Concentration for Trotter error

Chi-Fang Chen[1,*] and Fernando G.S.L. Brandão[1,2]

[1] *Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA*
[2] *AWS Center for Quantum Computing, Pasadena, CA*

(Dated: November 10, 2021)

Quantum simulation is expected to be one of the key applications of future quantum computers. Product formulas, or Trotterization, are the oldest and, still today, an appealing method for quantum simulation. For an accurate product formula approximation in the spectral norm, the state-of-the-art gate complexity depends on the number of terms in the Hamiltonian and a certain 1-norm of its local terms.

This work considers the *concentration* aspects of Trotter error: we prove that, typically, the Trotter error exhibits 2-norm (i.e., incoherent) scaling; the current estimate with 1-norm (i.e., coherent) scaling is for the worst cases. For general k-local Hamiltonians and higher-order product formulas, we obtain gate count estimates for input states drawn from a 1-design ensemble (which includes, e.g., computational basis states). Our gate count depends on the number of terms in the Hamiltonian but replaces the 1-norm quantity by its analog in 2-norm, giving significant speedup for systems with large connectivity. Our typical-case results generalize to Hamiltonians with Fermionic terms and when the input state is drawn from a low-particle number subspace. Further, when the Hamiltonian itself has Gaussian coefficients, such as the SYK models, we show the stronger result that the 2-norm behavior persists even for the worst input state.

Our primary technical tool is a family of simple but versatile inequalities from non-commutative martingales called *uniform smoothness*. We use them to derive *Hypercontractivity*, namely p-norm estimates for low-degree polynomials (including k-local Pauli operators), which implies concentration via Markov's inequality. In terms of optimality, we give examples that simultaneously match our p-norm estimates and the spectral norm estimates. Therefore, our improvement is due to asking a qualitatively different question from the spectral norm bounds. Our results give evidence that product formulas in practice may generically work much better than expected.

CONTENTS

* chifang@caltech.edu

.

I. INTRODUCTION

Given a k-local Hamiltonian

$$\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}_\gamma \tag{1.1}$$

composed of Γ terms, the goal of quantum simulation is to implement an approximation of its time evolution operator $\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}$ at time t, using a quantum computer. As a ubiquitous subroutine, it already finds promising applications in material science and quantum chemistry [1–5]. However, despite the simplicity of the problem statement, developing quantum simulation algorithms that minimize the required resources (e.g., the gate complexity) has drawn tremendous effort [6–8].

Albeit an old subject, *product formulas*, or *Trotterization*, have been recently revived for quantum computing due to their relatively simple implementation without ancillae. Intuitively, it simulates the exponential by a product of individual exponentials, for example

$$\mathrm{e}^{\mathrm{i}(\boldsymbol{H}_1+\boldsymbol{H}_2)t} = \mathrm{e}^{\mathrm{i}\boldsymbol{H}_1 t}\mathrm{e}^{\mathrm{i}\boldsymbol{H}_2 t} + \mathcal{O}(t^2). \tag{1.2}$$

Constructions such as the Lie-Trotter-Suzuki [9, 10] formulas generalize to Hamiltonians with many terms and to a higher-order approximation $\mathcal{O}(t^{\ell+1})$. Nevertheless, it was not until recently that the theoretical guarantees appropriately capture the structure of the problem, such as initial state knowledge [11] and spatial locality of the model [12]. Especially, the seminal work [8] put together an analytic framework that exploits commutation relations in k-local Hamiltonians. It showed that using higher-order formulas, the gate complexity

$$G \approx \Gamma \left\|\boldsymbol{H}\right\|_{(1),1} t, \quad \left\|\boldsymbol{H}\right\|_{(1),1} := \max_i \sum_{\gamma: i\in\gamma} \left\|\boldsymbol{H}_\gamma\right\|, \tag{1.3}$$

suffices for an accurate approximation in the spectral norm, where the maximization is taken over sites i and the sum over terms γ overlapping with i. The bound depends on the number of terms Γ in the Hamiltonian and $\left\|\boldsymbol{H}\right\|_{(1),1}$, a certain 1-norm quantity that captures commutation relations. This theoretical guarantee renders product formulas among the strongest candidates for quantum simulation of various systems.

Given the above developments, one may ask: is constant overhead improvement all that remains in product formulas? In some other contexts, the folklore [3] suggests errors in quantum computing might in practice add up *incoherently*, which is more desirable than coherent noise [13–15]. Such a *2-norm* behavior for Trotter error, if true, is not manifested in the existing gate complexity in 1-norm ($\left\|\boldsymbol{H}\right\|_{(1),1}$).

A. Summary of Results

This work presents the *concentration* aspects of Trotter error that complement the state-of-the-art [8]: we show quantitatively that the Trotter error exhibits 2-norm scaling "typically", and the current estimates in 1-norm are for "worst case" inputs. For general k-local Hamiltonians, we obtain gate count estimates for typical input states. Further, motivated by quantum chaos and the famous SYK models [16, 17], we show that when the Hamiltonian itself has random coefficients, even the worst input states enjoy a 2-norm scaling. Our results give evidence that, in practice, product formula may generically work much better than expected.

1. *Non-random Hamiltonians*

Consider a k-local (i.e., a sum of Pauli strings of length k) Hamiltonian on n-qubits with Γ terms $\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}_\gamma$ and normalization $\|\boldsymbol{H}_\gamma\| \le b_\gamma$. To present our main results, define the normalized Schatten p-norms $\|\boldsymbol{O}\|_{\bar{p}} := \frac{\|\boldsymbol{O}\|_p}{\|\boldsymbol{I}\|_p}$ and the *2-norm* like quantities,

$$\|\boldsymbol{H}\|_{(0),2} := \sqrt{\sum_\gamma b_\gamma^2}, \quad \|\boldsymbol{H}\|_{(1),2} := \max_i \sqrt{\sum_{\gamma: i\subset\gamma} b_\gamma^2}, \tag{1.4}$$

where the maximization is taken over single sites i.

Theorem I.1 (Trotter error in k-local models)**.** *To simulate a k-local Hamiltonian using ℓ-th order Suzuki formula, the gate count*

$$G \approx \left(\frac{\|\boldsymbol{H}\|_{(0),2}\, t}{\epsilon}\right)^{1/\ell} \Gamma \|\boldsymbol{H}\|_{(1),2}\, t \quad \textit{ensures} \quad \left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_\ell(t/r)^r\right\|_{\bar{p}} \le \epsilon. \tag{1.5}$$

The p-norm estimate implies concentration for typical input states via Markov's inequality.

Corollary I.1.1. *Draw $|\psi\rangle$ from a 1-design ensemble (i.e., $\mathbb{E}[|\psi\rangle\langle\psi|] = \boldsymbol{I}/\mathrm{Tr}[\boldsymbol{I}]$), then with high probability, gate count*

$$G \approx \left(\frac{\|\boldsymbol{H}\|_{(0),2}\, t}{\epsilon}\right)^{1/\ell} \cdot \Gamma \|\boldsymbol{H}\|_{(1),2}\, t \quad \textit{ensures} \quad \left\|(\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}(t/r)^r)\,|\psi\rangle\right\|_{\ell_2} \le \epsilon. \tag{1.6}$$

See Table I A 2 for the gate counts in various models and Section III for the explicit dependence on p and the failure probability hidden in (1.5) and (1.6). When the Hamiltonian contains Fermionic terms and/or the input is restricted to a low-particle number subspace, see Proposition III.3.1 and Proposition III.3.2 for analogous results[1].

Considering optimality of the results (Section IV), we construct a Hamiltonian that demonstrates a separation between the worst case and the typical case bounds: its Schatten p-norm saturates our estimates, while the operator norm saturates the state-of-the-art bound [8]. Namely, our 2-norm to 1-norm improvement is due to asking a qualitatively different question and not because the previous spectral norm bounds were loose.

Proposition I.1.1 (A Hamiltonian with different p-norms and spectral norm of Trotter error)**.** *For 2-local Hamiltonian on three subsystems of qubits $\mathcal{H} = \mathcal{H}_{S_1} \otimes \mathcal{H}_{S_2} \otimes \mathcal{H}_{S_3}$,*

$$\boldsymbol{H} = \sum_{s_1\in S_1, s_2\in S_2} \boldsymbol{Z}_{s_1}\boldsymbol{X}_{s_2} + \sum_{s_2\in S_2, s_3\in S_3} \boldsymbol{Y}_{s_2}\boldsymbol{Z}_{s_3}, \tag{1.7}$$

Take $|S_1| = |S_2| = |S_3| \to \infty$, then the first and second-order Trotter at short enough times matches the p-norm estimates in Theorem III.1 and also the spectral norm estimates [8].

We note that when the terms in the Hamiltonian are of uneven strength, the dependence on the number of terms Γ is not optimal; we can use a truncation argument [8] to improve the gate complexity at early times (Table I A 2). Interestingly, the error due to truncation also enjoys concentration (by directly using Hypercontractivity; see Appendix A).

---

[1] This applies to the electronic structure Hamiltonian [1, 18], but there the error is dominated by single site terms (1-local Pauli Z s), i.e., $\|\boldsymbol{H}\|_{(1),2} \sim \|\boldsymbol{H}\|_{(1),1}$. We only get improvement at lower order product formulas by $\|\boldsymbol{H}\|_{(0),2} \ll \|\boldsymbol{H}\|_{(0),1}$.

qDRIFT [7]	qubitization [6]	higher-orders	first-order Trotter	
$\lVert \boldsymbol{H} \rVert^2_{(0),1} t^2/\epsilon$	$\Gamma' \lVert \boldsymbol{H} \rVert_{(0),1} t$	$\Gamma \lVert \boldsymbol{H} \rVert_{(1),1} t$	$\Gamma \lVert \boldsymbol{H} \rVert_{(0),1} \lVert \boldsymbol{H} \rVert_{(1),1} t^2/\epsilon$	spectral norm [8]
		$\sqrt{n}\Gamma \lVert \boldsymbol{H} \rVert_{(1),2} t$	$n\Gamma \lVert \boldsymbol{H} \rVert_{(0),2} \lVert \boldsymbol{H} \rVert_{(1),2} t^2/\epsilon$	(all inputs)
		$\Gamma \lVert \boldsymbol{H} \rVert_{(1),2} t$	$\Gamma \lVert \boldsymbol{H} \rVert_{(0),2} \lVert \boldsymbol{H} \rVert_{(1),2} t^2/\epsilon$	(fixed input;typical inputs)
$n^{k+1}t^2/\epsilon$	$n^{\frac{3k+1}{2}} t$	$n^{\frac{3k-1}{2}} t$	$n^{2k}t^2/\epsilon$	k-local ($\sum_\gamma \lVert \boldsymbol{H}_\gamma \rVert^2 = \mathcal{O}(n)$)
		$n^{k+\frac{1}{2}} t$	$n^{k+\frac{3}{2}} t^2/\epsilon$	
		$n^k t$	$n^{k+\frac{1}{2}} t^2/\epsilon$	
$n^{4-2\alpha/d}t^2/\epsilon$	$n^{4-\frac{\alpha}{d}} t$	$n^{3-\frac{\alpha}{d}} t$	$n^{5-2\frac{\alpha}{d}} t^2/\epsilon$	Power-law ($d/2 \le \alpha \le d$)
		$n^{2+\frac{1}{2}} t$	$n^{2+\frac{3}{2}} t^2/\epsilon$	
		$n^2 t$	$n^{2+\frac{1}{2}} t^2/\epsilon$	
		$nt(\frac{nt^2}{\epsilon})^{\frac{d}{2\alpha-d}}$		$d < \alpha$(fixed;typical)

Table I. Comparison of gate complexities for simulation time t and system size n. $o(1/\ell)$ dependence is dropped for the higher-order formulas. Note that the "fixed input" is for random Hamiltonians and "typical inputs" is for random 1-design inputs, but we combine them as they have essentially the same gate complexities. The k-local models have uniform weights for each term, with an SYK-like normalization; the power-law interacting system is characterized by the dimension d and the decay exponent α (i.e., $\lVert \boldsymbol{H}_{xy} \rVert \le |x-y|^{-\alpha}$). In the qubitization gate counts, we plugged in $\Gamma' = \Gamma$ for comparison, but there are many particular examples, especially in quantum chemistry [3, 20] that Γ' can be potentially smaller.

2. *Random Hamiltonians*

Sometimes, Hamiltonians are defined as an ensemble, most notably the Sachdev-Ye-Kiteav [16, 17] models with random coefficients. The intrinsic randomness of the Hamiltonian allows us to obtain similar but stronger results. We consider random Hamiltonians $\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}_\gamma = \sum_{\gamma=1}^{\Gamma} g_\gamma \boldsymbol{Z}_\gamma$, where g_γ are i.i.d standard Gaussian $\mathbb{E}[g_\gamma^2] = 1$ and $\lVert \boldsymbol{Z}_\gamma \rVert \le b_\gamma$ are bounded deterministic matrices.

Theorem I.2 (Trotter error in random models)**.** *Simulating random k-local models with Gaussian coefficients via higher-order Suzuki formulas, the asymptotic gate count*

$$G \approx \left(\frac{\sqrt{n} \lVert \boldsymbol{H} \rVert^2_{(0),2} t}{\epsilon \lVert \boldsymbol{H} \rVert_{(1),2}} \right)^{1/\ell} \Gamma \lVert \boldsymbol{H} \rVert_{(1),2} t\sqrt{n} \text{ ensures } \left\lVert e^{i\boldsymbol{H}t} - \boldsymbol{S}(t/r)^r \right\rVert \le \epsilon \qquad \text{(all inputs)}, \tag{1.8}$$

$$G \approx \left(\frac{\lVert \boldsymbol{H} \rVert^2_{(0),2} t}{\epsilon \lVert \boldsymbol{H} \rVert_{(1),2}} \right)^{1/\ell} \Gamma \lVert \boldsymbol{H} \rVert_{(1),2} t \text{ ensures } \left\lVert e^{-i\boldsymbol{H}t} \boldsymbol{\rho} e^{i\boldsymbol{H}t} - \boldsymbol{S}(t/r)^{\dagger r} \boldsymbol{\rho} \boldsymbol{S}(t/r)^r \right\rVert_1 \le \epsilon \qquad \text{(fixed input state)}, \tag{1.9}$$

with high probability drawing from the random Hamiltonian ensemble. $\boldsymbol{\rho}$ denotes an arbitrary fixed input state.

See Section VIII for the complete theorem with explicit dependence on the failure probability and Theorem VI.1 for a precise gate count for the first-order Trotter. In other words, when the Hamiltonian is random, an arbitrary fixed input state exhibits 2-norm scaling of Trotter error. A slightly higher gate count (by a factor $\sqrt{n}$) would control the performance for the worst inputs that may correlate with the Hamiltonian (e.g., the Gibbs state or the ground state of the model).

Proposition I.2.1 (Distinct Hamiltonians)**.** *There exists $e^{\Omega(\Gamma)}$ Hamiltonians $\{\boldsymbol{H}^{(i)}\}$ such that each satisfies $\boldsymbol{H}^{(i)} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}^{(i)}_\gamma$, $\lVert \boldsymbol{H}^{(i)}_\gamma \rVert \le \mathcal{O}(1/\sqrt{n}^{k-1})$, but for times $t = \Omega(1)$ they are pairwise distinct*

$$\forall i \ne j, \left\lVert \boldsymbol{H}^{(i)} - \boldsymbol{H}^{(j)} \right\rVert_\infty t \ge \Omega(\sqrt{n}). \tag{1.10}$$

If we further assume the exponentials are also distinct (which is believable but hard to prove)

$$\left\lVert e^{\boldsymbol{H}^{(i)}t} - e^{\boldsymbol{H}^{(j)}t} \right\rVert_\infty \overset{?}{\ge} \Omega(1), \tag{1.11}$$

then this implies a circuit lower bound[2] $G = \Omega(\Gamma) = \Omega(n^k)$, which matches our gate complexity for fixed inputs (1.9) and typical input (1.5) at $t = \Omega(1)$ times.

[2] For SYK ($k = 4$, Majorana, Gaussian coefficients), [19] uses qubitization to obtain a gate count $G = n^{7/2}t + n^{5/2}t\log(n/\epsilon)$, which is lower than our $n^k t$. However, their n^4 Gaussians coefficients are not independent and hence not controlled by our circuit lower bounds. Physically, it is not clear whether SYK models with pseudorandom coefficients mimic the original ones.

B. Proof ingredients

The Trotter error is a complicated function of matrices. Recall the first order Lie-Trotter formula and the second-order Suzuki formula

$$\boldsymbol{S}_1(\tau) := \prod_{\gamma=1}^{\Gamma} \exp(\mathrm{i}\tau \boldsymbol{H}_\gamma), \quad \boldsymbol{S}_2(\tau) := \prod_{\gamma=\Gamma}^{1} \exp(\mathrm{i}(\tau/2)\boldsymbol{H}_\gamma) \cdot \prod_{\gamma=1}^{\Gamma} \exp(\mathrm{i}(\tau/2)\boldsymbol{H}_\gamma), \tag{1.12}$$

and the higher order Suzuki formulas constructed recursively

$$\boldsymbol{S}_{2p}(\tau) := \boldsymbol{S}_{2p-2}(q_p\tau)^2 \cdot \boldsymbol{S}_{2p-2}((1-4q_p)\tau) \cdot \boldsymbol{S}_{2p-2}(q_p\tau)^2, \tag{1.13}$$

where $q_p := 1/(4-4^{1/(2p-1)})$ [9].

Then, the leading order Trotter error is a commutator; for example in the first-order product formula

$$\boldsymbol{S}_1(t) - \mathrm{e}^{\sum_{\gamma=1}^{\Gamma} \mathrm{i}\boldsymbol{H}_\gamma t} = \frac{t^2}{2} \sum_{\gamma'>\gamma\geq 1}^{\Gamma} [\mathrm{i}\boldsymbol{H}_{\gamma'}, \mathrm{i}\boldsymbol{H}_\gamma] + O(t^3). \tag{1.14}$$

Analogously, the ℓ-th order product formulas have leading order error as a degree $\ell+1$ polynomial of commutators [8].

There are two main technical steps: First, how to take care of the infinite series of higher-order terms? Second, how to deliver concentration bounds for the commutator?

1. *A Good Presentation of Error*

The Trotter error has a rather nasty higher-order dependence on time, and a good expansion simplifies the proof. Here we build upon the framework from [8]. Starting with the target Hamiltonian $\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}_\gamma$, the amount that the product formula deviates from the target can be captured in the time-ordered exponential form

$$\prod_{j=1}^{J} \mathrm{e}^{\mathrm{i}a_j \boldsymbol{H}_{\gamma(j)} t} = \mathrm{e}^{\mathrm{i}a_J \boldsymbol{H}_{\gamma(J)} t} \cdots \mathrm{e}^{\mathrm{i}a_1 \boldsymbol{H}_{\gamma(1)} t} = \exp_{\mathcal{T}}\left(\mathrm{i}\int (\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_\Gamma, \Upsilon t) + \boldsymbol{H}) dt\right). \tag{1.15}$$

The error is now represented as a sum of nested commutators

$$\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_\Gamma, \Upsilon t) : = \sum_{j=1}^{J} \left(\prod_{k=j+1}^{J} \mathrm{e}^{a_k \mathcal{L}_{\gamma(k)} t} [a_j \boldsymbol{H}_{\gamma(j)}] - a_j \boldsymbol{H}_{\gamma(j)} \right), \tag{1.16}$$

where $\mathcal{L}_\gamma[O] := \mathrm{i}[\boldsymbol{H}_\gamma, O]$, and Υ is the number of stages. For the first-order Lie-Trotter formula, $\Upsilon = 1$; for ℓ-th order ($\ell \geq 2$) Suzuki formulas, $\Upsilon = J/\Gamma = 2 \cdot 5^{\ell/2-1}$, and the coefficients $|a_j| = \mathcal{O}(1)$ are of comparable sizes. In our proof, we will "beat $\boldsymbol{\mathcal{E}}$ to death"; do a Taylor expansion for the nested commutators, and each order will be a polynomial of matrices. We then apply our matrix concentration tools and go through a nasty combinatorics (which is much more involved than obtaining the 1-norm quantity $\|\boldsymbol{H}\|_{(1),1}$ in [8]).

2. *Uniform Smoothness, Matrix Martingales, and Hypercontractivity*

To obtain quantitative control of a complicated matrix function, let us begin with an instructive example that captures the different perspectives. Consider a Hamiltonian as a sum of single-site Paulis,

$$\boldsymbol{H} = \boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_n, \tag{1.17}$$

where each $\boldsymbol{Z}_i$ is supported on qubit i. How "big" is the sum?

(1) Take the spectral norm

$$\|\boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_n\| = n. \tag{1.18}$$

(2) Interpret the trace as an expectation, then its eigenvalue distribution is equivalent to a sum of independent random variables $S_n := x_1 + \cdots + x_n$, each drawn from the Rademacher distribution $\Pr(x_i = 1) = \Pr(x_i = -1) = 1/2$. Then we can use a *concentration inequality* to describe how rarely the random variable deviates from its expectation

$$\Pr(\lambda_i \geq \epsilon) \equiv \Pr(S_n \geq \epsilon) \leq \mathrm{e}^{-\epsilon^2/2n} \quad \text{(Hoeffding's inequality)}. \tag{1.19}$$

In other words, the *typical* eigenvalues $\lambda = \mathcal{O}(\sqrt{n}) \ll n$ are much smaller than the *worst-case* spectral norm. This simple example captures the overarching theme of this work: *Concentration is ubiquitous but often unspoken.*

To go beyond the above example, we rely on a family of recursive inequalities for their p-norms, which leads to concentration by Markov's inequality. We begin with reviewing the ancestral scalar version, often called the *two-point inequality* or *Bonami's inequality* (See, e.g., [21]).

Fact I.3 (Two-point inequality)**.** *For real numbers a, b,*

$$\frac{(a+b)^p + (a-b)^p}{2} \leq \left(a^2 + (p-1)b^2\right)^{p/2}. \tag{1.20}$$

This can be seen by expanding the binomial. This seemingly trivial inequality turns out to have far-reaching consequences for scalars and beyond, and its simplicity becomes its strength. The same form of inequality has an exact matrix analog, often called *uniform smoothness*.

Fact I.4 (Uniform smoothness for Schatten Classes [22])**.**

$$\left[\frac{1}{2}(\|\boldsymbol{X}+\boldsymbol{Y}\|_p^p + \|\boldsymbol{X}-\boldsymbol{Y}\|_p^p)\right]^{2/p} \leq \|\boldsymbol{X}\|_p^2 + (p-1)\,\|\boldsymbol{Y}\|_p^2. \tag{1.21}$$

The above form is not directly applicable, but its alternative forms with a martingale flavor streamline most of our proofs. For k-local operators (which is in fact closely related to *non-commutative martingales*), we derive and make heavy usage of the following:

Proposition I.4.1 (Uniform smoothness for subsystems)**.** *Consider matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{B}(\mathcal{H}_i \otimes \mathcal{H}_j)$ that satisfy $\mathrm{Tr}_i(\boldsymbol{Y}) = 0$ and $\boldsymbol{X} = \boldsymbol{X}_j \otimes \boldsymbol{I}_i$. For $p \geq 2$,*

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_p^2 \leq \|\boldsymbol{X}\|_p^2 + (p-1)\|\boldsymbol{Y}\|_p^2. \tag{1.22}$$

For example, this applies to the 2-local operator

$$\left\|\sum_{j<i} \boldsymbol{X}_i\boldsymbol{Y}_j\right\|_p^2 \leq (p-1)\sum_j \left\|\sum_{i:j<i} \boldsymbol{X}_i\boldsymbol{Y}_j\right\|_p^2 \tag{1.23}$$

$$\leq (p-1)^2 \sum_{j<i} \|\boldsymbol{X}_i\boldsymbol{Y}_j\|_p^2, \tag{1.24}$$

and more generally this gives concentration of k-local operators, or *Hypercontractivity* (Section II).

For random Hamiltonians, the flavor of the problem changed slightly; we can think of adding Gaussian coefficients in our guiding example

$$\boldsymbol{H} = g_1\boldsymbol{Z}_1 + \cdots + g_n\boldsymbol{Z}_n. \tag{1.25}$$

The Gaussian coefficient (i.e., external randomness) requires the following version of uniform smoothness regarding the expected p-norm $\|\!|\boldsymbol{X}|\!\|_p := (\mathbb{E}[\|\boldsymbol{X}\|_p^p])^{1/p}$ that will allow us to control the spectral norm, i.e., the worst input states. Initially, this featured in simple derivations of matrix concentration for martingales [23, 24].

Fact I.5 (Uniform smoothness for expected p-norm[24, Proposition 4.3])**.** *Consider random matrices $\boldsymbol{X}, \boldsymbol{Y}$ of the same size that satisfy $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}] = 0$. When $2 \leq p$,*

$$\|\!|\boldsymbol{X}+\boldsymbol{Y}|\!\|_p^2 \leq \|\!|\boldsymbol{X}|\!\|_p^2 + (p-1)\|\!|\boldsymbol{Y}|\!\|_p^2. \tag{1.26}$$

See Section V for the relevant background and an alternative norm for arbitrary fixed input states. Beyond the scope of this work, we emphasize these robust and straightforward martingale inequalities should find applications in versatile quantum information settings, whether by exploiting the tensor product structure of the Hilbert space or the randomness in matrix summands. See, e.g., [25, 26] for applications in operator growth and [27] in randomized quantum simulation.

C. Discussion

For many physical systems (i.e., non-random k-local Hamiltonians), our concentration results hold for typical input states, with a rapidly decaying tail of failure probability. Unfortunately, our current methods do not label for us the exceptional states. For quantum chemistry applications, one may use Trotter for phase estimation for (approximate) ground states. Our results strongly suggest a gate complexity with 2-norm scaling but not a guarantee for a particular state. Another natural question is whether the typical case mindset applies to qubitization methods, which we did not discuss in this work.

D. Acknowledgments

We thank Yuan Su and Mario Berta for helpful discussions and Joel Tropp and Andrew Lucas for related collaborations. CFC is especially thankful for Joel Tropp for introducing to him the subject of matrix concentration inequalities. After this work was completed, we became aware of related work by Qi Zhao, You Zhou, Alexander F. Shaw, Tongyang Li, and Andrew M. Childs that also studies Hamiltonian simulation with random input states. We thank them for letting us know about their work. CFC is supported by Caltech RA fellowship and the Eddleman Fellowship.

II. PRELIMINARY: k-LOCALITY, UNIFORM SMOOTHNESS, AND HYPERCONTRACTIVITY

In quantum information, we often encounter an operator $\boldsymbol{F} \in \mathcal{B}(\mathcal{H}(2^n))$ acting on n-qubits, with a "locality" guarantee that its expansion in Paulis is mostly supported on few qubits

$$\boldsymbol{F} = \boldsymbol{I} + \sum_{|S|=1} \sum_{\alpha=1}^{3} b^{\alpha}_{s_1} \boldsymbol{\sigma}^{\alpha}_{s_1} + \sum_{|S|=2} \sum_{\alpha=1}^{3} \sum_{\beta=1}^{3} c^{\alpha\beta}_{s_1 s_2} \boldsymbol{\sigma}^{\alpha}_{s_1} \boldsymbol{\sigma}^{\beta}_{s_2} + \cdots := \sum_{S} \boldsymbol{F}_S. \tag{2.1}$$

(In this section we use $\boldsymbol{\sigma}^x, \boldsymbol{\sigma}^y, \boldsymbol{\sigma}^z$ for Paulis as $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ is allocated for general matrices.) The sizes of the operator can be quantified in several ways. The operator norm $\|\boldsymbol{F}\| = \sup_{\psi} \|\boldsymbol{F}|\psi\rangle\|_{\ell_2}$ quantifies its strength over the worst possible state $|\psi\rangle$. One may also ask for a "typical" statement: for a state drawn randomly from a 1-design, how large is $\|\boldsymbol{F}|\psi\rangle\|_{\ell_2}$? A natural norm would now be the (normalized) p-norm that converts to concentration via a Markov's inequality.

Proposition II.0.1 (Typical states and Schatten p-norms)**.** *For a pure state $|\psi\rangle$ drawn from an ensemble $\mathbb{E}_{\psi}[|\psi\rangle\langle\psi|] = \boldsymbol{\rho}$,*

$$\Pr\left(\|\boldsymbol{F}|\psi\rangle\|_{\ell_2} \geq \epsilon\right) \leq \left(\frac{\left\|\boldsymbol{F}\boldsymbol{\rho}^{1/p}\right\|_p}{\epsilon}\right)^p. \tag{2.2}$$

In particular, for the maximally mixed state $\boldsymbol{\rho} = \boldsymbol{I}/\mathrm{Tr}[\boldsymbol{I}]$, this is the normalized Schatten p-norm $\left\|\boldsymbol{F}\boldsymbol{\rho}^{1/p}\right\|_p = \|\boldsymbol{F}\|_{\bar{p}}$.

