 \otimes \tilde{\boldsymbol{A}}_1 = \mathrm{const} &\Longrightarrow f(t)^3 + g(t)^3 = 1.\end{aligned}$$

There are only discrete solutions to both algebraic equations, and no continuous path can be established between $f = 0$ and $f = 1$. Indeed, the standard interpolant $f(t) = \sqrt{1-t}, g(t) = \sqrt{t}$ only cancels out the second moments. Therefore, if we want interpolation methods to capture higher-moments matching conditions, we need to go beyond the form of Eq. (F1). On the contrary, the Lindeberg principle appears more natural for this task.


[1] P. Shor, *Algorithms for quantum computation: Discrete logarithms and factoring*, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (1994), pp. 124–134.

[2] R. P. Feynman, *Simulating physics with computers*, Int. J. Theor. Phys. **21**, 467 (1982).

[3] S. Lloyd, *Universal quantum simulators*, Science **273**, 1073 (1996).

[4] G. H. Low and I. L. Chuang, *Optimal Hamiltonian simulation by quantum signal processing*, Phys. Rev. Lett. **118**, 010501 (2017).

[5] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, 2002), ISBN: 0821832298.

[6] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe, *The power of quantum systems on a line*, Commun. Math. Phys. **287**, 41 (2009).

[7] D. Gottesman and S. Irani, *The quantum and classical complexity of translationally invariant tiling and Hamiltonian problems*, in *2009 50th Annual IEEE Symposium on Foundations of Computer Science, Atlanta, GA* (IEEE, 2009), pp. 95–104.

[8] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, *Quantum computational chemistry*, Rev. Mod. Phys. **92**, 015003 (2020).

[9] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, *Low-depth quantum simulation of materials*, Phys. Rev. X **8**, 011044 (2018).

[10] C. Chamberland, K. Noh, P. Arrangoiz-Arriola, E. T. Campbell, C. T. Hann, J. Iverson, H. Putterman, T. C. Bohdanowicz, S. T. Flammia, A. Keller, G. Refael, J. Preskill, L. Jiang, A. H. Safavi-Naeini, O. Painter, and F. G. S. L. Brandão, *Building a fault-tolerant quantum computer using concatenated CAT codes*, 2020.

[11] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, *Even more efficient quantum computations of chemistry through tensor hypercontraction*, PRX Quantum **2**, 030305 (2021).

[12] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, *Quantum computing enhanced computational catalysis*, Phys. Rev. Res. **3** (2021) 033055.

[13] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, M. Kastoryano, R. Babbush, J. Preskill, D. R. Reichman, E. T. Campbell, E. F. Valeev, L. Lin, and G. K.-L. Chan, *Is there evidence for exponential quantum advantage in quantum chemistry?*, 2022.

[14] N.-H. Chia, A. Gilyén, T. Li, H.-H. Lin, E. Tang, and C. Wang, *Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning*, *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, 2019.

[15] A. Gilyén, S. Lloyd, and E. Tang, *Quantum-inspired low-rank stochastic regression with logarithmic dependence on the dimension*, arXiv:abs/1811.04909.

[16] E. Tang, *Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions*, Phys. Rev. Lett. **127**, 060503 (2021).

[17] E. Tang, *A quantum-inspired classical algorithm for recommendation systems*, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, Phoenix, AZ* (ACM, New York, 2019), pp. 217–228, 10.1145/3313276.3316310.

[18] S. Gharibian and F. L. Gall, *Dequantizing the quantum singular value transformation: Hardness and applications to quantum chemistry and the quantum PCP conjecture*, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, 2022.

[19] A candidate answer to our question would be to encode an arbitrary quantum computation (e.g., Shor's algorithm) into the ground state of a local Hamiltonian. Classical hardness would follow from the complexity-theoretic assumption BQP $\neq$ BPP (or, in the specific case of Shor's, that factoring is hard for classical computers). Indeed, this was the idea behind the proof that adiabatic evolution is universal for quantum computation [20]. However, we exclude this example from the main exposition because the resulting Hamiltonians are contrived and offer no *independent* insights about quantum advantage: This argument does not teach us *how* quantum computation is more powerful than classical computation for finding low-energy states of natural Hamiltonians. This conceptual question is precisely what our work aims to address.

[20] D. Aharonov, W. Van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, *Adiabatic quantum computation is equivalent to standard quantum computation*, SIAM Rev. **50**, 755 (2008).

