

Trotter error and gate complexity of the SYK and sparse SYK models

Yiyuan Chen¹, Jonas Helsen^{2,3}, and Maris Ozols^{4,1,2}

¹Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands

²QuSoft, Amsterdam, The Netherlands

³CWI, Amsterdam, The Netherlands

⁴Institute for Logic, Language, and Computation, University of Amsterdam, The Netherlands

The Sachdev–Ye–Kitaev (SYK) model is a prominent model of strongly interacting fermions that serves as a toy model of quantum gravity and black hole physics. In this work, we study the Trotter error and gate complexity of the quantum simulation of the SYK model using Lie–Trotter–Suzuki formulas. Building on recent results by Chen and Brandão [CB24] – in particular their uniform smoothing technique for random matrix polynomials – we derive bounds on the first- and higher-order Trotter error of the SYK model, and subsequently find near-optimal gate complexities for simulating these models using Lie–Trotter–Suzuki formulas. For the k -local SYK model on n Majorana fermions, at time t , our gate complexity estimates for the first-order Lie–Trotter–Suzuki formula scales with $\tilde{\mathcal{O}}(n^{k+\frac{5}{2}}t^2)$ for even k and $\tilde{\mathcal{O}}(n^{k+3}t^2)$ for odd k , and the gate complexity of simulations using higher-order formulas scales with $\tilde{\mathcal{O}}(n^{k+\frac{1}{2}}t)$ for even k and $\tilde{\mathcal{O}}(n^{k+1}t)$ for odd k . Given that the SYK model has $\Theta(n^k)$ terms, these estimates are close to optimal. These gate complexities can be further improved upon in the context of simulating the time evolution of an arbitrary fixed input state $|\psi\rangle$, leading to a $\mathcal{O}(n^2)$ -reduction in gate complexity for first-order formulas and $\mathcal{O}(\sqrt{n})$ -reduction for higher-order formulas.

We also apply our techniques to the sparse SYK model, which is a simplified variant of the SYK model obtained by deleting all but a $\Theta(n)$ fraction of the terms in a uniformly i.i.d. manner. We find the average (over the random term removal) gate complexity for simulating this model using higher-order formulas scales with $\tilde{\mathcal{O}}(n^{1+\frac{1}{2}}t)$ for even k and $\tilde{\mathcal{O}}(n^2t)$ for odd k . Similar to the full SYK model, we obtain a $\mathcal{O}(\sqrt{n})$ -reduction simulating the time evolution of an arbitrary fixed input state $|\psi\rangle$.

Our results highlight the potential of Lie–Trotter–Suzuki formulas for efficiently simulating the SYK and sparse SYK models, and our analytical methods can be naturally extended to other Gaussian random Hamiltonians.

Contents

1	Introduction	3
1.1	Previous work	3
1.2	Summary of results for the SYK model	5
1.2.1	Error bounds	5
1.2.2	Gate complexity	6
1.3	Summary of results for the sparse SYK model	7
1.4	Comparison with numerical simulations	8
1.4.1	Scaling in n	8
1.4.2	Scaling in t	10
1.5	Structure of the paper	11
1.6	Acknowledgements	11

2	Preliminaries	11
2.1	The SYK and sparse SYK models	12
2.2	Representation and norm of Majoranas	12
2.3	Notation for local Hamiltonians	13
2.4	Trotter error and product formulas	14
2.5	Probability statements and concentration inequalities	15
2.6	Gate complexity	16
2.7	Error representation	17
2.8	Random matrix polynomials	17
2.8.1	Rademacher expansion	18
2.8.2	Resummation by subsets	19
2.8.3	Remark for notations	20
3	First-order Trotter error of the SYK model	21
3.1	Error bound	21
3.2	Gate complexity	24
4	Higher-order Trotter error of the SYK model	25
4.1	Error bound	26
4.1.1	g -th order error generator	27
4.1.2	Beyond g -th order	30
4.1.3	Total error	31
4.2	Gate complexity	32
5	Higher-order Trotter error of the sparsified SYK model	34
5.1	Average Trotter error	34
5.2	Average gate complexity	40
6	Discussion	43
6.1	Gate complexities for simulating the SYK model	43
6.2	Gate complexities for simulating the sparse SYK model	45
6.3	Error bounds for general Gaussian models	45
6.4	Outlook and open questions	46
A	Lemmas for evaluating the probability statements	51
B	Anti-commutation lemma for Majorana fermions	52
C	Random matrix polynomials with Gaussian coefficients	53
D	Evaluating G_w for general Hamiltonians	59
E	Average G_w for sparse Hamiltonians	62

1 Introduction

The *Sachdev–Ye–Kitaev* (SYK) model is a randomized model of n strongly interacting fermions that has recently gained significant interest as a toy model for holography [Kita, Kitb] in the field of quantum gravity. Notably, it can capture certain characteristics of black holes such as Hawking radiation [Mon20]. Although this model is solvable in certain limits [Ros19, MS16, JY17], its behavior at finite n remains nontrivial. Simulating the SYK model on a quantum computer would help us gain insights into its properties [JZL⁺22]. *Out-of-time-order correlators* (OTOC), which are an important measurable quantity that serve as a measure for quantum chaos [HMY17, GMJW23], can be measured for the SYK model at finite n on a quantum computer [GAEL⁺17]. Such measurements have already been performed for small numbers of n on noisy superconducting quantum computers [AJS24].

The main goal of this work is to study the quantum simulation of the SYK model using *Lie–Trotter–Suzuki* formulas, also referred to as *product formulas*. These formulas approximate the time evolution of an arbitrary Hamiltonian, and the error of these approximations is referred to as *Trotter error*. Recently Chen and Brandão [CB24] developed sophisticated tools for bounding the Trotter error of *random Hamiltonians*. Using their toolkit, adapted to fermionic Hamiltonians, we derive error bounds for the first-order and higher-order Trotter errors of the SYK model. Based on these error bounds, we subsequently estimate the gate complexity of the quantum simulation of the SYK model using product formulas.

Additionally, we study quantum simulation of the *sparse* SYK model proposed by Xu, Susskind, Su, and Swingle [XSSS20]. This is a modified version of the SYK model that requires fewer resources to simulate, but still captures several essential properties. In particular, it exhibits a maximally chaotic gravitational sector at low temperature [XSSS20, OGP24]. Moreover, [AGK25] demonstrates that there exists a certain correspondence between the sparse and dense SYK models in terms of limiting maximum energy, and similar correspondence can be found across spin glasses [Cra07, AGK25]. The sparse SYK model can be obtained by probabilistically removing each local interaction in the SYK model with a given probability, such that on average there are only $\mathcal{O}(n)$ terms left. Due to the probabilistic nature of the removal, we study the Trotter error (and the corresponding gate complexity) in terms of average-case statements.

1.1 Previous work

The notion of Trotter error dates back to the late 1800s (in the context of Lie algebra theory) and has been widely studied in the context of quantum computation [CST⁺21]. In recent years, significant progress has been made in bounding the Trotter error [CB24, CST⁺21, BACS06, Som16, CS19]. In 2019, Childs et al. showed that the l -th order Trotter error for any Hamiltonian in the form $H = \sum_{i=1}^{\Gamma} H_i$ scales with the sum of (spectral) norms of nested commutators [CST⁺21, Theorem 6]:¹

$$\|e^{iHt} - S_l(t)\| = \mathcal{O} \left(\sum_{i_1, \dots, i_{l+1}=1}^{\Gamma} \|[H_{i_{l+1}}, \dots [H_{i_2}, H_{i_1}] \dots]\| t^{l+1} \right). \quad (1)$$

With this, they showed that the gate complexity G of simulating the time-evolution operator using higher-order product formulas scales as

$$G \approx \Gamma \|H\|_{\text{global},1} t, \quad (2)$$

¹Here, e^{iHt} denotes the time-evolution operator where we hide the minus-sign in the Hamiltonian. $S_l(t)$ represents the Trotterized unitary of order l .

where Γ is the number of terms in $H = \sum_{i=1}^{\Gamma} H_i$ [CST+21], and the global 1-norm is defined as

$$\|H\|_{\text{global},1} = \sum_{i=1}^{\Gamma} \|H_i\|. \quad (3)$$

In 2023, Chen and Brandão developed an improved bound on the Trotter error of *random* k -local qubit Hamiltonians [CB24]. They studied two cases extensively. First, for a random qubit Hamiltonian $H = \sum_{i=1}^{\Gamma} H_i$ such that each term is independent, bounded, and has zero mean, i.e.

$$\mathbb{E}(H_i) = 0, \quad \text{and} \quad \|H_i\| \leq b_i < \infty, \quad (4)$$

the gate complexity G of implementing the time-evolution operator using first-order product formula scales as

$$G \approx \Gamma \|H\|_{\text{global},2} \|H\|_{\text{local},2} t^2, \quad (5)$$

where the global and local 2-norms are defined as

$$\|H\|_{\text{global},2} = \sqrt{\sum_{i=1}^{\Gamma} b_i^2} \quad \text{and} \quad \|H\|_{\text{local},2} = \max_{\text{site } s} \sqrt{\sum_{i: H_i \text{ acts on site } s} b_i^2}. \quad (6)$$

This result applies even when the Hamiltonian is non-local, but does require each term H_i to be bounded.

The second case applies to random k -local qubit Hamiltonians with Gaussian coefficients

$$H = \sum_{i=1}^{\Gamma} g_i K_i \quad (7)$$

where g_i 's are i.i.d. standard Gaussian coefficients and K_i deterministic matrices bounded by $\|K_i\| \leq b_i$. Chen and Brandão found that the gate complexity of implementing the time-evolution operator using higher-order product formulas in this case enjoys an even better scaling:

$$G \approx \Gamma \|H\|_{\text{local},2} t. \quad (8)$$

Contrary to the previous case, this result does not require each local term $g_i K_i$ to be bounded, as the Gaussian coefficient g_i is an unbounded random variable. However, this result does require the Hamiltonian to be local. Both cases provide a mean square root improvement over the results of Childs et al. [CB24].

However, Chen and Brandão's results [CB24] are not directly applicable to the SYK model, because their theorems are derived in the context of local qubit Hamiltonians while the SYK model is a fermionic Hamiltonian. In particular, the Majorana fermions obey a different algebra than qubits. Contrary to qubits, Majorana fermions anti-commute when they do not collide. Mapping Majorana fermions to qubit operators (through e.g. the Jordan–Wigner transform) incurs inevitable locality overheads that make the bounds in [CB24] suboptimal. Still, [CB24] contains useful theorems and tools, such as uniform smoothness and Rademacher expansions, which apply in a broader context for operators as opposed to scalars. Using these results directly in the fermionic setting we can study the Trotter error of the SYK and the sparse SYK model.

There are other works studying the Trotter error of the SYK model. García-Álvarez et al. [GAEL+17] gave a rough estimate of the gate complexity for simulating the SYK model with $k = 4$ using the first-order product formula. Their result estimated that the gate complexity scales as $\mathcal{O}(n^{10}t^2)$. Asaduzzaman et al. [AJS24] found an improved gate complexity of $\mathcal{O}(n^5t^2)$ by using a graph-coloring algorithm to minimize the number of commuting clusters of terms in the qubitized Hamiltonian, although this result seems to be formulated under a norm assumption we could not verify. Both of these results only consider the first-order product formula specifically

applied to the SYK model with $k = 4$. For the SYK model with general k , Xu, Susskind, Su, and Swingle [XSSS20] give a gate complexity estimate of $\mathcal{O}(n^{2k-1+(k+1)/l} \log(n)t^{1+1/l})$ for simulations using higher-order product formulas (where l is the degree of the Trotter formula). They have also given a gate complexity estimate of $\mathcal{O}(n^{1+1/l} \log(n)t^{1+1/l})$ for simulating the sparse SYK model using higher-order product formulas. All the works above evaluate the complexity scaling under an average case statement, i.e. they compute the gate complexity for which the average Trotter error of the corresponding simulation is sufficiently small. In this work, we present an improved result in gate complexity using higher-order product formulas for both the SYK model and the sparse SYK model with general k , formulated under a more precise probability statement to characterize sufficiently small Trotter error.

Product formulas are not the only way to simulate the SYK model. For instance, asymmetric qubitization allows one to reach a gate complexity of $\mathcal{O}(n^{7/2}t + n^{5/2}t \cdot \text{polylog}(n/\epsilon))$ for simulating² the SYK model with $k = 4$ [BBN19], which is lower than the theoretical limit of product formulas. However, unlike product formulas, this algorithm is less suitable for current noisy quantum devices [CB24, AJS24].

1.2 Summary of results for the SYK model

1.2.1 Error bounds

We present a summary of our main results here. First, we have obtained bounds for the first- and higher-order Trotter error of the SYK model, established in Theorems 3.1 and 4.1 respectively (we state simplified versions below). In the statements below, n denotes the number of Majorana fermions, k is the locality of the SYK model, p refers to the Schatten p -norm, while r and l denote the number of rounds (also referred to as *Trotter number*) and the order of the Lie–Trotter–Suzuki formula, respectively.

Theorem (Simplified version of Theorem 3.1). *The first-order Trotter error of the SYK model for fixed locality k and $p \geq 2$ is bounded by*

$$\frac{\| \| e^{iHt} - S_1(t/r)^r \| \|_{*,p}}{\| \| I \| \|_{*,p}} \leq \Delta_1^{\text{dense}}, \quad (9)$$

where

$$\Delta_1^{\text{dense}} = \begin{cases} \mathcal{O}(p^2 \sqrt{nt}^2 [\frac{1}{r} + \frac{t}{r^2}]) & \text{if } k \text{ is even,} \\ \mathcal{O}(p^2 nt^2 [\frac{1}{r} + \sqrt{n} \frac{t}{r^2}]) & \text{if } k \text{ is odd.} \end{cases} \quad (10)$$

Theorem (Simplified version of Theorem 4.1). *The (even) l -th order Trotter error of the SYK model for fixed k and $l, p \geq 2$ is bounded by*

$$\frac{\| \| e^{iHt} - S_l(t/r)^r \| \|_{*,p}}{\| \| I \| \|_{*,p}} \leq \Delta_l^{\text{dense}}, \quad (11)$$

where

$$\Delta_l^{\text{dense}} = \begin{cases} \mathcal{O}(\sqrt{pnt} \left([\sqrt{p} \frac{t}{r}]^l + n^k [\sqrt{p} \frac{t}{r}]^{l+1} \right)) & \text{if } k \text{ is even,} \\ \mathcal{O}(\sqrt{pnt} \left([\sqrt{pn} \frac{t}{r}]^l + n^k [\sqrt{pn} \frac{t}{r}]^{l+1} \right)) & \text{if } k \text{ is odd.} \end{cases} \quad (12)$$

In these theorems the norm $\| \| A \| \|_{*,p}$ for a random operator A on a given Hilbert space \mathcal{H} is either the expected Schatten p -norm

$$\| \| A \| \|_p = (\mathbb{E}(\| \| A \| \|_p^p))^{\frac{1}{p}}, \quad (13)$$

²Here ϵ denotes the accuracy.

or the L^p -expected operator norm

$$\|A\|_{\text{fix},p} = \sup \left\{ \left(\mathbb{E}(\|A|\psi\rangle\|_{l_2}^p) \right)^{\frac{1}{p}} : |\psi\rangle \in \mathcal{H}, \|\psi\rangle\|_{l_2} = 1 \right\}. \quad (14)$$

1.2.2 Gate complexity

Orders of product formulas	Gate complexity	
	Even k	Odd k
First-order	$n^{k+\frac{5}{2}}t^2$	$n^{k+3}t^2$
Higher-order	$n^{k+\frac{1}{2}}t$	$n^{k+1}t$

Table 1: The gate complexity of simulating the SYK model on a quantum computer such that $\|e^{iHt} - S_l(t/r)^r\| < \epsilon$. We omitted the big- \mathcal{O} notation in the cells. The ϵ -scaling of these bounds go with $\sim \frac{1}{\epsilon}$ for first-order formulas and $\sim \left(\frac{1}{\epsilon}\right)^{\frac{1}{l}}$ for l -th order formulas.

Based on the above error bounds, we can estimate the corresponding gate complexity of the quantum simulation of the k -local SYK model of n Majorana fermions using product formulas. We define the gate complexity as the number of gates required to approximate the time-evolution operator of the SYK model using product formulas such that the Trotter error is sufficiently small. We quantify small Trotter error from two different perspectives. The first is to characterize it by the spectral norm of the difference between the time-evolution operator and the product formula approximation, which we can informally write down as

$$\|e^{iHt} - S_l(t/r)^r\| < \epsilon. \quad (15)$$

A precise statement of this is given in equation (49). Under this characterization of small Trotter error, we get the gate complexities shown in Table 1. Notice the different scaling behaviors for the gate complexities depending on the parity of the locality k of the SYK model, which we indeed expect as the behavior of the SYK model is different for even k and odd k .

We find that for the SYK model with $k = 4$, the gate complexity using first-order product formula scales as $n^{6.5}t^2$, and the gate complexity using higher-order product formulas scales as $n^{4.5}t$. This offers an improvement of $\mathcal{O}(n^{3.5})$ and $\mathcal{O}(n^{5.5})$ in n respectively compared to the previous result $\mathcal{O}(n^{10}t^2)$ obtained by [GAEL⁺17]. Our higher-order result also offers an improvement of $\mathcal{O}(n^{2.5})$ compared to the $\mathcal{O}(n^7t)$ result obtained by [XSS20]³, and a slight improvement of $\mathcal{O}(\sqrt{n})$ compared to the $\mathcal{O}(n^5t^2)$ result obtained by [AJS24], although the latter result seems to assume that the Trotter number r is constant in n , which in general is not the case when simulating the SYK model.

Moreover, as simulating the SYK model with product formulas of any order requires at least $\Omega(n^k)$ gates to capture all Hamiltonian terms, our estimated gate complexities $\mathcal{O}(n^{k+\frac{1}{2}}t)$ and $\mathcal{O}(n^{k+1}t)$ for implementing the SYK model using higher-order product formulas with even and odd k respectively, are close to optimal.

The other way to characterize small Trotter error is by the l^2 -norm deviation of approximating the time evolution of an arbitrary fixed input state $|\psi\rangle$ compared to its real time evolution, which we can informally write down as

$$\|(e^{iHt} - S_l(t/r)^r)|\psi\rangle\|_{l_2} < \epsilon. \quad (16)$$

³The original statement is $\mathcal{O}(n^{2k-1+(k+1)\delta} \log(n)t^{1+\delta})$, but we have dropped $\log(n)$ and taken δ to be sufficiently small for simplicity.

Orders of product formulas	Gate complexity	
	Even k	Odd k
First-order	$n^{k+\frac{1}{2}}t^2$	$n^{k+1}t^2$
Higher-order	$n^k t$	$n^{k+\frac{1}{2}}t$

Table 2: The gate complexity of simulating the SYK model on a quantum computer such that for an arbitrary fixed input state $|\psi\rangle$ we have $\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l_2} < \epsilon$ with high probability. We omitted the big- \mathcal{O} notation in the cells. The ϵ -scaling of these bounds go with $\sim \frac{1}{\epsilon}$ for first-order formulas and $\sim \left(\frac{1}{\epsilon}\right)^{\frac{1}{l}}$ for l -th order formulas.

A precise statement of this is given in equation (50). Under this characterization of small Trotter error, we get the gate complexities shown in Table 2. As we can see, simulating the SYK model in this case requires fewer gates, as the gate complexity of using first-order formulas gets reduced by a factor of $\mathcal{O}(n^2)$ and the gate complexity using higher-order formulas gets reduced by a factor of $\mathcal{O}(\sqrt{n})$ compared to Table 1. Moreover, given that simulating the SYK model requires at least $\Omega(n^k)$ gates for any order or product formula, we find that higher-order formulas are optimal for approximating the time evolution of an arbitrary fixed input state $|\psi\rangle$ under the even- k SYK Hamiltonians.

1.3 Summary of results for the sparse SYK model

We also extend our analysis to the sparse SYK model. The probabilistic removal of terms for obtaining the sparse SYK model can be equivalently achieved by attaching a Bernoulli variable $B_i \in \{0, 1\}$ to each term of the SYK model, such that every B_i takes the value 1 with a certain probability p_B and the value 0 with probability $1 - p_B$. The value of p_B is tuned in such a way that on average, there are $\mathcal{O}(n)$ terms in the sparse SYK model.

Due to the probabilistic removal, the number of remaining terms in the sparse SYK model depends on the specific sample of the Bernoulli variables. Hence, both the Trotter error and the corresponding gate complexity will also depend on the specific sample of the Bernoulli variables. Therefore, we study the sparse SYK model in the following setting. First, we treat the Bernoulli variables as coefficients and compute the Trotter error and the gate complexity as functions depending on these coefficients. Then, by averaging over these Bernoulli coefficients, we end up with average-case statements for both the Trotter error and gate complexity of the sparse SYK model.

We first state the following theorem bounding the average Trotter error:

$$\left\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \right\rangle := \left\langle \left\| e^{iH_{\text{SSYK}}(\mathbf{B})t} - S_l(t/r)^r \right\|_{*,p} \right\rangle, \quad (17)$$

where we use $\langle \cdot \rangle$ to denote averaging over the Bernoulli variables.

Theorem (Simplified version of Theorem 5.1). *Consider the sparse SYK model $H_{\text{SSYK}}(\mathbf{B})$. The l -th order average Trotter error (209) is bounded by*

$$\frac{\left\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \right\rangle}{\|I\|_{*,p}} \leq \Delta_l^{\text{sparse}} \quad (18)$$

where the error bound Δ_l^{sparse} is given by

$$\Delta_l^{\text{sparse}} := \begin{cases} n^{\frac{3-k}{2}} \sqrt{pt} \cdot \left(\left[\sqrt{pn}^{\frac{1}{2} - \frac{k}{2} \frac{t}{r}} \right]^l + n^k \left[\sqrt{pn}^{\frac{1}{2} - \frac{k}{2} \frac{t}{r}} \right]^{l+1} \right), & \text{if } k \text{ is even,} \\ n^{\frac{3-k}{2}} \sqrt{pt} \cdot \left(\left[\sqrt{pn}^{1 - \frac{k}{2} \frac{t}{r}} \right]^l + n^k \left[\sqrt{pn}^{1 - \frac{k}{2} \frac{t}{r}} \right]^{l+1} \right), & \text{if } k \text{ is odd.} \end{cases} \quad (19)$$

Similarly, we find the following *average* gate complexity of the simulation of the sparse SYK model using higher-order product formulas:

$$\overline{G} \sim \begin{cases} n^{1+\frac{1}{2}} \mathcal{J}t & \text{if } k \text{ is even,} \\ n^2 \mathcal{J}t & \text{if } k \text{ is odd.} \end{cases} \quad (20)$$

such that the spectral norm difference is sufficiently small. This gate complexity, compared to the (regular) SYK model, is significantly smaller. Similar to the argument made before for the SYK model, our result for the sparse SYK model is near optimal, since simulating the sparse SYK model using any order of product formulas requires on average $\Omega(n)$ gates.

Similar to the full SYK Hamiltonian, if we only aim for a slight deviation when approximating the time-evolution of an arbitrary fixed input state $|\psi\rangle$, we get the following improved average gate complexity

$$\overline{G} \sim \begin{cases} n \mathcal{J}t & \text{if } k \text{ is even,} \\ n^{1+\frac{1}{2}} \mathcal{J}t & \text{if } k \text{ is odd.} \end{cases} \quad (21)$$

which shows that higher-order formulas are also near optimal when it comes to approximating the time evolution of an arbitrary fixed input state $|\psi\rangle$ under even k sparse SYK Hamiltonians. The above two results also hold for a general class of sparse Gaussian random Hamiltonians where the sparsification is regulated by the Bernoulli variables similar to the sparse SYK model.⁴

1.4 Comparison with numerical simulations

We also study the performance of our bounds relative to actual numerical computations of the normalized Trotter error,

$$\| \| e^{iH_{\text{SYK}}t} - S_l(t/r)^r \| \|_{\overline{p}} := \frac{\| \| e^{iH_{\text{SYK}}t} - S_l(t/r)^r \| \|_p}{\| \| I \| \|_p} \quad (22)$$

for the dense SYK model, and the average normalized Trotter error,

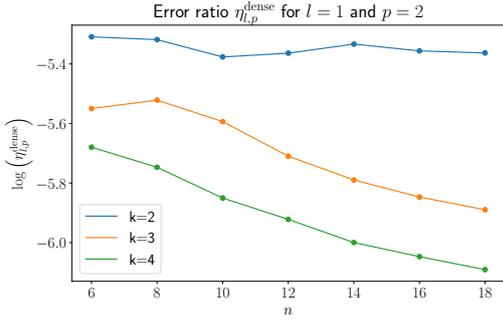
$$\left\langle \| \| e^{iH_{\text{SSYK}}t} - S_l(t/r)^r \| \|_{\overline{p}} \right\rangle := \frac{\left\langle \| \| e^{iH_{\text{SYK}}t} - S_l(t/r)^r \| \|_p \right\rangle}{\| \| I \| \|_p} \quad (23)$$

for the sparse SYK model. The $\langle \dots \rangle$ averages are taken over the Bernoulli variables. We perform two types of test of our error bounds by comparing their scaling in n , the system size, and in t , the evolution time, with the numerical values. The code we used for the numerics here can be found in GitHub repository [Che].

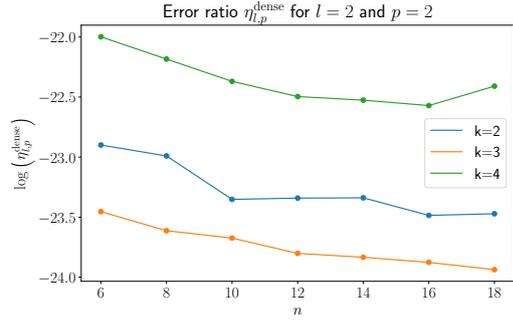
1.4.1 Scaling in n

Due to hardware limitations, our simulations could not be performed beyond $n = 18$. Consequently, it is not meaningful to compare the asymptotic scaling exponents of the analytical error bounds

⁴See equation (288).

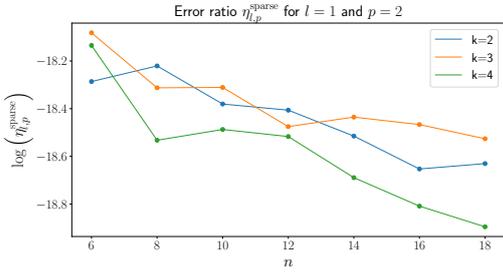


(a) Error ratio of the first-order Trotter error.

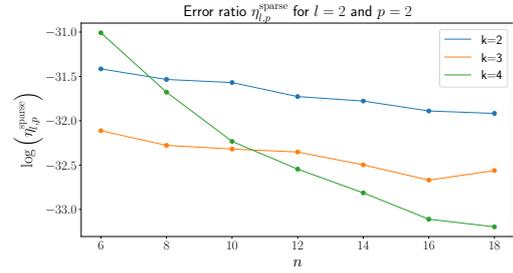


(b) Error ratio of the second-order Trotter error.

Figure 1: Error ratios of the dense SYK models of different localities for even n from $n = 6$ to $n = 18$.



(a) Error ratio of the first-order Trotter error.



(b) Error ratio of the second-order Trotter error.

Figure 2: Error ratios of the sparse SYK models of different localities for even n from $n = 6$ to $n = 18$.

with the numerical results directly. To enable a more reliable comparison at small system sizes, we introduce the following *error ratios*:

$$\eta_{l,p}^{\text{dense}} := \frac{1}{\Delta_l^{\text{dense}}} \|e^{iH_{\text{SYK}}t} - S_l(t/r)^r\|_{\bar{p}}, \quad (24)$$

for the dense SYK model, and

$$\eta_{l,p}^{\text{sparse}} := \frac{1}{\Delta_l^{\text{sparse}}} \left\langle \|e^{iH_{\text{SSYK}}t} - S_l(t/r)^r\|_{\bar{p}} \right\rangle, \quad (25)$$

for the sparse SYK model.

If the analytical bounds correctly capture the finite- n scaling behavior of the actual Trotter error, these ratios should remain approximately constant as n varies. In particular, a constant $\eta_{l,p}$ indicates that the bound differs from the true Trotter error only by a fixed multiplicative factor. Throughout our numerical analysis, we set $p = 2$, use $r = 10^5$ Trotter steps, and focus exclusively on the first- and second-order Trotter errors.

For the dense SYK model with $k = 2, 3, 4$, we plot the error ratio $\eta_{l,p}^{\text{dense}}$ for both the first- and second-order Trotter errors in Figure 1. As shown, the scaling behavior predicted by our analytical bounds Δ_l^{dense} agrees well with the numerical Trotter errors at small n . Overall, the plots display a decreasing trend, meaning that we are slightly overestimating the actual Trotter error. We do not know if this remains the case for larger n , as computing the Trotter error for $n > 18$ lies beyond our current computational capabilities. We nevertheless encourage future work to explore higher n -regime further.

We performed a similar numerical analysis for the sparse SYK model and plotted the corresponding error ratios $\eta_{l,p}^{\text{sparse}}$ for the first- and second-order Trotter errors in Figure 2. In this case, the analytical bounds Δ_l^{sparse} tend to overestimate the actual Trotter errors more than the dense

case, particularly for the second-order cases with $k = 4$. It remains uncertain whether these ratios would approach a constant value at larger n , as our computational resources do not allow us to probe beyond this regime. We encourage interested readers to explore this question further.

1.4.2 Scaling in t

To examine the scaling behavior with respect to t , we fix the system size to $n = 10$. In this setting, the operator dimension remains constant as t varies, allowing us to compute the Trotter error for times up to $t \sim 1000$, where an asymptotic scaling exponent becomes meaningful.

For the dense SYK model, we plot the first- and second-order normalized Trotter errors together with the analytical bounds Δ_l^{dense} in Figure 3. The scaling exponent in t is obtained via a linear fit $y = ax + b$, where y corresponds to either $\log(\|e^{iH_{\text{SYK}}t} - S_l(t/r)^r\|_{\mathbb{F}_2})$ or $\log(\Delta_l^{\text{dense}})$, and $x = \log(t)$. Once again, we find that our error bounds accurately capture the scaling behavior in t .

Similarly, for the sparse SYK model, we plot the first- and second-order average normalized Trotter errors together with the analytical bounds Δ_l^{sparse} in Figure 4. As in the dense case, the bounds Δ_l^{sparse} provide accurate predictions for $k = 2, 3, 4$.

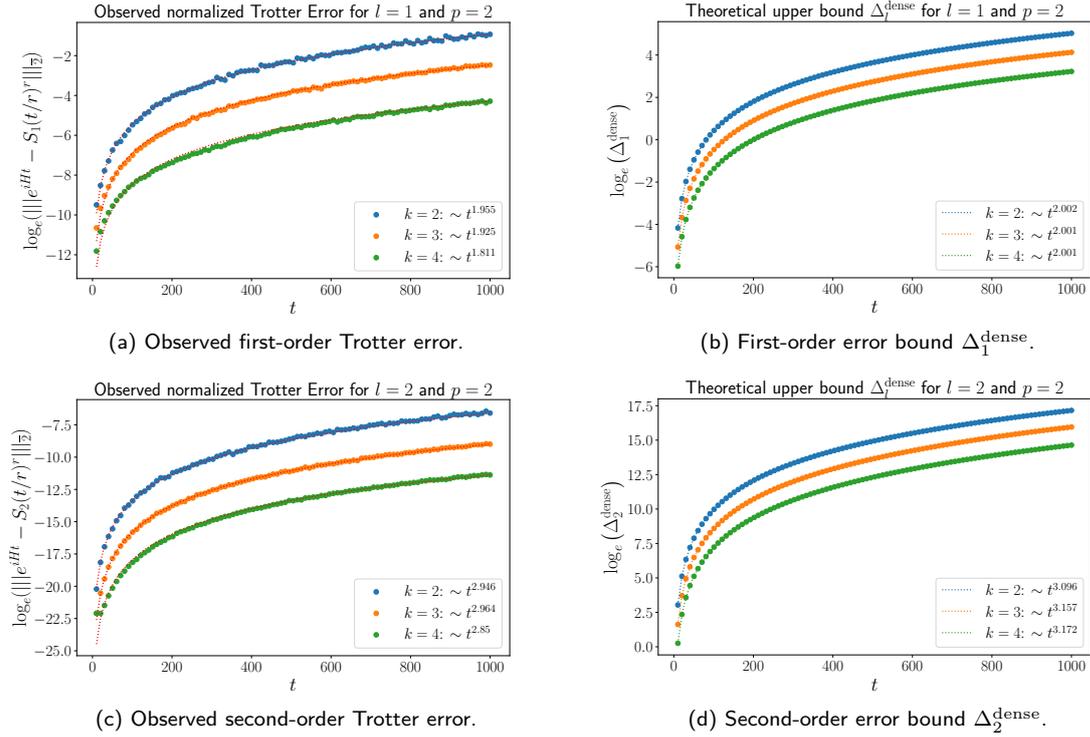


Figure 3: Observed normalized Trotter error and the theoretical upper bounds Δ_l^{dense} for $l = 1, 2$. We fix $n = 10$, and study the dense SYK models of localities $k = 2, 3, 4$ with $t \in [0, 1000]$.

