Optimizing sparse fermionic Hamiltonians

Yaroslav Herasymenko[1,2], Maarten Stroeks[2,3], Jonas Helsen[1], and Barbara Terhal[2,3]

[1]QuSoft & CWI, Science Park 123 1098 XG Amsterdam, The Netherlands
[2]QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
[3]EEMCS Faculty, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Delft, The Netherlands

We consider the problem of approximating the ground state energy of a fermionic Hamiltonian using a Gaussian state. In sharp contrast to the dense case [1, 2], we prove that strictly q-local *sparse* fermionic Hamiltonians have a constant Gaussian approximation ratio; the result holds for any connectivity and interaction strengths. Sparsity means that each fermion participates in a bounded number of interactions, and strictly q-local means that each term involves exactly q fermionic (Majorana) operators. We extend our proof to give a constant Gaussian approximation ratio for sparse fermionic Hamiltonians with both quartic and quadratic terms. With additional work, we also prove a constant Gaussian approximation ratio for the so-called sparse SYK model with strictly 4-local interactions (sparse SYK-4 model). In each setting we show that the Gaussian state can be efficiently determined. Finally, we prove that the $O(n^{-1/2})$ Gaussian approximation ratio for the normal (dense) SYK-4 model extends to SYK-q for even $q > 4$, with an approximation ratio of $O(n^{1/2-q/4})$. Our results identify non-sparseness as the prime reason that the SYK-4 model can fail to have a constant approximation ratio [1, 2].

Contents

1 Introduction

Approximating the ground state energy of a local Hamiltonian is a central problem in both physics and computer science. In computer science it plays a key role in complexity theory [3], while in physics ground states capture the behaviour of systems at low energy. Two common families of Hamiltonians of interest are those defined on collections of qubits and those acting on fermionic degrees of freedom. Fermionic Hamiltonians model various physical systems, such as electrons in condensed matter and quantum chemistry — prime targets for quantum simulation. Fermions also define a model of quantum computation, equivalent to the one based on qubits [4]. Despite its practical and conceptual relevance, the general problem of approximating fermionic ground state energies is currently less well understood than its qubit counterpart.

Some rigorous progress in studying this problem – both for qubits and for fermions – was made from the perspective of optimization. In this subfield of computer science, one of the central tasks is efficiently finding problem solutions that are provably close to optimal [5]. The closeness is usually quantified by an approximation ratio, i.e. the ratio between the value attained by an algorithm and the optimal value for a given problem. For the classical equivalent of the ground state energy finding – Constraint Satisfaction Problems (CSPs) – such approximation ratios have been extensively studied [6].

For quantum Hamiltonians, an interesting question is how well the ground state energy can be approximated using "classical" or "mean-field" states. For qubit Hamiltonians the natural choice of classical states are product states, while for fermionic Hamiltonians they are Gaussian states. Gaussian states play a prominent role in fermionic optimization problems using the mean-field Hartree-Fock method, see e.g. [7], or dynamical mean-field theory via solving impurity problems [8], or the simulation of free fermionic computation [9, 10].

Formal guarantees on approximation ratios characterize numerical simulation methods using classical states and outline their limitations compared to quantum computing. For qubit Hamiltonians, it was first proved by Lieb [11] (see [12] for a simplified proof) that there always exists a product state which approximates the ground state energy of a traceless 2-local qubit Hamiltonian by a factor of 1/9. Many more results on approximating ground state energies of many-body systems by product states can be found in [13, 14, 15, 16, 17, 18]. In [12] it was shown, through the Goemans-Williamson method, that for a 2-local traceless qubit Hamiltonian a product state can always be *efficiently* found with approximation ratio $O(1/\log(n))$ where n is the number of qubits. Ref. [12] also considered fermionic Hamiltonians with quadratic ($q = 2$) and quartic ($q = 4$) fermionic terms. They left as an open question whether all 4-local fermionic Hamiltonians have a constant approximation ratio with respect to Gaussian states (a Gaussian approximation ratio).

A surprising counterexample to this conjecture was recently presented in Refs. [1, 2] — the family of SYK-4 models (Sachdev-Ye-Kitaev models with quartic fermionic interactions, see Definition 2). It was shown that with high probability, SYK-4 Hamiltonians admits a Gaussian approximation ratio no better than $O(1/\sqrt{n})$ where n is the number of fermionic modes. Contrasting this result to Refs. [11, 14], it means that qubit and fermionic ground states strongly differ in their approximability by classical states. Moreover, this opens up the question of which fermionic Hamiltonians *do* allow finite Gaussian approximation ratios.

This is the question that we aim to answer here. We do this by considering *sparse* Hamiltonians, i.e.

Hamiltonians where each fermionic mode participates in a bounded number of interactions. Sparsity holds for many physically relevant Hamiltonians, such as the Fermi-Hubbard model. It also holds for exotic Hamiltonians, such as those determined by constant-degree expander hypergraphs; notably, it does not hold for the SYK model. Sparsity of interactions has been considered in the classical CSP literature. It was shown in [19] that the MaxQP problem has an efficient constant approximation ratio algorithm on graphs of bounded chromatic number, in particular graphs with bounded degree. We show that a similar assumption of sparsity is enough to guarantee constant Gaussian approximation ratios for 4-local and strictly q-local Hamiltonians. Moreover, we show that a constant Gaussian approximation ratio can be achieved for the *sparse* SYK-4 model [20] (which has a logarithmically growing interaction participation and is thus not sparse by our definition). Finally, we consider in more detail the optimal approximation ratio for the *dense* SYK-q model for $q > 4$ (thus extending the work of [2]). We show that the shortfall of Gaussian states is even more pronounced in this setting.

To avoid confusion, we note that instead of the *ground state* energy, existing works often consider approximating the *maximal* eigenvalue of the Hamiltonian $\lambda_{\max}(H)$. These two optimization problems are equivalent if the family of Hamiltonians considered is invariant under a change of sign (e.g. traceless q-local Hamiltonians). For mathematical convenience and consistency with the literature, in the rest of the text, we will also be formulating our results in terms of approximating $\lambda_{\max}(H)$.

2 Statement of results

2.1 Preliminaries

Before surveying our results, we introduce the basic setup of fermionic Hamiltonians and q-locality. This subsection also defines the SYK-q model and spells out the previous result of a vanishing Gaussian approximation ratio for SYK-4.

We consider a system of $2n$ traceless Majorana fermion operators c_i, $i = 1, \ldots, 2n$ with $c_i^2 = \mathbb{I}, c_i^\dagger = c_i$, forming a Clifford algebra, i.e., $\{c_j, c_k\} = 2\delta_{j,k}\mathbb{I}$ and representing n fermionic modes. We denote as I an ordered subset $I = \{i_1, i_2, ..i_q\} \subseteq [2n] \equiv \{1, \ldots, 2n\}$ where $i_1 < i_2 < \ldots i_q$ with q even. We denote C_I as the Hermitian Majorana monomial

$$C_I \equiv i^{q/2} c_{i_1}..c_{i_q}, \tag{1}$$

and one can verify that

$$C_I^2 = \mathbb{I}.$$

We can think about a subset I as corresponding to a term or interaction in a Hamiltonian. Indeed, it is natural to impose some form of locality:
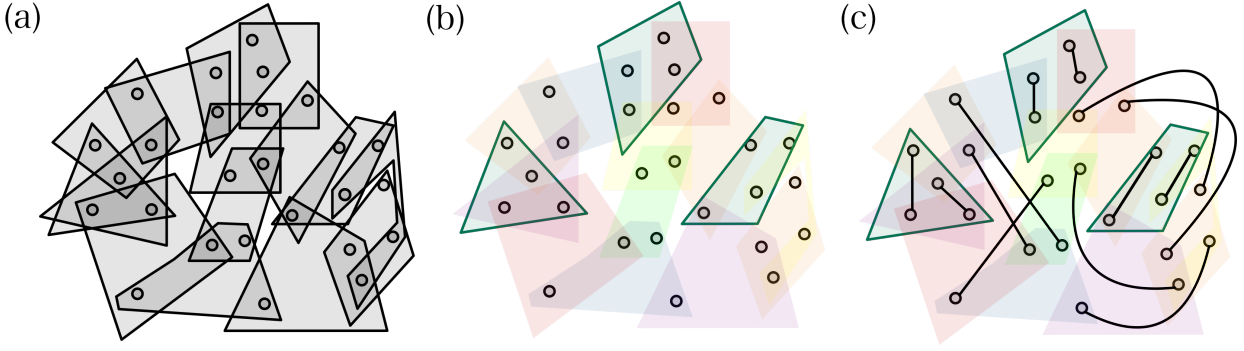
Figure 1: Illustrating the key idea of the proof of Theorem 5. An example of a strictly 4-local Hamiltonian is given in (a), vertices and faces representing Majorana operators and their interactions. The Hamiltonian is split into sets of terms – different colors in (b) – well separated from each other inside each set (so-called diffuse sets, see Definition 15). The next step is to match all Majorana operators, i.e., split the vertices into disjoint pairs, each connected by an edge (see panel (c)). We separately match the support of each term in one *targeted* set of terms (the color highlighted in (b) and (c)). The remaining vertices are matched in such a way that no two vertices connected by an edge belong to the same term. The Gaussian state is then created from the resulting matching, with only terms from the *targeted* set contributing to the energy. By optimizing the choice of the targeted set, a finite approximation ratio can be guaranteed.

Definition 1 (q-local fermionic Hamiltonian). *Let H be a fermionic Hamiltonian on $2n$ Majorana operators. We say that H is q-local if H is a sum of Hermitian traceless terms C_I of weight at most q, i.e. each term is proportional to a product of at most q operators c_i. H is said to be strictly q-local when all terms have exactly weight q.*

A local traceless fermionic Hamiltonian $H = \sum_{I \in \mathcal{I}} J_I C_I$ is thus characterized by an interaction set $\mathcal{I}$ and the coefficients $J_I \in \mathbb{R}$. The maximum eigenvalue of H is denoted by $\lambda_{\max}(H) := \max_\rho \mathrm{Tr}(H\rho)$. Sometimes we will refer to a collection of sets I denoted as $\mathcal{I} = \{I_1, I_2, \ldots\}$. The support of $\mathcal{I}$ is defined as $\mathrm{Sup}(\mathcal{I}) = \cup_i I_i$ and $\mathcal{I}' \subseteq \mathcal{I}$ implies that the sets in $\mathcal{I}'$ are also sets in $\mathcal{I}$.

Definition 2 (SYK-q Model). *A q-local (with q even) SYK model on $2n$ Majoranas is defined as a family of Hamiltonians*

$$H = \binom{2n}{q}^{-1/2} \sum_{I \subseteq [2n], |I|=q} J_I \, C_I, \qquad (2)$$

where each J_I is a Gaussian random variable (i.e., with zero mean and unit variance) and each C_I is the product of the q distinct Majorana operators as in Eq. (1). We normalize the model in expectation, i.e., $\mathbb{E}\left(\mathrm{Tr}(H^2)\right) = \binom{2n}{q}^{-1} \sum_{I \subseteq [2n], |I|=q} \mathbb{E}(J_I^2) \, \mathrm{Tr}(\mathbb{I}) = 2^n$.

In [1] it was shown that with high probability (over the draw of J_Is) for the SYK-4 model, one has

$$\max_{\rho \text{ Gaussian}} \mathrm{Tr}(H\rho) = O(1).$$

In order to thus provide a counterexample to a constant Gaussian approximation ratio, one needs to prove a lower bound on $\lambda_{\max}(H)$ for the SYK-4 model, which holds with high probability, which was done in [2]:

Theorem 3. *[2] There is a $\mathrm{poly}(n)$-time quantum algorithm that, given any SYK-4 Hamiltonian H, returns a quantum state ρ. With probability $1 - exp\left(-\Omega(n)\right)$ (over the draw of the J_Is), this state ρ has $\mathrm{Tr}(H\rho) = \Omega(\sqrt{n})$.*

2.2 Sparse fermionic Hamiltonians

Key to our work is the notion of a sparse Hamiltonian.

Definition 4. *Let H be a local traceless fermionic Hamiltonian of $2n$ Majorana operators. We say that H is k-sparse, for an integer k, if no Majorana operator c_i occurs in more than k terms of the Hamiltonian.*

Using graph theoretic terminology, one may say that interactions in a k-sparse Hamiltonian form a hypergraph of bounded degree k. This condition allows us to efficiently find Gaussian states with constant approximation ratio. We have the following theorem, which is the main result of our work:

Theorem 5. *Let H be a traceless fermionic Hamiltonian on $2n$ Majorana operators with maximal eigenvalue $\lambda_{\max}(H)$. If H is k-sparse and strictly q-local and $n > (q^2-1)k$, a Gaussian state ρ can be efficiently constructed such that*

$$\frac{\mathrm{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{Q}, \qquad (3)$$

for $Q = q(q-1)(k-1)^2 + q(k-1) + 2$.

The proof of this theorem is given in Section 5; its basic idea is explained in Figure 1.

We note that this proof only holds for Hamiltonians with terms of *exactly* weight q. Typical physical Hamiltonians, however, have quadratic (kinetic energy of the electrons) and quartic terms (potential energy due to Coulomb interaction). Fortunately, we can also show that in the $q = 4$ case we can include $q = 2$ terms. For this we use a trick from [12] to lift such a 4-local Hamiltonian to a strictly 4-local Hamiltonian. This trick makes the Hamiltonian non-sparse. However, we show in Section 6 that, in this special case, we can circumvent the non-sparseness of the Hamiltonian and achieve a constant Gaussian approximation ratio.

Theorem 6. *Let H be a traceless fermionic Hamiltonian on $2n$ Majorana operators with maximal eigenvalue $\lambda_{\max}(H)$. If H is k-sparse with terms of weight 2 and 4 and $2n > 15k$, a Gaussian state ρ can be efficiently constructed, such that*

$$\frac{\text{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{2Q} \qquad (4)$$

for $Q = 12(k-1)^2 + 4(k-1) + 2$.

2.3 The sparse $q = 4$ SYK model

In view of Theorem 5 it is worth revisiting the lack of a constant Gaussian approximation for the SYK model. The SYK-q model in Definition 2 is extremely non-sparse, in the sense that every Majorana operator occurs together with all other Majorana operators. This makes the SYK model somewhat unphysical, and several *sparse* versions of the model have been considered [20, 21]. Such sparse models intend to produce the same (low energy) physics, while being easier to simulate on both quantum and classical computers (see sections III and V in [20]). The sparse SYK model is generated by including terms by a Bernoulli trial with a certain probability p tuned such that the expected sparsity is bounded:

Definition 7. *The sparse SYK-4 or SSYK-4 model on $2n$ Majorana operators with expected sparsity $k = O(1)$ is given as*

$$H = \frac{1}{\sqrt{2kn}} \sum_{I \subset [2n], |I|=4} X_I J_I C_I \qquad (5)$$

where the X_I are i.i.d. Bernoulli random variables with $p = \Pr(X_I = 1) = \frac{k}{\binom{2n-1}{3}}$ and the J_I are i.i.d. Gaussian random variables with mean 0 and variance 1.

Unlike the full SYK model with $\binom{2n}{4}$ terms in H, the sparse SYK model has a number of terms $\sim n$ in expectation. Note that the SSYK-4 model is only k-sparse in expectation, and with high probability there is a Majorana operator with degree $\Omega\left(\frac{\log(n)}{\log\log(n)}\right)$ (the degree distribution follows that of an Erdős-Renyi hypergraph. See Theorem 3.4 in [22] for a proof of the

statement for Erdős-Renyi graphs. The hypergraph version follows by the same logic). This means that Theorem 5 does not directly apply. However, one can show, through a truncation argument, that almost all instantiations of SSYK-4 can be *sparsified*, giving rise to a constant approximation ratio result that holds with high probability.

Theorem 8. *Let H be a SSYK-4 Hamiltonian in Eq. (5) with expected degree $k = O(1)$, such that $n > 120(k + 1)$. With probability at least $1 - 4\exp\left[-\frac{e^{-16(k+1)}k^3}{64(8k+7)}n\right]$, a Gaussian state ρ can be efficiently constructed such that*

$$\frac{\text{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{Q}, \qquad (6)$$

where $Q = 1236 + 2752k + 1536k^2$.

Thus we arrive at the surprising conclusion that the SSYK-4 model has a constant Gaussian approximation ratio, while the dense SYK-4 model does not — even though SSYK-4 has similar physical properties as SYK-4.

2.4 Higher-q SYK models

We investigate what Gaussian approximation ratios can be achieved for the *dense* SYK model of even weight $q > 4$, as this was left as an open question in [2]. We establish an upper bound on the largest Gaussian expectation value of SYK-q, which behaves rather dramatically for $q > 4$. We prove the following Lemma employing a method similar to the one used in [1].

Lemma 9. *Let H be the dense SYK-q Hamiltonian (with even $q \geq 4$ and $q = O(1)$). With probability at least $1 - \exp(-\Omega(n))$ over the draw of SYK-q Hamiltonians, the expectation value of every Gaussian state ρ is bounded, more precisely:*

$$\max_{\rho \text{ Gaussian}} \text{Tr}(H\rho) = O(n^{1-q/4}). \qquad (7)$$

This Lemma is proved in Section 8. Our second result establishes a lower bound on the largest eigenvalue for SYK-q, essentially generalizing what was established in [2] for $q = 4$. We prove the following Lemma (its proof can be found in Section 8):

Lemma 10. *Let H be the dense SYK-q Hamiltonian with even $q \geq 4$ (and $q = O(1)$). With probability at least $1 - \exp(-\Omega(n))$ over the draw of SYK-q Hamiltonians, $\lambda_{\max}(H) = \Omega(\sqrt{n})$.*

As an immediate consequence of the previous results, we see that the Gaussian approximation ratio of the dense SYK-q model can be no better than $O(n^{1/2-q/4})$:

Theorem 11. *Let H be the* dense *SYK-q Hamiltonian (with even $q \geq 4$ and $q = O(1)$). With probability at least $1 - exp\big(-\Omega(n)\big)$ over the draw of SYK-q Hamiltonians, we have*

$$\max_{\rho \text{ Gaussian}} \frac{\text{Tr}(H\rho)}{\lambda_{\max}(H)} = O\big(n^{1/2-q/4}\big). \qquad (8)$$

Proof. Theorem 11 follows from combining Lemma 9 and Lemma 10 and applying the union bound. □

3   Discussion

The goal of this section is to place our results in a broader context and mention a few open questions.

First, let us discuss the relation between this work and the fermion-to-qubit mapping methods. As was shown in [4], one can map a *sparse* $O(1)$-local fermionic Hamiltonian onto a sparse $O(1)$-local qubit Hamiltonian (BK-superfast encoding). However, for this mapping one needs to enforce parity checks which are in general nonlocal; therefore, we cannot obtain our Theorem 5 in this way. There is also an additional obstacle: using the BK-superfast encoding, an approximating product state for the qubit Hamiltonian does not necessarily map back to a Gaussian fermionic state.

Ref. [4] also showed that one can map a general local fermionic Hamiltonian (like a SYK model) onto a qubit Hamiltonian with terms which are $O(\log n)$-local. Such qubit Hamiltonian is generally not expected to have a constant approximation ratio by a product state due to its n-dependent locality. In fact, one can easily prove that a dense model like the SYK model can only be mapped onto a qubit Hamiltonian which is $\Omega(\log n)$-local. We give the argument in Appendix A.

These observations suggest that approximation ratios by classical states such as Gaussian states or product states are likely to be affected by sparsity in the case of fermions, which is consistent with our new results.

Another question which is raised by our work and that of [2] and [15], is whether studying fermionic Hamiltonians can lead to new insights into the possibility of a quantum PCP theorem [23]. In this context it is important to mention that, besides the lower bound in Theorem 3, Ref. [2] also determined an upper bound on $\lambda_{\max}$ of the SYK-4 model showing that with high probability $\lambda_{\max} = \Theta(\sqrt{n})$. This shows that the SYK-4 model is extremely frustrated: the maximal average expected energy per term, *the energy density*, is only $\Theta(n^{-3/2})$. In contrast, our results for the sparse SYK model (see Lemma 23) show that the maximal average expected energy per term is $\Omega(1)$, which is the more 'natural' physical scaling. A simple fermionic toy model in which the maximal average energy per term decreases is a model in which an extensive set of Majorana operators is mutually anti-commuting, see Lemma 29 in Appendix A. The presence of many such fully-anticommuting sets in the SYK model can be seen as one of the intuitive reasons why the maximal energy density achieved is so low.

For k-local qubit Hamiltonians researchers have looked at the hardness of approximating the maximal energy density with constant error ϵ: showing that this problem is QMA-complete would prove the quantum PCP theorem. For dense (non-sparse) k-local qubit Hamiltonians, it was proved in [15] (Theorem 13) that there is a polynomial-time classical algorithm to approximate the maximal energy density, using product state approximations. Ref. [24] generalized this result and formulated an efficient classical algorithm which approximately estimates the free energy of a 2-local dense qubit Hamiltonian.

One can similarly ask the question of approximating the maximal energy density for dense q-local fermionic Hamiltonians. Observe that the question is moot if the maximal energy density decreases as a function of n (as in the SYK model), since for large enough n (depending on ϵ) the classical algorithm could always output 0 and make an error less than ϵ. However, other dense $O(1)$-local fermionic Hamiltonians could exist for which this question is nontrivial and not already covered by the dense qubit case.

There are further open directions that are more practically-oriented. One of these is achieving finite approximation ratios for at least some classes of *non-sparse* fermionic Hamiltonians (e.g., quantum chemistry or lattice systems with long-range Coulomb interactions). Furthermore, in most applications, one is interested in obtaining approximation ratios as close to 1 as possible. Although for most systems of interest one cannot expect ratios that are ϵ-close to 1, our theoretical lower bounds could still be vastly improved. For instance, for sparse SYK with $k = 10$, the guaranteed ratio is only $\simeq 5 \times 10^{-6}$ (cf. Theorem 8). This can be contrasted to the Hartree-Fock applications to quantum chemistry systems, which usually achieve approximation ratios of > 0.9. Improving our results to derive more realistic lower bounds could be of great value; some possible approaches are as follows.

One option is to extend the interaction subsets targeted by the constructed Gaussian state beyond the diffuse subsets considered here. If the overlapping interactions in the problem Hamiltonian are not prone to frustration, including them in the targeted set may dramatically increase the approximation ratio. The proof of Theorem 6 (Section 6) is a special case of this approach, with the constructed Gaussian state targeting multiple overlapping terms at the same time.

Another option for improvement is to minimize the contribution from frustration terms instead of avoiding frustration altogether. This could both improve the eventual approximation ratio by targeting a larger pool of interactions, as well as allowing to mitigate the

5

issue of non-sparsity. An example of this approach is the proof of Theorem 8 (Section 7), where the contributions from the non-sparse part of the Hamiltonian are shown to be small compared to the energy achieved by the Gaussian state.

As a third option, one can modify the basis of fermionic modes so that non-sparsity and frustration in the Hamiltonian are minimized. In the simplest case of $q = 2$, such a basis rotation can always turn all interactions into a diffuse set (simply by diagonalizing the Hamiltonian). A similar improvement may be possible for some classes of q-local Hamiltonians with $q \geq 4$.

Developing these and other directions for efficient Gaussian ground state approximation are interesting possibilities for future research.

Finally, it would be interesting to provide a non-random family of fermionic Hamiltonians without a constant approximation ratio with respect to Gaussian states.

4 Background on Gaussian states

In this section, we first provide some background and definitions that will be used throughout the remainder of this text.

4.1 Gaussian states

We define the class of fermionic Gaussian states, which are ground states and thermal states of non-interacting, quadratic ($q = 2$), fermionic Hamiltonians, and give some of their useful properties.

We first note that any transformation by a real orthogonal matrix $R \in SO(2n)$, i.e.,

$$\tilde{c}_i = \sum_j R_{ij} c_j, \qquad (9)$$

preserves the properties of Majorana operators and hence gives rise to a new set of $2n$ Majorana operators $\{\tilde{c}_j\}_{j=1}^{2n}$.