[21] L. Erdős and D. Schröder, *Phase transition in the density of states of quantum spin glasses*, Math. Phys. Appl. Math. **17**, 441 (2014).

[22] D. Sherrington and S. Kirkpatrick, *Solvable model of a spin-glass*, Phys. Rev. Lett. **35**, 1792 (1975).

[23] C. L. Baldwin and B. Swingle, *Quenched vs annealed: Glassiness from SK to SYK*, Phys. Rev. X **10**, 031026 (2020).

[24] Heuristically, if typical realizations of few-body spin glass models capture the complexity of NP-hard problems, we do not expect quantum computers to solve them efficiently.

[25] A. W. Harrow and A. Montanaro, *Extremal eigenvalues of local Hamiltonians*, Quantum **1**, 6 (2017).

[26] A. Kitaevj, *A simple model of quantum holography talk 1 and talk 2 at KITP*, 2015.

[27] S. Sachdev and J. Ye, *Gapless spin-fluid ground state in a random quantum Heisenberg magnet*, Phys. Rev. Lett. **70**, 3339 (1993).

[28] G. Gur-Ari, R. Mahajan, and A. Vaezi, *Does the SYK model have a spin glass phase?*, J. High Energy Phys. 11 (2018) 070.

[29] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, Phys. Rev. D **94** (2016) 106002.

[30] Chaos in the sense of fast thermalization means efficient preparation of Gibbs states via coupling to a bath. See Refs. [31,32] for a quantitative statement connecting the eigenstate thermalization hypothesis and thermalization.

[31] C.-F. Chen and F. G. Brandão, *Fast thermalization from the eigenstate thermalization hypothesis*, arXiv:2112.07646.

[32] O. Shtanko and R. Movassagh, *Algorithms for Gibbs state preparation on noiseless and noisy random quantum circuits*, 2021.

[33] M. B. Hastings and R. O'Donnell, *Optimizing strongly interacting fermionic Hamiltonians*, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy* (Association for Computing Machinery, New York, NY, 2022), pp. 776–789, 10.1145/3519935.3519960.

[34] Our circuit lower bound also fails (see Appendix D).

[35] Y. Herasymenko, M. Stroeks, J. Helsen, and B. Terhal, *Optimizing sparse fermionic Hamiltonians*, Quantum **7**, 1081 (2023).

[36] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. Math. **67**, 325 (1958).

[37] G. E. Mitchell, A. Richter, and H. A. Weidenmüller, *Random matrices and chaos in nuclear physics: Nuclear reactions*, Rev. Mod. Phys. **82**, 2845 (2010).

[38] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, England, 2009).

[39] B. Collins and I. Nechita, *Random matrix techniques in quantum information theory*, J. Math. Phys. (N.Y.) **57**, 015215 (2016).

[40] E. Campbell, *Random compiler for fast Hamiltonian simulation*, Phys. Rev. Lett. **123**, 070503 (2019).

[41] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, *Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics*, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019.

[42] This statement is analogous to the no-low energy-trivial-state conjecture (NLTS) [43]. As a disclaimer, we are far from the original context of NLTS, where the Hamiltonians are $\mathcal{O}(1)$ local, frustration-free, and topologically ordered. The proof we provide fails for constant $\mathcal{O}(1)$-local random

Hamiltonians. Indeed, as we mentioned, the two-body random Pauli model has an efficient product state constant-ratio approximation of the operator norm [25]. Recently, the NLTS conjecture for circuit depth $\log(n)$ was proven for certain $\mathcal{O}(1)$-local Hamiltonians arising from quantum error-correcting codes [44].

[43] M. H. Freedman and M. B. Hastings, *Quantum systems on non-k-hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs*, 2013.

[44] A. Anshu, N. P. Breuckmann, and C. Nirkhe, *NLTS Hamiltonians from good quantum codes*, in Proceedings of the 55th Annual ACM Symposium on Theory of Computing, Orlando, FL (Association for Computing Machinery, New York, NY, 2022), pp. 1090–1096, 10.1145/3564246.3585114.

[45] Asymptotically, a Gaussian distribution was known for lattice Hamiltonians [46] and random $k$-local Pauli Hamiltonians (1) for $k = o(\sqrt{n})$ [21]. Nonasymptotically, a spectral tail appears in the 4-local SYK model [33].

[46] F. G. S. L. Brandão and M. Cramer, *Equivalence of statistical mechanical ensembles for non-critical quantum systems*, arXiv:1502.03263.