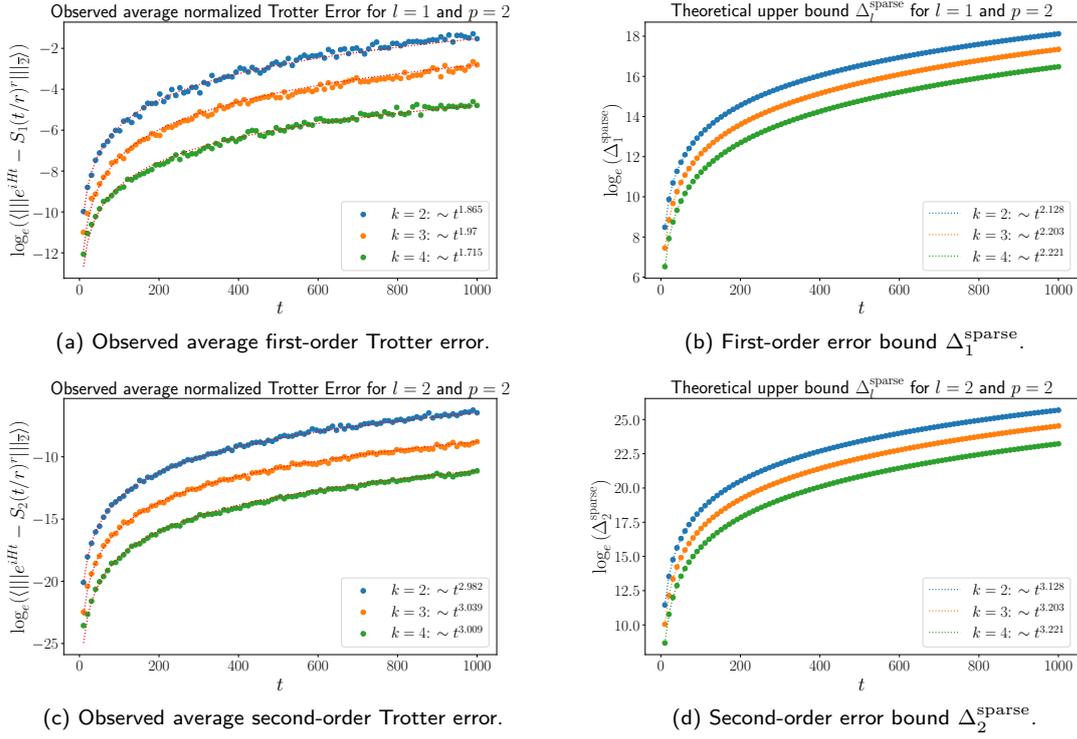


Figure 4: Observed average normalized Trotter error and the theoretical upper bounds Δ_t^{sparse} for $l = 1, 2$. We fix $n = 10$, and study the sparse SYK models of localities $k = 2, 3, 4$ with $t \in [0, 1000]$.

1.5 Structure of the paper

The paper is organized as follows. In Section 2, we establish notation and introduce several mathematical tools to study the Trotter error and the gate complexity of the associated simulation. After that, we first tackle the regular SYK model – we study the first-order and higher-order Trotter errors and their gate complexity in Sections 3 and 4 respectively. In Section 5, we study the higher-order Trotter error and gate complexity of the sparse SYK model. We conclude with a summary and open questions in Section 6.

1.6 Acknowledgements

We thank Chi-Fang Chen for explaining the proof methods of Ref. [CB24]. JH acknowledges funding from the Dutch Research Council (NWO) through a Veni grant (No.VI.Veni.222.331) and from the Quantum Software Consortium (NWO Zwaartekracht Grant No.024.003.037). MO was supported by an NWO Vidi grant (No.VI.Vidi.192.109) and by a National Growth Fund grant (NGF.1623.23.025). This work was primarily performed in the context of YC’s bachelor thesis at the University of Amsterdam.

2 Preliminaries

In this section, we formally describe the problem, provide sufficient mathematical background and introduce some key definitions and notation that will be used throughout.

2.1 The SYK and sparse SYK models

The *SYK model* is defined as a random k -local Majorana fermionic Hamiltonian⁵

$$H_{\text{SYK}} = i^{k(k-1)/2} \sum_{1 \leq i_1 < \dots < i_k \leq n} J_{i_1, \dots, i_k} \chi_{i_1} \cdots \chi_{i_k}, \quad (26)$$

consisting of (even) n Majorana fermions. The coefficients $J_{i_1, \dots, i_k} \in \mathbb{R}$ are i.i.d. Gaussian variables such that

$$\mathbb{E}(J_{i_1, \dots, i_k}) = 0, \quad \text{and} \quad \mathbb{E}(J_{i_1, \dots, i_k}^2) = \frac{(k-1)! \mathcal{J}^2}{kn^{k-1}} =: \sigma^2 \quad (27)$$

with a certain energy constant $\mathcal{J} < \infty$. The χ_i 's are the Majorana fermions that obey the following anti-commutation relation:

$$\{\chi_i, \chi_j\} = 2\delta_{ij}. \quad (28)$$

The *sparse SYK model* is a modified version of the SYK model proposed by Xu, Susskind, Su, and Swingle [XSSS20], that requires fewer resources to simulate, but still captures several essential properties [XSSS20, OGP24]. The model is obtained by removing each local interaction in the SYK model with probability $1 - p_B$, where p_B is usually defined as

$$p_B = \frac{\kappa n}{\binom{n}{k}} \quad (29)$$

with a constant $\kappa \geq 0$ (not to be confused with locality k) controlling the sparsity of the model. According to [XSSS20, OGP24], gravitational physics emerges for κ between $\frac{1}{4}$ and 4 at low temperatures.

Removing the local terms in the SYK model is equivalent to attaching a Bernoulli variable $B_{i_1, \dots, i_k} \in \{0, 1\}$ to each term such that $\mathbb{P}(B_{i_1, \dots, i_k} = 1) = p_B$. Hence, the sparse SYK model can be written as

$$H_{\text{SSYK}} = i^{k(k-1)/2} \sum_{1 \leq i_1 < \dots < i_k \leq n} B_{i_1, \dots, i_k} J_{i_1, \dots, i_k} \chi_{i_1} \cdots \chi_{i_k}. \quad (30)$$

To ensure that this model still has extensive energy, we renormalize the variance of the Gaussian couplings J_{i_1, \dots, i_k} as

$$\sigma^2 \mapsto \frac{\sigma^2}{p_B}, \quad (31)$$

to keep the model extensive.

2.2 Representation and norm of Majoranas

Evaluating the norm of the Trotter error of the SYK model inevitably requires evaluating the norm of the Majorana fermions χ_i appearing in equation (26). For even n , the Majoranas χ_1, \dots, χ_n can be understood as generators of the Clifford algebra $\text{Cl}(\mathbb{R}^{n/2})$. However, $\text{Cl}(\mathbb{R}^{n/2})$ is inherently not a normed space. Hence, when computing the norm of any element of this algebra, we implicitly compute the norm of its representation in $\text{GL}(2^{n/2}, \mathbb{C})$. A concrete representation we have in mind here is the Jordan–Wigner transformation⁶ [Nie05]

$$\rho_{\text{JW}} : \text{Cl}(\mathbb{R}^{n/2}) \longrightarrow \text{GL}(2^{n/2}, \mathbb{C}). \quad (32)$$

⁵For qubit Hamiltonians the underlying n -qubit Hilbert space $(\mathbb{C}^2)^{\otimes n}$ has tensor product structure, and a term is “ k -local” if it acts as identity on all but k qubits. For fermionic Hamiltonians the underlying Hilbert space need not have such structure, and the notion of locality is different – it refers to the number of Majoranas involved in a given term.

⁶One can choose another representation, as long as it satisfies the two properties in equation (33).

It is a representation of $\text{Cl}(\mathbb{R}^{n/2})$ because it preserves anti-commutativity and normalization of the generators χ_i :

$$\rho(\{\chi_i, \chi_j\}) = \{\rho(\chi_i), \rho(\chi_j)\}, \quad \rho(\chi_i)\rho(\chi_i) = I_{2^{n/2}} \quad (33)$$

for all $1 \leq i, j \leq n$. From now on, when we write the norm $\|\chi\|$ for any $\chi \in \text{Cl}(\mathbb{R}^{n/2})$, we implicitly mean the norm $\|\rho_{\text{JW}}(\chi)\|$ of its Jordan–Wigner representation. In other words, we have introduced a norm on $\text{Cl}(\mathbb{R}^{n/2})$ through a norm on $\text{GL}(2^{n/2}, \mathbb{C})$.

2.3 Notation for local Hamiltonians

The summation indices of the SYK and sparse SYK models can be cumbersome when computing product formula approximations of these models. Hence, in this section we introduce an alternative way of indexing terms in a local Hamiltonian by assigning an arbitrary ordering to hyperedges so that we can index terms by integers instead of subsets.

Let H be a k -local Hamiltonian acting on n particles labeled by $V = \{1, \dots, n\}$. We can write H as

$$H = \sum_{\eta \in E} H_\eta, \quad (34)$$

where $E \subseteq 2^V$ is some set of size- k subsets of V , and each H_η is a local interaction on particles $\eta \subseteq V$. This notation naturally induces a hypergraph (V, E) representing the local terms of H . The particles V are the vertices of the hypergraph, while E are the hyperedges. Each hyperedge denotes the particles on which the corresponding local interaction acts on. As we are using the hyperedges as labels for the local terms in the Hamiltonian, we denote the total number of local terms as $\Gamma := |E|$.

Throughout this paper, we will encounter situations where we need to partially sum over hyperedges in E , or sum over pairs of hyperedges in certain ordering. In those cases, the above notation can be cumbersome. To get a clearer overview for bookkeeping, we choose an arbitrary bijection

$$\phi : \{1, \dots, \Gamma\} \rightarrow E \quad (35)$$

to identify each hyperedge with a number (the explicit definition of ϕ does not matter). Observe that ϕ induces an ordering on E , such that for $\eta_1, \eta_2 \in E$,

$$\eta_1 > \eta_2 \iff \phi^{-1}(\eta_1) > \phi^{-1}(\eta_2). \quad (36)$$

In certain situations, we need to repeatedly sum over E . Hence, we can periodically extend ϕ to \mathbb{N} by considering the map

$$\gamma : \mathbb{N} \rightarrow E, \quad \gamma(r + q \cdot \Gamma) := \phi(r) \quad (37)$$

with $1 \leq r \leq \Gamma$ and $q \in \mathbb{Z}_{\geq 0}$. We will refer to this map γ as the *ordering map*. For example, if $f : E^4 \rightarrow \mathbb{C}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$, the ordering map allows us to write down sums such as

$$\sum_{i=1}^{3\Gamma} \sum_{j=i+1}^{4\Gamma} \sum_{\substack{k,l=1 \\ k+l \leq \Gamma}}^{\Gamma} f(H_{\gamma(g(i))}, H_{\gamma(j)}, H_{\gamma(g(k))}, H_{\gamma(l)}), \quad (38)$$

which would be quite cumbersome if we used summation over hyperedges. More importantly, we can write the SYK model in this notation as

$$H_{\text{SYK}} = \sum_{i=1}^{\Gamma} H_{\gamma(i)}, \quad (39)$$

where γ is any ordering map of our choice. The hyperedge $\gamma(i) = \{\gamma(i)_1, \dots, \gamma(i)_k\} \in E$ denotes the Majorana fermions acted upon by the local term

$$H_{\gamma(i)} = i^{k(k-1)/2} J_{\gamma(i)} (\chi_{\gamma(i)_1} \cdots \chi_{\gamma(i)_k})_{\text{lex}}, \quad (40)$$

where the product of Majorana fermions is ordered lexicographically to match the ordering of indices in the original definition of the SYK model in equation (26).

2.4 Trotter error and product formulas

Simulating the SYK model on a quantum computer means implementing its time-evolution operator $e^{itH_{\text{SYK}}}$ for any choice of the random couplings. In general, it is unclear how to exactly implement this matrix exponential, but there are several efficient approximation methods, one of them being product formulas. In essence, product formulas allow us to decompose an exponential of a sum of matrices into products of exponentials. Suppose H is a local qubit Hamiltonian defined as

$$H = \sum_{i=1}^{\Gamma} H_{\gamma(i)}, \quad (41)$$

with a certain ordering map γ defined in equation (37). Then we can decompose the unitary time evolution as a product plus some error

$$e^{iHt} = e^{iH_{\gamma(1)}t} e^{iH_{\gamma(2)}t} \cdots e^{iH_{\gamma(\Gamma)}t} + \mathcal{O}(t^2). \quad (42)$$

The product

$$S_1(t) = \prod_{i=1}^{\Gamma} e^{itH_{\gamma(i)}} \quad (43)$$

is called the *first-order product formula*. The error term $\mathcal{O}(t^2)$ associated with this formula is called the *first-order Trotter error*.

We can also write down product formulas of higher orders – the most prominent family of examples being the *Lie–Trotter–Suzuki formulas* [Suz91]. The *second-order formula* is defined as

$$S_2(t) = \prod_{i=\Gamma}^1 e^{iH_{\gamma(i)}t/2} \prod_{j=1}^{\Gamma} e^{iH_{\gamma(j)}t/2}, \quad (44)$$

and *higher-order formulas* are defined recursively:

$$S_{2p}(t) = S_{2p-2}(q_p t)^2 \cdot S_{2p-2}((1 - 4q_p)t) \cdot S_{2p-2}(q_p t)^2 \quad (45)$$

with $q_p = 1/(4 - 4^{1/(2p-1)})$. This recursive structure only defines product formulas of even order. Therefore, when we refer to higher-order product formulas in this work, we only consider even orders. In general, an l -th order product formula has Trotter error of leading order $\mathcal{O}(t^{l+1})$.

Another way to write down the l -th order product formula is to factor the recursive equation (45) into a product of Υ stages:

$$S_l(t) = \prod_{a=1}^{\Upsilon} \prod_{b=1}^{\Gamma} e^{ic(a,b)H_{\gamma(d(a,b))}t} \quad (46)$$

where $c(a,b) \in \mathbb{R}$ and $d(a,b)$ are chosen such that the above equals equation (45). By the recursive definition of l -th order product formulas, the number of stages is $\Upsilon = 2 \cdot 5^{\frac{l}{2}-1}$. We can write this even more compactly with $J = \Upsilon \cdot \Gamma$:

$$S_l(t) = \prod_{j=1}^J e^{ia_j H_{\gamma(b_j)}t} \quad (47)$$

where $a_j \in \mathbb{R}$ and $b_j \in \mathbb{N}$ are chosen such that the above expression is the same as equation (45).

In general, one uses $S_l(t/r)^r$ to approximate the Hamiltonian, instead of just $S_l(t)$. We refer to the *number of rounds* r of applying the product formulas as *Trotter number* for short. For further reading on this topic, we refer the reader to [CB24, CST⁺21, Suz91].

2.5 Probability statements and concentration inequalities

Because the SYK and sparse SYK models are random Hamiltonians, there are many ways to characterize what sufficiently small Trotter error means. The first one, appearing in [AJS24, XSSS20, GAEL⁺17], is the average case statement

$$\mathbb{E} \left(\|e^{iHt} - S_l(t/r)^r\| \right) < \epsilon, \quad (48)$$

where the expectation is over the randomness in H . Because this formulation of small Trotter error only covers the average case, we use the following two more precise statements given by [CB24]. The first one is stated using the operator norm:

$$\mathbb{P} \left(\|e^{iHt} - S_l(t/r)^r\| \geq \epsilon \right) < \delta. \quad (49)$$

The other one quantifies small Trotter error by looking at the difference between the actual time-evolution of an arbitrary fixed input state $|\psi\rangle$ and its approximated time evolution in l^2 -norm:

$$\mathbb{P} \left(\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l^2} \geq \epsilon \right) < \delta. \quad (50)$$

In addition to yielding more precise characterization of small Trotter error, these two probability statements also connect to the following two norms for random-valued operators A on a given finite-dimensional Hilbert space \mathcal{H} :

$$\| \|A\| \|_p := \left(\mathbb{E}(\|A\|_p^p) \right)^{\frac{1}{p}}, \quad (51)$$

$$\| \|A\| \|_{\text{fix},p} := \sup \left\{ \left(\mathbb{E}(\|A|\phi\rangle\|_{l^2}^p) \right)^{\frac{1}{p}} : |\phi\rangle \in \mathcal{H}, \| |\phi\rangle \|_{l^2} = 1 \right\}. \quad (52)$$

These norms obey certain uniform smoothness properties, which lead to better scaling in the Trotter error bounds and the corresponding gate complexities. Structuring our proofs around these properties ensures that the results that we derive in terms of $\| \cdot \|_p$ also hold in terms of $\| \cdot \|_{\text{fix},p}$. Hence, following the notation in [CB24], we will introduce the notation $\| \cdot \|_{*,p}$ ⁷ as a placeholder for both of these norms so that we can formulate our results more concisely. Whenever we state a result in the form

$$\| \|A\| \|_{*,p} \leq C \| \|B\| \|_{*,p} \quad (53)$$

for A, B operators on our Hilbert space and C a positive real constant, what we mean is that $\| \cdot \|_{*,p}$ can be replaced with either $\| \cdot \|_p$ or $\| \cdot \|_{\text{fix},p}$ throughout. In other words, both

$$\| \|A\| \|_p \leq C \| \|B\| \|_p, \quad \text{and} \quad \| \|A\| \|_{\text{fix},p} \leq C \| \|B\| \|_{\text{fix},p} \quad (54)$$

are true. For the scope of this work, we shall not discuss these uniform smoothness properties in detail. We refer interested readers to [TJ74, BCL94, HNWTW22, CB24] for more details.

To evaluate the probability statements in equations (49) and (50), we use the following concentration inequalities:

⁷The notation used in [CB24] is $\| \cdot \|_*$ which hides the p -dependence. Here, we will denote this dependence explicitly.

Lemma 2.1 (Concentration inequality for operator norm). *Let H be a Hamiltonian on a Hilbert space. Let $l \in \{1\} \cup 2\mathbb{N}$ be the order of the product formula approximating the Hamiltonian. If the Trotter error of this product formula is bounded by a positively valued function $\lambda(p, r)$, in the sense that*

$$\|e^{iHt} - S_l(t/r)^r\|_p \leq p\lambda(p, r)\|I\|_p \quad (55)$$

for all $p \geq 2$, then choosing the Trotter number r so that it obeys the inequality

$$\frac{\lambda\left(\log\left(e^2\|I\|_p^p/\delta\right), r\right)}{\epsilon} \leq \frac{1}{e \cdot \log\left(e^2\|I\|_p^p/\delta\right)} \quad (56)$$

ensures that

$$\mathbb{P}\left(\|e^{iHt} - S_l(t/r)^r\|_p \geq \epsilon\right) < \delta. \quad (57)$$

Lemma 2.2 (Concentration inequality for fixed input state). *Let H be a Hamiltonian on a Hilbert space of dimension D and $|\psi\rangle$ an arbitrary fixed input state in this space. Let $l \in \{1\} \cup 2\mathbb{N}$ be the order of the product formula approximating the Hamiltonian. If the Trotter error of this product formula is bounded by a positively valued function $\lambda(p, r)$, in the sense that*

$$\|e^{iHt} - S_l(t/r)^r\|_{\text{fix}, p} \leq p\lambda(p, r) \quad (58)$$

for all $p \geq 2$, then choosing the Trotter number r so that it obeys the inequality

$$\frac{\lambda\left(\log(e^2/\delta), r\right)}{\epsilon} \leq \frac{1}{e \cdot \log(e^2/\delta)} \quad (59)$$

ensures that

$$\mathbb{P}\left(\|(e^{iHt} - S_l(t/r)^r)|\psi\rangle\|_{l_2} \geq \epsilon\right) < \delta. \quad (60)$$

The proof of these lemmas can be found in Appendix A. We will refer to equations (55) and (58) as concentration inequalities. The positively-valued function $\lambda(p, r)$ typically takes the form of a positive rational power function of p and a negative rational power function of r , obtained from explicitly evaluating the representations of Trotter error defined in Section 2.7.

2.6 Gate complexity

After finding the Trotter number r through the concentration inequalities (depending on the probability statement we choose), we can compute the number of local exponentials in the corresponding approximation $S_l(t/r)^r$ as

$$C = \Upsilon\Gamma r, \quad (61)$$

where Υ is the number of stages and Γ the number of local terms, see equation (46). However, C is not the actual gate complexity⁸ of simulating a fermionic Hamiltonian. As we mentioned earlier, mapping fermions to qubits inevitably introduces extra overhead for the implementation of each local exponential. For instance, one could consider the standard Jordan–Wigner transform that gives each local exponential an $\mathcal{O}(n)$ overhead. In that case, the actual gate complexity should scale with

$$G = \mathcal{O}(n \cdot \Upsilon\Gamma r). \quad (62)$$

Fortunately, using the ternary tree mapping [JKMN20] incurs only a $\log(n)$ overhead for each local exponential, giving a total gate complexity that scales with

$$G = \mathcal{O}(\log(n) \cdot \Upsilon\Gamma r). \quad (63)$$

⁸We defined gate complexity as the number of gates required to simulate the Hamiltonian such that the Trotter error is sufficiently small, as characterized by either equation (49) or (50).

Because $\log(n)$ is a slow-growing function, we opt to omit it when writing down the gate complexity. To distinguish from the actual gate complexity, our result will be reported as

$$G = \tilde{O}(\Upsilon\Gamma r), \quad (64)$$

but the reader should always keep in mind the overhead from fermion-to-qubit mapping.

2.7 Error representation

We will need a more nuanced error representation to obtain better bounds on the Trotter error $\|e^{iHt} - S_l(t/r)^r\|$. Instead of measuring the error by a single number, we will consider an operator $\mathcal{E}(t)$ that captures the deviation from the exact evolution e^{iHt} at any given time t . More specifically, we will use the exponential form of the error from [CST⁺21, CB24] to capture the deviation of the product formula (47) at any given t :

$$S_l(t) = \prod_{j=1}^J e^{ia_j H_{\gamma(b_j)} t} = \exp_{\mathcal{T}} \left(i \int (\mathcal{E}(t) + H) dt \right), \quad (65)$$

where the right-hand side involves an ordered exponential. Hence, we can think of every product formula as the time evolution of our Hamiltonian H with some time-dependent error $\mathcal{E}(t)$ added to it. We call $\mathcal{E}(t)$ the *error generator*, and it takes the following form [CB24, CST⁺21]:

$$\mathcal{E}(t) := \sum_{j=1}^J \left(\prod_{k=j+1}^J e^{a_k \mathcal{L}_{\gamma(b_k)} t} [a_j H_{\gamma(b_j)}] - a_j H_{\gamma(b_j)} \right), \quad (66)$$

where $\mathcal{L}_{\gamma(j)}$ sends an operator O to the commutator with $H_{\gamma(j)}$:

$$\mathcal{L}_{\gamma(j)}[O] := i[H_{\gamma(j)}, O]. \quad (67)$$

This error generator will be our primary target when deriving bounds for the Trotter error.

2.8 Random matrix polynomials

A very useful method for the study of higher-order error generator is to view it as *random matrix polynomials* of local terms in the Hamiltonian. Hence, we introduce some tools for treating random matrix polynomials in this subsection.

We define a degree- g Gaussian random matrix polynomial of m matrix variables X_1, \dots, X_m as

$$F(X_m, \dots, X_1) = \sum_{\mathbf{i}} T_{\mathbf{i}} X_{i_1} \cdots X_{i_g} \quad (68)$$

where $\sum_{\mathbf{i}}$ is the abbreviated notation for the summation

$$\sum_{\mathbf{i}} := \sum_{i_1=1}^m \cdots \sum_{i_g=1}^m, \quad (69)$$

and $T_{\mathbf{i}} := T_{(i_1, \dots, i_g)}$ are scalar coefficients. For notation here, we will consistently use bold text to denote vectors. In this case \mathbf{j} is the vector of the summation indices (i_1, \dots, i_g) . For simplicity, we pose that the variables X_1, \dots, X_m are independent from each other, and each X_i can be written as

$$X_i = g_i K_i \quad (70)$$

with g_i a standard Gaussian variable and K_i a deterministic matrix with bounded norm $\|K_i\| \leq \sigma_i$. We refer to the set that collects the subscripts of the random variables F depends on as the index set \mathcal{I} . In this case, we have

$$\mathcal{I} := \{1, 2, 3, \dots, m\} = [m]. \quad (71)$$

To study the concentration bounds on these polynomials, [CB24, Proposition VII.2.1] proposed the following (sum of squares) uniform smoothing inequality:

Theorem 2.3. *Let $F(X_{i_m}, \dots, X_{i_1})$ be a random matrix polynomial, with index set $\mathcal{I} = \{i_1, \dots, i_m\}$ and each X_i is independent. Then, for $p \geq 2$ and $C_p = p - 1$,*

$$\|F\|_p^2 \leq \sum_{S \subseteq \mathcal{I}} C_p^{|S|} \|[F]_S\|_p^2 \quad (72)$$

where

$$[F]_S = \prod_{s \in S} (1 - \mathbb{E}_s) \prod_{s' \in S^c} \mathbb{E}_{s'}[F], \quad (73)$$

with \mathbb{E}_s being the entry-wise expectation applied only on X_s . The notation S^c is the complement $\mathcal{I} \setminus S$.

We refer to $[\dots]_S$ as the smoothing operator. This operator has a central part given by

$$(1 - \mathbb{E})_S[\cdot] := \prod_{s \in S} (1 - \mathbb{E}_s)[\cdot] \quad (74)$$

and an outer part given by

$$\mathbb{E}_{S^c}[\cdot] := \prod_{s' \in S^c} \mathbb{E}_{s'}[\cdot]. \quad (75)$$

Using our notation, the smoothing operator could be written as

$$[\cdot]_S = (1 - \mathbb{E})_S \circ \mathbb{E}_{S^c}[\cdot]. \quad (76)$$

We would like to apply Theorem 2.3 to the random matrix polynomial F defined in equation (68). This requires us to identify, for each subset $S \subseteq \mathcal{I}$, all terms $T_i X_{i_1} \cdots X_{i_g}$ in the polynomial F such that the corresponding smoothing operator is non-vanishing, i.e.

$$[T_i X_{i_1} \cdots X_{i_g}]_S \neq 0. \quad (77)$$

Notice that the Gaussian matrices X_i can appear with powers up to g in F , which makes them difficult to handle when applying the smoothing operator. To address this, we first expand each X_i into a product of Rademacher variables.

2.8.1 Rademacher expansion

Recall that the central limit theorem tells us that a standard Gaussian variable g_i can be written as the limit of a sum of i.i.d. Rademacher variables:

$$g_i = \lim_{N \rightarrow \infty} \sum_{j=1}^N \frac{\epsilon_{i,j}}{\sqrt{N}}, \quad (78)$$

where $\epsilon_{i,j} \in \{\pm 1\}$ are Rademacher variables. Taking N to be a sufficiently large integer motivates us to express each X_i as a sum of Rademacher variables $Y_{i,j}$:

$$X_i \approx \sum_{j=1}^N \frac{\epsilon_{i,j}}{\sqrt{N}} K_i =: \sum_{j=1}^N Y_{i,j}. \quad (79)$$

Under this expansion, we regard F as a polynomial in the variables

$$\begin{aligned} F &= F(Y_{1,1}, Y_{1,2}, \dots, Y_{1,N}, \dots, Y_{m,1}, Y_{m,2}, \dots, Y_{m,N}) \\ &:= \sum_{\mathbf{i}} \sum_{\mathbf{j}} T_{(i_1, \dots, i_g)} Y_{i_1, j_1} Y_{i_2, j_2} \dots Y_{i_g, j_g}. \end{aligned} \quad (80)$$

Although this representation makes F appear more complicated, it offers a key simplification: the non-subleading terms of F in the large- N limit contain Rademacher variables with exponents only in $\{0, 1, 2\}$, rather than arbitrary powers up to g . This significantly simplifies the application of the smoothing operator. To illustrate this subleading argument, we consider the following example.

For the first non-trivial case $m = 3$ (the analysis for $m > 3$ is identical), we have

$$\sum_{j_1=1}^N \sum_{j_2=1}^N \sum_{j_3=1}^N Y_{1,j_1} Y_{1,j_2} Y_{1,j_3}. \quad (81)$$

How many terms of the form Y_{1,j_i}^3 are present in this sum? The answer is N , corresponding to $j_1 = j_2 = j_3$. Summing over these terms gives

$$\sum_{j_1=1}^N Y_{1,j_1}^3 = \sum_{j_1=1}^N \frac{\epsilon_{1,j_1}^3}{N^{3/2}} K_1^3. \quad (82)$$

Since the Rademacher variable ϵ_{1,j_1}^3 only takes values ± 1 , we obtain the inequality

$$\left| \sum_{j_1=1}^N \frac{\epsilon_{1,j_1}^3}{N^{3/2}} \right| \leq \sum_{j_1=1}^N \frac{1}{N^{3/2}}. \quad (83)$$

Simplifying yields

$$\left| \sum_{j_1=1}^N \frac{\epsilon_{1,j_1}^3}{N^{3/2}} \right| \leq \frac{1}{N^{1/2}}. \quad (84)$$

Clearly, this expression vanishes as $N \rightarrow \infty$. Hence, each Rademacher variable in a non-subleading term of F , say Y_{i,j_i} , may only have exponent 0, 1, or 2.

2.8.2 Resummation by subsets

While this provides useful restrictions on the exponents of the Rademacher variables, it also introduces some complications: the index set of F becomes $\mathcal{I} = [m] \times [N]$, rather than simply $[m]$, because we now regard F as a polynomial depending on the variables $Y_{1,1}, \dots, Y_{m,N}$. Consequently, the subset S in the smoothing operator $[\dots]_S$ must now be taken from the Cartesian product $[m] \times [N]$. Given a subset $S \subseteq [m] \times [N]$, we can rewrite F as follows:

$$F = \sum_{\substack{x+y=g \\ x,y \in \{0, \dots, g\}}} \sum_{s_1, \dots, s_x \in S} \sum_{s'_1, \dots, s'_y \in S^c} \sum_{\pi \in \mathcal{S}_g / \mathcal{S}_x \times \mathcal{S}_y} T_{\pi[\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_y)]} \pi \left[Y_{s_1} \dots Y_{s_x} Y_{s'_1} \dots Y_{s'_y} \right]. \quad (85)$$

Let us unpack the notation. Each term in F takes the form

$$\pi \left[Y_{s_1} \dots Y_{s_x} Y_{s'_1} \dots Y_{s'_y} \right], \quad (86)$$

where $s_1, \dots, s_x \in S$ and $s'_1, \dots, s'_y \in S^c$. The sums over x and y specify how many terms in the product $Y_{s_1} \dots Y_{s_x} Y_{s'_1} \dots Y_{s'_y}$ have subscripts in S and S^c , respectively. The permutation $\pi \in \mathcal{S}_g$ reorders the positions of the Rademacher variables in the product to account for non-commutativity

of the random variables. To avoid double counting, we exclude permutations that only reorder Y_{s_1}, \dots, Y_{s_x} among themselves and/or $Y_{s'_1}, \dots, Y_{s'_y}$ among themselves, since such symmetries are already present in the summations over s_1, \dots, s_x and s'_1, \dots, s'_y . Thus, we sum over the quotient set $\mathcal{S}_g/\mathcal{S}_x \times \mathcal{S}_y$, which removes redundant symmetries from \mathcal{S}_g . We also extend the definition of the coefficients T so that they vanish whenever their subscript $\pi[\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_y)]$ does not correspond to a nonzero term in F .