Definition 12. *Fermionic Gaussian state.* *Given $2n$ Majorana operators denoted by $c_1, \ldots, c_{2n}$. A fermionic Gaussian state is a – generally mixed – state of the form*

$$\rho = \frac{1}{2^n} \exp\Big(-i \sum_{i \neq j} \beta_{ij} c_i c_j\Big), \qquad (10)$$

where $(\beta_{ij})_{i,j=1}^{2n}$ is a real anti-symmetric matrix and the normalization is such that $\operatorname{Tr} \rho = 1$.

Fermionic Gaussian states have a number of useful properties, which we list here for future use.

1. The matrix β can be block-diagonalized by a real orthogonal matrix $R \in SO(2n)$ such that

$$\beta = R \bigoplus_{j=1}^n \begin{pmatrix} 0 & b_j \\ -b_j & 0 \end{pmatrix} R^T, \qquad (11)$$

with $b_j \geq 0$. Therefore, ρ can be brought to the following standard form

$$\rho = \frac{1}{2^n} \prod_{j=1}^n \Big(\mathbb{I} + i\lambda_j \tilde{c}_{2j-1} \tilde{c}_{2j}\Big), \qquad (12)$$

where $\tilde{c}_i = \sum_j R_{ij} c_j$ and $\lambda_j = \tanh(2b_j) \in [-1, +1]$.

2. Each fermionic Gaussian state can be associated with a $2n \times 2n$ correlation matrix Γ, with

$$\Gamma_{ij} = \frac{i}{2} \operatorname{Tr}\big(\rho[c_i, c_j]\big). \qquad (13)$$

Γ is a real anti-symmetric matrix and hence there is a real orthogonal matrix $R \in SO(2n)$ such that

$$\Gamma = R \bigoplus_{j=1}^n \begin{pmatrix} 0 & \lambda_j \\ -\lambda_j & 0 \end{pmatrix} R^T, \qquad (14)$$

where the λ_j are in Eq. (12).

3. For pure fermionic Gaussian states, $\lambda_j \in \{-1, +1\}$ and hence for pure Gaussian states $\Gamma^T \Gamma = \mathbb{I}$. For mixed fermionic Gaussian states, $\Gamma^T \Gamma \leq \mathbb{I}$.

4. The Pfaffian of a $2k \times 2k$ anti-symmetric matrix A is defined as

$$\operatorname{Pf}(A) = \frac{1}{2^k k!} \sum_{\pi \in S_{2k}} \operatorname{sign}(\pi) \Pi_{i=1}^k A_{\pi(2i-1), \pi(2i)}.$$

Alternatively, we can see the Pfaffian as a sum over perfect matchings in a graph of $2k$ vertices where an edge $(i < j)$ has weight A_{ij} and each matching contributes the products of these weights to the sum. For a Gaussian state with correlation matrix Γ, one has for even $|I|$:

$$\operatorname{Tr}(C_I \rho) = \operatorname{Pf}(\Gamma_I), \qquad (15)$$

where Γ_I is the $|I| \times |I|$ submatrix of Γ restricted to rows and columns in the ordered set I.

A special class of a pure Gaussian states is given by a perfect matching M of Majorana operators. Such matching M is specified by n disjoint pairs (m_1, m_2) with $m_1 < m_2$. For each pair we have a coefficient $\lambda_{(m_1, m_2)} = \pm 1$, together forming the n-dimensional vector $\vec{\lambda}$. The class of states are of the form

$$\rho(M, \vec{\lambda}) = \frac{1}{2^n} \Pi_{(m_1, m_2) \in M} (\mathbb{I} + i\lambda_{(m_1, m_2)} c_{m_1} c_{m_2}). \qquad (16)$$

It is useful to introduce a notion of *consistency* between this class of Gaussian states specified by a matching M and an interaction subset I.

Definition 13. *An (even) interaction subset $I \subseteq [2n]$ and a perfect matching M on $[2n]$ are called consistent if M contains a perfect matching of the elements of I. Given a set of interactions $\mathcal{I}$, we say that M is consistent (resp. inconsistent) with $\mathcal{I}$ if M is consistent (resp. inconsistent) with **each** interaction in $\mathcal{I}$.*

The following Lemma is straightforward

Lemma 14. *Consider a matching M and an interaction $I = \{i_1, i_2, ..i_q\}$.*

1. *If M is consistent with interaction I, let the perfect matching on the subset I be given by pairs $(i_{\pi(2l-1)}, i_{\pi(2l)})$ for $l = 1, \ldots, q/2$ and a permutation $\pi \in S_q$ where $i_{\pi(2l-1)} < i_{\pi(2l)}$. Then, the following holds:*

$$\mathrm{Tr}(C_I \rho(M, \vec{\lambda})) = \mathrm{sign}(\pi) \prod_{l \in \{1, \ldots, q/2\}} \lambda_{(i_{\pi(2l-1)}, i_{\pi(2l)})}.$$

where $\mathrm{sign}(\pi) = \pm 1$.

2. *If M is inconsistent with I, then*

$$\mathrm{Tr}(C_I \rho(M, \vec{\lambda})) = 0. \tag{17}$$

Proof. In order for the trace to be nonzero, one needs to exactly match the Majorana operators in C_I with some in the expansion of $\rho(M, \vec{\lambda})$ since $\mathrm{Tr}(C_{I'}) = 0$ for any non-empty subset I'. If M is inconsistent, there is no term in the expansion of ρ which precisely matches C_I, so the expectation vanishes. If M is consistent, we have

$$\mathrm{Tr}(C_I \rho(M, \vec{\lambda}))$$
$$= \frac{1}{2^n} \mathrm{Tr}\big(C_I \Pi_{(m_1, m_2) \in M}(\mathbb{I} + i\lambda_{(m_1, m_2)} c_{m_1} c_{m_2})\big)$$
$$= \frac{1}{2^n} \mathrm{Tr}\big(C_I \Pi_{(m_1, m_2) \in M, (m_1, m_2) \cap I \neq \emptyset}$$
$$\times (i\lambda_{(m_1, m_2)} c_{m_1} c_{m_2})\big)$$
$$= \mathrm{sign}(\pi) \prod_{l \in [1, .., q/2]} \lambda_{(i_{\pi(2l-1)}, i_{\pi(2l)})}. \tag{18}$$

Here we have used that one can first reorder C_I such that the pairs in the perfect matching are adjacent, i.e. $C_I = \mathrm{sign}(\pi) i^{q/2} c_{i_{\pi(1)}} c_{i_{\pi(2)}} \ldots c_{i_{\pi(q)}}$, then one can commute through each pair to its matching pair in ρ and use $(c_i c_j)^2 = -\mathbb{I}$, $i^q = (-1)^{q/2}$ and $\mathrm{tr}(\mathbb{I}) = 2^n$. □

5 Approximation ratios for sparse fermionic Hamiltonians

In this section we prove Theorem 5. We begin by setting up needed definitions and stating several technical Lemmas (which are proved in the Appendices).

The key auxiliary notion in the proof of Theorem 5 is that of a *diffuse* subset of Hamiltonian terms. Intuitively, the terms in a diffuse subset are well separated from each other while covering only a limited part of the system. This idea is formalized as follows:

Definition 15. *Consider a set of q-local interactions $\mathcal{I}$ on $2n$ Majorana operators. A subset of these interactions $\mathcal{I}' \subset \mathcal{I}$ is diffuse with respect to $\mathcal{I}$, if the following three conditions apply:*

1. *$\forall I_1, I_2 \in \mathcal{I}'$, I_1 and I_2 don't share any Majorana operators, i.e. $I_1 \cap I_2 = \emptyset$.*

2. *$\forall I_1, I_2 \in \mathcal{I}'$, there exists no $I_3 \in \mathcal{I}$ which shares Majorana operators with both I_1 and I_2 (if $I_3 \cap I_1 \neq \emptyset$ then $I_3 \cap I_2 = \emptyset$ and vice versa).*

3. *The size of support of $\mathcal{I}'$, i.e. $|\mathrm{Sup}(\mathcal{I}')|$, is smaller than $\frac{2qn}{q+1}$.*

In the setting of Theorem 5, diffuse sets of terms appear naturally due to the following Lemma.

Lemma 16. *Consider a k-sparse strictly q-local fermionic Hamiltonian H on $2n$ Majoranas. The interaction set $\mathcal{I}$ of H can be split into Q disjoint subsets $\mathcal{I}_\alpha$ ($\alpha \in [Q]$) all of which are diffuse with respect to $\mathcal{I}$ such that*

$$\mathcal{I} = \bigcup_{\alpha=1}^{Q} \mathcal{I}_\alpha. \tag{19}$$

The parameter Q is given as $Q = q(q-1)(k-1)^2 + q(k-1) + 2$ and does not depend on n. The construction of this splitting can be done efficiently, in time $\mathrm{poly}(n)$.

Lemma 16 is a special case of Lemma 19, which is proven in Appendix B. The proof relies on a combinatorial argument on a graph that takes Hamiltonian terms as vertices and connects them with an edge if the pair violates conditions 1 or 2 of Definition 15. By the sparsity assumption, this graph has an efficiently constructable coloring with a bounded number of colors, from which the split $\mathcal{I} = \bigcup_{\alpha=1}^{Q} \mathcal{I}_\alpha$ can be constructed.

The usefulness of diffuse sets comes from Lemma 20, see its proof in Appendix C. Here we state its corollary, relevant to proving Theorem 5:

Lemma 17. *Let the interaction set $\mathcal{I}'$ be diffuse w.r.t. $\mathcal{I} \supset \mathcal{I}'$ ($\mathcal{I}'$ and $\mathcal{I}$ are strictly q-local and k-sparse). If $n > (q^2 - 1)k$, one can efficiently construct a matching M of the set $[2n]$ that is consistent with each interaction in $\mathcal{I}'$ and inconsistent with each interaction in $\mathcal{I} \backslash \mathcal{I}'$.*

With matchings introduced above, one can construct useful Gaussian states. The tool to do so is given by the following statement:

Lemma 18. *Let $H = \sum_{I \in \mathcal{I}} J_I C_I$ be strictly q-local and $\mathcal{I}'$ be a diffuse subset of $\mathcal{I}$. Let M be a matching of $[2n]$ as guaranteed by Lemma 17. One can efficiently construct a Gaussian state $\rho_{\mathcal{I}'}$ with the property:*

$$\mathrm{Tr}(H\rho_{\mathcal{I}'}) = \sum_{I \in \mathcal{I}'} |J_I|. \tag{20}$$

Lemma 18 is a specific case of a slightly more general Lemma 21, which is stated and proven in Appendix D. We denote

$$\mathbf{J}(\mathcal{I}') \equiv \sum_{I \in \mathcal{I}'} |J_I|. \tag{21}$$

As shown below, Theorem 5 can be proven by constructing a diffuse $\mathcal{I}' \subset \mathcal{I}$ and a corresponding Gaussian state $\rho_{\mathcal{I}'}$ with large enough $\mathrm{Tr}(H\rho_{\mathcal{I}'}) = \mathbf{J}(\mathcal{I}')$.

Theorem (Repetition of Theorem 5). *Let H be a traceless fermionic Hamiltonian on $2n$ Majoranas with maximal eigenvalue $\lambda_{\max}(H)$. If H is k-sparse and strictly q-local and $n > (q^2 - 1)k$, a Gaussian state ρ can be efficiently constructed, such that*

$$\frac{\mathrm{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{Q}, \tag{22}$$

for $Q = q(q-1)(k-1)^2 + q(k-1) + 2$.

Proof. For a Hamiltonian $H = \sum_{I \in \mathcal{I}} J_I C_I$, we construct the splitting of $\mathcal{I}$ into diffuse subsets $\mathcal{I} = \cup_\alpha \mathcal{I}_\alpha$ as guaranteed by Lemma 16. Next, find $\alpha = \mathrm{argmax}_{\alpha'} \mathbf{J}(\mathcal{I}_{\alpha'})$; since Q in Lemma 16 is constant, α can be found efficiently. Next, use Lemma 17 to construct a matching $M(\mathcal{I}_\alpha)$ (the condition $n > (q^2 - 1)k$ is satisfied by assumptions of Theorem 5). Since $\mathcal{I}_\alpha$ is diffuse with respect to $\mathcal{I}$, the Gaussian state $\rho_{\mathcal{I}_\alpha}$ can be efficiently constructed from $M(\mathcal{I}_\alpha)$ via Lemma 18. Using $\mathrm{Tr}(H\rho_{\mathcal{I}_\alpha}) = \mathbf{J}(\mathcal{I}_\alpha)$, the following inequality can be obtained for the resulting approximation ratio:

$$\frac{\mathrm{Tr}(H\rho_{\mathcal{I}_\alpha})}{\lambda_{\max}(H)} \geq \frac{\mathbf{J}(\mathcal{I}_\alpha)}{\sum_{\alpha'} \mathbf{J}(\mathcal{I}_{\alpha'})} \geq \frac{1}{Q}. \tag{23}$$

For the first inequality, note that $\lambda_{\max}(H) \leq \sum_{I \in \mathcal{I}} |J_I| = \sum_\alpha \mathbf{J}(\mathcal{I}_\alpha)$. The second inequality comes from a pigeonhole-type argument: if $\mathbf{J}(\mathcal{I}_\alpha) = \max_{\alpha'} \mathbf{J}(\mathcal{I}_{\alpha'})$, it directly follows that $\mathbf{J}(\mathcal{I}_\alpha) \geq \frac{1}{Q} \sum_{\alpha'} \mathbf{J}(\mathcal{I}_{\alpha'})$. Inequality (23) concludes the proof, as it asserts the approximation ratio bound claimed in the Theorem. $\square$

6 Sparse Hamiltonians with terms of weight 2 and 4

In this section we prove Theorem 6. We will again need to use the concept of diffuse subsets in Definition 15. The proof of Theorem 6 is similar in its basic idea to that of Theorem 5. The main obstacle in this case is the presence of terms of different weight, which does not allow one to use Lemmas 16-18 directly. This can be resolved by a slightly more elaborate construction and applying the more general Lemmas 19-21 which are proved in the Appendices and Lemmas 16-18 directly follow as special cases.

Lemma 19 (Generalization of Lemma 16). *Let $\mathcal{I}$ be the interaction set of a k-sparse q-local Hamiltonian on the set of Majorana fermions $[2n]$. The set $\mathcal{I}$ can be split into $(qQ)/2$ disjoint, strictly $2q'$-local subsets $\mathcal{I}_\alpha^{(2q')}$ (with $\alpha \in [Q]$ and $q' \in [q/2]$) each of which is diffuse with respect to $\mathcal{I}$:*

$$\mathcal{I} = \bigcup_{q'=1}^{q/2} \bigcup_{\alpha=1}^{Q} \mathcal{I}_\alpha^{(2q')}. \tag{24}$$

The parameter $Q = q(q-1)(k-1)^2 + q(k-1) + 2$ does not grow with n. The construction of this splitting can be done efficiently, in time $\mathrm{poly}(n)$.

Lemma 20 (Generalization of Lemma 17). *Let a strictly q'-local $\mathcal{I}'$ be diffuse w.r.t. q-local k-sparse $\mathcal{I}$ on $[2n]$, such that $n > (q^2 - 1)k$. One can efficiently construct a matching M of $[2n]$ that is consistent with $\mathcal{I}'$ and inconsistent with all interactions $I \in \mathcal{I} \backslash \mathcal{I}'$ such that (1) $|I| \geq q'$ or (2) $I \not\subset \mathrm{Sup}(\mathcal{I}')$.*

Lemma 21 (Generalization of Lemma 18). *Let $H = \sum_{I \in \mathcal{I}} J_I C_I$ on $[2n']$ be q-local and $\mathcal{I}'$ be a diffuse subset of $\mathcal{I}$. Consider a matching M of $[2n']$. If M is consistent with $\mathcal{I}'$ and inconsistent with $\mathcal{I} \backslash \mathcal{I}'$, one can efficiently construct a Gaussian state $\rho_{\mathcal{I}'}$ with the property:*

$$\mathrm{Tr}(H\rho_{\mathcal{I}'}) = \sum_{I \in \mathcal{I}'} |J_I|. \tag{25}$$

In Lemma 21, we use n' instead of n to avoid confusion, as it will also be used for $n' \neq n$. The Lemmas above are proven in Appendices B-D. With these in hand, we are ready to proceed with the proof of Theorem 6.

Theorem (Repetition of Theorem 6). *Let H be a traceless fermionic Hamiltonian on $[2n]$ with maximal eigenvalue $\lambda_{\max}(H)$. If H is k-sparse with terms of weight 2 and 4 and $2n > 15k$, a Gaussian state ρ can be efficiently constructed, such that*

$$\frac{\mathrm{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{2Q} \tag{26}$$

with $Q = 12(k-1)^2 + 4(k-1) + 2$.

Proof. We make use of the construction in Ref. [12] which relates a Hamiltonian with weights 2 and 4 on a set of fermionic modes $[2n]$, that is,

$$H = \sum_{I \in \mathcal{I}^{(2)}} J_I C_I + \sum_{I \in \mathcal{I}^{(4)}} J_I C_I, \tag{27}$$

to a strictly 4-local Hamiltonian $\tilde{H}$ on an extended set of fermions $[2n + 2]$:

$$\tilde{H} = \sum_{I \in \mathcal{I}^{(2)}} (-ic_{2n+1}c_{2n+2}) J_I C_I + \sum_{I \in \mathcal{I}^{(4)}} J_I C_I. \tag{28}$$

Introducing $\tilde{\mathcal{I}}^{(2)} \equiv \{(2n+1, 2n+2) \cup I \mid I \in \mathcal{I}^{(2)}\}$, $\tilde{H}$ can be also written as:

$$\tilde{H} = -\sum_{I \in \tilde{\mathcal{I}}^{(2)}} J_I C_I + \sum_{I \in \mathcal{I}^{(4)}} J_I C_I. \tag{29}$$

The relation between $\tilde{H}$ and H is via the following property:

Lemma 22 (Lemma 6 of [12]). *For H and $\tilde{H}$ introduced above, $\lambda_{\max}(H) = \lambda_{\max}(\tilde{H})$. Moreover, for any Gaussian state $\tilde{\rho}$ of $2n + 2$ Majorana modes, one can efficiently compute a Gaussian state ρ of $2n$ Majorana modes s.t. $\mathrm{Tr}(H\rho) \geq \mathrm{Tr}(\tilde{H}\tilde{\rho})$.*

Although strictly 4-local, Hamiltonian $\tilde{H}$ is no longer sparse since the operators c_{2n+1} and c_{2n+2} participate in $|\mathcal{I}^{(2)}|$ terms (which is generally $O(n)$). This prevents a direct application of Lemma 16 to $\tilde{H}$. We resolve the issue as follows.

Similarly to the proof of Theorem 5, we start by splitting each set of the original interactions $\mathcal{I}^{(2,4)}$ in H into subsets diffuse w.r.t. $\mathcal{I}^{(2)} \cup \mathcal{I}^{(4)}$: $\mathcal{I}^{(2)} = \cup_\alpha \mathcal{I}_\alpha^{(2)}$, $\mathcal{I}^{(4)} = \cup_\alpha \mathcal{I}_\alpha^{(4)}$. Each of the two splittings exists and can be done efficiently, as guaranteed by Lemma 16 (since the original H is sparse). Since $\mathcal{I}^{(2)} \cup \mathcal{I}^{(4)}$ is k-sparse and 4-local, we can bound $|\{\mathcal{I}_\alpha^{(2)}\}| < Q$, $|\{\mathcal{I}_\alpha^{(4)}\}| < Q$ for $Q = 12(k-1)^2 + 4(k-1) + 2$. In what follows, we will use the splittings $\mathcal{I} = \bigcup_\alpha \mathcal{I}_\alpha^{(2)} \cup \bigcup_\alpha \mathcal{I}_\alpha^{(4)}$ to construct two Gaussian states $\tilde{\rho}(\mathcal{I}_\alpha^{(2)})$ and $\tilde{\rho}(\mathcal{I}_\alpha^{(4)})$ on $[2n+2]$ with good properties relative to $\tilde{H}$, that is,

$$\text{Tr}(\tilde{H}\tilde{\rho}(\mathcal{I}_\alpha^{(2,4)})) = \sum_{I \in \mathcal{I}_\alpha^{(2,4)}} |J_I| \equiv \mathbf{J}(\mathcal{I}_\alpha^{(2,4)}). \qquad (30)$$

With these Gaussian states, we will then show that the Gaussian state $\tilde{\rho}(\mathcal{I}_\alpha^{(2,4)})$ for $q, \alpha = \text{argmax}_{q', \alpha'}(\mathbf{J}(\mathcal{I}_{\alpha'}^{(q')}))$ is efficiently constructable and yields the desired approximation ratio for $\tilde{H}$. We will then apply Lemma 22 and extend the statement to the original Hamiltonian H, thus finishing the proof.

Following the outline above, we now move to construct the Gaussian state $\tilde{\rho}(\mathcal{I}_\alpha^{(2)})$. Consider an ansatz of the form $\tilde{\rho} \equiv \rho_{[2n]}\sigma_{\{2n+1,2n+2\}}$, where $\rho_{[2n]}$ is itself a Gaussian state of $[2n]$. To construct $\rho_{[2n]}$, note that each $\mathcal{I}_\alpha^{(2)}$ is 2$-$local and diffuse w.r.t. $\mathcal{I}^{(2)} \cup \mathcal{I}^{(4)}$ which is 4-local. Since $2n > 15k$ by assumptions of Theorem 6, we can apply Lemma 20 with $q = 4$ to construct a matching $M(\mathcal{I}_\alpha^{(2)})$ that is consistent with $\mathcal{I}_\alpha^{(2)}$. Since $\mathcal{I}_\alpha^{(2)}$ is 2-local, Lemma 20 also implies that the matching $M(\mathcal{I}_\alpha^{(2)})$ is inconsistent with the entirety of $\mathcal{I}^{(4)} \cup \mathcal{I}^{(2)}\backslash\mathcal{I}_\alpha^{(2)}$. We then use $M(\mathcal{I}_\alpha^{(2)})$ in Lemma 21 (substituting $n' = n$ for n' used in the Lemma) to construct $\rho_{[2n]}$ in $\tilde{\rho} = \rho_{[2n]}\sigma_{\{2n+1,2n+2\}}$. This implies the following expression (using Eq. (28) for $\tilde{H}$):

$$\text{Tr}(\tilde{H}\tilde{\rho}) = \text{Tr}(\sigma(-ic_{2n+1}c_{2n+2})) \sum_{I \in \mathcal{I}_\alpha^{(2)}} |J_I|. \qquad (31)$$

By choosing σ to be the $+1$ eigenstate projector of operator $-ic_{2n+1}c_{2n+2}$, we arrive at the desired outcome:

$$\text{Tr}(\tilde{H}\tilde{\rho}) = \sum_{I \in \mathcal{I}_\alpha^{(2)}} |J_I| = \mathbf{J}(\mathcal{I}_\alpha^{(2)}). \qquad (32)$$

The constructed Gaussian state $\tilde{\rho}$ we will denote as $\tilde{\rho}(\mathcal{I}_\alpha^{(2)})$.