Proof. Consider a pure state ensemble whose average is some mixed state $\mathbb{E}_{\psi}[|\psi\rangle\langle\psi|] = \boldsymbol{\rho}$, then

$$\mathbb{E}_{\psi}\left[\|\boldsymbol{F}|\psi\rangle\|^p_{\ell_2}\right] = \mathbb{E}_{\psi}\mathrm{Tr}\left[\left(\boldsymbol{F}|\psi\rangle\langle\psi|\boldsymbol{F}^{\dagger}\right)^{p/2}\right] \leq \mathrm{Tr}\left[\left(\boldsymbol{F}\rho^{\frac{1}{p}}\boldsymbol{F}^{\dagger}\right)^{p/2}\right], \tag{2.3}$$

where the inequality is due to certain form of concavity (Fact II.5). □

The rest of our discussions boil down to estimating the p-norm of k-local operators. Our main techniques are *Hypercontractivity*, which is more "global", and *uniform smoothness*, which is more "local".

Fact II.1 (non-commutative Hypercontractivity [28, Theorem 46])**.** *For $p \geq 2$, an operator acting on qubits $\boldsymbol{F} \in \mathcal{B}(\mathcal{H}(2^n))$, $C_p := p - 1$,*

$$\|\boldsymbol{F}\|_{\bar{p}} \leq \left\|\sum_{S} \sqrt{C_p}^{|S|} \boldsymbol{F}_S\right\|_{\bar{2}}. \tag{2.4}$$

In other words, so long as $|S|$ is small, e.g., the operator is k-local for a fixed k, we obtain all p-norm estimates from the 2-norm calculation. See also [29] for some applications of this fact, and note the equivalent inverted form is more common

$$\left\|\sum_{S} \frac{1}{\sqrt{C_p}^{|S|}} \boldsymbol{F}_S\right\|_{\bar{p}} \leq \|\boldsymbol{F}\|_{\bar{2}}. \tag{2.5}$$

However, in this work, we focus on the following recursive approach called uniform smoothness:

Proposition II.1.1 (Uniform smoothness for subsystems)**.** *Consider matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{B}(\mathcal{H}_i \otimes \mathcal{H}_j)$ that satisfy $\mathrm{Tr}_i(\boldsymbol{Y}) = 0$ and $\boldsymbol{X} = \boldsymbol{X}_j \otimes \boldsymbol{I}_i$. For $p \geq 2$, $C_p = p - 1$,*

$$\|\boldsymbol{X} + \boldsymbol{Y}\|^2_p \leq \|\boldsymbol{X}\|^2_p + C_p\|\boldsymbol{Y}\|^2_p. \tag{2.6}$$

The partially traceless assumption $\mathrm{Tr}_i(\boldsymbol{Y}) = 0$ makes it a *non-commutative martingale*.[3] The martingale requirement, being partially traceless, holds for a wide range of applications, while Hypercontractivity as a black box is more "global" and requires strict k-locality. Although these two ideas are intimately related, we emphasize uniform smoothness is a versatile[4] and transparent driving horse, which implies Hypercontractivity (Corollary II.2.1).

---

[3] $\mathcal{B}(\mathcal{H}_j) \otimes \boldsymbol{I}_i \subset \mathcal{B}(\mathcal{H}_i \otimes \mathcal{H}_j)$ is the filtration of von Neumann algebras. We will stick to the usual qubit picture and not take this formal route any further.

[4] In particular the above Hypercontractivity (Fact II.1) is restricted to qubits, ([30] generalizes to other unital noise operators, on qubits), while uniform smoothness generalizes fairly easily.

A. Uniform Smoothness for Subsystems

We first prove II.1.1 by adapting the argument in [24, Prop 4.3][5]. We first require the primitive form of uniform smoothness.

Fact II.2 (Uniform smoothness for Schatten Classes, recap [22]).

$$\left[\frac{1}{2}\left(\|\boldsymbol{X}+\boldsymbol{Y}\|_p^p+\|\boldsymbol{X}-\boldsymbol{Y}\|_p^p\right)\right]^{2/p} \le \|\boldsymbol{X}\|_p^2 + C_p\|\boldsymbol{Y}\|_p^2. \tag{2.7}$$

Proposition II.2.1 (Monotonicity of p-norm w.r.t partial trace). *For* $\mathrm{Tr}_i(\boldsymbol{Y})=0; \boldsymbol{X}=\boldsymbol{X}_j\otimes\boldsymbol{I}_i$, $p\ge 2$,

$$\|\boldsymbol{X}\|_p \le \|\boldsymbol{X}+\boldsymbol{Y}\|_p. \tag{2.8}$$

This is a non-commutative analog of convexity $\|\boldsymbol{X}+\mathbb{E}_{\boldsymbol{Y}}\boldsymbol{Y}\|_p \le \mathbb{E}_{\boldsymbol{Y}}\|\boldsymbol{X}+\boldsymbol{Y}\|_p$.

Proof. Recall the variational expression [31, Sec 12.2.1] for Schatten p-norms $\|\boldsymbol{X}_j\|_p = \sup_{\|\boldsymbol{B}_j\|_q\le 1}\mathrm{Tr}(\boldsymbol{X}_j^\dagger\boldsymbol{B}_j)$ for $1/p+1/q=1$. Then

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_p = \sup_{\|\boldsymbol{B}\|_q\le 1}\mathrm{Tr}[(\boldsymbol{X}+\boldsymbol{Y})^\dagger\boldsymbol{B}] \ge \mathrm{Tr}\left[(\boldsymbol{X}^\dagger+\boldsymbol{Y}^\dagger)\boldsymbol{B}_j\otimes\frac{\boldsymbol{I}_i}{\|\boldsymbol{I}_i\|_q}\right] = \|\boldsymbol{X}_j\otimes\boldsymbol{I}_i\|_p. \tag{2.9}$$

The last equality is that $\mathrm{Tr}_i\boldsymbol{Y}=0$ and that $\boldsymbol{X}_j\otimes\boldsymbol{I}_i$ has maximizer $\boldsymbol{B}_j\otimes\boldsymbol{I}_i/\|\boldsymbol{I}_i\|_q$.
An alternative proof is by averaging over Haar unitary on i

$$\|\boldsymbol{X}\|_p = \left\|\boldsymbol{X}+\mathbb{E}_{\boldsymbol{U}}[\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\dagger]\right\|_p \le \mathbb{E}_{\boldsymbol{U}}\left\|\boldsymbol{X}+\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\dagger\right\|_p \le \|\boldsymbol{X}+\boldsymbol{Y}\|_p. \tag{2.10}$$

The first equality is Schur's lemma, then convexity, and lastly we used unitary invariance of p-norm. □

We can almost prove Proposition II.1.1.

$$\frac{\|\boldsymbol{X}+\boldsymbol{Y}\|_p^2+\|\boldsymbol{X}\|_p^2}{2} \le \frac{\|\boldsymbol{X}+\boldsymbol{Y}\|_p^2+\|\boldsymbol{X}-\boldsymbol{Y}\|_p^2}{2} \tag{2.11}$$

$$\le \left(\frac{\|\boldsymbol{X}+\boldsymbol{Y}\|_p^p+\|\boldsymbol{X}-\boldsymbol{Y}\|_p^p}{2}\right)^{2/p} \le \|\boldsymbol{X}\|_p^2 + C_p\|\boldsymbol{Y}\|_p^2. \tag{2.12}$$

The last inequality is Lyapunov's and then Fact I.4. Rearranging terms yields a slightly worse constant $2C_p$. The advertised constant can be obtained via another elementary but insightful trick [24, Lemma A.1], which we reproduce as follows.

Proof of Proposition II.1.1. The proof considers a rescaling argument. Let $\boldsymbol{Z}:=\frac{1}{n}\boldsymbol{Y}$. We have just obtained

$$\|\boldsymbol{X}+\boldsymbol{Z}\|_p^2 - \|\boldsymbol{X}\|_p^2 \le 2C_p\|\boldsymbol{Z}\|_p^2. \tag{2.13}$$

Rearranging Fact I.4,

$$\|\boldsymbol{X}+k\boldsymbol{Z}\|_p^2 - \|\boldsymbol{X}-(k-1)\boldsymbol{Z}\|_p^2 \le \left(\|\boldsymbol{X}+(k-1)\boldsymbol{Z}\|_p^2 - \|\boldsymbol{X}+(k-2)\boldsymbol{Z}\|_p^2\right) + 2C_p\|\boldsymbol{Z}\|_p^2 \tag{2.14}$$

$$\le 2C_pk\|\boldsymbol{Z}\|_p^2. \tag{2.15}$$

The last inequality recursively applies of the first line for $n\ge k\ge 2$ and (2.13) at base case[6] $k=1$. Therefore,

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_p^2 = \sum_k\left(\|\boldsymbol{X}+k\boldsymbol{Z}\|_p^2 - \|\boldsymbol{X}-(k-1)\boldsymbol{Z}\|_p^2\right) \le \sum_k 2C_pk\|\boldsymbol{Z}\|_p^2 \tag{2.16}$$

$$= C_p\frac{n+1}{n}\|\boldsymbol{Y}\|_p^2. \tag{2.17}$$

Take $n\to\infty$ to obtain the sharp constant. □

---

[5] Uniform smoothness for subsystems was conceived during this and other work [25]. We include the proof in both.
[6] The quadratic rescaling argument inherits its uniform smoothness constant from *Fact I.4*, and the dependence of the base case constant vanishes in the limit. We just need some constant at the base. $\|\boldsymbol{X}+\boldsymbol{Z}\|_p^2 - \|\boldsymbol{X}\|_p^2 \le f(p)\|\boldsymbol{Z}\|_p^2$.

1. Subalgebras

Let us briefly point out that the same strategies work for a subalgebra $\mathcal{N} \subset \mathcal{M}$, which captures non-commutative martingales in full generality. This also provides a unifying perspective for the manipulations we are doing. For subalgebras $\mathcal{N} \subset \mathcal{M} \subset B(\mathcal{H})$, let $E : \mathcal{M} \to \mathcal{N}$ be the *projection to subalgebra* $\mathcal{N}$ (or *the trace-preserving conditional expectation*), with the defining properties: $E^\dagger[\boldsymbol{I}] = \boldsymbol{I}$ and for $\boldsymbol{X} \in \mathcal{N}, \boldsymbol{Z} \in \mathcal{M}$

$$\mathrm{Tr}[\boldsymbol{Z}\boldsymbol{X}] = \mathrm{Tr}[E[\boldsymbol{Z}]\boldsymbol{X}]. \tag{2.18}$$

Intuitively, E is the analog of normalized partial trace $\boldsymbol{I}_j \frac{\mathrm{Tr}_j[\cdot]}{\mathrm{Tr}[\boldsymbol{I}_j]}$. Using the notation natural in this setting, we reproduce the monotonicity.

Proposition II.2.2 (Monotonicity of p-norm w.r.t projection to subalgebra)**.** *Consider finite dimensional subalgebras $\mathcal{N} \subset \mathcal{M} \subset B(\mathcal{H})$ and the corresponding projection to subalgebra $E : \mathcal{M} \to \mathcal{N}$. Then, for any $\boldsymbol{Z} \subset \mathcal{M}$, $p \geq 2$,*

$$\left\| E[\boldsymbol{Z}] \right\|_p \leq \left\| \boldsymbol{Z} \right\|_p. \tag{2.19}$$

Proof. Again, consider the variational expression

$$\left\| \boldsymbol{X} \right\|_p = \sup_{\left\| \boldsymbol{B}_j \right\|_q \leq 1, \boldsymbol{B} \in \mathcal{N}} \mathrm{Tr}(\boldsymbol{X}^\dagger \boldsymbol{B}),\ 1/p + 1/q = 1. \tag{2.20}$$

Note that the maximum is attained in the same algebra $\boldsymbol{B} \in \mathcal{N}$ (This can be seen by the structure theorem of finite-dimensional von Neumann algebra. $\mathcal{N}$ is a direct sum of subsystems.). Then

$$\left\| \boldsymbol{Z} \right\|_p = \sup_{\left\| \boldsymbol{B}' \right\|_q \leq 1} \mathrm{Tr}[\boldsymbol{Z}^\dagger \boldsymbol{B}'] \geq \mathrm{Tr}\left[\boldsymbol{Z}^\dagger \boldsymbol{B}\right] = \mathrm{Tr}\left[E[\boldsymbol{Z}^\dagger]\boldsymbol{B}\right] = \mathrm{Tr}\left[E[\boldsymbol{Z}]^\dagger \boldsymbol{B}\right] = \left\| E[\boldsymbol{Z}] \right\|_p, \tag{2.21}$$

which is the advertised result. □

Through the same arguments, we conclude the discussion for subalgebras by the following.

Proposition II.2.3 (Uniform smoothness for subalgebras)**.** *Consider finite dimensional subalgebras $\mathcal{N} \subset \mathcal{M} \subset B(\mathcal{H})$ and the corresponding projection to subalgebra $E_{\mathcal{N}} : \mathcal{M} \to \mathcal{N}$. Then, for any $\boldsymbol{X} \in \mathcal{N}$, $\boldsymbol{Y} \in \mathcal{M}$, $E_{\mathcal{N}}[\boldsymbol{Y}] = 0$, and $p \geq 2$, $C_p = p - 1$,*

$$\|\boldsymbol{X} + \boldsymbol{Y}\|_p^2 \leq \|\boldsymbol{X}\|_p^2 + C_p \|\boldsymbol{Y}\|_p^2. \tag{2.22}$$

B. Deriving Hypercontractivity

Uniform smoothness, through a recursion, implies Hypercontractivity-like global formulas.

Proposition II.2.4 (Moment estimates for local operator)**.** *On n-qudits, consider an operator $\boldsymbol{F} = \sum_{S \subset \{m,\cdots,1\}} \boldsymbol{F}_S$ expanded in subsets S of subsystem $\{m, \cdots, 1\} \subset \{n, \cdots, 1\}$. Then, for $p \geq 2$, with $C_p := p - 1$,*

$$\|\boldsymbol{F}\|_p^2 \leq \sum_{S \subset \{m,\cdots,1\}} (C_p)^{|S|} \left\| \prod_{s \in S} (\boldsymbol{I} - \boldsymbol{I}_s \otimes \bar{\mathrm{Tr}}_s) \prod_{s' \in S^c} \boldsymbol{I}_{s'} \otimes \bar{\mathrm{Tr}}_{s'}[\boldsymbol{F}] \right\|_p^2 \tag{2.23}$$

$$:= \sum_{S \subset \{m,\cdots,1\}} (C_p)^{|S|} \left\| \boldsymbol{F}_S \right\|_p^2 \tag{2.24}$$

where $\bar{\mathrm{Tr}}_S[\cdot] := \frac{1}{\mathrm{Tr}_S[\boldsymbol{I}_S]} \mathrm{Tr}[\cdot]$ is the normalized trace on subsystem S, and S^c denotes the complement of set S w.r.t to subsystem $\{m, \cdots, 1\}$.

Notice the expansion is "stable" in the sense that it can be restricted to any subsystem $m \leq n$. Also, we do not require the constituents to be qubits, contrasting with existing specialized result Fact II.1. Intuitively, in the particular case $p = 2$, it becomes an equality since different sets are orthogonal in the Hilbert-Schmidt inner product. The point is, at general values of p where the Banach space loses the inner product structure, operators supported on different subsets S still interact "orthogonally", albeit with a factor $C_p^{|S|}$ that depends on the size of their supports.

This recovers the usual concentration for bounded independent summand (e.g., Hoeffdings' inequality).

Example II.2.1.

$$\|\boldsymbol{H}\|_p^2 = \left\| \sum_i \alpha_i \boldsymbol{\sigma}_i^x \right\|_p^2 \leq C_p \sum_i \alpha_i^2 \left\| \boldsymbol{\sigma}_i^x \right\|_p^2 = C_p \sum_i |\alpha_i|^2 \left\| \boldsymbol{\sigma}_i^x \right\| \|\boldsymbol{I}\|_p^2. \tag{2.25}$$

By Markov's inequality, we obtain concentration for eigenvalue distribution (as well as for typical state in a 1-design)

$$\Pr[\nu \geq \epsilon] \leq \frac{\mathbb{E}[\nu^p]}{\epsilon^p} = (\frac{(\mathbb{E}[\nu^p])^{1/p}}{\epsilon})^p \leq (\frac{\sqrt{(p-1)\sum_i |\alpha_i|^2}}{\epsilon})^p \leq \mathrm{e}\cdot\exp(-\frac{\epsilon^2}{\mathrm{e}\sum_i |\alpha_i|^2}), \tag{2.26}$$

where the expectation is taken as the normalized trace $\mathbb{E} := \mathrm{Tr}[\cdot]/\mathrm{Tr}[\boldsymbol{I}]$, hence $(\mathbb{E}[\nu^p])^{1/p} = \|\boldsymbol{H}\|_p / \|\boldsymbol{I}\|_p$.

Moreover, we obtain a similar sum-of-squares behavior for 4-local Pualis, albeit with heavier tails

Example II.2.2.

$$\left\| \sum_{i>j>k>\ell} \alpha_{ijk\ell} \boldsymbol{\sigma}^x_i \boldsymbol{\sigma}^x_j \boldsymbol{\sigma}^x_k \boldsymbol{\sigma}^x_\ell \right\|_p^2 \leq (C_p)^4 \sum_{i>j>k>\ell} |\alpha_{ijk\ell}|^2 \left\| \boldsymbol{\sigma}^x_i \boldsymbol{\sigma}^x_j \boldsymbol{\sigma}^x_k \boldsymbol{\sigma}^x_\ell \right\|^2 \|\boldsymbol{I}\|_p^2 \tag{2.27}$$

By Markov's inequality, we obtain

$$\Pr[\nu \geq \epsilon] \leq \frac{\mathbb{E}[\nu^p]}{\epsilon^p} \leq \left(\frac{\sqrt{C_p^4 \sum_{i>j>k>\ell} |\alpha_{ijk\ell}|^2}}{\epsilon} \right)^p \leq \exp\left(- \sqrt{\frac{\epsilon}{\mathrm{e}\sqrt{\sum_{i>j>k>\ell} |\alpha_{ijk\ell}|^2}}} \right) \tag{2.28}$$

which does not have a Gaussian tail anymore but still decays super-polynomially.

Let us now present the proof.

Proof. By induction, start with decomposing the component on the first qubit $m = 1$:

$$\|\boldsymbol{F}\|_p^2 \leq \left\| \boldsymbol{I}_1 \bar{\mathrm{Tr}}_1[\boldsymbol{F}] \right\|_p^2 + C_p \left\| (\boldsymbol{I} - \boldsymbol{I}_1 \bar{\mathrm{Tr}}_1)[\boldsymbol{F}] \right\|_p^2 \tag{2.29}$$

$$= \sum_{S\subset\{1\}} (C_p)^{|S|} \left\| (\boldsymbol{I} - \boldsymbol{I}_S \bar{\mathrm{Tr}}_S) \boldsymbol{I}_{S^c} \bar{\mathrm{Tr}}_{S^c}[\boldsymbol{F}] \right\|_p^2 = \sum_{S\subset\{1\}} (C_p)^{|S|} \|\boldsymbol{F}_S\|_p^2 \tag{2.30}$$

where we used Proposition II.1.1 for $\bar{\mathrm{Tr}}_1[(\boldsymbol{I}_1 - \boldsymbol{I}_1 \otimes \bar{\mathrm{Tr}}_1)\boldsymbol{F}] = 0$. Given the result for m, prove for $m+1$:

$$\|\boldsymbol{F}\|_p^2 \leq \sum_{S\subset\{m,\cdots,1\}} (C_p)^{|S|} \|\boldsymbol{F}_S\|_p^2 \tag{2.31}$$

$$\leq \sum_{S\subset\{m,\cdots,1\}} (C_p)^{|S|} \left(\left\| \boldsymbol{I}_{m+1} \bar{\mathrm{Tr}}_{m+1}[\boldsymbol{F}_S] \right\|_p^2 + C_p \left\| (\boldsymbol{I}_{m+1} - \boldsymbol{I}_{m+1} \bar{\mathrm{Tr}}_{m+1})[\boldsymbol{F}_S] \right\|_p^2 \right) \tag{2.32}$$

$$= \sum_{S'\subset\{m+1,m,\cdots,1\}} (C_p)^{|S'|} \|\boldsymbol{F}_{S'}\|_p^2 . \tag{2.33}$$

In the first inequality, we use the induction hypothesis; in the second, we decompose the operator w.r.t to the $m+1$-qubit. Lastly rewrite in terms of the set $S' \subset \{m+1, m, \cdots, 1\}$ to complete the induction. ☐

To compare with Hypercontractivity, it is worth bringing Proposition II.2.4 to the following form.

Corollary II.2.1 (Non-commutative Hypercontractivity)**.** *In the setting of Proposition II.2.4,*

$$\|\boldsymbol{F}\|_{\bar{p}}^2 \leq \sum_{S\subset\{m,\cdots,1\}} (3C_p)^{|S|} \|\boldsymbol{F}_S\|_{\bar{2}}^2 = \left\| \sqrt{3C_p}^{|S|} \boldsymbol{F}_S \right\|_{\bar{2}} . \tag{2.34}$$

This is equivalent to the existing bound (Fact II.1), up to slightly worse constants. However, the martingale formulation streamlines a simple proof (Proposition II.1.1) and, more importantly, allows us to adapt to different settings in the subsequent sections.

Proof. Bound the normalized p-norm $\|\boldsymbol{F}\|_{\bar{p}} := \frac{\|\boldsymbol{F}\|_p}{\|\boldsymbol{I}\|_p}$ by Pauli expansion

$$\|\boldsymbol{F}_S\|_{\bar{p}}^2 \leq \left(\sum_{\sigma_S} \|\boldsymbol{F}_{\sigma_S}\|_{\bar{p}} \right)^2 \leq \left(\sum_{\sigma_S} \|\boldsymbol{F}_{\sigma_S}\|_{\bar{2}} \right)^2 \leq (\sum_{\sigma_S}) \cdot \sum_{\sigma_S} \|\boldsymbol{F}_{\sigma_S}\|_{\bar{2}}^2 = 3^{|S|} \cdot \|\boldsymbol{F}_S\|_{\bar{2}}^2 , \tag{2.35}$$

which is the advertised result. ☐

C. Product Background States

Uniform smoothness generalizes to the $\boldsymbol{\rho}$-weighted p-norm for factorized state $\boldsymbol{\rho} = \otimes_i \boldsymbol{\rho}_i$. Recall [32] for $0 \le s \le 1$,

$$\|\boldsymbol{O}\|_{q,\boldsymbol{\rho},s} := \left\| \boldsymbol{\rho}^{\frac{1-s}{p}} \boldsymbol{O} \boldsymbol{\rho}^{\frac{s}{p}} \right\|_p, \tag{2.36}$$

where $s = 0, 1/2, 1$ are the more notable cases

$$\|\boldsymbol{O}\|_{q,\boldsymbol{\rho},\frac{1}{2}} := \left\| \boldsymbol{\rho}^{\frac{1}{2p}} \boldsymbol{O} \boldsymbol{\rho}^{\frac{1}{2p}} \right\|_p, \quad \|\boldsymbol{O}\|_{q,\boldsymbol{\rho},1} := \left\| \boldsymbol{O} \boldsymbol{\rho}^{\frac{1}{p}} \right\|_p. \tag{2.37}$$

Then, we have analogously the following.

Proposition II.2.5 (Uniform smoothness for subsystem, weighted)**.** *Consider background state* $\boldsymbol{\rho} = \boldsymbol{\rho}_j \otimes \boldsymbol{\rho}_i$ *and matrices* $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{B}(\mathcal{H}_i \otimes \mathcal{H}_j)$ *that satisfy* $\mathrm{Tr}_i(\boldsymbol{\rho}_i \boldsymbol{Y}) = 0; \boldsymbol{X} = \boldsymbol{X}_j \otimes \boldsymbol{I}_i$*. For* $p \ge 2$*,* $C_p = p - 1$*,*

$$\|\boldsymbol{X} + \boldsymbol{Y}\|^2_{p,\boldsymbol{\rho},s} \le \|\boldsymbol{X}\|^2_{p,\boldsymbol{\rho},s} + C_p \|\boldsymbol{Y}\|^2_{p,\boldsymbol{\rho},s}. \tag{2.38}$$

All we need is the following modification of monotonicity.

Fact II.3 (monotonicity w.r.t partial trace)**.** *For* $\mathrm{Tr}_i(\boldsymbol{\rho}_i \boldsymbol{Y}) = 0; \boldsymbol{X} = \boldsymbol{X}_j \otimes \boldsymbol{I}_i$*. For* $p \ge 2$*,*

$$\|\boldsymbol{X}\|_{p,\boldsymbol{\rho},s} \le \|\boldsymbol{X} + \boldsymbol{Y}\|_{p,\boldsymbol{\rho},s}. \tag{2.39}$$

Proof. Once again, plug in the variational expression

$$\left\| \boldsymbol{\rho}_j^{\frac{1-s}{2p}} \boldsymbol{X}_j \boldsymbol{\rho}_j^{\frac{s}{2p}} \right\|_p = \sup_{\|\boldsymbol{B}_j\|_q \le 1} \mathrm{Tr}\left[\boldsymbol{\rho}_j^{\frac{1-s}{2p}} \boldsymbol{X}_j^\dagger \boldsymbol{\rho}_j^{\frac{s}{2p}} \boldsymbol{B}_j \right] \tag{2.40}$$

for $1/p + 1/q = 1$, which will be attained at some $\boldsymbol{B}_j$. Then by Proposition II.2.1,

$$\|\boldsymbol{X} + \boldsymbol{Y}\|_{p,\boldsymbol{\rho},s} = \sup_{\|\boldsymbol{B}\|_q \le 1} \mathrm{Tr}\left(\boldsymbol{\rho}_j^{\frac{1-s}{2p}} (\boldsymbol{X}_j + \boldsymbol{Y}_j)^\dagger \boldsymbol{\rho}_j^{\frac{s}{2p}} \boldsymbol{B} \right) \tag{2.41}$$

$$\ge \mathrm{Tr}\left[\boldsymbol{\rho}_j^{\frac{1-s}{2p}} (\boldsymbol{X}_j + \boldsymbol{Y}_j)^\dagger \boldsymbol{\rho}_j^{\frac{s}{2p}} \cdot \boldsymbol{B}_j \otimes \frac{\boldsymbol{\rho}_i^{\frac{1}{q}}}{\|\boldsymbol{\rho}_i^{\frac{1}{q}}\|_q} \right] = \|\boldsymbol{X}_j \otimes \boldsymbol{I}_i\|_{p,\boldsymbol{\rho},s}. \tag{2.42}$$

In the last inequality we used that $\mathrm{Tr}_i[\boldsymbol{\rho}_i \boldsymbol{Y}] = 0$ and that $\boldsymbol{\rho}^{\frac{1-s}{2p}} (\boldsymbol{X}_j \otimes \boldsymbol{I}_i) \boldsymbol{\rho}^{\frac{s}{2p}}$ has maximizer $\boldsymbol{B}_j \otimes \frac{\boldsymbol{\rho}_i^{1/q}}{\|\boldsymbol{\rho}_i^{1/q}\|_q}$). ☐

Combining the above with Fact I.4, we obtain Proposition II.2.5. We automatically get a weighted version of a Hypercontractivity-like formula. Let us first define the appropriate operator re-centered w.r.t. the background $\mathrm{Tr}[\boldsymbol{\rho}_i \boldsymbol{O}_i^\eta] = 0$

$$\boldsymbol{O}^\eta := (1 - \eta) |1\rangle \langle 1| - \eta |0\rangle \langle 0| \tag{2.43}$$

as a "shifted" Pauli $\boldsymbol{\sigma}^z$.

Proposition II.3.1 (Moment estimates for local operator, $\boldsymbol{\rho}$-weighted)**.** *Consider an operator* $\boldsymbol{F} = \sum_{S \subset \{m, \cdots, 1\}} \boldsymbol{F}_S$ *expanded in terms of local operators* $\{\boldsymbol{\sigma}^y, \boldsymbol{\sigma}^x, \boldsymbol{O}^\eta, \boldsymbol{I}\}$ *of subsystem* $\{m, \cdots, 1\} \subset \{n, \cdots, 1\}$*. For* $p \ge 2$*,* $C_p = p - 1$

$$\|\boldsymbol{F}\|^2_{p,\boldsymbol{\rho},s} \le \sum_{S \subset \{m, \cdots, 1\}} (C_p)^{|S|} \|\boldsymbol{F}_S\|^2_{p,\boldsymbol{\rho},s}. \tag{2.44}$$

1. *Low-particle number subspace*

Interestingly, the product-state-weighted p-norms tell us about concentration restricting to *low particle number subspaces*. Consider the projector $\boldsymbol{P}_m$ to m-particle subspace of n-qubit Hilbert space

$$\boldsymbol{P}_m := \sum_{\#(|1\rangle) = m} (|0\rangle \cdots |1\rangle)(\cdot)^\dagger = \sum_{\#(|1\rangle) = m} |0\rangle \langle 0| \cdots |1\rangle \langle 1| \tag{2.45}$$

and the product state

$$\boldsymbol{\rho}_{\frac{m}{n}} = \otimes_i \boldsymbol{\rho}_i = \bigotimes_i \left(\frac{m}{n} |1\rangle \langle 1| + (1 - \frac{m}{n}) |0\rangle \langle 0| \right)_i. \tag{2.46}$$

Then by Stirling's approximation,

$$\bar{\boldsymbol{P}} := \frac{\boldsymbol{P}_m}{\mathrm{Tr}[\boldsymbol{P}_m]} \le \boldsymbol{\rho}_{\frac{m}{n}} \cdot b(n,m) \tag{2.47}$$

where $b(n,m) = (\frac{m}{n})^m(1-\frac{m}{n})^{n-m}\binom{n}{m} = \mathcal{O}(\mathrm{Poly}(n,m))$. This gives the conversion

$$\|\boldsymbol{F}\|_{p,\bar{\boldsymbol{P}},s} \le \|\boldsymbol{F}\|_{p,\boldsymbol{\rho}_{\frac{m}{n}},s} \cdot (\mathrm{Poly}(n,m))^{1/p}\,, \tag{2.48}$$

where the $\mathrm{Poly}(n,m)$ factors are suppressed as long as $p \gtrsim \log(\mathrm{Poly}(n,m))$. The first inequality, proved below, is that weighted norms are monotone w.r.t the state. In our Trotter application, the Hamiltonian is often particle number preserving and the following becomes trivial. But for potential applications in other contexts, we include a quick proof when $\boldsymbol{O}$ and $\boldsymbol{\rho}$ is not commuting.