[47] This is closely related to the phase estimation amplitude profile where the width $\eta$ is roughly the resolution (inversely proportional to the runtime).

[48] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics, 2nd ed. (Springer, New York, 2010).

[49] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, Mathematical Surveys and Monographs, Vol. 171 (American Mathematical Society, Providence, RI, 2011).

[50] S. Chatterjee, *A simple invariance theorem*, 2005.

[51] S. B. Korada and A. Montanari, *Applications of the Lindeberg principle in communications and statistical learning*, IEEE Trans. Inf. Theory **57**, 2440 (2011).

[52] J. A. Tropp, *An introduction to matrix concentration inequalities*, 2015.

[53] J. A. Tropp, *Second-order matrix concentration inequalities*, Appl. Comput. Harmon. Anal. **44**, 700 (2018).

[54] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel, Matrix concentration inequalities and free probability, 2021.

[55] T. Brailovskaya and R. van Handel, *Universality and sharp matrix concentration inequalities*, 2022.

[56] A. Lytova and L. Pastur, *Central limit theorem for linear eigenvalue statistics of random matrices with independent entries*, Ann. Probab. **37**, 1778 (2009).

[57] The random signed permutation is defined by $\boldsymbol{Q}' \coloneqq \boldsymbol{D}'\boldsymbol{P}$, where $\boldsymbol{D}'_{ab} = \delta_{ab} r_a$ and $r_a \overset{\text{i.i.d.}}{\sim} \{1, -1\}$.

[58] 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).

[59] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, K. Zuo, S. Tamate, Y. Tabuchi, and Y. Nakamura, *Quantum circuits for exact unitary-designs and applications to higher-order randomized benchmarking*, PRX Quantum **2** (2021) 030339.

[60] N. Alon and S. Lovett, *Almost k-wise vs. k-wise independent permutations, and uniformity for general group actions*, Theory Comput. **9**, 559 (2013).

[61] N.-H. Chia, H.-H. Lin, and C. Wang, *Quantum-inspired sublinear classical algorithms for solving low-rank linear systems*, arXiv:abs/1811.04852.

[62] More carefully, our current results require the number of terms $m$ to grow with the desired accuracy.

[63] Taking $O(\delta^{-1}\epsilon^{-2})$ samples, one can estimate the number of states to a multiplicative error $\epsilon$.

[64] S. Oymak and J. A. Tropp, *Universality laws for randomized dimension reduction, with applications*, Inf. Inference **7**, 337 (2018).

[65] J. A. Tropp, *The expected norm of a sum of independent random matrices: An elementary approach*, in *High Dimensional Probability VII: The Cargèse Volume* (Springer, Birkhäuser, Cham, 2016), pp. 173–202.

[66] R. Bhatia, *Matrix Analysis* (Springer, New York, 1997).

[67] The odd $p$-norms contain an absolute value $|||\boldsymbol{S}|||_p^p = \mathbb{E}\overline{\mathrm{Tr}}|\boldsymbol{S}|^p$, so they are less suitable for Taylor expansion.

[68] D. Williams, *Probability with Martingales* (Cambridge University Press, Cambridge, England, 1991).

[69] P. Hitczenko, *On the behavior of the constant in a decoupling inequality for martingales*, Proc. Am. Math. Soc. **121**, 253 (1994).

[70] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics, Vol. 118 (Cambridge University Press, Cambridge, England, 2010).

[71] G. Pisier, *Introduction to Operator Space Theory*, London Mathematical Society Lecture Note Series (Cambridge University Press, Cambridge, England, 2003).

[72] L. Erdős and H.-T. Yau, *A dynamical approach to random matrix theory*, 2017.

[73] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford University Press, New York, 2013).

[74] G. H. Low and I. L. Chuang, *Hamiltonian simulation by qubitization*, Quantum **3**, 163 (2019).

[75] Strictly speaking, the process we illustrate is a quantum channel involving both quantum gates and classical randomness (i.e., repeating until success is observed). A fixed, deterministic circuit could be constructed by performing phase estimation on half of an input maximally entangled state (for which the reduced density matrix is maximally mixed) and performing fixed-point amplitude amplification [76] to coherently boost the probability of success.

[76] T. J. Yoder, G. H. Low, and I. L. Chuang, *Fixed-point quantum search with an optimal number of queries*, Phys. Rev. Lett. **113**, 210501 (2014).