For a diffuse $\mathcal{I}_\alpha^{(4)} \subset \mathcal{I}^{(2)} \cup \mathcal{I}^{(4)}$, we construct the Gaussian states $\tilde{\rho}(\mathcal{I}_\alpha^{(4)})$ in a different way. First we use Lemma 20 to construct a matching $M(\mathcal{I}_\alpha^{(4)})$ of $[2n]$. This matching is guaranteed to be consistent with $\mathcal{I}_\alpha^{(4)}$. However, since $\mathcal{I}^{(2)}$ is 2-local and $\mathcal{I}^{(4)}$ is
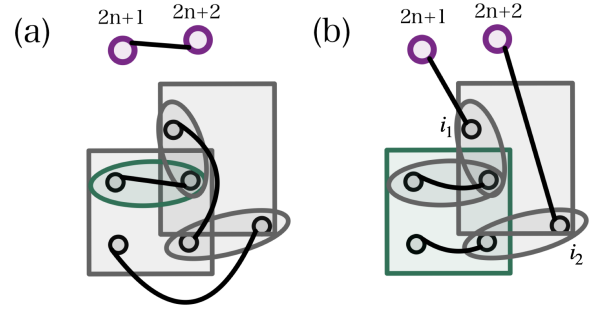

Figure 2: Demonstration of the method in the proof of Theorem 6. (a) Matching $M(\mathcal{I}_\alpha^{(2)})$ for $\mathcal{I}_\alpha^{(2)}$, here comprised of a single term (shown in green). To ensure consistency with $\mathcal{I}_\alpha^{(2)}$ in $\tilde{H}$, $M(\mathcal{I}_\alpha^{(2)})$ perfectly matches these terms and the pair $(2n+1, 2n+2)$. The rest of the vertices are matched so that each pair does not belong to the same term in $\mathcal{I}\backslash\mathcal{I}_\alpha^{(2)}$ (grey). (b) Matching $M(\mathcal{I}_\alpha^{(4)})$ for $\mathcal{I}_\alpha^{(4)}$ shown in green. Vertices i_1, i_2 are chosen not to belong to the same term in $\mathcal{I}^{(2)}$, ensuring no accidental consistency with a term in $\tilde{H}$. Note the special status of the term from $\mathcal{I}^{(2)}$ that is a subset of the $\mathcal{I}_\alpha^{(4)}$ term. From the perspective of $\tilde{H}$, it is not consistent with $M(\mathcal{I}_\alpha^{(4)})$ although it coincides with an edge from $M(\mathcal{I}_\alpha^{(4)})$. This is due to the intentional absence of the edge $(2n+1, 2n+2)$ in $M(\mathcal{I}_\alpha^{(4)})$.

4-local, while in general $\text{Sup}(\mathcal{I}^{(2)}) \cap \text{Sup}(\mathcal{I}_\alpha^{(4)}) \neq \emptyset$, Lemma 20 implies that $M(\mathcal{I}_\alpha^{(4)})$ is inconsistent with $\mathcal{I}^{(4)}\backslash\mathcal{I}_\alpha^{(4)}$ but may be consistent with some terms in $\mathcal{I}^{(2)}$ (as those I don't obey the $|I| \geq q' = 4$ condition). At the same time, we aim to achieve $\text{Tr}(\tilde{H}\tilde{\rho}(\mathcal{I}_\alpha^{(4)})) = \mathbf{J}(\mathcal{I}_\alpha^{(4)})$ which excludes contributions from $\mathcal{I}^{(2)}$. Thus we cannot extend $M(\mathcal{I}_\alpha^{(4)})$ to the extended set $[2n+2]$ directly, as it was done for $\mathcal{I}_\alpha^{(2)}$. Instead, we will create a matching of $[2n+2]$ using a reduced version of $M(\mathcal{I}_\alpha^{(4)})$ which inherits its beneficial properties, and then complete the matching by making it inconsistent with $\tilde{\mathcal{I}}^{(2)}$ – eliminating the difficulty described above.

To enable this, we find and mark an edge $(i_1, i_2) \in M(\mathcal{I}_\alpha^{(4)})$, such that $i_1 \notin \text{Sup}(\mathcal{I}_\alpha^{(4)})$. This is always possible since $\mathcal{I}_\alpha^{(4)}$ is diffuse and thus $[2n]\backslash\text{Sup}(\mathcal{I}_\alpha^{(4)})$ is non-empty (cf. Condition 3 in Definition 15). Note that $M(\mathcal{I}_\alpha^{(4)})$ is constructed via Lemma 20 and $\{i_1, i_2\} \not\subset \text{Sup}(\mathcal{I}_\alpha^{(4)})$. This implies that as a two-fermion interaction, $\{i_1, i_2\}$ is guaranteed not to belong to $\mathcal{I}^{(2)}$. The latter statement is the key property of the marked edge (i_1, i_2) that we will employ momentarily.

We construct a matching $\tilde{M}(\mathcal{I}_\alpha^{(4)})$ of $[2n+2]$ in two stages. First we construct an intermediate matching $M'(\mathcal{I}_\alpha^{(4)})$ of $[2n+2]\backslash\{i_1, i_2, 2n+1, 2n+2\}$ by removing the edge from $M(\mathcal{I}_\alpha^{(4)})$:

$$M'(\mathcal{I}_\alpha^{(4)}) = M(\mathcal{I}_\alpha^{(4)})\backslash(i_1, i_2). \qquad (33)$$

Since $\{i_1, i_2\} \not\subset \mathrm{Sup}(\mathcal{I}_\alpha^{(4)})$, we are guaranteed that $M'(\mathcal{I}_\alpha^{(4)})$ is consistent with $\mathcal{I}_\alpha^{(4)}$ and inconsistent with $\mathcal{I}^{(4)} \backslash \mathcal{I}_\alpha^{(4)}$ (from the construction of $M(\mathcal{I}_\alpha^{(4)})$). In the second stage, we complete $\tilde{M}(\mathcal{I}_\alpha^{(4)})$ to the entire set of $2n+2$ modes by adding two edges: $(i_1, 2n+1)$ and $(i_2, 2n+2)$:

$$\tilde{M}(\mathcal{I}_\alpha^{(4)}) = M(\mathcal{I}_\alpha^{(4)}) \cup \{(i_1, 2n+1), (i_2, 2n+2)\}. \tag{34}$$

These new edges render $\tilde{M}(\mathcal{I}_\alpha^{(4)})$ inconsistent with $\tilde{\mathcal{I}}^{(2)}$. To see it, note that all interactions in $\tilde{\mathcal{I}}^{(2)}$ take the form $I = \{j_1, j_2, 2n+1, 2n+2\}$ where $\{j_1, j_2\} \in \mathcal{I}^{(2)}$. By construction $\{i_1, i_2\} \notin \mathcal{I}^{(2)}$, thus we have $\{j_1, j_2\} \neq \{i_1, i_2\}$. As a result, matching $\tilde{M}(\mathcal{I}_\alpha^{(4)})$ of $[2n+2]$ is consistent with $\mathcal{I}_\alpha^{(4)}$ and inconsistent with $\tilde{\mathcal{I}}^{(2)} \cup \mathcal{I}^{(4)} \backslash \mathcal{I}_\alpha^{(4)}$. We continue by applying Lemma 21 to such $\tilde{M}(\mathcal{I}_\alpha^{(4)})$ and $\tilde{H}$ (substituting $n' = n+1$ for n' used in the Lemma). This efficiently constructs a Gaussian state $\tilde{\rho}(\mathcal{I}_\alpha^{(4)})$ that yields:

$$\mathrm{Tr}(\tilde{H}\tilde{\rho}(\mathcal{I}_\alpha^{(4)})) = \sum_{I \in \mathcal{I}_\alpha^{(4)}} |J_I| \equiv \mathbf{J}(\mathcal{I}_\alpha^{(4)}), \tag{35}$$

as desired.

The Gaussian state claimed in Theorem 6 is to be chosen among the states $\tilde{\rho}(\mathcal{I}_\alpha^{(2,4)})$ whose existence we've proven above. We make the choice by identifying the highest energy in the respective Gaussian state: $(q, \alpha) = \mathrm{argmax}_{(q,\alpha)} \mathbf{J}_\alpha^{(q)}$. As we showed, the respective Gaussian state $\tilde{\rho}(\mathcal{I}_\alpha^{(q)})$ can be efficiently constructed and the following is guaranteed:

$$\frac{\mathrm{Tr}(\tilde{H}\tilde{\rho}(\mathcal{I}_\alpha^{(q)}))}{\lambda_{\max}(\tilde{H})} \geq \frac{\mathbf{J}_\alpha^{(q)}}{\sum_{q',\alpha'} \mathbf{J}_{\alpha'}^{(q')}} \geq \frac{1}{2Q}. \tag{36}$$

Here we used that $\lambda_{\max}(\tilde{H}) \leq \sum_{q',\alpha'} \mathbf{J}_{\alpha'}^{(q')}$ and that $\mathbf{J}_\alpha^{(q)} = \max_{(q',\alpha')} \mathbf{J}_{\alpha'}^{(q')}$.

With the state $\tilde{\rho}(\mathcal{I}_\alpha^{(q)})$ on $[2n+2]$ fermions at hand, we finalize the proof by an application of Lemma 22. This relates $\lambda_{\max}(H)$ to $\lambda_{\max}(\tilde{H})$ and allows us to efficiently construct the Gaussian state $\rho(\mathcal{I}_\alpha^{(q)})$ of $[2n]$, with the desired property:

$$\frac{\mathrm{Tr}(H\rho(\mathcal{I}_\alpha^{(q)}))}{\lambda_{\max}(H)} \geq \frac{\mathrm{Tr}(\tilde{H}\tilde{\rho}(\mathcal{I}_\alpha^{(q)}))}{\lambda_{\max}(\tilde{H})} \geq \frac{1}{2Q}. \tag{37}$$

□

7 The sparse SYK-4 model

Theorem (Repetition of Theorem 8)**.** *Let H be a SSYK-4 Hamiltonian in Eq. (5) with expected degree $k = O(1)$, such that $n > 120(k+1)$. With probability*

at least $1 - 4\,exp\left[-\frac{e^{-16(k+1)}k^3}{64(8k+7)}n\right]$, a Gaussian state ρ can be efficiently constructed such that

$$\frac{\mathrm{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{Q}, \tag{38}$$

where $Q = 1236 + 2752k + 1536k^2$.

Proof. In what follows we will omit the normalization $1/\sqrt{2kn}$ in Eq. (5), of course this normalization is irrelevant for lowerbounding the Gaussian approximation ratio. We split H as $H = H^{(k')} + h^{(k')}$, s.t. the Hamiltonian $H^{(k')}$ is k'-sparse and the residual Hamiltonian $h^{(k')}$ contains the rest of H. The term sets are denoted as follows:

$$H^{(k')} = \sum_{I \in \mathcal{I}^{(k')}} J_I C_I, \quad h^{(k')} = \sum_{I \in \bar{\mathcal{I}}^{(k')}} J_I C_I, \tag{39}$$

i.e. $\mathcal{I} = \mathcal{I}^{(k')} \cup \bar{\mathcal{I}}^{(k')}$. To define such a split, we use the following deterministic algorithm. For every given Majorana, we list the interactions $I \in \mathcal{I}$ which involve that Majorana using a lexicographical order for the words $I = \{i_1, i_2, i_3, i_4\}$. For each Majorana where such a list is longer than k', we mark all elements except for the first k'. All terms of H which were marked this way at least once, we include into $h^{(k')}$. The rest of the terms enter $H^{(k')}$, which by this construction is k'-sparse. To continue the proof we need a pair of Lemmas. The first lower bounds the total interaction strength of the SSYK-4 Hamiltonian:

Lemma 23. *With probability at least $1 - 2e^{-\frac{kn}{32}}$, we have*

$$\sum_{I \in \mathcal{I}} |J_I| \geq kn/8. \tag{40}$$

This statement is proven in Appendix E, by splitting the problem into upper bounding $|\mathcal{I}|$ separately from $\sum_I |J_I|$, and then applying the Chernoff bound for both.

The second lemma shows that the total interaction strength of the residual Hamiltonian $h^{(k')}$ is bounded from above with high probability:

Lemma 24. *If $k' \geq e^2 k + 1$, we have with probability at least $1 - 2\,exp\left[-\frac{e^{-2k'}k^3}{64(k'-1)}n\right]$ that*

$$\sum_{I \in \bar{\mathcal{I}}^{(k')}} |J_I| \leq \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n. \tag{41}$$

Lemma 24 is proven in Appendix E. The key technical difficulty is bounding the random variable $\bar{\mathcal{I}}^{(k')}$, which does not reduce to a sum of independent variables and thus a simple Chernoff bound cannot be applied. Instead, we apply an exponential version of Efron-Stein inequality [25].

To build a Gaussian state with finite approximation ratio, we apply the construction of Theorem 5 to $H^{(k')}$, which is k'-sparse and strictly 4-local. If n

is large enough (i.e. $n > (q^2 - 1)k'$ for $q = 4$), this state ρ is guaranteed to yield energy $\text{Tr}(H^{(k')}\rho) > \frac{1}{Q'}\sum_{I\in\mathcal{I}^{(k')}}|J_I|$ for $Q' = 12k'^2 - 20k' + 10$ (see Eq. (23) in the proof of Theorem 5). At the same time, with high probability $|\text{Tr}(h^{(k')}\rho)| \leq \sum_{I\in\bar{\mathcal{I}}^{(k')}}|J_I| \leq \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n$ and $\sum_{I\in\mathcal{I}}|J_I| \geq \frac{kn}{8}$ (Lemmas 23 and 24). The resulting approximation ratio is then:

$$
\begin{aligned}
\frac{\text{Tr}(H\rho)}{\lambda_{\max}(H)} &\geq \frac{\text{Tr}(H^{(k')}\rho) - |\text{Tr}(h^{(k')}\rho)|}{\sum_{I\in\mathcal{I}}|J_I|} \\
&\geq \frac{\frac{1}{Q'}\sum_{I\in\mathcal{I}}|J_I| - (1 + \frac{1}{Q'})\sum_{I\in\bar{\mathcal{I}}^{(k')}}|J_I|}{\sum_{I\in\mathcal{I}}|J_I|} \\
&\geq \frac{1}{Q'} - \frac{32(Q'+1)ke^{-k'}}{\sqrt{k'-1}Q'}.
\end{aligned} \tag{42}
$$

Crucially, the second term decays exponentially with k' and the first term only algebraically (note here the definition of Q'). We now fix $k' = 8(k+1)$, consistent with the requirement $k' \geq e^2 k + 1$ of Lemma 24. In this case $\frac{32(Q'+1)ke^{-k'}}{\sqrt{k'-1}Q'}$ as a function of k is always smaller than $\frac{1}{2Q'}$. This allows us to bound the right hand side of Eq. (42) as $\frac{1}{2Q'}$, and substituting $k' = 8(k+1)$ we obtain the bound claimed in the Theorem:

$$
\frac{\text{Tr}(H\rho)}{\lambda_{\max}(H)} \geq \frac{1}{1236 + 2752k + 1536k^2}. \tag{43}
$$

The earlier assumed condition $n > (q^2 - 1)k'$ for $q = 4$ and k' translates into $n > 120(k+1)$. Given the conditions of Lemmas 23 and 24, the bound in Eq. (43) holds with the probability:

$$
\begin{aligned}
&\left(1 - 2\exp\left[-\frac{e^{-16(k+1)}k^3}{64(8k+7)}n\right]\right)\left(1 - 2e^{-\frac{kn}{32}}\right) \\
&\geq 1 - 4\exp\left[-\frac{e^{-16(k+1)}k^3}{64(8k+7)}n\right].
\end{aligned} \tag{44}
$$

$\square$

8 Upper bound on Gaussian approximation ratio for SYK-q Hamiltonians

8.1 Gaussian upper bound for SYK-q models

We consider the expectation value of a SYK-q Hamiltonian H with respect to fermionic Gaussian states and we obtain an upper bound on its expectation value, with high probability over the random couplings J_I.

Lemma (Repetition of Lemma 9). *Let H denote a Hamiltonian drawn from the q-local SYK Hamiltonians (with $q \geq 4$ even and $q = O(1)$), i.e. the coupling strengths J_I are drawn according to their distribution. With probability at least $1 - \exp(-\Omega(n))$, H has the*

property that, for any fermionic Gaussian state ρ

$$
\begin{aligned}
\text{Tr}(H\rho) &\leq (q-1)!!\, 2^{1/2 - q/4}q^{1/2 + q/2} \\
&\quad \times \sqrt{\log[q/\log(3/2)]}\,(2n)^{1 - q/4}. \tag{45}
\end{aligned}
$$

Proof. We first use Wick's theorem on the expectation of a product of Majorana operators w.r.t. a fermionic Gaussian state ρ characterized by a correlation matrix Γ, see Eq. (15). Note that the correlation matrix $\Gamma_{i<j}$ can be viewed as a real $d := (2n^2 - n)$-dimensional vector. We note that $\sum_{i<j}\Gamma_{ij}^2 = \frac{1}{2}\text{Tr}(\Gamma^T\Gamma) \leq \frac{1}{2}\text{Tr}(\mathbb{I}) = n$ so that $\|\Gamma\| \leq n^{1/2}$.

Let $M(I)$ be a perfect matching of the indices in I ($|I|$ even), there are $(q-1)!!$ such matchings. We have

$$
\begin{aligned}
\text{Tr}(C_I\rho) = i^{q/2} \sum_{M(I)} &\text{sign}(M(I))\, \text{Tr}(c_{i_1(M)}c_{i_2(M)}\rho) \\
&\times \text{Tr}(c_{i_3(M)}c_{i_4(M)}\rho)\ldots\text{Tr}(c_{i_{q-1}(M)}c_{i_q(M)}\rho). \tag{46}
\end{aligned}
$$

Here we have assumed that for each matching $M(I)$; $i_1(M) < i_2(M)$, $i_3(M) < i_4(M)$, ..., $i_{q-1}(M) < i_q(M)$, i.e. any sign arising from getting the expression to this form is absorbed in $\text{sign}(M(I))$.

The expectation of H in Eq. (2) w.r.t. fermionic Gaussian states ρ can be written as:

$$
\begin{aligned}
\text{Tr}(H\rho) &= \binom{2n}{q}^{-1/2}i^{q/2}\sum_{I\subseteq[2n],\,|I|=q}J_I\Bigg[\sum_{M(I)}\text{sign}(M(I)) \\
&\quad\times\prod_{t=1}^{q/2}\text{Tr}(c_{i_{2t-1}(M)}c_{i_{2t}(M)}\rho)\Bigg] \\
&= \binom{2n}{q}^{-1/2}\sum_{I\subseteq[2n],\,|I|=q}J_I\sum_{M(I)}\text{sign}(M(I)) \\
&\quad\times\prod_{t=1}^{q/2}\Gamma_{i_{2t-1}(M),i_{2t}(M)}. \tag{47}
\end{aligned}
$$

We note that we can view $\text{Tr}(H\rho)$ as a sum of $(q-1)!!$ terms, one for each matching M of some subset of indices I, i.e. $\text{Tr}(H\rho) = \sum_M \text{Tr}(H_M\rho)$ where $H_M = \sum_I \tilde{J}(M,I)\prod_{t=1}^{q/2}\Gamma_{i_{2t-1}(M),i_{2t}(M)}$. We have defined the $q/2$-way, $d \times d \times \ldots \times d$, tensor $\tilde{J}(M,I)$, whose entries are equal to either zero (when the indices coincide or are not ordered properly) or to a standard Gaussian random variable. Each J_I appears only once in $\text{Tr}(H_M\rho)$ and therefore all entries of $\tilde{J}(M,I)$ are statistically independent. We note that $\text{sign}(M)$ does not depend on which (ordered) subset I one chooses. To bound each term $\text{Tr}(H_M\rho)$, with high probability, we invoke the following Lemma:

Lemma 25. *(Theorem 1 in [26].) Let A be a random K-way tensor $\in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_K}$ and w_i be vectors $\in \mathbb{R}^{d_i}$ and*

$$
A(w_1, w_2, \ldots, w_K) := \sum_{k_1,\ldots,k_K} A_{k_1,\ldots,k_K}(w_1)_{k_1}\ldots(w_{q/2})_{k_K}.
$$

If we have for each fixed unit vector $w_i/\|w_i\|$ ($i \in \{1,\dots,K\}$):

$$\Pr\Big(|A(w_1/\|w_1\|,\dots,w_K/\|w_K\|)| \geq t\Big) \leq 2\,exp\Big(-t^2/(2\sigma^2)\Big), \quad (48)$$

then the spectral norm $\|A\| := \max_{w_1,\dots,w_K} A\ \big(w_1/\|w_1\|,\dots,w_K/\|w_K\|\big)$ (with $w_i \in \mathbb{R}^{d_i}$) can be bounded as follows:

$$\|A\| \leq \Big[8\sigma^2\Big[\Big(\sum_{i=1}^{K} d_i\Big)\log\big[2K/\log(3/2]\big] + \log\big(\tfrac{2}{\delta}\big)\Big]\Big]^{\frac{1}{2}},$$

with probability at least $1-\delta$.

To apply the Lemma, note that the vectors w_i correspond to $\Gamma_{i<j}$ viewing $i < j$ as a single index and we can use their norm $\|\Gamma\| \leq n^{1/2}$. In addition, for each entry in the tensor we have $\mathbb{E}\big[\exp\big(t\tilde{J}(M,I)_{k_1,\dots,k_{q/2}}\big)\big] \leq \exp\big(t^2/2\big)$ (for $t \geq 0$) as the entry is zero or a Gaussian variable with variance 1 and mean zero. Using Chernoff's bound and the fact that all entries of $\tilde{J}(M,I)$ are statistically independent, we conclude that for any set of real vectors $w_1,\dots,w_{q/2}$ one has

$$\Pr\Big[\Big|\sum_{k_1,\dots,k_{q/2}}\tilde{J}(M,I)_{k_1,\dots,k_{q/2}}\frac{(w_1)_{k_1}}{\|w_1\|}\dots\frac{(w_{q/2})_{k_{q/2}}}{\|w_{q/2}\|}\Big| \geq t\Big] \leq 2\exp(-t^2/2). \quad (49)$$

Therefore, for each term H_M we can apply Lemma 25 and, using $K = q/2$ and $\sigma = 1$, obtain

$$\|\tilde{J}(M,I)\| \leq \Big[4q(2n^2-n)\log\big[q/\log(3/2)\big] + 8\log\big(2\delta^{-1}\big)\Big]^{1/2}, \quad (50)$$

with probability at least $1-\delta$. Then we can first bound

$$\max_{\rho \text{ Gaussian}}\operatorname{Tr}(H\rho) \leq \binom{2n}{q}^{-1/2}\|\Gamma\|^{q/2}\sum_M \|\tilde{J}(M,I)\|$$
$$\leq \big(q/\sqrt{2}\big)^{q/2}(2n)^{-q/4}\sum_M \|\tilde{J}(M,I)\|, \quad (51)$$

where we have used that $\binom{2n}{q} \geq (2n/q)^q$. We can now combine the upper bound in Eq. (50) and Eq. (51). Applying the union bound, we have with probability at least $1-(q-1)!!\,\delta$, that

$$\max_{\rho \text{ Gaussian}}\operatorname{Tr}(H\rho) \leq (q-1)!!\Big[2^{1-q/2}q^{q+1}$$
$$\times \big[(2n)^{2-q/2}-(2n)^{1-q/2}\big]\log\big[q/\log(3/2)\big]$$
$$+ 2^{3-q/2}q^q(2n)^{-q/2}\log\big(2\delta^{-1}\big)\Big]^{1/2}. \quad (52)$$

Therefore, we can take $\delta = \exp\big(-\Omega(n)\big)$ such that, asymptotically, we have (assuming $q = O(1)$):

$$\max_{\rho \text{ Gaussian}}\operatorname{Tr}(H\rho) \leq (q-1)!!\,2^{1/2-q/4}q^{1/2+q/2}$$
$$\times \sqrt{\log[q/\log(3/2)]}\,(2n)^{1-q/4}, \quad (53)$$

with probability at least $1-\delta$. Note that in deriving this upper bound we only use the norm of the correlation matrix Γ, hence this upper bound is not necessarily achievable by a Gaussian state as the constraint $\Gamma^T\Gamma \leq \mathbb{I}$ imposes more conditions on Γ than just an upper bound on its norm. $\qquad\square$

8.2 Maximum eigenvalue lower bound for q-local SYK Hamiltonians

To show that fermionic Gaussian states cannot achieve a constant approximation ratio for $q \geq 4$ SYK models, we derive a lower bound on the maximum eigenvalue of the Hamiltonians H in Eq. (2):

Lemma (Repetition of Lemma 10). *For the class of q-local SYK Hamiltonians (with even $q \geq 4$) in Eq. (2), $\lambda_{\max}(H) = \Omega(\sqrt{n})$ with probability at least $1 - exp\big(-\Omega(n)\big)$ over the draw of Hamiltonians.*

The remainder of this section will be devoted to proving this Lemma. The techniques used are similar to those used in Section 6 of Ref. [2]. We note that throughout this section, we shall use C to denote a quantity that is constant in n or is bounded from above and below by a constant in n, and it will generally differ from appearance to appearance (for the sake of clarity). Importantly, C can contain factors of q (note that $q = O(1)$).