Fact II.4 (monotonicity of weighted p-norm)**.** *For $\boldsymbol{\rho} \ge \boldsymbol{\sigma} \ge 0$ (presumably not normalized),*

$$\|\boldsymbol{O}\|_{p,\boldsymbol{\rho},s} \ge \|\boldsymbol{O}\|_{p,\boldsymbol{\sigma},s}\,. \tag{2.49}$$

This is closely related to a polynomial version of Lieb's concavity.

Fact II.5 ([33, Theorem 1.1])**.** *For $\boldsymbol{A} \ge 0$, $q \ge 1$, $r \le 1$, the function*

$$f(\boldsymbol{A}) := \mathrm{Tr}[(\boldsymbol{B}^\dagger \boldsymbol{A}^{\frac{1}{q}} \boldsymbol{B})^{rq}] \tag{2.50}$$

is concave (and hence monotone) in $\boldsymbol{A}$.

We can now quickly adapt to our settings to present a proof.

Proof.

$$\begin{aligned}\|\boldsymbol{O}\|_{p,\boldsymbol{\rho}}^p = \mathrm{Tr}\left[\left(\boldsymbol{\rho}^{\frac{s}{2p}}\boldsymbol{O}^\dagger\boldsymbol{\rho}^{\frac{1-s}{p}}\boldsymbol{O}\boldsymbol{\rho}^{\frac{s}{2p}}\right)^{\frac{p}{2}}\right] &\ge \mathrm{Tr}\left[\left(\boldsymbol{\rho}^{\frac{s}{2p}}\boldsymbol{O}^\dagger\boldsymbol{\sigma}^{\frac{1-s}{p}}\boldsymbol{O}\boldsymbol{\rho}^{\frac{s}{2p}}\right)^{\frac{p}{2}}\right] \\ &= \mathrm{Tr}\left[\left(\boldsymbol{\sigma}^{\frac{1-s}{2p}}\boldsymbol{O}\boldsymbol{\rho}^{\frac{s}{p}}\boldsymbol{O}^\dagger\boldsymbol{\sigma}^{\frac{1-s}{2p}}\right)^{\frac{p}{2}}\right] \ge \mathrm{Tr}\left[\left(\boldsymbol{\sigma}^{\frac{1-s}{2p}}\boldsymbol{O}\boldsymbol{\sigma}^{\frac{s}{p}}\boldsymbol{O}^\dagger\boldsymbol{\sigma}^{\frac{1-s}{2p}}\right)^{\frac{p}{2}}\right] = \|\boldsymbol{O}\|_{p,\boldsymbol{\sigma},s}^p\,.\end{aligned} \tag{2.51}$$

Both inequalities use Fact II.5 for $q = \frac{p}{1-s} \ge 1, r = \frac{1-s}{2} \le 1$ and for $q = \frac{p}{s} \ge 1, r = \frac{s}{2} \le 1$. The second equality is $\left\|\boldsymbol{X}^\dagger\boldsymbol{X}\right\|_{\frac{p}{2}} = \left\|\boldsymbol{X}\boldsymbol{X}^\dagger\right\|_{\frac{p}{2}}$. □

D. Fermionic Operators

Uniform smoothness applies to Fermions. Consider the Jordan-Wigner transform

$$\boldsymbol{a}_s := -\prod_{j=1}^{s-1}\boldsymbol{I}_j \cdot \boldsymbol{\sigma}_s^- \cdot \prod_{i=s+1}^{n}(\boldsymbol{\sigma}_i^z) \tag{2.52}$$

$$\boldsymbol{a}_s^\dagger := -\prod_{j=1}^{s-1}\boldsymbol{I}_j \cdot \boldsymbol{\sigma}_s^+ \cdot \prod_{i=s+1}^{n}(\boldsymbol{\sigma}_i^z) \tag{2.53}$$

where $\boldsymbol{\sigma}^- = |0\rangle\langle 1|\,, \boldsymbol{\sigma}^+ = |1\rangle\langle 0|$ are the lowering and raising operators. These operators also linearly span the full algebra on n-qubits $\mathcal{B}(\mathcal{H}(2^n))$ by $\otimes_s(\boldsymbol{a}_s, \boldsymbol{a}_s^\dagger, \boldsymbol{a}_s\boldsymbol{a}_s^\dagger, \boldsymbol{I}_s)$. In this form, Fermions are not local operator due to the Pauli-Z strings. Fortunately, all we need for uniform smoothness is the martingale property (conditionally zero-mean). We derive an analogous 2-norm-like bound with a minor tweak due to Jordan-Wigner strings. The following result was known in [34, Theorem 4][7] but we emphasize our derivation is elementary. We will also extend it in Corollary II.5.2.

Corollary II.5.1 (Uniform smoothness for Fermions)**.** *On n-qubits, consider an operator without any Fermionic square on the same site $\boldsymbol{a}_i\boldsymbol{a}_i^\dagger$. Expand it $\boldsymbol{A} = \sum_{S\subset\{m,\cdots,1\}}\boldsymbol{A}_S$ by subsets S indicated by Fermionic operators $\{\boldsymbol{a}^\dagger, \boldsymbol{a}\}$. Then, for $p \ge 2$, $C_p = p-1$,*

$$\|\boldsymbol{A}\|_p^2 \le \sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\|\boldsymbol{A}_S\|_p^2\,. \tag{2.54}$$

[7] It uses the primitive uniform smoothness (Fact I.4).

Proof. WLG, assume the Fermionic operators are ordered such that the larger index appears on the right (e.g.$\boldsymbol{a}_1\boldsymbol{a}_3\boldsymbol{a}_n$).

$$\|\boldsymbol{A}\|_p^2 = \left\|\boldsymbol{a}_1\boldsymbol{A}_{>1} + \boldsymbol{a}_1^\dagger\boldsymbol{A}'_{>1} + \boldsymbol{I}_1\otimes\boldsymbol{B}_{>1}\right\|_p^2 \le \|\boldsymbol{I}_1\otimes\boldsymbol{B}_{>1}\|_p^2 + C_p\left\|\boldsymbol{a}_1\boldsymbol{A}_{>1} + \boldsymbol{a}_1^\dagger\boldsymbol{A}'_{>1}\right\|_p^2. \tag{2.55}$$

To complete the induction as in II.2.4, apply a gauge transformation to change the Jordan-Winger string such that only a_2 is nontrivial on site 2. Then we can repeat the above inequality. Note that the background $\boldsymbol{\rho}_\eta$ is invariant under gauge transformations, and the Pauli strings of $\boldsymbol{\sigma}^z$ do not blow up the weighted p-norm.

□

Example II.5.1 (2-local Fermionic operators)**.**

$$\left\|\sum_{i<j}\alpha_{ij}\boldsymbol{a}_j\boldsymbol{a}_i\right\|_p^2 \le \sum_{i<j}(C_p)^2\left|\alpha_{ij}\right|^2\|\boldsymbol{a}_j\boldsymbol{a}_i\|_p^2 \le \sum_{i<j}(C_p)^2\left|\alpha_{ij}\right|^2\|\boldsymbol{a}_i\boldsymbol{a}_j\|\,\|\boldsymbol{I}\|_p^2 \tag{2.56}$$

However, when multiplying fermion operators we may get even powers $\boldsymbol{a}_i^\dagger\boldsymbol{a}_i = (\boldsymbol{I}+\boldsymbol{\sigma}_i)/2, \boldsymbol{a}_i\boldsymbol{a}_i^\dagger = (\boldsymbol{I}-\boldsymbol{\sigma}_i)/2$ where the Pauli string $\boldsymbol{\sigma}^z$ cancels. Let us quickly extend to the cases with the presence of $\boldsymbol{\sigma}_i^z$. As a bonus, we also generalize when the background is a product state in the computational basis (which implies concentration for low particle number subspace).

Corollary II.5.2 (uniform smoothness for Fermions and $\boldsymbol{O}^\eta$)**.** *On n qubits, consider a product state diagonal in the computational basis $\boldsymbol{\rho}_\eta = \otimes_i\boldsymbol{\rho}_i = \otimes_i(\eta\left|1\right\rangle\left\langle 1\right| + (1-\eta)\left|0\right\rangle\left\langle 0\right|)_i$. Expand the operator $\boldsymbol{A} = \sum_{S\subset\{m,\cdots,1\}}\boldsymbol{A}_S$ into subsets S indicated by non-trivial operators $(\boldsymbol{a},\boldsymbol{a}^\dagger,\boldsymbol{O}^\eta)$. Then, for $p\ge 2$, $C_p = p-1$,*

$$\|\boldsymbol{A}\|_{p,\boldsymbol{\rho}_\eta}^2 \le \sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\|\boldsymbol{A}_S\|_{p,\boldsymbol{\rho}_\eta}^2. \tag{2.57}$$

The proof is also elementary.

Proof.

$$\|\boldsymbol{A}\|_{p,\boldsymbol{\rho}_\eta}^2 = \left\|\boldsymbol{a}_1\boldsymbol{A}_{>1} + \boldsymbol{a}_1^\dagger\boldsymbol{A}'_{>1} + \boldsymbol{O}_1^\eta\otimes\boldsymbol{C}_{>1} + \boldsymbol{I}_1\otimes\boldsymbol{B}_{>1}\right\|_{p,\boldsymbol{\rho}_\eta}^2 \tag{2.58}$$

$$\le \|\boldsymbol{I}_1\otimes\boldsymbol{B}_{>1}\|_{p,\boldsymbol{\rho}_\eta}^2 + C_p\left\|\boldsymbol{a}_1\boldsymbol{A}_{>1} + \boldsymbol{a}_1^\dagger\boldsymbol{A}'_{>1} + \boldsymbol{O}_1^\eta\otimes\boldsymbol{C}_{>1}\right\|_{p,\boldsymbol{\rho}_\eta}^2. \tag{2.59}$$

The rest gauge transformation argument follows Corollary II.5.1. Note that $\boldsymbol{O}^\eta$ is invariant under gauge transformations.

□

III. NON-RANDOM k-LOCAL HAMILTONIANS

This section applies the framework developed in Section II to show the main result for non-random Hamiltonians.

Theorem III.1 (Trotter error in k-local models)**.** *To simulate a k-local Hamiltonian using ℓ-th order Suzuki formula, the gate complexity*

$$G = \Omega\left(\left(\frac{p^{\frac{k}{2}}\|\boldsymbol{H}\|_{(0),2}\,t}{\epsilon}\right)^{1/\ell}\Gamma p^{\frac{k-1}{2}}\|\boldsymbol{H}\|_{(1),2}\,t\right) \quad \textit{ensures} \quad \left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_\ell(t/r)^r\right\|_{\bar{p}} \le \epsilon. \tag{3.1}$$

The p-norm estimate and Proposition II.0.1 imply concentration for typical input states via Markov's inequality.

Corollary III.1.1. *Draw $\left|\psi\right\rangle$ from a 1-design ensemble (i.e., $\mathbb{E}[\left|\psi\right\rangle\left\langle\psi\right|] = \boldsymbol{I}/\mathrm{Tr}[\boldsymbol{I}]$), then*

$$G = \Omega\left(\left(\frac{\sqrt{\log(1/\delta)}^k\|\boldsymbol{H}\|_{(0),2}\,t}{\epsilon}\right)^{1/\ell}\sqrt{\log(1/\delta)}^{k-1}\Gamma\|\boldsymbol{H}\|_{(1),2}\,t\right) \quad \textit{ensures} \quad \Pr\left(\left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}(t/r)^r\left|\psi\right\rangle\right\|_{\ell_2} \ge \epsilon\right) \le \delta.$$

This quickly converts to the trace distance between the difference of pure states

$$\|\left|\psi_1\right\rangle\left\langle\psi_1\right| - \left|\psi_2\right\rangle\left\langle\psi_2\right|\|_1 \le \left\|\left|\psi_1\right\rangle\big(\left\langle\psi_1\right| - \left\langle\psi_2\right|\big) + \big(\left|\psi_1\right\rangle - \left|\psi_2\right\rangle\big)\left\langle\psi_2\right|\right\|_1 \tag{3.2}$$

$$\le 2\|\left|\psi_1\right\rangle - \left|\psi_2\right\rangle\|_{\ell_2}. \tag{3.3}$$

We sketch the proof at Section III B. In Section III C and Section III D, we do the main calculation and conclude the proof with explicit constants in Section III E. See Section III G for the analogous result for Fermions.

A. Heuristic Argument for First-Order Trotter

Let us play with the first-order Trotter error in leading order expansion

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}_\Gamma t}\cdots \mathrm{e}^{\mathrm{i}\boldsymbol{H}_1} - \mathrm{e}^{\mathrm{i}\sum_{\gamma=1}^{\Gamma}\boldsymbol{H}_\gamma} = \frac{t^2}{2}\sum_{\gamma_2>\gamma_1}[\boldsymbol{H}_{\gamma_2},\boldsymbol{H}_{\gamma_1}] + \mathcal{O}(t^3). \tag{3.4}$$

Let us consider the cases when the supports coincide at one site, where each commutator is $2k-1$-local.

$$\left\|\sum_{\gamma_2>\gamma_1}[\boldsymbol{H}_{\gamma_2},\boldsymbol{H}_{\gamma_1}]\right\|_p^2 \lesssim C_p^{2k-1}\sum_{|S|=2k-1}\left(\sum_{\gamma_2\cup\gamma_2=S}\|[\boldsymbol{H}_{\gamma_2},\boldsymbol{H}_{\gamma_1}]\|\right)^2\|\boldsymbol{I}\|_p^2 \tag{3.5}$$

$$\le C_p^{2k-1}\sum_{|S|=2k-1}\sum_{\gamma_2\cup\gamma_2=S}(\|[\boldsymbol{H}_{\gamma_2},\boldsymbol{H}_{\gamma_1}]\|)^2\binom{2k-1}{k}^2\|\boldsymbol{I}\|_p^2 \tag{3.6}$$

$$\le C_p^{2k-1}\sum_{\gamma_1}\sum_{\gamma_2\cap\gamma_1\neq\emptyset}(\|[\boldsymbol{H}_{\gamma_2},\boldsymbol{H}_{\gamma_1}]\|)^2.\binom{2k-1}{k}^2\|\boldsymbol{I}\|_p^2 \tag{3.7}$$

$$= \mathcal{O}\left(C_p^{2k-1}\|\!|\boldsymbol{H}|\!\|_{(0),2}^2\|\!|\boldsymbol{H}|\!\|_{(1),2}^2\|\boldsymbol{I}\|_p^2\right) \tag{3.8}$$

The second is Cauchy-Schwartz, and $\binom{2k-1}{k}^2$ is a rough estimate for the possible ways that $\gamma_1\cup\gamma_2=S$. However, to complete the proof, we must deal with the $\mathcal{O}(t^3)$ terms and the cases when $\boldsymbol{H}_{\gamma_2}$ and $\boldsymbol{H}_{\gamma_1}$ overlap on multiple sites. This is the content of the following sections.

B. Proof Outline

Our primary tool is the following form of Hypercontractivity we derived (Proposition II.2.4)

$$\|\boldsymbol{F}\|_p^2 \le \sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\|\boldsymbol{F}_S\|_p^2, \tag{3.9}$$

which straightforwardly generalizes to the case of Fermions (Section III G), and is not restricted to the case of qubits. See Section III E for comments on how much constant overhead improvement is possible using the other Hypercontractivity $\|\boldsymbol{F}\|_{\bar{p}}^2 \le \sum_S C_p^{|S|}\|\boldsymbol{F}_S\|_2^2$ (Proposition II.1). Recall the Trotter error can be represented in a time-ordered exponential

$$\prod_{j=1}^{J}\mathrm{e}^{\mathrm{i}a_j\boldsymbol{H}_{\gamma(j)}t} = \mathrm{e}^{\mathrm{i}a_J\boldsymbol{H}_{\gamma(J)}t}\cdots\mathrm{e}^{\mathrm{i}a_1\boldsymbol{H}_{\gamma(1)}t} = \exp_{\mathcal{T}}(\mathrm{i}\int(\boldsymbol{\mathcal{E}}+\boldsymbol{H})dt), \tag{3.10}$$

where the error takes the form of commutator

$$\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{E}}(\boldsymbol{H}_1,\cdots,\boldsymbol{H}_\Gamma,\Upsilon t) := \sum_{j=1}^{J}\left(\prod_{k=j+1}^{J}\mathrm{e}^{a_k\mathcal{L}_{\gamma(k)}t}[a_j\boldsymbol{H}_{\gamma(j)}] - a_j\boldsymbol{H}_{\gamma(j)}\right) \tag{3.11}$$

and $\mathcal{L}_\gamma[O] := \mathrm{i}[\boldsymbol{H}_\gamma, O]$. We will drop the a_j, which is $\mathcal{O}(1)$ for Suzuki formulas, by absorbing them into the Hamiltonian (our arguments do not use any delicate cancellations). We will "beat $\boldsymbol{\mathcal{E}}$ to death": by Taylor expansion, from right to left

Fact III.2 ([8, Theorem 10])**.**

$$\mathrm{e}^{\mathcal{L}_J t}\cdots\mathrm{e}^{\mathcal{L}_{j+1}t} = \sum_{g=0}^{g'}\sum_{g_J+\cdots+g_{j+1}=g}\mathcal{L}_J^{g_J}\cdots\mathcal{L}_2^{g_{j+1}}\frac{t^g}{g_J!\cdots g_{j+1}!}$$
$$+\sum_{m=j+1}^{J}\mathrm{e}^{\mathcal{L}_J t}\cdots\mathrm{e}^{\mathcal{L}_{m+1}t}\int_0^t dt_1\sum_{g_m+\cdots+g_{j+1}=g',g_m\ge 1}e^{\mathcal{L}_m t_1}\mathcal{L}_m^{g_m}\cdots\mathcal{L}_{j+1}^{g_{j+1}}\frac{(t-t_1)^{g_m-1}t^{g'-g_m}}{(g_m-1)!\cdots g_{j+1}!}.$$

We control the p-norm of each g-th order in Section III C, and the edge case g'-th order in Section III D. We combine the estimates and apply Markov's inequality in Section III E.

C. Bounds on the g-th Order

We proceed by controlling each g-th order polynomial (for $\ell < g < g'$). Notice that we are granted the product formula matches $t, \cdots, t^\ell$ orders. By Proposition II.1.1,

$$\|[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_\Gamma, \Upsilon t)]_g\|_p^2 = \left\| \sum_{j=1}^{J} \sum_{g_J+\cdots+g_{j+1}=g-1} \mathcal{L}_J^{g_J} \cdots \mathcal{L}_{j+1}^{g_{j+1}}[\boldsymbol{H}_{j+1}] \frac{t^{g-1}}{g_J! \cdots g_{j+1}!} \right\|_p^2 \tag{3.12}$$

$$\leq \sum_{S \subset \{n,\cdots,1\}} (C_p)^{|S|} \left\| \left[\sum_{j=1}^{J} \sum_{g_J+\cdots+g_{j+1}=g-1} \mathcal{L}_J^{g_J} \cdots \mathcal{L}_{j+1}^{g_{j+1}}[\boldsymbol{H}_{j+1}] \frac{t^{g-1}}{g_J! \cdots g_{j+1}!} \right]_S \right\|_p^2. \tag{3.13}$$

Note that the time only appears at order t^{g-1} because there is an additional integral over t in the time-ordered integral. We should keep in mind that there will be a global uniform bound on the k-locality $|S|_2 + |S|_1 - 1 \leq S_{max}$, while the system size n can be large. Let us ignore the $(C_p)^{|S|}$ for now and gain some intuition by studying

$$\sum_{S \subset \{n,\cdots,1\}} \left\| \left[\sum_{S_2} \mathcal{L}_{S_2} [\sum_{S_1} \boldsymbol{O}_{S_1}] \right]_S \right\|_p^2 \tag{3.14}$$

where $\mathcal{L}_{S_2} = \mathrm{i}[\boldsymbol{H}_{S_2}, \cdot]$, and $\boldsymbol{H}_{S_2}$ are $|S_2|$-local operator on S_2 (similiarly for $\boldsymbol{O}_{S_1}$). The guiding principle is to account for the non-empty overlaps $S_2 \cap S_1 \neq \emptyset$. However, there are various ways S_2 may overlap with S_1 and the simplest case is when they overlap on a single site. This is the formal version of the heuristic argument III A.

Example III.2.1 (Overlapping at 1 site.)**.** What we will see is that the multiplicative growth due to acting $\sum_{S_2} \mathcal{L}_{S_2}$ is controlled by the succinct norm $\|\boldsymbol{H}\|_{(1),2}^2$, multiplied by some function of the locality $|S|$. This is analogous to the story for 1-norm [8].

$$\sum_{S \subset \{n,\cdots,1\}} \left\| \left[\sum_{\substack{|S_1 \cap S_2|=1 \\ S_1 \cup S_2 = S}} \mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}] \right]_S \right\|_p^2 \leq \sum_{S \subset \{n,\cdots,1\}} \left(\sum_{\substack{|S_1 \cap S_2|=1 \\ S_1 \cup S_2 = S}} \|\mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}]\|_p \right)^2 \tag{3.15}$$

$$\leq \sum_{S \subset \{n,\cdots,1\}} \sum_{\substack{|S_1 \cap S_2|=1 \\ S_1 \cup S_2 = S}} \|\mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}]\|_p^2 \Big(\sum_{\substack{|S_1' \cap S_2'|=1 \\ S_1' \cup S_2' = S}} \Big) \tag{3.16}$$

$$\leq \sum_{S_1} \sum_{|S_1 \cap S_2|=1} 4\|\boldsymbol{H}_{S_2}\|^2 \, \|\boldsymbol{O}_{S_1}\|_p^2 \cdot (\binom{|S|}{|S_1|} |S_1|) \tag{3.17}$$

$$\leq 4 \cdot \max(|S_1| \cdot \binom{S}{|S_1|} |S_1|) \cdot \|\boldsymbol{H}\|_{(1),2}^2 \cdot \left(\sum_{S_1} \|\boldsymbol{O}_{S_1}\|_p^2 \right). \tag{3.18}$$

The second inequality is Cauchy-Schwartz. In the third inequality, we rearrange the sum over S_1, S_2, use Holder's inequality, and then evaluate the combinations S_1, S_2 giving rise to S. In the last inequality, we make the 2-norm $\|\boldsymbol{H}\|_{(1),2}$ explicit by

$$\sum_{S_2 : |S_1 \cap S_2|=1} \|\boldsymbol{H}_{S_2}\|^2 = \sum_{s_1 \in S_1} \sum_{S_+ \cap S_1 = \emptyset} \|\boldsymbol{H}_{S_+ \cup \{s_1\}}\|^2 \tag{3.19}$$

$$\leq |S_1| \cdot \max_{s_1} \sum_{s_1 \in S_1} \sum_{S_+ \cap S_1 = \emptyset} \|\boldsymbol{H}_{S_+ \cup \{s_1\}}\|^2 =: |S_1| \cdot \|\boldsymbol{H}\|_{(1),2}^2, \tag{3.20}$$

and use with some uniform upper-bound for function of $|S|, |S_1|$.

Unfortunately, we have to do more work when S_2 overlaps with S_1 at multiple sites, as we may lose all but 1 site. For example, [8]

$$[\boldsymbol{X}_4 \boldsymbol{Z}_3 \boldsymbol{Y}_2, \boldsymbol{Y}_3 \boldsymbol{Y}_2 \boldsymbol{X}_1] = 2\boldsymbol{X}_4 \boldsymbol{X}_3 \boldsymbol{X}_1. \tag{3.21}$$

[8] Since $\mathrm{Tr}_S[\boldsymbol{O}_{AS}, \boldsymbol{O}_{SB}] = 0$, for operators $\boldsymbol{O}_{AS}, \boldsymbol{O}_{SB}$ partially traceless on S; we must have at least 1 Pauli left. This is not the case for Fermions. See III G.

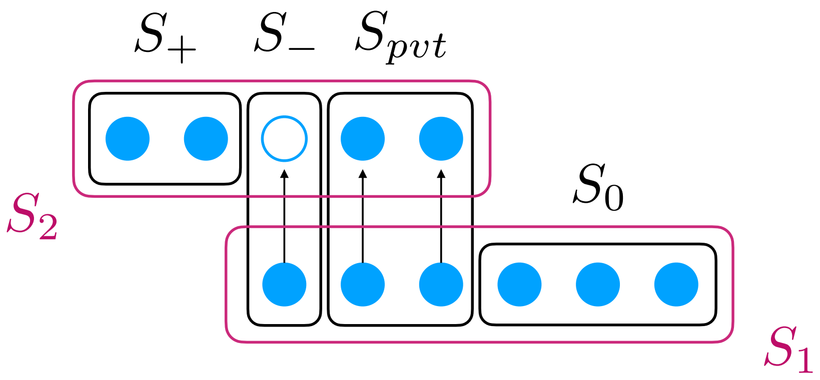

Figure 1. The possible sets that can be produced by $\mathcal{L}_{S_2}$ acting on $\boldsymbol{O}_{S_1}$. In their intersection, some subset S_- becomes identity, some site S_{pvt} remains occupied. For Pauli strings, S_{pvt} must be non-empty; in the fermonic case S_{pvt} may be empty.

To proceed, we have to first identify the ways S_2, S_1 may produce S. Suppose we lose some set in the overlap $S - S_1 =: S_- \subset S_1 \cap S_2$ and gain $S_2/S_1 =: S_+$ (Figure 1)

$$S = S_0 \perp \quad S_{pvt} \perp S_+, \tag{3.22}$$
$$S_2 = \quad S_- \perp S_{pvt} \perp S_+, \tag{3.23}$$
$$S_1 = S_0 \perp S_- \perp S_{pvt}. \tag{3.24}$$

Fixing some integers $|S_{pvt}| = k_{pvt}$, we formally prove the following

Proposition III.2.1 (Effective $2-2$ norm of commutator $\sum_{S_2} \mathcal{L}_{S_2}$)**.** *Fixing* $|S_{pvt}| = k_{pvt}$*,*

$$\sum_{S\subset\{n,\cdots,1\}} \left(\sum_{\substack{|S_{pvt}|=k_{pvt} \\ S=S_0\perp S_{pvt}\perp S_+}} \|[\mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}]]_S\|_p \right)^2 \le 2^{k+k_{pvt}+2} \cdot \left(\frac{(|S|_{max})^k}{(k-k_{pvt})!(k-1)!} \right)^2 \cdot \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{H} \vert\kern-0.25ex\vert\kern-0.25ex\vert^2_{(k_{pvt}),2} \cdot \left(\sum_{S_1} \|\boldsymbol{O}_{S_1}\|_p^2 \right),$$

where

$$\vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{H} \vert\kern-0.25ex\vert\kern-0.25ex\vert_{(c),2} := \sqrt{\max_{|S_c|=c} \sum_{\gamma:S_c\subset \mathrm{Supp}(\boldsymbol{H}_\gamma)} b_\gamma^2}. \tag{3.25}$$

In other words, we are effectively calculating the 2-2 norm of $\sum_{S_2} \mathcal{L}_{S_2}$, where the "2-norm" is $\sum_{S_1} \|[\boldsymbol{O}]_{S_1}\|_p^2$. We will keep this at an analogy level to avoid introducing extra notations.