To match the subscripts of coefficients $T(\dots)$, we introduce the projection from a direct product to the first set $\text{pr}_1 : A \times B \rightarrow A$ defined by

$$\text{pr}_1(i, j) = i \in A. \quad (87)$$

The bold symbol \mathbf{pr}_1 denotes a vector of projectors $(\text{pr}_1, \dots, \text{pr}_1)$ with g entries. Hence,

$$\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_y) = (\text{pr}_1 s_1, \dots, \text{pr}_1 s_x, \text{pr}_1 s'_1, \dots, \text{pr}_1 s'_y). \quad (88)$$

Because we have separated the subscripts in parts that belong to S and parts that belong to S^c , the action of the smoothing operator is simply

$$[F]_S = \sum_{x+y=g} \sum_{s_1, \dots, s_x \in S} \sum_{s'_1, \dots, s'_y \in S^c} \sum_{\pi \in \mathcal{S}_g/\mathcal{S}_x \times \mathcal{S}_y} T_{\pi[\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_y)]} \pi \left[(1 - \mathbb{E})_S [Y_{s_1} \cdots Y_{s_x}] \mathbb{E}_{S^c} [Y_{s'_1} \cdots Y_{s'_y}] \right] \quad (89)$$

Evaluating the above summations and combining it with Theorem 2.3, we get the following result

Theorem 2.4. For $p \geq 2$ and $C_p = p - 1$,

$$\|F\|_{*,p}^2 \leq g^g C_p^g \sum_{\mathbf{w}} \left(\sum_{\mathbf{v}} \sum_{\pi \in \mathcal{S}_g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i+2v_i} \right)^2 \|I\|_{*,p}^2, \quad (90)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^m$ are such that $|\mathbf{w}| + 2|\mathbf{v}| = g$.⁹ The unary encoding $\boldsymbol{\eta}$ is defined as

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := (\underbrace{1, \dots, 1}_{a_1 \text{ times}}, \underbrace{2, \dots, 2}_{a_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{a_m \text{ times}}, \underbrace{1, \dots, 1}_{b_1 \text{ times}}, \underbrace{2, \dots, 2}_{b_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{b_m \text{ times}}) \quad (91)$$

for $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$.

This is a significant result originally from the work of [CB24, Theorem VII.3] with a slight variation of prefactors compared to their statement. However, we prove this result via a slightly different route by considering the resummation in equation (89). For technical details, we refer the reader to Appendix C. Intuitively, Theorem 2.4 upper bounds the square of the norm of a Gaussian random matrix polynomial F by a sum of squares of its individual terms given by

$$\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}: |\mathbf{w}|+2|\mathbf{v}|=g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i+2v_i}. \quad (92)$$

Therefore, this theorem provides us a square root improvement compared to triangle inequality for upper bounding the norm of a random matrix polynomial. Since the SYK model is Gaussian random, we will use this result to evaluate the expected Schatten norm of higher-order Trotter errors as we are going to see in the next section.

2.8.3 Remark for notations

Most of our notation follows closely to that of [CB24]. In particular, the norms and error representations follow their conventions. Since Theorem 2.4 is a slight variant of Theorem VII.3 in [CB24],

⁹If $\mathbf{u} = (u_1, \dots, u_m) \in \mathbb{Z}_{\geq 0}^m$ is a vector then $|\mathbf{u}| := \sum_{i=1}^m |u_i|$.

we follow the same notation for the coefficients $T_{\pi(\dots)}$ and the vector summations over \mathbf{w}, \mathbf{v} take over the role of \mathbf{u}, \mathbf{v} in the original statement.

We also use our own notation apart from writing down local Hamiltonians. For instance, we use $[N]$ to denote the set $\{1, \dots, N\}$ for any positive integer N . Unless explicitly stated otherwise, all the vectors inside summations should be interpreted as non-negatively valued integer vectors. We shall introduce more notations as we progress through the following chapters.

3 First-order Trotter error of the SYK model

In this section, we will derive a bound for the first-order Trotter error of the SYK model, and estimate the corresponding gate complexity through the concentration inequality (55). To derive the error bound, we use the following result in [CB24, Lemma VI.2]: for $p \geq 2$ and $C_p = p - 1$,

$$\|\mathcal{E}(t)\|_{*,p}^2 \leq 2C_p \sum_{i=2}^{\Gamma} \left[4C_p t^2 \sum_{j=1}^{i-1} \|[H_{\gamma(j)}, H_{\gamma(i)}]\|_{*,p}^2 + \left(\sum_{j=1}^{i-1} \frac{t^2}{2} \|[H_{\gamma(j)}, [H_{\gamma(j)}, H_{\gamma(i)}]]\|_{*,p} \right)^2 \right], \quad (93)$$

with an arbitrary choice of ordering map γ , see equation (37).

Although this result is derived in the context of qubit Hamiltonians, its proof generalizes to fermionic Hamiltonians such as the SYK model, with a slightly modified interpretation of the expected Schatten p -norm. Because the Majorana fermions are not normed objects, one could first transform them to qubit operators using, for instance, the Jordan–Wigner transformation. The Schatten norm is then taken over these qubit operators transformed from the Majoranas.

3.1 Error bound

Equation (93) can be explicitly evaluated because we can directly compute the commutator using the commutation relation (28) of Majorana fermions. From the generator, we can derive an error bound on the first-order Trotter error of the SYK model. We state our result below:

Theorem 3.1. *The first-order Trotter error of the SYK model is bounded by*

$$\frac{\|e^{iHt} - S_1(t/r)^r\|_{*,p}}{\|I\|_{*,p}} \leq \Delta_1(n, k, t, r, p), \quad (94)$$

where

$$\Delta_1(n, k, t, r, p) := 4\sqrt{2}p^2\sigma^2 \sqrt{\binom{n}{k}} Q(n, k) t^2 \left[\frac{1}{2r} + \sigma \sqrt{Q(n, k)} \frac{t}{3r^2} \right]. \quad (95)$$

Here σ^2 is the variance of the coupling coefficients, see equation (27), and $Q(n, k)$ is defined as

$$Q(n, k) := \begin{cases} \sum_{\substack{s \text{ odd}, \\ s \geq \mu}}^{k-1} \binom{n-k}{k-s} \binom{k}{s} & \text{if } k \text{ is even,} \\ \sum_{\substack{s \text{ even}, \\ s \geq \mu}}^{k-1} \binom{n-k}{k-s} \binom{k}{s} & \text{if } k \text{ is odd,} \end{cases} \quad (96)$$

where $\mu := \max\{0, k - (n - k)\}$ is the minimal overlap of two hyperedges.

Proving the above result revolves around evaluating the commutators in equation (93). To do that, we use the following anti-commutation lemma for local terms in the SYK model. The proof of this lemma is presented in Appendix B. The same result is also recorded in [GGJV18].

Lemma 3.2. *Let $H = \sum_{i=1}^{\Gamma} H_{\gamma(i)}$ be the k -local SYK model with $\gamma : [\Gamma] \rightarrow E$ an arbitrary ordering of hyperedges. Then for all $1 \leq i < j \leq \Gamma$,*

$$H_{\gamma(i)}H_{\gamma(j)} = (-1)^{k+m}H_{\gamma(j)}H_{\gamma(i)}, \quad (97)$$

where $m = |\gamma(i) \cap \gamma(j)|$ is the overlap between hyperedges $\gamma(i)$ and $\gamma(j)$.

The above lemma leads to the following commutation relations if k is even:

$$[H_{\gamma(i)}, H_{\gamma(j)}] = \begin{cases} 0 & \text{if } |\gamma(i) \cap \gamma(j)| \text{ is even,} \\ 2H_{\gamma(i)}H_{\gamma(j)} & \text{else,} \end{cases} \quad (98)$$

$$[H_{\gamma(i)}, [H_{\gamma(i)}, H_{\gamma(j)}]] = \begin{cases} 0 & \text{if } |\gamma(i) \cap \gamma(j)| \text{ is even,} \\ 4H_{\gamma(i)}^2H_{\gamma(j)} & \text{else,} \end{cases} \quad (99)$$

and if k is odd:

$$[H_{\gamma(i)}, H_{\gamma(j)}] = \begin{cases} 0 & \text{if } |\gamma(i) \cap \gamma(j)| \text{ is odd,} \\ 2H_{\gamma(i)}H_{\gamma(j)} & \text{else,} \end{cases} \quad (100)$$

$$[H_{\gamma(i)}, [H_{\gamma(i)}, H_{\gamma(j)}]] = \begin{cases} 0 & \text{if } |\gamma(i) \cap \gamma(j)| \text{ is odd,} \\ 4H_{\gamma(i)}^2H_{\gamma(j)} & \text{else.} \end{cases} \quad (101)$$

Proof (of Theorem 3.1). Note from the above expressions that if $2 \leq i \leq \Gamma$ and k is even then

$$\sum_{j=1}^{i-1} \|\| [H_{\gamma(j)}, H_{\gamma(i)}] \|\|_{*,p}^2 \leq \sum_{\substack{j=1 \\ j \neq i}}^{\Gamma} \|\| [H_{\gamma(j)}, H_{\gamma(i)}] \|\|_{*,p}^2 \quad (102)$$

$$\leq 4M_1^2 \left\{ \sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=1}}^{\Gamma} 1 + \sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=3}}^{\Gamma} 1 + \cdots + \sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=k-1}}^{\Gamma} 1 \right\}, \quad (103)$$

while if k is odd then

$$\sum_{j=1}^{i-1} \|\| [H_{\gamma(j)}, H_{\gamma(i)}] \|\|_{*,p}^2 \leq 4M_1^2 \left\{ \sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=0}}^{\Gamma} 1 + \sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=2}}^{\Gamma} 1 + \cdots + \sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=k-1}}^{\Gamma} 1 \right\}, \quad (104)$$

with

$$M_1 := \frac{1}{\|\| I \|\|_{*,p}} \max \left\{ \|\| H_{\gamma(k)} H_{\gamma(l)} \|\|_{*,p} : k, l \in [\Gamma], k \neq l \right\}. \quad (105)$$

Observe that for any $0 \leq s \leq k-1$,

$$\sum_{\substack{j=1 \\ |\gamma(j) \cap \gamma(i)|=s}}^{\Gamma} 1 = \binom{n-k}{k-s} \binom{k}{s}, \quad (106)$$

since we can construct the hyperedge $\gamma(j)$ in two steps: first we choose $k-s$ elements out of $n-k$ that do not belong to $\gamma(i)$, and then we choose s additional elements from $\gamma(i)$. Therefore,

$$\sum_{j=1}^{i-1} \|\| [H_{\gamma(j)}, H_{\gamma(i)}] \|\|_{*,p}^2 \leq 4M_1^2 Q(n, k) \quad (107)$$

with $Q(n, k)$ defined in equation (96).¹⁰ Similarly, observe that for $2 \leq i \leq \Gamma$,

$$\left(\sum_{j=1}^{i-1} \frac{\tau^2}{2} \lll [H_{\gamma(j)}, [H_{\gamma(j)}, H_{\gamma(i)}]] \lll_{*,p} \right)^2 \leq (2\tau^2 M_2 Q(n, k))^2, \quad (108)$$

where

$$M_2 := \frac{1}{\lll I \lll_{*,p}} \max \left\{ \lll H_{\gamma(k)}^2 H_{\gamma(l)} \lll_{*,p} : k, l \in [\Gamma], k \neq l \right\}. \quad (109)$$

Putting all of the above together,

$$\lll \mathcal{E}(\tau) \lll_{*,p}^2 \leq 2 \lll I \lll_{*,p}^2 C_p \sum_{i=2}^{\Gamma} [16C_p \tau^2 M_1^2 Q(n, k) + 4\tau^4 M_2^2 Q(n, k)^2]. \quad (110)$$

It remains to upper bound the constants M_1 and M_2 . When we map the SYK model to a qubit Hamiltonian through, for instance, the Jordan–Wigner transformation, each local term of the SYK model can be written as an i.i.d. standard Gaussian coefficient multiplied by a deterministic Pauli string:

$$H_{\gamma(i)} \mapsto i^{k(k-1)/2} J_{\gamma(i)} \sigma_{\gamma(i)}. \quad (111)$$

For every Pauli string, its $*$ -norm is exactly equal to $\lll I \lll_{*,p}$ by unitarity. Therefore,

$$M_1 = \max \left\{ \left(\mathbb{E}(|J_{\gamma(k)} J_{\gamma(l)}|^p) \right)^{\frac{1}{p}} : k, l \in [\Gamma], k \neq l \right\}, \quad (112)$$

$$M_2 = \max \left\{ \left(\mathbb{E}(|J_{\gamma(k)}^2 J_{\gamma(l)}|^p) \right)^{\frac{1}{p}} : k, l \in [\Gamma], k \neq l \right\}. \quad (113)$$

Since all $J_{\gamma(i)}$ are i.i.d., let \tilde{J} denote a Gaussian coefficient as in equation (27). Then M_1 and M_2 can be bounded using central absolute moments of the Gaussian distribution \tilde{J} :

$$M_1 = (\mathbb{E}(|\tilde{J}|^p))^{1/p} \cdot (\mathbb{E}(|\tilde{J}|^p))^{1/p} \leq (\sigma^p (p-1)!!)^{2/p} \leq \sigma^2 p, \quad (114)$$

$$M_2 = (\mathbb{E}(|\tilde{J}|^{2p}))^{1/p} \cdot (\mathbb{E}(|\tilde{J}|^p))^{1/p} \leq (\sigma^{2p} (2p-1)!!)^{1/p} \cdot (\sigma^p (p-1)!!)^{1/p} \leq \sigma^2 (2p) \cdot \sigma \sqrt{p}, \quad (115)$$

where $\sigma = \mathcal{J} \sqrt{\frac{(k-1)!}{kn^{k-1}}}$ is the standard deviation of \tilde{J} .

Inserting this back into equation (110),

$$\lll \mathcal{E}(\tau) \lll_{*,p}^2 \leq 2 \lll I \lll_{*,p}^2 C_p \sum_{i=2}^{\Gamma} [16C_p \tau^2 \sigma^4 p^2 Q(n, k) + 16\tau^4 \sigma^6 p^3 Q(n, k)^2] \quad (116)$$

$$\leq 32 \lll I \lll_{*,p}^2 p^4 \binom{n}{k} [\tau^2 \sigma^4 Q(n, k) + \tau^4 \sigma^6 Q(n, k)^2] \quad (117)$$

where in the second inequality we used $C_p < p$ to pull all the factors of p to the front, as well as $\Gamma = \binom{n}{k}$ to remove the sum.

Hence, by subadditivity of the square root,

$$\lll \mathcal{E}(\tau) \lll_{*,p} \leq 4\sqrt{2} \lll I \lll_{*,p} p^2 \sigma^2 \sqrt{\binom{n}{k} Q(n, k) [\tau + \tau^2 \sigma \sqrt{Q(n, k)}]}. \quad (118)$$

¹⁰The same $Q(n, k)$ turns out to be important also in the higher-order case since it captures the commutation relation of the Majoranas in the SYK model.

If we divide the total evolution time t into r rounds, where each round has duration t/r , the error per round can be bounded using the triangle inequality:

$$\| \| e^{iHt/r} - S_1(t/r) \| \|_{*,p} \leq \int_0^{t/r} \| \| \mathcal{E}(\tau) \| \|_{*,p} d\tau \quad (119)$$

$$\leq 4\sqrt{2} \| \| I \| \|_{*,p} p^2 \sigma^2 \sqrt{\binom{n}{k} Q(n,k)} \left[\frac{t^2}{2r^2} + \sigma \sqrt{Q(n,k)} \frac{t^3}{3r^3} \right]. \quad (120)$$

The total first-order Trotter error is then bounded by

$$\| \| e^{iHt} - S_1(t/r)^r \| \|_{*,p} \leq r \cdot \| \| e^{iHt/r} - S_1(t/r) \| \|_{*,p} \quad (121)$$

$$\leq 4\sqrt{2} \| \| I \| \|_{*,p} p^2 \sigma^2 \sqrt{\binom{n}{k} Q(n,k)} t^2 \left[\frac{1}{2r} + \sigma \sqrt{Q(n,k)} \frac{t}{3r^2} \right] \quad (122)$$

which concludes the proof. \square

3.2 Gate complexity

Now that we have a bound on the first-order Trotter error, we can bound the associated gate complexity.

Corollary 3.2.1. *When simulating the k -local SYK model (26) with r rounds of the first-order product formula, the error bound*

$$\mathbb{P}(\| \| e^{iHt} - S_1(t/r)^r \| \| \geq \epsilon) < \delta \quad (123)$$

can be ensured with gate complexity

$$G = \tilde{\mathcal{O}} \left(n^{g(k)} (n + \log(e^2/\delta))^2 \frac{\mathcal{J}^2 t^2}{\epsilon} + n^{g(k)-\frac{1}{4}} (n + \log(e^2/\delta)) \frac{\mathcal{J}^{\frac{3}{2}} t^{\frac{3}{2}}}{\sqrt{\epsilon}} \right), \quad (124)$$

where \mathcal{J} controls the variance of the coupling coefficients, see equation (27), and

$$g(k) := \begin{cases} k + \frac{1}{2} & \text{if } k \text{ is even,} \\ k + 1 & \text{if } k \text{ is odd.} \end{cases} \quad (125)$$

In terms of n , the gate complexity G scales as

$$G \sim \begin{cases} n^{k+\frac{5}{2}} \mathcal{J}^2 t^2 & \text{if } k \text{ is even,} \\ n^{k+3} \mathcal{J}^2 t^2 & \text{if } k \text{ is odd.} \end{cases} \quad (126)$$

Proof. By Theorem 3.1, the first-order Trotter error is bounded by

$$\| \| e^{iHt} - S_1(t/r)^r \| \|_p \leq p \| \| I \| \|_p \lambda(p, r) \quad (127)$$

where

$$\lambda(p, r) := 4\sqrt{2} p \sigma^2 \sqrt{\binom{n}{k} Q(n,k)} t^2 \left[\frac{1}{2r} + \sigma \sqrt{Q(n,k)} \frac{t}{3r^2} \right]. \quad (128)$$

This gives us an explicit expression for the concentration inequality (55). By Lemma 2.1, finding a Trotter number r that ensures the error bound (123) is equivalent to finding r that solves the inequality

$$e4\sqrt{2} p_D^2 \cdot \sigma^2 \sqrt{\binom{n}{k} Q(n,k)} \frac{t^2}{\epsilon} \left[\frac{1}{2r} + \sigma \sqrt{Q(n,k)} \frac{t}{3r^2} \right] \leq 1 \quad (129)$$

where $p_D := \log(e^2 D/\delta)$ with $D := \|I\|_p^p$.

To ensure the error bound (123), we need

$$r \geq e4\sqrt{2}p_D^2 \cdot \sigma^2 \sqrt{\binom{n}{k} Q(n, k)} \frac{t^2}{2\epsilon} + \left(e4\sqrt{2}p_D^2 \cdot \sigma^3 \sqrt{\binom{n}{k} Q(n, k)} \frac{t^3}{3\epsilon} \right)^{\frac{1}{2}}. \quad (130)$$

Recall from Section 2.2 that for (even) n Majoranas $\|I\|_p$ denotes the Schatten p -norm on $\text{GL}(2^{n/2}, \mathbb{C})$, hence $D = \|I\|_p^p = 2^{n/2}$ and $p_D = \frac{n}{2} \log 2 + \log(e^2/\delta)$.

Recall from Section 2.6 that the gate complexity scales as $G = \tilde{\mathcal{O}}(\Upsilon \Gamma r)$, where $\Upsilon = \mathcal{O}(1)$ for first-order formulas, $\Gamma = \binom{n}{k}$ is the number of terms, and r is the number of rounds. Hence,

$$G = \tilde{\mathcal{O}} \left(\binom{n}{k} \cdot \left[e4\sqrt{2} \left(\frac{\log 2}{2} n + \log(e^2/\delta) \right)^2 \sigma^2 \sqrt{\binom{n}{k} Q(n, k)} \frac{t^2}{2\epsilon} + \left(e4\sqrt{2} \left(\frac{\log 2}{2} n + \log(e^2/\delta) \right)^2 \sigma^3 \sqrt{\binom{n}{k} Q(n, k)} \frac{t^3}{3\epsilon} \right)^{\frac{1}{2}} \right] \right). \quad (131)$$

To analyze the asymptotic scaling of G for fixed k , note that $\binom{n}{k} = \mathcal{O}(n^k)$. Moreover, note from equation (96) that $Q(n, k) = \mathcal{O}(n^{k-1})$ for even k while $Q(n, k) = \mathcal{O}(n^k)$ for odd k . Finally, note from equation (27) that $\sigma = \mathcal{O}(\mathcal{J}/\sqrt{n^{k-1}})$. Putting this all together,

$$G = \tilde{\mathcal{O}} \left(n^{g(k)} (n + \log(e^2/\delta))^2 \frac{\mathcal{J}^2 t^2}{\epsilon} + n^{g(k)-\frac{1}{4}} (n + \log(e^2/\delta)) \frac{\mathcal{J}^{\frac{3}{2}} t^{\frac{3}{2}}}{\sqrt{\epsilon}} \right) \quad (132)$$

where $g(k)$ is defined in equation (125). The asymptotic scaling of G is as stated in equation (126) since the first term in (132) is dominant. \square

Alternatively, we can consider the error bound in equation (50) for a fixed input state $|\psi\rangle$. In that case, we use Lemma 2.2 instead of Lemma 2.1, which leads us to the following result:¹¹

Corollary 3.2.2. *When simulating the k -local SYK model (26) with r rounds of the first-order product formula, the error bound*

$$\mathbb{P}(\| (e^{iHt} - S_1(t/r)^r) |\psi\rangle \|_{l_2} \geq \epsilon) < \delta \quad (133)$$

for an arbitrary input state $|\psi\rangle$ can be ensured with gate complexity

$$G = \tilde{\mathcal{O}} \left(n^{g(k)} (\log(e^2/\delta))^2 \frac{\mathcal{J}^2 t^2}{\epsilon} + n^{g(k)-\frac{1}{4}} (\log(e^2/\delta)) \frac{\mathcal{J}^{\frac{3}{2}} t^{\frac{3}{2}}}{\sqrt{\epsilon}} \right), \quad (134)$$

where \mathcal{J} controls the variance of the coupling coefficients, see equation (27), and $g(k)$ is defined in equation (125). In terms of n , the gate complexity G scales as

$$G \sim \begin{cases} n^{k+\frac{1}{2}} \mathcal{J}^2 t^2 & \text{if } k \text{ is even,} \\ n^{k+1} \mathcal{J}^2 t^2 & \text{if } k \text{ is odd.} \end{cases} \quad (135)$$

4 Higher-order Trotter error of the SYK model

In this section, we will extend our first-order Trotter error results from the Section 3 to higher-order Trotter error. As our main tool we will use the uniform smoothness Theorem 2.4 to evaluate the

¹¹Using Lemma 2.2 is ultimately the same as replacing the factors of $n + \log(e^2/\delta)$ with $\log(e^2/\delta)$.

error generator as random matrix polynomials. Based on the error generator, we will then derive bounds for higher-order Trotter error, from which we will derive the gate complexity for simulating the SYK Hamiltonian using higher-order product formulas.

In equation (65) in Section 2.7, we introduced a general exponential representation of the Trotter error, where we wrote the product formula as

$$\prod_{j=1}^J e^{ia_j H_{\gamma(b_j)} t} = \exp_{\mathcal{T}} \left(i \int (\mathcal{E}(t) + H) dt \right), \quad (136)$$

where the error generator $\mathcal{E}(t)$ defined in equation (66) takes the form

$$\mathcal{E}(t) := \sum_{j=1}^J \left(\prod_{k=j+1}^J e^{a_k \mathcal{L}_{\gamma(b_k)} t} [a_j H_{\gamma(b_j)}] - a_j H_{\gamma(b_j)} \right). \quad (137)$$

Here $\mathcal{L}_{\gamma(j)}[O] := i[H_{\gamma(j)}, O]$, see equation (67), and $J = \Gamma \cdot \Upsilon$ with Γ the number of terms in the k -local Hamiltonian and Υ the number of stages of the product formula.

Following the approach in [CB24], we can do a Taylor expansion on the error generator $\mathcal{E}(t)$ and collect terms by the degree of product of local terms $H_{\gamma(1)}, \dots, H_{\gamma(\Gamma)}$, which leads to

$$\mathcal{E}(t) = \sum_{g=l+1}^{g'-1} \mathcal{E}_g(t) + \mathcal{E}_{\geq g'}(t), \quad (138)$$

where g' is some cutoff and the sum starts at $l+1$ since the l -th order product formula is exact up to order l . The $\mathcal{E}_g(t)$ term is the g -th order error generator given by

$$\mathcal{E}_g(t) := \sum_{j=1}^J \sum_{g_j + \dots + g_{j+1} = g-1} \mathcal{L}_{\gamma(j)}^{g_j} \dots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{t^{g-1}}{g_j! \dots g_{j+1}!}, \quad (139)$$

and $\mathcal{E}_{\geq g'}(t)$ is the residual term given by

$$\begin{aligned} \mathcal{E}_{\geq g'}(t) &:= \sum_{j=1}^J \sum_{m=j+1}^J e^{\mathcal{L}_{\gamma(j)} t} \dots e^{\mathcal{L}_{\gamma(m+1)} t} \\ &\times \int_0^t d\xi \sum_{g_m + \dots + g_{j+1} = g'-1, g_m \geq 1} e^{\mathcal{L}_{\gamma(m)} \xi} \mathcal{L}_{\gamma(m)}^{g_m} \dots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{(t-\xi)^{g_m-1} t^{g'-g_m-1}}{(g_m-1)! g_{m-1}! \dots g_{j+1}!}. \end{aligned} \quad (140)$$

4.1 Error bound

We will proceed by separately evaluating the norms of \mathcal{E}_g and $\mathcal{E}_{\geq g'}$ for the SYK model, and then collecting the results together. We will try to write everything down in terms of $Q(n, k)$ as defined in equation (96). In the end, we get the following result:

Theorem 4.1. *The (even) l -th order Trotter error of the SYK model for $l \geq 2$ is bounded by*

$$\frac{\| \| e^{iHt} - S_l(t/r)^r \| \|_{*,p}}{\| \| I \| \|_{*,p}} \leq \Delta_l(n, k, t, r, p), \quad (141)$$

where

$$\Delta_l := \frac{\mathcal{C}(l) \sqrt{p\sigma} t}{\sqrt{Q(n, k)}} \left(\binom{n}{k} \left[\sqrt{p\sigma} \sqrt{Q(n, k)} \frac{t}{r} \right]^l + \binom{n}{k}^2 \left[\sqrt{p\sigma} \sqrt{Q(n, k)} \frac{t}{r} \right]^{l+1} \right), \quad (142)$$

and $D(l)$ is a factor only depending on l .

The rest of Section 4.1 constitutes a proof of this theorem.

Proof. When proving this result, we assume that k is even, as the odd k case follows the same argument. We shall first compute the g -th order error generator \mathcal{E}_g and the tail bound $\mathcal{E}_{\geq g'}$ separately, and then join them together to obtain the total error. Our steps follow largely the approach in [CB24], but with our own computation of coefficients specific to the SYK model.

4.1.1 g -th order error generator

In this subsection, we will upper bound

$$\|\mathcal{E}_g\|_{*,p} = \left\| \sum_{j=1}^J \sum_{g_j+\dots+g_{j+1}=g-1} \mathcal{L}_{\gamma(j)}^{g_j} \cdots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{t^{g-1}}{g_j! \cdots g_{j+1}!} \right\|_{*,p} \quad (143)$$

by using uniform smoothing in the form of Theorem 2.4.

As the summation indices are difficult to work with here, we consider the following symmetrization:

$$\|\mathcal{E}_g\|_{*,p} = t^{g-1} \left\| \sum_{j_{g-1}=1}^J \cdots \sum_{j_0=1}^J \tilde{\theta}_j \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \right\|_{*,p}, \quad (144)$$

where the sum is over $\mathbf{j} = (j_{g-1}, \dots, j_0) \in [J]^g$, and we have included coefficients $\tilde{\theta}_j \in [0, 1]$ that absorb the factorials $\frac{1}{g_j! \cdots g_{j+1}!}$ and reduce the range of the sum to match with equation (143). We will not write an explicit expression for these coefficients since later we will ignore them for simplicity anyway. All that matters is that $\tilde{\theta}_j \leq 1$ and they impose an extra summation constraint. As $J = \Upsilon \cdot \Gamma$, with $\Gamma = \binom{n}{k}$ the total number of local terms in the Hamiltonian, we have the following expression to bound:

$$\|\mathcal{E}_g\|_{*,p} \leq \Upsilon^g t^{g-1} \left\| \sum_{j_{g-1}=1}^{\Gamma} \cdots \sum_{j_0=1}^{\Gamma} \tilde{\theta}_j \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \right\|_{*,p}, \quad (145)$$

where the number of stages Υ accounts for repeated values of $\gamma(\mathbf{j})$ thanks to the periodicity of the ordering map γ , see equation (37).

We will think of the expression we are dealing with as a non-commutative polynomial F_g in the local terms H_i of the Hamiltonian. Following our discussion in Section 2.8, we write it as

$$F_g(H_{\gamma(\Gamma)}, \dots, H_{\gamma(1)}) = \sum_{j_{g-1}=1}^{\Gamma} \cdots \sum_{j_0=1}^{\Gamma} \tilde{\theta}_j \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \quad (146)$$

$$=: \sum_{j_{g-1}=1}^{\Gamma} \cdots \sum_{j_0=1}^{\Gamma} T_{\mathbf{j}} H_{\gamma(j_{g-1})} \cdots H_{\gamma(j_0)}. \quad (147)$$

While the coefficients $T_{\mathbf{j}}$ can be calculated explicitly, we will not need to do that since for applying Theorem 2.4 it is sufficient to only have an upper bound on their norm.

Note that the commutator chain $\mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}]$ is a polynomial in the local terms $H_{\gamma(j_{g-1})}, \dots, H_{\gamma(j_0)}$, with an overall pre-factor of i^g :

$$\mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] = i^g \sum_{\mathbf{k} \in \mathbf{j}^g} c_{\mathbf{j}, \mathbf{k}} H_{\gamma(k_g)} \cdots H_{\gamma(k_1)}, \quad (148)$$

where $c_{\mathbf{j},\mathbf{k}}$ are some coefficients and the sum runs over vectors $\mathbf{k} = (k_g, \dots, k_1)$ in $\mathbf{j}^g := (j_{g-1}, \dots, j_0)^g$. As each term $H_{\gamma(k_g)} \cdots H_{\gamma(k_1)}$ emerges from a commutator, its coefficient $c_{\mathbf{j},\mathbf{k}}$ is either zero or ± 1 (it is zero when $\mathbf{j} \neq \mathbf{k}$ as sets). Since the local terms $H_{\gamma(j_{g-1})}, \dots, H_{\gamma(j_0)}$ are proportional to Pauli strings, each pair of them either commutes or anti-commutes. Hence, every term $H_{\gamma(k_g)} \cdots H_{\gamma(k_1)}$ with a non-zero $c_{\mathbf{j},\mathbf{k}}$ can be reordered to the product $H_{\gamma(j_{g-1})} \cdots H_{\gamma(j_0)}$ that matches the ordering \mathbf{j} of the commutator chain in equation (148), with an extra pre-factor $d_{\mathbf{j},\mathbf{k}}$:

$$H_{\gamma(k_g)} \cdots H_{\gamma(k_1)} = d_{\mathbf{j},\mathbf{k}} H_{\gamma(j_{g-1})} \cdots H_{\gamma(j_0)}. \quad (149)$$

The pre-factor $d_{\mathbf{j},\mathbf{k}}$ is either $+1$ or -1 depending on the (anti-)commutation between the local terms. Therefore,

$$\mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)}[H_{\gamma(j_0)}] = i^g \left(\sum_{\mathbf{k} \in \mathbf{j}^g} c_{\mathbf{j},\mathbf{k}} d_{\mathbf{j},\mathbf{k}} \right) H_{\gamma(j_{g-1})} \cdots H_{\gamma(j_0)}. \quad (150)$$

Substituting this in equation (146), we can write the $T_{\mathbf{j}}$ introduced in equation (147) as

$$T_{\mathbf{j}} = i^g \tilde{\theta}_{\mathbf{j}} \sum_{\mathbf{k} \in \mathbf{j}^g} c_{\mathbf{j},\mathbf{k}} d_{\mathbf{j},\mathbf{k}}. \quad (151)$$

To give an upper bound of $T_{\mathbf{j}}$, notice that the sum in equation (148) vanishes if the commutator chain evaluates to zero. We can also remove the coefficients $\tilde{\theta}_{\mathbf{j}}$ from equation (151) since they are, by construction, at most one. Hence, by triangle inequality,

$$|T_{\mathbf{j}}| \leq \left(\sum_{\mathbf{k} \in \mathbf{j}^g} 1 \right) \cdot \mathbb{1}(\mathbf{j}) = g^g \mathbb{1}(\mathbf{j}), \quad (152)$$

where the indicator function $\mathbb{1}(\mathbf{j})$ defined for all $\mathbf{j} \in [J]^g$ enforces the commutation constraints:

$$\mathbb{1}(\mathbf{j}) := \begin{cases} 1 & \text{if } \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)}[H_{\gamma(j_0)}] \neq 0, \\ 0 & \text{else.} \end{cases} \quad (153)$$

Lastly, before we can apply Theorem 2.4, we rewrite each local term $H_{\gamma(i)}$ in the SYK Hamiltonian (39) as

$$H_{\gamma(i)} = J_{\gamma(i)} K_{\gamma(i)}, \quad (154)$$

where each $J_{\gamma(i)} \in \mathbb{R}$ is an i.i.d. Gaussian coefficient and $K_{\gamma(i)}$ is a deterministic matrix bounded as $\|K_{\gamma(i)}\| = 1$.¹² However, note that Theorem 2.4 is only applicable when the mean of the coefficients $J_{\gamma(i)}$ is zero and the standard deviation is 1, but the standard deviation of each $J_{\gamma(i)}$ is σ , see equation (27). Thus, we multiply each $K_{\gamma(i)}$ by σ and divide $J_{\gamma(i)}$ by σ :

$$H_{\gamma(i)} = \left(\frac{J_{\gamma(i)}}{\sigma} \right) \cdot \sigma K_{\gamma(i)} =: g_{\gamma(i)} \tilde{K}_{\gamma(i)}. \quad (155)$$

This way, each $g_{\gamma(i)}$ is a standard Gaussian coefficient with 0 mean and standard deviation 1. Since each $K_{\gamma(i)}$ has norm 1, each $\tilde{K}_{\gamma(i)}$ is a deterministic matrix bounded by $\|\tilde{K}_{\gamma(i)}\| \leq \sigma$.