We start by obtaining a lower bound on the maximum eigenvalue of a so-called 2-colored SYK model and will use this to prove Lemma 10. The Hamiltonian of such a 2-colored SYK model is slightly different from the standard SYK model Hamiltonian in Eq. (2). We divide the $2n$ Majorana operators into two subsets, with sizes n_1 and n_2 ($n_2 \leq n_1$), and denote the operators in the first set by $\phi_1,\dots,\phi_{n_1}$ and the ones in the second set by $\chi_1,\dots,\chi_{n_2}$. The Hamiltonian is now given by[1]:

$$H^{(2)} = \frac{i}{\sqrt{n_2}}\sum_{j=1}^{n_2}\tau_j\chi_j, \quad (54)$$

where

$$\tau_j = \binom{n_1}{q-1}^{-1/2}i^{q/2-1}\sum_{\substack{S\subseteq[n_1]\\|S|=q-1}}J_{S,j}\,\phi^S. \quad (55)$$

Here ϕ^S the product of $q-1$ of the ϕ Majorana operators in subset S, and $J_{S,j}$ are independent Gaussian

---

[1]We denote Hamiltonians from the class of 2-colored SYK Hamiltonians by $H^{(2)}$.

random variables. The subset S labels an ordered subset of $q-1$ Majorana operators (note that these are different from the subsets I defined before that correspond to ordered subsets of q Majorana operators). We note that the (Hermitian) τ_j operators do not necessarily obey $\{\tau_j, \tau_k\} = 2\delta_{jk}\mathbb{I}$, but instead satisfy $\mathbb{E}(\{\tau_j, \tau_k\}) = -i^{q-2}\delta_{jk}\mathbb{I}$.

Lemma 26. *Let $\{\phi_i\}_{i=1}^{n_1}$ and $\{\chi_i\}_{i=1}^{n_2}$ be $n_1 + n_2$ Majorana operators. For the class of q-local 2-colored SYK Hamiltonians (with even $q \geq 4$) in Eq. (54) defined in terms of these Majorana operators, the maximum eigenvalue of the Hamiltonian $\lambda_{\max}(H)$ is lower bounded by $C\sqrt{n}$ (with C a constant) with probability at least $1 - exp(-\Omega(n))$ over the draw of Hamiltonians.*

Proof. We introduce a new set of Majorana operators (again of size n_2) $\sigma_1, \ldots, \sigma_{n_2}$ (which do obey $\{\sigma_j, \sigma_k\} = 2\delta_{jk}\mathbb{I}$) and we define the *quadratic* Hamiltonian H':

$$H' = \frac{i}{\sqrt{n_2}} \sum_{j=1}^{n_2} \sigma_j \chi_j. \qquad (56)$$

This quadratic Hamiltonian H' is optimized by the fermionic Gaussian state $\rho_0 = \frac{1}{2^{n_2 + n_1/2}} \prod_{j=1}^{n_2} (\mathbb{I} + i\sigma_j\chi_j)$, which achieves $\mathrm{Tr}(H'\rho_0) = \sqrt{n_2}$. The idea is now to construct a new state ρ_θ obtained from ρ_0 by applying a unitary transformation to ρ_0, and to find a lower bound for the expectation value of $H^{(2)}$ w.r.t. ρ_θ.

$$\rho_\theta := e^{-\theta\zeta}\rho_0 e^{+\theta\zeta}, \quad \text{where } \zeta := \sum_{j=1}^{n_2} \tau_j\sigma_j \text{ and } \theta \in \mathbb{R}. \qquad (57)$$

The expectation value of $H^{(2)}$ w.r.t. ρ_θ is:

$$\mathrm{Tr}(H^{(2)}\rho_\theta) = \mathrm{Tr}(H^{(2)}_\theta\rho_0),$$
$$\text{where } H^{(2)}_\theta := e^{+\theta\zeta}H^{(2)}e^{-\theta\zeta}. \qquad (58)$$

Using the BCH expansion of H_θ and $\mathrm{Tr}(H^{(2)}\rho_0) = 0$, we obtain:

$$\begin{aligned}
\mathrm{Tr}(H^{(2)\rho_\theta}) &= \theta\,\mathrm{Tr}([\zeta, H^{(2)}]\rho_0) \\
&+ \theta^2 \int_0^1 (1-s)\mathrm{Tr}([\zeta, [\zeta, H^{(2)}]]\rho_{s\theta})ds \\
&= \theta\,\mathrm{Tr}([\zeta, H^{(2)}]\rho_0) \\
&+ \theta^2 \underset{s\sim[0,1]}{\mathbb{E}}\Big[(1-s)\mathrm{Tr}([\zeta, [\zeta, H^{(2)}]]\rho_{s\theta})\Big] \\
&\geq \theta\,\mathrm{Tr}([\zeta, H^{(2)}]\rho_0) - \theta^2\|[\zeta, [\zeta, H^{(2)}]]\|, \quad (59)
\end{aligned}$$

where we have used the triangle inequality and $\|\cdot\|$ denotes the spectral norm. To lower bound $\mathrm{Tr}(H^{(2)}\rho_\theta)$, one now has to (i) lower bound $\theta\,\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)$ and (ii) upper bound $\theta^2\|[\zeta, [\zeta, H^{(2)}]]\|$. This proof technique is similar in spirit to the proof in [27], although their proof is for qubit Hamiltonians with bounded-degree interactions.

First, we find a lower bound for $\theta\,\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)$ which holds with high probability:

$$\begin{aligned}
\mathrm{Tr}([\zeta, H^{(2)}]\rho_0) &= \frac{i}{\sqrt{n_2}} \sum_{j,k=1}^{n_2} \mathrm{Tr}([\tau_j\sigma_j, \tau_k\chi_k]\rho_0) \\
&= \frac{i}{\sqrt{n_2}} \sum_{j=1}^{n_2} \mathrm{Tr}([\tau_j\sigma_j, \tau_j\chi_j]\rho_0) \\
&= \frac{2i}{\sqrt{n_2}} \sum_{j=1}^{n_2} \mathrm{Tr}(\sigma_j\chi_j\,\tau_j^2\rho_0) \\
&= \frac{2}{\sqrt{n_2}} 2^{-(n_2+n_1/2)} \sum_{j=1}^{n_2} \mathrm{Tr}(\mathbb{I}_{n_2}\,\tau_j^2) \\
&= \frac{2(-1)^{q/2}}{\sqrt{n_2}\binom{n_1}{q-1}} \sum_{j=1}^{n_2} \sum_{\substack{S\subseteq[n_1]\\|S|=q-1}} (J_{S,j})^2, \quad (60)
\end{aligned}$$

where we have used that $\mathrm{Tr}([\tau_j\sigma_j, \tau_k\chi_k]\rho_0)$ is non-zero only for $j = k$, and the definition of τ_j. The quantity $\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)$ is thus a chi-squared random variable (up to normalization factors and potentially a sign) with $n_2\binom{n_1}{q-1}$ degrees of freedom and its expectation value is given by:

$$\begin{aligned}
\mathbb{E}\big[\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)\big] &= \frac{2(-1)^{q/2}}{\sqrt{n_2}\binom{n_1}{q-1}} \sum_{j=1}^{n_2} \sum_{\substack{S\subseteq[n_1]\\|S|=q-1}} \mathbb{E}\big[(J_{S,j})^2\big] \\
&= 2\sqrt{n_2}\,(-1)^{q/2}, \qquad (61)
\end{aligned}$$

where we have used that $\mathbb{E}\big[(J_{S,j})^2\big] = 1$. We note that in order to obtain a positive first-order contribution to $\mathrm{Tr}(H^{(2)}\rho_\theta)$, one should take θ positive for $q/2$ even, and one should take θ negative for $q/2$ odd. Since $\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)$ is a chi-squared random variable with $n_2\binom{n_1}{q-1}$ degrees of freedom, the following tail bounds can be obtained [28]:

$$\Pr\big[\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)\leq\sqrt{n_2}\big] \leq \exp\Big(-\Omega\big(n_2\,n_1^{q-1}\big)\Big), \quad (62)$$

for $q/2$ even, and

$$\Pr\big[\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)\geq-\sqrt{n_2}\big] \leq \exp\Big(-\Omega\big(n_2\,n_1^{q-1}\big)\Big), \qquad (63)$$

for $q/2$ odd. The random variable $\mathrm{Tr}([\zeta, H^{(2)}]\rho_0)$ is thus equal to $2\sqrt{n_2}(-1)^{q/2}$ in expectation and the probability that – for any even $q \geq 4$ – its norm is smaller than half the norm of this expectation is at most exponentially small in the system size.

In order to upper bound $\theta^2\|[\zeta, [\zeta, H^{(2)}]]\|$, we first evaluate $[\zeta, [\zeta, H^{(2)}]]$:

$$[\zeta, [\zeta, H^{(2)}]]$$
$$= \frac{i^{q/2}}{\sqrt{n_2 \binom{n_1}{q-1}}} \sum_{j=1}^{n_2} \sum_{\substack{S \subseteq [n_1] \\ |S| = q-1}} J_{S,j} [\zeta, [\zeta, \phi^S \chi_j]]$$
$$= \frac{i^{q/2}}{\sqrt{n_2 \binom{n_1}{q-1}}} \sum_{j,k,l=1}^{n_2} \sum_{\substack{S \subseteq [n_1] \\ |S| = q-1}} J_{S,j} [\tau_k \sigma_k, [\tau_l \sigma_l, \phi^S \chi_j]]$$
$$= \frac{i^{3q/2-2}}{\sqrt{n_2 \binom{n_1}{q-1}}^3} \sum_{j,k,l=1}^{n_2} \sum_{S,S',S''} J_{S,j} J_{S',k} J_{S'',l}$$
$$\times [\phi^{S'}\sigma_k, [\phi^{S''}\sigma_l, \phi^S \chi_j]], \quad (64)$$

where the final sum over S, S', S'' is over all $S, S', S'' \subseteq [n_1]$ with $|S| = |S'| = |S''| = q - 1$ (all sums over S, S', S'' will implicitly have this constraint from now on). The nested commutator in this expression simplifies as follows (note that the product of $i^{3q/2-2}$ and the nested commutator is Hermitian):

$$i^{3q/2-2}[\phi^{S'}\sigma_k, [\phi^{S''}\sigma_l, \phi^S \chi_j]] =$$

$$\begin{cases} C(\phi^K \sigma_k \sigma_l \chi_j)_H, & \text{if } (|S'' \cap S| \text{ is odd}) \\ & \wedge (|S' \cap (S'' \triangle S)| + \delta_{k,l} \text{ is odd}) \\ & \wedge (|S| = |S'| = |S''| = q - 1), \\ \\ 0, & \text{otherwise,} \end{cases}$$
$$(65)$$

where $(\phi^K \sigma_k \sigma_l \chi_j)_H$ denotes a Hermitian version of $\phi^K \sigma_k \sigma_l \chi_j$ (i.e., $\phi^K \sigma_k \sigma_l \chi_j$ up to potential integer powers of i) and $K := (S \triangle S' \triangle S'') \cup (S \cap S' \cap S'')$ (note that $|K|$ is odd). We therefore have:

$$[\zeta, [\zeta, H^{(2)}]] =$$
$$C \frac{1}{\sqrt{n_2}} \binom{n_1}{q-1}^{-3/2} \sum_{j,k,l=1}^{n_2} \sum_{S,S',S''} J_{S,j} J_{S',k} J_{S'',l}$$
$$\times \Big((\phi^K \sigma_k \sigma_l \chi_j)_H f(S, S', S'', j, k, l) \Big), \quad (66)$$

where we have defined

$$f(S, S', S'', j, k, l) :=$$

$$\begin{cases} 1, & \text{if } (|S'' \cap S| \text{ is odd}) \\ & \wedge (|S' \cap (S'' \triangle S)| + \delta_{k,l} \text{ is odd}) \\ & \wedge (|S| = |S'| = |S''| = q - 1), \\ \\ 0, & \text{otherwise.} \end{cases}$$
$$(67)$$

We now wish to find an upper bound on the expected value of the spectral norm of $[\zeta, [\zeta, H^{(2)}]]$. And in addition, we would like to show that the spectral

norm exceeds twice the value of this upper bound with probability that is at most exponentially small in the system size. To establish this, we will have to show the following:

$$\mathbb{E}\Big(\| [\zeta, [\zeta, H^{(2)}]] \|^k\Big) \leq \alpha^k, \quad (68)$$

for *even* k proportional to the system size and for some α. Eq. (68) implies two things: First, since $\mathbb{E}\big(\| [\zeta, [\zeta, H^{(2)}]] \|\big)^k \leq \mathbb{E}\big(\| [\zeta, [\zeta, H^{(2)}]] \|^k\big)$ (using Jensen's inequality), it implies $\mathbb{E}\big(\|[\zeta, [\zeta, H^{(2)}]]\|\big) \leq \alpha$ (i.e., α is the upper bound on the expected value of the spectral norm). Second, applying Markov's inequality to the random variable $\| [\zeta, [\zeta, H^{(2)}]] \|$ and using Eq. (68) yields

$$\Pr\Big[\| [\zeta, [\zeta, H^{(2)}]] \| \geq \alpha'\Big] = \Pr\Big[\| [\zeta, [\zeta, H^{(2)}]] \|^k \geq (\alpha')^k\Big]$$
$$\leq \big(\alpha/\alpha'\big)^k, \quad (69)$$

with $\alpha' \geq \alpha$. So taking $\alpha' = 2\alpha$ and k equal to the system size $2n$ $(= 2n_2 + n_1)$ yields the desired result of the probability of the spectral norm exceeding twice the value of the upper bound being at most exponentially small in the system size.

For convenience, we define $A := [\zeta, [\zeta, H^{(2)}]]$. Since A is Hermitian (by direct calculation), the spectrum of A^2 is non-negative and therefore we have $\|A\|^k = \lambda_{\max}(A^2)^{k/2} \leq \text{Tr}(A^k)$ (for even k). Using Eq. (66), we express A as $C \sum_{\tilde{S} \subseteq [2n_2 + n_1]} Q_{\tilde{S}} C_{\tilde{S}}$ for convenience, where C is a non-negative constant, $Q_{\tilde{S}}$ are real random variables, and $C_{\tilde{S}}$ denotes a Hermitian (even) Majorana monomial. In addition, we define the random variable (which is obtained by replacing Majorana monomials in A with 1)

$$A(1) := C \sum_{\tilde{S} \subseteq [2n_2 + n_1]} Q_{\tilde{S}}$$
$$= C \frac{1}{\sqrt{n_2}} \binom{n_1}{q-1}^{-3/2} \sum_{j,k,l=1}^{n_2} \sum_{S,S',S''} J_{S,j} J_{S',k} J_{S'',l}$$
$$\times f(S, S', S'', j, k, l). \quad (70)$$

If we now *assume* that

$$\mathbb{E}\big(Q_{\tilde{S}_1} \dots Q_{\tilde{S}_k}\big) \geq 0 \quad \text{and}$$
$$\mathbb{E}\Big(A(1)^k\Big)/\alpha^k \leq 1/2^{n_2 + n_1/2}, \quad (71)$$

both hold for some even k and some constant α (note that the first condition will automatically be satisfied since $\{J_{S,j}\}$ is a collection of independent standard Gaussian random variables), then for even k we can establish

$$
\begin{aligned}
\mathbb{E}\big(\|A\|\big)^k &\leq \mathbb{E}\big(\|A\|^k\big) \\
&\leq \mathbb{E}\big(\mathrm{Tr}(A^k)\big) \\
&= C^k \sum_{\substack{\tilde{S}_1,\ldots,\tilde{S}_k \\ \subseteq [2n_2+n_1]}} \mathbb{E}\big(Q_{\tilde{S}_1}...Q_{\tilde{S}_k}\big)\mathrm{Re}\big[\mathrm{Tr}\big(C_{\tilde{S}_1}...C_{\tilde{S}_k}\big)\big] \\
&\leq 2^{n_2+n_1/2}\, C^k \sum_{\substack{\tilde{S}_1,\ldots,\tilde{S}_k \\ \subseteq [2n_2+n_1]}} \mathbb{E}\big(Q_{\tilde{S}_1}...Q_{\tilde{S}_k}\big) \\
&= 2^{n_2+n_1/2}\, \mathbb{E}\big(A(1)^k\big) \leq \alpha^k, \quad\quad (72)
\end{aligned}
$$

where the first inequality is again Jensen's inequality and we have also used that $\mathbb{E}\big(\mathrm{Tr}(A^k)\big)$ is real (since A is Hermitian) and that $\mathrm{Re}\big[\mathrm{Tr}\big(C_{\tilde{S}_1}...C_{\tilde{S}_k}\big)\big]$ is always at most $2^{n_2+n_1/2}$ (note that $\mathrm{Tr}\big(C_{\tilde{S}_1}...C_{\tilde{S}_k}\big)$ equals $2^{n_2+n_1/2}$ up to integer powers of i, but imaginary contributions vanish in the sum). This establishes Eq. (68), and thereby the desired result. Therefore, what is left is to show that the second condition in Eq. (71) is satisfied.

From this point onward, we shall take n_1 and n_2 proportional to n, where $2n = 2n_2 + n_1$ denotes the total number of Majorana operators. We now show that the second condition in Eq. (71) is satisfied for $k = 2n$ and $\alpha = C\sqrt{n}$. In order to do so, we show that $\mathbb{E}\big(A(1)^{2n}\big) \leq (C\sqrt{n})^{2n}$ (where the factor of $2^{n_2+n_1/2}$ is absorbed in C^{2n}). To that end, we thus need to find an upper bound on the $(2n)$th moment of the random variable $A(1)$ in Eq. (70).

In Appendix F, we derive this upper bound and indeed show that $\mathbb{E}\big(A(1)^{2n}\big) \leq (C\sqrt{n})^{2n}$. Therefore,

$$
\begin{aligned}
&\mathbb{E}\big(\|\,[\zeta,[\zeta,H^{(2)}]]\,\|\big) \leq C\sqrt{n} \quad \text{and} \\
&\Pr\Big[\|\,[\zeta,[\zeta,H^{(2)}]]\,\| \geq 2C\sqrt{n}\Big] \leq \exp\big(-\Omega(n)\big), \quad (73)
\end{aligned}
$$

which is the desired result.

Combining Eq. (59), Eqs. (62),(63) and Eq. (73), we conclude that there exists a $\theta = O(1)$ such that

$$
\mathrm{Tr}(H^{(2)}\rho_\theta) \geq C\sqrt{n}, \quad\quad (74)
$$

with probability at least $1 - \exp\big(-\Omega(n)\big)$. $\quad\square$

What is left is to show that this result also holds for the standard SYK Hamiltonian. This translation from 2-colored SYK Hamiltonian to standard SYK Hamiltonian is given in Lemma 27 below, and its proof is given in Appendix G.

Lemma 27. *For the class of q-local SYK Hamiltonians (with even $q \geq 4$) in Eq. (2), ρ_θ (defined in Eq. (57)) achieves $\mathrm{Tr}(H\rho_\theta) \geq C\sqrt{n}$ with probability at least $1 - exp\big(-\Omega(n)\big)$ over the draw of Hamiltonians, provided that ρ_θ achieves $\mathrm{Tr}(H^{(2)}\rho_\theta) \geq C\sqrt{n}$ (with $H^{(2)}$ the 2-coloured SYK Hamiltonian defined in Eq. (54)) with probability at least $1 - exp\big(-\Omega(n)\big)$ over the draw of 2-coloured Hamiltonians.*

This also concludes the proof of Lemma 10, i.e., that $\lambda_{\max}(H) = \Omega(\sqrt{n})$ with probability at least $1 - \exp\big(-\Omega(n)\big)$ over the draw of standard SYK Hamiltonians.

Acknowledgements

References

[1] Arijit Haldar, Omid Tavakol, and Thomas Scaffidi. Variational wave functions for Sachdev-Ye-Kitaev models. *Phys. Rev. Research*, 3: 023020, 2021. URL https://doi.org/10.1103/PhysRevResearch.3.023020.

[2] Matthew B. Hastings and Ryan O'Donnell. Optimizing strongly interacting fermionic Hamiltonians. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 776–789, New York, NY, USA, 2022. ACM. URL https://doi.org/10.1145/3519935.3519960.

[3] Sevag Gharibian, Yichen Huang, Zeph Landau, and Seung Woo Shin. Quantum Hamiltonian Complexity. *Foundations and Trends in Theoretical Computer Science*, 10(3):159–282, 2015. URL https://doi.org/10.1561%2F0400000066.

[4] Sergey B. Bravyi and Alexei Yu. Kitaev. Fermionic quantum computation. *Annals of Physics*, 298(1):210–226, 2002. URL https://doi.org/10.1006/aphy.2002.6254.

[5] Sanjeev Khanna, Madhu Sudan, Luca Trevisan, and David P Williamson. The approximability of constraint satisfaction problems. *SIAM Journal on Computing*, 30(6):1863–1920, 2001. URL https://doi.org/10.1137/S0097539799349948.

[6] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6): 1115–1145, 1995. URL https://doi.org/10.1145/227683.227684.

[7] Christina V Kraus and J Ignacio Cirac. Generalized Hartree–Fock theory for interacting fermions in lattices: numerical methods. *New Journal of Physics*, 12(11):

Accepted in ⟨Quantum 2023-06-07, click title to verify. Published under CC-BY 4.0.

15

113004, 2010. URL https://doi.org/10.1088/1367-2630/12/11/113004.

[8] Sergey Bravyi and David Gosset. Complexity of quantum impurity problems. *Communications in Mathematical Physics*, 356(2): 451–500, 2017. URL https://doi.org/10.1007/s00220-017-2976-9.

[9] Sergey Bravyi and Robert Koenig. Classical simulation of dissipative fermionic linear optics. *Quant. Inf. Comp.*, 12(11–12):925–943, 2012. URL https://doi.org/10.48550/arXiv.1112.2184.

[10] Fernando de Melo, Piotr Ćwikliński, and Barbara M Terhal. The power of noisy fermionic quantum computation. *New Journal of Physics*, 15(1):013015, 2013. URL https://doi.org/10.1088/1367-2630/15/1/013015.

[11] Elliott H. Lieb. The classical limit of quantum spin systems. *Communications in Mathematical Physics*, 31(4):327 – 340, 1973. URL https://doi.org/10.1007/BF01646493.

[12] Sergey Bravyi, David Gosset, Robert König, and Kristan Temme. Approximation algorithms for quantum many-body problems. *Journal of Mathematical Physics*, 60(3):032203, 2019. URL https://doi.org/10.1063/1.5085428.

[13] Nikhil Bansal, Sergey Bravyi, and Barbara M Terhal. Classical approximation schemes for the ground-state energy of quantum and classical Ising spin Hamiltonians on planar graphs. *Quantum Inf. Comp.*, 9(7-8):701–720, 2009. URL https://doi.org/10.48550/arXiv.0705.1115.

[14] Sevag Gharibian and Julia Kempe. Approximation algorithms for QMA-complete problems. *SIAM Journal on Computing*, 41(4):1028–1050, 2012. URL https://doi.org/10.1137/110842272.

[15] Fernando G.S.L. Brandao and Aram W. Harrow. Product-state approximations to quantum ground states. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 871–880, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450320290. URL https://doi.org/10.1145/2488608.2488719.

[16] Aram W. Harrow and Ashley Montanaro. Extremal eigenvalues of local Hamiltonians. *Quantum*, 1:6, April 2017. URL https://doi.org/10.22331/q-2017-04-25-6.

[17] Thiago Bergamaschi. Improved product-state approximation algorithms for quantum local Hamiltonians, 2022. URL https://doi.org/10.48550/arXiv.2210.08680.

[18] Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of max-

cut. 2019. URL https://doi.org/10.4230/LIPICS.APPROX-RANDOM.2019.31.

[19] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *Inventiones mathematicae*, 163(3): 499–522, 2006. URL https://doi.org/10.1007/s00222-005-0465-9.

[20] Shenglong Xu, Leonard Susskind, Yuan Su, and Brian Swingle. A sparse model of quantum holography, 2020. URL https://doi.org/10.48550/arXiv.2008.02303.

[21] Antonio M. García-García, Yiyang Jia, Dario Rosa, and Jacobus J. M. Verbaarschot. Sparse Sachdev-Ye-Kitaev model, quantum chaos, and gravity duals. *Phys. Rev. D*, 103:106002, 2021. URL https://doi.org/10.1103/PhysRevD.103.106002.