Proof.

$$\sum_{S\subset\{n,\cdots,1\}} \left(\sum_{\substack{|S_{pvt}|=k_{pvt} \\ S=S_0\perp S_{pvt}\perp S_+}} \|[\mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}]]_S\|_p \right)^2 \le \sum_{S\subset\{n,\cdots,1\}} \left(2^{|S_{pvt}|} \sum_{S_{pvt}S_+} \sum_{S_-} 2\|\boldsymbol{H}_{S_+S_{pvt}S_-}\| \left\|\boldsymbol{O}_{S_{pvt}S_-S_0}\right\|_p \right)^2 \tag{3.26}$$

$$\le (\cdot) \sum_{S\subset\{n,\cdots,1\}} \sum_{S_{pvt}S_+} \left(\sum_{S_-} \|\boldsymbol{H}_{S_+S_{pvt}S_-}\| \left\|\boldsymbol{O}_{S_{pvt}S_-S_0}\right\|_p \right)^2 \Big(\sum_{S'_0\perp S'_{pvt}\perp S'_+=S} 1\Big) \tag{3.27}$$

$$\le (\cdot) \sum_{S\subset\{n,\cdots,1\}} \sum_{S_{pvt}S_+} \left(\sum_{S_-} \|\boldsymbol{H}_{S_+S_{pvt}S_-}\|^2 \right) \left(\sum_{S'_-} \left\|\boldsymbol{O}_{S_{pvt}S'_-S_0}\right\|_p^2 \right) (\cdot) \tag{3.28}$$

$$= (\cdot\cdot) \sum_{S_{pvt}} \left(\sum_{S_-,S_+} \|\boldsymbol{H}_{S_+S_{pvt}S_-}\|^2 \right) \left(\sum_{S'_-,S_0} \left\|\boldsymbol{O}_{S_{pvt}S'_-S_0}\right\|_p^2 \right) \tag{3.29}$$

$$\le 2^{k+k_{pvt}+2} \cdot \left(\frac{(|S|_{max})^k}{(k-k_{pvt})!(k-1)!} \right)^2 \cdot \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{H} \vert\kern-0.25ex\vert\kern-0.25ex\vert^2_{(k_{pvt}),2} \cdot \left(\sum_{S_1} \|\boldsymbol{O}_{S_1}\|_p^2 \right) \tag{3.30}$$

The first inequality removes $[\cdot]_S$, the projection onto operators partially traceless on S, via the following. We first observe $[\mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}]]_S = [\mathcal{L}_{S_2}[\boldsymbol{O}_{S_1}]]_{S_{pvt}}$ to reduce constant overheads.

Fact III.3. *For any set* S,$\left\|[\boldsymbol{O}]_S\right\|_p \le 2^{|S|}\left\|\boldsymbol{O}\right\|_p$.

Proof.

$$\left\|[\boldsymbol{O}]_S\right\|_p = \left\| \prod_{s'\in S^c} (\boldsymbol{I}_{s'}\frac{\mathrm{Tr}_{s'}[\cdot]}{\mathrm{Tr}[\boldsymbol{I}_{s'}]}) \prod_{s\in S}(1-\boldsymbol{I}_s\frac{\mathrm{Tr}_s[\cdot]}{\mathrm{Tr}[\boldsymbol{I}_s]})[\boldsymbol{O}]\right\|_p \tag{3.31}$$

$$\le \left\| \prod_{s\in S}\left(1-\boldsymbol{I}_s\frac{\mathrm{Tr}_s[\cdot]}{\mathrm{Tr}[\boldsymbol{I}_s]}\right)[\boldsymbol{O}]\right\|_p \tag{3.32}$$

$$\le 2^{|S|}\left\|\boldsymbol{O}\right\|_p. \tag{3.33}$$

Throughout the inequalities we used Proposition II.2.1, i.e., $\boldsymbol{I}_{s'}\mathrm{Tr}_{s'}[\cdot]/\mathrm{Tr}_{s'}[\boldsymbol{I}_{s'}]$ is norm non-increasing. The last $2^{|S|}$ is due to a brutal triangle inequality. ☐

The second and third inequalities are Cauchy-Schwartz, w.r.t to the sum over S_{pvt}, S_+, S_0 to associated with a given S, and w.r.t the sum over S_-. We also evaluate the elementary sum

$$\sum_{S'_0\perp S'_{pvt}\perp S'_+=S} 1 = \sum_{|S_+|=0}^{k-|S_{pvt}|}\binom{S}{S_{pvt}}\binom{S-S_{pvt}}{S_+} = \sum_{|S_+|=0}^{k-|S_{pvt}|}\frac{|S|!}{|S_{pvt}|!\,|S_+|!(|S|-|S_+|-|S_{pvt}|)!} \tag{3.34}$$

$$\le \sum_{|S_+|=0}^{k-|S_{pvt}|}\frac{|S|^{|S_+|}}{|S_+|!}\frac{|S|^{|S_{pvt}|}}{|S_{pvt}|!} \tag{3.35}$$

$$\le (k-k_{pvt}+1)\frac{|S|^k}{(k-k_{pvt})!k!}. \tag{3.36}$$

The equality in the fourth line is a rearrangement, and in the last inequality, we do over-counting estimates

$$\max_{S_{pvt}}\left(\sum_{S_-,S_+}\left\|\boldsymbol{H}_{S_+S_{pvt}S_-}\right\|^2\right) = \max_{S_{pvt}}\left(\sum_{S'}\sum_{S_-\cup S_+=S'}\left\|\boldsymbol{H}_{S_+S_{pvt}S_-}\right\|^2\right) \tag{3.37}$$

$$\le 2^{k-k_{pvt}}\cdot \left\|\!\left|\boldsymbol{H}\right|\!\right\|^2_{(k_{pvt}),2}, \tag{3.38}$$

$$\sum_{S_{pvt},S'_-,S_0}\left\|\boldsymbol{O}_{S_{pvt}S'_-S_0}\right\|_p^2 \le \sum_{S_1}\sum_{S_{pvt}\perp S'_-\perp S_0=S_1}\left\|\boldsymbol{O}_{S_1}\right\|_p^2 \le \sum_{|S'_-|=0}^{k-k_{pvt}}\binom{|S_1|}{S_{pvt}}\binom{|S_1|-S_{pvt}}{S'_-}\cdot\sum_{S_1}\left\|\boldsymbol{O}_{S_1}\right\|_p^2 \tag{3.39}$$

$$\le (k-k_{pvt}+1)\frac{(|S|_{max})^k}{(k-k_{pvt})!k!}\cdot\sum_{S_1}\left\|\boldsymbol{O}_{S_1}\right\|_p^2. \tag{3.40}$$

These, together with constants hidden in $(\cdot\cdot)$, give the ultimate prefactors. ☐

What we will quote in the Trotter calculation is the following refinement (with an additional innermost sum $\sum_\alpha$ and a sum over values of k_{pvt}).

Corollary III.3.1.

$$\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_2,S_1}\sum_{\alpha}\left\|\left[\mathcal{L}_{S_2}[\boldsymbol{O}^\alpha_{S_1}]\right]_S\right\|_p\right)^2 \le \lambda(k)^2(|S|_{max})^{2k}\sum_{S_1\subset\{n,\cdots,1\}}\left(\sum_\alpha\left\|[\boldsymbol{O}^\alpha]_{S_1}\right\|_p\right)^2 \tag{3.41}$$

where

$$\lambda(k) := \frac{2^{k/2+1}}{(k-1)!}\sum_{k_{pvt}=1}^{k}\frac{2^{k_{pvt}/2}}{(k-k_{pvt})!}\left\|\!\left|\boldsymbol{H}\right|\!\right\|_{(k_{pvt}),2}. \tag{3.42}$$

Proof.

$$\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_2,S_1}\sum_{\alpha}\left\|[\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_1}]]_S\right\|_p\right)^2=\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_2,S_1}\mathbb{1}(S_2,S_1\to S)\sum_{\alpha}\left\|[\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_1}]]_S\right\|_p\right)^2 \tag{3.43}$$

$$=\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{|S_{pvt}|=1}^{k}\sum_{S_{pvt}\perp S_+\perp S_0=S}\sum_{S_-\notin S}\sum_{\alpha}\left\|\left[\mathcal{L}_{S_+S_{pvt}S_-}[\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}]\right]_S\right\|_p\right)^2 \tag{3.44}$$

$$\le\left(\sum_{|S_{pvt}|=1}^{k}\sqrt{\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_{pvt}\perp S_+\perp S_0=S}\sum_{S_-\notin S}\sum_{\alpha}\left\|\left[\mathcal{L}_{S_+S_{pvt}S_-}[\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}]\right]_S\right\|_p\right)^2}\right)^2 \tag{3.45}$$

$$\le\left(\sum_{|S_{pvt}|=1}^{k}\sqrt{2}^{k+k_{pvt}+2}\frac{(|S|_{max})^k}{(k-k_{pvt})!(k-1)!}\cdot\vert\!\vert\!\vert\boldsymbol{H}\vert\!\vert\!\vert_{(k_{pvt}),2}\right)^2\sum_{S_1\subset\{n,\cdots,1\}}\left(\sum_{\alpha}\|[\boldsymbol{O}^{\alpha}]_{S_1}\|_p\right)^2. \tag{3.46}$$

The second equality is presenting S_1,S_2 by S_{pvt},S_-,S_+,S_0. The last inequality might look intimidating, but is nothing more than triangle inequality (over values $|S_{pvt}|$) for the 2-norm

$$|\sum_k f(k,S)|_2:=\sqrt{\sum_S(\sum_k f(k,S))^2}\le\sum_k\sqrt{\sum_S f(k,S)^2}=\sum_k|f(k,S)|_2. \tag{3.47}$$

We skip the steps towards the last inequality as in (3.26). Note that the sum $\sum_\alpha$ stays at the innermost sum, sticking to $\boldsymbol{O}^\alpha$, i.e., with the replacement

$$\left\|\boldsymbol{O}_{S_{pvt}S_-S_0}\right\|_p\to\left(\sum_{\alpha}\left\|\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}\right\|_p\right), \tag{3.48}$$

which completes the proof. □

Finally, back to our Trotter error,

$$\|[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1,\cdots,\boldsymbol{H}_\Gamma,\Upsilon t)]_g\|_p^2\le\sum_{S\subset\{n,\cdots,1\}}(C_p)^{|S|}\left\|\left[\sum_{j=1}^{J}\sum_{g_J+\cdots+g_{j+1}=g-1}\mathcal{L}_J^{g_J}\cdots\mathcal{L}_{j+1}^{g_{j+1}}[\boldsymbol{H}_{j+1}]\frac{t^{g-1}}{g_J!\cdots g_{j+1}!}\right]_S\right\|_p^2 \tag{3.49}$$

$$\le(C_p)^{g(k-1)+1}\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{\gamma_{g-1}}^{\Gamma}\sum_{S_1}\cdots\sum_{\gamma_0}^{\Gamma}\left\|\left[\mathcal{L}_{\gamma_{g-1}}\left[\mathcal{L}_{\gamma_{g-2}}\cdots\mathcal{L}_{\gamma_1}[\boldsymbol{H}_{\gamma_0}]\right]_{S_1}\frac{(t\Upsilon)^{g-1}}{(g-1)!}\right]_S\right\|_p\right)^2 \tag{3.50}$$

$$\le\frac{(\Upsilon t)^{2(g-1)}}{(g-1)!^2}(C_p)^{g(k-1)+1}\cdot(g(k-1)+1)^{2k}\lambda(k)^2\cdot\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{\gamma_{g-2}}^{\Gamma}\cdots\sum_{\gamma_0}^{\Gamma}\left\|[\mathcal{L}_{\gamma_{g-2}}\cdots\mathcal{L}_{\gamma_1}[\boldsymbol{H}_{\gamma_0}]]_S\right\|_p\right)^2 \tag{3.51}$$

$$\le\frac{(\Upsilon t)^{2(g-1)}}{(g-1)!^2}(C_p)^{g(k-1)+1}\Big((g(k-1)+1)\cdots(2(k-1)-1)\Big)^{2k}\lambda(k)^{2g}\cdot\|\boldsymbol{H}\|^2_{(0),2}\|\boldsymbol{I}\|_p^2 \tag{3.52}$$

$$\le g^2\cdot(C_p)^{g(k-1)+1}g^{2g(k-1)}\left(c(k)\Upsilon t\right)^{2(g-1)}\cdot\|\boldsymbol{H}\|^2_{(0),2}\|\boldsymbol{I}\|_p^2. \tag{3.53}$$

In the second inequality, we throw in terms to "complete the exponential" and use a uniform bound on locality $|S|\le g(k-1)+1$; the third is Corollary III.3.1 ; the fourth calls Corollary III.3.1 a few more times and uses

$$\sum_{S\subset\{n,\cdots,1\}}\left\|[\boldsymbol{H}_{\gamma_0}]_S\right\|_p^2\le\|\boldsymbol{H}\|_{(0),2}\|\boldsymbol{I}\|_p^2. \tag{3.54}$$

Lastly, we hide constants depending only on k in c_k. Unfortunately, the series is not summable over g due to the factor $g^{g(k-1)}$; therefore, we would stop the expansion at some properly chosen order g'. This is the content of the following section.

D. Bounds for g'-th Order and Beyond.

The higher-order terms in the Trotter error have infinite-order dependence on time, and we have to tweak the calculations.

$$
\begin{aligned}
&\left\|[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1,\cdots,\boldsymbol{H}_\Gamma,\Upsilon t)]_{\geq g}\right\|_p \\
&= \left\| \sum_{j=1}^{J}\sum_{m=j+1}^{J} \mathrm{e}^{\mathcal{L}_J t}\cdots \mathrm{e}^{\mathcal{L}_{m+1}t}\int_0^t dt_1 \sum_{g_m+\cdots+g_{j+1}=g-1,g_m\geq 1} e^{\mathcal{L}_m t_1}\mathcal{L}_m^{g_m}\cdots\mathcal{L}_{j+1}^{g_{j+1}}[\boldsymbol{H}_j]\frac{(t-t_1)^{g_m-1}t^{g'-g_m}}{(g_m-1)!\cdots g_{j+1}!}\right\|_p &(3.55)\\
&\leq \sum_{m=2}^{J}\left\|\sum_{j=m-1}^{J}\sum_{g_m+\cdots+g_{j+1}=g-1,g_m\geq 1}\mathcal{L}_m^{g_m}\cdots\mathcal{L}_{j+1}^{g_{j+1}}[\boldsymbol{H}_j]\frac{t^{g-1}}{g_m!\cdots g_{j+1}!}\right\|_p &(3.56)\\
&\leq \sqrt{C_p}^{g(k-1)+1}\sum_{\gamma_{g-1}}^{\Gamma}\sqrt{\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_1}\sum_{\gamma_{g-2}}^{\Gamma}\cdots\sum_{\gamma_0}^{\Gamma}\left\|\left[\mathcal{L}_{\gamma_{g-1}}\left[\mathcal{L}_{\gamma_{g-2}}\cdots\mathcal{L}_{\gamma_1}[\boldsymbol{H}_{\gamma_0}]\right]_{S_1}\frac{(t\Upsilon)^{g-1}}{(g-2)!}\right]_S\right\|_p\right)^2} &(3.57)\\
&\leq \frac{(\Upsilon t)^{g-1}}{(g-2)!}\sqrt{C_p}^{g(k-1)+1}(g(k-1)+1)^{k/2}\lambda'(k)\cdot\Big(((g-1)(k-1)+1)\cdots(2(k-1)-1)\Big)^k\lambda(k)^{g-2}\cdot\left\|\boldsymbol{H}\right\|_{(0),2}\left\|\boldsymbol{I}\right\|_p\\
&\leq g\cdot\sqrt{C_p}^{g(k-1)+1}c'(k)\cdot g^{g(k-1)}\left(c(k)\Upsilon t\right)^{g-1}\left\|\boldsymbol{I}\right\|_p &(3.58)
\end{aligned}
$$

The first inequality exchanges the summation order and removes the unitary by invariance of p-norm, with a price of the triangle inequality. The second inequality is again completing the exponential for $\gamma_{g-2}\cdots\gamma_0$. In the third inequality, we apply Corollary III.3.2 for the outer-most sum $\sum_{\gamma_{g-1}}^{\Gamma}$, Corollary III.3.1 for $\gamma_{g-2},\cdots\gamma_1$, and rearrange the expression. Lastly we absorb constants into $c'(k), c(k)$

$$
\begin{aligned}
c'(k) &= \frac{\lambda'(k)}{\lambda(k)}\frac{1}{\sqrt{k}^k}\left\|\boldsymbol{H}\right\|_{(0),2} &(3.59)\\
c(k) &= 2\lambda(k), &(3.60)
\end{aligned}
$$

and the 2 in $c(k)$ comes from crude estimates $g\leq 2^{g-1}$. (Also $g(k-1)+1\leq gk$, $g!\leq g^g$)[9]

Corollary III.3.2.

$$
\begin{aligned}
\sum_{S_2}\sqrt{\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_1}\sum_{\alpha}\left\|\left[\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_1}]\right]_S\right\|_p\right)^2} &\leq 2\cdot\sum_{k'=1}^{k}\binom{k}{k'}\sqrt{20}^{k'}\sqrt{\binom{|S_{max}|}{k'}}\left\|\boldsymbol{H}\right\|_{(k'),1}\left\|\boldsymbol{H}\right\|_{(0),1}\\
&\quad\cdot\sqrt{\sum_{S_1\subset\{n,\cdots,1\}}\left(\sum_{\alpha}\left\|[\boldsymbol{O}^{\alpha}]_{S_1}\right\|_p\right)^2}. &(3.61)\\
&\leq |S_{max}|^{k/2}\cdot\lambda'(k)\cdot\sqrt{\sum_{S_1\subset\{n,\cdots,1\}}\left(\sum_{\alpha}\left\|[\boldsymbol{O}^{\alpha}]_{S_1}\right\|_p\right)^2}. &(3.62)
\end{aligned}
$$

where $\lambda'(k) = 2\cdot\sum_{k'=1}^{k}\binom{k}{k'}\sqrt{20}^{k'}\sqrt{\frac{1}{k'!}}\left\|\boldsymbol{H}\right\|_{(k'),1}\left\|\boldsymbol{H}\right\|_{(0),1}$.

---

[9] e^g is wasted here!

Proof. This is the tweak we needed to take care of the outermost sum, which has a sum outside of square root.

$$\sum_{S_2}\sqrt{\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_1}\sum_{\alpha}\left\|[\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_1}]]_S\right\|_p\right)^2} \le \sum_{S_2}\sqrt{\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{|S_-|+|S_{pvt}|=1}^{k}\sum_{S_-,S_{pvt}\subset S_2}2^{S_{pvt}}\sum_{\alpha}\left\|\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}]\right\|_p\right)^2} \tag{3.63}$$

$$\le \sum_{|S_-|+|S_{pvt}|=1}^{k}\sum_{S_2}\sqrt{\sum_{S\subset\{n,\cdots,1\}}\left(\sum_{S_-,S_{pvt}\subset S_2}2^{S_{pvt}}\sum_{\alpha}\left\|\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}]\right\|_p\right)^2} \tag{3.64}$$

$$\le \sum_{|S_-|+|S_{pvt}|=1}^{k}\sum_{S_2}\sqrt{\sum_{S_0}\sum_{S_-,S_{pvt}\subset S_2}\left(\sum_{\alpha}\left\|\mathcal{L}_{S_2}[\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}]\right\|_p\right)^2(\sum_{S'_{pvt},S'_-}2^{2S'_{pvt}})} \tag{3.65}$$

$$\le \sum_{|S_-|+|S_{pvt}|=1}^{k}\sum_{S_+,S_-,S_{pvt}}2\|\boldsymbol{H}_{S_+S_-S_{pvt}}\|\sqrt{\sum_{S_0}\left(\sum_{\alpha}\left\|\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}\right\|_p\right)^2(\cdot)} \tag{3.66}$$

$$\le \sum_{|S_-|+|S_{pvt}|=1}^{k}2\sqrt{\sum_{S_-,S_{pvt}}\left(\sum_{S_+}\|\boldsymbol{H}_{S_+S_-S_{pvt}}\|\right)^2}\sqrt{\sum_{S_-,S_{pvt},S_0}\left(\sum_{\alpha}\left\|\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}\right\|_p\right)^2(\cdot)} \tag{3.67}$$

The first inequality parameterizes the $S_1 = S_{pvt}S_-S_0$ that could give rise to S after $\mathcal{L}_{S_2}$. The $2^{|S_{pvt}|}$ is due to Fact III.3. The second inequality is a triangle inequality to postpone the sum over $|S_{pvt}|+|S_-|$. The third inequality is Cauchy-Schwartz, where the combinatorial sum evaluates to

$$(\cdot) = \sum_{S'_{pvt},S'_-}2^{2S'_{pvt}} = \binom{k}{k'}\cdot\sum_{k_{pvt}=1}^{k'}\binom{k'}{k_{pvt}}2^{2S'_{pvt}} = \binom{k}{k'}5^{k'} \tag{3.68}$$

for $|S_{pvt}|+|S_-| = k'$. The fourth inequality is a triangle inequality for the sum over $S_-, S_{pvt}\subset S_2$, which then combines with the sum over S_2. The fifth inequality is Cauchy-Schwartz's. Lastly, we evaluate the combinatorial factors for each term

$$\sum_{S_-,S_{pvt}}\left(\sum_{S_+}\|\boldsymbol{H}_{S_+S_-S_{pvt}}\|\right)^2 \le \sum_{S_-,S_{pvt}}\sum_{S_+}\|\boldsymbol{H}_{S_+S_-S_{pvt}}\|\cdot\max_{S_-,S_{pvt}}\sum_{S_+}\|\boldsymbol{H}_{S_+S_-S_{pvt}}\| \tag{3.69}$$

$$= \binom{k}{k'}2^{k'}\,\|\boldsymbol{H}\|_{(0),1}\cdot\|\!|\boldsymbol{H}|\!\|_{(k'),1}, \tag{3.70}$$

and

$$\sum_{S_-,S_{pvt},S_0}\left(\sum_{\alpha}\left\|\boldsymbol{O}^{\alpha}_{S_{pvt}S_-S_0}\right\|_p\right)^2 = \binom{|S_{max}|}{k'}2^{k'}\cdot\sum_{S}\left(\sum_{\alpha}\|\boldsymbol{O}^{\alpha}_S\|_p\right)^2 \tag{3.71}$$

to conclude the proof.

□

E. Proof of Theorem III.1

Proof. For a short time $\tau = \Upsilon t$, we arrange and perform the last integral using estimate $\int(\tau')^{g-1}d\tau' \le \tau'^g/g$

$$\frac{\left\|e^{\mathrm{i}\boldsymbol{H}\tau}-\boldsymbol{S}_\ell(\tau)\right\|_p}{\|\boldsymbol{I}\|_p} \le \int_0^\tau \frac{\|\boldsymbol{\mathcal{E}}(\tau')\|_p}{\|\boldsymbol{I}\|_p}d\tau' \tag{3.72}$$

$$\le \frac{\sqrt{C_p}}{c(k)}\|\boldsymbol{H}\|_{(0),2}\cdot\sum_{g=\ell+1}^{g'-1}\left(g^{k-1}\sqrt{C_p}^{k-1}c(k)\tau\right)^g+\sqrt{C_p}\frac{c'(k)}{c(k)}\cdot\left(g'^{k-1}\sqrt{C_p}^{k-1}c(k)\tau\right)^{g'} \tag{3.73}$$

$$:= c'_{1,p}\sum_{g=\ell+1}^{g'-1}\left(g^{k-1}b_p\tau\right)^g+c'_{2,p}\left(g'^{k-1}b_p\tau\right)^{g'} \tag{3.74}$$

$$\le \frac{c'_{1,p}}{1-1/\mathrm{e}}\left((\ell+1)^{(k-1)}b_p\tau\right)^{\ell+1}+c'_{2,p}\exp\left(-\frac{1}{\mathrm{e}(b_p\tau)^{1/(k-1)}}+1\right) \tag{3.75}$$

$$:= c_{1,p}(b_p\tau)^{\ell+1}+c_{2,p}\exp\left(-\frac{1}{\mathrm{e}(b_p\tau)^{1/(k-1)}}\right). \tag{3.76}$$

In the second inequality we call the bounds for each $g-th$ order (3.53) and the g'-th order (3.58) for a good value of $g' = \left\lfloor\frac{1}{\mathrm{e}(b_p\tau)^{1/(k-1)}}\right\rfloor$. This is possible as long as

Constraint III.3.1. $(\frac{1}{b_p\tau})^{1/(k-1)} \ge \mathrm{e}(\ell+3)$.

Then, the total Trotter error at a long time $t = r\cdot\tau$ is bounded by a telescoping sum

$$\frac{\|\boldsymbol{\mathcal{E}}_{tot}\|_p}{\|\boldsymbol{I}\|_p} := \frac{\left\|e^{\mathrm{i}\boldsymbol{H}t}-\boldsymbol{S}_\ell(t/r)^r\right\|_p}{\|\boldsymbol{I}\|_p} \le r\cdot\frac{\left\|e^{\mathrm{i}\boldsymbol{H}t/r}-\boldsymbol{S}_\ell(t/r)\right\|_p}{\|\boldsymbol{I}\|_p} \le c_{1,p}\frac{(b_pt)^{\ell+1}}{r^\ell}+rc_{2,p}\exp\left(-\frac{1}{\mathrm{e}}(\frac{r}{b_pt})^{1/(k-1)}\right) \tag{3.77}$$

$$\le 2c_{1,p}\frac{(b_pt)^{\ell+1}}{r^\ell} \le p^\eta 2c_1\frac{(bt)^{\ell+1}}{r^\ell}. \tag{3.78}$$

At the second line we restrict to sufficiently large values of r that the first term dominates.[10]

Constraint III.3.2. $(\frac{1}{b_p\tau})^{1/(k-1)} \ge \mathrm{e}\ln\left(\frac{c_2}{c_1}(\frac{1}{b_p\tau})^{\ell+1}\right)$.

The last inequality isolates the p-dependence and we set $\eta := \frac{(\ell+1)(k-1)+1}{2}$ and use $C_p = p-1 \le p$. For each gate count r, pick p_r that

$$\frac{\epsilon}{\mathrm{e}^\eta} = p_r^\eta\cdot 2c_1\frac{(bt)^{\ell+1}}{r^\ell} \tag{3.79}$$

Via Markov's inequality, this gives concentration for its singular values(or over any 1-design inputs)

$$\hat{\Pr}(\nu(\boldsymbol{\mathcal{E}}_{tot})\ge\epsilon) \le \frac{\|\boldsymbol{\mathcal{E}}_{tot}\|_p^p}{\epsilon^p} \le \exp\left(-\frac{\eta}{\mathrm{e}}(\frac{\epsilon r^\ell}{2c'_1(bt)^{\ell+1}})^{1/\eta}\right) = \delta. \tag{3.80}$$

Choose r such that $p_r = \left(\frac{\epsilon r^\ell}{2c_1(bt)^{\ell+1}}\right)^{1/\eta}\frac{1}{\mathrm{e}} \ge \max(2,\log(1/\delta)/\eta)$, which is

$$r \ge \left(\frac{2\sqrt{\mathrm{e}\log(1/\delta)/\eta}}{\mathrm{e}-1}\left((\ell+1)\sqrt{\mathrm{e}\log(1/\delta)/\eta}\right)^{(\ell+1)(k-1)}\cdot\frac{\|\boldsymbol{H}\|_{(0),2}\,t}{\epsilon}\right)^{1/\ell}2\lambda(k)t. \tag{3.81}$$

We also need to comply with both Constraint III.3.1 and Constraint III.3.2 by

$$r \ge a^{2\eta/k}\left(2\lambda(k)t^{1/k}\right)\left(\frac{1-1/\mathrm{e}}{2\mathrm{e}^\eta(\ell+1)^{(\ell+1)(k-1)}}\frac{\epsilon}{\|\boldsymbol{H}\|_{(0),2}}\right)^{\frac{k-1}{k}}, \tag{3.82}$$

[10] In obtaining Constraint VIII.1.2, note that factors of p cancels out $c_{2,p}/c_{1,p} = c_2/c_1$.

where

$$a = \max\left[(\mathrm{e}(\ell+3))^{k-1}, 2\left(\mathrm{e}\ln\left(\frac{\mathrm{e}-1}{\sqrt{k}^k(\ell+1)^{(\ell+1)(k-1)}} \frac{\lambda'(k)}{\lambda(k)} \right)\right)^{k-1}, x \right] \tag{3.83}$$

and x is the unique solution to the transcendental equation

$$x = 2(\mathrm{e}(\ell+1))^{k-1} \cdot \ln^{k-1}(x). \tag{3.84}$$

And recall

$$\lambda'(k) = 2 \cdot \sum_{k'=1}^{k} \binom{k}{k'} \sqrt{20}^{k'} \sqrt{\frac{1}{k'!} \left\lVert \boldsymbol{H} \right\rVert_{(k'),1} \left\lVert \boldsymbol{H} \right\rVert_{(0),1}} \tag{3.85}$$

$$\lambda(k) = \frac{2^{k/2+1}}{(k-1)!} \sum_{k'=1}^{k} \frac{2^{k'/2}}{(k-k')!} \left\lVert \boldsymbol{H} \right\rVert_{(k'),2}. \tag{3.86}$$

The above expressions for gate count are for numerical evaluation; for comprehension, use $\Omega(\cdot)$ to suppress functions of k, ℓ and note the local norms are decreasing with k' that $k_1 \le k_2 \implies \left\lVert \boldsymbol{H} \right\rVert_{(k_1),1} \ge \left\lVert \boldsymbol{H} \right\rVert_{(k_2),1}, \left\lVert \boldsymbol{H} \right\rVert_{(k_1),2} \ge \left\lVert \boldsymbol{H} \right\rVert_{(k_2),2}$.

$$\begin{aligned} r = \Omega\Bigg[& \ln(\delta)^{\eta/\ell} \left(\frac{\left\lVert \boldsymbol{H} \right\rVert_{(0),2}}{\epsilon \left\lVert \boldsymbol{H} \right\rVert_{(1),2}} \right)^{\frac{1}{\ell}} \left(\left\lVert \boldsymbol{H} \right\rVert_{(1),2} t \right)^{1+\frac{1}{\ell}} \\ & + (\left\lVert \boldsymbol{H} \right\rVert_{(1),2} t)^{\frac{1}{k}} \left(\frac{\epsilon \left\lVert \boldsymbol{H} \right\rVert_{(1),2}}{\left\lVert \boldsymbol{H} \right\rVert_{(0),2}} \right)^{\frac{k-1}{k}} \left(\ln\left(\frac{\sqrt{\left\lVert \boldsymbol{H} \right\rVert_{(1),1} \left\lVert \boldsymbol{H} \right\rVert_{(0),1}}}{\left\lVert \boldsymbol{H} \right\rVert_{(1),2}} \right) \right)^{2\eta\frac{k-1}{k}} \Bigg] \end{aligned} \tag{3.87}$$

$$= \Omega\left[\ln(\delta)^{\frac{k-1}{2}} \left\lVert \boldsymbol{H} \right\rVert_{(1),2} t \cdot \left(\ln(\delta)^{k/2} \frac{\left\lVert \boldsymbol{H} \right\rVert_{(0),2} t}{\epsilon} \right)^{\frac{1}{\ell}} \right]. \tag{3.88}$$

This is the advertised result. □

1. *Constant overhead improvement from another Hypercontractivity*

One may consider directly the existing Hypercontractivity $\left\lVert \boldsymbol{F} \right\rVert_{\bar{p}}^2 \le \sum_S C_p^{|S|} \left\lVert \boldsymbol{F}_S \right\rVert_2^2$(Proposition II.1). However, one needs to go through the same combinatorial estimates, with minor constant overheads improvements by replacing $\left\lVert \boldsymbol{O}_{S_1} \right\rVert_p^2 \to \left\lVert \boldsymbol{O}_{S_1} \right\rVert_2^2$ and discarding Fact III.3. Unfortunately, what comes into the ultimate quantity $\left\lVert \boldsymbol{H} \right\rVert_{(0),2}$ is the spectral norm $\left\lVert \boldsymbol{H}_\gamma \right\rVert$ coming from a Holder's inequality

$$\left\lVert \mathcal{L}_\gamma[\boldsymbol{O}_{S_1}] \right\rVert_p \le 2 \left\lVert \boldsymbol{H}_\gamma \right\rVert \left\lVert \boldsymbol{O}_{S_1} \right\rVert_p, \tag{3.89}$$

and it requires more accounting to get better estimates for the commutator.