We can now view the F_g in equation (146) as a standard Gaussian random matrix polynomial. We can directly apply Theorem 2.4 with $p \geq 2$ and $m = \Gamma$ to obtain

$$\|F_g\|_{*,p}^2 \leq C_p^g g^g \sum_{\mathbf{w}:|\mathbf{w}|\leq g} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}+2\mathbf{v}|=g} |T_{\pi[\eta(\mathbf{w},2\mathbf{v})]}| \prod_{i=1}^m \sigma^{w_i+2v_i} \right)^2 \|I\|_{*,p}^2, \quad (156)$$

¹²Recall from equation (40) that the matrices $K_{\gamma(i)}$ are products of Jordan–Wigner representations of Majorana fermions $\gamma(i)$ ordered lexicographically, and Jordan–Wigner representations of Majorana fermions are Pauli strings, which are unitaries.

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^\Gamma$ such that $|\mathbf{w}| + 2|\mathbf{v}| = g$. The coefficients $T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}$ are inherited from the corresponding $T_{\mathbf{j}}$ in equation (151) by replacing the vector \mathbf{j} with $\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]$ where $\boldsymbol{\eta}$ is the unary encoding function defined in equation (91). By the bound in equation (152),

$$|T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \leq g^g \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]), \quad (157)$$

where the indicator $\mathbb{1}(\mathbf{a})$ is defined in equation (153) on all vectors $\mathbf{a} \in [\Gamma]^g$. Since $|\mathbf{w}| + 2|\mathbf{v}| = g$ by construction, $\prod_{i=1}^m \sigma^{w_i + 2v_i} = \sigma^g$ and hence equation (156) simplifies to

$$\|F_g\|_{*,p}^2 \leq C_p^g g^{3g} \sigma^{2g} \sum_{\mathbf{w}:|\mathbf{w}| \leq g} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}|+2|\mathbf{v}|=g} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2 \|I\|_{*,p}^2. \quad (158)$$

Let us split the sum over \mathbf{w} as

$$\sum_{\mathbf{w}:|\mathbf{w}| \leq g} = \sum_{w=0}^g \sum_{\mathbf{w}:|\mathbf{w}|=w} \quad (159)$$

and rewrite equation (158) as

$$\|F_g\|_{*,p}^2 \leq C_p^g g^{3g} \sigma^{2g} \sum_{w=0}^g G_w(H) \|I\|_{*,p}^2 \quad (160)$$

where

$$G_w(H) := \sum_{\mathbf{w}:|\mathbf{w}|=w} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}|+2|\mathbf{v}|=g} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2. \quad (161)$$

We can then fix a $w \in \{0, \dots, g\}$ and focus on upper bounding $G_w(H)$. We can do this using the following lemma:

Lemma 4.2. *Let $H = \sum_{i=1}^m H_{\gamma(i)}$ be a Hamiltonian of m terms with arbitrary ordering γ , such that for all $1 \leq i < j \leq m$, $H_{\gamma(i)}$ and $H_{\gamma(j)}$ either commute or anti-commute. Fix $g \geq 1$ and $w \in \{0, \dots, g\}$, and consider the H -dependent sum*

$$G_w(H) := \sum_{\mathbf{w}:|\mathbf{w}|=w} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}|+2|\mathbf{v}|=g} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2, \quad (162)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^m$ such that $|\mathbf{w}| = w$ and $|\mathbf{w}| + 2|\mathbf{v}| = g$. The permutation $\pi \in \mathcal{S}_g$ reorders the entries of vectors in $[m]^g$. The unary encoding $\boldsymbol{\eta}$ and the indicator $\mathbb{1}$ are defined on vectors $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$ and $\mathbf{j} \in [m]^g$ as follows:

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := (\underbrace{1, \dots, 1}_{a_1 \text{ times}}, \underbrace{2, \dots, 2}_{a_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{a_m \text{ times}}, \underbrace{1, \dots, 1}_{b_1 \text{ times}}, \underbrace{2, \dots, 2}_{b_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{b_m \text{ times}}), \quad (163)$$

and

$$\mathbb{1}(\mathbf{j}) := \begin{cases} 1 & \text{if } \mathcal{L}_{\gamma(j_g)} \cdots \mathcal{L}_{\gamma(j_2)}[H_{\gamma(j_1)}] \neq 0, \\ 0 & \text{else.} \end{cases} \quad (164)$$

Then

$$G_w(H) \leq g^{3g-2} m^2 Q_{\max}(H)^{g-2} \quad (165)$$

where

$$Q_{\max}(H) := \max_{1 \leq i \leq m} |\{j \in [m] : [H_{\gamma(j)}, H_{\gamma(i)}] \neq 0\}|. \quad (166)$$

The proof of this lemma can be found in Appendix D. For the SYK model, we have $m = \binom{n}{k}$ and $Q_{\max}(H) = Q(n, k)$. Hence, by Lemma 4.2

$$G_w(H) \leq g^{3g-2} \binom{n}{k}^2 Q(n, k)^{g-2}. \quad (167)$$

Plugging this back into equation (160) gives us

$$\|F_g\|_{*,p}^2 \leq C_p^g g^{3g} \sigma^{2g} \cdot (g+1) \cdot g^{3g-2} \binom{n}{k}^2 Q(n, k)^{g-2} \|I\|_{*,p}^2 \quad (168)$$

where the factor $g+1$ comes from the sum over w . Taking the square root of both sides we get

$$\|F_g\|_{*,p} \leq C_p^{\frac{g}{2}} (g+1)^{\frac{1}{2}} g^{3g-1} \sigma^g \binom{n}{k} Q(n, k)^{\frac{g}{2}-1} \|I\|_{*,p}. \quad (169)$$

Plugging this back into equation (145) gives us

$$\|\mathcal{E}_g\|_{*,p} \leq \Upsilon g t^{g-1} C_p^{\frac{g}{2}} (g+1)^{\frac{1}{2}} g^{3g-1} \sigma^g \binom{n}{k} Q(n, k)^{\frac{g}{2}-1} \|I\|_{*,p}. \quad (170)$$

4.1.2 Beyond g -th order

Next, we want to upper bound the residual error term defined in equation (140):

$$\begin{aligned} \|\mathcal{E}_{\geq g'}\|_{*,p} &= \left\| \sum_{j=1}^J \sum_{m=j+1}^J e^{\mathcal{L}_{\gamma(j)} t} \dots e^{\mathcal{L}_{\gamma(m+1)} t} \right. \\ &\quad \times \int_0^t d\xi \sum_{g_m + \dots + g_{j+1} = g' - 1, g_m \geq 1} e^{\mathcal{L}_{\gamma(m)} \xi} \mathcal{L}_{\gamma(m)}^{g_m} \dots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{(t-\xi)^{g_m-1} t^{g'-g_m-1}}{(g_m-1)! g_{m-1}! \dots g_{j+1}!} \left. \right\|_{*,p}. \end{aligned} \quad (171)$$

We would first like to remove the exponentials inside the expected Schatten norm. This is hindered by the fact that they depend on the summation index m . The workaround is to replace the triangular double summation by

$$\sum_{j=1}^J \sum_{m=j+1}^J \mapsto \sum_{m=2}^J \sum_{j=1}^{m-1}. \quad (172)$$

Hence,

$$\begin{aligned} \|\mathcal{E}_{\geq g'}\|_{*,p} &= \left\| \sum_{m=2}^J \sum_{j=1}^{m-1} e^{\mathcal{L}_{\gamma(j)} t} \dots e^{\mathcal{L}_{\gamma(m+1)} t} \right. \\ &\quad \times \int_0^t d\xi \sum_{g_m + \dots + g_{j+1} = g' - 1, g_m \geq 1} e^{\mathcal{L}_{\gamma(m)} \xi} \mathcal{L}_{\gamma(m)}^{g_m} \dots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{(t-\xi)^{g_m-1} t^{g'-g_m-1}}{(g_m-1)! g_{m-1}! \dots g_{j+1}!} \left. \right\|_{*,p}. \end{aligned} \quad (173)$$

By triangle inequality (and taking the integral outside the sum),

$$\begin{aligned} \|\mathcal{E}_{\geq g'}\|_{*,p} &\leq \sum_{m=2}^J \int_0^t d\xi \left\| \sum_{j=1}^{m-1} e^{\mathcal{L}_{\gamma(j)} t} \dots e^{\mathcal{L}_{\gamma(m+1)} t} \right. \\ &\quad \times \sum_{g_m + \dots + g_{j+1} = g' - 1, g_m \geq 1} e^{\mathcal{L}_{\gamma(m)} \xi} \mathcal{L}_{\gamma(m)}^{g_m} \dots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{(t-\xi)^{g_m-1} t^{g'-g_m-1}}{(g_m-1)! g_{m-1}! \dots g_{j+1}!} \left. \right\|_{*,p}. \end{aligned} \quad (174)$$

Since the exponentials inside the expected Schatten norm do not depend on the summations over j and g_i 's, we can pull them in front of both sums and then by unitary invariance of the norm remove altogether:

$$\|\mathcal{E}_{\geq g'}\|_{*,p} \leq \sum_{m=2}^J \int_0^t d\xi \left\| \sum_{j=1}^{m-1} \sum_{g_m + \dots + g_{j+1} = g' - 1, g_m \geq 1} \mathcal{L}_{\gamma(m)}^{g_m} \dots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{(t-\xi)^{g_m-1} t^{g'-g_m-1}}{(g_m-1)! g_{m-1}! \dots g_{j+1}!} \right\|_{*,p}. \quad (175)$$

Consider the same type of symmetrization as in the previous section:

$$\|\mathcal{E}_{\geq g'}\|_{*,p} \leq \sum_{m=2}^J \int_0^t d\xi \left\| \sum_{j_{g'-1}=1}^J \dots \sum_{j_0=1}^J \tilde{\theta}_j \mathcal{L}_{\gamma(j_{g'-1})} \dots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \right\|_{*,p}, \quad (176)$$

where we have absorbed $\frac{(t-\xi)^{g_m-1} t^{g'-g_m-1}}{(g_m-1)! g_{m-1}! \dots g_{j+1}!}$ into the coefficients $\tilde{\theta}_j$, which are set to zero as needed to reduce this sum to the previous one. Compared to the previous section where $\tilde{\theta}_j \in [0, 1]$, the coefficients $\tilde{\theta}_j$ are now bounded by

$$|\tilde{\theta}_j| \leq t^{g'-2} \quad (177)$$

due to the extra factor of $(t-\xi)^{g_m-1} t^{g'-g_m-1}$ in their definition.

Similar to our strategy for the g -th order terms in equation (147), we can write the sums inside the norm in equation (176) as a random matrix polynomial with coefficients T_j :

$$F_{g'}(H_{\gamma(\Gamma)}, \dots, H_{\gamma(1)}) := \sum_{j_{g'-1}=1}^{\Gamma} \dots \sum_{j_0=1}^{\Gamma} T_j H_{\gamma(j_{g'-1})} \dots H_{\gamma(j_0)} \quad (178)$$

where the coefficients T_j are now bounded by

$$|T_j| \leq (g')^{g'} t^{g'-1} \mathbf{1}(\mathbf{j}), \quad (179)$$

where the indicator function $\mathbf{1}(\mathbf{j})$ is defined for all vectors $\mathbf{j} \in [J]^{g'}$ in the same way as in equation (153). By the same argument as in the previous section, we get the following bound that follows directly from equation (169) with the updated upper bound of the coefficients T_j in equation (179):

$$\|F_{g'}\|_{*,p} \leq C_p^{\frac{g'}{2}} (g'+1)^{\frac{1}{2}} (g')^{3g'-1} \sigma^{g'} \binom{n}{k} Q(n, k)^{\frac{g'}{2}-1} \|I\|_{*,p}. \quad (180)$$

Plug this back into equation (175) (the integral over ξ simply becomes a factor t), we get

$$\|\mathcal{E}_{\geq g'}\|_{*,p} \leq \Upsilon^{g'+1} t^{g'-1} C_p^{\frac{g'}{2}} (g'+1)^{\frac{1}{2}} (g')^{3g'-1} \sigma^{g'} \binom{n}{k}^2 Q(n, k)^{\frac{g'}{2}-1} \|I\|_{*,p}. \quad (181)$$

4.1.3 Total error

Now, we can finally put everything together. For an l -th order Υ -stage product formula, the norm of the error generator \mathcal{E} of the SYK model can be bounded by applying the triangle inequality in equation (138) and then using the bounds from equations (170) and (181):

$$\|\mathcal{E}\|_{*,p} \leq \sum_{g=l+1}^{g'-1} \|\mathcal{E}_g\|_{*,p} + \|\mathcal{E}_{\geq g'}\|_{*,p} \quad (182)$$

$$\leq \|I\|_{*,p} \sum_{g=l+1}^{g'-1} \Upsilon^g t^{g-1} C_p^{\frac{g}{2}} (g+1)^{\frac{1}{2}} g^{3g-1} \sigma^g \binom{n}{k} Q(n, k)^{\frac{g}{2}-1} \quad (183)$$

$$+ \|I\|_{*,p} \Upsilon^{g'+1} t^{g'-1} C_p^{\frac{g'}{2}} (g'+1)^{\frac{1}{2}} (g')^{3g'-1} \sigma^{g'} \binom{n}{k}^2 Q(n, k)^{\frac{g'}{2}-1}.$$

For $g' = l + 2$, we have

$$\|\mathcal{E}\|_{*,p} \leq \|I\|_{*,p} \mathcal{A}(l) \left(t^l \left[\sqrt{C_p} \sigma \sqrt{Q(n,k)} \right]^{l+1} \binom{n}{k} Q(n,k)^{-1} + t^{l+1} \left[\sqrt{C_p} \sigma \sqrt{Q(n,k)} \right]^{l+2} \binom{n}{k}^2 Q(n,k)^{-1} \right), \quad (184)$$

where $\mathcal{A}(l) = \Upsilon^{l+3} (l+3)^{\frac{1}{2}} (l+2)^{3(l+2)-1}$. As we divide the total evolution time t into r rounds, with each round having duration $\frac{t}{r}$, the error per round can be bounded as

$$\|e^{iHt/r} - S_l(t/r)\|_{*,p} \leq \int_0^{\frac{t}{r}} \|\mathcal{E}(\tau)\| d\tau \quad (185)$$

$$\begin{aligned} &\leq \|I\|_{*,p} \frac{\mathcal{A}(l)}{l+1} \left(\frac{t^{l+1}}{r^{l+1}} \left[\sqrt{C_p} \sigma \sqrt{Q(n,k)} \right]^{l+1} \binom{n}{k} Q(n,k)^{-1} \right. \\ &\quad \left. + \frac{t^{l+2}}{r^{l+2}} \left[\sqrt{C_p} \sigma \sqrt{Q(n,k)} \right]^{l+2} \binom{n}{k}^2 Q(n,k)^{-1} \right). \end{aligned} \quad (186)$$

Using $C_p < p$, the total error is then bounded as

$$\|e^{iHt} - S_l(t/r)\|_{*,p} \leq r \cdot \|e^{iHt/r} - S_l(t/r)\|_{*,p} \quad (187)$$

$$\begin{aligned} &\leq \|I\|_{*,p} \mathcal{C}(l) \sqrt{p} \sigma Q(n,k)^{-\frac{1}{2}} t \\ &\quad \cdot \left(\left[\sqrt{p} \sigma \sqrt{Q(n,k)} \frac{t}{r} \right]^l \binom{n}{k} + \left[\sqrt{p} \sigma \sqrt{Q(n,k)} \frac{t}{r} \right]^{l+1} \binom{n}{k}^2 \right) \end{aligned} \quad (188)$$

with $\mathcal{C}(l) = \frac{\mathcal{A}(l)}{l+1}$. This concludes the proof of Theorem 4.1. \square

4.2 Gate complexity

Corollary 4.2.1. *Simulating a k -local SYK model with (even) l -th order product formula, the gate complexity*

$$G = \tilde{\mathcal{O}} \left[n^{g_1(k)} \sqrt{n + \log(e^2/\delta)} \mathcal{J} t \cdot \left(\left[n^{g_2(k)} \sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^{g_3(k)} \sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right], \quad (189)$$

where

$$(g_1(k), g_2(k), g_3(k)) := \begin{cases} (k, 1, k+1) & \text{if } k \text{ is even,} \\ (k + \frac{1}{2}, \frac{1}{2}, k + \frac{1}{2}) & \text{if } k \text{ is odd,} \end{cases} \quad (190)$$

ensures the probability statement

$$\mathbb{P} (\|e^{iHt} - S_l(t/r)^r\| \geq \epsilon) < \delta. \quad (191)$$

When l is sufficiently large, this gate complexity scales in n as

$$G_{l \gg 2} \sim \begin{cases} n^{k+\frac{1}{2}} \mathcal{J} t & \text{if } k \text{ is even,} \\ n^{k+1} \mathcal{J} t & \text{if } k \text{ is odd.} \end{cases} \quad (192)$$

Proof. To derive the l -th order complexity considering the probability statement in (49), we first bound the l -th Trotter error using Theorem 4.1:

$$\|e^{iHt} - S_l(t/r)^r\|_p \leq \|I\|_p \cdot p \cdot \lambda(p, r), \quad (193)$$

where

$$\lambda(p, r) = \frac{1}{\sqrt{p}} \mathcal{C}(l) \sigma Q(n, k)^{-\frac{1}{2}} t \left(\binom{n}{k} \left[\sqrt{p} \sigma \sqrt{Q(n, k)} \frac{t}{r} \right]^l + \binom{n}{k}^2 \left[\sqrt{p} \sigma \sqrt{Q(n, k)} \frac{t}{r} \right]^{l+1} \right). \quad (194)$$

This gives us an explicit expression for the concentration inequality (55). By Lemma 2.1, finding the Trotter number r that ensures the probability statement (49) is equivalent to finding an r that solves the inequality

$$e \sqrt{p_D} \mathcal{C}(l) \sigma Q(n, k)^{-\frac{1}{2}} \frac{t}{\epsilon} \left(\binom{n}{k} \left[\sqrt{p_D} \sigma \sqrt{Q(n, k)} \frac{t}{r} \right]^l + \binom{n}{k}^2 \left[\sqrt{p_D} \sigma \sqrt{Q(n, k)} \frac{t}{r} \right]^{l+1} \right) \leq 1, \quad (195)$$

where $p_D := \log(e^2 D / \delta)$ with $D := \|I\|_p^p$.

Hence,

$$r \geq \sqrt{\log(e^2 D / \delta)} \sigma \sqrt{Q(n, k)} t \cdot \left(\left[2e \sqrt{\log(e^2 D / \delta)} \mathcal{C}(l) \binom{n}{k} \sigma Q(n, k)^{-\frac{1}{2}} \frac{t}{\epsilon} \right]^{\frac{1}{l}} + \left[2e \sqrt{\log(e^2 D / \delta)} \mathcal{C}(l) \binom{n}{k}^2 \sigma Q(n, k)^{-\frac{1}{2}} \frac{t}{\epsilon} \right]^{\frac{1}{l+1}} \right). \quad (196)$$

Recall that the norm $\|I\|_p$ should be interpreted as the Schatten p -norm on $\text{GL}(n/2, \mathbb{C})$ because we have (even) n Majoranas. Hence, $D = \|I\|_p^p = 2^{\frac{n}{2}}$. Substituting $D = 2^{n/2}$ and multiply r with the number of local terms Γ and the number of stages $\Upsilon = 2 \cdot 5^{\frac{l}{2}-1}$, the l -th order gate complexity scales with

$$G = \tilde{\mathcal{O}} \left(\Upsilon \sqrt{\frac{\log(2)}{2} n + \log(e^2 / \delta)} \sigma \binom{n}{k} \sqrt{Q(n, k)} t \cdot \left(\left[2e \sqrt{\frac{\log(2)}{2} n + \log(e^2 / \delta)} \mathcal{C}(l) \binom{n}{k} \sigma Q(n, k)^{-\frac{1}{2}} \frac{t}{\epsilon} \right]^{\frac{1}{l}} + \left[2e \sqrt{\frac{\log(2)}{2} n + \log(e^2 / \delta)} \mathcal{C}(l) \binom{n}{k}^2 \sigma Q(n, k)^{-\frac{1}{2}} \frac{t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right). \quad (197)$$

For fixed k and l , the k and l dependent factors can be removed from the big- $\tilde{\mathcal{O}}$ notation. Asymptotically, we have $\binom{n}{k} = \mathcal{O}(n^k)$, and $Q(n, k) = \mathcal{O}(n^{k-1})$ for even k while $Q(n, k) = \mathcal{O}(n^k)$ for odd k . By writing out $\sigma = \mathcal{J} \sqrt{\frac{(k-1)!}{kn^{k-1}}}$, we can write G asymptotically as

$$G = \tilde{\mathcal{O}} \left[n^{g_1(k)} \sqrt{n + \log(e^2 / \delta)} \mathcal{J} t \cdot \left(\left[n^{g_2(k)} \sqrt{n + \log(e^2 / \delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^{g_3(k)} \sqrt{n + \log(e^2 / \delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right], \quad (198)$$

where $g_1(k)$, $g_2(k)$, $g_3(k)$ are defined in equation (190). This concludes the proof. \square

Similar to Section 3, we can consider the probability statement in equation (50), and apply Lemma 2.2 instead of Lemma 2.1, which leads us to the following result:

Corollary 4.2.2. *For an arbitrary fixed input state $|\psi\rangle$, simulating an even k -local SYK model with the l -th order product formula (for even $l \geq 2$), the gate complexity*

$$G = \tilde{\mathcal{O}} \left[n^{g_1(k)} \sqrt{\log(e^2 / \delta)} \mathcal{J} t \cdot \left(\left[n^{g_2(k)} \sqrt{\log(e^2 / \delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^{g_3(k)} \sqrt{\log(e^2 / \delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right] \quad (199)$$

where

$$(g_1(k), g_2(k), g_3(k)) = \begin{cases} (k, 1, k+1) & \text{if } k \text{ is even,} \\ (k + \frac{1}{2}, \frac{1}{2}, k + \frac{1}{2}) & \text{if } k \text{ is odd} \end{cases} \quad (200)$$

ensures the probability statement

$$\mathbb{P} \left(\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l^2} \geq \epsilon \right) < \delta. \quad (201)$$

When l is sufficiently large, this gate complexity scales in n as

$$G \sim \begin{cases} n^k \mathcal{J}t & \text{if } k \text{ is even,} \\ n^{k+\frac{1}{2}} \mathcal{J}t & \text{if } k \text{ is odd.} \end{cases} \quad (202)$$

5 Higher-order Trotter error of the sparsified SYK model

Our methods for evaluating higher-order Trotter errors also work in the case of the *sparse SYK model* which is defined as the regular SYK model (26) but with an additional Bernoulli variable $B_{i_1, \dots, i_k} \in \{0, 1\}$ attached to each term:

$$H_{\text{SSYK}} = \sum_{1 \leq i_1 < \dots < i_k \leq n} B_{i_1, \dots, i_k} J_{i_1, \dots, i_k} \chi_{i_1} \cdots \chi_{i_k}. \quad (203)$$

Similar to our treatment for the regular SYK model in equation (39), we write the sparse SYK model in the local Hamiltonian form with an arbitrary ordering γ

$$H_{\text{SSYK}} = \sum_{i=1}^{\Gamma} B_i H_{\gamma(i)}. \quad (204)$$

The sparse SYK model has two sources of randomness, one from the Gaussian variables $J_{\gamma(i)}$ within $H_{\gamma(i)}$ as in the regular SYK model, the other from the Bernoulli variables B_i . We pose that the Bernoulli variables are independent of each other and also independent of the Gaussian variables $J_{\gamma(i)}$. For simplicity, each Bernoulli variable B_i is equal to 1 with probability p_B and equal to 0 with probability $1 - p_B$ for some $p_B \in [0, 1]$. Usually, we set

$$p_B = \frac{\kappa n}{\Gamma} \quad (205)$$

where $\kappa \in \mathbb{R}_{\geq 0}$ is the degree of sparsity, and relative to equation (27) we renormalize the variance of the Gaussian variables $J_{\gamma(i)}$ to

$$\sigma^2 = \frac{1}{p_B} \frac{(k-1)! \mathcal{J}^2}{kn^{k-1}}. \quad (206)$$

When evaluating the expected Schatten norm of the Trotter error of the sparse SYK model, we will deal with these two types of randomness separately. We will write \mathbf{B} to denote the Bernoulli variables B_1, \dots, B_{Γ} . When we want to specify that this series of random variables takes a certain configuration \mathbf{b} , we will write $\mathbf{B} = \mathbf{b}$ where \mathbf{b} is a fixed vector in $\{0, 1\}^{\Gamma}$. This is equivalent to $B_1 = b_1, B_2 = b_2, \dots, B_{\Gamma} = b_{\Gamma}$. Using this notation, we can write the Trotter error given a specific sample $\mathbf{b} \in \{0, 1\}^{\Gamma}$ as

$$\mathcal{D}(\mathbf{B} = \mathbf{b}) := e^{iH_{\text{SSYK}}(\mathbf{B}=\mathbf{b})t} - S_l(t/r)^r, \quad (207)$$

indicating that the Trotter error depends on the specific sample \mathbf{b} of the Bernoulli variables.

5.1 Average Trotter error

Let $b := |\mathbf{b}| = \sum_{i=1}^{\Gamma} b_i$ denote the number of ones in the sample \mathbf{b} of the Bernoulli variables \mathbf{B} . When the Bernoulli variables take the value of this sample, we get a configuration of the sparse

SYK model where we treat the sample values as coefficients:

$$H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}) = \sum_{i=1}^{\Gamma} b_i H_{\gamma(i)}, \quad (208)$$

which contains exactly b terms. Of course, this configuration is still a k -local Hamiltonian, although it becomes more difficult to count the anti-commuting terms, since the all-to-all connected symmetry in the full SYK model is broken. Therefore, we have to resort to average statements about both the Trotter error and gate complexity of these sparse models. The following theorem bounds the Trotter error averaged over the Bernoulli variables, defined as¹³

$$\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle := \langle \|e^{iH_{\text{SSYK}}(\mathbf{B})t} - S_l(t/r)^r\|_{*,p} \rangle, \quad (209)$$

where we use $\langle \cdot \rangle$ to denote averaging over the Bernoulli variables, while the norm $\|\cdot\|_{*,p}$ only concerns the Gaussian variables and treats the Bernoulli variables as coefficients.

Theorem 5.1. *Consider the sparse SYK model $H_{\text{SSYK}}(\mathbf{B})$ with Bernoulli variables $\mathbf{B} = (B_1, \dots, B_{\Gamma})$ such that $\mathbb{P}(B_i = 1) = p_B$ for all $1 \leq i \leq \Gamma$. Given $l \geq 2$ even, if $p_B Q(n, k) \geq 1$, the l -th order average Trotter error (209) is bounded by*

$$\frac{\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle}{\|I\|_{*,p}} \leq \beta(l) \frac{\Gamma \sqrt{p} \sigma \sqrt{p_B} t}{\sqrt{Q(n, k)}} \cdot \left(\left[\sqrt{p} \sigma \sqrt{p_B Q(n, k)} \frac{t}{r} \right]^l + \Gamma \cdot \left[\sqrt{p} \sigma \sqrt{p_B Q(n, k)} \frac{t}{r} \right]^{l+1} \right), \quad (210)$$

where $p \geq 2$ describes the norm, σ is the variance in (206), $\beta(l)$ is a purely l -dependent factor, and $Q(n, k)$ is given in equation (96). If $p_B Q(n, k) \leq 1$ instead, the l -th order average Trotter error (209) is bounded by

$$\frac{\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle}{\|I\|_{*,p}} \leq \beta(l) \frac{\Gamma \sqrt{p} \sigma t}{Q(n, k)} \cdot \left(\left[\sqrt{p} \sigma \frac{t}{r} \right]^l + \Gamma \cdot \left[\sqrt{p} \sigma \frac{t}{r} \right]^{l+1} \right). \quad (211)$$

Proof. Let us first compute the error generator $\|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p}$ given a specific configuration $\mathbf{b} \in \{0, 1\}^{\Gamma}$:

$$\|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p} = \left\| \sum_{j=1}^J \sum_{g_j + \dots + g_{j+1} = g-1} \prod_{i=j}^J b_i^{g_i} \mathcal{L}_{\gamma(j)}^{g_j} \cdots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{t^{g-1}}{g_j! \cdots g_{j+1}!} \right\|_{*,p}, \quad (212)$$

where we have included the product of coefficients b_i from each local term. However, since each $b_i \in \{0, 1\}$, it suffices to consider each of them to be linear and we can remove their exponent:

$$\|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p} = \left\| \sum_{j=1}^J \sum_{g_j + \dots + g_{j+1} = g-1} \prod_{i=j}^J b_i \mathcal{L}_{\gamma(j)}^{g_j} \cdots \mathcal{L}_{\gamma(j+1)}^{g_{j+1}} [H_{\gamma(j)}] \frac{t^{g-1}}{g_j! \cdots g_{j+1}!} \right\|_{*,p}. \quad (213)$$

Similar to Section 4, we consider the symmetrization

$$\|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p} = t^{g-1} \left\| \sum_{j_{g-1}=1}^J \cdots \sum_{j_0=1}^J \tilde{\theta}_j \prod_{i=0}^{g-1} b_{j_i} \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \right\|_{*,p}, \quad (214)$$

where we have included extra coefficients $\tilde{\theta}_j \in [0, 1]$ that reduces this expression exactly to the previous one. As $J = \Upsilon \cdot \Gamma$, we get the following upper bound

$$\|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p} = \Upsilon^g t^{g-1} \left\| \sum_{j_{g-1}=1}^{\Gamma} \cdots \sum_{j_0=1}^{\Gamma} \tilde{\theta}_j \prod_{i=0}^{g-1} b_{j_i} \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \right\|_{*,p} \quad (215)$$

¹³For the rest of this work, we will use $\langle \cdot \rangle$ to denote averaging over the Bernoulli variables.

thanks to the periodicity of the ordering map γ in equation (37).