[22] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016. URL https://doi.org/10.1017/CBO9781316339831.

[23] Dorit Aharonov, Itai Arad, and Thomas Vidick. The quantum PCP conjecture. *ACM SIGACT News*, 44:47–79, 2013. URL https://doi.org/10.48550/arXiv.1309.7495.

[24] Sergey Bravyi, Anirban Chowdhury, David Gosset, and Pawel Wocjan. On the complexity of quantum partition functions, 2021. URL https://doi.org/10.48550/arXiv.2110.15466.

[25] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. *The Annals of Probability*, 31(3): 1583–1614, 2003. URL https://doi.org/10.1214/aop/1055425791.

[26] Ryota Tomioka and Taiji Suzuki. Spectral norm of random tensors. *arXiv*, 2014. URL https://doi.org/10.48550/arXiv.1407.1870.

[27] Anurag Anshu, David Gosset, Karen J. Morenz Korol, and Mehdi Soleimanifar. Improved approximation algorithms for bounded-degree local Hamiltonians. *Phys. Rev. Lett.*, 127: 250502, 2021. URL https://doi.org/10.1103/PhysRevLett.127.250502.

[28] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302 – 1338, 2000. URL https://doi.org/10.1214/aos/1015957395.

[29] Béla Bollobás. *Modern Graph Theory*. Graduate Texts in Mathematics 184. Springer-Verlag New York, 1998. URL https://doi.org/10.1007/978-1-4612-0619-4.

[30] Gabriel Andrew Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952. URL https://doi.org/10.1112/plms/s3-2.1.69.

[31] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas,*

Accepted in ⟨ quantum 2023-06-07, click title to verify. Published under CC-BY 4.0.

16

Graphs, and Mathematical Tables. Dover, New York, 1964. ISBN 0-486-61272-4.

[32] Rafal Latala. Estimates of moments and tails of Gaussian chaoses. *The Annals of Probability*, 34 (6), nov 2006. URL https://doi.org/10.1214%2F009117906000000421.

[33] V. H. de la Pena, S. J. Montgomery-Smith, and Jerzy Szulga. Contraction and decoupling inequalities for multilinear forms and u-statistics. *The Annals of Probability*, 22(4):1745–

1765, 1994. URL https://doi.org/10.48550/arXiv.math/9406214.

[34] Radosław Adamczak and Paweł Wolff. Concentration inequalities for non-lipschitz functions with bounded derivatives of higher order. *Probability Theory and Related Fields*, 162(3):531–586, 2015. URL https://doi.org/10.1007/s00440-014-0579-3.

A Extensive sets of all anti-commuting terms

One can easily prove that when one maps a dense, non-sparse, fermionic model such as the SYK model onto a qubit Hamiltonian, the locality of the resulting Hamiltonian has to grow as some function of n, due to the following Lemma:

Lemma 28. *Any set of all-mutually anti-commuting Pauli strings $\{Q_i\}_{i=0}^{m-1}$, each of weight at most k, on n qubits has cardinality m bounded as*

$$m \leq 3 \times 2^{k(3k-1)}, \tag{75}$$

assuming that $k(k-1) < n$.

Proof. Take Q_0 of weight at most k and let $m-1$ Paulis Q_i anticommute with it. We can represent each Pauli string as a $2n$-bit string y, say $Q_0 = y_x y_z$ where the Hamming weight $|y_x| \leq k, |y_z| \leq k$. Any other Q_i in the set has to anti-commute with Q_0 on the support of the string y. First, note that the set of strings of length at most $2k$ which have symplectic inner product equal to 1 (so anti-commute) to a given string of length $2k$ is at most 2^{2k-1}. Now we pick the largest subset $\mathcal{M}_1$ of the set of elements $Q_1, \ldots Q_{m-1}$ such that all elements in the subset act identically on the support of Q_0, i.e. are represented by the same string of length at most $2k$ while differing beyond the support of Q_0. Let the cardinality of this set be $|\mathcal{M}_1| = m_1 \leq m-1$ and $m_1 \geq \frac{m-1}{2^{2k-1}} \geq \frac{m}{2^{2k}}$ as the largest set should at least be a fraction $1/2^{2k-1}$ of the total. So now we consider this set $\mathcal{M}_1$ and their action on the remaining $n-k$ qubits (outside the support of Q_0), where these elements all have to anti-commute. In addition, each element has Pauli weight at most $k-1$ (as we had to overlap with at least one Pauli with Q_0). We then reapply this argument on this set, leading to a new set $\mathcal{M}_2$ with $|\mathcal{M}_2| = m_2 \geq \frac{m_1-1}{2^{2(k-1)-1}}$ acting on $n-2k$ qubits and having weight $k-2$ etc. We can reiterate this process l times so that the remaining weight of the set of Pauli strings $\mathcal{M}_l$ has $k-l = 1$. This implies that $\mathcal{M}_l$ can contain at most 3 elements since they all need to anti-commute on a single qubit (assuming that $n-kl > 0$ or $n-k(k-1) > 0$). So we have

$$3 \geq |\mathcal{M}_{l=k-1}| = m_{k-1} \geq \frac{m}{4^{k+(k-1)+\ldots+l}} = \frac{m}{2^{k(3k-1)}}. \tag{76}$$

$\square$

The SYK-4 model contains large (of size n) sets of mutually anti-commuting terms. An example is the set of all terms which only overlap on one fixed Majorana. Lemma 28 then shows that any fermion-to-qubit mapping (an encoding possibly using more qubits) will require the weight of some of the resulting Pauli terms to grow as a function of n. Note that the actual mapping by Bravyi and Kitaev [4] with $k = O(\log n)$ shows that the upper bound in Eq. (75) is not completely tight.

Another straightforward observation on the energy scaling of a model where all terms anti-commute is that $\lambda_{\max}$ does not necessarily scale with the number of terms, as captured by the following Lemma

Lemma 29. *Let $H = \sum_{i \in \mathcal{I}} J_I C_I$ where the $\{C_I\}$ are a set of all-mutually anti-commuting Majorana operators on $[2n]$ (each C_I has even support). Then*

$$\lambda_{\max}(H) = \sqrt{\sum_I J_I^2}. \tag{77}$$

Proof. We have $H = \sum_I J_I C_I = \sqrt{\sum_I J_I^2} \sum_I \beta_I C_I$ with $\sum_I \beta_I^2 = 1$. Take the state $\rho = \frac{1}{2^n}(\mathbb{I} + \sum_I \beta_I C_I)$ and thus $\text{Tr}(H\rho) = \sqrt{\sum_I J_I^2} \sum_I \beta_I^2 = \sqrt{\sum_I J_I^2}$. This is the maximal eigenvalue that can be reached since one can map each c_I onto a single Majorana operator $c_{i(I)}$ as these sets form identical algebras. Then we can use

the normalization of β_I to view $\sum_i \beta_I c_{i(I)} = \tilde{c}_1$ with single Majorana operator $\tilde{c}_1$ (this is an example of the transformation in Eq. (9)). A single Majorana $\tilde{c}_1$ has spectrum ± 1 and hence the (hugely degenerate) spectrum of H is simply $\pm\sqrt{\sum_I J_I^2}$. $\qquad\square$

Thus, if all J_I are of similar strength, we observe that the overall maximal energy scales as $\sqrt{|\mathcal{I}|}$ rather than $|\mathcal{I}|$.

B Splitting sparse Hamiltonians into diffuse interaction sets

Lemma (Repetition of Lemma 19). *Let $\mathcal{I}$ be the interaction set of a k-sparse q-local Hamiltonian on the set of fermions $[2n]$. The set $\mathcal{I}$ can be split into $(qQ)/2$ disjoint, strictly $2q'$-local subsets $\mathcal{I}_\alpha^{(2q')}$ (with $\alpha \in [Q]$ and $q' \in [q/2]$) each of which is diffuse with respect to $\mathcal{I}$:*

$$\mathcal{I} = \bigcup_{q'=1}^{q/2} \bigcup_{\alpha=1}^{Q} \mathcal{I}_\alpha^{(2q')}. \tag{78}$$

The parameter $Q = q(q-1)(k-1)^2 + q(k-1) + 2$ does not grow with n. The construction of this splitting can be done efficiently, in time $\mathrm{poly}(n)$.

Proof. Consider a graph $\mathcal{G}$ with vertices corresponding to interaction sets $I \in \mathcal{I}$, where two interaction sets I_1, I_2 are connected with an edge if either 1. they share at least one Majorana operator or 2. I_1 and I_2 both share Majorana operators with another set $I' \neq I_1, I_2$. For a q-local k-sparse Hamiltonian, $\mathcal{G}$ has maximal degree Q' with $Q' = q(q-1)(k-1)^2 + q(k-1)$. Here $q(k-1)$ is the maximal number of interactions I_2 directly sharing a Majorana fermion with any given interaction I_1, and $q(q-1)(k-1)^2$ is the maximal number of interactions satisfying condition 2. Since a Q'-sparse graph is vertex-colorable by at most $(Q'+1)$ colors [29], we can split $\mathcal{I}$ into $(Q'+1)$ subsets $\mathcal{I}_\alpha$, s.t. any two interactions I_1, I_2 from a set $\mathcal{I}_\alpha$ are not connected by an edge in $\mathcal{G}$. By definition of $\mathcal{G}$, this amounts to sets $\mathcal{I}_\alpha$ satisfying the first two conditions of Definition 15. A greedy algorithm can be used to assign the vertices $\mathcal{G}$ with $(Q'+1)$ colors, so $\mathcal{I}_\alpha$ can be constructed efficiently.

Each interaction set $\mathcal{I}_\alpha$ can contain terms of different weight. For each value of α we define strictly $2q'$-local sets $\mathcal{I}_\alpha^{(2q')}$ (for $q' = 1, .., q/2$) by restricting to the strictly $2q'$-local part of $\mathcal{I}_\alpha$. This gives a splitting of $\mathcal{I}$ into efficiently constructable subsets $\mathcal{I}_\alpha^{(2q')}$:

$$\mathcal{I} = \bigcup_{q'=1}^{q/2} \bigcup_{\alpha=1}^{Q'+1} \mathcal{I}_\alpha^{(2q')}, \tag{79}$$

where all sets $\mathcal{I}_\alpha^{(2q')}$ satisfy conditions 1 and 2 in Definition 15.

The rest of the proof is concerned with the third condition of a diffuse set in Definition 15, for all sets $\mathcal{I}_\alpha^{(2q')}$. This means ensuring that for all values of α and q', the support size $|\mathrm{Sup}(\mathcal{I}_\alpha^{(2q')})|$ is smaller than $2n\frac{q}{q+1}$. Fix q' and consider sets $\mathcal{I}_\alpha^{(2q')}$ for $\forall \alpha \in [Q'+1]$.

Consider the case where $|\mathrm{Sup}(\mathcal{I}_\alpha^{(2q')})| < 2n\frac{q}{q+1}$ does not hold for at least one value of α, which we set to be $\alpha = Q'+1$ without loss of generality.

Let us prove that the violation $|\mathrm{Sup}(\mathcal{I}_\beta^{(2q')})| \geq 2n\frac{q}{q+1}$ cannot hold for any $\beta \neq Q'+1$. Firstly, no interaction I from $\mathcal{I}_\beta^{(2q')}$ can be a strict subset of an interaction in $\mathcal{I}_{Q'+1}^{(2q')}$ or share Majoranas with two terms in $\mathcal{I}_{Q'+1}^{(2q')}$ simultaneously. The first scenario is excluded since $\mathcal{I}_{Q'+1}^{(2q')}$ and $\mathcal{I}_\beta^{(2q')}$ are both strictly $2q'$-local and the second scenario is excluded because $\mathcal{I}_{Q'+1}^{(2q')}$ satisfies condition 2 of Definition 15. From these two facts it follows that each interaction in $\mathcal{I}_\beta^{(2q')}$ must involve at least one Majorana from $[2n]\backslash\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})$. This implies

$$|\mathrm{Sup}(\mathcal{I}_\beta^{(2q')})| \leq 2q'|[2n]\backslash\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})|, \tag{80}$$

This can be further bounded as $|\mathrm{Sup}(\mathcal{I}_\beta^{(2q')})| \leq q|[2n]\backslash\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})|$, because $2q' \leq q$. Since we assumed $|\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})| \geq 2nq/(q+1)$ and thus $|[2n]\backslash\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})| < 2n/(q+1)$, it follows that $|\mathrm{Sup}(\mathcal{I}_\beta^{(2q')})| \leq q|[2n]\backslash\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})| < 2nq/(q+1)$. Thus we have shown that for a given q', the condition 3 of Definition 15 – indeed cannot be violated by more than one $\mathcal{I}_\alpha^{(2q')}$.

Consider all q' for which there exists a violation $|\mathrm{Sup}(\mathcal{I}_{Q'+1}^{(2q')})| \geq 2n\frac{q}{q+1}$. Since $\frac{q}{q+1} > \frac{1}{2}$ for any q, this violation can be fixed by splitting $\mathcal{I}_{Q'+1}^{(2q')}$ in half. Introduce non-overlapping sets $\tilde{\mathcal{I}}_{Q'+1}^{(2q')}$ and $\tilde{\tilde{\mathcal{I}}}_{Q'+1}^{(2q')}$ of sizes $\lfloor |\mathcal{I}_{Q'+1}^{(2q')}|/2 \rfloor$
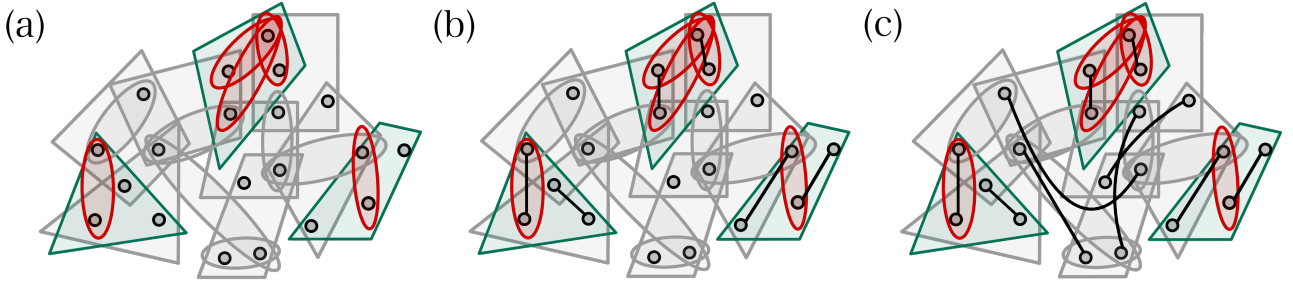
Figure 3: Example of the construction from the proof of Lemma 20. (a) The q-local set of interactions $\mathcal{I}$ ($q = 4$). Highlighted in green is diffuse and strictly q'-local $\mathcal{I}'$ ($q' = 4$), in red are the interactions in $\mathrm{Sup}(\mathcal{I}')$ with weight less than q', in grey are the rest of interactions in $\mathcal{I}$. The goal is to create a matching M consistent with green-colored terms, inconsistent with grey-colored terms, and with no guaranteed relation to the red-colored terms. (b) Matching M' on $\mathrm{Sup}(\mathcal{I}')$, consistent with $\mathcal{I}'$ by construction. Ensuring inconsistency with all red-colored terms is in general impossible. For example, consider the three overlapping red-colored terms at the top center. (c) Completing $M = M' \cup M''$ with a matching M'' on $[2n]\backslash\mathrm{Sup}(\mathcal{I}')$, ensuring inconsistency with all grey-colored terms. For this, the vertices are matched only if they belong to different interactions.

and $\lceil|\mathcal{I}_{Q'+1}^{(2q')}|/2\rceil$: $\mathcal{I}_{Q'+1}^{(2q')} = \tilde{\mathcal{I}}_{Q'+1}^{(2q')} \cup \tilde{\tilde{\mathcal{I}}}_{Q'+1}^{(2q')}$. By implication, $|\mathrm{Sup}(\tilde{\mathcal{I}}_{Q'+1}^{(2q')})| \leq 2n/2 \leq 2nq/(q+1)$ and similarly $|\mathrm{Sup}(\tilde{\tilde{\mathcal{I}}}_{Q'+1}^{(2q')})| \leq 2nq/(q+1)$. We conclude the construction by modifying the set $\mathcal{I}_\alpha^{(2q')}$ for the considered q': we redefine $\mathcal{I}_{Q'+1}^{(2q')} \equiv \tilde{\mathcal{I}}_{Q'+1}^{(2q')}$, and introduce one extra interaction set $\mathcal{I}_{Q'+2}^{(2q')} \equiv \tilde{\tilde{\mathcal{I}}}_{Q'+1}^{(2q')}$.

The proof can now be finalized. Performing the above procedure for all q' where a violation was present, and completing the $\{\mathcal{I}_\alpha^{(2q')}\}$ without such violations with $\mathcal{I}_{\alpha=Q'+2}^{(2q')} = \emptyset$, we arrive at the splitting

$$\mathcal{I} = \bigcup_{q'=1}^{q/2} \bigcup_{\alpha=1}^{Q} \mathcal{I}_\alpha^{(2q')}, \tag{81}$$

where $Q = Q' + 2 = q(q-1)(k-1)^2 + q(k-1) + 2$. Interaction sets $\mathcal{I}_\alpha^{(2q')}$ are diffuse (satisfying all three conditions of Definition 15) with respect to $\mathcal{I}$ for all q' and α. The construction of $\mathcal{I}_\alpha^{(2q')}$ is efficient, because each step can be implemented in time poly(n). $\qquad\square$

C Majorana matchings from diffuse interaction sets

Lemma (Repetition of Lemma 20). *Let a strictly q'-local $\mathcal{I}'$ be diffuse w.r.t. q-local k-sparse $\mathcal{I}$ on $[2n]$, such that $n > (q^2-1)k$. One can efficiently construct a matching M of $[2n]$ that is consistent with $\mathcal{I}'$ and inconsistent with all interactions $I \in \mathcal{I}\backslash\mathcal{I}'$ such that (1) $|I| \geq q'$ or (2) $I \not\subset \mathrm{Sup}(\mathcal{I}')$.*

Proof. We first note that for $I \in \mathcal{I}\backslash\mathcal{I}'$ the condition $|I| \geq q'$ implies $I \not\subset \mathrm{Sup}(\mathcal{I}')$. Indeed, there are two possible options for $I \in \mathcal{I}\backslash\mathcal{I}'$ such that $I \subset \mathrm{Sup}(\mathcal{I}')$. The first option is that I is a strict subset of a single interaction from $\mathcal{I}'$. However, this is not possible given $|I| \geq q'$, because $\mathcal{I}'$ is q'-local. The second option is for I to share Majorana modes with two or more interactions in $\mathcal{I}'$. This is ruled out because $\mathcal{I}'$ is diffuse with respect to $\mathcal{I}$ (cf. Condition 2 in Definition 15). The above implies that it is sufficient to construct the matching M that is consistent with $\mathcal{I}'$ and inconsistent with $\{I \in \mathcal{I}\backslash\mathcal{I}' | I \not\subset \mathrm{Sup}(\mathcal{I}')\}$.

We construct M in two steps. First we construct a matching M' of $\mathrm{Sup}(\mathcal{I}')$ (note $|\mathrm{Sup}(\mathcal{I}')|$ is always even). Next, we construct a matching M'' of the remaining Majorana modes $[2n]\backslash\mathrm{Sup}(\mathcal{I}')$. The desired matching of $[2n]$ is the union $M = M' \cup M''$.

To construct M', we match vertices of each $I \in \mathcal{I}'$ in an arbitrary way: for every such $I = \{i_1, .. i_{q'}\}$, $\{i_{2l-1}, i_{2l}\} \in M'$ for $l \in [1, .. q'/2]$. This matching is always possible, since $\mathcal{I}'$ is diffuse and thus different interactions from $\mathcal{I}'$ do not overlap. Thus constructed M' (and therefore also $M = M' \cup M''$) is explicitly consistent with all $I \in \mathcal{I}'$.

To construct a matching M'' of $[2n]\backslash\mathrm{Sup}(\mathcal{I}')$, we aim to ensure that no $(m_1, m_2) \in M''$ is a subset of any interaction in $\mathcal{I}$. For this, consider a 'permitted edge' graph $\mathcal{P}$ with vertices $[2n]\backslash\mathrm{Sup}(\mathcal{I}')$, and edges inserted between every pair (i_1, i_2) *unless* they belong to the same interaction in $\mathcal{I}$. We aim to construct M'' as a perfect matching of $\mathcal{P}$. Note that since $\mathcal{I}$ is q-local and k-sparse, the graph $\mathcal{P}$ has degree bounded from below as

$|[2n]\backslash\text{Sup}(\mathcal{I}')| - (q-1)k$. At the same time, since $\mathcal{I}'$ is diffuse, we're guaranteed by Condition 3 in Definition 15 that $|[2n]\backslash\text{Sup}(\mathcal{I}')| \geq \frac{2n}{q+1}$. Therefore, since $n > (q^2-1)k$ by assumption, the degree of the vertices in $\mathcal{P}$ is lower bounded as $|[2n]\backslash\text{Sup}(\mathcal{I}')| - (q-1)k \geq |[2n]\backslash\text{Sup}(\mathcal{I}')|/2 + \frac{n}{q+1} - (q-1)k > |[2n]\backslash\text{Sup}(\mathcal{I}')|/2$. Given this lower bound, we apply Dirac's theorem [30], which yields an efficiently constructable Hamiltonian cycle in the graph $\mathcal{P}$. Matching M'' is then obtained by pairing the sequential vertices in this cycle, making it a perfect matching of $\mathcal{P}$. By definition of $\mathcal{P}$, M'' is guaranteed to contain at least one outgoing edge from every interaction in $\{I \in \mathcal{I}\backslash\mathcal{I}' | I \not\subset \text{Sup}(\mathcal{I}')\}$. This makes $M = M' \cup M''$ inconsistent with $\{I \in \mathcal{I}\backslash\mathcal{I}' | I \not\subset \text{Sup}(\mathcal{I}')\}$, as desired.

□

Lemma 17, which is used in the proof of Theorem 5, is a special case of Lemma 20. To obtain Lemma 17, one sets $q' = q$ and considers strictly q-local $\mathcal{I}$ instead of simply q-local. In this case all terms in $\mathcal{I}\backslash\mathcal{I}'$ satisfy the first condition of the Lemma, and therefore the constructed M is inconsistent with the entirety of $\mathcal{I}\backslash\mathcal{I}'$.

D Matchings and Gaussian states

Lemma (Repetition of Lemma 21). *Let $H = \sum_{I \in \mathcal{I}} J_I C_I$ on $[2n']$ be q-local and $\mathcal{I}'$ be a diffuse subset of $\mathcal{I}$. Consider a matching M of $[2n']$. If M is consistent with $\mathcal{I}'$ and inconsistent with $\mathcal{I}\backslash\mathcal{I}'$, one can efficiently construct a Gaussian state $\rho_{\mathcal{I}'}$ with the property:*

$$\text{Tr}(H\rho(\mathcal{I}')) = \sum_{I \in \mathcal{I}'} |J_I|. \tag{82}$$

Proof. For the given matching M, consider its associated Gaussian state pure $\rho(M, \vec{\lambda})$ of the form:

$$\rho(M, \vec{\lambda}) = \frac{1}{2^n} \Pi_{\{m_1,m_2\} \in M} (\mathbb{I} + i\lambda_{(m_1,m_2)} c_{m_1} c_{m_2}). \tag{83}$$

Lemma 14 implies that the contribution to $\text{Tr}(H\rho(M, \vec{\lambda}))$ from inconsistent interactions $\mathcal{I}\backslash\mathcal{I}'$ vanishes and contributions from $\mathcal{I}'$ yield:

$$\text{Tr}(H\rho(M, \vec{\lambda})) = \sum_{I \in \mathcal{I}'} J_I \text{sign}(\pi) \prod_{l \in \{1,..,|I|/2\}} \lambda_{(i_{\pi(2l-1)}, i_{\pi(2l)})}. \tag{84}$$

The proof is completed by choosing an appropriate value for $\vec{\lambda}$. Since $\mathcal{I}'$ is diffuse, by Condition 1 of Definition 15, distinct interactions from $\mathcal{I}'$ do not share Majorana fermions. This means that the values $\lambda_{(m_1,m_2)}$ for different I in Eq. (84) can be chosen independently. In particular, by picking appropriate $\lambda_{(m_1,m_2)} = \pm 1$, one can eliminate the sign of $J_I \text{sign}(\pi)$ and achieve a contribution $|J_I|$ for each $I \in \mathcal{I}'$. Note that this procedure can be done efficiently, as it is simply a matter of choosing at most n ± 1 values by checking the sign of most $|\mathcal{I}'|$ terms. Denoting the thus chosen $\rho(M, \vec{\lambda})$ as $\rho(\mathcal{I}')$, this yields Eq. (82).