F. Spin Models at a Low Particle Number

Consider simulating a Hamiltonian where each term $\boldsymbol{H}_\gamma$ preserves the particle number[11] with an input state from the m-particle subspace (in the $\boldsymbol{Z}$ basis). Formally, denote the m-particle subspace by the orthogonal projector

$$\boldsymbol{P}_m := \sum_{\#(|1\rangle)=m} \left(|0\rangle \cdots |1\rangle \right) (\cdot)^\dagger = \sum_{\#(|1\rangle)=m} |0\rangle\langle 0| \cdots |1\rangle\langle 1|, \tag{3.90}$$

then for all m', γ, particle number preserving means

$$[\boldsymbol{P}_{m'}, \boldsymbol{H}_\gamma] = 0. \tag{3.91}$$

[11] Or perhaps those Hamiltonians that grow particle very slowly.

Can we employ similar concentration arguments, restricting to input states drawn randomly from the m-particle subspace $\boldsymbol{P}_m/\mathrm{Tr}[\boldsymbol{P}_m]$? The answer is affirmative. We need to first define the appropriate k-locality in this case by expanding the Hamiltonian in the basis

$$\boldsymbol{H} = \sum_{S_-\perp S_+\perp S_z \subset\{m,\cdots,1\}} b_{S_+S_-S_z} \prod_{s_+\in S_+} \boldsymbol{Z}^+_{s_+} \prod_{s_-\in S_-} \boldsymbol{Z}^-_{s_-} \prod_{s_z\in S_z} \boldsymbol{O}^\eta_{s_z}, \tag{3.92}$$

where $\boldsymbol{O}^\eta := (1-\eta)\left|1\right\rangle\left\langle 1\right| - \eta\left|0\right\rangle\left\langle 0\right|$ is the analog of Pauli $\boldsymbol{Z}$ in a biased background $\eta\left|1\right\rangle\left\langle 1\right| + (1-\eta)\left|0\right\rangle\left\langle 0\right|$, and $\boldsymbol{Z}^+ := \left|1\right\rangle\left\langle 0\right|$, $\boldsymbol{Z}^- := \left|0\right\rangle\left\langle 1\right|$ are the raising and lower operators. k-locality can be defined in this basis by $|S_-|+|S_+|+|S_z| = k$, and note that particle number preserving enforces the number of raising and lower operators match $|S_+| = |S_-|$. This expansion is motivated by an auxiliary product state $\boldsymbol{\rho}_{\frac{m}{n}}$ and see Section II C for the details.

Proposition III.3.1 (Trotter error in k-local models)**.** *To simulate a number preserving k-local Hamiltonian using the ℓ-th order Suzuki formula on the m-particle subspace $\boldsymbol{P}_m$, the gate complexity*

$$G = \Omega\left(\left(\mathrm{Poly}(n,m)^{1/p}\frac{p^{\frac{k}{2}}\left\|\boldsymbol{H}\right\|_{(0),2} t}{\epsilon}\right)^{1/\ell} \Gamma p^{\frac{k-1}{2}}\left\|\boldsymbol{H}\right\|_{(1),2} t\right) \quad \textit{ensures} \quad \left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_\ell(t/r)^r\right\|_{p,\bar{\boldsymbol{P}}_m} \le \epsilon, \tag{3.93}$$

where $\left\|\boldsymbol{H}\right\|_{(0),2}, \left\|\boldsymbol{H}\right\|_{(1),2}$ *is defined w.r.t to* (3.92)*.*

Note that we have drop the parameter s in $\left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_\ell(t/r)^r\right\|_{p,\bar{\boldsymbol{P}}_m,s}$ since every term commutes with $\boldsymbol{P}_m$ (and the auxiliary state $\boldsymbol{\rho}_{\frac{m}{n}}$).

Proof. The result quickly follows by converting to the p-norm w.r.t. the auxiliary product state $\boldsymbol{\rho}_\eta = \otimes_i \boldsymbol{\rho}_i = \otimes_i(\eta\left|1\right\rangle\left\langle 1\right| + (1-\eta)\left|0\right\rangle\left\langle 0\right|)_i$ defined by the filling ratio $\eta = \frac{m}{n}$. For $\boldsymbol{F} = [\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1,\cdots,\boldsymbol{H}_\Gamma,\Upsilon t)]_g$,

$$\left\|\boldsymbol{F}\right\|_{p,\bar{\boldsymbol{P}}} \le \left\|\boldsymbol{F}\right\|_{p,\boldsymbol{\rho}_\eta}\cdot(\mathrm{Poly}(n,m))^{1/p} \le \sqrt{\sum_{S\subset\{m,\cdots,1\}} (C_p)^{|S|}\left\|\boldsymbol{F}_S\right\|^2_{p,\boldsymbol{\rho}_\eta}}\cdot(\mathrm{Poly}(n,m))^{1/p}. \tag{3.94}$$

Some technical notes: Holder's inequality still works for the $\boldsymbol{\rho}_\eta-$ weighted norms[12] $\left\|\boldsymbol{H}_\gamma\boldsymbol{O}\right\|_{p,\boldsymbol{\rho}_\eta} \le \left\|\boldsymbol{O}\right\|_{p,\boldsymbol{\rho}_\eta}\left\|\boldsymbol{H}_\gamma\right\|$ (which needs not be true for general $\boldsymbol{\rho}$); if $\boldsymbol{O}$ is particle number preserving, then $[\boldsymbol{O}]_S$ is also particle number preserving [13]. □

Via Markov's inequality (plug $\boldsymbol{\rho} = \bar{\boldsymbol{P}}_m$ into Proposition II.0.1), we obtain concentration.

Corollary III.3.3. *Draw $\left|\psi\right\rangle$ from a $m = \eta n$ - particle subspaces (i.e., $\mathbb{E}[\left|\psi\right\rangle\left\langle\psi\right|] = \boldsymbol{P}_m/\mathrm{Tr}[\boldsymbol{P}_m]$), then*

$$G = \Omega\left(\left(\frac{\sqrt{\log(\mathrm{Poly}(n,m)/\delta)}^k\left\|\boldsymbol{H}\right\|_{(0),2} t}{\epsilon}\right)^{1/\ell}\sqrt{\log(\mathrm{Poly}(n,m)/\delta)}^{k-1}\Gamma\left\|\boldsymbol{H}\right\|_{(1),2} t\right)$$
$$\textit{ensures} \quad \Pr\left(\left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}(t/r)^r\left|\psi\right\rangle\right\|_{\ell_2} \ge \epsilon\right) \le \delta.$$

G. k-locality for Fermions

Analogously, we generalize to Hamiltonians with Fermionic terms. We begin with defining k-locality for Fermionic systems. Suppose the particle-number preserving Fermionic Hamiltonian can be written as

$$\boldsymbol{H} = \sum_{S_-\perp S_+\perp S_z \subset\{m,\cdots,1\}} b_{S_+S_-S_z} \prod_{s_+\in S_+} \boldsymbol{a}^\dagger_{s_+} \prod_{s_-\in S_-} \boldsymbol{a}_{s_-} \prod_{s_z\in S_z} \boldsymbol{O}^\eta_{s_z}. \tag{3.95}$$

Again, particle number preserving enforces $|S_+| = |S_-|$.[14] Recall the second quantization commutation relations (following [8])

$$[\boldsymbol{a}^\dagger, \boldsymbol{O}^\eta] = -\boldsymbol{a}^\dagger, \tag{3.96}$$
$$[\boldsymbol{a}, \boldsymbol{O}^\eta] = \boldsymbol{a}, \tag{3.97}$$

[12] Due to particle number preserving, i.e., $\boldsymbol{\rho}_\eta$ commutes with any term produced by the Hamiltonian.
[13] This can be seen by $\boldsymbol{O}$ is sum of terms that each has the same number of $\boldsymbol{Z}^+$ and $\boldsymbol{Z}^-$. Removing some of them by $[\boldsymbol{O}]_S$ does not change this structure.
[14] Even worse, odd Fermionic terms, e.g., a single fermion $\boldsymbol{a}^\dagger_i$, anti-commute with each other even if they have no overlapping sites.

and

$$[\boldsymbol{a}_j^\dagger \boldsymbol{a}_k, \boldsymbol{a}_\ell^\dagger \boldsymbol{a}_m] = \delta_{kl}\boldsymbol{a}_j^\dagger \boldsymbol{a}_m - \delta_{jm}\boldsymbol{a}_\ell^\dagger \boldsymbol{a}_k, \tag{3.98}$$

$$[\boldsymbol{a}_j^\dagger \boldsymbol{a}_k, \boldsymbol{O}_\ell^\eta] = \delta_{kl}\boldsymbol{a}_j^\dagger \boldsymbol{a}_\ell - \delta_{j\ell}\boldsymbol{a}_\ell^\dagger \boldsymbol{a}_k, \tag{3.99}$$

$$[\boldsymbol{a}_j^\dagger \boldsymbol{a}_k, \boldsymbol{O}_\ell^\eta \boldsymbol{O}_m^\eta] = \left(\delta_{kl}\boldsymbol{a}_j^\dagger \boldsymbol{a}_\ell - \delta_{j\ell}\boldsymbol{a}_\ell^\dagger \boldsymbol{a}_k\right)\boldsymbol{O}_m^\eta + \boldsymbol{O}_\ell^\eta\left(\delta_{km}\boldsymbol{a}_j^\dagger \boldsymbol{a}_m - \delta_{jm}\boldsymbol{a}_m^\dagger \boldsymbol{a}_k\right). \tag{3.100}$$

Compared with k-local Paulis, the only difference for the Fermionic case is (3.98): commuting two Fermionic operators on the same site ℓ can produce an identity $\boldsymbol{I}_\ell$. This would add an extra term in our effective 2-2 norm calculation (Corollary III.3.1)

$$\lambda_{ferm}(k) := \frac{2^{k/2+1}}{(k-1)!}\sum_{k_{pvt}=1}^{k}\frac{2^{k_{pvt}/2}}{(k-k_{pvt})!}\left\|\boldsymbol{H}\right\|_{(k_{pvt}),2} + \frac{2^{k/2+1}}{(k-1)!}\frac{1}{k!}\left\|\boldsymbol{H}_{ferm}\right\|_{(0),2}, \tag{3.101}$$

where the "global" 2-norm $\left\|\cdot\right\|_{(0),2}$ only contains Fermionic operators

$$\boldsymbol{H}_{ferm} := \sum_{|S_-|+|S_+|\neq 0, S_-\perp S_+\perp S_z\subset\{m,\cdots,1\}} b_{S_+S_-S_z}\prod_{s_+\in S_+}\boldsymbol{a}_{s_+}^\dagger\prod_{s_-\in S_-}\boldsymbol{a}_{s_-}\prod_{s_z\in S_z}\boldsymbol{O}_{s_z}^\eta. \tag{3.102}$$

Intuitively, when identity is produced at the overlapping site, more terms may collide, i.e., add coherently. See Section IV B for an example where this term is necessary. Otherwise, the rest of the calculation is identical ($\lambda'(k)$ remains the same). Note that we would use a Fermionic version of Fact III.3, which can be shown by a gauge transformation argument.

Proposition III.3.2 (k-local Fermionic Hamiltonians)**.** *To simulate a k-local, particle number preserving Fermionic Hamiltonian using ℓ-th order Suzuki formula on m-particle subspace $\boldsymbol{P}_m$, the gate complexity*

$$G = \Omega\left(\left(\frac{\mathrm{Poly}(n,m)^{1/p}p^{k/2}\left\|\boldsymbol{H}\right\|_{(0),2}t}{\epsilon}\right)^{1/\ell}\Gamma p^{k/2}(\left\|\boldsymbol{H}\right\|_{(1),2}+\left\|\boldsymbol{H}_{ferm}\right\|_{(0),2})t\right) \quad \textit{ensures} \quad \left\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\boldsymbol{S}(t/r)^r\right\|_{\bar{p},\boldsymbol{P}_m}\le\epsilon,$$

where $\left\|\boldsymbol{H}\right\|_{(0),2}, \left\|\boldsymbol{H}\right\|_{(1),2}$ is defined w.r.t to (3.95).

Corollary III.3.4. *Draw $|\psi\rangle$ from a $m=\eta n$ - particle subspaces (i.e., $\mathbb{E}[|\psi\rangle\langle\psi|] = \boldsymbol{P}_m/\mathrm{Tr}[\boldsymbol{P}_m]$), then*

$$G = \tilde{\Omega}\left(\left(\frac{\log(\mathrm{Poly}(n,m)/\delta)^{k/2}\left\|\boldsymbol{H}\right\|_{(0),2}t}{\epsilon}\right)^{1/\ell}\log(\mathrm{Poly}(n,m)/\delta)^{k/2}\Gamma\left(\left\|\boldsymbol{H}\right\|_{(1),2}+\left\|\boldsymbol{H}_{ferm}\right\|_{(0),2}\right)t\right)$$

$$\textit{ensures} \quad \Pr\left(\left\|(\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\boldsymbol{S}(t/r)^r)|\psi\rangle\right\|_{\ell_2}\ge\epsilon\right)\le\delta.$$

IV. OPTIMALITY FOR FIRST-ORDER AND SECOND-ORDER FORMULAS

We demonstrate the optimality of our p-norm estimates for a particular 2-local Hamiltonian, at short times, for the first and second-order Lie-Trotter-Suzuki formulas. The $k\ge 2$ cases can also be constructed analogously. Consider the Hamiltonian

$$\boldsymbol{H} = \sum_{i>j}\alpha_{ij}\boldsymbol{Z}_i\boldsymbol{Z}_j + \sum_{i>j}\alpha_{ij}\boldsymbol{X}_i\boldsymbol{X}_j =: \boldsymbol{A}+\boldsymbol{B} \tag{4.1}$$

for the first-order Trotter formula

$$\mathrm{e}^{\mathrm{i}(\boldsymbol{A}+\boldsymbol{B})t} - \mathrm{e}^{\mathrm{i}\boldsymbol{A}}\mathrm{e}^{\mathrm{i}\boldsymbol{B}t} = \frac{1}{2}[\boldsymbol{A},\boldsymbol{B}]t^2 + \mathcal{O}(t^3). \tag{4.2}$$

$$\tag{4.3}$$

We can exactly compute its 2-norm due to the orthogonality of Paulis

$$\left\|[\boldsymbol{A},\boldsymbol{B}]\right\|_2^2 = \left\|\left[\sum_{k>\ell}\alpha_{k\ell}\boldsymbol{Z}_k\boldsymbol{Z}_\ell, \sum_{i>j}\alpha_{ij}\boldsymbol{X}_i\boldsymbol{X}_j\right]\right\|_2^2 \tag{4.4}$$

$$= \sum_{\{i,j,k\}}\left\|\alpha_{ij}\alpha_{jk}\boldsymbol{Z}_i\boldsymbol{Y}_j\boldsymbol{X}_k\right\|_2^2 = \sum_{\{i,j,k\}}\alpha_{ij}^2\alpha_{jk}^2. \tag{4.5}$$

For our upper bounds (3.53),

$$\|\boldsymbol{H}\|_{(0),2}^2 = \sum_{ij} 4\alpha_{ij}^2, \|\boldsymbol{H}\|_{(1),2}^2 = \max_i \sum_j 4\alpha_{ij}^2, \tag{4.6}$$

which means when $\alpha_{ij} = 1$ are equal strength,

$$\|\boldsymbol{H}\|_{(0),2}^2 \|\boldsymbol{H}\|_{(1),2}^2 = \theta\left(\|[\boldsymbol{A},\boldsymbol{B}]\|_2^2\right). \tag{4.7}$$

It is less obvious how to calculate its p-norm or operator norm.

To obtain tight p-norm and spectral norm estimates, we construct another Hamiltonian on three set of qubits $\mathcal{H} = \mathcal{H}_{S_1} \otimes \mathcal{H}_{S_2} \otimes \mathcal{H}_{S_3}$

$$\boldsymbol{H} = \sum_{s_1\in S_1, s_2\in S_2} \boldsymbol{Z}_{s_1}\boldsymbol{X}_{s_2} + \sum_{s_2\in S_2, s_3\in S_3} \boldsymbol{Y}_{s_2}\boldsymbol{Z}_{s_3} := \boldsymbol{A}+\boldsymbol{B}. \tag{4.8}$$

The commutator evaluates to a factorized commuting sum

$$[\boldsymbol{A},\boldsymbol{B}] = \left[\sum_{s_1\in S_1, s_2\in S_2} \boldsymbol{Z}_{s_1}\boldsymbol{X}_{s_2}, \sum_{s_2\in S_2, s_3\in S_3} \boldsymbol{Y}_{s_2}\boldsymbol{Z}_{s_3}\right] \tag{4.9}$$

$$= 2\sum_{s_1\in S_1, s_2\in S_2, s_2\in S_3} \boldsymbol{Z}_{s_1}\boldsymbol{Z}_{s_2}\boldsymbol{Z}_{s_3} = 2(\sum_{s_1\in S_1}\boldsymbol{Z}_{s_1})\cdot(\sum_{s_2\in S_2}\boldsymbol{Z}_{s_2})\cdot(\sum_{s_3\in S_3}\boldsymbol{Z}_{s_3}). \tag{4.10}$$

Its p-norms can be obtain by central limit theorem at large $|S_1|, |S_2|, |S_3|$

$$\left\|\sum_{s_1\in S_1}\boldsymbol{Z}_{s_1}\right\|_p = \Omega(\sqrt{p\,|S_1|})\,\|\boldsymbol{I}\|_p\,, \tag{4.11}$$

where we recall the p-th moment of standard Gaussian $|g|_p = \theta(\sqrt{p})$. Now, let $|S_1| = |S_2| = |S_3| = \theta(n)$, then it saturates our first-order p-norm upper bound (3.53).

$$\|[\boldsymbol{A},\boldsymbol{B}]\|_p = \Omega(\sqrt{pn})^3\,\|\boldsymbol{I}\|_p \tag{4.12}$$

$$\sqrt{C_p}^3\,\|\boldsymbol{H}\|_{(0),2}\,\|\boldsymbol{H}\|_{(1),2}\,\|\boldsymbol{I}\|_p = \mathcal{O}\left(\sqrt{p}^3\cdot\sqrt{n^2}\cdot\sqrt{n}\right)\|\boldsymbol{I}\|_p\,. \tag{4.13}$$

At the same time, its spectral norm

$$\|[\boldsymbol{A},\boldsymbol{B}]\| = \theta(n^3) = \|\boldsymbol{H}\|_{(0),1}\,\|\boldsymbol{H}\|_{(1),1} \tag{4.14}$$

matches the triangle inequality bound in [8].

A. Second-order Suzuki Formulas

For the second-order Trotter error, recall the expansion [8, Appendix L],

$$\mathrm{e}^{\mathrm{i}(\boldsymbol{A}+\boldsymbol{B})t} - \mathrm{e}^{\mathrm{i}\boldsymbol{A}t/2}\mathrm{e}^{\mathrm{i}\boldsymbol{B}t}\mathrm{e}^{\mathrm{i}\boldsymbol{A}t/2} = -\frac{\mathrm{i}}{12}\left([\boldsymbol{B},[\boldsymbol{B},\boldsymbol{A}]-\frac{1}{2}[\boldsymbol{A},[\boldsymbol{A},\boldsymbol{B}]\right)t^3 + \mathcal{O}(t^4) \tag{4.15}$$

with the same Hamiltonian (4.8). Due to the symmetry, we know $[\boldsymbol{B},[\boldsymbol{B},\boldsymbol{A}]]$ has the same p-norm as $[\boldsymbol{A},[\boldsymbol{A},\boldsymbol{B}]]$. Conveniently, the factor $\frac{1}{2}$ allows us to consider only one term (at most losing a constant overhead $\frac{1}{2}$)

$$[\boldsymbol{B},[\boldsymbol{B},\boldsymbol{A}]] = -4\sum_{s_1\in S_1, s_2\in S_2, s_3, s_3'\in S_3} \boldsymbol{Z}_{s_1}\boldsymbol{X}_{s_2}\boldsymbol{Z}_{s_3}\boldsymbol{Z}_{s_3'}. \tag{4.16}$$

This converges to a function of three independent Gaussians (note that the s_3, s_3' are two dummy indexes in the same set S_3)

$$\|[\boldsymbol{B},[\boldsymbol{B},\boldsymbol{A}]]\|_p = \Omega(\left|g_1g_2g_3^2\right|p)\,\|\boldsymbol{I}\|_p = \Omega(\sqrt{pn})^4\,\|\boldsymbol{I}\|_p\,, \tag{4.17}$$

$$\sqrt{C_p}^4\,\|\boldsymbol{H}\|_{(0),2}\,\|\boldsymbol{H}\|_{(1),2}^2\,\|\boldsymbol{I}\|_p = \mathcal{O}\left(\sqrt{p}^4\cdot\sqrt{n^2}\cdot\sqrt{n}^2\right)\|\boldsymbol{I}\|_p\,, \tag{4.18}$$

matching our p-norm bound. The spectral norm

$$\|[\boldsymbol{B},[\boldsymbol{B},\boldsymbol{A}]\| = \theta(n^4) = \|\boldsymbol{H}\|_{(0),1}\,\|\boldsymbol{H}\|_{(1),1}^2 \tag{4.19}$$

again matches the triangle inequality bounds in [8].

B. Fermionic Hamiltonians

To demonstrate the need for the extra term for Fermionic Hamiltonians $\|\boldsymbol{H}_{ferm}\|_{(0),2}$, consider a Hamiltonian of the form

$$\boldsymbol{H} = \sum_{s_1\in S_1, s_2\in S_2} \boldsymbol{a}_{s_1}\boldsymbol{a}^{\dagger}_{s_2} + \boldsymbol{a}^{\dagger}_{s_1}\boldsymbol{a}_{s_2} + \sum_{s_2\in S_2, s_3\in S_3} \boldsymbol{a}_{s_2}\boldsymbol{a}^{\dagger}_{s_3} + \boldsymbol{a}^{\dagger}_{s_2}\boldsymbol{a}_{s_3} := \boldsymbol{A} + \boldsymbol{B}. \tag{4.20}$$

The commutator evaluates to

$$[\boldsymbol{B}, \boldsymbol{A}] = \sum_{s_1\in S_1, s_2\in S_2, s_2\in S_3} \boldsymbol{a}_{s_1}\boldsymbol{a}^{\dagger}_{s_3} - \boldsymbol{a}^{\dagger}_{s_1}\boldsymbol{a}_{s_3} \tag{4.21}$$

$$\|[\boldsymbol{A}, \boldsymbol{B}]\|_2 = 2\sqrt{|S_1|}\,|S_2|\,\sqrt{|S_3|} = \theta(\|\boldsymbol{H}\|_{(0),2}\,\|\boldsymbol{H}\|_{(0),2}). \tag{4.22}$$

And for the second-order Suzuki,

$$[\boldsymbol{B}, [\boldsymbol{B}, \boldsymbol{A}]] = -|S_2| \cdot \sum_{s_1\in S_1, s_2\in S_2, s_2\in S_3} \boldsymbol{a}^{\dagger}_{s_2}\boldsymbol{a}_{s_1} + \boldsymbol{a}_{s_1}\boldsymbol{a}^{\dagger}_{s_2} \tag{4.23}$$

$$\|[\boldsymbol{B}, [\boldsymbol{B}, \boldsymbol{A}]\|_2 = 2\sqrt{|S_1|}\,|S_2|^2\,\sqrt{|S_3|} = \theta(\|\boldsymbol{H}\|^2_{(0),2}\,\|\boldsymbol{H}\|_{(0),2}). \tag{4.24}$$

V. PRELIMINARY: MATRIX-VALUED MARTINGALES

Concentration inequalities are well known for i.i.d sum of random scalars and matrices. However, beyond sums, we would need to address concentration for matrix-valued functions. For this, we rely on tools from *(matrix-valued) martingales.* For a minimal technical introduction (following Tropp [35] and Huang et. al [24]), consider a filtration of the master sigma algebra $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \cdots \subset \mathcal{F}_t \subset \cdots \mathcal{F}$, where for each $\mathcal{F}_j$ we denote the conditional expectation $\mathbb{E}_j$. A martingale is a sequence of random variable Y_t adapted to the filtration $\mathcal{F}_t$ such that

$$\sigma(Y_t) \subset \mathcal{F}_t \qquad \text{(causality)}, \tag{5.1}$$

$$\mathbb{E}_{t-1} Y_t = Y_{t-1} \qquad \text{(status quo)}. \tag{5.2}$$

Intuitively, we can think of t is a 'time' index and $\mathcal{F}_t$ hosts possible events happening before t. The present depends on the past ('causality'), and tomorrow has the same expectation as today ('status quo'). For simplicity, we often subtract the mean to obtain a *martingale difference* sequence $D_t := Y_t - Y_{t-1}$ such that

$$\mathbb{E}_{t-1} D_t = 0. \tag{5.3}$$

A. Useful Norms and Recursive Bounds

In our case, the martingale would be *matrix-valued.* Historically, the earliest general results were established in [36–38], and more recent works and applications include [24, 26, 27, 35, 39, 40]. Throughout this work, the martingale tool we use restricts to uniform smoothness (in slightly different forms for non-random Hamiltonians and random Hamiltonians). It is not the tightest kind of martingale inequality but arguably the simplest and most robust when matrices are bounded (or with Gaussian coefficients via the central limit theorem).

Let us begin by picking a suitable norm. Keeping in mind the goal to quantify the error between the ideal unitary $\boldsymbol{U} := e^{\mathrm{i}\boldsymbol{H}t}$ and the product formula $\boldsymbol{S}$, there are two plausible error metrics with different operational meaning.

1. *The operator norm*

If we are interested in the concentration in operator norm $\|\boldsymbol{U} - \boldsymbol{V}\|$, it suffices to control its moments by the expected Schatten p-norm

$$(\mathbb{E}\|\boldsymbol{Y}\|^p)^{1/p} \le (\mathbb{E}\|\boldsymbol{Y}\|_p^p)^{1/p} =: |\!|\!|\boldsymbol{Y}|\!|\!|_p. \tag{5.4}$$

where $\|\boldsymbol{X}\|_p := \mathrm{Tr}[(\boldsymbol{X}^{\dagger}\boldsymbol{X})^{p/2}]^{1/p}$. To bound the RHS, the workhorse is the following bound with only a martingale requirement ("conditionally zero-mean").