Keeping the coefficients b_i and following our discussion in Section 2.8, we can write the sum of commutator chains as the following random matrix polynomial:

$$F_g^{(\mathbf{b})}(H_{\gamma(\Gamma)}, \dots, H_{\gamma(1)}) := \sum_{j_{g-1}=1}^{\Gamma} \cdots \sum_{j_0=1}^{\Gamma} \tilde{\theta}_{\mathbf{j}} \prod_{i=0}^{g-1} b_{j_i} \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \quad (216)$$

$$=: \sum_{j_{g-1}=1}^{\Gamma} \cdots \sum_{j_0=1}^{\Gamma} T_{\mathbf{j}}^{(\mathbf{b})} H_{\gamma(j_{g-1})} \cdots H_{\gamma(j_0)}, \quad (217)$$

where the coefficients $T_{\mathbf{j}}^{(\mathbf{b})}$ are defined as

$$T_{\mathbf{j}}^{(\mathbf{b})} := T_{\mathbf{j}} \cdot \prod_{i=0}^{g-1} b_{j_i} \quad (218)$$

with $T_{\mathbf{j}}$ defined in equation (147). Hence, these coefficients are bounded by

$$|T_{\mathbf{j}}^{(\mathbf{b})}| \leq g^g \mathbf{1}(\mathbf{j}) \cdot \prod_{i=0}^{g-1} b_{j_i}, \quad (219)$$

where the indicator is defined exactly the same as before for vectors $\mathbf{j} \in \{1, \dots, \Gamma\}^g$

$$\mathbf{1}(\mathbf{j}) = \begin{cases} 1 & \text{if } \mathcal{L}_{\gamma(j_{g-1})} \cdots \mathcal{L}_{\gamma(j_1)} [H_{\gamma(j_0)}] \neq 0, \\ 0 & \text{else.} \end{cases} \quad (220)$$

By standardizing the local terms $H_{\gamma(i)}$ as in Section 4 and apply Theorem 2.4 with $p \geq 2$ and $m = \Gamma$, we get

$$\|F_g^{(\mathbf{b})}\|_{*,p}^2 \leq C_p^g g^g \sum_{\mathbf{w}:|\mathbf{w}|\leq g} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}+2|\mathbf{v}|=g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}^{(\mathbf{b})}| \prod_{i=1}^{\Gamma} \sigma^{w_i+2v_i} \right)^2 \|I\|_{*,p}^2, \quad (221)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^{\Gamma}$ such that $|\mathbf{w}| + 2|\mathbf{v}| = g$. The coefficients $T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}^{(\mathbf{b})}$ are inherited from the corresponding $T_{\mathbf{j}}^{(\mathbf{b})}$ via the unary encoding function $\boldsymbol{\eta}$ defined in equation (91) by replacing \mathbf{j} with $\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]$. By equation (219),

$$|T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}^{(\mathbf{b})}| \leq g^g \mathbf{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \cdot \prod_{i \in \text{supp}(\mathbf{w}+2\mathbf{v})} b_i. \quad (222)$$

Since $|\mathbf{w}| + 2|\mathbf{v}| = g$ by construction, $\prod_{i=1}^{\Gamma} \sigma^{w_i+2v_i} = \sigma^g$ and hence

$$\|F_g^{(\mathbf{b})}\|_{*,p}^2 \leq C_p^g g^{3g} \sigma^{2g} \sum_{\mathbf{w}:|\mathbf{w}|\leq g} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}+2|\mathbf{v}|=g} \prod_{i \in \text{supp}(\mathbf{w}+2\mathbf{v})} b_i \mathbf{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2 \|I\|_{*,p}^2. \quad (223)$$

Because \mathbf{w} and \mathbf{v} do not necessarily have complementary support, we consider the following splitting by moving all coefficients b_i for $i \in \text{supp}(\mathbf{w})$ outside of the squared parenthesis. The remaining coefficients b_j are now taken with $j \in \text{supp}(\mathbf{v}) \setminus \text{supp}(\mathbf{w})$. This gives us

$$\|F_g^{(\mathbf{b})}\|_{*,p}^2 \leq C_p^g g^{3g} \sigma^{2g} \sum_{\mathbf{w}:|\mathbf{w}|\leq g} \prod_{i \in \text{supp}(\mathbf{w})} b_i \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}+2|\mathbf{v}|=g} \prod_{j \in \text{supp}(\mathbf{v}) \setminus \text{supp} \mathbf{w}} b_j \mathbf{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2 \|I\|_{*,p}^2. \quad (224)$$

By splitting $\sum_{\mathbf{w}:|\mathbf{w}|\leq g}$ into $\sum_{w=0}^g \sum_{\mathbf{w}:|\mathbf{w}|=w}$ and taking square roots on both sides, we arrive at

$$\|F_g(\mathbf{b})\|_{*,p} \leq C_p^{\frac{g}{2}} g^{3g/2} \sigma^g \sqrt{\sum_{w=0}^g G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \|I\|_{*,p}. \quad (225)$$

where

$$G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b})) := \sum_{\mathbf{w}:|\mathbf{w}|=w} \prod_{i \in \text{supp}(\mathbf{w})} b_i \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{v}|+2|\mathbf{w}|=g} \prod_{j \in \text{supp}(\mathbf{v}) \setminus \text{supp}(\mathbf{w})} b_j \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2. \quad (226)$$

We will attempt to evaluate G_w explicitly later. For now, let us keep G_w and write the g -th order generator as

$$\|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p} \leq \Upsilon t^{g-1} C_p^{\frac{g}{2}} g^{3g/2} \sigma^g \sqrt{\sum_{w=0}^g G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \|I\|_{*,p}. \quad (227)$$

Similarly, the residual error is bounded by

$$\|\mathcal{E}_{\geq g'}(\mathbf{B} = \mathbf{b})\|_{*,p} \leq \Gamma \Upsilon^{g'+1} t^{g'-1} C_p^{\frac{g'}{2}} (g')^{3g'/2} \sigma^{g'} \sqrt{\sum_{w=0}^{g'} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \|I\|_{*,p}. \quad (228)$$

Hence, for $g' = l + 2$ and use $C_p < p$, the l -th order error generator is bounded by

$$\begin{aligned} \|\mathcal{E}(\mathbf{B} = \mathbf{b})\|_{*,p} &\leq \sum_{g=l+1}^{g'-1} \|\mathcal{E}_g(\mathbf{B} = \mathbf{b})\|_{*,p} + \|\mathcal{E}_{\geq g'}(\mathbf{B} = \mathbf{b})\|_{*,p} \\ &\leq \|I\|_{*,p} \left(\Upsilon^{l+1} t^l p^{\frac{l+1}{2}} (l+1)^{3(l+1)/2} \sigma^{l+1} \sqrt{\sum_{w=0}^{l+1} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right. \\ &\quad \left. + \Gamma \Upsilon^{l+3} t^{l+1} p^{\frac{l+2}{2}} (l+2)^{3(l+2)/2} \sigma^{l+2} \sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right). \end{aligned} \quad (229)$$

Then, as we did in Section 4, upper bounding by collecting common prefactors, integrating over time from 0 to t/r , and multiplying by r gives us the Trotter error:

$$\begin{aligned} \|\mathcal{D}(\mathbf{B} = \mathbf{b})\|_{*,p} &\leq \|I\|_{*,p} \alpha(l) \sqrt{\bar{p}} \sigma t \cdot \left(\left[\sqrt{\bar{p}} \sigma \frac{t}{r} \right]^l \sqrt{\sum_{w=0}^{l+1} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right. \\ &\quad \left. + \Gamma \cdot \left[\sqrt{\bar{p}} \sigma \frac{t}{r} \right]^{l+1} \sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right) \end{aligned} \quad (230)$$

where $\alpha(l)$ is a purely l -dependent factor given by

$$\alpha(l) := \frac{\Upsilon^{l+3} (l+2)^{3(l+2)/2}}{l+1}. \quad (231)$$

Equation (230) gives the l -th order Trotter error of a given configuration of the sparse SYK model. However, it is difficult to parse this expression as evaluating $G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))$ for specific sample \mathbf{b} is quite complicated, but giving an average statement is tractable thanks to the following lemma:

Lemma 5.2. Let $H = \sum_{i=1}^m B_i H_{\gamma(i)}$ be a sparse Hamiltonian of m terms with arbitrary ordering γ , such that for all $1 \leq i < j \leq m$, $H_{\gamma(i)}$ and $H_{\gamma(j)}$ either commute or anti-commute. The B_i 's are independent Bernoulli variables with $\mathbb{P}(B_i = 1) = p_B$ for all i . Fix $g \geq 1$ and $w \in \{0, \dots, g\}$, and consider the average H -dependent sum over all samples of the Bernoulli variables

$$\left\langle G_w(H(\mathbf{B})) \right\rangle = \left\langle \sum_{\mathbf{w}: |\mathbf{w}|=w} \prod_{i \in \text{supp}(\mathbf{w})} B_i \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}: |\mathbf{w}|+2|\mathbf{v}|=g} \prod_{j \in \text{supp}(\mathbf{v})} B_j \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2 \right\rangle, \quad (232)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^m$ such that $|\mathbf{w}| = w$ and $|\mathbf{w}| + 2|\mathbf{v}| = g$. The permutation $\pi \in \mathcal{S}_g$ reorders the entries of vectors in $[m]^g$. The unary encoding $\boldsymbol{\eta}$ and the indicator $\mathbb{1}$ are defined on vectors $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$ and $\mathbf{j} \in [m]^g$ as follows:

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := (\underbrace{1, \dots, 1}_{a_1 \text{ times}}, \underbrace{2, \dots, 2}_{a_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{a_m \text{ times}}, \underbrace{1, \dots, 1}_{b_1 \text{ times}}, \underbrace{2, \dots, 2}_{b_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{b_m \text{ times}}), \quad (233)$$

and

$$\mathbb{1}(\mathbf{j}) := \begin{cases} 1 & \text{if } \mathcal{L}_{\gamma(j_g)} \cdots \mathcal{L}_{\gamma(j_2)}[H_{\gamma(j_1)}] \neq 0, \\ 0 & \text{else.} \end{cases} \quad (234)$$

Then

$$\left\langle G_w(H(\mathbf{B})) \right\rangle \leq g^{4g} m^2 Q_{\max}(H)^{-2} \sum_{s=0}^w \sum_{q, q'=0}^{(g-w)/2 \min(q, q')} \sum_{c=0}^{\min(q, q')} (p_B Q_{\max}(H))^{s+q+q'-c}. \quad (235)$$

where

$$Q_{\max}(H) := \max_{1 \leq i \leq m} |\{j \in [m] : [H_{\gamma(j)}, H_{\gamma(i)}] \neq 0\}|. \quad (236)$$

To get some intuition for this lemma, recall that for the full Hamiltonian, we could upper-bound $G_w(H)$ by

$$G_w(H) \leq g^{3g-2} m^2 Q_{\max}(H)^{g-2} \quad (237)$$

in Lemma 4.2. Since we perform the sparsification by attaching i.i.d. Bernoulli variables to the summands of the Hamiltonian, and the average of each Bernoulli variable is p_B , we could expect the average $\left\langle G_w(H(\mathbf{B})) \right\rangle$ to scale with some polynomial (of utmost degree g)

$$\left\langle G_w(H(\mathbf{B})) \right\rangle \sim \sum_s a_s (p_B Q_{\max}(H))^s. \quad (238)$$

To illustrate this, let us naively forget the sum of squares in G_w and pretend it scales with the following sum of commutator chains (each commutator chain has weight 1):

$$G_w(H(\mathbf{B})) \sim \sum_{\pi} \sum_{i_1, \dots, i_g} \left(\prod_{j=1}^g B_{i_j} \right) \mathbb{1}(\pi(i_1, \dots, i_g)). \quad (239)$$

If we take the expectation now, we get

$$\left\langle G_w(H(\mathbf{B})) \right\rangle \sim \sum_{\pi} \sum_{i_1, \dots, i_g} \left\langle \prod_{j=1}^g B_{i_j} \right\rangle \mathbb{1}(\pi(i_1, \dots, i_g)). \quad (240)$$

Note that the average of the product $\prod_{j=1}^g B_i$ is not p_B^g , because we do not know how many distinct elements there are in the product. Therefore, we need to write our summation more carefully to control the distinct elements:

$$G_w(H(\mathbf{B})) \sim \sum_{\pi} \sum_{s=1}^g \sum_{\phi: \text{surj}} \sum_{i_1, \dots, i_s} \left(\prod_{j=1}^s B_{i_j} \right) \mathbb{1}(\pi(i_{\phi(1)}, \dots, i_{\phi(g)})), \quad (241)$$

where we now sum over s distinct elements i_1, \dots, i_s . We use surjective functions $\phi : \{1, \dots, g\} \rightarrow \{1, \dots, s\}$ to distribute i_1, \dots, i_s over the indicator. In this way, the average becomes:

$$\langle G_w(H(\mathbf{B})) \rangle \sim \sum_{\pi} \sum_{s=1}^g \sum_{\phi: \text{surj}} \sum_{\text{distinct } i_1, \dots, i_s} p_B^s \mathbf{1}(\pi(i_{\phi(1)}, \dots, i_{\phi(g)})), \quad (242)$$

The right-hand side is upper-bounded by

$$\sum_{\pi} \sum_{s=1}^g \sum_{\phi: \text{surj}} \sum_{\text{distinct } i_1, \dots, i_s} p_B^s \mathbf{1}(\pi(i_{\phi(1)}, \dots, i_{\phi(g)})) \leq g! \cdot g^g m Q_{\max}(H)^{-1} \sum_{s=1}^g (p_B Q_{\max}(H))^s \quad (243)$$

where m denotes the total number of terms in the Hamiltonian. We use $g!$ to bound the number of permutations $\pi \in \mathcal{S}_g$ and g^g to bound the number of surjective functions from $\{1, \dots, g\} \rightarrow \{1, \dots, s\}$. The above bound in equation (243) is indeed a polynomial as expected, and resembles the bound in Lemma 5.2 apart from additional sums there. In reality, we cannot simply ignore the sum of squares in G_w as it plays a crucial role in capturing the scaling behavior of the Trotter error. To get the correct bound, one needs to expand the sum of squares and treat the overlapping summations carefully. (This is taken care of by the additional summations over x, q and q' in Lemma 5.2.) For details of this treatment, we refer the reader to the proof in Appendix E.

To apply Lemma 5.2, let us first put expectations on both side of equation (230) and apply Jensen's inequality:

$$\begin{aligned} \langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle &\leq \|I\|_{*,p} \alpha'(l) \sqrt{p\sigma t} \cdot \left(\left[\sqrt{p\sigma} \frac{t}{r} \right]^l \sqrt{\sum_{w=0}^{l+1} \langle G_w(H_{\text{SSYK}}(\mathbf{B})) \rangle} \right. \\ &\quad \left. + \Gamma \cdot \left[\sqrt{p\sigma} \frac{t}{r} \right]^{l+1} \sqrt{\sum_{w=0}^{l+2} \langle G_w(H_{\text{SSYK}}(\mathbf{B})) \rangle} \right). \end{aligned} \quad (244)$$

For the SYK model, we have $Q_{\max} = Q(n, k)$, and therefore, by Lemma 5.2, we get

$$\begin{aligned} \langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle &\leq \|I\|_{*,p} \alpha'(l) \frac{\Gamma \sqrt{p\sigma t}}{Q(n, k)} \cdot \left(\left[\sqrt{p\sigma} \frac{t}{r} \right]^l \sqrt{\sum_{w=0}^{l+1} \sum_{s=0}^w \sum_{q, q'=0}^{(l+1-w)/2 \min(q, q')} \sum_{c=0}^s (p_B Q(n, k))^{s+q+q'-c}} \right. \\ &\quad \left. + \Gamma \cdot \left[\sqrt{p\sigma} \frac{t}{r} \right]^{l+1} \sqrt{\sum_{w=0}^{l+2} \sum_{s=0}^w \sum_{q, q'=0}^{(l+2-w)/2 \min(q, q')} \sum_{c=0}^s (p_B Q(n, k))^{s+q+q'-c}} \right), \end{aligned} \quad (245)$$

where

$$\alpha'(l) := (l+2)^{2(l+2)} \cdot \frac{\Upsilon^{l+3} (l+2)^{3(l+2)/2}}{l+1}. \quad (246)$$

We can simplify the expression above depending on how we tune the probability p_B . If $p_B Q(n, k) \geq 1$, then

$$\begin{aligned} \sum_{s=0}^w \sum_{q, q'=0}^{(g-w)/2 \min(q, q')} \sum_{c=0}^s (p_B Q(n, k))^{s+q+q'-c} &\leq (p_B Q(n, k))^g \sum_{s=0}^w \sum_{q, q'=0}^{(g-w)/2 \min(q, q')} \sum_{c=0}^s 1 \\ &\leq (p_B Q(n, k))^g (w+1) \left(\frac{g-w}{2} + 1 \right)^3 \\ &\leq (p_B Q(n, k))^g \cdot (g+1)^4. \end{aligned} \quad (247)$$

Plugging these into the expression above, we get

$$\frac{\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle}{\|I\|_{*,p}} \leq \beta(l) \frac{\Gamma \sqrt{p\sigma} \sqrt{p_B t}}{\sqrt{Q(n, k)}} \cdot \left(\left[\sqrt{p\sigma} \sqrt{p_B Q(n, k)} \frac{t}{r} \right]^l + \Gamma \cdot \left[\sqrt{p\sigma} \sqrt{p_B Q(n, k)} \frac{t}{r} \right]^{l+1} \right), \quad (248)$$

where

$$\beta(l) = (l+3)^{\frac{5}{2}}(l+2)^{2(l+2)} \cdot \frac{\Upsilon^{l+3}(l+2)^{3(l+2)/2}}{l+1}. \quad (249)$$

If p_B is tuned such that $p_B Q(n, k) \leq 1$, then

$$\begin{aligned} \sum_{s=0}^w \sum_{q, q'=0}^{(g-w)/2 \min(q, q')} \sum_{c=0}^{(g-w)/2} (p_B Q(n, k))^{s+q+q'-c} &\leq \sum_{s=0}^w \sum_{x=0}^{(g-w)/2} \sum_{q, q'=0}^{(g-w)/2} 1 \\ &\leq (g+1)^4. \end{aligned} \quad (250)$$

This gives us

$$\frac{\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle}{\|\mathcal{I}\|_{*,p}} \leq \beta(l) \frac{\Gamma \sqrt{p} \sigma t}{Q(n, k)} \cdot \left(\left[\sqrt{p} \sigma \frac{t}{r} \right]^l + \Gamma \cdot \left[\sqrt{p} \sigma \frac{t}{r} \right]^{l+1} \right). \quad (251)$$

This proves the statement. \square

5.2 Average gate complexity

As we have seen in equation (230), the Trotter error depends on the specific sample \mathbf{b} of the Bernoulli variables \mathbf{B} , so will the gate complexity. If the support of \mathbf{b} happens to have size Γ , we recover the full SYK Hamiltonian. If the support of \mathbf{b} is much smaller than Γ , we get a Hamiltonian with much fewer terms that requires much fewer gates to simulate. It is therefore tricky to derive one gate complexity that covers all the cases. Moreover, due to the difficulty of evaluating $G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))$ for specific sample \mathbf{b} in equation (230), it is also difficult to formulate a gate count specific to each sample. Hence, we decided on the following approach, leading us to a notion of average gate complexity. Let

$$\mathcal{G} := \{G(\mathbf{B} = \mathbf{b}) : \mathbf{b} \in \{0, 1\}^\Gamma\} \quad (252)$$

be the set of gate complexities for all configurations \mathbf{b} , where each $G(\mathbf{B} = \mathbf{b})$ is the gate complexity of the l -th order product formula for approximating the configuration $H_{\text{SSYK}}(\mathbf{B} = \mathbf{b})$ and ensuring the probability statement in equation (49) or (50). Each gate complexity $G(\mathbf{B} = \mathbf{b})$ is the minimal gate complexity derived from the Trotter error of the corresponding sample $\|\mathcal{D}(\mathbf{B} = \mathbf{b})\|_{*,p}$. If we take the expectation of \mathcal{G} over all samples Bernoulli variables, the quantity

$$\bar{G} := \langle G(\mathbf{B}) \rangle \quad (253)$$

yields a measure for the average gate complexity when simulating the sparse SYK model. We shall formulate our result in the following corollary:

Corollary 5.2.1. *Simulating the k -local κ -sparse SYK model using (even) l -th order product formula has average gate complexity*

$$\bar{G} = \tilde{\mathcal{O}} \left(n \left(\sqrt{p_B Q(n, k)} \right)^{1+\frac{2}{l}} \sqrt{n + \log(e^2/\delta)} \mathcal{J} t \cdot \left(\left[\sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^k \sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right), \quad (254)$$

in the sense that \bar{G} upper bounds the average of the set

$$\mathcal{G} = \{G(\mathbf{B} = \mathbf{b}) : \mathbf{b} \in [0, 1]^\Gamma\}, \quad (255)$$

where $G(\mathbf{B} = \mathbf{b})$ is the minimal gate complexity of $S_l(t/r)^r$ approximating the configuration $H_{\text{SSYK}}(\mathbf{B} = \mathbf{b})$ that ensures the probability statement

$$\mathbb{P} \left(\left\| e^{iH_{\text{SSYK}}(\mathbf{B}=\mathbf{b})t} - S_l(t/r)^r \right\| \geq \epsilon \right) < \delta. \quad (256)$$

The term $Q(n, k)$ is given by equation (96). When l is sufficiently large, the average gate complexity scales in n as

$$\bar{G} \sim \begin{cases} n^{1+\frac{1}{2}} \mathcal{J}t & \text{if } k \text{ is even,} \\ n^2 \mathcal{J}t & \text{if } k \text{ is odd.} \end{cases} \quad (257)$$

This result is irrespective of the size of locality k , but rather its parity.

Proof. Given a sample $\mathbf{b} \in \{0, 1\}^\Gamma$, let us start from equation (230)

$$\|\mathcal{D}(\mathbf{B} = \mathbf{b})\|_{*,p} \leq \|I\|_{*,p} \cdot p \cdot \lambda(p, r) \quad (258)$$

where

$$\begin{aligned} \lambda(p, r) := & \frac{1}{\sqrt{p}} \alpha(l) \sigma t \cdot \left(\left[\sqrt{p} \sigma \frac{t}{r} \right]^l \sqrt{\sum_{w=0}^{l+1} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right. \\ & \left. + \Gamma \cdot \left[\sqrt{p} \sigma \frac{t}{r} \right]^{l+1} \sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right). \end{aligned} \quad (259)$$

By Lemma (2.1), finding the Trotter number r to ensure the probability statement

$$\mathbb{P}(\|\mathcal{D}(\mathbf{B} = \mathbf{b})\| \geq \epsilon) \leq \delta, \quad (260)$$

is equivalent to finding r such that

$$\begin{aligned} 1 \geq & e\sqrt{p_D} \alpha(l) \sigma \frac{t}{\epsilon} \cdot \left(\left[\sqrt{p_D} \sigma \frac{t}{r} \right]^l \sqrt{\sum_{w=0}^{l+1} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right. \\ & \left. + \Gamma \cdot \left[\sqrt{p_D} \sigma \frac{t}{r} \right]^{l+1} \sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right). \end{aligned} \quad (261)$$

where $p_D = \log(e^2 D / \delta)$ with $D = \|I\|_p^p$. Hence, we find that

$$\begin{aligned} r \geq & \sqrt{p_D} \sigma t \cdot \left(\left[2e\sqrt{p_D} \alpha(l) \sigma \frac{t}{\epsilon} \sqrt{\sum_{w=0}^{l+1} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right]^{\frac{1}{l}} \right. \\ & \left. + \left[2e\sqrt{p_D} \alpha(l) \sigma \frac{\Gamma t}{\epsilon} \sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right]^{\frac{1}{l+1}} \right). \end{aligned} \quad (262)$$

For simplicity, we can move the sum over G_w out of the square brackets by considering a higher r :

$$r \geq \left(\sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right)^{\frac{1}{l}} \cdot R(n) \quad (263)$$

with $R(n)$ defined as

$$R(n) := \sqrt{p_D} \sigma t \cdot \left(\left[2e\sqrt{p_D} \alpha(l) \sigma \frac{t}{\epsilon} \right]^{\frac{1}{l}} + \left[2e\sqrt{p_D} \alpha(l) \sigma \frac{\Gamma t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \quad (264)$$

which is a coefficient that appears in the Trotter number of all configurations.

The gate complexity specific to this configuration is then upper bounded by

$$G(\mathbf{B} = \mathbf{b}) \leq |\mathbf{b}| \cdot r = R(n) \cdot |\mathbf{b}| \left(\sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B} = \mathbf{b}))} \right)^{\frac{1}{l}}. \quad (265)$$

It is an upper bound because we did not consider the minimal r .

Averaging over all Bernoulli variables gives us

$$\langle G(\mathbf{B}) \rangle \leq R(n) \left\langle |\mathbf{B}| \left(\sqrt{\sum_{w=0}^{l+2} G_w(H_{\text{SSYK}}(\mathbf{B}))} \right)^{\frac{1}{l}} \right\rangle. \quad (266)$$

By the Cauchy–Schwarz inequality, and applying Jensen’s inequality for the square root term, we get

$$\langle G(\mathbf{B}) \rangle \leq R(n) \sqrt{\langle |\mathbf{B}|^2 \rangle} \sqrt{\left\langle \left(\sum_{w=0}^{l+2} \langle G_w(H_{\text{SSYK}}(\mathbf{B})) \rangle \right)^{\frac{1}{2}} \right\rangle}. \quad (267)$$

The first expectation can be computed from the second moment of the binomial distribution:

$$\begin{aligned} \langle |\mathbf{B}|^2 \rangle &= \sum_{b=0}^{\Gamma} b^2 \binom{\Gamma}{b} p_B^b (1-p_B)^{\Gamma-b} \\ &= \Gamma(\Gamma-1)p_B^2 + \Gamma p_B. \end{aligned} \quad (268)$$

For the second expectation, we get

$$\langle G_w(H_{\text{SSYK}}(\mathbf{B})) \rangle \leq (p_B Q(n, k))^g \cdot (g+1)^4 \quad (269)$$

if $p_B Q(n, k) \geq 1$, and

$$\langle G_w(H_{\text{SSYK}}(\mathbf{B})) \rangle \leq (g+1)^4 \quad (270)$$

if $p_B Q(n, k) \leq 1$.

We can finally put everything together. If $p_B Q(n, k) \geq 1$, we get

$$\langle G(\mathbf{B}) \rangle \leq A(n) \sqrt{\Gamma(\Gamma-1)p_B^2 + \Gamma p_B} \left(\sqrt{p_B Q(n, k)} \right)^{1+\frac{2}{l}} \quad (271)$$

where

$$A(n) := (l+3)^{\frac{5}{2l}} R(n) = (l+3)^{\frac{5}{2l}} \sqrt{p_D} \sigma t \cdot \left(\left[2e\sqrt{p_D} \alpha(l) \sigma \frac{t}{\epsilon} \right]^{\frac{1}{l}} + \left[2e\sqrt{p_D} \alpha(l) \sigma \frac{\Gamma t}{\epsilon} \right]^{\frac{1}{l+1}} \right). \quad (272)$$

If $p_B Q(n, k) \leq 1$, we get

$$\langle G(\mathbf{B}) \rangle \leq A(n) \sqrt{\Gamma(\Gamma-1)p_B^2 + \Gamma p_B}. \quad (273)$$

Since we want to study the behavior of the gate complexity asymptotic in n , while fixing other parameters such as the sparsity κ , locality k , and evolution time t , we can safely assume that

$$p_B Q(n, k) \geq 1 \quad \text{and} \quad (p_B \Gamma)^m \geq p_B \Gamma \quad (274)$$

for all $m \geq 1$. Since $D = \|I\|_p^p = 2^{n/2}$, and $\sigma = \mathcal{O}(\mathcal{J})$, suppressing all l - and k -dependent factors gives us

$$A(n) = \tilde{\mathcal{O}} \left(\sqrt{n + \log(e^2/\delta)} \mathcal{J} t \cdot \left(\left[\sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^k \sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right). \quad (275)$$

Using this, we get the following asymptotic average gate complexities:

$$\bar{G} = \tilde{\mathcal{O}} \left(n \left(\sqrt{p_B Q(n, k)} \right)^{1+\frac{2}{l}} \sqrt{n + \log(e^2/\delta)} \mathcal{J} t \cdot \left(\left[\sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^k \sqrt{n + \log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right), \quad (276)$$

As for $Q(n, k)$, it's behavior depends on the parity of locality k . If k is even, $Q(n, k) = \mathcal{O}(n^{k-1})$, and if k is odd, $Q(n, k) = \mathcal{O}(n^k)$. Suppressing all constants other than n , t and \mathcal{J} , we get the following scaling for the asymptotic average gate complexities when l is sufficiently large:

$$\bar{G} \sim \begin{cases} n^{1+\frac{1}{2}} \mathcal{J} t & \text{if } k \text{ is even,} \\ n^2 \mathcal{J} t & \text{if } k \text{ is odd.} \end{cases} \quad (277)$$

This proves the statement. □

Similar to Section 3, by consider the probability statement in equation (50) and applying Lemma 2.2 instead of Lemma 2.1, we get the following result for simulating an arbitrary fixed input state:

Corollary 5.2.2. *For an arbitrary fixed input state $|\psi\rangle$, simulating the k -local κ -sparse SYK model using (even) l -th order product formula has average gate complexity*

$$\bar{G} = \tilde{\mathcal{O}} \left(n \left(\sqrt{p_B Q(n, k)} \right)^{1+\frac{2}{l}} \sqrt{\log(e^2/\delta)} \mathcal{J} t \cdot \left(\left[\sqrt{\log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l}} + \left[n^k \sqrt{\log(e^2/\delta)} \frac{\mathcal{J} t}{\epsilon} \right]^{\frac{1}{l+1}} \right) \right), \quad (278)$$

in the sense that \bar{G} upper bounds the average of the set

$$\mathcal{G} = \{G(\mathbf{B} = \mathbf{b}) : \mathbf{b} \in [0, 1]^\Gamma\}, \quad (279)$$

where $G(\mathbf{B} = \mathbf{b})$ is the minimal gate complexity of $S_l(t/r)^r$ approximating the configuration $H_{\text{SSYK}}(\mathbf{B} = \mathbf{b})$ that ensures the probability statement

$$\mathbb{P} \left(\left\| \left(e^{iH_{\text{SSYK}}(\mathbf{B}=\mathbf{b})t} - S_l(t/r)^r \right) |\psi\rangle \right\|_{l^2} \geq \epsilon \right) < \delta. \quad (280)$$

The term $Q(n, k)$ is given by equation (96). When l is sufficiently large, the average gate complexity scales in n as

$$\bar{G} \sim \begin{cases} n \mathcal{J} t & \text{if } k \text{ is even,} \\ n^{1+\frac{1}{2}} \mathcal{J} t & \text{if } k \text{ is odd.} \end{cases} \quad (281)$$

This result is irrespective of the size of locality k , but rather its parity.

6 Discussion

6.1 Gate complexities for simulating the SYK model

We obtained bounds for various orders l of the Trotter error, and estimated the gate complexity of simulating the SYK model with accuracy described by the probability statements in equations (49) and (50).

The probability statement in equation (49) characterizes small Trotter error by the difference in spectral norm between the time-evolution operator and its product formula approximation. The gate complexities, displayed in Table 3, for simulating the SYK model ensure the probability statement in equation (49). In Table 3, we only focus on the scaling in n and t while keeping all other parameters fixed. As mentioned earlier, the $\log n$ overhead from the fermion-to-qubit mapping is omitted. The results in the large l limit are obtained by taking l sufficiently large, such that the terms with $\frac{1}{l}$ and $\frac{1}{l+1}$ in their exponent become irrelevant to the scaling behavior.

Orders of product formulas	Gate complexity G	
	Even k	Odd k
First-order	$n^{k+\frac{5}{2}}t^2$	$n^{k+3}t^2$
l -th order	$n^{k+\frac{1}{2}}t \cdot \left(\left[n^{\frac{3}{2}}t \right]^{\frac{1}{l}} + \left[n^{k+\frac{3}{2}}t \right]^{\frac{1}{l+1}} \right)$	$n^{k+1}t \cdot \left([nt]^{\frac{1}{l}} + [n^{k+1}t]^{\frac{1}{l+1}} \right)$
Large l limit	$n^{k+\frac{1}{2}}t$	$n^{k+1}t$

Table 3: The gate complexity of using different orders of product formulas to simulate the SYK model of locality k , ensuring the probability statement in equation (49), for fixed $\epsilon > 0$, $\delta \in (0, 1)$ and $k \leq 1$, omitting the $\log n$ overhead produced by the fermion-to-qubit mapping. We omitted the big- \tilde{O} notation in the cells. The ϵ -scaling of these bounds go with $\sim \frac{1}{\epsilon}$ for first-order formulas and $\sim \left(\frac{1}{\epsilon}\right)^{\frac{1}{l}}$ for l -th order formulas.

We observe that the scaling behavior of the gate complexities is different for even k compared to odd k . This is a direct consequence of $Q(n, k)$, see equation (96), which behaves differently for even and odd k . As a result, the Trotter error, and subsequently the corresponding gate complexities, will have different scaling behaviors. We expect this phenomenon, since the SYK model behaves differently depending on the parity of its locality k . It is also reasonable that the gate complexity for simulating odd- k SYK models is higher than even- k SYK models. As $Q(n, k) = \mathcal{O}(n^k)$ for odd k , most of the terms will anti-commute in the odd- k SYK model, whereas even- k SYK models have $Q(n, k) = \mathcal{O}(n^{k-1})$, meaning that most of the terms commute as n grows. Since the Trotter error depends on the number of anti-commuting pairs of terms, simulating the odd- k SYK models will naturally require a higher gate complexity than the even- k SYK models.