□

A special case of Lemma 21 is Lemma 18 used in the proof of Theorem 5.

E Concentration bounds for sparse SYK-4

Here we derive the concentration bounds for the SSYK- 4 Hamiltonian that were used in the proof of Theorem 8 (Section 7). We first prove an auxiliary Lemma that will be used later in this Section, allowing to separate the statistics of interaction selection and interaction strength:

Lemma 30. *For $a \in [D]$, let X_a be i.i.d. Bernoulli random variables $X_a \sim \text{Bern}(p)$ and J_a i.i.d. Gaussian random variables $J_a \sim N(0,1)$. Then for any integer $d \in [D]$*

$$\mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| < y\right] \leq \mathbb{P}\left[\sum_{a=1}^{D} X_a < d\right] + \mathbb{P}\left[\sum_{a=1}^{d} |J_a| < y\right], \tag{85}$$

$$\mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| > y\right] \leq \mathbb{P}\left[\sum_{a=1}^{D} X_a > d\right] + \mathbb{P}\left[\sum_{a=1}^{d} |J_a| > y\right]. \tag{86}$$

Proof. To prove Eq. (85), first show

$$\mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| \geq y\right] = \sum_{d'=1}^{D} \left(\mathbb{P}\left[\sum_{a=1}^{D} X_a = d'\right] \mathbb{P}\left[\sum_{a=1}^{d'} |J_a| \geq y\right]\right)$$

$$\geq \sum_{d'=d}^{D} \left(\mathbb{P}\left[\sum_{a=1}^{D} X_a = d'\right] \mathbb{P}\left[\sum_{a=1}^{d'} |J_a| \geq y\right]\right)$$

$$\geq \left(\sum_{d'=d}^{D} \mathbb{P}\left[\sum_{a=1}^{D} X_a = d'\right]\right) \mathbb{P}\left[\sum_{a=1}^{d} |J_a| \geq y\right]$$

$$= \mathbb{P}\left[\sum_{a=1}^{D} X_a \geq d\right] \mathbb{P}\left[\sum_{a=1}^{d} |J_a| \geq y\right]. \tag{87}$$

It follows that

$$\mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| < y\right] = 1 - \mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| \geq y\right]$$

$$\leq 1 - \mathbb{P}\left[\sum_{a=1}^{D} X_a \geq d\right] \mathbb{P}\left[\sum_{a=1}^{d} |J_a| \geq y\right]$$

$$= 1 - \left(1 - \mathbb{P}\left[\sum_{a=1}^{D} X_a < d\right]\right)\left(1 - \mathbb{P}\left[\sum_{a=1}^{d} |J_a| < y\right]\right)$$

$$\leq \mathbb{P}\left[\sum_{a=1}^{D} X_a < d\right] + \mathbb{P}\left[\sum_{a=1}^{d} |J_a| < y\right]. \tag{88}$$

This ends the proof of Eq. (85). In the same vein, one derives Eq. (86). Namely, we first have (cf. Eq. (87)):

$$\mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| \leq y\right] \geq \sum_{d'=1}^{d} \left(\mathbb{P}\left[\sum_{a=1}^{D} X_a = d'\right] \mathbb{P}\left[\sum_{a=1}^{d'} |J_a| \leq y\right]\right)$$

$$\geq \mathbb{P}\left[\sum_{a=1}^{D} X_a \leq d\right] \mathbb{P}\left[\sum_{a=1}^{d} |J_a| \leq y\right]. \tag{89}$$

Similarly to Eq. (88), one obtains Eq. (86) from Eq. (89):

$$\mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| > y\right] = 1 - \mathbb{P}\left[\sum_{a=1}^{D} X_a |J_a| \geq y\right]$$

$$\leq 1 - \left(1 - \mathbb{P}\left[\sum_{a=1}^{D} X_a > d\right]\right)\left(1 - \mathbb{P}\left[\sum_{a=1}^{d} |J_a| > y\right]\right)$$

$$\leq \mathbb{P}\left[\sum_{a=1}^{D} X_a > d\right] + \mathbb{P}\left[\sum_{a=1}^{d} |J_a| > y\right]. \tag{90}$$

$\square$

We proceed with the proof of Lemmas 23 and 24, which were used in Section 7 to prove Theorem 8.

Lemma (Repetition of Lemma 23). *Let interactions $\mathcal{I}$ and interaction strengths $\{J_I\}$ be those of the SSYK-4 model with average degree k. With probability at least $1 - 2e^{-\frac{kn}{32}}$ we have*

$$\sum_{I \in \mathcal{I}} |J_I| \geq kn/8. \tag{91}$$

Proof. The random variable $\sum_{I \in \mathcal{I}} |J_I|$ is a function of two sets of random variables. The first set is $X_I \in \{0, 1\}$ for all possible 4-Majorana interactions $I \subset [2n]$, $|I| = 4$, indicating the presence of I in $\mathcal{I}$. Denoting

$$D \equiv \binom{2n-1}{3}, \tag{92}$$

X_I is drawn from a Bernoulli distribution with probability $p = kD^{-1}$, i.e. $X_I \sim \text{Bern}(kD^{-1})$. The second set is J_I for all $I \in \mathcal{I}$, distributed normally $J_I \sim \text{N}(0, 1)$. We introduce auxiliary variables J_a for $a \in [\lceil kn/4 \rceil]$ and $J_a \sim \text{N}(0, 1)$. Then by Lemma 30:

$$\mathbb{P}\left[\sum_{I \in \mathcal{I}} |J_I| < kn/8\right] \leq \mathbb{P}\left[\sum_{I \subset [2n], \, |I|=4} X_I < \left\lceil \frac{kn}{4} \right\rceil\right] + \mathbb{P}\left[\sum_{a \in [\lceil kn/4 \rceil]} |J_a| < kn/8\right]. \tag{93}$$

We can bound the first term using the Chernoff bound for sums of Bernoulli random variables. Substituting $\binom{2n}{4} = \binom{2n-1}{3}\frac{n}{2}$, we get

$$\mathbb{P}\left[\sum_{I \subset [2n], \, |I|=4} X_I < \left\lceil \frac{kn}{4} \right\rceil\right] = \mathbb{P}\left[\sum_{I \subset [2n], \, |I|=4} X_I < \frac{kn}{4}\right] \leq \exp\left(-\frac{kn}{4}(1 - \log 2)\right). \tag{94}$$

On the other hand, standard concentration properties of Gaussian random variables imply, see Lemma 31 at the end of this Appendix,

$$\mathbb{P}\left[\sum_{a=1}^{\lceil kn/4 \rceil} |J_a| < \frac{kn}{8}\right] < e^{-kn/32}. \tag{95}$$

Since $\exp\left(-\frac{kn}{4}(1 - \log 2)\right) \leq e^{-kn/32}$, the bound in Eq. (93) yields

$$\mathbb{P}\left[\sum_{I \in \mathcal{I}} |J_I| < kn/8\right] < 2e^{-kn/32}, \tag{96}$$

as desired.

$\square$

Lemma (Repetition of Lemma 24). *If $k' \geq e^2 k + 1$, we have with probability at least $1 - 2\exp\left[-\frac{e^{-2k'}k^3}{64(k'-1)}n\right]$ that*

$$\sum_{I \in \bar{\mathcal{I}}^{(k')}} |J_I| \leq \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n. \tag{97}$$

Proof. The random variable $\sum_{I \in \bar{\mathcal{I}}^{(k')}} |J_I|$ is a function of random variables $X_I \sim \text{Bern}(kD^{-1})$ for $I \subset [2n]$, $|I| = 4$ and $J_I \sim \text{N}(0, 1)$ for all $I \in \mathcal{I}$. We introduce auxiliary random variables $J_a' \sim \text{N}(0, 1)$ for $a \in [K]$ where

$$K \equiv \lfloor \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n \rfloor. \tag{98}$$

By Lemma 30, one can upperbound

$$\mathbb{P}\left[\sum_{I \in \bar{\mathcal{I}}^{(k')}} |J_I| > \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n\right] \leq \mathbb{P}\left[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n\right] + \mathbb{P}\left[\sum_{a \in [K]} |J_a'| > \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n\right]. \tag{99}$$

We now proceed with upper bounding $\mathbb{P}\left[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n\right]$ and then $\mathbb{P}\left[\sum_{a \in [K]} |J_a'| > \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n\right]$.

To bound $\mathbb{P}\left[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}} e^{-k'} n\right]$, we introduce the Majorana degree function $k_i = k_i(\{X_I\}) \in [D]$, which is a random variable that counts the number of interactions in $\mathcal{I}$ involving a given Majorana c_i. Since $X_I \sim \text{Bern}(kD^{-1})$, k_i follows the binomial distribution $\text{Bin}(D, kD^{-1})$ (note however that different k_i and k_j are not necessarily independent). Given the construction of $h^{(k')}$, it is clear that $|\bar{\mathcal{I}}^{(k')}|$ can be bounded by the 'excess degree' summed over all Majoranas. Concretely, using the Majorana degree function k_i we define a random variable

$$Z \equiv Z(\{X_I\}) \equiv \frac{1}{2n} \sum_{i=1}^{2n} (k_i - k') \mathbb{I}_{k_i > k'}, \tag{100}$$

which has the immediate property

$$|\bar{\mathcal{I}}^{(k')}| \le 2nZ. \tag{101}$$

Here we used the indicator function $\mathbb{I}_{k_i > k'} = 1$ when $k_i > k'$ and 0 otherwise. Given Eq. (101), $\mathbb{P}[Z > \frac{2k^2}{\sqrt{k'-1}}e^{-k'}] \ge \mathbb{P}[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n]$ and thus it suffices to bound the former. We begin by calculating its mean:

$$\mathbb{E}[Z] = \frac{1}{2n} \sum_{i=1}^{2n} \mathbb{E}[(k_i - k')\mathbb{I}_{k_i > k'}] = \mathbb{E}[(k_1 - k')\mathbb{I}_{k_1 > k'}], \tag{102}$$

where we used linearity of $\mathbb{E}(.)$ and the permutation symmetry of the SSYK ensemble. Hence we now need to calculate $\mathbb{E}[(k_1 - k')\mathbb{I}_{k_1 > k'}]$ for a single Majorana (w.l.o.g. c_1). Since the associated degree $k_1 \sim \text{Bin}(D, kD^{-1})$, we calculate directly (denoting $p = kD^{-1}$):

$$\mathbb{E}[(k_1 - k')\mathbb{I}_{k_1 > k'}] \le \mathbb{E}[k_1 \mathbb{I}_{k_1 > k'}] = \sum_{k_1 = k'+1}^{D} p^{k_1}(1-p)^{D-k_1} \frac{D!}{(D-k_1)!k_1!} k_1$$

$$= Dp \sum_{k_1 = k'}^{D-1} p^{k_1}(1-p)^{D-1-k_1} \frac{(D-1)!}{(D-1-k_1)!k_1!}. \tag{103}$$

The following identity holds [31]:

$$\sum_{x=y}^{z-1} w^x (1-w)^{z-1-x} \frac{(z-1)!}{(z-1-x)!x!} = \beta_w(y, z-y), \tag{104}$$

where $\beta_w(y, z-y)$ is the regularized incomplete beta function. For integer $y, z > y$ it is defined as

$$\beta_w(y, z-y) = \frac{(z-1)!}{(y-1)!(z-y-1)!} \int_{t=0}^{w} t^{y-1}(1-t)^{z-y-1} dt.$$

Using the Stirling bound $x! \ge \sqrt{2\pi} x^{x+1/2} e^{-x}, x \in \mathbb{N}$, one bounds $\beta_w(y, z-y)$ as:

$$\beta_w(y, z-y) < \frac{w(z-1)}{\sqrt{2\pi(y-1)}} \left(e\frac{w(z-1)}{(y-1)} \right)^{y-1}. \tag{105}$$

Substituting $p = kD^{-1}$ and using Eqs. (104), (105) in Eq. (103) for $k' > e^2 k + 1$ we obtain

$$\mathbb{E}[(k_1 - k')\mathbb{I}_{k_i > k'}] < Dp \frac{p(D-1)}{\sqrt{2\pi(k'-1)}} \left(e\frac{p(D-1)}{k'-1} \right)^{k'-1}$$

$$< \sqrt{\frac{e^2 k^4}{2\pi(k'-1)}} e^{-k'}$$

$$\Rightarrow \mathbb{E}[Z] < \sqrt{\frac{e^2 k^4}{2\pi(k'-1)}} e^{-k'}. \tag{106}$$

We now aim to apply the Efron-Stein inequality [25] to bound deviations from the mean $\mathbb{E}(Z)$. For this, we introduce an additional set of independent random variables $\{X_I'\}$ such that $X_I' \sim \text{Bern}\left(kD^{-1}\right)$. This allows to define auxiliary functions Z_I'

$$Z_I' \equiv Z\big|_{X_I \to X_I'} \tag{107}$$

where for a single interaction I only, the variable X_I is replaced by X_I'. Using the indicator function $\mathbb{I}_{Z > Z_I'}$, a further auxiliary function $V = V(\{X_I\})$ can be defined:

$$V \equiv \mathbb{E}_{\{X_I'\}} \left[\sum_{I \subset [2n], \, |I|=4} (Z - Z_I')^2 \mathbb{I}_{Z > Z_I'} \right], \tag{108}$$

where the averaging is performed over the additional random variables $\{X'_I\}$ alone. An exponential version of the Efron-Stein inequality (Theorem 2 of [25]) states for all $\theta > 0$ and $\lambda \in (0, \theta^{-1})$:

$$\log \mathbb{E}[\exp(\lambda(Z - \mathbb{E}[Z]))] \leq \frac{\lambda\theta}{1 - \lambda\theta} \log \mathbb{E}\left[\exp\left(\frac{\lambda V}{\theta}\right)\right]. \tag{109}$$

To employ Eq. (109), we have to bound $\mathbb{E}\left[\exp\left(\frac{\lambda V}{\theta}\right)\right]$. First we upper bound $V(\{X_I\})$ as a function. For all interactions I we claim, independent of $\{X_I\}$ and $\{X'_I\}$:

$$(Z - Z'_I)^2 \mathbb{I}_{Z > Z'_I} \leq \frac{4}{n^2} \mathbb{I}_{X_I = 1}. \tag{110}$$

To show this, we will go through four possible cases: $(X_I, X'_I) = (0, 0)$, $(1, 0)$, $(0, 1)$, or $(1, 1)$. If $X_I = X'_I$, the left hand side of Eq. (110) vanishes, reproducing Eq. (110) for the cases $(X_I, X'_I) = (0, 0)$ and $(1, 1)$. For $(X_I, X'_I) = (0, 1)$, Z is smaller than Z'_I, because replacing $X_I = 0$ by $X_I = 1$ cannot decrease the excess degree for any Majorana (cf. definition of k_i and $2nZ = \sum_{i=1}^{2n} (k_i - k')\mathbb{I}_{k_i > k'}$). Due to the factor $\mathbb{I}_{Z > Z'_I}$, $(Z - Z'_I)^2 \mathbb{I}_{Z > Z'_I}$ in this case is zero, in agreement with Eq. (110). The last case is $(X_I, X'_I) = (1, 0)$. As any interaction I only involves 4 fermions, the reduction of total excess degree $2n(Z|_{X_I=1} - Z|_{X_I=0})$ is at most equal to 4, independent of the rest of the variables $\{X_I\}$. Therefore $(Z - Z'_I)^2 \mathbb{I}_{Z > Z'_I}$ for $(X_I, X'_I) = (1, 0)$ is at most equal to $\frac{4}{n^2}$, proving Eq. (110). From Eq. (110) it follows that $\mathbb{E}_{\{X'_I\}}[(Z - Z'_I)^2 \mathbb{I}_{Z > Z'_I}] \leq \frac{4}{n^2} \mathbb{I}_{X_I = 1}$, which we can use to bound $V(\{X_I\})$. From the definition stated in Eq. (108) we get:

$$V(\{X_I\}) \leq \frac{4}{n^2} \sum_{I \subset [2n], \, |I| = 4} X_I. \tag{111}$$

Since $X_I \sim \mathrm{Bin}(1, kD^{-1})$, we have

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(\frac{\lambda V}{\theta}\right)\right] &\leq \mathbb{E}\left[\exp\left(\frac{4\lambda}{\theta n^2} \sum_{I \subset [2n], \, |I| = 4} X_I\right)\right] \\
&= \left(\mathbb{E}\left[\exp\left(\frac{4\lambda}{\theta n^2} X_1\right)\right]\right)^{\binom{2n}{4}} \\
&= \left((1 - kD^{-1}) + kD^{-1} \exp\left(\frac{4\lambda}{\theta n^2}\right)\right)^{\binom{2n}{4}} \\
&\leq \exp\left(k\binom{2n}{4} D^{-1}\left(\exp\left(\frac{4\lambda}{\theta n^2}\right) - 1\right)\right) \\
&= \exp\left(\frac{kn}{2}\left(\exp\left(\frac{4\lambda}{\theta n^2}\right) - 1\right)\right).
\end{aligned}
\tag{112}
$$

We further assume a constraint $\lambda < \frac{n^2 \theta}{4}$, which implies the inequality $\exp(\frac{4\lambda}{\theta n^2}) - 1 < \frac{8\lambda}{\theta n^2}$. This allows to further bound $\mathbb{E}\left[\exp\left(\frac{\lambda V}{\theta}\right)\right]$:

$$\mathbb{E}\left[\exp\left(\frac{\lambda V}{\theta}\right)\right] \leq \exp\left[\frac{4\lambda k}{\theta n}\right]. \tag{113}$$

We now assume an additional constraint $\lambda < \frac{1}{2\theta}$, which strengthens the condition $\lambda < \theta^{-1}$ of Eq. (109). With this constraint, using Eq. (113) in Eq. (109), we obtain:

$$
\begin{aligned}
\log \mathbb{E}[\exp(\lambda(Z - \mathbb{E}[Z]))] &\leq \frac{4\lambda^2}{1 - \lambda\theta} \frac{k}{n} \\
&\leq \frac{8\lambda^2 k}{n}.
\end{aligned}
\tag{114}
$$

This inequality is true regardless of θ and λ, insofar both numbers are positive and satisfy the constraints we introduced:

$$\frac{4\lambda}{n^2} < \theta < \frac{1}{2\lambda}. \tag{115}$$

For a valid θ to exist, it's necessary and sufficient that λ belongs to the interval $(0, \frac{n}{2\sqrt{2}})$. For such λ, Eq. (114) holds, and combined with a Markov inequality it implies for any $t > 0$:

$$\mathbb{P}[Z > \mathbb{E}[Z] + t] < \exp\left[\frac{8\lambda^2 k}{n} - \lambda t\right]. \tag{116}$$

We next choose the value of $\lambda \in (0, \frac{n}{2\sqrt{2}})$ that optimizes the right hand side. If $\frac{t}{2\sqrt{2}k} < 1$, this is achieved with $\lambda = \frac{tn}{16k}$. This yields the result

$$\mathbb{P}[Z > \mathbb{E}[Z] + t] < \exp\left[-\frac{nt^2}{32k}\right]. \tag{117}$$

We choose $t = \sqrt{\frac{k^4}{2(k'-1)}}e^{-k'}$, which automatically ensures the desired condition $\frac{t}{2\sqrt{2}k} < 1$ because of the constraint $k' > e^2k + 1$ that we assumed in the Lemma statement. We obtain:

$$\mathbb{P}\left[Z > \mathbb{E}[Z] + \sqrt{\frac{k^4}{2(k'-1)}}e^{-k'}\right] < \exp\left[-\frac{e^{-2k'}k^3}{64(k'-1)}n\right]. \tag{118}$$

Since $\mathbb{E}[Z] < \sqrt{\frac{e^2k^4}{2\pi(k'-1)}}e^{-k'}$ (Eq. (106)) and $\sqrt{\frac{e^2}{2\pi}} + \sqrt{\frac{1}{2}} < 2$, we arrive at an upper bound for the probability $\mathbb{P}\left[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right]$:

$$\mathbb{P}\left[Z > \frac{2k^2}{\sqrt{k'-1}}e^{-k'}\right] < \exp\left[-\frac{e^{-2k'}k^3}{64(k'-1)}n\right]$$

$$\Rightarrow \mathbb{P}\left[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right] < \exp\left[-\frac{e^{-2k'}k^3}{64(k'-1)}n\right]. \tag{119}$$

To bound $\mathbb{P}\left[\sum_{a\in[K]}|J'_a| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right]$, we use the concentration properties of Gaussian random variables (see Lemma 31 at the end of this Appendix). Using $K = \lfloor\frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\rfloor$ in Lemma 31.1:

$$\mathbb{P}\left[\sum_{a\in[K]}|J'_a| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right] \leq \mathbb{P}\left[\sum_{a\in[K]}|J'_a| > K\right] < e^{-K/20}. \tag{120}$$

Note that our bound for $\mathbb{P}\left[|\bar{\mathcal{I}}^{(k')}| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right]$ in Eq. (119) is always greater than our bound for $\mathbb{P}\left[\sum_{a\in[K]}|J'_a| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right]$ in Eq. (120). This allows us to conclude the proof of the Lemma, as Eqs. (99) and (119) imply:

$$\mathbb{P}\left[\sum_{I\in\bar{\mathcal{I}}^{(k')}}|J_I| > \frac{4k^2}{\sqrt{k'-1}}e^{-k'}n\right] \leq 2\exp\left[-\frac{e^{-2k'}k^3}{64(k'-1)}n\right]. \tag{121}$$

$\square$

Lemma 31. *Let $J_a \sim N(0,1)$ for $a \in [A]$, $A \in \mathbb{N}$. Then*

1. $\mathbb{P}\left[\sum_{a=1}^{A}|J_a| \geq A\right] \leq e^{-A/20}$.

2. $\mathbb{P}\left[\sum_{a=1}^{A}|J_a| \leq \frac{A}{2}\right] \leq e^{-A/8}$.

Proof. For $J \sim N(0,1)$, we have

$$\mathbb{E}\left[\exp\left(\lambda|J|\right)\right] = e^{\frac{\lambda^2}{2}}\left(1 + \text{Erf}\left(\frac{\lambda}{\sqrt{2}}\right)\right). \tag{122}$$

The Chernoff bound then implies:

$$\mathbb{P}\left[\sum_{a=1}^{A}|J_a| \geq A\right] \leq \left(e^{\frac{\lambda^2}{2}-\lambda}\left(1 + \text{Erf}\left(\frac{\lambda}{\sqrt{2}}\right)\right)\right)^A,$$

$$\mathbb{P}\left[\sum_{a=1}^{A}|J_a| \leq A/2\right] \leq \left(e^{\frac{\lambda^2}{2}+\frac{\lambda}{2}}\left(1 - \text{Erf}\left(\frac{\lambda}{\sqrt{2}}\right)\right)\right)^A. \tag{123}$$

Evaluating the two expressions at $\lambda = \frac{1}{2}$ and $\lambda = 1$ respectively and using basic inequalities for the resulting constants, we obtain the two bounds claimed in the Lemma. $\square$
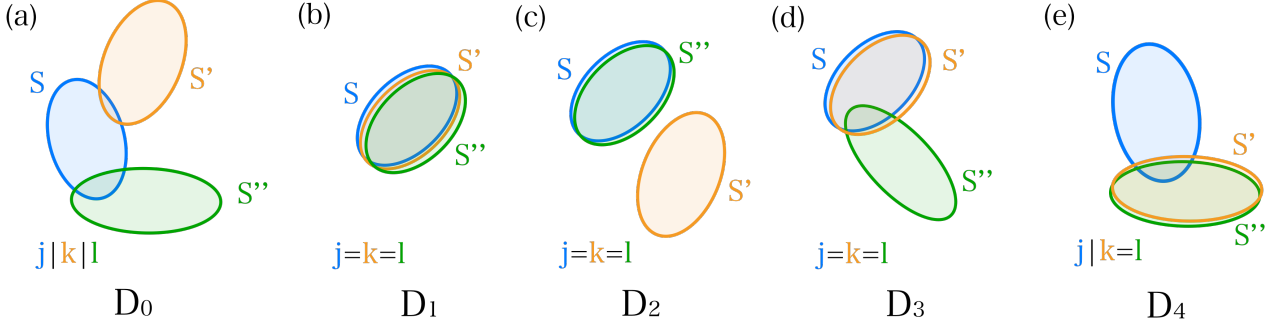
Figure 4: Illustration of *examples* of index sets (S, j), (S', k) and (S'', l) (corresponding to non-zero values of f in Eq. (125)) associated with the different classes of contributions to $A(1)$ in Eq. (124). The D_0 contribution in (a) is the diagonal-free contribution (i.e., (S, j), (S', k) and (S'', l) are unequal). The D_1, D_2, D_3 and D_4 contributions in resp. (b), (c), (d) and (e) are the diagonal contributions (i.e., at least two of (S, j), (S', k) and (S'', l) are equal).