Fact V.1 (Uniform smoothness for Schatten classes [24, Proposition 4.3])**.** *Consider random matrices $\boldsymbol{X}, \boldsymbol{Y}$ of the same size that satisfy $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}] = 0$. When $2 \le p$,*

$$|\!|\!|\boldsymbol{X} + \boldsymbol{Y}|\!|\!|_p^2 \le |\!|\!|\boldsymbol{X}|\!|\!|_p^2 + C_p|\!|\!|\boldsymbol{Y}|\!|\!|_p^2 \tag{5.5}$$

The constant $C_p = p - 1$ is the best possible.

Uniform smoothness for Schatten classes was first proven by [22], with optimal constants determined by [41]. The above martingale form is due to [23] and [24, Proposition 4.3].

2. Fixed input state

Sometimes we only care about a fixed input state. This turns out deserves a different norm (following [27])

$$\frac{1}{2}\left(\sup_{\rho}\mathbb{E}\left\|\boldsymbol{U}\rho\boldsymbol{U}^{\dagger}-\boldsymbol{S}\rho\boldsymbol{S}^{\dagger}\right\|_1^p\right)^{1/p} \le \frac{1}{2}\left(\sup_{|\psi\rangle}\mathbb{E}\left\|\boldsymbol{U}\left|\psi\right\rangle\left\langle\psi\right|\boldsymbol{U}^{\dagger}-\boldsymbol{S}\left|\psi\right\rangle\left\langle\psi\right|\boldsymbol{S}^{\dagger}\right\|_1^p\right)^{1/p} \tag{5.6}$$

$$\le \sup_{|\psi\rangle}\left(\mathbb{E}\left\|(\boldsymbol{U}-\boldsymbol{S})\left|\psi\right\rangle\left\langle\psi\right|\right\|_1^p\right)^{1/p} = \sup_{|\psi\rangle}\left(\mathbb{E}\left\|(\boldsymbol{U}-\boldsymbol{S})\left|\psi\right\rangle\left\langle\psi\right|\right\|_p^p\right)^{1/p}. \tag{5.7}$$

The first inequality uses convexity, the second is a telescoping sum, and the third is the fact that the p-norms $\|\|_p = \|\|_1$ are equal for rank 1 matrices. Notice this differs from the spectral norm by order of quantifier

$$(\mathbb{E}\|\boldsymbol{Y}\|^p)^{1/p} := \left(\mathbb{E}\sup_{|\psi\rangle}\|\boldsymbol{Y}\left|\psi\right\rangle\left\langle\psi\right|\|_1^p\right)^{1/p} \neq \sup_{|\psi\rangle}\left(\mathbb{E}\|\boldsymbol{Y}\left|\psi\right\rangle\left\langle\psi\right|\|_1^p\right)^{1/p}. \tag{5.8}$$

Formally, we need to define another norm

$$|||\boldsymbol{X}|||_{\mathrm{fix},p} := \sup_{rank(\boldsymbol{P})=1}\left(\mathbb{E}[\|\boldsymbol{X}\boldsymbol{P}\|_p^p]\right)^{1/p}, \tag{5.9}$$

where the supremum is taken over projectors $\boldsymbol{P}$ with rank 1. We therefore extend uniform smoothness to the following:

Corollary V.1.1 (Uniform smoothness, fixed input [26])**.** *Consider random matrices $\boldsymbol{X},\boldsymbol{Y}$ of the same size that satisfy $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}]=0$. When $2\le p$,*

$$|||\boldsymbol{X}+\boldsymbol{Y}|||_{fix,p}^2 = |||\boldsymbol{X}|||_{fix,p}^2 + C_p|||\boldsymbol{Y}|||_{fix,p}^2 \tag{5.10}$$

with constant $C_p = p-1$.

This can be seen by

$$|||\boldsymbol{X}+\boldsymbol{Y}|||_{\mathrm{fix},p}^2 = \sup_{rank(\boldsymbol{P})=1}|||\boldsymbol{X}\boldsymbol{P}+\boldsymbol{Y}\boldsymbol{P}|||_p^2 \tag{5.11}$$

$$\le \sup_{rank(\boldsymbol{P})=1}\left(|||\boldsymbol{X}\boldsymbol{P}|||_p^2 + C_p|||\boldsymbol{Y}\boldsymbol{P}|||_p^2\right) \tag{5.12}$$

$$\le \sup_{rank(\boldsymbol{P})=1}|||\boldsymbol{X}\boldsymbol{P}|||_p^2 + C_p\sup_{rank(\boldsymbol{P})=1}|||\boldsymbol{Y}\boldsymbol{P}|||_p^2. \tag{5.13}$$

These simple recursive inequalities streamline our proof for concentration for matrix polynomial(Section VII). These inequalities deliver sum-of-square ("incoherent") estimates sharper than triangle inequality, which is linear ("coherent").

B. Reminders of Useful Facts

Before we turn to the proof, let us remind ourselves the useful properties for the underlying norms $|||\cdot|||_* := |||\cdot|||_{p,q}, |||\cdot|||_{\mathrm{fix},p}$ for $p,q\ge 2$. They are largely inherited from the (non-random) Schatten p-norm. Following [26],

Fact V.2 (non-commutative Minkowski)**.** *Each of the expected moments satisfies the triangle inequality and thus is a valid norm. For any random matrix $\boldsymbol{X},\boldsymbol{Y}$*

$$|||\boldsymbol{X}+\boldsymbol{Y}|||_* \le |||\boldsymbol{X}|||_* + |||\boldsymbol{Y}|||_*. \tag{5.14}$$

Fact V.3 (operator ideal norms)**.** *For operators $\boldsymbol{A}$ deterministic and $\boldsymbol{O}$ random*

$$|||\boldsymbol{A}\boldsymbol{O}|||_*, |||\boldsymbol{O}\boldsymbol{A}|||_* \le \|\boldsymbol{A}\|\cdot|||\boldsymbol{O}|||_*. \tag{5.15}$$

Fact V.4 (unitary invariant norms)**.** *For $\boldsymbol{U},\boldsymbol{V}$ deterministic unitaries and random operator $\boldsymbol{O}$*

$$|||\boldsymbol{U}\boldsymbol{O}\boldsymbol{V}|||_* = |||\boldsymbol{O}|||_*. \tag{5.16}$$

Being operator ideal already implies unitary invariance, but we state it regardless. As the norm $|||\cdot|||_{\mathrm{fix},p}$ defined for low rank input is somewhat non-standard, we include a proof as follows.

Proof of Fact V.3 for fixed inputs.

$$\|\!|\boldsymbol{X}\boldsymbol{A}|\!\|_{\mathrm{fix},p} = \sup_{rank(\boldsymbol{P})=1} (\mathbb{E}[\|\boldsymbol{X}\boldsymbol{A}\boldsymbol{P}\|_p^p])^{1/p} \tag{5.17}$$

$$= \sup_{rank(\boldsymbol{P})=1} (\mathbb{E}[\|\boldsymbol{X}\boldsymbol{P}'\boldsymbol{A}'\|_p^p])^{1/p} \tag{5.18}$$

$$= \sup_{rank(\boldsymbol{P}')=1} (\mathbb{E}[\|\boldsymbol{X}\boldsymbol{P}'\|_p^p])^{1/p}\|\boldsymbol{A}\| \tag{5.19}$$

$$= \|\!|\boldsymbol{X}|\!\|_{\mathrm{fix},p}\|\boldsymbol{A}\|. \tag{5.20}$$

In the second line, we use the singular value decomposition

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S}_1\boldsymbol{S}_{\boldsymbol{A}'}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S}_1\boldsymbol{U}^\dagger\cdot\boldsymbol{U}\boldsymbol{S}_{\boldsymbol{A}'}\boldsymbol{V} := \boldsymbol{P}'\boldsymbol{A}', \tag{5.21}$$

where we rewrite the diagonal elements as product $\boldsymbol{S} = \boldsymbol{S}_1\boldsymbol{S}_{\boldsymbol{A}'}$, where $\boldsymbol{S}_1$ is a rank 1 projector and $\|\boldsymbol{S}_{\boldsymbol{A}'}\| \le \|\boldsymbol{S}\| \le \|\boldsymbol{A}\|$. This is possible since $\boldsymbol{S}$ is bounded by $\|\boldsymbol{S}\| \le \|\boldsymbol{P}\boldsymbol{A}\| \le \|\boldsymbol{A}\|$. This is the advertised result. □

VI. FIRST-ORDER TROTTER FOR RANDOM HAMILTONIANS

We will use the above tools to derive bounds for first-order Trotter on a random Hamiltonian. It suffices to control the error defined in the exponentiated form [8]

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}_\Gamma t}\cdots\mathrm{e}^{\mathrm{i}\boldsymbol{H}_1} = \exp_{\mathcal{T}}(\mathrm{i}\int(\boldsymbol{\mathcal{E}}(t)+\boldsymbol{H})dt), \tag{6.1}$$

which is given by commutators

$$\boldsymbol{\mathcal{E}}(t) := \sum_{k=2}^{\Gamma}\left(\prod_{\gamma=k-1}^{1}\mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k\right), \tag{6.2}$$

Theorem VI.1 (First-order Trotter for random Hamiltonians)**.** *For Hamiltonian $\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma}\boldsymbol{H}_\gamma$ on n qudits, where (A) each term $\boldsymbol{H}_\gamma$ is independent, zero mean $\mathbb{E}\boldsymbol{H}_\gamma = 0$ and bounded $\|\boldsymbol{H}_\gamma\| \le b_\gamma$ almost surely. or (B) $\boldsymbol{H}_\gamma = g_\gamma\boldsymbol{Z}_\gamma$, with g_γ i.i.d Gaussian and $\|\boldsymbol{Z}_\gamma\| \le b_\gamma$ deterministic bounded matrix. Then gate count*

$$G = 2\sqrt{2}\Gamma(n\ln(d)+\log(\mathrm{e}^2/\delta))\frac{\|\boldsymbol{H}\|_{(0),2}\,\|\boldsymbol{H}\|_{(1),2}\,t^2}{\epsilon} \quad \textit{ensures} \quad \Pr\left(\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_1(t/r)^r\| \ge \epsilon\right) \le \delta.$$

For arbitrary fixed input state ρ, gate count

$$G = 2\sqrt{2}\Gamma(\log(\mathrm{e}^2/\delta)\frac{\|\boldsymbol{H}\|_{(0),2}\,\|\boldsymbol{H}\|_{(1),2}\,t^2}{\epsilon} \quad \textit{ensures} \quad \Pr\left(\frac{1}{2}\left\|(\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\rho\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_1(t/r)^{\dagger r}\rho\boldsymbol{S}_1(t/r)^r)\right\|_1 \ge \epsilon\right) \le \delta.$$

For example, for SYK-like models on n-qudits, $\Gamma \le n^k/k!$, and

$$\|\boldsymbol{H}\|_{(0),2}^2 \le \frac{J^2(k-1)!}{kn^{k-1}}\cdot\frac{n^k}{k!} = \frac{J^2 n}{k^2} \tag{6.3}$$

$$\|\boldsymbol{H}\|_{(1),2}^2 \le \frac{J^2(k-1)!}{kn^{k-1}}\cdot k\frac{n^{k-1}}{(k-1)!} = J^2. \tag{6.4}$$

Corollary VI.1.1 (First-order Trotter for SYK models)**.** *For SYK-like k-local random model,*

$$G = \frac{2\sqrt{2}}{k\cdot k!}(n\ln(d)+\log(\mathrm{e}^2/\delta))\frac{n^{k+1/2}(Jt)^2}{\epsilon} \quad \textit{ensures} \quad \Pr(\|e^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_1(t/r)^r\| \ge \epsilon) \le \delta.$$

For arbitrary fixed input state,

$$G = \frac{2\sqrt{2}}{k\cdot k!}\log(\mathrm{e}^2/\delta)\frac{n^{k+1/2}(Jt)^2}{\epsilon} \quad \textit{ensures} \quad \Pr\left(\frac{1}{2}\left\|(\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\rho\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_1(t/r)^{\dagger r}\rho\boldsymbol{S}_1(t/r)^r)\right\|_1 \ge \epsilon\right) \le \delta.$$

The proof of Theorem VI.1 is mainly due to the following.

Lemma VI.2. *For both p-norms $|\!|\!|\cdot|\!|\!|_* = |\!|\!|\cdot|\!|\!|_p, |\!|\!|\cdot|\!|\!|_{fix,p}$,*

$$\left|\!\left|\!\left|\sum_{k=2}^{\Gamma}\left(\prod_{\gamma=k-1}^{1}\mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k\right)\right|\!\right|\!\right|_*^2 \le 2|\!|\!|\boldsymbol{I}|\!|\!|_*^2 C_p\sum_{k=1}^{\Gamma-1}\left[4C_p t^2\sum_{n=1}^{k-1}\left|\|[\boldsymbol{H}_n,\boldsymbol{H}_k]\|_\infty\right|_p^2 + \left(\sum_{n=1}^{k-1}\frac{t^2}{2}\left|\|[\boldsymbol{H}_n,[\boldsymbol{H}_n,\boldsymbol{H}_k]]\|_\infty\right|_p\right)^2\right].$$

Proof of Thoerem VI.1. For a total evolution time t, use the Trotter formula for r rounds with duration $\tau = t/r$ each. By Lemma VI.2, each round give an error

$$\vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{\mathcal{E}}(\tau) \vert\kern-0.25ex\vert\kern-0.25ex\vert_*^2 \le 2 \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_*^2 C_p \sum_{k=1}^{\Gamma-1} \left[4C_p\tau^2 \sum_{n=1}^{k-1} \left| \vert\kern-0.25ex\vert\kern-0.25ex\vert [\boldsymbol{H}_n, \boldsymbol{H}_k] \Vert_\infty \right|_p^2 + \left(\sum_{n=1}^{k-1} \frac{\tau^2}{2} \left| \vert\kern-0.25ex\vert\kern-0.25ex\vert [\boldsymbol{H}_n, [\boldsymbol{H}_n, \boldsymbol{H}_k]] \Vert_\infty \right|_p \right)^2 \right] \tag{6.5}$$

$$\le 2 \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_*^2 C_p \sum_{k=1}^{\Gamma-1} \left[16 C_p b_k^2 \tau^2 \left\Vert \boldsymbol{H} \right\Vert_{(1),2}^2 + \left(b_k \frac{\tau^2}{2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2}^2 \right)^2 \right] \tag{6.6}$$

$$\le 32 \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_*^2 C_p \left\Vert \boldsymbol{H} \right\Vert_{(0),2}^2 \tau^2 \left\Vert \boldsymbol{H} \right\Vert_{(1),2}^2 \left[C_p + \left(\frac{\tau}{4} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \right)^2 \right] \tag{6.7}$$

$$\le 32 \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_*^2 p^2 \tau^2 \left\Vert \boldsymbol{H} \right\Vert_{(0),2}^2 \left\Vert \boldsymbol{H} \right\Vert_{(1),2}^2 \tag{6.8}$$

where we used $\frac{\tau}{4} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \le 1$ to simplify the subleading effects. Intergrate along τ and telescope,

$$\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} - \boldsymbol{S}_1(t/r)^r \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_* = \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{\mathcal{E}}_{tot}(t) \vert\kern-0.25ex\vert\kern-0.25ex\vert_* \le r \cdot 4\sqrt{2} \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_* p \frac{\tau^2}{2} \left\Vert \boldsymbol{H} \right\Vert_{(0),2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \tag{6.9}$$

$$= 2\sqrt{2} \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_* p \frac{t^2}{r} \left\Vert \boldsymbol{H} \right\Vert_{(0),2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} =: \lambda \vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{I} \vert\kern-0.25ex\vert\kern-0.25ex\vert_* p \tag{6.10}$$

The rest will be evaluating Markov's inequality.
(i) For the spectral norm, set $\vert\kern-0.25ex\vert\kern-0.25ex\vert \cdot \vert\kern-0.25ex\vert\kern-0.25ex\vert_* = (\mathbb{E}\Vert \cdot \Vert_p^p)^{1/p}$

$$\Pr(\Vert \boldsymbol{\mathcal{E}}_{tot} \Vert \ge \epsilon) \le \frac{\mathbb{E}\Vert \boldsymbol{\mathcal{E}}_{tot} \Vert^p}{\epsilon^p} \le \frac{\mathbb{E} \left\Vert \boldsymbol{\mathcal{E}}_{tot} \right\Vert_p^p}{\epsilon^p} \tag{6.11}$$

$$\le D \left(p \frac{\lambda}{\epsilon} \right)^p \tag{6.12}$$

$$\le D \exp(-\frac{\epsilon}{\lambda} + 2) \tag{6.13}$$

where the factor of dimension $D = \Vert I \Vert_p^p$ is due to trace, and used $+2$ to ensure $p \ge 2$. Therefore, we need $\lambda \le \frac{\epsilon}{\log(\mathrm{e}^2 D/\delta)}$, or

$$G = \Gamma r = 2\sqrt{2} \log(\mathrm{e}^2 D/\delta) \Gamma \frac{t^2}{\epsilon} \left\Vert \boldsymbol{H} \right\Vert_{(0),2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \tag{6.14}$$

to ensure the Trotter error is at most ϵ with failure probability δ.
(ii) For fixed input state, the factor of $\log(D)$ disappear since $\Vert I \Vert_{\mathrm{fix},p}^p = \sup_{rank(P)=1} \Vert P \Vert_p^p = 1$

$$G = \Gamma r \ge 2\sqrt{2} \log(\mathrm{e}^2/\delta) \Gamma \frac{t^2}{\epsilon} \left\Vert \boldsymbol{H} \right\Vert_{(0),2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \tag{6.15}$$

which is already better than qDRIFT. Note, in both (i) and (ii), the choices of r guarantee that

$$\frac{t^2}{r^2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \left\Vert \boldsymbol{H} \right\Vert_{(1),2} \le \frac{\epsilon}{2\sqrt{2} \log(\mathrm{e}^2/\delta)} \le 16, \tag{6.16}$$

which is what we needed for (6.8).

Gaussian coefficients.
So far, we have shown for bounded, zero-mean random matrices. It is only one step away from matrices with Gaussian coefficients by central limit theorem. Represent Gaussian by Rademacher

$$\boldsymbol{H}_\gamma = g_\gamma \boldsymbol{Z}_\gamma = (\lim_{N\to\infty} \sum_j^N \frac{\epsilon_{\gamma,j}}{\sqrt{N}}) \boldsymbol{Z}_\gamma := \lim_{N\to\infty} \sum_j^N \boldsymbol{Y}_{\gamma,j}, \tag{6.17}$$

we obtain an Hamiltonian as sum over bounded, zero mean summands

$$\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}_\gamma = \sum_{j=1}^{N} \sum_{\gamma=1}^{\Gamma} \boldsymbol{Y}_{\gamma,j} = \sum_{\gamma'=1}^{\Gamma'} \boldsymbol{Y}_{\gamma'} := \boldsymbol{H}' \tag{6.18}$$

where we use another notation $\boldsymbol{H}'$ to refer to the data about summand $\boldsymbol{Y}_{\gamma'}$. We then need to control the performance of the first-order formula

$$\boldsymbol{S}_1(t) = \mathrm{e}^{\mathrm{i}\boldsymbol{H}'_{\Gamma'}t} \cdots \mathrm{e}^{\mathrm{i}\boldsymbol{H}'_1} = \prod_{\gamma=1}^{\Gamma} \mathrm{e}^{i\boldsymbol{Y}_{\gamma,j}} = \exp_{\mathcal{T}}(\mathrm{i} \int (\boldsymbol{\mathcal{E}}(t) + \boldsymbol{H}) dt), \tag{6.19}$$

by the commutators

$$\boldsymbol{\mathcal{E}}(t) := \sum_{k=2}^{\Gamma'} \left(\prod_{\gamma'=k-1}^{1} \mathrm{e}^{\mathcal{L}_{\gamma'} t}[\boldsymbol{Y}_k] - \boldsymbol{Y}_k \right). \tag{6.20}$$

Thankfully, we just need to call Lemma VI.2 and evaluate

$$\|\boldsymbol{H}'\|_{(0),2}^2 = \|\boldsymbol{H}\|_{(0),2}^2 \tag{6.21}$$

$$\|\boldsymbol{H}'\|_{(1),2}^2 = \|\boldsymbol{H}\|_{(0),2}^2 \tag{6.22}$$

to conclude the proof that the Gaussian case indeed has the advertised expression w.r.t to b_γ. □

A. Proof of Lemma VI.2

It remains to prove Lemma VI.2.

Proof. By uniform smoothness,

$$\left\|\left\|\boldsymbol{\mathcal{E}}(t)\right\|\right\|_*^2 = \left\|\left\|\sum_{k=2}^{\Gamma}\left(\prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k\right)\right\|\right\|^2 \le \left\|\left\|\sum_{k=2}^{\Gamma-1}\left(\prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k\right)\right\|\right\|^2 + C_p \left\|\left\|\prod_{\gamma=\Gamma-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_\Gamma] - \boldsymbol{H}_\Gamma\right\|\right\|_*^2 \tag{6.23}$$

$$\le C_p \sum_{k=1}^{\Gamma-1} \left\|\left\|\prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k\right\|\right\|_*^2 \tag{6.24}$$

where we crucially used the conditional independence for each of $k = 1, \cdots, \Gamma$

$$\mathbb{E}_{k-1} \prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k = 0. \tag{6.25}$$

Next, we will expand to the next order. Keeping the martingale calculation tidy requires some foresight to subtract a term ('the bias'). For each k, consider a telescoping sum

$$\prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k = \sum_{n=1}^{k-1} \prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] \tag{6.26}$$

$$= \sum_{n=1}^{k-1} \left(\prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] - \prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}\mathbb{E}_{n-1}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] \right) \tag{6.27}$$

$$+ \sum_{n=1}^{k-1} \prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}\mathbb{E}_{n-1}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] \tag{6.28}$$

$$:= \sum_{n=1}^{k-1} \boldsymbol{D}_n + \boldsymbol{B}_n \tag{6.29}$$

Where the first terms desirably form a martingale difference sequence that $\mathbb{E}_{n-1}\boldsymbol{D}_n = 0$, which brings about the sum-of-squares behavior

$$\left\|\left\|\prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k\right\|\right\|_*^2 = \left\|\left\|\sum_{n=1}^{k-1} \boldsymbol{D}_n + \boldsymbol{B}_n\right\|\right\|_*^2 \le \left(\left\|\left\|\sum_{n=1}^{k-1} \boldsymbol{D}_n\right\|\right\|_* + \left\|\left\|\sum_{n=1}^{k-1} \boldsymbol{B}_n\right\|\right\|_* \right)^2 \tag{6.30}$$

$$\le 2\left\|\left\|\sum_{n=1}^{k-1} \boldsymbol{D}_n\right\|\right\|_*^2 + 2\left\|\left\|\sum_{n=1}^{k-1} \boldsymbol{B}_n\right\|\right\|_*^2 \tag{6.31}$$

$$\le 2C_p \sum_{n=1}^{k-1} \left\|\left\|\boldsymbol{D}_n\right\|\right\|_*^2 + 2\left\|\left\|\sum_{n=1}^{k-1} \boldsymbol{B}_n\right\|\right\|_*^2 \tag{6.32}$$

where we used an elementary inequality $(a+b)^2 \le 2a^2 + 2b^2$ and lastly called uniform smoothness.

$$\left\|\left\|\left\| \boldsymbol{D}_n \right\|\right\|\right._*^2 = \left|\left|\left| \prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] - \prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}\mathbb{E}_{n-1}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] \right|\right|\right|_*^2 \tag{6.33}$$

$$\le \left|\left|\left| \left((\mathrm{e}^{\mathcal{L}_n t} - I) - \mathbb{E}_{n-1}[e^{\mathcal{L}_n t} - I] \right) [\boldsymbol{H}_k] \right|\right|\right|_*^2 \tag{6.34}$$

$$\le 4 \left|\left|\left| \left(\mathrm{e}^{\mathcal{L}_n t} - I\right) [\boldsymbol{H}_k] \right|\right|\right|_*^2 \tag{6.35}$$

$$\le 4t^2 \left| \left\| [\boldsymbol{H}_n, \boldsymbol{H}_k] \right\|_\infty \right|_p^2 |||\boldsymbol{I}|||_*^2 \tag{6.36}$$

where we sacrifice a factor of 2^2 by using convexity of $|||\cdot|||_*^2$. And for the bias,

$$\left|\left|\left| \sum_{n=1}^{k-1} \boldsymbol{B}_n \right|\right|\right|_*^2 = \left|\left|\left| \sum_{n=1}^{k-1} \prod_{\gamma=n-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}\mathbb{E}_{n-1}(\mathrm{e}^{\mathcal{L}_n t} - I)[\boldsymbol{H}_k] \right|\right|\right|_*^2 \tag{6.37}$$

$$\le \left(\sum_{n=1}^{k-1} \frac{t^2}{2} \left| \left\| [\boldsymbol{H}_n, [\boldsymbol{H}_n, \boldsymbol{H}_k]] \right\|_\infty \right|_p \right)^2 |||\boldsymbol{I}|||_*^2 \tag{6.38}$$

which is high-order in t and will be truly subleading in the end. Plug in the complete expression to arrive at the total error,

$$|||\boldsymbol{\mathcal{E}}(t)|||_*^2 \le C_p \sum_{k=1}^{\Gamma-1} \left|\left|\left| \prod_{\gamma=k-1}^{1} \mathrm{e}^{\mathcal{L}_\gamma t}[\boldsymbol{H}_k] - \boldsymbol{H}_k \right|\right|\right|_*^2 \tag{6.39}$$

$$\le C_p \sum_{k=1}^{\Gamma-1} \left[2 \left|\left|\left| \sum_{n=1}^{k-1} \boldsymbol{D}_n^{(k)} \right|\right|\right|_*^2 + 2 \left|\left|\left| \sum_{n=1}^{k-1} \boldsymbol{B}_n^{(k)} \right|\right|\right|_*^2 \right] \tag{6.40}$$

$$\le 2|||\boldsymbol{I}|||_*^2 C_p \sum_{k=1}^{\Gamma-1} \left[4C_p t^2 \sum_{n=1}^{k-1} \left| \left\| [\boldsymbol{H}_n, \boldsymbol{H}_k] \right\|_\infty \right|_p^2 + \left(\sum_{n=1}^{k-1} \frac{t^2}{2} \left| \left\| [\boldsymbol{H}_n, [\boldsymbol{H}_n, \boldsymbol{H}_k]] \right\|_\infty \right|_p \right)^2 \right]. \tag{6.41}$$

We declare victory that the sub-leading terms are mild, and the dominating source of error coincides with the leading order Taylor expansion. □

VII. PRELIMINARY: CONCENTRATION FOR MULTIVARIATE POLYNOMIALS

This section prepares us for the proof of Trotter error in random Hamiltonians (Section VIII). We use the "local" martingale inequality recursively to derive the "global" concentration for multivariate matrix polynomials.

A. Scalars

For a polynomial of independent scalars, the general results are relatively new and multifaceted [42–44]. The problem is better understood for Rademachers and Gaussians, captured in the form of Hypercontractivity [45, 46]. It relates the p-norm $|f|_p := (\mathbb{E}[|f|^p])^{1/p}$ to the 2-norm, i.e., the typical fluctuation is well-captured by the variance.

Fact VII.1 (Hypercontractivity for rademacher polynomial [45])**.** *For a degree-r polynomial of rademachers* $f(z_m, \cdots, z_1) = \sum_{S\subset\{0,1\}^m} f_S \prod_{s\in S} z_s$*. Then*

$$|f|_p \le \left| \sum_S \sqrt{C_p}^{|S|} f_S \prod_{s\in S} z_s \right|_2 \le \sqrt{C_p}^{\,r} |f|_2 . \tag{7.1}$$

Fact VII.2 (Hypercontractivity for a multivariate polynomial of independent Gaussians [46, Theorem 6.12])**.** *For a degree-r polynomial of i.i.d Gaussian variables* $f(g_m, \cdots, g_1)$*. Then*

$$|f|_p \le \sqrt{C_p}^{\,r} |f|_2 . \tag{7.2}$$

We do not present the intermediate bound for the Gaussian case because it requires expansion by the orthogonal Hermite polynomials, which complicates the picture. Note that we can WLG assumed the above to have zero mean.

B. Matrices

Unlike the scalar cases, concentration for a multivariate polynomial of matrices is relatively unexplored; even the i.i.d cases are fairly modern (See, e.g., [47]) and there it remains what the appropriate matrix analog quantity (such as the variance) is. For multivariate polynomials, the problem seems too general in terms of how matrices may interact with each other and how randomness is involved.

Nevertheless, we will derive concentration results that arguably match the best-known scalar results. What enables this is that we specialize in polynomials of *bounded* matrices with Gaussian coefficients, motivated by concrete applications in physics and quantum information (e.g., Hamiltonian with Paulis strings with Gaussian coefficients). We will derive comparable Hypercontractivity-like results for matrices using *uniform smoothness*, the matrix version of scalar two-point inequality (Section I B 2). The familiar reader may find the style of proof analogous to the scalar cases.