By ignoring the $\log n$ overhead produced by the fermion-to-qubit mapping, we found that for the SYK model with $k = 4$, the gate complexity of its digital quantum simulation using first-order product formula scales with $\tilde{O}(n^{6.5}t^2)$, while using higher-order product formulas (given the order l is sufficiently large) it scales with $\tilde{O}(n^{4.5}t)$. This offers an improvement of $\mathcal{O}(n^{3.5})$ and $\mathcal{O}(n^{5.5})$ in n (up to log-factors) respectively compared to the previous result $\mathcal{O}(n^{10}t^2)$ obtained by [GAEL+17]. Compared to the result $\mathcal{O}(n^5t^2)$ [AJS24], our higher-order bound offers a slight improvement of $\mathcal{O}(\sqrt{n})$ in n .

Additionally, any order of product formula simulating the SYK model requires at least $\Omega(n^k)$ gates, because there are $\mathcal{O}(n^k)$ local term exponentials per round in the product formula due to the fact that the SYK model has $\Gamma = \binom{n}{k} = \mathcal{O}(n^k)$ local terms. In this sense, one could interpret our results $\tilde{O}(n^{k+1/2}t)$ and $\tilde{O}(n^{k+1}t)$ in the large l limit as close to optimal.

These gate complexities can be further improved upon, if we consider the probability statement in equation (50), which characterizes small Trotter error by the difference in the l^2 -norm between the time evolution of an arbitrary fixed input state $|\psi\rangle$ and the state obtained by applying the product formula approximation on $|\psi\rangle$. Overall, considering this probability statement allows us to reduce the previous gate complexities by $\mathcal{O}(n^2)$ for first-order product formulas and by $\mathcal{O}(\sqrt{n})$ for higher-order formulas. These results are presented in Table 4, where we only focus on the scaling in n and t while keeping all other parameters fixed. The results in the large l limit are obtained in the same way by taking l sufficiently large, such that the terms with $\frac{1}{l}$ and $\frac{1}{l+1}$ in their exponent vanish. As we can see in Table 4, simulating the SYK model with even k using higher-order formulas requires only $\tilde{O}(n^k t)$ gates in this case. This means that for any arbitrary fixed input state $|\psi\rangle$, higher-order product formulas are actually optimal for simulating the time-evolution of $|\psi\rangle$ under the SYK Hamiltonian with even k , apart from the inevitable $\log n$ overhead from the fermion-to-qubit mapping.

Orders of product formulas	Gate complexity G	
	Even k	Odd k
First-order	$n^{k+\frac{1}{2}}t^2$	$n^{k+1}t^2$
l -th order	$n^k t \cdot \left([nt]^{\frac{1}{l}} + [n^{k+1}t]^{\frac{1}{l+1}} \right)$	$n^{k+\frac{1}{2}} t \cdot \left(\left[n^{\frac{1}{2}}t \right]^{\frac{1}{l}} + \left[n^{k+\frac{1}{2}}t \right]^{\frac{1}{l+1}} \right)$
Large l limit	$n^k t$	$n^{k+\frac{1}{2}} t$

Table 4: The gate complexities of using different orders of product formulas to simulate the SYK model of locality k on an arbitrary fixed input state $|\psi\rangle$, ensuring the probability statement in equation (50), for fixed $\epsilon > 0$, $\delta \in (0, 1)$ and $k \leq 1$, omitting the $\log n$ overhead produced by the fermion-to-qubit mapping. We omitted the big- \tilde{O} notation in the cells. The ϵ -scaling of these bounds go with $\sim \frac{1}{\epsilon}$ for first-order formulas and $\sim \left(\frac{1}{\epsilon}\right)^{\frac{1}{l}}$ for l -th order formulas.

6.2 Gate complexities for simulating the sparse SYK model

Additionally, using similar methods, we derived average higher-order Trotter error bounds for simulating the sparse SYK model. They lead to the following average gate complexity in the large l -limit¹⁴:

$$\overline{G} \sim \begin{cases} n^{1+\frac{1}{2}}\mathcal{J}t & \text{if } k \text{ is even,} \\ n^2\mathcal{J}t & \text{if } k \text{ is odd.} \end{cases} \quad (282)$$

Compared to the full SYK model, this result does not depend on the locality k , but rather its parity. Since the sparse SYK model on average contains κn terms in the Hamiltonian, the gate complexity of product formulas scales at least with κn . Hence, similar to the case for the dense SYK model, one could interpret our result in the large l limit as close to optimal.

Similarly, if we consider the probability statement for an arbitrary fixed input state $|\psi\rangle$ in equation (50) instead of the statement in equation (49), we get

$$\overline{G} \sim \begin{cases} n\mathcal{J}t & \text{if } k \text{ is even,} \\ n^{1+\frac{1}{2}}\mathcal{J}t & \text{if } k \text{ is odd.} \end{cases} \quad (283)$$

This shows that for any fixed input state $|\psi\rangle$, higher-order product formulas are near-optimal on average for simulating the time evolution of $|\psi\rangle$ under the sparse SYK Hamiltonian with even k , apart from the inevitable $\log n$ overhead cost from the fermion-to-qubit mapping. Together with the results in the previous section, we demonstrated the potential of product formulas to efficiently simulate the SYK and sparse SYK models.

6.3 Error bounds for general Gaussian models

Due to the generality of Lemma 4.2, our Trotter error bounds for the SYK Hamiltonians in Theorems 3.1 and 4.1 can be easily generalized to other Gaussian models by simply replacing $Q(n, k)$ with Q_{\max} (equation (166)) and $\binom{n}{k}$ with the total number of terms in the specific Hamiltonian. In other words, consider a Hamiltonian of the form

$$H = \sum_{i=1}^{\Gamma} J_i H_i, \quad (284)$$

¹⁴The complexities present here have an ϵ -scaling that goes with $\sim \left(\frac{1}{\epsilon}\right)^{\frac{1}{l}}$ for l -th order product formulas.

where the J_i 's are i.i.d. Gaussian variables with zero mean and standard deviation σ , and the H_i 's are linearly independent Pauli strings or Pauli string representations of fermionic operators of arbitrary locality. For $p \geq 2$, the first-order normalized Trotter error is bounded by

$$\Delta_1(H) := 4\sqrt{2}p^2\sigma^2\sqrt{\Gamma Q_{\max}(H)}t^2\left[\frac{1}{2r} + \sigma\sqrt{Q_{\max}(H)}\frac{t}{3r^2}\right], \quad (285)$$

and the l -th order normalized Trotter error is bounded by

$$\Delta_l(H) := \frac{\mathcal{D}(l)\sqrt{p}\sigma t}{\sqrt{Q_{\max}(H)}}\left(\Gamma\left[\sqrt{p}\sigma\sqrt{Q_{\max}(H)}\frac{t}{r}\right]^l + \Gamma^2\left[\sqrt{p}\sigma\sqrt{Q_{\max}(H)}\frac{t}{r}\right]^{l+1}\right), \quad (286)$$

where $\mathcal{D}(l)$ is a factor only depending on l . Hence, bounding the Trotter error of a Hamiltonian of the form (284) reduces to the problem of counting the maximal number of anti-commuting terms in the Hamiltonian. Subsequently, one can use equations (285) and (286) to estimate the corresponding gate complexity as we did for the SYK models.

Similarly, we can generalize our results for the sparse SYK model to a broader class of sparse Hamiltonians where the sparsification is governed by Bernoulli processes similar to the sparse SYK model. To be specific, given a full Hamiltonian

$$H_{\text{full}} := \sum_{i=1}^m J_i H_i, \quad (287)$$

we consider its sparsification of the form

$$H_{\text{sparse}} = \sum_{i=1}^m B_i J_i H_i \quad (288)$$

where B_i 's are i.i.d. Bernoulli variables and J_i 's are i.i.d. Gaussian variables, such that given constants $\kappa, \mathcal{J} \geq 0$,

$$\mathbb{P}(B_i = 1) = p_B, \quad \mathbb{E}(J_i) = 0, \quad \text{and} \quad \mathbb{E}(J_i^2) = \mathcal{O}(\mathcal{J}^2) \quad (289)$$

for all $1 \leq i \leq m$. The H_i 's are linearly independent Pauli strings (or can be represented by Pauli strings, such as the Majoranas) of arbitrary locality. Then, by simply replacing $Q(n, k)$ with $Q_{\max}(H_{\text{full}})$ in equations (210) and (211), the average Trotter error is bounded by

$$\frac{\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle}{\|I\|_{*,p}} \leq \beta(l) \frac{\Gamma\sqrt{p}\sigma\sqrt{p_B}t}{\sqrt{Q_{\max}(H_{\text{full}})}} \cdot \left(\left[\sqrt{p}\sigma\sqrt{p_B Q_{\max}(H_{\text{full}})}\frac{t}{r} \right]^l + \Gamma \cdot \left[\sqrt{p}\sigma\sqrt{p_B Q_{\max}(H_{\text{full}})}\frac{t}{r} \right]^{l+1} \right), \quad (290)$$

where $\beta(l)$ is a purely l -dependent factor. If $p_B Q_{\max}(H_{\text{full}}) \leq 1$ instead, the l -th order average Trotter error is bounded by

$$\frac{\langle \|\mathcal{D}(\mathbf{B})\|_{*,p} \rangle}{\|I\|_{*,p}} \leq \beta(l) \frac{\Gamma\sqrt{p}\sigma t}{Q_{\max}(H_{\text{full}})} \cdot \left(\left[\sqrt{p}\sigma\frac{t}{r} \right]^l + \Gamma \cdot \left[\sqrt{p}\sigma\frac{t}{r} \right]^{l+1} \right). \quad (291)$$

Subsequently, gate complexities can be derived in a similar fashion as Corollaries 5.2.1 and 5.2.2.

6.4 Outlook and open questions

Although we have derived bounds for several orders of Trotter error and found the near-optimal complexity for the product formula simulation of both the SYK and sparse SYK models, there are still some open questions left.

The first one concerns the error bound Δ_l . Recall that by Theorem 4.1,

$$\frac{\|e^{iHt} - S_l(t/r)^r\|_p}{\|I\|_p} \leq \Delta_l(n, k, t, r, p), \quad (292)$$

where

$$\Delta_l := \frac{\mathcal{D}(l)\sqrt{p}\sigma t}{\sqrt{Q(n, k)}} \left(\binom{n}{k} \left[\sqrt{p}\sigma\sqrt{Q(n, k)}\frac{t}{r} \right]^l + \binom{n}{k}^2 \left[\sqrt{p}\sigma\sqrt{Q(n, k)}\frac{t}{r} \right]^{l+1} \right). \quad (293)$$

Asymptotically for even k , note that $\sigma = \mathcal{O}(n^{-\frac{k-1}{2}})$ and $Q(n, k) = \mathcal{O}(n^{k-1})$. Hence, $\sigma\sqrt{Q(n, k)} = \mathcal{O}(1)$. Therefore, for any even $l > 1$, the scaling of Δ_l in n will asymptotically not depend on the terms in the square brackets. In other words, fix k, t, r , and p , then for any even $l > 1$, Δ_l will have the following asymptotic scaling in n :

$$\Delta_l(n, k, t, r, p) = \mathcal{O}\left(\frac{\sigma}{\sqrt{Q(n, k)}} \binom{n}{k}^2\right) = \mathcal{O}(n^{k+1}), \quad (294)$$

which is independent of l . This behavior is illustrated in Figure 5, where the Δ_l for different l start with different curves, but after some n_0 (in Figure 5 we could take $n_0 \in [e^3, e^4]$), they all become parallel straight lines for $n \geq n_0$, indicating the same scaling behavior in n .

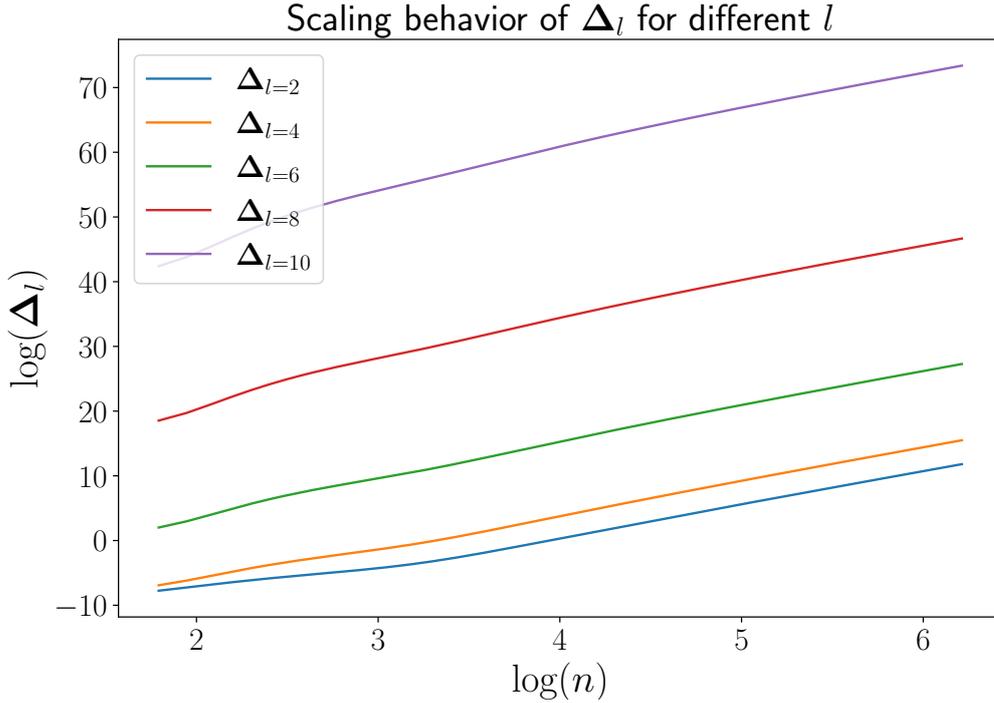


Figure 5: Scaling behavior of Δ_l in $n \in [6, 500]$ for $l = 2, 4, 6, 8, 10$ at fixed $k = 4$, $t = 10$, $r = 100000$, $p = 2$. When plotting these, we omitted the calculation of the prefactor $\mathcal{D}(l)$ because it gets too large for us to compute. Therefore, this plot does not give the correct prefactors but still captures the correct scaling behavior in n .

Since Δ_l is an upper bound for the l -th Trotter error, this indicates that the l -th order Trotter error of the SYK model with even k for any even $l > 1$ could also have this uniform asymptotic scaling independent of l . Numerically confirming this interesting observation by directly computing the Trotter error is beyond the reach of our computational capability. We leave the verification of this result to further research.

The second question addresses the gate complexity of the sparse SYK model. The quantity we provide does not yield the full picture, as it only gives a measure for the gate complexity averaged over all samples of the Bernoulli variables. To obtain a fuller picture, one could consider computing higher-order moments of the gate complexity, such as the variance, and give a probabilistic description via Chebyshev's inequality in the form

$$\mathbb{P}\left(|G(\mathbf{B}) - \overline{G}| \geq c\sqrt{\text{Var}(G(\mathbf{B}))}\right) \leq \frac{1}{c^2}, \quad (295)$$

where $G(\mathbf{B})$ is the gate complexity dependent on the Bernoulli variables, \overline{G} is the average gate complexity we computed, and c is a constant strictly larger than 1. Alternatively, one could define a typical set for the samples of the Bernoulli variables, and provide a statement about gate complexities for typical samples. We leave both approaches for future research.

Lastly, inspired by [AJS24], we could consider a graph-coloring optimization to reduce the number of gates per round in a product formula. This idea is based on the fact that we can decompose the SYK Hamiltonian into clusters of mutually commuting terms. If one considers the anti-commutation graph A of the SYK Hamiltonian, i.e. the graph obtained by placing each local term as a vertex, and an edge between two vertices if the corresponding local terms anti-commute, then the chromatic number χ_A is the least number of commuting clusters we can decompose the SYK Hamiltonian into. The commuting clusters have the advantage that they can be simultaneously diagonalized. Implementing each commuting cluster as Pauli strings requires $\mathcal{O}(n^2)$ gates, coming from the simultaneous diagonalization of the cluster [MCK⁺22, vdBT20, MFL⁺24]. This leads to a complexity of $\mathcal{O}(n^2\chi(A))$ gates per round. Because the anti-commutation graph A is a $Q(n, k)$ -regular graph, its chromatic number χ_A is upper bounded by $Q(n, k) + 1$. Hence, using this approach, the gate complexity per round would scale at most with $\mathcal{O}(n^{k+1})$, which is the same as the gate complexity where we use the Jordan–Wigner transformation to map each Majorana fermion to a Pauli string instead of the ternary tree mapping. However, the chromatic number χ_A can potentially exhibit a better scaling than $Q(n, k)$, as we can see in Figure 6. Hence, although decomposing the SYK Hamiltonian into commuting clusters does not impact the Trotter number as it is not ordering dependent, this approach could potentially lead to a more efficient implementation of single rounds of the product formulas.

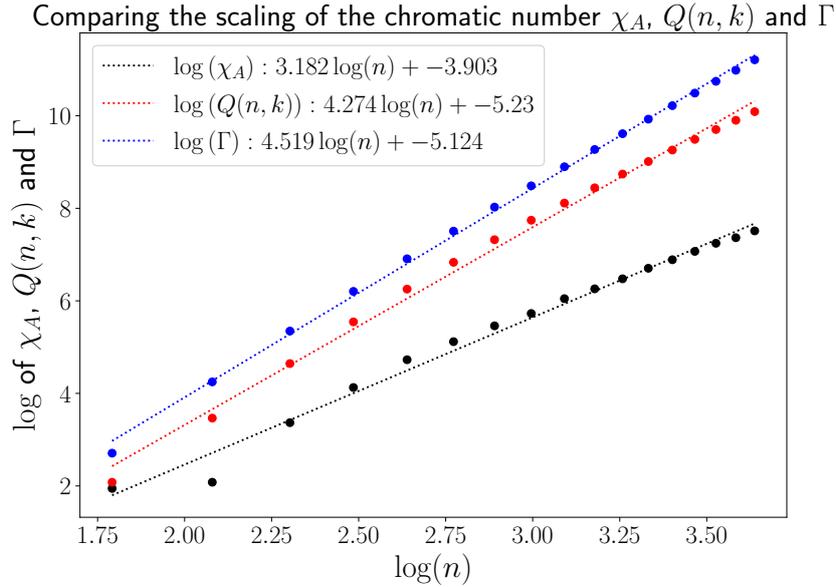


Figure 6: Scaling comparison between the chromatic number χ_A of the anti-commutation graph of the SYK model A and $Q(n, k)$ for even $n \in [6, 38]$ and $k = 4$. The chromatic number is estimated using a greedy coloring algorithm.

References

- [AGK25] Eric R. Anschuetz, David Gamarnik, and Bobak T. Kiani. Bounds on the ground state energy of quantum p -spin Hamiltonians. *Communications in Mathematical Physics*, 406(10), September 2025. [arXiv:2404.07231](#), [doi:10.1007/s00220-025-05412-4](#).
- [AJS24] Muhammad Asaduzzaman, Raghav G. Jha, and Bharath Sambasivam. Sachdev-Ye-Kitaev model on a noisy quantum computer. *Phys. Rev. D*, 109:105002, May 2024. [arXiv:2311.17991](#), [doi:10.1103/PhysRevD.109.105002](#).
- [BACS06] Dominic W. Berry, Graeme Ahokas, Richard Cleve, and Barry C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Communications in Mathematical Physics*, 270(2):359–371, December 2006. [arXiv:quant-ph/0508139](#), [doi:10.1007/s00220-006-0150-x](#).
- [BBN19] Ryan Babbush, Dominic W. Berry, and Hartmut Neven. Quantum simulation of the Sachdev-Ye-Kitaev model by asymmetric qubitization. *Phys. Rev. A*, 99:040301, Apr 2019. [arXiv:1806.02793](#), [doi:10.1103/PhysRevA.99.040301](#).
- [BCL94] Keith Ball, Eric A. Carlen, and Elliott H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones mathematicae*, 115:463–482, 1994. [doi:10.1007/BF01231769](#).
- [CB24] Chi-Fang Chen and Fernando Brandão. Average-case speedup for product formulas. *Communications in Mathematical Physics*, 405, 02 2024. [arXiv:2111.05324](#), [doi:10.1007/s00220-023-04912-5](#).
- [Che] Yiyuan Chen. GitHub repository for the source codes of the numerics. [github.com/cyy020726/Trotter-error-and-gate-complexity-of-the-SYK-and-sparse-SYK-models](#). Last updated on 2025-11-02.
- [Cra07] Nicholas Crawford. Thermodynamics and universality for mean field quantum spin glasses. *Communications in Mathematical Physics*, 274(3):821–839, May 2007. [arXiv:math-ph/0610031](#), [doi:10.1007/s00220-007-0263-x](#).
- [CS19] Andrew M. Childs and Yuan Su. Nearly optimal lattice simulation by product formulas. *Phys. Rev. Lett.*, 123:050503, Aug 2019. [arXiv:1901.00564](#), [doi:10.1103/PhysRevLett.123.050503](#).
- [CST⁺21] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. Theory of Trotter error with commutator scaling. *Phys. Rev. X*, 11:011020, Feb 2021. [arXiv:1912.08854](#), [doi:10.1103/PhysRevX.11.011020](#).
- [GAEL⁺17] L. García-Álvarez, I. L. Egusquiza, L. Lamata, A. del Campo, J. Sonner, and E. Solano. Digital quantum simulation of minimal AdS/CFT. *Physical Review Letters*, 119(4), July 2017. [arXiv:1607.08560](#), [doi:10.1103/physrevlett.119.040501](#).
- [GGJV18] Antonio M. García-Garcá, Yiyang Jia, and Jacobus J. M. Verbaarschot. Exact moments of the Sachdev-Ye-Kitaev model up to order $1/N^2$. *Journal of High Energy Physics*, 2018(4), April 2018. [arXiv:1801.02696](#), [doi:10.1007/jhep04\(2018\)146](#).
- [GMJW23] Ignacio García-Mata, Rodolfo Jalabert, and Diego Wisniacki. Out-of-time-order correlations and quantum chaos. *Scholarpedia*, 18(4):55237, 2023. [arXiv:1607.08560](#), [doi:10.4249/scholarpedia.55237](#).
- [HMY17] Koji Hashimoto, Keiju Murata, and Ryosuke Yoshii. Out-of-time-order correlators in quantum mechanics. *Journal of High Energy Physics*, 2017(10), October 2017. [arXiv:2209.07965](#), [doi:10.1007/jhep10\(2017\)138](#).
- [HNWTW22] De Huang, Jonathan Niles-Weed, Joel A. Tropp, and Rachel Ward. Matrix concentration for products. *Found. Comput. Math.*, 22(6):1767–1799, December 2022. [arXiv:2003.05437](#), [doi:10.1007/s10208-021-09533-9](#).

- [JKMN20] Zhang Jiang, Amir Kalev, Wojciech Mruczkiewicz, and Hartmut Neven. Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning. *Quantum*, 4:276, June 2020. [arXiv:1910.10746](https://arxiv.org/abs/1910.10746), [doi:10.22331/q-2020-06-04-276](https://doi.org/10.22331/q-2020-06-04-276).
- [JY17] Shao-Kai Jian and Hong Yao. Solvable Sachdev-Ye-Kitaev models in higher dimensions: From diffusion to many-body localization. *Physical Review Letters*, 119(20), November 2017. [arXiv:1703.02051](https://arxiv.org/abs/1703.02051), [doi:10.1103/physrevlett.119.206602](https://doi.org/10.1103/physrevlett.119.206602).
- [JZL⁺22] Daniel Jafferis, Alexander Zlokapa, Joseph D. Lykken, David K. Kolchmeyer, Samantha I. Davis, Nikolai Lauk, Hartmut Neven, and Maria Spiropulu. Traversable wormhole dynamics on a quantum processor. *Nature*, 612(7938):51–55, Dec 2022. [doi:10.1038/s41586-022-05424-3](https://doi.org/10.1038/s41586-022-05424-3).
- [Kita] Alexei Kitaev. A simple model of quantum holography (part 1). <https://online.kitp.ucsb.edu/online/entangled15/kitaev/>. Retrieved on 29-4-2024.
- [Kitb] Alexei Kitaev. A simple model of quantum holography (part 2). <https://online.kitp.ucsb.edu/online/entangled15/kitaev2/>. Retrieved on 29-4-2024.
- [MCK⁺22] Edison M. Murairi, Michael J. Cervia, Hersh Kumar, Paulo F. Bedaque, and Andrei Alexandru. How many quantum gates do gauge theories require? *Phys. Rev. D*, 106:094504, Nov 2022. [arXiv:2208.11789](https://arxiv.org/abs/2208.11789), [doi:10.1103/PhysRevD.106.094504](https://doi.org/10.1103/PhysRevD.106.094504).
- [MFL⁺24] Daniel Miller, Laurin E. Fischer, Kyano Levi, Eric J. Kuehnke, Igor O. Sokolov, Panagiotis Kl. Barkoutsos, Jens Eisert, and Ivano Tavernelli. Hardware-tailored diagonalization circuits. *npj Quantum Information*, 10(1), November 2024. [arXiv:2203.03646](https://arxiv.org/abs/2203.03646), [doi:10.1038/s41534-024-00901-1](https://doi.org/10.1038/s41534-024-00901-1).
- [Mon20] Swapnamay Mondal. A simple model for Hawking radiation. *Journal of High Energy Physics*, 2020(3), March 2020. [arXiv:1901.03483](https://arxiv.org/abs/1901.03483), [doi:10.1007/jhep03\(2020\)119](https://doi.org/10.1007/jhep03(2020)119).
- [MS16] Juan Maldacena and Douglas Stanford. Remarks on the Sachdev-Ye-Kitaev model. *Physical Review D*, 94(10), November 2016. [arXiv:1604.07818](https://arxiv.org/abs/1604.07818), [doi:10.1103/physrevd.94.106002](https://doi.org/10.1103/physrevd.94.106002).
- [Nie05] Michael A. Nielsen. The Fermionic canonical commutation relations and the Jordan–Wigner transform. 2005. URL: https://futureofmatter.com/assets/fermions_and_jordan_wigner.pdf.
- [OGP24] Patrick Orman, Hrant Gharibyan, and John Preskill. Quantum chaos in the sparse SYK model, 2024. [arXiv:2403.13884](https://arxiv.org/abs/2403.13884), [doi:10.48550/arXiv.2403.13884](https://doi.org/10.48550/arXiv.2403.13884).
- [Ros19] Vladimir Rosenhaus. An introduction to the SYK model. *Journal of Physics A: Mathematical and Theoretical*, 52(32):323001, July 2019. [arXiv:1807.03334](https://arxiv.org/abs/1807.03334), [doi:10.1088/1751-8121/ab2ce1](https://doi.org/10.1088/1751-8121/ab2ce1).
- [Som16] Rolando D. Somma. A Trotter-Suzuki approximation for Lie groups with applications to Hamiltonian simulation. *Journal of Mathematical Physics*, 57(6), June 2016. [arXiv:1512.03416](https://arxiv.org/abs/1512.03416), [doi:10.1063/1.4952761](https://doi.org/10.1063/1.4952761).
- [Suz91] Masuo Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *Journal of Mathematical Physics*, 32(2):400–407, February 1991. [doi:10.1063/1.529425](https://doi.org/10.1063/1.529425).
- [TJ74] Nicole Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of the trace classes S_p ($1 \leq p < \infty$). *Studia Mathematica*, 50(2):163–182, 1974. URL: <http://eudml.org/doc/217886>.
- [vdBT20] Ewout van den Berg and Kristan Temme. Circuit optimization of Hamiltonian simulation by simultaneous diagonalization of Pauli clusters. *Quantum*, 4:322, September 2020. [arXiv:2003.13599](https://arxiv.org/abs/2003.13599), [doi:10.22331/q-2020-09-12-322](https://doi.org/10.22331/q-2020-09-12-322).
- [XSS20] Shenglong Xu, Leonard Susskind, Yuan Su, and Brian Swingle. A sparse model of quantum holography, 2020. [arXiv:2008.02303](https://arxiv.org/abs/2008.02303), [doi:10.48550/arXiv.2008.02303](https://doi.org/10.48550/arXiv.2008.02303).

A Lemmas for evaluating the probability statements

Lemma 2.1 (Concentration inequality for operator norm). *Let H be a Hamiltonian on a Hilbert space. Let $l \in \{1\} \cup 2\mathbb{N}$ be the order of the product formula approximating the Hamiltonian. If the Trotter error of this product formula is bounded by a positively valued function $\lambda(p, r)$, in the sense that*

$$\| \|e^{iHt} - S_l(t/r)^r\| \|I\|_p \leq p\lambda(p, r) \|I\|_p \quad (55)$$

for all $p \geq 2$, then choosing the Trotter number r so that it obeys the inequality

$$\frac{\lambda\left(\log\left(e^2 \|I\|_p^p / \delta\right), r\right)}{\epsilon} \leq \frac{1}{e \cdot \log\left(e^2 \|I\|_p^p / \delta\right)} \quad (56)$$

ensures that

$$\mathbb{P}\left(\| \|e^{iHt} - S_l(t/r)^r\| \geq \epsilon\right) < \delta. \quad (57)$$

Proof. Let H be a Hamiltonian on a Hilbert space of dimension D . Suppose its l -th order Trotter error is bounded by

$$\| \|e^{iHt} - S_l(t/r)^r\| \|I\|_p \leq p\lambda(p, r) \|I\|_p \quad (296)$$

for all $p \geq 2$. By applying Markov's inequality, we obtain

$$\mathbb{P}\left(\| \|e^{iHt} - S_l(t/r)^r\| \geq \epsilon\right) \leq \frac{\| \|e^{iHt} - S_l(t/r)^r\|_p^p}{\epsilon^p} \leq D \left(p \frac{\lambda(p, r)}{\epsilon} \right)^p, \quad (297)$$

where $D = \|I\|_p^p$. We want to achieve $D \left(p \frac{\lambda(p, r)}{\epsilon} \right)^p < \delta$. Let $p_D = \log(e^2 D / \delta)$. Since equation (297) holds for all $p \geq 2$ and $p_D \geq 2$, it must also hold for p_D .¹⁵

Observe the following equivalence:

$$\left(p_D \frac{\lambda(p_D, r)}{\epsilon} \right)^{p_D} \leq \frac{\delta}{D} \iff \log\left(\left(p_D \frac{\lambda(p_D, r)}{\epsilon} \right)^{p_D}\right) \leq \log\left(\frac{\delta}{D}\right) \quad (298)$$

$$\iff \log\left(p_D \frac{\lambda(p_D, r)}{\epsilon}\right) \cdot p_D \leq \log\left(\frac{\delta}{D}\right) \quad (299)$$

$$\iff \log\left(\frac{\lambda(p_D, r)}{\epsilon}\right) \leq \frac{-\log(D/\delta)}{p_D} - \log(p_D) \quad (300)$$

$$\iff \frac{\lambda(p_D, r)}{\epsilon} \leq \exp\left(\frac{-\log(D/\delta)}{\log(D/\delta) + 2}\right) \cdot \frac{1}{p_D}. \quad (301)$$

Now, the minimal value of $e^{\frac{-x}{x+2}}$ for $x \geq 0$ is at $x \rightarrow \infty$ with $e^{-1} \leq e^{\frac{-x}{x+2}}$ for all $x \geq 0$. Thus,

$$\frac{\lambda(p_D, r)}{\epsilon} \leq \frac{1}{e} \cdot \frac{1}{p_D} \implies \frac{\lambda(p_D, r)}{\epsilon} \leq \exp\left(\frac{-\log D/\delta}{p_D}\right) \cdot \frac{1}{p_D} \iff \left(p_D \frac{\lambda(p_D, r)}{\epsilon} \right)^{p_D} \leq \frac{\delta}{D} \quad (302)$$

which means

$$\frac{\lambda(p_D, r)}{\epsilon} \leq \frac{1}{e} \cdot \frac{1}{p_D} \implies \mathbb{P}\left(\| \|e^{iHt} - S_l(t/r)^r\| > \epsilon\right) \leq \delta. \quad (303)$$

In other words, solving

$$\frac{\lambda(p_D, r)}{\epsilon} \leq \frac{1}{e} \cdot \frac{1}{p_D} \quad (304)$$

for r ensures the concentration inequality. This concludes the lemma. \square

¹⁵We have $p_D \geq 2$ because $D > \delta$, since $D > 1$ and $\delta \in (0, 1)$ for it bounds a probability.