F   Moment bound for dense SYK-q

In this Appendix, we establish the moment bound $\mathbb{E}\big(A(1)^{2n}\big) \leq \big(C\sqrt{n}\big)^{2n}$, where $A(1)$ is defined as (in Eq. (70)):

$$A(1) = C\frac{1}{\sqrt{n}}\binom{n}{q-1}^{-3/2}\sum_{j,k,l=1}^{n}\sum_{S,S',S''} J_{S,j}J_{S',k}J_{S'',l}\, f(S, S', S'', j, k, l). \tag{124}$$

The function f in this expression is defined as (in Eq. (67)):

$$f(S, S', S'', j, k, l) = \begin{cases} 1, & \text{if} \quad (|S'' \cap S| \text{ is odd}) \\ & \qquad \wedge\, (|S' \cap (S'' \triangle S)| + \delta_{k,l} \text{ is odd}) \\ & \qquad \wedge\, (|S| = |S'| = |S''| = q - 1), \\ \\ 0, & \text{otherwise.} \end{cases} \tag{125}$$

We classify the terms in the sum in Eq. (124) into five classes whose total contributions to the sum are denoted by D_0, D_1, D_2, D_3 and D_4. D_0 comprises of all terms for which the three J's are distinct. We shall therefore call the call the D_0 contribution the *diagonal-free* contribution. D_1 comprises of all terms for which the three J's are equal. D_2, D_3 and D_4 comprise of all terms for which exactly two out of three J's are equal.

Taking f into account, and thereby the terms that actually appear in $A(1)$, we conclude that the terms appearing in each class D_0, D_1, D_2, D_3 and D_4 correspond to the index sets given in Table 1. An illustration of *examples* of the index sets (S, j), (S', k) and (S'', l) associated with these different classes of contributions to $A(1)$ is given in Figure 4.

class	associated index sets of terms	associated index sets of terms in $A(1)$
D_0	$(S, j) \neq (S', k) \neq (S'', l) \neq (S, j)$	$(S, j) \neq (S', k) \neq (S'', l) \neq (S, j)$
D_1	$(S, j) = (S', k) = (S'', l)$	$(S, j) = (S', k) = (S'', l)$
D_2	$(S, j) = (S'', l) \neq (S', k)$	$S = S'' \neq S',\ j = l = k$
D_3	$(S, j) = (S', k) \neq (S'', l)$	$S = S' \neq S'',\ j = l = k$
D_4	$(S', k) = (S'', l) \neq (S, j)$	$(S', k) = (S'', l) \neq (S, j)$

Table 1: The index sets associated with each class of terms, and the index sets associated with each class of terms that appear in the expression for $A(1)$ (i.e., taking f into account).

To upper bound the $(2n)$th moment of $A(1)_{\min}$, we upper bound the rth moments (for even $r \leq 16 \cdot 2n$) of D_0, D_1, D_2, D_3, D_4 separately. In particular, if $\mathbb{E}\big((D_i)^r\big) \leq \big(C\sqrt{n}\big)^r$ for $i = 0, 1, ..., 4$ and all even $r \leq 16 \cdot 2n$, then $\mathbb{E}\big(A(1)^{2n}\big) \leq \big(C\sqrt{n}\big)^{2n}$. Note that through the multinomial expansion and successive application of Cauchy-Schwarz inequality these former bounds indeed give an upper bound on the $(2n)$th moment of $A(1)$:

$$\mathbb{E}\big(A(1)^{2n}\big) = \mathbb{E}\Big((D_0 + D_1 + D_2 + D_3 + D_4)^{2n}\Big) = \sum_{k_0+\ldots+k_4=2n} \frac{2n!}{k_0!\ldots k_4!} \mathbb{E}\Big(D_0^{k_0} D_1^{k_1} \ldots D_4^{k_4}\Big)$$

$$\leq \sum_{k_0+\ldots+k_4=2n} C^{2n}\, \mathbb{E}\big(D_0^{2k_0}\big)^{1/2} \mathbb{E}\big(D_1^{4k_1}\big)^{1/4} \mathbb{E}\big(D_2^{8k_2}\big)^{1/8} \mathbb{E}\big(D_3^{16k_3}\big)^{1/16} \mathbb{E}\big(D_4^{16k_4}\big)^{1/16}$$

$$\leq \sum_{k_0+\ldots+k_4=2n} C^{2n}\, \big(C\sqrt{n}\big)^{k_0+\ldots+k_4} \leq \big(C\sqrt{n}\big)^{2n}, \tag{126}$$

where we have used that the multinomial coefficient can be upper bounded by C^{2n} and that the number of 5-tuples of non-negative integers whose sum equals $2n$ is upper bounded by Cn^4 (which is smaller than C^{2n} for some constant C). Although clearly the rth moments of e.g. D_0 have to only be bounded for even $r \leq 2 \cdot 2n$, we bound – for the sake of clarity – the rth moments for even $r \leq 16 \cdot 2n$ for *all* D_i's. We first deal with the case of D_0, since the fact that this contribution is diagonal-free allows one to employ a decoupling technique. Afterwards, we will consider the D_1, D_2, D_3 and D_4 contributions.

First, we state the following lemma, which will be useful throughout this appendix.

Lemma 32. *Let P and P' be two polynomials of centered Gaussian random variables (i.e., the monomials are formed by products of elements from a sequence of independent centered Gaussian random variables, and each variable is allowed to appear in a monomial multiple times) with non-negative coefficients. Then, for any even r, $\mathbb{E}\big(|P + P'|^r\big) \geq \mathbb{E}\big(|P|^r\big)$.*

Proof. We have $\mathbb{E}\big(|P + P'|^r\big) = \mathbb{E}\big(|P|^r\big) + \sum_{k=1}^{r} \binom{r}{k} \mathbb{E}\big(P^{r-k}(P')^k\big)$, and $\mathbb{E}\big(P^{r-k}(P')^k\big)$ is non-negative (for any integers r, k) since P and P' have non-negative coefficients and all moments of centered Gaussian random variables are non-negative. $\square$

F.1 Upper bound for moments of D_0 (diagonal-free contribution)

We start by noting that the function f takes on values 0 or 1, dependent on the index sets S, S', S'', j, k, l labeling the Majorana operators. We consider replacing f in each term of D_0 (Eq. (124)) with $\delta_{a,b}\delta_{c,d}$, where either

$$a \in (S' \cup k),\ b \neq c \in (S'' \cup l),\ d \in (S \cup j), \tag{option 1}$$

or

$$a \in (S' \cup k),\ b \neq c \in (S \cup j),\ d \in (S'' \cup l). \tag{option 2}$$

We denote this modified sum as $D_{0,\delta\delta}$. By inspection, the index sets for which f is non-zero all correspond to a non-zero contribution for $\delta_{a,b}\delta_{c,d}$. Note that those index sets for which $\delta_{a,b}\delta_{c,d}$ is non-zero also include index sets for which f is zero. Hence, the terms associated with non-zero $\delta_{a,b}\delta_{c,d}$ (for the two options listed above) are a superset of the terms that correspond to non-zero values of f. Therefore, by Lemma 32, the upper bounds on even moments of D_0 can be obtained by upper bounding the even moments of $D_{0,\delta\delta}$.

We will denote the part of the sum $D_{0,\delta\delta}$ corresponding to option 1 as $D_{0,\min}$:

$$D_{0,\min} := C\frac{1}{\sqrt{n}} \binom{n}{q-1}^{-3/2} \sum_{\substack{j,k,l, \\ S,S',S'', \\ \text{s.t. } |(S'\cup k)\cap(S''\cup l)|\geq 1 \\ \text{and } |(S''\cup l)\cap(S\cup j)|\geq 1}} J_{S,j} J_{S',k} J_{S'',l}, \tag{127}$$

where the sum is over indices such that $(S,j) \neq (S',k) \neq (S'',l) \neq (S,j)$ (by definition of D_0) and such that $(S' \cup k) \cap (S'' \cup l)$ and $(S'' \cup l) \cap (S \cup j)$ differ by at least one element. Any bound for all even moments of $D_{0,\min}$ also holds for $D_{0,\delta\delta} - D_{0,\min}$ which corresponds to option 2, due to the symmetry $(S,j) \leftrightarrow (S'',l)$ between the two options. An upper bound on all even moments of $D_{0,\delta\delta}$ (and, by implication, D_0) then follows from binomial expansion and application of the Cauchy-Schwarz inequality, similarly to Eq. (126). Thus it only remains to prove $\mathbb{E}\big(|D_{0,\min}|^r\big) < (C\sqrt{n})^r$ for all even r.

To upper bound the even moments of $D_{0,\min}$, we are going to employ a decoupling technique. To that end, we will study the even moments of a related decoupled quantity. This decoupled quantity is defined as $D_{0,\min}$ but with the standard Gaussian random variables $J_{S,j}$, $J_{S',k}$ and $J_{S'',l}$ (selected from a single sequence of standard

Gaussian random variables) being replaced by their *decoupled* versions $J_{S,j}^{(1)}$, $J_{S',k}^{(2)}$ and $J_{S'',l}^{(3)}$ (selected from *three* independent sequences of standard Gaussian random variables). The related decoupled quantity is given by (where the sum is again over indices $(S,j) \neq (S',k) \neq (S'',l) \neq (S,j)$ and again such that $(S' \cup k) \cap (S'' \cup l)$ and $(S'' \cup l) \cap (S \cup j)$ differ by at least one element):

$$C \, \frac{1}{n^{3q/2-1}} \sum_{\substack{j,k,l, \\ S,S',S'', \\ \text{s.t. } |(S'\cup k)\cap(S''\cup l)|\geq 1 \\ \text{and } |(S''\cup l)\cap(S\cup j)|\geq 1}} J_{S,j}^{(1)} J_{S',k}^{(2)} J_{S'',l}^{(3)}, \tag{128}$$

where we have additionally used that $(k/l)^l \leq \binom{k}{l} \leq (e\, k/l)^l$.

To upper bound the even moments of this decoupled quantity, we will make use of Lemma 33 below from [32]. The even moments of this decoupled sum are upper bounded by upper bounding the even moments of a decoupled sum whose terms are a superset of the terms in the sum in Eq. 128. Through Lemma 32, the even moments of the latter sum are larger than those of the former sum. For each $J_{S,j}^{(i)}$, we introduce $q! - 1$ additional independent standard Gaussian random variables associated with the $q!$ permutations of the indices in the subsets of size q. Furthermore, we introduce additional independent standard Gaussian random variables for which some or all of the q indices that label them are equal. We consider a sum over lists of $3q$ indices (which label the independent standard Gaussian random variables) $i_1^{(1)}, \ldots, i_q^{(1)}$, $i_1^{(2)}, \ldots, i_q^{(2)}$ and $i_1^{(3)}, \ldots, i_q^{(3)}$ (with each index in $[2n]$), instead of the sum over subsets of $[2n]$ in Eq. (128). Note that the sum over lists, by definition, can contain terms for which two (or three) of the Gaussian random variables have equal index sets.

The index lists $i_1^{(1)}, \ldots, i_q^{(1)}$ and $i_1^{(2)}, \ldots, i_q^{(2)}$ that are summed over each have *any* one index (denoted by resp. x and y) that is equal to an index in $i_1^{(3)}, \ldots, i_q^{(3)}$. If we additionally sum over all 'positions' of the x and y indices (where p_1, p_2, p_3 and $p_4 \in \{1, \ldots, q\}$ label these positions), we obtain the sum (see Eq. (129) below) whose terms are a superset of those in the sum in Eq. (128). Note that this sum in Eq. (129) contains all the contributions from a sum over lists of indices, and contains some terms multiple times that would occur only once in a sum over lists of indices: For example, in the hypothetical case $q = 2$, one could have a contribution $J_{3,3}^{(1)} J_{6,7}^{(2)} J_{3,7}^{(3)}$ that would appear once in the sum over lists of indices but appears twice in the sum in Eq. (129) (once for $p_1 = 1$ and once for $p_1 = 2$). Through Lemma 32, the even moments of the sum in Eq. (129) will therefore be larger than those of the sum over lists of indices (and therefore larger than those of the sum in Eq. (128)), and it will thus suffice to upper bound the even moments of the sum in Eq. (129).

$$D_{0,\min}^{\text{decoupled}} := C \, \frac{1}{n^{3q/2-1}} \sum_{\substack{p_1,\ldots,p_4=1 \\ \text{s.t. } p_3 \neq p_4}}^{q} \Bigg[$$

$$\times \sum_{\substack{i_1^{(1)},..,i_{p_1-1}^{(1)},x, \\ i_{p_1+1}^{(1)},..,i_q^{(1)}=1}}^{2n} \sum_{\substack{i_1^{(2)},..,i_{p_2-1}^{(2)},y, \\ i_{p_2+1}^{(2)},..,i_q^{(2)}=1}}^{2n} \sum_{\substack{i_1^{(3)},..,i_{p_3-1}^{(3)},x,i_{p_3+1}^{(3)},.., \\ i_{p_4-1}^{(3)},y,i_{p_4+1}^{(3)},..,i_q^{(3)}=1}}^{2n} \left(J_{\substack{i_1^{(1)},..,i_{p_1-1}^{(1)},x, \\ i_{p_1+1}^{(1)},..,i_q^{(1)}}}^{(1)} J_{\substack{i_1^{(2)},..,i_{p_2-1}^{(2)},y, \\ i_{p_2+1}^{(2)},..,i_q^{(2)}}}^{(2)} J_{\substack{i_1^{(3)},..,i_{p_3-1}^{(3)},x,i_{p_3+1}^{(3)},.., \\ i_{p_4-1}^{(3)},y,i_{p_4+1}^{(3)},..,i_q^{(3)}}}^{(3)} \right) \Bigg]. \tag{129}$$

The free indices ($q-1$ indices of $i_1^{(1)}, \ldots, i_q^{(1)}$ and $i_1^{(2)}, \ldots, i_q^{(2)}$, and $q-2$ indices of $i_1^{(3)}, \ldots, i_q^{(3)}$) can be summed over to obtain new independent standard Gaussian random variables denoted by $K_{x,p_1}^{(1)}$, $K_{y,p_2}^{(2)}$ and $K_{x,p_3;y,p_4}^{(3)}$:

$$K_{x,p_1}^{(1)} := 1/\sqrt{(2n)^{q-1}} \sum_{i_1^{(1)},..,i_{p_1-1}^{(1)},i_{p_1+1}^{(1)},..,i_q^{(1)}=1}^{2n} J_{i_1^{(1)},..,i_{p_1-1}^{(1)},x,i_{p_1+1}^{(1)},..,i_q^{(1)}}^{(1)}, \tag{130a}$$

$$K_{y,p_2}^{(2)} := 1/\sqrt{(2n)^{q-1}} \sum_{i_1^{(2)},..,i_{p_2-1}^{(2)},i_{p_2+1}^{(2)},..,i_q^{(2)}=1}^{2n} J_{i_1^{(2)},..,i_{p_2-1}^{(2)},y,i_{p_2+1}^{(2)},..,i_q^{(2)}}^{(2)}, \tag{130b}$$

$$K_{x,p_3;y,p_4}^{(3)} := 1/\sqrt{(2n)^{q-2}} \sum_{\substack{i_1^{(3)},..,i_{p_3-1}^{(3)},i_{p_3+1}^{(3)},.., \\ i_{p_4-1}^{(3)},i_{p_4+1}^{(3)},..,i_q^{(3)}=1}}^{2n} J_{i_1^{(3)},..,i_{p_3-1}^{(3)},x,i_{p_3+1}^{(3)},..,i_{p_4-1}^{(3)},y,i_{p_4+1}^{(3)},..,i_q^{(3)}}^{(3)}, \tag{130c}$$

where we have used that the normalized sum $1/\sqrt{m} \sum_{i=1}^{m} J_i$ of a sequence of standard Gaussian random variables

$J_1, \ldots, J_m$ is again a standard Gaussian random variable. We now obtain the following expression for $D_{0,\min}^{\text{decoupled}}$:

$$D_{0,\min}^{\text{decoupled}} := C \sum_{\substack{p_1,\ldots,p_4=1 \\ \text{s.t. } p_3 \neq p_4}}^{q} \left[\frac{1}{n} \sum_{x,y=1}^{2n} K_{x,p_1}^{(1)} K_{y,p_2}^{(2)} K_{x,p_3;y,p_4}^{(3)} \right]. \tag{131}$$

The sum over all free indices gives an extra total factor of $n^{3q/2-2}$, which partially cancels against $n^{3q/2-1}$ in Eq. (129). Importantly, we note that now the random variables $K_{x,p_1}^{(1)}$ and $K_{x',p_1}^{(1)}$ are independent for $x \neq x'$ (and equivalently for $K_{y,p_2}^{(2)}$ and $K_{x,p_3;y,p_4}^{(3)}$). We will apply Lemma 33 from [32] separately to each contribution to $D_{0,\min}^{\text{decoupled}}$ in Eq. (131) (with a contribution corresponding to one combination of p_i's).

Lemma 33 (Theorem 1 in [32]). *Let* $Y \in \mathbb{R}^{N \times \cdots \times N}$ *be a d-dimensional matrix and define:*

$$F\left(\{K_1^{(j)}, \ldots, K_N^{(j)}\}_{j=1}^d\right) := \sum_{i_1,\ldots,i_d=1}^{N} Y_{i_1,\ldots,i_d} \prod_{j=1}^{d} K_{i_j}^{(j)}, \tag{132}$$

where $\{K_1^{(j)}, \ldots, K_N^{(j)}\}_{j=1}^d$ *are d independent sequences of N standard Gaussian random variables. Then for any integer* $k \geq 2$:

$$\mathbb{E}\left(\left|F\left(\{K_1^{(j)}, \ldots, K_N^{(j)}\}_{j=1}^d\right)\right|^k\right) \leq \left(C \sum_{\mathcal{P}} k^{|\mathcal{P}|/2} \|Y\|_{\mathcal{P}}\right)^k$$

$$\leq \left(C \max_{\mathcal{P}} \left[k^{|\mathcal{P}|/2} \|Y\|_{\mathcal{P}}\right]\right)^k,$$

where $\mathcal{P}$ *are partitions of* $[d]$ *into non-empty parts* $(P_1, \ldots, P_s)$. *The second inequality holds because the number of partitions of* $[d]$ *into non-empty parts is constant in n (since d is constant in n). The quantity* $\|Y\|_{\mathcal{P}}$ *is defined as:*

$$\|Y\|_{\mathcal{P}} = \|Y\|_{(P_1,\ldots,P_s)} := \max\left\{\sum_{i_1,\ldots,i_d=1}^{N} Y_{i_1,\ldots,i_d} x_{i_{P_1}}^{(1)} \ldots x_{i_{P_s}}^{(s)} : \sum_{i_{P_1}} \left(x_{i_{P_1}}^{(1)}\right)^2 \leq 1, \ldots, \sum_{i_{P_k}} \left(x_{i_{P_k}}^{(k)}\right)^2 \leq 1\right\}, \tag{133}$$

with each $x \in \mathbb{R}$.

Remark. *If F in Eq. (132) is diagonal-free (i.e.,* $Y_{i_1,\ldots,i_d} = 0$ *if* $i_j = i_k$ *for any* $j \neq k$*) then the moments of the 'decoupled' F in Eq. (132) are (up to constants only depending on d) an upper bound for the moments of its 'coupled' counterpart* $F'(K_1, \ldots, K_N) := \sum_{i_1,\ldots,i_d=1}^{N} Y_{i_1,\ldots,i_d} \prod_{j=1}^{d} K_{i_j}$ *(i.e., where the random variables are all taken from the same sequence of N standard Gaussian random variables):*

$$\mathbb{E}\left((F')^r\right) \leq C\mathbb{E}\left((F)^r\right).$$

See e.g. Theorem 2.1 in [33].

The fact that this decoupling inequality only holds for diagonal-free polynomials is exactly the reason for differentiating between the diagonal-free contribution D_0 and the diagonal contributions D_1, D_2, D_3, D_4 to $A(1)$.

For $\sum_{x,y=1}^{2n} K_{x,p_1}^{(1)} K_{y,p_2}^{(2)} K_{x,p_3;y,p_4}^{(3)}$ in Eq. (131), we see that $d = 3$ and hence the possible partitions $\mathcal{P}$ are $\{1,2,3\}, \{1\}\{2,3\}, \{2\}\{1,3\}, \{1,2\}\{3\}, \{1\}\{2\}\{3\}$. The associated $\|Y\|_{\mathcal{P}}$ values can be (straightforwardly) calculated and are given in Table 2. Using Table 2 and Lemma 33, we find the following upper bound on $\mathbb{E}\left(\left[C/n \sum_{x,y=1}^{2n} K_{x,p_1}^{(1)} K_{y,p_2}^{(2)} K_{x,p_3;y,p_4}^{(3)}\right]^r\right)$ (for all even r):

$$\left(C/n \max\left(\sqrt{r}n, r\sqrt{n}, r\sqrt{n}, r, r^{3/2}\right)\right)^r \leq \left(C\sqrt{n}\right)^r. \tag{134}$$

Note that $D_{0,\min}^{\text{decoupled}}$ in Eq. 131 consists of q^4 (with $q = O(1)$) contributions, each corresponding to a given combination of p_i's. We can again use the multinomial expansion and successive application of the Cauchy-Schwarz inequality (together with the fact that the multinomial coefficients can be upper bounded by C^r and that the number of q^4-tuples of non-negative integers whose sum equals r is upper bounded by C^r for some constant C) to conclude that the upper bounds of $(C\sqrt{n})^r$ for rth moments (for all even r) of these contributions imply an upper bound of $(C\sqrt{n})^r$ for rth moments (for all even r) of $D_{0,\min}^{\text{decoupled}}$.

We now employ the decoupling inequality from the above remark to obtain

$$\mathbb{E}\left(|D_{0,\min}|^r\right) \leq C\mathbb{E}\left(\left|D_{0,\min}^{\text{decoupled}}\right|^r\right) \leq \left(C\sqrt{n}\right)^r.$$

From the arguments given previously, this implies the desired bound $\mathbb{E}\left(|D_0|^r\right) \leq \left(C\sqrt{n}\right)^r$ for all even r, in particular for $r \leq 16 \cdot 2n$.

$\mathcal{P}$	$|\mathcal{P}|$	$\|Y\|_{\mathcal{P}}$
$\{1,2,3\}$	1	$2n$
$\{1\}\{2,3\}$	2	$\sqrt{2n}$
$\{2\}\{1,3\}$	2	$\sqrt{2n}$
$\{1,2\}\{3\}$	2	1
$\{1\}\{2\}\{3\}$	3	1

Table 2: The different partitions $\mathcal{P}$ of [3] into non-empty parts, with the associated number of parts $|\mathcal{P}|$, and the associated $\|Y\|_{\mathcal{P}}$ for $\sum_{x,y=1}^{2n} K_{x,p_1}^{(1)} K_{y,p_2}^{(2)} K_{x,p_3;y,p_4}^{(3)}$ in Eq. (131). $\|Y\|_{\mathcal{P}}$ for the first four partitions can be straightforwardly evaluated by applying Eq. (133) to Eq. (128), and the fifth $\|Y\|_{\mathcal{P}}$ can be evaluated by additional application of the Cauchy-Schwarz inequality.