Fact VII.3 (Uniform smoothness for Schatten classes and with fixed input)**.** *Consider random matrices* $\boldsymbol{X}, \boldsymbol{Y}$ *of the same size that satisfy* $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}] = 0$*. When* $2 \le p$*,*

$$\left\lvert\!\left\lvert\!\left\lvert \boldsymbol{X}+\boldsymbol{Y} \right\rvert\!\right\rvert\!\right\rvert_*^2 \le \left\lvert\!\left\lvert\!\left\lvert \boldsymbol{X} \right\rvert\!\right\rvert\!\right\rvert_*^2 + C_p \left\lvert\!\left\lvert\!\left\lvert \boldsymbol{Y} \right\rvert\!\right\rvert\!\right\rvert_*^2 \tag{7.3}$$

with constant $C_p = p-1$ *and p-th moments* $\lvert\!\lvert\!\lvert\cdot\rvert\!\rvert\!\rvert_* = \lvert\!\lvert\!\lvert\cdot\rvert\!\rvert\!\rvert_p, \lvert\!\lvert\!\lvert\cdot\rvert\!\rvert\!\rvert_{fix,p}$*.*

This can be applied recursively to obtain the concentration for a multivariate polynomial of random matrices.

Proposition VII.3.1 (Concentration for matrix function)**.** *For a matrix-valued function* $\boldsymbol{F}(\boldsymbol{X}_m, \cdots, \boldsymbol{X}_1)$*, with matrix-valued variables* $\boldsymbol{X}_i$*. Then*

$$\lvert\!\lvert\!\lvert \boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) \rvert\!\rvert\!\rvert_*^2 \le \sum_{S\subset\{m,\cdots,1\}} (C_p)^{|S|} \left\lvert\!\left\lvert\!\left\lvert \prod_{s\in S}(1-\mathbb{E}_s) \prod_{s'\in S^c} (\mathbb{E}_{s'}) \boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) \right\rvert\!\right\rvert\!\right\rvert_*^2 \tag{7.4}$$

where $\mathbb{E}_s$ *denotes expectation over* $\boldsymbol{X}_s$ *and* S^c *denotes the complement of set* S*.*

Note the expectation $\mathbb{E}_s$ should not be confused with conditional expectation.

Proof. By induction, start with $m=1$:

$$\lvert\!\lvert\!\lvert \boldsymbol{F}(\boldsymbol{X}_1) \rvert\!\rvert\!\rvert_*^2 \le \lvert\!\lvert\!\lvert \mathbb{E}_1\boldsymbol{F}(\boldsymbol{X}_1) \rvert\!\rvert\!\rvert_*^2 + C_p \lvert\!\lvert\!\lvert (1-\mathbb{E}_1)\boldsymbol{F}(\boldsymbol{X}_1) \rvert\!\rvert\!\rvert_*^2 = \sum_{S\subset\{1\}} (C_p)^{|S|} \left\lvert\!\left\lvert\!\left\lvert \prod_{s\in S}(1-\mathbb{E}_s) \prod_{s'\in S^c} (\mathbb{E}_{s'}) \boldsymbol{F}(\boldsymbol{X}_1) \right\rvert\!\right\rvert\!\right\rvert_*^2 \tag{7.5}$$

where we used the proposition for $\mathbb{E}_1[(1-\mathbb{E}_1)\boldsymbol{F}(\boldsymbol{X}_1)] = 0$, without the need of conditioning. Given the result for m for arbitrary function, prove for $m+1$:

$$\begin{aligned}
\left\lvert\!\left\lvert\!\left\lvert \boldsymbol{F}\big(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1\big) \right\rvert\!\right\rvert\!\right\rvert_*^2 &\le \left\lvert\!\left\lvert\!\left\lvert \mathbb{E}_{m+1}\boldsymbol{F}\big(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1\big) \right\rvert\!\right\rvert\!\right\rvert_*^2 + C_p \left\lvert\!\left\lvert\!\left\lvert (1-\mathbb{E}_{m+1})\boldsymbol{F}\big(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1\big) \right\rvert\!\right\rvert\!\right\rvert_*^2 \\
&\le \sum_{S\subset\{m,\cdots,1\}} (C_p)^{|S|} \left\lvert\!\left\lvert\!\left\lvert \prod_{s\in S}(1-\mathbb{E}_s) \prod_{s'\in S^c} (\mathbb{E}_{s'})\mathbb{E}_{m+1}\boldsymbol{F}\big(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1\big) \right\rvert\!\right\rvert\!\right\rvert_*^2 \\
&+ \sum_{S\subset\{m,\cdots,1\}} (C_p)^{|S|+1} \left\lvert\!\left\lvert\!\left\lvert \prod_{s\in S}(1-\mathbb{E}_s) \prod_{s'\in S^c} (\mathbb{E}_{s'})(1-\mathbb{E}_{m+1})\boldsymbol{F}\big(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1\big) \right\rvert\!\right\rvert\!\right\rvert_*^2 \\
&= \sum_{S\subset\{m+1,m,\cdots,1\}} (C_p)^{|S|} \left\lvert\!\left\lvert\!\left\lvert \prod_{s\in S}(1-\mathbb{E}_s) \prod_{s'\in S^c} (\mathbb{E}_{s'})\boldsymbol{F}(\boldsymbol{X}_{m+1},\boldsymbol{X}_i,\ldots,\boldsymbol{X}_1) \right\rvert\!\right\rvert\!\right\rvert_*^2 .
\end{aligned}$$

First, we centered the conditional expectation $\mathbb{E}[(1-\mathbb{E}_{m+1})\boldsymbol{F}|\boldsymbol{X}_i,\cdots,\boldsymbol{X}_1] = \mathbb{E}_{m+1}(1-\mathbb{E}_{m+1})\boldsymbol{F} = 0$ and used Fact VII.3. Second, plug in the induction hypothesis for functions $\mathbb{E}_{m+1}\boldsymbol{F}(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)$ and $(1-\mathbb{E}_{m+1})\boldsymbol{F}(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)$.[15] Lastly, rearrange the expectations to complete the induction. □

To give a concrete example, we take $\boldsymbol{F}$ to be multi-linear.

---

[15] Note that the function depends on variable $\boldsymbol{X}_{m+1}$. Rigorously we would use induction hypothesis conditioned on $\boldsymbol{X}_{m+1}$ and then take expectation over $\boldsymbol{X}_{m+1}$.

Corollary VII.3.1 (multi-linear functions of bounded matrices)**.** *Let* $\boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)=\sum_{i_r,\cdots,i_i} T_{i_r,\cdots,i_1}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}$ *a degree* r *multi-linear polynomial. Suppose each* $\boldsymbol{X}_i$ *is zero mean* $\mathbb{E}\boldsymbol{X}_i=0$*, bounded* $\|\boldsymbol{X}_i\|\le b_i$*, and independent. Then*

$$\left\|\left\|\sum_{\boldsymbol{i}} T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2 \le (C_p)^r\sum_{S\subset\{m,\cdots,1\}}\left\|\left\|\sum_{\boldsymbol{i}}\mathbb{1}(\boldsymbol{i}\sim S)T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2 \tag{7.6}$$

$$\le (C_p)^r\sum_{S\subset\{m,\cdots,1\}}\left(b_{s_r}\cdots b_{s_1}\sum_{\boldsymbol{i}}\mathbb{1}(\boldsymbol{i}\sim S)\left|T_{\boldsymbol{i}}\right|\right)^2 \left\|\left\|\boldsymbol{I}\right\|\right\|_*^2 \tag{7.7}$$

where $\mathbb{1}(\boldsymbol{i}\sim S)$ *indicates the tuple* $\boldsymbol{I}=i_r,\cdots,i_1$ *coincides (up to relabeling) with set* $S=\{s_r,\cdots,s_1\}$*.*

Intuitively, the sum over different sets S exhibits a sum-of-squares behavior. Within each set S, the reordering of the polynomial is summed via a triangle inequality ($\mathbb{1}(\boldsymbol{i}\sim S)T_{\boldsymbol{i}}$), reflecting the fact we are treated the matrices $\boldsymbol{X}_i$ by their scalar absolute bound b_i. This may seem wasteful to matrix concentration specialists but turns out to be a mild overhead for our applications.

Proof. By Proposition VII.3.1,

$$\left\|\left\|\sum_{\boldsymbol{i}} T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2 \le \sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\left\|\left\|\prod_{s\in S}(1-\mathbb{E}_s)\prod_{s'\in S^c}(\mathbb{E}_{s'})\sum_{\boldsymbol{i}}T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2 \tag{7.8}$$

$$=\sum_{S\subset\{m,\cdots,1\}}(C_p)^r\left\|\left\|\sum_{\boldsymbol{i}}\mathbb{1}(\{i_r,\ldots,i_1\}=S)T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2. \tag{7.9}$$

In the second line, we used the expectation vanishing for $i\in S^c$ to convert to indicator. Lastly, use $\|\|\cdot\|\|_*$ is operator ideal (Fact V.3) to convert to the bounds on matrices as advertised. ☐

1. Deterministic matrix with Gaussian coefficients

Thus far, we have shown for a polynomial of random bounded, zero mean matrices. In physics (such as the SYK model), randomness often comes in via adding Gaussian coefficients to a deterministic matrix.

Proposition VII.3.2. *Consider random matrices* $\boldsymbol{X},\boldsymbol{Y}$ *of the same size and a Gaussian* $g\sim\mathcal{N}(0,1)$ *being independent of* $\boldsymbol{Y},\boldsymbol{X}$*. For* $2\le p$*,*

$$\|\|\boldsymbol{X}+g\boldsymbol{Y}\|\|_*^2\le\|\|\boldsymbol{X}\|\|_*^2+C_p\|\|\boldsymbol{Y}\|\|_*^2. \tag{7.10}$$

Proof. By central limit theorem, represent Gaussian as i.i.d Rademachers $g=\lim_{N\to\infty}\sum_i^N\frac{\epsilon_i}{\sqrt{N}}$.

$$\|\|\boldsymbol{X}+g\boldsymbol{Y}\|\|_*^2=\left\|\left\|\boldsymbol{X}+(\lim_{N\to\infty}\sum_i^N\frac{\epsilon_i}{\sqrt{N}})\boldsymbol{Y}\right\|\right\|_*^2=\lim_{N\to\infty}\left\|\left\|\boldsymbol{X}+(\sum_i^N\frac{\epsilon_i}{\sqrt{N}})\boldsymbol{Y}\right\|\right\|_*^2 \tag{7.11}$$

$$\le\|\|\boldsymbol{X}\|\|_*^2+\lim_{N\to\infty}\sum_i^N\frac{1}{N}C_p\|\|\boldsymbol{Y}\|\|_*^2=\|\|\boldsymbol{X}\|\|_*^2+C_p\|\|\boldsymbol{Y}\|\|_*^2. \tag{7.12}$$

Note, this is better than directly applying Fact VII.3

$$\|\|\boldsymbol{X}+g\boldsymbol{Y}\|\|_*^2\le\|\|\boldsymbol{X}\|\|_*^2+C_p\|\|g\boldsymbol{Y}\|\|_*^2=\|\|\boldsymbol{X}\|\|_*^2+C_p\mathcal{O}(p)\|\|\boldsymbol{Y}\|\|_*^2 \tag{7.13}$$

where the Gaussian moments sinfully appear $\|g\|_p^2=\mathcal{O}(p)$. ☐

It would be tempting to guess that g only needs to be subgaussian, but it is not evident from the proof. At least, one can obtain comparable results if willing to sacrifice factors of p, i.e., heavier tails. Back to the discussion, as a corollary, we can upgrade the premise to allow Gaussian coefficients.

Corollary VII.3.2 (multi-linear function of matrices with Gaussian coefficients)**.** *Let*

$$\boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)=\sum_{i_r,\cdots,i_i}T_{i_r,\cdots,i_1}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1} \tag{7.14}$$

a degree r multi-linear polynomial. Suppose each matrix has i.i.d standard Gaussian coefficient $\boldsymbol{X}_i = g_i\boldsymbol{Z}_i$ with a bounded deterministic matrix $\|\boldsymbol{Z}_i\| \le \sigma_i$. Then

$$\left|\left|\left|\sum_{\boldsymbol{i}} T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right|\right|\right|_*^2 \le (C_p)^r \sum_{S\subset\{m,\cdots,1\}} \left|\left|\left|\sum_{\boldsymbol{i}} \mathbb{1}(\boldsymbol{i}\sim S)T_{\boldsymbol{i}}\boldsymbol{Z}_{i_r}\cdots\boldsymbol{Z}_{i_1}\right|\right|\right|_*^2 \tag{7.15}$$

$$\le (C_p)^r \sum_{S\subset\{m,\cdots,1\}} \left(\sigma_{s_r}\cdots\sigma_{s_1}\sum_{\boldsymbol{i}}\mathbb{1}(\boldsymbol{i}\sim S)\,|T_{\boldsymbol{i}}|\right)^2 |||\boldsymbol{I}|||_*^2. \tag{7.16}$$

We will present two proofs.

Proof based on Proposition VII.3.2:

Proof. The proof is the same as Proposition VII.3.1. At the critical inductive step,

$$|||\boldsymbol{F}(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)|||_*^2 \le |||\mathbb{E}_{m+1}\boldsymbol{F}(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) + (1-\mathbb{E}_{m+1})\boldsymbol{F}(\boldsymbol{X}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)|||_*^2 \tag{7.17}$$

$$\le |||\boldsymbol{F}(0,\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) + g_{i+1}[\boldsymbol{F}(\boldsymbol{Z}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) - \boldsymbol{F}(0,\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)]|||_*^2 \tag{7.18}$$

$$\le |||\boldsymbol{F}(0,\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)|||_*^2 + C_p|||\boldsymbol{F}(\boldsymbol{Z}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) - \boldsymbol{F}(0,\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)|||_*^2 \tag{7.19}$$

$$\le |||(1-\mathbb{E}_{m+1})\boldsymbol{F}(\epsilon_{m+1}\boldsymbol{Z}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)|||_*^2 + C_p|||\mathbb{E}_{m+1}\boldsymbol{F}(\epsilon_{m+1}\boldsymbol{Z}_{m+1},\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)|||_*^2. \tag{7.20}$$

In the second, the multi-linear property simplifies the expression; in the third, we called Proposition VII.3.2; lastly, we conveniently insert Rademacher $\epsilon_{m+1}\boldsymbol{Z}_{m+1}$ so that the rest follows the proof of Proposition VII.3.1 and Corollary VII.3.1. Crucially, we avoided an extra factor of $\mathcal{O}(p)^r$ that could have appeared by directly plugging in Proposition VII.3.1. □

Proof based on Rademacher and central limit theorem:

Proof. We can employ the central limit theorem mindset from the ground up. For each $\boldsymbol{X}_i$, consider i.i.d Rademachers $\epsilon_{i,j}$ to converge to Gaussian.

$$\boldsymbol{X}_i = g_i\boldsymbol{Z}_i = (\lim_{N\to\infty}\sum_j^N \frac{\epsilon_{i,j}}{\sqrt{N}})\boldsymbol{Z}_i := \lim_{N\to\infty}\sum_j^N \boldsymbol{Y}_{i,j}. \tag{7.21}$$

Then,

$$\boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1) = \sum_{\boldsymbol{i}} T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1} = \sum_{j_r,\cdots,j_1}^N \sum_{\boldsymbol{i}} T_{\boldsymbol{i}}\boldsymbol{Y}_{i_r,j_r}\cdots\boldsymbol{Y}_{i_1,j_1} = h(\boldsymbol{Y}_{m,N},\cdots,\boldsymbol{Y}_{1,N},\cdots,\boldsymbol{Y}_{1,N},\cdots,\boldsymbol{Y}_{1,1}) \tag{7.22}$$

is again a multi-linear function.

$$\left|\left|\left|\sum_{j_r,\cdots,j_1}^N\sum_{\boldsymbol{i}} T_{\boldsymbol{i}}\boldsymbol{Y}_{i_r,j_r}\cdots\boldsymbol{Y}_{i_1,j_1}\right|\right|\right|_*^2 \le (C_p)^r \sum_{S'\subset\{mN,\cdots,1\}} \left|\left|\left|\sum_{\boldsymbol{j}\boldsymbol{i}} \mathbb{1}((\boldsymbol{i},\boldsymbol{j})\sim S')T_{\boldsymbol{i}}\boldsymbol{Y}_{i_r,j_r}\cdots\boldsymbol{Y}_{i_1,j_1}\right|\right|\right|_*^2 \tag{7.23}$$

$$\le (C_p)^r \sum_{S\subset\{m,\cdots,1\}}\sum_{j_{s_r}}\cdots\sum_{j_{s_1}} \left|\left|\left|\sum_{\boldsymbol{i}} \mathbb{1}(\boldsymbol{i}\sim S)T_{\boldsymbol{i}}\boldsymbol{Y}_{i_r,j_{i_r}}\cdots\boldsymbol{Y}_{i_1,j_{i_1}}\right|\right|\right|_*^2 \tag{7.24}$$

$$\le (C_p)^r \sum_{S\subset\{m,\cdots,1\}} \left(\sigma_{s_r}\cdots\sigma_{s_1}\sum_{\boldsymbol{i}}\mathbb{1}(\boldsymbol{i}\sim S)\,|T_{\boldsymbol{i}}|\right)^2 |||\boldsymbol{I}|||_*^2. \tag{7.25}$$

First, we called Corollary VII.3.1. Second, the subsets S' can be indexed by $S\subset\{m,\cdots,1\}$ and j_{s_r} for each element s_r. Once fixing the pairs $(s_r,j_{s_r}),\cdots(s_1,j_{s_1})$, the index j is a function of i and hence we only need to look for reordering of $i_r,\cdots i_1$. Also note that $T_{\boldsymbol{i}}$ is independent of $\boldsymbol{j}$. □

2. *Beyond multi-linear function*

The story was clean and straightforward for multi-linear functions, but we eventually need to go beyond that. Our following general results will be for bounded matrices (Corollary VII.3.3) and matrices with Gaussian coefficients (Theorem VII.4). The bound may appear complicated, but something roughly alike seems present even in the best-known scalar result [44, Theorem 1.4].

Corollary VII.3.3 (Polynomial of bounded matrices)**.** *Let* $\boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)=\sum_{i_r,\cdots,i_i}T_{i_r,\cdots,i_1}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}$ *a multivariate polynomial (potentially* $i_a=i_{a'}$*). Suppose each* $\boldsymbol{X}_i$ *is zero mean and bounded* $\mathbb{E}\boldsymbol{X}_i=0,\|\boldsymbol{X}_i\|\le b_i$*, and independent. Then*

$$\left\|\left\|\sum_{\boldsymbol{i}}T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2 \le \sum_{S\subset\{m,\cdots,1\}}(4C_p)^{|S|}\left(\sum_{\mathrm{Supp}(\boldsymbol{u})=S}\ \sum_{\mathrm{Supp}(\boldsymbol{v})=S^c,v_i\neq 1}b_1^{u_1+v_1}\cdots b_m^{u_m+v_m}\sum_{\pi}|T_{\pi(\boldsymbol{u},\boldsymbol{v})}|\right)^2\|\|\boldsymbol{I}\|\|_*^2$$

where π *enumerates reordering of polynomial* $\boldsymbol{X}_{b_1}^{u_1}\cdots\boldsymbol{X}_m^{u_m}\boldsymbol{X}_1^{v_1}\cdots\boldsymbol{X}_m^{v_m}$*.*

In other words, as usual, we have sum-of-square behavior across different sets S. For each set S, we first select polynomial powers $u_1,\cdots u_m$ within S, select v_i with disjoint support from S, and lastly enumerate reordering of the polynomial. Without explicit examples, this nasty bound might seem mundane. The takeaway for this calculation is that (1) the larger set $|S|$ corresponds to a heavier tail $C_p^{|S|}$, and (2) the dominating contribution often comes from larger sets S in settings with a fixed total degree and a growing number of summands. There, unevenly distributed values of $v_i\gg 2, v_i\gg 1$ suppress the count on other possible u_i,v_i.

Proof. By Proposition VII.3.1,

$$\left\|\left\|\sum_{\boldsymbol{i}}T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2\le\sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\left\|\left\|\prod_{s\in S}(1-\mathbb{E}_s)\prod_{s'\in S^c}(\mathbb{E}_{s'})\sum_{\boldsymbol{i}}T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_p^2 \tag{7.26}$$

$$\le\sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\left\|\left\|\sum_{\boldsymbol{i}}T_{\boldsymbol{i}}\left(\prod_{s\in S}(1-\mathbb{E}_s)\sum_{\mathrm{Supp}(\boldsymbol{u})=S}\right)\left(\prod_{s'\in S^c}(\mathbb{E}_{s'})\sum_{\mathrm{Supp}(\boldsymbol{v})=S^c,v_i\neq 2}\right)\mathbb{1}(\boldsymbol{i}\sim(\boldsymbol{u},\boldsymbol{v}))\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_p^2 \tag{7.27}$$

$$=\sum_{S\subset\{m,\cdots,1\}}(C_p)^{|S|}\left\|\left\|\sum_{\mathrm{Supp}(\boldsymbol{u})=S}\ \sum_{\mathrm{Supp}(\boldsymbol{v})=S^c,v_i\neq 2}\ \sum_{\pi}T_{\pi(\boldsymbol{u},\boldsymbol{v})}\pi\left((\boldsymbol{X}_1^{u_1}-\mathbb{E}\boldsymbol{X}_1^{u_1})\cdots(\boldsymbol{X}_m^{u_m}-\mathbb{E}\boldsymbol{X}_m^{u_m})\mathbb{E}\boldsymbol{X}_1^{v_1}\cdots\mathbb{E}\boldsymbol{X}_m^{v_m}\right)\right\|\right\|_*^2. \tag{7.28}$$

In the second line we inserted indicators for powers in polynomial (up to reordering) $\boldsymbol{X}_1^{u_1}\cdots\boldsymbol{X}_m^{u_m}\boldsymbol{X}_1^{v_1}\cdots\boldsymbol{X}_m^{v_m}$, with $\boldsymbol{v}$ for the powers of S and $\boldsymbol{u}$ for S^c. In the third we evaluated the expectations to give the constraint $q_c\neq 1$ and denoted by π the reordering of the non-commutative polynomial. Lastly we used $\|\|\cdot\|\|_*$ being operator ideal (Fact V.3) to convert to bounds on individual spectral norm. The factor $4^{|S|}$ is due to the crude estimate $\|\boldsymbol{X}^u-\mathbb{E}\boldsymbol{X}^u\|\le 2^u\|\boldsymbol{X}^u\|$. ☐

Next, we will use the central limit theorem to upgrade to Gaussian variables.

Theorem VII.4 (Polynomial of matrices with Gaussian coefficients)**.** *Let* $\boldsymbol{F}(\boldsymbol{X}_m,\cdots,\boldsymbol{X}_1)=\sum_{i_r,\cdots,i_i}T_{i_r,\cdots,i_1}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}$ *a degree* r *polynomial (potentially* $i_a=i_{a'}$*). Suppose each matrix has i.i.d standard Gaussian coefficient* $\boldsymbol{X}_i=g_i\boldsymbol{Z}_i$ *with a bounded deterministic matrix* $\|\boldsymbol{Z}_i\|\le\sigma_i$*. Then*

$$\left\|\left\|\sum_{\boldsymbol{i}}T_{\boldsymbol{i}}\boldsymbol{X}_{i_r}\cdots\boldsymbol{X}_{i_1}\right\|\right\|_*^2\le\sum_{\boldsymbol{u}}(C_p)^{|\boldsymbol{u}|}\left(\sum_{\boldsymbol{v}}\sigma_1^{u_1+2v_1}\cdots\sigma_i^{u_i+2v_i}\sum_{\pi}|T_{\pi(\boldsymbol{u},\boldsymbol{v})}|w(\boldsymbol{u},\boldsymbol{v})\right)^2\|\|\boldsymbol{I}\|\|_*^2 \tag{7.29}$$

where π *enumerates reordering of polynomial* $\boldsymbol{X}_1^{u_1+2v_1}\cdots\boldsymbol{X}_i^{u_i+2v_i}$ *and*

$$w(\boldsymbol{u},\boldsymbol{v})=\prod_i\frac{(u_i+2v_i-1)!!}{(u_i-1)!!}\le\prod_i(u_i+2v_i)^{v_i}. \tag{7.30}$$

The right-hand side should be understood as a variance proxy

$$\sum_{\boldsymbol{u}}\left(\sum_{\boldsymbol{v}}\sigma_1^{u_1+2v_1}\cdots\sigma_i^{u_i+2v_i}\sum_{\pi}|T_{\pi(\boldsymbol{u},\boldsymbol{v})}|w(\boldsymbol{u},\boldsymbol{v})\right)^2=\mathbb{E}\left[\left((\sum_{\boldsymbol{i}}|T_{\boldsymbol{i}}|)x_{i_r}\cdots x_{i_1}\right)^2\right] \tag{7.31}$$

where $x_i(=g_i\sigma_i)$ *are scalar Gaussians with variance* σ_i^2*.*

Intuitively, we see a sum over squares $\sum_{\boldsymbol{u}}$, but unlike the multi-linear case, we also allow the same term to have larger power $\sigma_1^{u_1}$. One may be concerned about the sum $\sum_{\boldsymbol{v}}$ *inside* the square, but looking more carefully we see each term is already squared $\sigma_i^{2v_i}$. Up to the reordering $\sum_{\boldsymbol{i}}|T_{\boldsymbol{i}}|$, this is known for scalars (Fact VII.2 replaces $(C_p)^{|\boldsymbol{u}|}$ with the uniform upper bound $(C_p)^r$.). When we replace Gaussians $x_i=g_i\sigma_i$ with $\boldsymbol{X}_i=g_i\boldsymbol{Z}_i$ for a bounded matrix $\|\boldsymbol{Z}_i\|\le\sigma_i$, uniform smoothness tells us that the analogous bound holds; we only need the degree r and an estimate of the variance proxy.

Proof. Let us painfully employ the central limit theorem as in the proof of Proposition VII.3.2. Recall

$$\boldsymbol{X}_i = g_i \boldsymbol{Z}_i = (\lim_{N\to\infty} \sum_j^N \frac{\epsilon_{i,j}}{\sqrt{N}}) \boldsymbol{Z}_i := \lim_{N\to\infty} \sum_j^N \boldsymbol{Y}_{i,j}. \tag{7.32}$$

$$\begin{aligned}
&\left|\left|\left| \sum_{j_r,\cdots,j_1}^N \sum_{\boldsymbol{i}} T_{\boldsymbol{i}} \boldsymbol{Y}_{i_r,j_r}\cdots \boldsymbol{Y}_{i_1,j_1} \right|\right|\right|_*^2 \\
&\overset{N\to\infty}{\le} \sum_{\boldsymbol{u}} (C_p)^{|\boldsymbol{u}|} \sum_{\boldsymbol{j}_1}\cdots\sum_{\boldsymbol{j}_m} \left|\left|\left| \sum_{\pi,\boldsymbol{v}} \sum_{\boldsymbol{k}_1}\cdots\sum_{\boldsymbol{k}_m} T_{\pi((\boldsymbol{u},\boldsymbol{j}),(\boldsymbol{v},\boldsymbol{k}))} \pi\Big((\boldsymbol{Y}_{1,j_{1,1}}\cdots \boldsymbol{Y}_{1,j_{1,u_1}})\cdots(\boldsymbol{Y}_{m,j_{m,1}}\cdots \boldsymbol{Y}_{m,j_{m,u_m}}) \right. \right.\right. \\
&\qquad\qquad \left.\left.\left. (\mathbb{E}\boldsymbol{Y}^2_{1,k_{1,1}}\cdots \mathbb{E}\boldsymbol{Y}^2_{1,k_{1,v_m}})\cdots(\mathbb{E}\boldsymbol{Y}^2_{m,k_{m,1}}\cdots \mathbb{E}\boldsymbol{Y}^2_{m,k_{m,v_m}})\Big) \right|\right|\right|_*^2 \\
&=: \sum_{\boldsymbol{u}} (C_p)^{|\boldsymbol{u}|} \sum_{\boldsymbol{j}_1}\cdots\sum_{\boldsymbol{j}_m} \left|\left|\left| \sum_{\pi,\boldsymbol{v}} \sum_{\boldsymbol{k}_1}\cdots\sum_{\boldsymbol{k}_m} T_{\pi((\boldsymbol{u},\boldsymbol{j}),(\boldsymbol{v},\boldsymbol{k}))} \pi\left(\boldsymbol{Y}_1^{(u_1)}\cdots \boldsymbol{Y}_m^{(u_m)} (\mathbb{E}\boldsymbol{Y}_1^2)^{(v_1)}\cdots(\mathbb{E}\boldsymbol{Y}_m^2)^{(v_m)}\right) \right|\right|\right|_*^2 && (7.33)\\
&= \sum_{\boldsymbol{u}} (C_p)^{|\boldsymbol{u}|} \left|\left|\left| \sum_{\pi,\boldsymbol{v}} T_{\pi(\boldsymbol{u},\boldsymbol{v})} w(\boldsymbol{u},\boldsymbol{v}) \pi\left(\boldsymbol{Z}_1^{u_1+2v_1}\cdots \boldsymbol{Z}_m^{u_m+2v_m}\right) \right|\right|\right|_*^2 && (7.34)\\
&\le \sum_{\boldsymbol{u}} (C_p)^{|\boldsymbol{u}|} \left(\sum_{\boldsymbol{v}} \sigma_1^{u_1+2v_1}\cdots\sigma_m^{u_m+2v_m} \sum_{\pi} |T_{\pi(\boldsymbol{u},\boldsymbol{v})}| w(\boldsymbol{u},\boldsymbol{v})\right)^2 |||\boldsymbol{I}|||_*^2. && (7.35)
\end{aligned}$$

First, we called Corollary VII.3.3. Importantly, in the large N limit, the only possible contribution would be the linear terms (e.g.$\boldsymbol{Y}_{1,j_{1,1}}$) and expected squares (e.g., $\mathbb{E}\boldsymbol{Y}^2_{1,j_{1,1}}$); any cubic term (e.g., $\boldsymbol{Y}^3_{1,j_{1,1}}$) is subleading in $1/N$. The array $\boldsymbol{u} = u_1, \cdots, u_m$ collects the number of duplicates u_i of each $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m$ and contributes to $(C_p)^{|\boldsymbol{u}|}$. Given $\boldsymbol{u}$, the sum $\sum_{\boldsymbol{j_1}} := \sum_{j_{1,1}}\cdots\sum_{j_{1,u_1}}$ runs through the duplicates $\boldsymbol{Y}_{1,j_{1,1}}\cdots\boldsymbol{Y}_{1,j_{1,u_1}}$ of $\boldsymbol{Y}_1$. Second, we compress the notation by grouping duplicates into the exponent. Third, we get rid of duplicates by summing over $\boldsymbol{j}, \boldsymbol{k}$ and counting the ways to get $\boldsymbol{Y}_1^{(u)}(\mathbb{E}\boldsymbol{Y}_1^2)^{(v)}$ from $\boldsymbol{X}_1^{u+2v}$.

$$w(u,v) := (u+2v-1)\cdot(u+2v-3)\cdots(u+1) = \frac{(u+2v-1)!!}{(u-1)!!} \le (u+2v)^v \tag{7.36}$$

$$w(\boldsymbol{u},\boldsymbol{v}) := w(u_1,v_1)\cdots w(u_m,v_m) \tag{7.37}$$

which are precisely the Wick contractions.