Lemma 2.2 (Concentration inequality for fixed input state). *Let H be a Hamiltonian on a Hilbert space of dimension D and $|\psi\rangle$ an arbitrary fixed input state in this space. Let $l \in \{1\} \cup 2\mathbb{N}$ be the order of the product formula approximating the Hamiltonian. If the Trotter error of this product formula is bounded by a positively valued function $\lambda(p, r)$, in the sense that*

$$\| \|e^{iHt} - S_l(t/r)^r \|_{\text{fix}, p} \leq p\lambda(p, r) \quad (58)$$

for all $p \geq 2$, then choosing the Trotter number r so that it obeys the inequality

$$\frac{\lambda(\log(e^2/\delta), r)}{\epsilon} \leq \frac{1}{e \cdot \log(e^2/\delta)} \quad (59)$$

ensures that

$$\mathbb{P}(\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l^2} \geq \epsilon) < \delta. \quad (60)$$

Proof. Let H be a Hamiltonian on a Hilbert space of dimension D and $|\psi\rangle$ an arbitrary fixed input state in the Hilbert space. Suppose the l -th order Trotter error is bounded by

$$\| \|e^{iHt} - S_l(t/r)^r \|_{\text{fix}, p} \leq p\lambda(p, r) \quad (305)$$

for all $p \geq 2$. Observe that by Markov's inequality

$$\mathbb{P}(\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l^2} \geq \epsilon) \leq \frac{\mathbb{E}(\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l^2}^p)}{\epsilon^p} \quad (306)$$

$$\leq \sup_{\|\phi\|_{l^2}=1} \left\{ \frac{\mathbb{E}(\| (e^{iHt} - S_l(t/r)^r) \phi\|_{l^2}^p)}{\epsilon^p} \right\}. \quad (307)$$

Hence,

$$\mathbb{P}(\| (e^{iHt} - S_l(t/r)^r) |\psi\rangle \|_{l^2} \geq \epsilon) \leq \frac{\| \|e^{iHt} - S_l(t/r)^r \|_{\text{fix}, p}^p}{\epsilon^p} \leq \left(p \frac{\lambda(p, r)}{\epsilon} \right)^p. \quad (308)$$

We want to achieve $\left(p \frac{\lambda(p, r)}{\epsilon} \right)^p < \delta$. Let $p_1 = \log(e^2/\delta)$. Since equation (308) holds for all $p \geq 2$ and $p_1 \geq 2$, it must also hold for p_1 . The rest of the proof is analogous to the previous one by replacing p_D with p_1 . This concludes the lemma. \square

B Anti-commutation lemma for Majorana fermions

Lemma 3.2. *Let $H = \sum_{i=1}^{\Gamma} H_{\gamma(i)}$ be the k -local SYK model with $\gamma : [\Gamma] \rightarrow E$ an arbitrary ordering of hyperedges. Then for all $1 \leq i < j \leq \Gamma$,*

$$H_{\gamma(i)} H_{\gamma(j)} = (-1)^{k+m} H_{\gamma(j)} H_{\gamma(i)}, \quad (97)$$

where $m = |\gamma(i) \cap \gamma(j)|$ is the overlap between hyperedges $\gamma(i)$ and $\gamma(j)$.

Proof. Let $\alpha = \gamma(i)$ and $\beta = \gamma(j)$ two different hyperedges. Suppose $H_\alpha = J_\alpha \chi_{\alpha_1} \cdots \chi_{\alpha_k}$ and $H_\beta = J_\beta \chi_{\beta_1} \cdots \chi_{\beta_k}$, where $\alpha_i \neq \alpha_j$ and $\beta_i \neq \beta_j$ for $i \neq j$. Given the anti-commutation relation

$$\{\chi_i, \chi_j\} = 2\delta_{ij} \quad (309)$$

we are going to show

$$H_\alpha H_\beta = (-1)^{k^2-m^2} H_\beta H_\alpha \quad (310)$$

where $m = |\alpha \cap \beta|$.

We write the product

$$H_\alpha H_\beta = J_\alpha J_\beta \chi_{\alpha_1} \cdots \chi_{\alpha_k} \chi_{\beta_1} \cdots \chi_{\beta_k} =: J_\alpha J_\beta(\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k) \quad (311)$$

as an ordered set. Now, suppose $|\alpha \cap \beta| = 0$. This means that H_α and H_β contain completely different fermions. Therefore, by the anti-commutation relation:

$$H_\alpha H_\beta = J_\alpha J_\beta(\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k) \quad (312)$$

$$= (-1)^k \cdot J_\alpha J_\beta(\beta_1, \alpha_1, \dots, \alpha_k, \beta_2, \dots, \beta_k) \quad (313)$$

$$= (-1)^{2k} \cdot J_\alpha J_\beta(\beta_1, \beta_2, \alpha_1, \dots, \alpha_k, \beta_3, \dots, \beta_k) \quad (314)$$

$$= (-1)^{k^2} \cdot J_\alpha J_\beta(\beta_1, \dots, \beta_k, \alpha_1, \dots, \alpha_k) \quad (315)$$

$$= (-1)^{k^2} H_\beta H_\alpha. \quad (316)$$

Now, suppose $|\alpha \cap \beta| = m > 0$. This means that H_α and H_β share common fermions $\{\gamma_1, \dots, \gamma_m\}$. Define

$$\tilde{H}_\alpha = J_\alpha(\alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m), \quad (317)$$

$$\tilde{H}_\beta = J_\beta(\beta_1, \dots, \beta_{k-m}, \gamma_1, \dots, \gamma_m). \quad (318)$$

Note that all permutations π of $(\alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m)$ only yield a parity factor $\sigma(\pi)$ at the front. In other words,

$$\pi(\alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m) = \sigma(\pi) \cdot (\alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m). \quad (319)$$

Since H_α and H_β can be written as permutations of \tilde{H}_α and \tilde{H}_β respectively, we have

$$H_\alpha = \sigma(\pi_1) \tilde{H}_\alpha, \quad \text{and} \quad H_\beta = \sigma(\pi_2) \tilde{H}_\beta \quad (320)$$

where π_1 and π_2 are the corresponding permutations that send H_α and H_β to \tilde{H}_α and \tilde{H}_β , respectively. Hence,

$$\begin{aligned} H_\alpha H_\beta &= \sigma(\pi_1) \tilde{H}_\alpha \cdot \sigma(\pi_2) \tilde{H}_\beta \\ &= \sigma(\pi_1) \sigma(\pi_2) \cdot J_\alpha J_\beta(\alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m, \beta_1, \dots, \beta_{k-m}, \gamma_1, \dots, \gamma_m) \\ &= \sigma(\pi_1) \sigma(\pi_2) \cdot (-1)^{(k-m)k} \cdot J_\alpha J_\beta(\beta_1, \dots, \beta_{k-m}, \alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m, \gamma_1, \dots, \gamma_m) \\ &= \sigma(\pi_1) \sigma(\pi_2) \cdot (-1)^{(k-m)k+m(k-m)} \cdot J_\alpha J_\beta(\beta_1, \dots, \beta_{k-m}, \gamma_1, \dots, \gamma_m, \alpha_1, \dots, \alpha_{k-m}, \gamma_1, \dots, \gamma_m) \\ &= \sigma(\pi_1) \sigma(\pi_2) \cdot (-1)^{k^2-m^2} \cdot \tilde{H}_\beta \tilde{H}_\alpha \\ &= (-1)^{k^2-m^2} H_\beta H_\alpha \\ &= (-1)^{k+m} H_\beta H_\alpha. \end{aligned} \quad \square$$

C Random matrix polynomials with Gaussian coefficients

In the main text we used 2.3, which comes from Chen & Brandão [CB24, Theorem VII.3] (with a slight variation in the prefactor). For completeness, we present a proof of this result in this appendix (which takes a slightly different track than [CB24]).

For context, consider a random matrix polynomial of degree g

$$F(X_m, \dots, X_1) = \sum_i T_i X_{i_g} \cdots X_{i_1} \quad (321)$$

where each term X_i can be written as

$$X_i = g_i K_i \quad (322)$$

with g_i i.i.d. standard Gaussian variables and K_i a deterministic matrix bounded by $\|K_i\| \leq \sigma_i$. By Rademacher expansion and resummation by a subset S of the index set $\mathcal{I} := [m] \times [N]$, we can write F as

$$F = \sum_{x+y=g} \sum_{s_1, \dots, s_x \in S} \sum_{s'_1, \dots, s'_y \in S^c} \sum_{\pi \in \mathcal{S}_g / \mathcal{S}_x \times \mathcal{S}_y} T_{\pi[\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_y)]} \pi \left[Y_{s_1} \cdots Y_{s_x} Y_{s'_1} \cdots Y_{s'_y} \right]. \quad (323)$$

which prepares us to act the smoothing operator on it. For the notation here, we refer the reader to Section 2.8. The rest of this Appendix constitutes the proof of the following theorem.

Theorem 2.4. *For $p \geq 2$ and $C_p = p - 1$,*

$$\|F\|_{*,p}^2 \leq g^g C_p^g \sum_{\mathbf{w}} \left(\sum_{\mathbf{v}} \sum_{\pi \in \mathcal{S}_g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i + 2v_i} \right)^2 \|I\|_{*,p}^2, \quad (90)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^m$ are such that $|\mathbf{w}| + 2|\mathbf{v}| = g$.¹⁶ The unary encoding $\boldsymbol{\eta}$ is defined as

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := \underbrace{(1, \dots, 1)}_{a_1 \text{ times}} \underbrace{(2, \dots, 2)}_{a_2 \text{ times}} \dots \underbrace{(m, \dots, m)}_{a_m \text{ times}} \underbrace{(1, \dots, 1)}_{b_1 \text{ times}} \underbrace{(2, \dots, 2)}_{b_2 \text{ times}} \dots \underbrace{(m, \dots, m)}_{b_m \text{ times}} \quad (91)$$

for $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$.

Proof. Applying \mathbb{E}_{S^c} to F gives

$$\mathbb{E}_{S^c}[F] = \sum_{x+y=g} \sum_{s_1, \dots, s_x \in S} \sum_{s'_1, \dots, s'_y \in S^c} \sum_{\pi \in \mathcal{S}_g / \mathcal{S}_x \times \mathcal{S}_y} T_{\pi[\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_y)]} \pi \left[Y_{s_1} \cdots Y_{s_x} \mathbb{E}(Y_{s'_1} \cdots Y_{s'_y}) \right]. \quad (324)$$

We now analyze the sum over s'_1, \dots, s'_y more carefully. Observe that

$$\mathbb{E}(Y_{s'_1} \cdots Y_{s'_y}) \neq 0 \quad (325)$$

if and only if y is even, since the product is nonzero in expectation only when it contains no odd moments of $Y_{s'_i}$. Thus, we scale y to $2y$ and update the summation constraints accordingly:

$$\mathbb{E}_{S^c}[F] = \sum_{x+2y=g} \sum_{s_1, \dots, s_x \in S} \sum_{s'_1, \dots, s'_{2y} \in S^c} \sum_{\pi \in \mathcal{S}_g / \mathcal{S}_x \times \mathcal{S}_{2y}} T_{\pi[\mathbf{pr}_1(s_1, \dots, s_x, s'_1, \dots, s'_{2y})]} \pi \left[Y_{s_1} \cdots Y_{s_x} \mathbb{E}(Y_{s'_1} \cdots Y_{s'_{2y}}) \right]. \quad (326)$$

Since each Rademacher variable contributes only through exponents 0, 1, or 2, any non-vanishing $\mathbb{E}(Y_{s'_1} \cdots Y_{s'_{2y}})$ must consist of a product of second moments of Rademachers. Therefore, the subscripts $\{s'_1, \dots, s'_{2y}\}$ contain exactly y distinct elements $\{s''_1, \dots, s''_y\}$, and the expectation takes the form

$$\mathbb{E}(Y_{s'_1} \cdots Y_{s'_{2y}}) = \pi' \left[\mathbb{E}(Y_{s''_1}^2) \cdots \mathbb{E}(Y_{s''_y}^2) \right], \quad (327)$$

where $\pi' \in \mathcal{S}_{2y}$ reorders the product. This reordering is necessary due to the non-commutativity of the Rademacher variables. Hence, the corresponding sum over expectations in equation (324) can be expressed as

$$\sum_{s'_1, \dots, s'_{2y} \in S^c} \mathbb{E}(Y_{s'_1} \cdots Y_{s'_{2y}}) = \sum_{\substack{s''_1, \dots, s''_y \in S^c \\ \text{distinct}}} \sum_{\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y} \pi' \left[\mathbb{E}(Y_{s''_1}^2) \cdots \mathbb{E}(Y_{s''_y}^2) \right]. \quad (328)$$

We do not sum over the entire group \mathcal{S}_{2y} for π' to avoid overcounting. Specifically, two classes of permutations must be excluded: (i) internal flips that leave the argument invariant, of the form $(12)^{\zeta_1} (34)^{\zeta_2} \cdots$ with $\zeta_i \in \{0, 1\}$, corresponding to the group

$$\mathcal{S}_2^y := \underbrace{\mathcal{S}_2 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_2}_{y \text{ times}}, \quad (329)$$

¹⁶If $\mathbf{u} = (u_1, \dots, u_m) \in \mathbb{Z}_{\geq 0}^m$ is a vector then $|\mathbf{u}| := \sum_{i=1}^m |u_i|$.

and (ii) pair permutations, i.e. permutations that exchange entire pairs, corresponding to \mathcal{S}_y . The composition of these two symmetry types (first the internal flips, then the pair permutations) forms the group

$$\mathcal{S}_2^y \rtimes \mathcal{S}_y. \quad (330)$$

The semi-direct product ensures the correct group structure.¹⁷ This explains the notation $\pi' \in \mathcal{S}_{2y}/\mathcal{S}_2^y \rtimes \mathcal{S}_y$ in the summation. Plugging this back into the sum in equation (324) gives

$$\begin{aligned} \mathbb{E}_{S^c}[F] = & \sum_{x+2y=g} \sum_{s_1, \dots, s_x \in S} \sum_{\pi \in \mathcal{S}_g/\mathcal{S}_x \times \mathcal{S}_{2y}} \sum_{\substack{s'_1, \dots, s'_y \in S^c \\ \text{distinct}}} \sum_{\pi' \in \mathcal{S}_{2y}/\mathcal{S}_2^y \rtimes \mathcal{S}_y} \\ & T_{\pi[\mathbf{pr}_1(s_1, \dots, s_x, \pi'[s'_1, s'_1, \dots, s'_y, s'_y])]} \pi \left[Y_{s_1} \cdots Y_{s_x} \pi' \left[\mathbb{E}(Y_{s'_1}^2) \cdots \mathbb{E}(Y_{s'_y}^2) \right] \right]. \end{aligned} \quad (334)$$

Applying the central part $(1 - \mathbb{E})_S$ to $\mathbb{E}_{S^c}[F]$ and using the triangle inequality yields

$$\begin{aligned} \|\| [F]_S \|\|_{*,p} \leq & \sum_{\pi \in \mathcal{S}_g/\mathcal{S}_{2y}} \sum_{\substack{s'_1, \dots, s'_y \in S^c \\ \text{distinct}}} \sum_{\pi' \in \mathcal{S}_{2y}/\mathcal{S}_2^y \rtimes \mathcal{S}_y} \\ & \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, \pi'[s'_1, s'_1, \dots, s'_y, s'_y])]} \right| \left\| \left\| \pi \left[Y_{s_1} \cdots Y_{s_{|S|}} \pi' \left[\mathbb{E}(Y_{s'_1}^2) \cdots \mathbb{E}(Y_{s'_y}^2) \right] \right] \right\| \right\|_{*,p}. \end{aligned} \quad (335)$$

Here y is fixed by $y = (g - |S|)/2$. This follows from the fact that

$$(1 - \mathbb{E})_S [Y_{s_1} \cdots Y_{s_x}] \neq 0 \quad (336)$$

if and only if all subscripts in S appear in the product $Y_{s_1} \cdots Y_{s_x}$ and each term has an odd exponent. Since each contributing Rademacher variable can only have exponent 0, 1, or 2, all must have exponent 1, fixing $x = |S|$ and thus $y = (g - |S|)/2$. We also extend the summation over π to the entire group \mathcal{S}_g ¹⁸, quotienting out permutations that reorder $Y_{s'_1}, \dots, Y_{s'_y}$ among themselves. This extension means that the part of \mathcal{S}_g we sum over now accounts for permutations that reorder $Y_{s_1}, \dots, Y_{s_{|S|}}$ among themselves, allowing us to drop the summation over $s_1, \dots, s_{|S|}$.

By construction, each Rademacher variable can be written as

$$Y_s = \frac{\epsilon_s}{\sqrt{N}} K_{\mathbf{pr}_1 s} \quad (337)$$

and recall that each $K_{\mathbf{pr}_1 s}$ is bounded by $\|\| K_{\mathbf{pr}[m]s} \|\|_{*,p} \leq \sigma_{\mathbf{pr}[m]s} \cdot \|\| I \|\|_{*,p}$. Plugging this back in

¹⁷Normally, such semi-direct product does not give rise to a group structure, but in this case the symmetries of \mathcal{S}_2^y and \mathcal{S}_y have trivial intersection. Therefore, the semi-direct product does yield a group structure. To visualize this, consider a vector of $2y$ entries divided into y columns where each column contains a pair:

$$(s''_1, s''_2 | s''_3, s''_4 | \cdots | s''_{2y-1}, s''_{2y}). \quad (331)$$

The symmetries of \mathcal{S}_2^y act on this vector by flipping the pairs inside each column. For instance, flipping the pair in the second column gives

$$(s''_1, s''_2 | s''_4, s''_3 | \cdots | s''_{2y-1}, s''_{2y}). \quad (332)$$

The symmetries of \mathcal{S}_y permute the columns among each other while leaving the pairs inside the columns invariant. For instance, permuting the first column and the last column gives us

$$(s''_{2y-1}, s''_{2y} | s''_3, s''_4 | \cdots | s''_1, s''_2). \quad (333)$$

Then it is clear that the symmetries of \mathcal{S}_2^y and \mathcal{S}_y have trivial intersection.

¹⁸We are allowed to do this here, because all the summands after triangle inequality in equation (335) are positive.

equation (335) gives us

$$\begin{aligned} \lll[F]_S \lll_{*,p} \leq & \sum_{\pi \in \mathcal{S}_g / \mathcal{S}_{2y}} \sum_{\substack{s''_1, \dots, s''_y \in S^c \\ \text{distinct}}} \sum_{\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, \pi'[s''_1, s''_1, \dots, s''_y, s''_y])]} \right| \\ & \frac{1}{\sqrt{N}^{|S|}} \frac{1}{N^{\frac{g-|S|}{2}}} \left[\sigma_{\mathbf{pr}_1 s_1} \cdots \sigma_{\mathbf{pr}_1 s_{|S|}} \sigma_{\mathbf{pr}_1 s''_1}^2 \cdots \sigma_{\mathbf{pr}_1 s''_y}^2 \right] \lll I \lll_{*,p} \end{aligned} \quad (338)$$

We can now simplify the right-hand side by making the summand independent of the summation over π' . To do so, we first extend the sum over π to the full group \mathcal{S}_g . Since we are summing positive terms, the upper bound remains valid. Notice that the sum over the entire \mathcal{S}_g commutes with the other two summations. Hence, we can write equation (335) as

$$\begin{aligned} \lll[F]_S \lll_{*,p} \leq & \sum_{\substack{s''_1, \dots, s''_y \in S^c \\ \text{distinct}}} \sum_{\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y} \sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, \pi'[s''_1, s''_1, \dots, s''_y, s''_y])]} \right| \\ & \frac{1}{\sqrt{N}^{|S|}} \frac{1}{N^{\frac{g-|S|}{2}}} \left[\sigma_{\mathbf{pr}_1 s_1} \cdots \sigma_{\mathbf{pr}_1 s_{|S|}} \sigma_{\mathbf{pr}_1 s''_1}^2 \cdots \sigma_{\mathbf{pr}_1 s''_y}^2 \right] \lll I \lll_{*,p}. \end{aligned} \quad (339)$$

Observe that the sum over $\pi \in \mathcal{S}_g$ has the following property: for any permutation $\tau \in \mathcal{S}_g$,

$$\sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(\tau[s_1, \dots, s_{|S|}, s''_1, s''_1, \dots, s''_y, s''_y])]} \right| = \sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, s''_1, s''_1, \dots, s''_y, s''_y)]} \right|, \quad (340)$$

because the composition of τ and π yields another permutation in \mathcal{S}_g . We can view each $\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y$ as a permutation in \mathcal{S}_g that leaves the first $|S|$ elements invariant and acts only on the last $2y$ elements. Hence, for every $\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y$, we have

$$\sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, \pi'[s''_1, s''_1, \dots, s''_y, s''_y])]} \right| = \sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, s''_1, s''_1, \dots, s''_y, s''_y)]} \right|. \quad (341)$$

Plugging this back into the sum in equation (339), we get

$$\begin{aligned} \lll[F]_S \lll_{*,p} \leq & \sum_{\substack{s''_1, \dots, s''_y \in S^c \\ \text{distinct}}} \sum_{\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y} \sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, s''_1, s''_1, \dots, s''_y, s''_y)]} \right| \\ & \frac{1}{\sqrt{N}^{|S|}} \frac{1}{N^{\frac{g-|S|}{2}}} \left[\sigma_{\mathbf{pr}_1 s_1} \cdots \sigma_{\mathbf{pr}_1 s_{|S|}} \sigma_{\mathbf{pr}_1 s''_1}^2 \cdots \sigma_{\mathbf{pr}_1 s''_y}^2 \right] \lll I \lll_{*,p} \end{aligned} \quad (342)$$

and find that the summand is completely independent of the summation over π' . Noting that

$$\sum_{\pi' \in \mathcal{S}_{2y} / \mathcal{S}_2^y \rtimes \mathcal{S}_y} 1 = \frac{(2y)!}{2^y \cdot y!} = (2y-1)!! \leq g^{g/2}. \quad (343)$$

Plugging this back into equation (342) gives us

$$\begin{aligned} \lll[F]_S \lll_{*,p} \leq & g^{g/2} \sum_{\substack{s''_1, \dots, s''_y \in S^c \\ \text{distinct}}} \sum_{\pi \in \mathcal{S}_g} \left| T_{\pi[\mathbf{pr}_1(s_1, \dots, s_{|S|}, s''_1, s''_1, \dots, s''_y, s''_y)]} \right| \\ & \frac{1}{\sqrt{N}^{|S|}} \frac{1}{N^{\frac{g-|S|}{2}}} \left[\sigma_{\mathbf{pr}_1 s_1} \cdots \sigma_{\mathbf{pr}_1 s_{|S|}} \sigma_{\mathbf{pr}_1 s''_1}^2 \cdots \sigma_{\mathbf{pr}_1 s''_y}^2 \right] \lll I \lll_{*,p}. \end{aligned} \quad (344)$$

To sum over all $\lll[F]_S \lll_{*,p}$ in Theorem 2.3, we parametrize every S by

$$S = \bigcup_{i=1}^m \{(i, j_1^{(i)}), \dots, (i, j_{w_i}^{(i)})\} \quad (345)$$

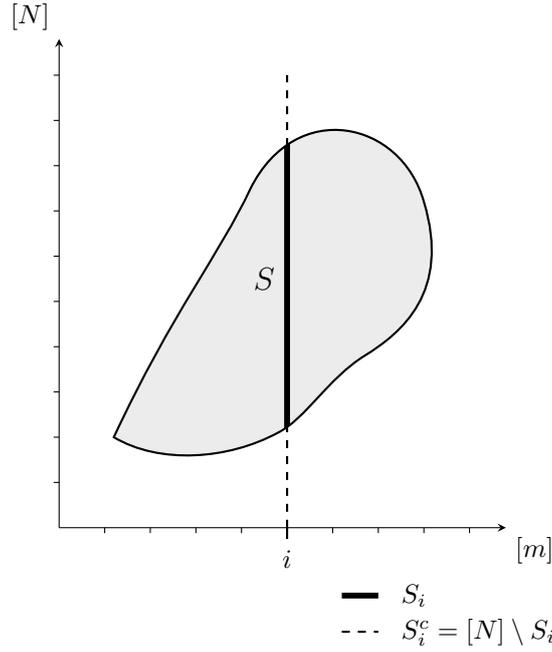


Figure 7: Depiction of the set S on the two-dimensional plane $[m] \times [N]$, the slice S_i , and slice complement $S_i^c = [N] \setminus S_i$. The same concepts also work for disconnected S .

with a vector $\mathbf{w} \in \{0, \dots, g\}^m$ with $|\mathbf{w}| = |S|$, and integers $j_l^{(i)} \in [N]$. When $w_i = 0$, the corresponding set $\{(i, j_1^{(i)}), \dots, (i, j_{w_i}^{(i)})\}$ is empty. One can view S as a union of vertical slices in the two-dimensional grid $[m] \times [N]$:

$$S = \bigcup_{i=1}^m S_i, \quad (346)$$

where each slice is determined by the entries of the vector $\mathbf{j}^{(i)} := (j_1^{(i)}, \dots, j_{w_i}^{(i)})$ via

$$S_i := \{(i, j_1^{(i)}), \dots, (i, j_{w_i}^{(i)})\}. \quad (347)$$

The size of each slice S_i is given by w_i , and S_i is empty when $w_i = 0$. We also define the slice complement

$$S_i^c := [N] \setminus S_i, \quad (348)$$

which should not be confused with S^c , the complement of S . These constructions are illustrated in Figure 7.

These concepts allow us to parametrize the summation over S by

$$\sum_S \mapsto \sum_{\mathbf{w}} \sum_{\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}} := \sum_{\substack{j_1^{(1)}, \dots, j_{w_1}^{(1)}=1 \\ \text{distinct}}}^N \cdots \sum_{\substack{j_1^{(m)}, \dots, j_{w_m}^{(m)}=1 \\ \text{distinct}}}^N. \quad (349)$$

When $w_i = 0$ for a certain $1 \leq i \leq m$, the sum $\sum_{\mathbf{j}^{(i)}} = \sum_{j_1^{(i)}, \dots, j_{w_i}^{(i)}=1}^N$ does not appear. Similarly, we can rewrite the summation $\sum_{\substack{s'_1, \dots, s'_y \in S^c \\ \text{distinct}}}$ as

$$\sum_{\substack{s'_1, \dots, s'_y \in S^c \\ \text{distinct}}} \mapsto \sum_{\mathbf{v}} \sum_{\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m)}} := \sum_{\substack{k_1^{(1)}, \dots, k_{v_1}^{(1)} \in S_1^c \\ \text{distinct}}} \cdots \sum_{\substack{k_1^{(m)}, \dots, k_{v_m}^{(m)} \in S_m^c \\ \text{distinct}}} \quad (350)$$

such that \mathbf{v} is a vector in $\{0, \dots, g\}^m$ such that $|\mathbf{w}| + 2|\mathbf{v}| = g$. The slice complements S_i^c here are understood as $[N] \setminus \{(i, j_1^{(i)}), \dots, (i, j_{w_i}^{(i)})\}$, which is fixed by the summation over \mathbf{w} and $\mathbf{j}^{(i)}$. When

$v_i = 0$, the sum $\sum_{\mathbf{k}^{(i)}} = \sum_{k_1^{(i)}, \dots, k_{v_i}^{(i)} \in S_i^c}$ does not appear. In this way, we can still pick y distinct elements from S^c as

$$\bigcup_{i=1}^m \{(i, k_1^{(i)}), \dots, (i, k_{v_i}^{(i)})\}. \quad (351)$$

Using these reparametrizations, we have

$$\begin{aligned} \sum_S C_p^{|S|} \|[F]_S\|_{*,p}^2 &\leq g^g C_p^g \sum_{\mathbf{w}} \sum_{\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}} \frac{1}{N^{|S|}} \frac{1}{N^{g-|S|}} \|[I]\|_{*,p}^2 \left(\sum_{\mathbf{v}} \sum_{\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m)}} \sum_{\pi \in S_g} \right. \\ &\quad \left. \left| T_{\pi[\text{pr}_1(1, j_1^{(1)}), \dots, \text{pr}_1(m, j_{w_m}^{(m)}), \text{pr}_1(1, k_1^{(1)}), \text{pr}_1(1, k_1^{(1)}), \dots, \text{pr}_1(m, k_{v_m}^{(m)}), \text{pr}_1(m, k_{v_m}^{(m)})]} \right| \right. \\ &\quad \left. \left[\sigma_{\text{pr}_1(1, j_1^{(1)})} \cdots \sigma_{\text{pr}_1(m, j_{w_m}^{(m)})} \sigma_{\text{pr}_1(1, k_1^{(1)})}^2 \cdots \sigma_{\text{pr}_1(m, k_{v_m}^{(m)})}^2 \right] \right)^2 \end{aligned} \quad (352)$$

The product over σ can be simply written as

$$\sigma_{\text{pr}_1(1, j_1^{(1)})} \cdots \sigma_{\text{pr}_1(m, j_{w_m}^{(m)})} \sigma_{\text{pr}_1(1, k_1^{(1)})}^2 \cdots \sigma_{\text{pr}_1(m, k_{v_m}^{(m)})}^2 = \prod_{i=1}^m \sigma_i^{w_i + 2v_i}. \quad (353)$$

To write the coefficients $T_{\pi[\dots]}$ more compactly, we introduce the following unary encoding $\boldsymbol{\eta}$

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := (\underbrace{1, \dots, 1}_{a_1 \text{ times}}, \underbrace{2, \dots, 2}_{a_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{a_m \text{ times}}, \underbrace{1, \dots, 1}_{b_1 \text{ times}}, \underbrace{2, \dots, 2}_{b_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{b_m \text{ times}}), \quad (354)$$

for vectors $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$. Using this, we can write the coefficients $T_{\pi[\dots]}$ as

$$T_{\pi[\text{pr}_1(1, j_1^{(1)}), \dots, \text{pr}_1(m, j_{w_m}^{(m)}), \text{pr}_1(1, k_1^{(1)}), \text{pr}_1(1, k_1^{(1)}), \dots, \text{pr}_1(m, k_{v_m}^{(m)}), \text{pr}_1(m, k_{v_m}^{(m)})]} = T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}. \quad (355)$$

Plugging this back into the right-hand side of equation (352) yields

$$\begin{aligned} \sum_S C_p^{|S|} \|[F]_S\|_{*,p}^2 &\leq g^g C_p^g \sum_{\mathbf{w}} \sum_{\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}} \frac{1}{N^{|S|}} \frac{1}{N^{g-|S|}} \|[I]\|_{*,p}^2 \\ &\quad \left(\sum_{\mathbf{v}} \sum_{\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m)}} \sum_{\pi \in S_g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i + 2v_i} \right)^2. \end{aligned} \quad (356)$$

Notice that the product over σ is independent of the summation over $\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m)}$. Since the size of each slice complement S_i^c is always less than N , we have

$$\sum_{\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m)}} 1 \leq \sum_{k_1^{(1)}, \dots, k_{v_1}^{(1)}=1}^N \cdots \sum_{k_1^{(m)}, \dots, k_{v_m}^{(m)}=1}^N 1 = N^{|\mathbf{v}|}. \quad (357)$$

By construction, $|\mathbf{v}| = \frac{g-|\mathbf{w}|}{2}$. Hence,

$$\begin{aligned} \sum_S C_p^{|S|} \|[F]_S\|_{*,p}^2 &\leq g^g C_p^g \sum_{\mathbf{w}} \sum_{\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}} \frac{1}{N^{|\mathbf{w}|}} \frac{1}{N^{g-|\mathbf{w}|}} \|[I]\|_{*,p}^2 \\ &\quad \left(\sum_{\mathbf{v}} N^{\frac{g-|\mathbf{w}|}{2}} \sum_{\pi \in S_g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i + 2v_i} \right)^2. \end{aligned} \quad (358)$$

By moving the $N^{\frac{g-|\mathbf{w}|}{2}}$ factor out of the parentheses in equation (358), we see that it exactly cancels the factor $\frac{1}{N^{g-|\mathbf{w}|}}$, leaving us with

$$\begin{aligned} \sum_S C_p^{|S|} \|[F]_S\|_{*,p}^2 &\leq g^g C_p^g \sum_{\mathbf{w}} \sum_{\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}} \frac{1}{N^{|\mathbf{w}|}} \|[I]\|_{*,p}^2 \\ &\quad \left(\sum_{\mathbf{v}} \sum_{\pi \in S_g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i + 2v_i} \right)^2. \end{aligned} \quad (359)$$

We can do a similar treatment to the summation over $\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}$, and notice that

$$\sum_{\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(m)}} 1 = N^{|\mathbf{w}|}, \quad (360)$$

which cancels out the $\frac{1}{N^{|\mathbf{w}|}}$ factor in equation (359). In the end, we obtain

$$\sum_S C_p^{|S|} \|\| [F]_S \|\|_{*,p}^2 \leq g^g C_p^g \sum_{\mathbf{w}} \left(\sum_{\mathbf{v}} \sum_{\pi \in \mathcal{S}_g} |T_{\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]}| \prod_{i=1}^m \sigma_i^{w_i + 2v_i} \right)^2 \|\| I \|\|_{*,p}^2. \quad (361)$$

Together with Theorem 2.3, we have proved the desired statement. \square

D Evaluating G_w for general Hamiltonians

Lemma 4.2. *Let $H = \sum_{i=1}^m H_{\gamma(i)}$ be a Hamiltonian of m terms with arbitrary ordering γ , such that for all $1 \leq i < j \leq m$, $H_{\gamma(i)}$ and $H_{\gamma(j)}$ either commute or anti-commute. Fix $g \geq 1$ and $w \in \{0, \dots, g\}$, and consider the H -dependent sum*

$$G_w(H) := \sum_{\mathbf{w}: |\mathbf{w}|=w} \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}: |\mathbf{w}|+2|\mathbf{v}|=g} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2, \quad (162)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^m$ such that $|\mathbf{w}| = w$ and $|\mathbf{w}| + 2|\mathbf{v}| = g$. The permutation $\pi \in \mathcal{S}_g$ reorders the entries of vectors in $[m]^g$. The unary encoding $\boldsymbol{\eta}$ and the indicator $\mathbb{1}$ are defined on vectors $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$ and $\mathbf{j} \in [m]^g$ as follows:

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := \underbrace{(1, \dots, 1)}_{a_1 \text{ times}}, \underbrace{(2, \dots, 2)}_{a_2 \text{ times}}, \dots, \underbrace{(m, \dots, m)}_{a_m \text{ times}}, \underbrace{(1, \dots, 1)}_{b_1 \text{ times}}, \underbrace{(2, \dots, 2)}_{b_2 \text{ times}}, \dots, \underbrace{(m, \dots, m)}_{b_m \text{ times}}, \quad (163)$$

and

$$\mathbb{1}(\mathbf{j}) := \begin{cases} 1 & \text{if } \mathcal{L}_{\gamma(j_g)} \cdots \mathcal{L}_{\gamma(j_2)} [H_{\gamma(j_1)}] \neq 0, \\ 0 & \text{else.} \end{cases} \quad (164)$$

Then

$$G_w(H) \leq g^{3g-2} m^2 Q_{\max}(H)^{g-2} \quad (165)$$

where

$$Q_{\max}(H) := \max_{1 \leq i \leq m} |\{j \in [m] : [H_{\gamma(j)}, H_{\gamma(i)}] \neq 0\}|. \quad (166)$$

Proof. Using Hölder's inequality with $p = 1$ and $q = \infty$,

$$G_w(H) \leq \left(\sum_{\mathbf{w}: |\mathbf{w}|=w} \sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}: |\mathbf{w}|+2|\mathbf{v}|=g} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right) \cdot \left(\max_{\mathbf{w}: |\mathbf{w}|=w} \sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}: |\mathbf{w}|+2|\mathbf{v}|=g} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right). \quad (362)$$

By using the definition of $\boldsymbol{\eta}$ from equation (91), we have

$$G_w(H) \leq \left(\sum_{\pi \in \mathcal{S}_g} \sum_{i_w, \dots, i_1=1}^{\Gamma} \sum_{j_v, \dots, j_1=1}^{\Gamma} \mathbb{1}(\pi(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)) \right) \cdot \left(\max_{i_w, \dots, i_1} \sum_{\pi \in \mathcal{S}_g} \sum_{j_v, \dots, j_1} \mathbb{1}(\pi(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)) \right). \quad (363)$$

where $v = (g - w)/2$. Instead of summing over the vectors \mathbf{w} and \mathbf{v} , we are now directly summing over the local terms in the Hamiltonian, where indices i_1, \dots, i_w come from \mathbf{w} and indices j_1, \dots, j_v come from \mathbf{v} . Because of this, each j_i occurs in pairs in the indicator $\mathbb{1}$. The permutation π is now a reordering of $(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)$.