F.2 Upper bound for moments of D_1, D_2, D_3 and D_4

In the previous section we used a decoupling inequality to upper bound the rth moments (for even $r \leq 16 \cdot 2n$) of D_0. These decoupling inequalities hold for (Gaussian) polynomials for which each Gaussian monomial is a product of *distinct* Gaussian random variables, i.e., *diagonal-free* polynomials. This holds indeed – by definition – for the D_0 contribution to $A(1)$, but not for contributions D_1, D_2, D_3 and D_4. For that reason, we cannot make use of the same decoupling inequality for the D_1, D_2, D_3 and D_4 contributions. Therefore, we have to resort to other methods to bound their rth moments (for even $r \leq 16 \cdot 2n$).

- The D_1 contribution can be written as:

$$D_1 := C \frac{1}{\sqrt{n}} \binom{n}{q-1}^{-3/2} \sum_{j,S} \left(J_{S,j}\right)^3. \tag{135}$$

The rth moment (with r even) of D_1 can be upper bounded as follows:

$$
\begin{aligned}
\mathbb{E}\left(|D_1|^r\right) &\leq C^r \left[\frac{1}{\sqrt{n}} \binom{n}{q-1}^{-3/2}\right]^r \mathbb{E}\left[\left(\sum_{S,j} \left(J_{S,j}\right)^3\right)^r\right] \\
&\leq C^r \left[\frac{1}{\sqrt{n}} \binom{n}{q-1}^{-3/2}\right]^r \left(\binom{n}{q-1} n \, r^{3/2}\right)^r \\
&\leq C^r n^{(5/2-q/2)r},
\end{aligned}
\tag{136}
$$

where we have used that $\mathbb{E}\left[\left(\sum_{i=1}^m K_i\right)^r\right] = \sum_{k_1+\ldots+k_m=r} \frac{r!}{k_1!\ldots k_m!} \mathbb{E}\left((K_1)^{k_1}\right) \ldots \mathbb{E}\left((K_m)^{k_m}\right)$ (for $K_1, \ldots, K_m$ independent random variables), the fact that (S,j) can take on $2n\binom{2n}{q-1}$ values and the fact that the pth moment of a standard Gaussian random variable is equal to $(p-1)!! \; (\leq p^{p/2})$. For even r, we therefore conclude that $\mathbb{E}\left(|D_1|^r\right) \leq \left(C\sqrt{n}\right)^r$.

- The D_2 and D_3 contributions are equivalent and can be written as:

$$D_2 = D_3 := C \frac{1}{\sqrt{n}} \binom{n}{q-1}^{-3/2} \sum_{j,S,S' \text{ s.t. } S \neq S'} \left(J_{S,j}\right)^2 J_{S',j}. \tag{137}$$

The rth moment (with r even) can be written as follows:

$$\mathbb{E}\left(|D_2|^r\right), \mathbb{E}\left(|D_3|^r\right) = C^r \left[\frac{1}{\sqrt{n}} \binom{n}{q-1}^{-3/2}\right]^r \mathbb{E}\left(\left(\sum_{j,S,S' \text{ s.t. } S \neq S'} \left(J_{S,j}\right)^2 J_{S',j}\right)^r\right). \tag{138}$$

We define

$$g := \sum_{j,S,S' \text{ s.t. } S \neq S'} \left(J_{S,j}\right)^2 J_{S',j}, \tag{139}$$

for which $\mathbb{E}(g) = 0$. We note that g is a homogeneous polynomial in standard Gaussian random variables of degree 3. To upper bound the moments of g, and thereby the moments of D_2 and D_3, we use the following result from [34]. This result is an extension of Lemma 33 from [32] to the setting where diagonal terms are allowed to appear in the polynomial. The extension also includes inhomogeneous polynomials, although in the current setting we are considering only homogeneous polynomials.

Lemma 34 (Theorem 1.3 in [34])**.** *Let* $\mathbf{K} := K_1 \dots, K_N$ *denote a sequence of* N *independent standard Gaussian random variables and* $g : \mathbb{R}^N \to \mathbb{R}$ *a polynomial of degree* D. *Then, for all* $r \geq 2$:

$$\mathbb{E}\left(\left[g(\mathbf{K}) - \mathbb{E}\big(g(\mathbf{K})\big)\right]^r\right) \leq C^r \left[\sum_{1 \leq d \leq 3} \sum_{\mathcal{P}([d])} r^{|\mathcal{P}|/2} \big\|\mathbb{E}\big(\mathbf{D}^d g(\mathbf{K})\big)\big\|_{\mathcal{P}}\right]^r, \quad (140)$$

where $\mathcal{P}$ *are partitions of* $[d]$ *into non-empty parts, and* $\|Y\|_{\mathcal{P}}$ *(with* Y *a* d-*way tensor) is defined in Eq.* (133). $\mathbf{D}^d g(\mathbf{K})$ *denotes the* dth *derivative of* $g(\mathbf{K})$, *which corresponds to a* d-*way tensor with entries equal to* $\big[\mathbf{D}^d g(\mathbf{K})\big]_{i_1,\dots,i_d} = \frac{\partial}{\partial K_{i_1}} \dots \frac{\partial}{\partial K_{i_d}} g(\mathbf{K})$. *For* $d = D$, $\mathbf{D}^d g(\mathbf{K})$ *is constant.*

For g in Eq. (139), we have that $N = n\binom{n}{q-1}$, since the sequence of Gaussian random variables corresponds to $\{J_{S,j}\}$. To find an upper bound for the rth moment of g using Eq. (140), we first calculate $\mathbf{D}^d g$ for $d = 1, 2, 3$. Then, for each d, we upper bound $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ for all partitions $\mathcal{P}$ of $[d]$. We will show that for all d and associated partitions $\mathcal{P}([d])$, $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ can be upper bounded in such a way that $\mathbb{E}\big(|D_2|^r\big), \mathbb{E}\big(|D_3|^r\big) \leq \big(C\sqrt{n}\big)^r$ for all even $2 \leq r \leq 16 \cdot 2n$. Finally, the 0th moment also (trivially) satisfies this upper bound, hence it holds for all even $r \leq 16 \cdot 2n$.

The derivatives of g are equal to:

$$D\,g = \Big(\sum_{S' : S' \neq S} J_{S',j}^2 + 2J_{S,j} \sum_{S' : S' \neq S} J_{S',j}\Big)_{(S,j)} \implies \mathbb{E}(D\,g) = \binom{n}{q-1}_{(S,j)}, \quad (141)$$

$$D^2 g = \begin{pmatrix} 2\sum_{S' : S' \neq S} J_{S',j}, & \text{if } (S,j) = (T,k) \\ 2(J_{T,j} + J_{S,j}), & \text{if } S \neq T \text{ and } j = k \\ 0, & \text{if } j \neq k \end{pmatrix}_{(S,j),(T,k)} \implies \mathbb{E}(D^2 g) = (0)_{(S,j),(T,k)}, \quad (142)$$

$$D^3 g = \begin{pmatrix} 2, & \text{if } (S = T \neq U \text{ or } S \neq T = U \text{ or } S = U \neq T) \text{ and } j = k = l \\ 0, & \text{if } (S = T = U \text{ or } S \neq T \neq U) \text{ and } j = k = l \\ 0, & \text{if } j, k, l \text{ are not all equal} \end{pmatrix}_{(S,j),(T,k),(U,l)} \implies \mathbb{E}(D^3\,g) = D^3\,g.$$
$$(143)$$

In Table 3, we give the values of $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ for all partitions $\mathcal{P}([d])$ for $d = 1, 2, 3$. $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ for $d = 1$ can be straightforwardly evaluated using Eq. (133) and for $d = 2$ can be trivially evaluated by using $\mathbb{E}\big(\mathbf{D}^2 g\big) = 0$. For $d = 3$, $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ can be upper bounded using Eq. (133), and the triangle and Cauchy-Schwarz inequalities (for illustration purposes, we provide an example of the derivation of this upper bound for $\mathcal{P} = \{1, 2\}\{3\}$ below).

$\mathcal{P}$	$|\mathcal{P}|$	$\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$
$\{1\}$	1	$\binom{n}{q-1}^{3/2}$
$\{1, 2\}$	1	0
$\{1\}\{2\}$	2	0
$\{1, 2, 3\}$	1	$C\sqrt{n}\binom{n}{q-1}$
$\{1\}\{2, 3\}$	2	$Cn\binom{n}{q-1}$
$\{2\}\{1, 3\}$	2	$Cn\binom{n}{q-1}$
$\{1, 2\}\{3\}$	2	$Cn\binom{n}{q-1}$
$\{1\}\{2\}\{3\}$	3	$C\binom{n}{q-1}$

Table 3: The different partitions $\mathcal{P}$ of $[3]$ into non-empty parts, with the associated number of parts $|\mathcal{P}|$, and (the upper bounds for) the associated $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ for g in Eq. (139).

Combining the upper bounds for $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ in Table 3 with the factor of $r^{|\mathcal{P}|/2}(\leq Cn^{|\mathcal{P}|/2})$ in Eq. (140) and the normalization factor in Eq. (138), we find – using $\mathbb{E}(g) = 0$ – that indeed $\mathbb{E}\big(|D_2|^r\big), \mathbb{E}\big(|D_3|^r\big) \leq \big(C\sqrt{n}\big)^r$ for all even $r \leq 16 \cdot 2n$.

Example: For illustration purposes, we give an explicit evaluation of $\big\|\mathbb{E}\big(\mathbf{D}^d g\big)\big\|_{\mathcal{P}}$ for $\mathcal{P} = \{1,2\}\{3\}$ (the evaluations for other $\mathcal{P}$'s follow using similar methods). By definition (Eq. (133)), we have:

$$\big\|\mathbb{E}\big(\mathbf{D}^3 g\big)\big\|_{\{1,2\}\{3\}} = \sup\Big\{ \sum_{(S,j),(T,k),(U,l)} \mathbb{E}\big(\mathbf{D}^3 g\big)_{(S,j),(T,k),(U,l)} x_{(S,j),(T,k)} y_{(U,l)}$$
$$: \sum_{(S,j),(T,k)} x^2_{(S,j),(T,k)} \leq 1, \sum_{(U,l)} y^2_{(U,l)} \leq 1 \Big\}. \quad (144)$$

Using the expression obtained for $\mathbb{E}\big(\mathbf{D}^3 g\big)_{(S,j),(T,k),(U,l)}$, we obtain:

$$\big\|\mathbb{E}\big(\mathbf{D}^3 g\big)\big\|_{\{1,2\}\{3\}} = \sup\Big\{ \sum_{j,S,T} 2\, x_{(S,j),(T,j)} y_{(T,j)} + \sum_{j,S,T} 2\, x_{(S,j),(T,j)} y_{(S,j)} + \sum_{j,S,U} 2\, x_{(S,j),(S,j)} y_{(U,j)}$$
$$: \sum_{(S,j),(T,k)} x^2_{(S,j),(T,k)} \leq 1, \sum_{(U,l)} y^2_{(U,l)} \leq 1 \Big\}$$
$$\leq \sup\Big\{ 2\sum_{S}\Big|\sum_{j,T} x_{(S,j),(T,j)} y_{(T,j)}\Big| + 2\sum_{T}\Big|\sum_{j,S} x_{(S,j),(T,j)} y_{(S,j)}\Big|$$
$$+ 2\sum_{j}\Big|\sum_{S} x_{(S,j),(S,j)}\Big| \Big|\sum_{U} y_{(U,j)}\Big| : \sum_{(S,j),(T,k)} x^2_{(S,j),(T,k)} \leq 1, \sum_{(U,l)} y^2_{(U,l)} \leq 1 \Big\}$$
$$\leq \sup\Big\{ 2\sum_{S} \|x\|\, \|y\| + 2\sum_{T} \|x\|\, \|y\|$$
$$+ 2\sum_{j} \|x\|\sqrt{\binom{n}{q-1}} \|y\|\sqrt{\binom{n}{q-1}} : \sum_{(S,j),(T,k)} x^2_{(S,j),(T,k)} \leq 1, \sum_{(U,l)} y^2_{(U,l)} \leq 1 \Big\}$$
$$\leq Cn\binom{n}{q-1}, \quad (145)$$

where we have used the triangle inequality in the first inequality, and the Cauchy-Schwarz inequality for the second inequality (and we note that e.g. $\sum_U y_{(U,j)}$ is simply equal to the inner product of $y_{(j)} := (y_{(U_1,j)}, y_{(U_2,j)}, \ldots)$ with the all-ones vector).

- The D_4 contribution can be written as:

$$D_4 := C\frac{1}{\sqrt{n}}\binom{n}{q-1}^{-3/2} \sum_{\substack{j,k,S,S' \\ \text{s.t. } 0<|S\cap S'|<q-1 \\ \text{is odd}}} J_{S,j}\big(J_{S',k}\big)^2. \quad (146)$$

We note that the main difference with the D_2 and D_3 contributions is that, for D_4, the sum is over the double index j,k (instead of over the single index j), and over a restricted sum over sets S, S' (instead of over a free sum over sets S, S'). To bound the moments of D_4, we will employ a similar method as for D_2 and D_3. The rth moment (with r even) can be upper bounded as follows (where we drop the '$|S \cap S'|$ is odd' constraint using Lemma 32 and denote the collection of subsets S' such that $0 < |S \cap S'| < q-1$ by $\sigma(S)$):

$$\mathbb{E}\big(|D_4|^r\big) \leq C^r\Big[\frac{1}{\sqrt{n}}\binom{n}{q-1}^{-3/2}\Big]^r \mathbb{E}\Big(\Big(\sum_{\substack{j,k,S, \\ S'\in\sigma(S)}} J_{S,j}\big(J_{S',k}\big)^2\Big)^r\Big). \quad (147)$$

We note that $|\sigma(S)|$ can be upper bounded and lower bounded by Cn^{q-2} (for some constants C). We define

$$h := \sum_{\substack{j,k,S, \\ S'\in\sigma(S)}} J_{S,j}\big(J_{S',k}\big)^2, \quad (148)$$

for which $\mathbb{E}(h) = 0$. We note that h is a homogeneous polynomial in standard Gaussian random variables of degree 3. To upper bound the moments of g, and thus the moments of D_4, we again use Lemma 32 from [34]. We use Eq. (140) to find an upper bound for the rth moment of h. We first calculate $\mathbf{D}^d h$ for $d = 1, 2, 3$. Then, for each d, we upper bound $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ for all partitions $\mathcal{P}$ of $[d]$. Thereby, we show that for all d and associated partitions $\mathcal{P}([d])$, $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ can be upper bounded such that $\mathbb{E}\left(|D_4|^r\right) \leq \left(C\sqrt{n}\right)^r$ for all even $2 \leq r \leq 16 \cdot 2n$. The 0th moment trivially satisfies this bound, and therefore it holds for all even $r \leq 16 \cdot 2n$.

The derivatives of h are equal to:

$$D\,h = \Big(\sum_{S'\in\sigma(S),p} J_{S',p}^2 + 2J_{S,j}\sum_{S'\in\sigma(S),p} J_{S',p}\Big)_{(S,j)} \implies \mathbb{E}(Dh) \leq \left(Cn^{q-1}\right)_{(S,j)}, \tag{149}$$

where the sum over p runs from 0 to n and we have used the bounds on $|\sigma(S)|$. Note that this is a pointwise upper bound on the entries of the vector $\mathbb{E}(Dh)$, which will be enough to bound the corresponding norm.

$$D^2 h = \begin{pmatrix} 2J_{T,k} + 2J_{S,j}, & \text{if } T \in \sigma(S) \text{ and } \forall k \\ 2\sum_{S'\in\sigma(S),p} J_{S',p}, & \text{if } (T,k) = (S,j) \\ 0, & \text{otherwise} \end{pmatrix}_{(S,j),(T,k)} \implies \mathbb{E}(D^2 h) = (0)_{(S,j),(T,k)}, \tag{150}$$

$$D^3 h = \begin{pmatrix} 2, & \text{if } ((U,l) = (T,k)\,(U,T \in \sigma(S))) \text{ or} \\ & ((U,l) = (S,j) \text{ and } T \in \sigma(S) \text{ and } \forall k) \text{ or} \\ & ((T,k) = (S,j) \text{ and } U \in \sigma(S) \text{ and } \forall l) \\ 0, & \text{otherwise} \end{pmatrix}_{(S,j),(T,k),(U,l)} \implies \mathbb{E}(D^3 h) = D^3 h. \tag{151}$$

In Table 4, we give the values of $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ for all partitions $\mathcal{P}([d])$ for $d = 1, 2, 3$. $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ for $d = 1$ can be straightforwardly evaluated using Eq. (133) and for $d = 2$ can be trivially evaluated by using $\mathbb{E}(\mathbf{D}^2 h) = 0$. For $d = 3$, $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ can be upper bounded using Eq. (133), and the triangle and Cauchy-Schwarz inequalities. To obtain these upper bounds, we have again used the bounds on $|\sigma(S)|$.

$\mathcal{P}$	$|\mathcal{P}|$	$\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$
$\{1\}$	1	$C\left(n^{q-1}\right)^{3/2}$
$\{1,2\}$	1	0
$\{1\}\{2\}$	2	0
$\{1,2,3\}$	1	$Cn^{q-1/2}$
$\{1\}\{2,3\}$	2	Cn^q
$\{2\}\{1,3\}$	2	Cn^q
$\{1,2\}\{3\}$	2	Cn^q
$\{1\}\{2\}\{3\}$	3	$Cn^{q/2}$

Table 4: The different partitions $\mathcal{P}$ of $[3]$ into non-empty parts, with the associated number of parts $|\mathcal{P}|$, and (the upper bounds for) the associated $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ for h in Eq. (148).

Combining the upper bounds for $\left\|\mathbb{E}(\mathbf{D}^d h)\right\|_{\mathcal{P}}$ in Table 4 with the factor of $r^{|\mathcal{P}|/2}(\leq Cn^{|\mathcal{P}|/2})$ in Eq. (140) and the normalization factor in Eq. (147), we find – using $\mathbb{E}(h) = 0$ – that indeed $\mathbb{E}\left(|D_4|^r\right) \leq \left(C\sqrt{n}\right)^r$ for all even $r \leq 16 \cdot 2n$.

In conclusion, we have shown that the rth moments (for even $r \leq 16 \cdot 2n$) of D_0, D_1, D_2, D_3 and D_4 can be upper bounded by $\left(C\sqrt{n}\right)^r$, and hence, by Eq. (126), the $(2n)$th moment of $A(1)$ can be upper bounded by $\left(C\sqrt{n}\right)^{2n}$. Thereby, we have also established that the second condition in Eq. (71) is satisfied.

G Two-colored SYK to standard SYK

In this Appendix, we give the proof of Lemma 27.

Accepted in ⟨ ⟩uantum 2023-06-07, click title to verify. Published under CC-BY 4.0.

33

Proof. To establish Lemma 27, we now show that the state ρ that achieves $\text{Tr}(H^{(2)}\rho) \geq C\sqrt{n}$ (with $H^{(2)}$ defined in Eq. (54)), with probability at least $1 - \exp\big(-\Omega(n)\big)$ also achieves $\text{Tr}(H\rho) \geq C\sqrt{n}$ for the standard SYK Hamiltonian. To that end, we consider a standard SYK model with $2n$ Majorana operators and partition these Majorana operators into a subset of size $\frac{2n(q-1)}{q}$ and a complementary subset of size $\frac{2n}{q}$. The standard SYK model Hamiltonian H, see Eq. (2), consists of $\binom{2n}{q}$ terms. These terms are labeled by all ordered subsets $\{j_1 < \ldots < j_q\}$, and $\mathcal{I}$ denotes the collection of these subsets. The terms in H for which $q-1$ Majorana operators are in the first subset, and the other Majorana operator is in the complementary subset, are labelled by ordered subsets $\{j_1 < \ldots < j_q : j_1 < \ldots < j_{q-1} < \frac{q-1}{q} \leq j_q\}$. We denote the collection of these subsets by $\mathcal{T}$. The collection of other subsets is denoted by $\mathcal{T}' = \mathcal{I}\backslash\mathcal{T}$. $\mathcal{T}$ and $\mathcal{T}'$ thus correspond to collections of terms in the Hamiltonian. We denote the Hamiltonian consisting of the collection $\mathcal{T}$ by $H_\mathcal{T}$ and the Hamiltonian consisting of terms $\mathcal{T}'$ by $H_{\mathcal{T}'}$, hence $H = H_\mathcal{T} + H_{\mathcal{T}'}$. $H_\mathcal{T}$ corresponds exactly to the 2-colored Hamiltonian in Eq. (54) when multiplied by

$$e^{-(q-1)/2} \leq \frac{\sqrt{\binom{2n}{q}}}{\sqrt{\frac{2n}{q}}\sqrt{\binom{\frac{q-1}{q}2n}{q-1}}} \leq e^{q/2}, \tag{152}$$

which, importantly, is lower bounded and upper bounded by a constant in n. Note that the sizes of the two subsets into which the Majorana operators are partitioned can in fact be *any* $c2n$ and $(1-c)2n$ for $c = O(1)$ (instead of just $\frac{2n(q-1)}{q}$ and $\frac{2n}{q}$). The factor in Eq. (152) is lower bounded and upper bounded by a constant in n as well for all of these other partitions. Hence n is not constrained to be a multiple of q.

For any state ρ, $\mathbb{E}\big(\text{Tr}(H\rho)\big) = 0$, where the expectation value is w.r.t. the couplings in H since the couplings are random variables with zero mean. The state ρ_θ defined in Eq. (57) is able to achieve $\text{Tr}(H^{(2)}\rho_\theta) \geq C\sqrt{n}$ (with high probability) since it is constructed using a circuit that itself *depends* on the random couplings J_I ($I \in \mathcal{T}$) appearing in $H^{(2)}$. Since ρ_θ does *not* depend on the couplings J_I with $I \in \mathcal{T}'$, we have $\text{Tr}(H_{\mathcal{T}'}\rho_\theta) = 0$. Since: (i) $|\text{Tr}(C_I\rho)| \leq 1$ (for any ρ) for $I \in \mathcal{T}'$, (ii) that each J_I is a standard Gaussian random variable, and (iii) that $|\mathcal{T}'| \leq \binom{2n}{q}$, the quantity

$$\text{Tr}\big(H_{\mathcal{T}'}\rho\big) = \binom{2n}{q}^{-1/2} \sum_{I \in \mathcal{T}'} J_I \text{Tr}(C_I\rho) \tag{153}$$

is a Gaussian random variable with zero mean and variance at most one, for any ρ. Then, $\mathbb{E}\big[\exp(t\,\text{Tr}(H_{\mathcal{T}'}\rho))\big] \leq \exp(\frac{1}{2}t^2)$ for all $t \geq 0$. Applying Chernoff's bound to $\text{Tr}(H_{\mathcal{T}'}\rho)$, and choosing $t = C\sqrt{n}$, we obtain:

$$\Pr\Big[|\text{Tr}(H_{\mathcal{T}'}\rho)| \geq C\sqrt{n}\Big] \leq 2\exp\big(-\Omega(n)\big), \tag{154}$$

for any constant C.

Using Eq. (154) and $\text{Tr}(H\rho) = \text{Tr}(H_\mathcal{T}\rho) + \text{Tr}(H_{\mathcal{T}'}\rho)$, we conclude that the state ρ_θ which achieves $\text{Tr}(H^{(2)}\rho_\theta) \geq C\sqrt{n}$ (i.e., for the 2-colored SYK Hamiltonian) with probability at least $1 - \exp\big(-\Omega(n)\big)$, also achieves $\text{Tr}(H\rho_\theta) \geq C\sqrt{n}$ (where H is the standard SYK Hamiltonian in Eq. (2)) with probability at least $1 - \exp\big(-\Omega(n)\big)$. Therefore, $\lambda_{\max}(H) \geq C\sqrt{n}$ with probability at least $1 - \exp\big(-\Omega(n)\big)$. $\qquad\square$