We are going to evaluate the two terms in equation (363) separately. The first term is

$$L_1 = \sum_{\pi \in \mathcal{S}_g} \sum_{i_w, \dots, i_1=1}^{\Gamma} \sum_{j_v, \dots, j_1=1}^{\Gamma} \mathbb{1}(\pi(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)), \quad (364)$$

where $g = w + 2v \geq 2$.

Fix a permutation $\pi \in \mathcal{S}_g$. The sum

$$S_1 = \sum_{i_w, \dots, i_1=1}^{\Gamma} \sum_{j_v, \dots, j_1=1}^{\Gamma} \mathbb{1}(\pi(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)) \quad (365)$$

is effectively counting the number of ways we can make a commutator chain of length g by selecting $w + v$ local terms of H while duplicating v selected terms. We shall first provide a concrete example of evaluating this sum before proceeding to the general argument.

Take $g = 6$, $w = 2$ and $v = 2$. Let $\pi(i_2, i_1, j_2, j_2, j_1, j_1) = (j_1, i_2, j_2, j_2, j_1, i_1)$. The sum S_1 corresponds to counting the number of commutator chains of the form

$$[H_{\gamma(j_1)}, [H_{\gamma(i_2)}, [H_{\gamma(j_2)}, [H_{\gamma(j_2)}, [H_{\gamma(j_1)}, H_{\gamma(i_1)}]]]]] \quad (366)$$

such that it is non-zero. We refer to the position of local terms $H_{\gamma(i)}$ in the commutator chain as “layer”. In this case, from the innermost to the outermost commutator, $H_{\gamma(i_1)}$ is on the first layer, $H_{\gamma(j_1)}$ the second layer, etc.. We are now going to count how many such non-zero commutators there are. Starting from the first layer, there should always be up to m choices. The local term on the second layer must anti-commute with the first one. Therefore, there should be up to $Q_{\max}(H)$ choices. The local term on the third layer must anti-commute with at least one of the local term of the previous layers, i.e. at least one of $H_{\gamma(j_1)}$ or $H_{\gamma(i_1)}$. This gives us up to $2Q_{\max}(H)$ choices. The fourth layer must be a duplicate of the third layer, leaving us with up to 1 choice. The fifth layer must also anti-commute with at least one of the local terms on the previous layers, giving us up to $3Q_{\max}(H)$ choices. Lastly, the sixth layer is a duplicate of the second layer, leaving us again with up to 1 choice. Thus, the number of such non-zero commutators are less or equal to $m \cdot Q_{\max}(H) \cdot 2Q_{\max}(H) \cdot 1 \cdot 3Q_{\max}(H) \cdot 1 = 3! \cdot mQ_{\max}(H)^3$. Observe that this bound is uniform for all permutations π . The only thing that changes is where the duplicate layers are in the commutator chain, but one could simply repeat the reasoning above to end up with the same upper bound. For instance, consider $\pi'(i_2, i_1, j_2, j_2, j_1, j_1) = (j_1, j_2, i_2, j_2, j_1, i_1)$ by swapping i_2 and j_2 , the fourth layer now has up to $3Q_{\max}(H)$ choices, while the fifth and the sixth layer only have up to 1 choice being duplicate layers. Hence, under this permutation, the bound stays the same. Even in the extreme case such as $\pi''(i_2, i_1, j_2, j_2, j_1, j_1) = (j_1, i_1, j_1, i_2, j_2, j_2)$, the innermost commutator automatically vanishes and the number of non-zero commutator chains of this form is 0, which is still bounded by $3! \cdot mQ_{\max}(H)$.

Now, consider general commutator chain of length g

$$[H_{\gamma(a_{g-1})}, [\dots, [H_{\gamma(a_2)}, [H_{\gamma(a_1)}, H_{\gamma(a_0)}]]] \dots] \quad (367)$$

with v duplicate layers distributed somewhere in the chain. When selecting local terms from the first layer to the last layer, we expect three constraints on each layer

1. it is on the first layer, or
2. it must anti-commute with at least one of the local terms on previous layers, or
3. it is a duplicate of one of the previous layers.

The first constraint leaves us with m choices, and the last constraint only up to 1 choice. Since there are $w+v$ non-duplicate layers, the second constraint leaves us with up to $(w+v-1)!Q_{\max}(H)^{w+v-1}$ choices. Thus,

$$S_1 \leq g^{w+v-1} m Q_{\max}(H)^{w+v-1}. \quad (368)$$

Because the above upper bound does not depend on the particular choice of $\pi \in \mathcal{S}_g$, we have

$$L_1 \leq g! g^{w+v-1} m Q_{\max}(H)^{w+v-1}. \quad (369)$$

Next, we move on to the second term:

$$L_2 = \max_{i_w, \dots, i_1} \sum_{\pi \in \mathcal{S}_g} \sum_{j_v, \dots, j_1} \mathbb{1}(\pi(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)) \quad (370)$$

Fix any configuration (i_w, \dots, i_1) and permutation $\pi \in \mathcal{S}_g$, the sum

$$S_2 = \sum_{j_v, \dots, j_1} \mathbb{1}(\pi(i_w, \dots, i_1, j_v, j_v, \dots, j_1, j_1)) \quad (371)$$

is effectively counting the number of ways of making a commutator chain of length g out of terms in H such that $H_{\gamma(i_w)}, \dots, H_{\gamma(i_1)}$ must be present in the commutator chain and the rest of the terms must be duplicates. Let

$$[H_{\gamma(b_{g-1})}, [\dots, [H_{\gamma(b_2)}, [H_{\gamma(b_1)}, H_{\gamma(b_0)}]]] \dots] \neq 0 \quad (372)$$

be a commutator chain that satisfies the constraints in S_2 . Now, when we select local terms from the first to last layer, we expect four constraints:

1. it is on the first layer, or
2. it must anti-commute with at least one of the local terms on previous layers, or
3. it is a duplicate of one of the previous layers, or
4. it is fixed for being one of $H_{\gamma(i_w)}, \dots, H_{\gamma(i_1)}$.

The last two cases only leave us with up to 1 choice, but the first and second cases leaves us with a total of up to $(v-1)! \cdot m Q_{\max}(H)^{v-1}$ choices. Therefore,

$$S_2 \leq g^{v-1} m Q_{\max}(H)^{v-1} \quad (373)$$

and because this upper bound is independent of the choice of (i_w, \dots, i_1) and $\pi \in \mathcal{S}_g$,

$$L_2 \leq g! g^{v-1} m Q_{\max}(H)^{v-1}. \quad (374)$$

Putting L_1 and L_2 together and using $g! < g^g$ gives us

$$\begin{aligned} G_w(H) &\leq L_1 \cdot L_2 \\ &\leq g^{3g-2} m^2 Q_{\max}(H)^{g-2}. \end{aligned} \quad (375)$$

This concludes the proof. \square

E Average G_w for sparse Hamiltonians

Lemma 5.2. *Let $H = \sum_{i=1}^m B_i H_{\gamma(i)}$ be a sparse Hamiltonian of m terms with arbitrary ordering γ , such that for all $1 \leq i < j \leq m$, $H_{\gamma(i)}$ and $H_{\gamma(j)}$ either commute or anti-commute. The B_i 's are independent Bernoulli variables with $\mathbb{P}(B_i = 1) = p_B$ for all i . Fix $g \geq 1$ and $w \in \{0, \dots, g\}$, and consider the average H -dependent sum over all samples of the Bernoulli variables*

$$\langle G_w(H(\mathbf{B})) \rangle = \left\langle \sum_{\mathbf{w}: |\mathbf{w}|=w} \prod_{i \in \text{supp}(\mathbf{w})} B_i \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}: |\mathbf{w}|+2|\mathbf{v}|=g} \prod_{j \in \text{supp}(\mathbf{v})} B_j \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right) \right\rangle, \quad (232)$$

where $\mathbf{w}, \mathbf{v} \in \{0, \dots, g\}^m$ such that $|\mathbf{w}| = w$ and $|\mathbf{w}| + 2|\mathbf{v}| = g$. The permutation $\pi \in \mathcal{S}_g$ reorders the entries of vectors in $[m]^g$. The unary encoding $\boldsymbol{\eta}$ and the indicator $\mathbb{1}$ are defined on vectors $\mathbf{a}, \mathbf{b} \in \{0, \dots, g\}^m$ and $\mathbf{j} \in [m]^g$ as follows:

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}) := (\underbrace{1, \dots, 1}_{a_1 \text{ times}}, \underbrace{2, \dots, 2}_{a_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{a_m \text{ times}}, \underbrace{1, \dots, 1}_{b_1 \text{ times}}, \underbrace{2, \dots, 2}_{b_2 \text{ times}}, \dots, \underbrace{m, \dots, m}_{b_m \text{ times}}), \quad (233)$$

and

$$\mathbb{1}(\mathbf{j}) := \begin{cases} 1 & \text{if } \mathcal{L}_{\gamma(j_g)} \cdots \mathcal{L}_{\gamma(j_2)}[H_{\gamma(j_1)}] \neq 0, \\ 0 & \text{else.} \end{cases} \quad (234)$$

Then

$$\langle G_w(H(\mathbf{B})) \rangle \leq g^{4g} m^2 Q_{\max}(H)^{-2} \sum_{s=0}^w \sum_{q, q'=0}^{(g-w)/2} \sum_{c=0}^{\min(q, q')} (p_B Q_{\max}(H))^{s+q+q'-c}. \quad (235)$$

where

$$Q_{\max}(H) := \max_{1 \leq i \leq m} |\{j \in [m] : [H_{\gamma(j)}, H_{\gamma(i)}] \neq 0\}|. \quad (236)$$

Proof. Consider a Hamiltonian of the form

$$H(\mathbf{B}) = \sum_{i=1}^m B_i H_{\gamma(i)} \quad (376)$$

where the B_i 's are independent Bernoulli variables such that $\mathbb{P}(B_i = 1) = p_B$ for all $1 \leq i \leq m$. We use the bolded text $\mathbf{B} = (B_1, \dots, B_m)$ to denote the dependency of these Bernoulli variables. Let $\mathbf{b} \in \{0, 1\}^m$ be a sample of the Bernoulli variables. When the Bernoulli variables take the

values of this sample, we get a configuration of the sparse Hamiltonian where we can treat \mathbf{b} as coefficients

$$H(\mathbf{B} = \mathbf{b}) = \sum_{i=1}^m b_i H_{\gamma(i)}. \quad (377)$$

Similar to the G_w for general Hamiltonians, we now consider the configuration specific sum:

$$G_w(H(\mathbf{B} = \mathbf{b})) = \sum_{\mathbf{w}:|\mathbf{w}|=w} \prod_{i \in \text{supp}(\mathbf{w})} b_i \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}|+2|\mathbf{v}|=g} \prod_{j \in \text{supp}(\mathbf{v}) \setminus \text{supp}(\mathbf{w})} b_j \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2 \quad (378)$$

which looks slightly more complicated than the original G_w due to the extra coefficients b_i . The other summations over vectors $\mathbf{w}, \mathbf{v} \in \mathbb{Z}_{\geq 0}^m$ work the same as in G_w , where $|\mathbf{w}| + 2|\mathbf{v}| = g$. The indicator $\mathbb{1}$ and unary encoding $\boldsymbol{\eta}$ work precisely the same way as before. We are now interested in computing the average of this object over all samples of the Bernoulli variables, denoted by

$$\langle G_w(H(\mathbf{B})) \rangle = \left\langle \sum_{\mathbf{w}:|\mathbf{w}|=w} \prod_{i \in \text{supp}(\mathbf{w})} B_i \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}|+2|\mathbf{v}|=g} \prod_{j \in \text{supp}(\mathbf{v}) \setminus \text{supp}(\mathbf{w})} B_j \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2 \right\rangle. \quad (379)$$

By construction, the product over B_i is completely independent from B_j , we have¹⁹

$$\langle G_w(H(\mathbf{B})) \rangle = \sum_{\mathbf{w}:|\mathbf{w}|=w} \prod_{i \in \text{supp}(\mathbf{w})} \langle B_i \rangle \langle S_w(\mathbf{B}) \rangle, \quad (380)$$

where

$$S_w(\mathbf{B}) := \left(\sum_{\pi \in \mathcal{S}_g} \sum_{\mathbf{v}:|\mathbf{w}|+2|\mathbf{v}|=g} \prod_{j \in \text{supp}(\mathbf{v}) \setminus \text{supp}(\mathbf{w})} B_j \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \right)^2. \quad (381)$$

Because the product of Bernoulli variables considers the support of \mathbf{w} , it is reasonable to reorganize the summation over \mathbf{w} by their support sizes, i.e.

$$\sum_{\mathbf{w}:|\mathbf{w}|=w} \mapsto \sum_{s=0}^w \sum_{\substack{\mathbf{w}:|\mathbf{w}|=w; \\ |\text{supp}(\mathbf{w})|=s}}. \quad (382)$$

Since $\langle B_i \rangle = p_B$ for all $i \in \{1, \dots, m\}$, and the product over the support of \mathbf{w} only contains distinct B_i 's, we have

$$\langle G_w(H(\mathbf{B})) \rangle = \sum_{s=1}^w \sum_{\substack{\mathbf{w}:|\mathbf{w}|=w; \\ |\text{supp}(\mathbf{w})|=s}} p_B^s \langle S_w(\mathbf{B}) \rangle. \quad (383)$$

Next, we want to evaluate $\langle S_w(\mathbf{B}) \rangle$. By first expanding the squared term, we get

$$\langle S_w(\mathbf{B}) \rangle = \sum_{\pi, \pi'} \sum_{\mathbf{v}, \mathbf{v}'} \left\langle \prod_{j, j'} B_j B_{j'} \right\rangle \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v})]) \mathbb{1}(\pi'[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{v}')]) \quad (384)$$

where we hide the summation constraints $\pi, \pi' \in \mathcal{S}_g$, $\mathbf{v}, \mathbf{v}' \in \mathbb{Z}_{\geq 0}^g$ such that $|\mathbf{w}| + 2|\mathbf{v}| = |\mathbf{w}| + 2|\mathbf{v}'| = g$, for notational convenience. The product over j and j' might overlap, since the supports of \mathbf{v} and \mathbf{v}' might overlap. At the same time, the two vectors could also overlap with \mathbf{w} . Because the

¹⁹ B_i is independent from B_j if $i \neq j$.

product over B_j and $B_{j'}$ is taken over the non-overlapping support of \mathbf{v} and \mathbf{v}' with \mathbf{w} , we should first split these two vectors accordingly. We consider:

$$\mathbf{v} = \mathbf{x} + \mathbf{y} \text{ and } \mathbf{v}' = \mathbf{x}' + \mathbf{y}' \quad (385)$$

such that $\text{supp}(\mathbf{x}), \text{supp}(\mathbf{x}') \subseteq \text{supp}(\mathbf{w})$ and $\text{supp}(\mathbf{x}) \cap \text{supp}(\mathbf{y}) = \text{supp}(\mathbf{x}') \cap \text{supp}(\mathbf{y}') = \emptyset$. In this way, we have²⁰

$$\langle S_{\mathbf{w}}(\mathbf{B}) \rangle = \sum_{\pi, \pi'} \sum_{\mathbf{x}, \mathbf{x}'} \sum_{\mathbf{y}, \mathbf{y}'} \left\langle \prod_{\substack{j \in \text{supp}(\mathbf{y}) \\ j' \in \text{supp}(\mathbf{y}')}} B_j B_{j'} \right\rangle \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{x}, 2\mathbf{y})]) \mathbb{1}(\pi'[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{x}', 2\mathbf{y}')]). \quad (387)$$

Notice that now the indices over which the product of B_j and $B_{j'}$ are taken, match exactly with the support of \mathbf{y} and \mathbf{y}' . Since we are taking this product over the supports, and its value only depends on the size of the supports and their overlap, we split the summations even further into

$$\sum_{\mathbf{x}, \mathbf{x}'} \sum_{\mathbf{y}, \mathbf{y}'} \mapsto \sum_{a=0}^{w/2} \sum_{a'=0}^{w/2} \sum_{b=0}^{(g-w)/2} \sum_{b'=0}^{(g-w)/2} \sum_{c=0}^{\min(b, b')} \sum_{\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}': * } \quad (388)$$

where the *-condition is given by

$$* : \begin{cases} |\mathbf{x}| = a \\ |\mathbf{x}'| = a' \\ |\text{supp}(\mathbf{y})| = b \\ |\text{supp}(\mathbf{y}')| = b' \\ |\text{supp}(\mathbf{y}) \cap \text{supp}(\mathbf{y}')| = c \\ |\mathbf{w}| + 2|\mathbf{x}| + 2|\mathbf{y}| = g \end{cases} . \quad (389)$$

In this way, given any b, b' and c , the average product over B_j and $B_{j'}$ is always

$$\left\langle \prod_{\substack{j \in \text{supp}(\mathbf{y}) \\ j' \in \text{supp}(\mathbf{y}')}} B_j B_{j'} \right\rangle = p_B^{b+b'-c} \quad (390)$$

by construction, and we get

$$\langle S_{\mathbf{w}}(\mathbf{B}) \rangle = \sum_{\pi, \pi'} \sum_{a, a'=0}^{w/2} \sum_{b, b'=0}^{(g-w)/2} \sum_{c=0}^{\min(b, b')} \sum_{\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}': *} p_B^{b+b'-c} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{x}, 2\mathbf{y})]) \mathbb{1}(\pi'[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{x}', 2\mathbf{y}')]). \quad (391)$$

However, the *-condition might not be achievable for some combinations of a, a', b, b' and c . In those cases, the contribution of the sum over $\mathbf{x}, \mathbf{x}', \mathbf{y}$ and \mathbf{y}' is simply zero.

Given a vector \mathbf{w} and achievable a, a', b, b' , and c , let us focus on the following sum of indicators:

$$\sum_{\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}': *} \mathbb{1}(\pi[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{x}, 2\mathbf{y})]) \mathbb{1}(\pi'[\boldsymbol{\eta}(\mathbf{w}, 2\mathbf{x}', 2\mathbf{y}')]). \quad (392)$$

²⁰We extend the definition of the unary encoding to

$$\boldsymbol{\eta}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = (\underbrace{1, \dots, 1}_{a_1 \text{ times}}, \dots, \underbrace{m, \dots, m}_{a_m \text{ times}}, \underbrace{1, \dots, 1}_{b_1 \text{ times}}, \dots, \underbrace{m, \dots, m}_{b_m \text{ times}}, \underbrace{1, \dots, 1}_{c_1 \text{ times}}, \dots, \underbrace{m, \dots, m}_{c_n \text{ times}}), \quad (386)$$

for vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \{0, \dots, g\}^m$.

By expanding the unary encoding, we get

$$\begin{aligned} & \sum_{\phi:\text{surj } k_1, \dots, k_a \in \{i_w, \dots, i_1\}} \sum_{k'_1, \dots, k'_{a'} \in \{i_w, \dots, i_1\} \text{ distinct}} \sum_{j_1, \dots, j_{b+b'-c}=1}^m \\ & \mathbb{1} \left(\pi \left(i_w, \dots, i_1, k_a, k_a, \dots, k_1, k_1, j_{\phi(\frac{g-w}{2}-a)}, j_{\phi(\frac{g-w}{2}-a)}, \dots, j_{\phi(1)}, j_{\phi(1)} \right) \right) \\ & \cdot \mathbb{1} \left(\pi' \left(i_w, \dots, i_1, k'_{a'}, k'_{a'}, \dots, k'_1, k'_1, j_{\phi(g-w-a-a')}, j_{\phi(g-w-a-a')}, \dots, j_{\phi(\frac{g-w}{2}+1-a)}, j_{\phi(\frac{g-w}{2}+1-a)} \right) \right), \end{aligned} \quad (393)$$

where the summations over k_i and k'_i come from \mathbf{x} and \mathbf{x}' whose support overlaps with the support of \mathbf{w} , and the summations over j_i and j'_i come from \mathbf{y} and \mathbf{y}' whose supports are disjoint with the support of \mathbf{w} . For the latter summations, we have introduced surjective functions $\phi : \{1, \dots, g-w-a-a'\} \rightarrow \{1, \dots, b+b'-c\}$ to control the subscripts of j , since we are organizing the summation over vectors \mathbf{y}, \mathbf{y}' by their support sizes. These functions distribute the summation indices $j_1, \dots, j_{b+b'-c} = 1$ over the j -part of the two indicators. To capture the overlap between the supports, we require these surjective functions to satisfy the condition such that $\{\phi(1), \dots, \phi(\frac{g-w}{2}-a)\}$ contains b distinct elements, while $\{\phi(\frac{g-w}{2}+1-a), \dots, \phi(g-w-a-a')\}$ contains b' distinct elements. In this way, since we are summing over $b+b'-c$ distinct elements, there is guaranteed to be an overlap of c elements. Note that by introducing the constraints above for ϕ , we are only summing over a subset of all possible surjective functions. As our aim is to deliver an upper bound, there is no need to count the number of ϕ 's exactly. We can safely ignore these constraints and upper bound our sum using the set of all surjective functions from $\{1, \dots, g-w-a-a'\}$ to $\{1, \dots, b+b'-c\}$.

By doing the same trick for the summation over \mathbf{w} in equation (383), we can now upper bound $\langle G_w(H(\mathbf{B})) \rangle$ by

$$\begin{aligned} & \sum_{s=0}^w \sum_{\psi:\text{surj } i_1, \dots, i_s=1} \sum_{\pi, \pi' a, a'=0}^m \sum_{b, b'=0}^{w/2} \sum_{c=0}^{(g-w)/2 \min(b, b')} p_B^{s+b+b'-c} \\ & \sum_{\phi:\text{surj } k_1, \dots, k_a \in \text{supp}(\mathbf{w})} \sum_{k'_1, \dots, k'_{a'} \in \text{supp}(\mathbf{w}) \text{ distinct}} \sum_{j_1, \dots, j_{b+b'-c}=1}^m \\ & \mathbb{1} \left(\pi \left(i_{\psi(w)}, \dots, i_{\psi(1)}, k_a, k_a, \dots, k_1, k_1, j_{\phi(\frac{g-w}{2}-a)}, j_{\phi(\frac{g-w}{2}-a)}, \dots, j_{\phi(1)}, j_{\phi(1)} \right) \right) \\ & \cdot \mathbb{1} \left(\pi' \left(i_{\psi(w)}, \dots, i_{\psi(1)}, k'_{a'}, k'_{a'}, \dots, k'_1, k'_1, j_{\phi(g-w-a-a')}, j_{\phi(g-w-a-a')}, \dots, j_{\phi(\frac{g-w}{2}+1-a)}, j_{\phi(\frac{g-w}{2}+1-a)} \right) \right) \end{aligned} \quad (394)$$

where $\psi : \{1, \dots, w\} \rightarrow \{1, \dots, s\}$ are surjective functions to organize the subscripts of i 's based on support size of \mathbf{w} (no need to impose extra conditions here), and the summations over i_1, \dots, i_s and $j_1, \dots, j_{b+b'-c}$ should be interpreted as summation over distinct elements. Based on these summations, we can reformulate the summation over k_1, \dots, k_a and $k'_1, \dots, k'_{a'}$ more nicely so that it does not explicitly depend on the i -indices, but as repetitions of these indices.

$$\begin{aligned} & \sum_{s=0}^w \sum_{\psi:\text{surj } \pi, \pi' a, a'=0} \sum_{b, b'=0}^{w/2} \sum_{c=0}^{(g-w)/2 \min(b, b')} p_B^{s+b+b'-c} \sum_{\phi:\text{surj } \sigma, \sigma':\text{func}} \sum_{\text{distinct } i_1, \dots, i_s=1} \sum_{\text{distinct } j_1, \dots, j_{b+b'-c}=1}^m \\ & \mathbb{1} \left(\pi \left(i_{\psi(w)}, \dots, i_{\psi(1)}, i_{\sigma(a)}, i_{\sigma(a)}, \dots, i_{\sigma(1)}, i_{\sigma(1)}, j_{\phi(\frac{g-w}{2}-a)}, j_{\phi(\frac{g-w}{2}-a)}, \dots, j_{\phi(1)}, j_{\phi(1)} \right) \right) \\ & \cdot \mathbb{1} \left(\pi' \left(i_{\psi(w)}, \dots, i_{\psi(1)}, i_{\sigma'(a')}, i_{\sigma'(a')}, \dots, i_{\sigma'(1)}, i_{\sigma'(1)}, j_{\phi(g-w-a-a')}, j_{\phi(g-w-a-a')}, \dots, j_{\phi(\frac{g-w}{2}+1-a)}, j_{\phi(\frac{g-w}{2}+1-a)} \right) \right) \end{aligned} \quad (395)$$

where we use arbitrary well-defined functions $\sigma : \{1, \dots, a\} \rightarrow \{1, \dots, s\}$ and $\sigma' : \{1, \dots, a\} \rightarrow \{1, \dots, s\}$ to capture the repetition of indices i_1, \dots, i_s . Similar to ψ and ϕ , these functions only

care about assigning the summation indices as labels rather than their actual value. We have also moved the summation over the i -indices down to the second line, as they commute with the summation over ϕ and σ .

We can now apply the same argument of counting commutator chains as we did previously for general G_w . Let us first evaluate the summation over the second indicator by fixing s, ψ, i -indices $i_1, \dots, i_s, \phi, \sigma, \sigma'$ and the j -indices $j_{\phi(\frac{g-w}{2}-a)}, \dots, j_{\phi(1)}$ in the first indicator. For the second indicator, we only sum over the j -indices that do not overlap with the j -indices in the first indicator. By construction, there are c j -indices in the second indicator that is already fixed due to the overlap. Thus, when only summing over the second indicator, we can expect the following constraints for selecting local terms on each layer of the corresponding commutator chain:

1. it is the first layer, or
2. it must anti-commute with one of the previous layers, or
3. it is fixed for being one of $H_{\gamma(i_1)}, \dots, H_{\gamma(i_s)}$, or
4. it is a duplicate layer, or
5. it is fixed for being in the overlap with the j -indices in the first indicator.

Constraints 3, 4 and 5 leave us with only 1 choice, while the first and second constraints leave us with up to $g^{b'-c-1} m Q_{\max}(H)^{b'-c-1}$. Then, for the first indicator, summing i_1, \dots, i_s and the previously fixed j -indices $j_{\phi(\frac{g-w}{2}-a)}, \dots, j_{\phi(1)}$ while keeping every other summation index still fixed, we expect the following constraints for selecting local terms on each layer of the corresponding commutator chain:

1. it is the first layer, or
2. it must anti-commute with one of the previous layers, or
3. it is a duplicate layer.

The last constraint leaves us with only up to 1 option. But the first and second constraint leaves us with up to $g^{s+b-1} m Q_{\max}(H)^{s+b-1}$ options because we are selecting s distinct elements in the i -indices and b distinct element sin the j -indices. Putting everything together, we can upper bound $\langle G_w(H(\mathbf{B})) \rangle$ by

$$\sum_{s=0}^w \sum_{\psi: \text{surj}} \sum_{\pi, \pi'} \sum_{a, a'=0}^{w/2} \sum_{b, b'=0}^{(g-w)/2} \sum_{c=0}^{\min(b, b')} p_B^{s+b+b'-c} \sum_{\phi: \text{surj}} \sum_{\sigma, \sigma': \text{func}} g^{s+b+b'-c-1} m^2 Q_{\max}(H)^{s+b+b'-c-1} \quad (396)$$

Since the summand now is independent of label-distributing functions $\psi, \phi, \sigma, \sigma'$ and permutations π, π' , we can contract these summations into the following factors. For the permutations, recall that $|\mathcal{S}_g| = g! \leq g^g$. Hence, $\sum_{\pi, \pi'} 1 \leq g^{2g}$. For the label-distributing functions, we use the fact that, in general, the number of arbitrary well-defined functions from $[N] \rightarrow [M]$ is upper bounded by M^N . Hence,

$$\sum_{\psi: \text{surj}} 1 \leq s^w, \quad \sum_{\phi: \text{surj}} 1 \leq (b+b'-c)^{g-w-a-a'}, \quad \sum_{\sigma, \sigma': \text{func}} 1 \leq s^{a+a'}. \quad (397)$$

By construction $s \leq g$ and $(b+b'-c) \leq g$. Therefore,

$$\sum_{\psi: \text{surj}} \sum_{\phi: \text{surj}} \sum_{\sigma, \sigma': \text{func}} 1 \leq g^{w+g-w-a-a'+a+a'} = g^g. \quad (398)$$

Thus, $\langle G_w(H(\mathbf{B})) \rangle$ is upper bounded by

$$g^{3g} \sum_{s=0}^w \sum_{a,a'=0}^{w/2} \sum_{b,b'=0}^{(g-w)/2} \sum_{c=0}^{\min(b,b')} p_B^{s+b+b'-c} g^{s+b+b'-c-2} m^2 Q_{\max}(H)^{s+b+b'-c-2} \quad (399)$$

Contract $\sum_{a,a'} 1 \leq w^2 \leq g^2$, note that by construction $s + b + b' - c \leq g$, and reorganize the summands in equation (396), we get the following upper bound

$$\langle G_w(H(\mathbf{B})) \rangle \leq g^{4g} m^2 Q_{\max}(H)^{-2} \sum_{s=0}^w \sum_{b,b'=0}^{(g-w)/2} \sum_{c=0}^{\min(b,b')} (p_B Q_{\max}(H))^{s+b+b'-c}. \quad (400)$$

The bound in the theorem uses a different labeling where q, q' takes over b, b' . This completes the proof. □