Lastly, to see how the RHS is the advertised variance (7.31), note that the two-point inequality (Fact I.3) becomes an equality with $C_p = 1$, which implies equality for a scalar version of Proposition VII.3.1. The rest is analogous as above. □

VIII. RANDOM k-LOCAL HAMILTONIANS

Sometimes, the object of interest is an ensemble of random Hamiltonians, such as the SYK-models. Formally, we consider random Hamiltonians with Γ k-local terms

$$\boldsymbol{H} = \sum_{\gamma=1}^{\Gamma} \boldsymbol{H}_\gamma = \sum_{\gamma=1}^{\Gamma} g_\gamma \boldsymbol{Z}_\gamma, \tag{8.1}$$

where g_γ are i.i.d standard Gaussian $\mathbb{E}[g_\gamma^2] = 1$ and $\|\boldsymbol{Z}_\gamma\| \le b_\gamma$ is a bounded deterministic matrix.

The intrinsic randomness makes its Trotter error amenable to matrix concentration treatments. This section applies the theory for multivariate polynomials of random matrices (Section VII) to prove the following theorem, slightly different from the non-random case (Section III).

Theorem VIII.1 (Trotter error in random Hamiltonians)**.** *Simulating random k-local models with Gaussian coefficients* *via higher-order Suzuki formulas, the asymptotic gate count*

$$G = \Omega\left[\left(\frac{\|\boldsymbol{H}\|^2_{(0),2}\, t\sqrt{n+\log(1/\delta)}}{\|\boldsymbol{H}\|_{(1),2}\,\epsilon}\right)^{1/\ell}\|\boldsymbol{H}\|_{(1),2}\, t\sqrt{n+\log(1/\delta)}\right]$$

$$\textit{ensures}\quad \Pr\left(\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\boldsymbol{S}(t/r)^r\|\ge\epsilon\right)\le\delta \qquad (\textit{all inputs}),$$

$$G = \Omega\left[\left(\frac{\|\boldsymbol{H}\|^2_{(0),2}\, t\sqrt{\log(1/\delta)}}{\|\boldsymbol{H}\|_{(1),2}\,\epsilon}\right)^{1/\ell}\Gamma\,\|\boldsymbol{H}\|_{(1),2}\, t\sqrt{\log(1/\delta)}\right]$$

$$\textit{ensures}\quad \Pr\left(\frac{1}{2}\left\|(\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\rho\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\boldsymbol{S}(t/r)^{\dagger r}\rho\boldsymbol{S}(t/r)^r)\right\|_1\ge\epsilon\right)\le\delta, \quad (\textit{fixed input state}),$$

where $\Pr(\cdot)$ *samples from the random Hamiltonian ensemble and* ρ *is a non-random input state.*

This is similar but stronger than the non-random k-local results (Theorem III.1): when the Hamiltonian is random, an arbitrary fixed input ρ already displays a 2-norm scaling; even the worst input states that correlate with the Hamiltonian, such as the Gibbs state, enjoys concentration with a price of $\sqrt{n}$. More carefully, the concentration here is stronger that the failure probability $(\log(1/\delta))$ has a lower power. However, the factor in $(\cdot)^{1/\ell}$ is somehow slightly worse (which is suppressed at large ℓ anyway).

The proof strategy is the same Taylor expansion as in Section III B. The calculations are combined in Section VIII C. Section (Section IX) presents an argument for lower bounds at short times.

A. Bounds on the g-th Order

We proceed by controlling each g-th order polynomial (for $\ell < g < g'$), and the product formula matches $t,\cdots,t^\ell$ orders.

$$\left|\!\left|\!\left|[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1,\cdots,\boldsymbol{H}_\Gamma,t)]_g\right|\!\right|\!\right|_*^2 = \left|\!\left|\!\left|\sum_{j=1}^{J}\sum_{g_J+\cdots+g_{j+1}=g-1}\mathcal{L}_J^{g_J}\cdots\mathcal{L}_{j+1}^{g_{j+1}}[\boldsymbol{H}_{j+1}]\frac{t^{g-1}}{g_J!\cdots g_{j+1}!}\right|\!\right|\!\right|_*^2 \tag{8.2}$$

$$\le (2t)^{2(g-1)}\sum_{\boldsymbol{u}}(C_p)^{|\boldsymbol{u}|}\left(\sum_{\boldsymbol{v}}\sigma_1^{u_1+2v_1}\cdots\sigma_\Gamma^{u_\Gamma+2v_\Gamma}w(\boldsymbol{u},\boldsymbol{v})\sum_{j,\boldsymbol{g}}\frac{\mathbb{1}(\boldsymbol{g}\sim\boldsymbol{u}+2\boldsymbol{v})}{g_J!\cdots g_{j+1}!}\right)^2 |\!|\!|\boldsymbol{I}|\!|\!|_*^2 \tag{8.3}$$

The first inequality is Theorem VII.4, and $\mathbb{1}(\boldsymbol{g}\sim\boldsymbol{u}+2\boldsymbol{v})$ indicates the occurrences of each term in $\boldsymbol{g}$, and enforces the commutation constraint, i.e., it returns zero if some $\mathcal{L}$ commute through all term on its right. This scalar sum (8.3) can be numerically evaluated to get explicit gate counts. To get analytic estimates, we proceed with the combinatorics. First, throw in extra terms to complete the exponential

$$(\textit{cont.}) \le |\!|\!|\boldsymbol{I}|\!|\!|_*^2\frac{(2\Upsilon t)^{2(g-1)}}{(g-1)!^2}\sum_{\boldsymbol{u}}(C_p)^{|\boldsymbol{u}|}\left(\sum_{\boldsymbol{v}}\sigma_1^{u_1+2v_1}\cdots\sigma_\Gamma^{u_\Gamma+2v_\Gamma}w(\boldsymbol{u},\boldsymbol{v})\sum_{\gamma_{g-1}=1}^{\Gamma}\cdots\sum_{\gamma_0=1}^{\Gamma}\mathbb{1}(\boldsymbol{\gamma}\sim\boldsymbol{u}+2\boldsymbol{v})\right)^2 \tag{8.4}$$

$$\le(\cdot)\sum_{|\boldsymbol{u}|=0}^{g}\left[\sum_{\boldsymbol{u}}(C_p)^{|\boldsymbol{u}|}\sigma_1^{2u_1}\cdots\sigma_\Gamma^{2u_\Gamma}\sum_{\boldsymbol{v}}\sigma_1^{2v_1}\cdots\sigma_\Gamma^{2v_\Gamma}w(\boldsymbol{u},\boldsymbol{v})\sum_{\boldsymbol{\gamma}}\mathbb{1}(\boldsymbol{\gamma}\sim\boldsymbol{u}+2\boldsymbol{v})\right.$$
$$\left.\cdot\max_{\boldsymbol{u}}\left(\sum_{\boldsymbol{v}'}\sigma_1^{2v'_1}\cdots\sigma_\Gamma^{2v'_\Gamma}w(\boldsymbol{u},\boldsymbol{v}')\sum_{\boldsymbol{\gamma}}\mathbb{1}(\boldsymbol{\gamma}\sim\boldsymbol{u}+2\boldsymbol{v}')\right)\right] \tag{8.5}$$

where $\frac{(\Upsilon t)^g}{(g-1)!}$ is the Taylor coefficients for exponential. The second inequality is Holder's w.r.t the sum over $\boldsymbol{u}$. For the first term,

$$\sum_{\boldsymbol{u}}(C_p)^{|\boldsymbol{u}|}\sigma_1^{2u_1}\cdots\sigma_\Gamma^{2u_\Gamma}\sum_{\boldsymbol{v}}\sigma_1^{2v_1}\cdots\sigma_\Gamma^{2v_\Gamma}w(\boldsymbol{u},\boldsymbol{v})\sum_{\boldsymbol{\gamma}}\mathbb{1}(\boldsymbol{\gamma}\sim\boldsymbol{u}+2\boldsymbol{v}) \tag{8.6}$$

$$\le g^{|\boldsymbol{v}|}\cdot g^{|\boldsymbol{v}|}\cdot(C_p)^g\cdot\sum_{\boldsymbol{u}}\sigma_1^{2u_1}\cdots\sigma_\Gamma^{2u_\Gamma}\sum_{\boldsymbol{v}}\sigma_1^{2v_1}\cdots\sigma_\Gamma^{2v_\Gamma}\sum_{\boldsymbol{\gamma}'}\mathbb{1}(\boldsymbol{\gamma}'\sim\boldsymbol{u}+\boldsymbol{v}) \tag{8.7}$$

$$\le g^{|\boldsymbol{v}|}\cdot g^{|\boldsymbol{v}|}\cdot(C_p)^g\cdot\sum_{\boldsymbol{\gamma}'}\sigma_{\gamma_0}^2\cdots\sigma_{\gamma_{g-1}}^2\mathbb{1}(\gamma'\sim \textit{Comm}) \tag{8.8}$$

$$\le g^g\cdot(C_p)^g\left(gk\,\|\boldsymbol{H}\|_{(1),2}\right)^{2(g-1-|\boldsymbol{v}|)}\|\boldsymbol{H}\|^2_{(0),2} \tag{8.9}$$

where in passing $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}'$ we first assign pairings of $\boldsymbol{v}$(which is bounded by $g^{|\boldsymbol{v}|}$) and then we sum over $\boldsymbol{\gamma}'$ (which has $|\boldsymbol{\gamma}'| = g - |\boldsymbol{v}|$ layers of sums). The other factor came from crude uniform estimates $w(\boldsymbol{u},\boldsymbol{v}) \le g^{|\boldsymbol{v}|}$. Now that these constant are outside the sum, in the second inequality we rearrange the sum in term of $\boldsymbol{\gamma}$, which was analyzed in [8], but we replace 1-norm to 2-norm $\|\boldsymbol{H}\|_{(1),1} \to \|\boldsymbol{H}\|_{(1),2}$. We used a crude bound $(g(k-1)+1) \le gk$ for the locality of commutators. For the second term,

$$\max_{\boldsymbol{u}} \left(\sum_{\boldsymbol{v}'} \sigma_1^{2v'_1} \cdots \sigma_\Gamma^{2v'_\Gamma} w(\boldsymbol{u},\boldsymbol{v}') \sum_{\boldsymbol{\gamma}} \mathbb{1}(\boldsymbol{\gamma} \sim \boldsymbol{u} + 2\boldsymbol{v}') \right) \le g^{|\boldsymbol{v}|} \cdot g^{|\boldsymbol{u}|} g^{|\boldsymbol{v}|} \cdot \max_{\boldsymbol{u}} \left(\sum_{\boldsymbol{v}'} \sigma_1^{2v'_1} \cdots \sigma_\Gamma^{2v'_\Gamma} \sum_{\boldsymbol{\gamma}''} \mathbb{1}(\boldsymbol{\gamma}'' \sim \boldsymbol{v}') \right) \tag{8.10}$$

$$\le g^{|\boldsymbol{v}|} \cdot g^{|\boldsymbol{u}|} g^{|\boldsymbol{v}|} \cdot \left(gk \|\boldsymbol{H}\|_{(1),2} \right)^{2(|\boldsymbol{v}|-1)} \|\boldsymbol{H}\|^2_{(0),2} \cdot \left(\frac{\|\boldsymbol{H}\|^2_{(1),2}}{\|\boldsymbol{H}\|^2_{(0),2}}\right)^{\mathbb{1}(|\boldsymbol{u}|=g)} \tag{8.11}$$

again, in passing $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}''$ we first select the locations $\boldsymbol{u}$, and then the pairing over $\boldsymbol{v}$. The indicator $\mathbb{1}(|\boldsymbol{u}| = g)$ takes care of the edge case $\boldsymbol{v} = 0$. Altogether, we obtain

$$|||[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_\Gamma, t)]_g|||_* \le \sqrt{C_p}^g |||\boldsymbol{I}|||_* \mathrm{e} g \cdot \left(2\mathrm{e} gk\Upsilon \|\boldsymbol{H}\|_{(1),2}\, t \right)^{g-1} \frac{\|\boldsymbol{H}\|^2_{(0),2}}{\|\boldsymbol{H}\|_{(1),2}} \tag{8.12}$$

for order $g \ge \ell + 1 \ge 3$. For $\ell = 1$ we have a sharper bound for the first order Trotter (Theorem VI.1).

B. Bounds for g'-th Order and Beyond

We have a polynomial with integral

$$\begin{aligned}
&|||[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_\Gamma, t)]_{>g}|||_* \\
&= \left|\left|\left| \sum_{j=1}^{J} \sum_{m=j+1}^{J} \mathrm{e}^{\mathcal{L}_J t} \cdots \mathrm{e}^{\mathcal{L}_{m+1} t} \int_0^t dt_1 \sum_{g_m + \cdots + g_{j+1} = g-1, g_m \ge 1} e^{\mathcal{L}_m t_1} \mathcal{L}_m^{g_m} \cdots \mathcal{L}_{j+1}^{g_{j+1}} [\boldsymbol{H}_j] \frac{(t-t_1)^{g_m - 1} t^{g' - g_m}}{(g_m - 1)! \cdots g_{j+1}!} \right|\right|\right|_* \\
&\le \sum_{m=2}^{J} \left|\left|\left| \sum_{j=m-1}^{J} \sum_{g_m + \cdots + g_{j+1} = g-1, g_m \ge 1} \mathcal{L}_m^{g_m} \cdots \mathcal{L}_{j+1}^{g_{j+1}} [\boldsymbol{H}_j] \frac{t^{g-1}}{g_m! \cdots g_{j+1}!} \right|\right|\right|_* \\
&\le |||\boldsymbol{I}|||_* \frac{(2\Upsilon t)^{(g-1)}}{\Upsilon (g-1)!} \sum_{m=2}^{J} \sqrt{ \sum_{\boldsymbol{u}} (C_p)^{|\boldsymbol{u}|} \left(\sum_{\boldsymbol{v}} \sigma_1^{u_1 + 2v_1} \cdots \sigma_\Gamma^{u_\Gamma + 2v_\Gamma} w(\boldsymbol{u},\boldsymbol{v}) \sum_{\gamma_{g-1}=1}^{\Gamma} \cdots \sum_{\gamma_0 = 1}^{\Gamma} \mathbb{1}(\boldsymbol{\gamma} \sim \boldsymbol{u} + 2\boldsymbol{v}, \gamma_{g-1} = \gamma(m)) \right)^2 }
\end{aligned}$$

Once removing the integral via unitary invariance of $|||\cdot|||_*$ norm, we can plug in Theorem VII.4. We have a similar expression except for the constraint $\gamma_{g-1} = \gamma(m)$ coming from $g_m \ge 1$, and the sum $\sum_{m=2}^{J}$ outside the square root. This amounts to minor tweaks in the calculation. For the first term,

$$\sum_{\boldsymbol{u}} (C_p)^{|\boldsymbol{u}|} \sigma_1^{2u_1} \cdots \sigma_\Gamma^{2u_\Gamma} \sum_{\boldsymbol{v}} \sigma_1^{2v_1} \cdots \sigma_\Gamma^{2v_\Gamma} w(\boldsymbol{u},\boldsymbol{v}) \sum_{\boldsymbol{\gamma}} \mathbb{1}(\boldsymbol{\gamma} \sim \boldsymbol{u} + 2\boldsymbol{v}, \gamma_{g-1} = \gamma(m)) \tag{8.13}$$

$$\le g^{|\boldsymbol{v}|} \cdot g^{|\boldsymbol{v}|} \cdot (C_p)^g \cdot \sum_{\boldsymbol{u}} \sigma_1^{2u_1} \cdots \sigma_\Gamma^{2u_\Gamma} \sum_{\boldsymbol{v}} \sigma_1^{2v_1} \cdots \sigma_\Gamma^{2v_\Gamma} \max_{0 \le j \le g-1} \sum_{\boldsymbol{\gamma}'} \mathbb{1}(\boldsymbol{\gamma}' \sim \boldsymbol{u} + \boldsymbol{v}, \gamma'_j = \gamma(m)) \tag{8.14}$$

$$\le g^{|\boldsymbol{v}|} \cdot g^{|\boldsymbol{v}|} \cdot (C_p)^g \cdot \sum_{0 \le j \le g-1} \sum_{\boldsymbol{\gamma}'} \sigma^2_{\gamma_0} \cdots \sigma^2_{\gamma_{g-1-|\boldsymbol{v}|}} \mathbb{1}(\gamma'_j = \gamma(m)) \tag{8.15}$$

$$\le g^g \cdot (C_p)^g g \left(2gk \|\boldsymbol{H}\|_{(1),2} \right)^{2(g-|\boldsymbol{v}|-1)}. \tag{8.16}$$

The only difference from (8.6) is that once a pairing of $\boldsymbol{v}$ is chosen, we lose one choice of γ', but this could happen at any γ'_j. In the second inequality, we bound $\max_j$ by $\sum_j$ to express the sum in terms of over $\boldsymbol{\gamma}'$. In the last inequality, we used

$$\sum_{\boldsymbol{\gamma}} \sigma^2_{\gamma_0} \cdots \sigma^2_{\gamma_{g-1}} \mathbb{1}(\gamma'_j = \gamma(m)) \le \sum_{i=0}^{j-2} \binom{j-2}{i} \left(k \|\boldsymbol{H}\|_{(1),2} \right)^{2(i+1)} \cdot \left(gk \|\boldsymbol{H}\|_{(1),2} \right)^{2(g-|\boldsymbol{v}|-1-(i+1))} \tag{8.17}$$

$$\le g 2^{g-1} \left(gk \|\boldsymbol{H}\|_{(1),2} \right)^{2(g-|\boldsymbol{v}|-1)}. \tag{8.18}$$

To derive this, note there must exists a subsequence $\gamma'_{s_0}\cdots\gamma'_{s_i}$ that $\gamma'_{s_0}=\gamma'_0$, $\gamma'_{s_n}\cap\gamma'_{s_{n+1}}\neq\emptyset$ and $\gamma'_{s_i}\cap\gamma'_j\neq\emptyset$, otherwise the commutator vanishes. The positions of this subsequence gives the binomial, and the choices of γ'_{s_n} gives $\left(k\left\|\boldsymbol{H}\right\|_{(1),2}\right)^{2(i+1)}$. Similarly

$$\max_{\boldsymbol{u}}\sum_{\boldsymbol{v}}\sigma_1^{2v_1}\cdots\sigma_\Gamma^{2v_\Gamma}w(\boldsymbol{u},\boldsymbol{v})\sum_{\boldsymbol{\gamma}}\mathbb{1}(\boldsymbol{\gamma}\sim\boldsymbol{u}+2\boldsymbol{v},\gamma_{g-1}=\gamma(m))\le g^{|\boldsymbol{v}|}g^{|\boldsymbol{u}|}\cdot g^{|\boldsymbol{v}|}\cdot\sum_{\boldsymbol{v}}\sigma_1^{2v_1}\cdots\sigma_\Gamma^{2v_\Gamma}\sum_{\boldsymbol{\gamma''}}\mathbb{1}(\boldsymbol{\gamma''}\sim\boldsymbol{v}) \tag{8.19}$$

$$\le g^g\cdot\sum_{\boldsymbol{\gamma''}}\sigma^2_{\gamma''_0}\cdots\sigma^2_{\gamma''_{\boldsymbol{v}-1}} \tag{8.20}$$

$$\le g^g\cdot g\left(2gk\left\|\boldsymbol{H}\right\|_{(1),2}\right)^{2(|\boldsymbol{v}|-1)}\left\|\boldsymbol{H}\right\|^2_{(0),2}. \tag{8.21}$$

In the first inequality, γ_{g-1} may or may not be in $\boldsymbol{v}$, and we upper bound by the latter case. Finally,

$$|||[\boldsymbol{\mathcal{E}}(\boldsymbol{H}_1,\cdots,\boldsymbol{H}_\Gamma,t)]_{>g}|||_*\le(C_p)^{g/2}|||\boldsymbol{I}|||_*\left(4egk\Upsilon\left\|\boldsymbol{H}\right\|_{(1),2}t\right)^{(g-1)}\frac{eg}{2k}\cdot\frac{\left\|\boldsymbol{H}\right\|_{(0),1}\left\|\boldsymbol{H}\right\|_{(0),2}}{\left\|\boldsymbol{H}\right\|_{(1),2}} \tag{8.22}$$

C. The Proof

The proof is analogous to Section III E, with minor changes.

Proof. From the bounds for each $g-th$ order (8.12) and the g'-th order (8.22), define

$$c(k):=4ek\left\|\boldsymbol{H}\right\|_{(1),2}, \tag{8.23}$$

$$c'(k):=\frac{e}{2k}\cdot\frac{\left\|\boldsymbol{H}\right\|_{(0),1}\left\|\boldsymbol{H}\right\|_{(0),2}}{\left\|\boldsymbol{H}\right\|_{(1),2}}. \tag{8.24}$$

For a short time τ, we arrange and perform the last integral using estimate $\int(\tau')^{g-1}d\tau'\le\tau'^g/g$

$$\frac{\left\|e^{i\boldsymbol{H}\tau}-\boldsymbol{S}_\ell(\tau)\right\|_p}{\left\|\boldsymbol{I}\right\|_p}\le\int_0^\tau\frac{\left\|\boldsymbol{\mathcal{E}}(\tau')\right\|_p}{\left\|\boldsymbol{I}\right\|_p}d\tau' \tag{8.25}$$

$$\le\frac{e}{c(k)}\frac{\left\|\boldsymbol{H}\right\|^2_{(0),2}}{\left\|\boldsymbol{H}\right\|_{(1),2}}\cdot\sum_{g=\ell+1}^{g'-1}\left(g^{k-1}\sqrt{C_p}c(k)\tau\right)^g+\frac{c'(k)}{c(k)}\cdot\left(g'^{k-1}\sqrt{C_p}c(k)\tau\right)^{g'} \tag{8.26}$$

$$:=c'_1\sum_{g=\ell+1}^{g'-1}\left(g^{k-1}b_p\tau\right)^g+c'_2\left(g'^{k-1}b_p\tau\right)^{g'} \tag{8.27}$$

$$\le\frac{c'_1}{1-1/e}\left((\ell+1)^{(k-1)}b_p\tau\right)^{\ell+1}+c'_2\exp\left(-\frac{1}{e(b_p\tau)^{1/(k-1)}}+1\right) \tag{8.28}$$

$$:=c_1(b_p\tau)^{\ell+1}+c_2\exp\left(-\frac{1}{e(b_p\tau)^{1/(k-1)}}\right). \tag{8.29}$$

In the second inequality we call a good value of $g'=\left\lfloor\frac{1}{e(b_p\tau)^{1/(k-1)}}\right\rfloor$. This is possible as long as

Constraint VIII.1.1. $(\frac{1}{b_p\tau})^{1/(k-1)}\ge e(\ell+3)$.

Then, the total Trotter error at a long time $t=r\cdot\tau$ is bounded by a telescoping sum

$$\frac{\left\|\boldsymbol{\mathcal{E}}_{tot}\right\|_p}{\left\|\boldsymbol{I}\right\|_p}:=\frac{\left\|e^{i\boldsymbol{H}t}-\boldsymbol{S}_\ell(t/r)^r\right\|_p}{\left\|\boldsymbol{I}\right\|_p}\le r\cdot\frac{\left\|\boldsymbol{\mathcal{E}}(t/r)\right\|_p}{\left\|\boldsymbol{I}\right\|_p}\le c_1\frac{(b_pt)^{\ell+1}}{r^\ell}+rc_2\exp\left(-\frac{1}{e}(\frac{r}{b_pt})^{1/(k-1)}\right) \tag{8.30}$$

$$\le 2c_1\frac{(b_pt)^{\ell+1}}{r^\ell}\le p^\eta 2c_1\frac{(bt)^{\ell+1}}{r^\ell}. \tag{8.31}$$

At the second line, we restrict to sufficiently large values of r that the first term dominates.

Constraint VIII.1.2. $(\frac{1}{b_p\tau})^{1/(k-1)}\ge e\ln\left(\frac{c_2}{c_1}(\frac{1}{b_p\tau})^{\ell+1}\right)$.

The last inequality isolates the p-dependence and we set $\eta := \frac{\ell+1}{2}$ and use $C_p = p-1 \leq p$. For each gate count r, pick p_r that

$$\frac{\epsilon}{\mathrm{e}^{\eta}} = p_r^{\eta} \cdot 2c_1 \frac{(bt)^{\ell+1}}{r^{\ell}} \tag{8.32}$$

which, via Markov's inequality, gives concentration.

(I) for fixed input(suffice to consider pure states $|\psi\rangle$),

$$\Pr(\left\lVert \boldsymbol{\mathcal{E}}_{tot} \left|\psi\right\rangle \right\rVert_{\ell_2} \geq \epsilon) \leq \frac{\vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{\mathcal{E}}_{tot} \vert\kern-0.25ex\vert\kern-0.25ex\vert^p_{fix,p}}{\epsilon^p} \leq \exp\left(-\frac{\eta}{\mathrm{e}}(\frac{\epsilon r^{\ell}}{2c_1'(bt)^{\ell+1}})^{1/\eta}\right) = \delta. \tag{8.33}$$

Choose r such that $p_r = \left(\frac{\epsilon r^{\ell}}{2c_1(bt)^{\ell+1}}\right)^{1/\eta}\frac{1}{\mathrm{e}} \geq \max(2, \log(1/\delta)/\eta)$, which is

$$r \geq \Omega\left(b\sqrt{\log(1/\delta)}t \cdot \left(\frac{c_1 b\sqrt{\log(1/\delta)}t}{\epsilon}\right)^{\frac{1}{\ell}}\right). \tag{8.34}$$

We also need to comply with both Constraint VIII.1.1 and Constraint VIII.1.2 by

$$r = \Omega\left(\frac{\epsilon}{c_1}\left(\ln(\frac{c_2}{c_1})\right)^{(k-1)(\ell+1)}\right). \tag{8.35}$$

Altogether,

$$r = \Omega\Bigg[\left\lVert \boldsymbol{H} \right\rVert_{(1),2}\sqrt{\log(1/\delta)}t \cdot \left(\frac{\left\lVert \boldsymbol{H} \right\rVert^2_{(0),2}\sqrt{\log(1/\delta)}t}{\left\lVert \boldsymbol{H} \right\rVert_{(1),2}\,\epsilon}\right)^{\frac{1}{\ell}} + \frac{\epsilon\left\lVert \boldsymbol{H} \right\rVert^2_{(1),2}}{\left\lVert \boldsymbol{H} \right\rVert^2_{(0),2}}\ln^{(k-1)(\ell+1)}\left(\frac{\left\lVert \boldsymbol{H} \right\rVert_{(0),1}}{\left\lVert \boldsymbol{H} \right\rVert_{(0),2}}\right)\Bigg] \tag{8.36}$$

and the first term dominates.

(II) For the spectral norm,

$$\hat{\Pr}(\left\lVert \boldsymbol{\mathcal{E}}_{tot} \right\rVert \geq \epsilon) \leq \frac{\vert\kern-0.25ex\vert\kern-0.25ex\vert \boldsymbol{\mathcal{E}}_{tot} \vert\kern-0.25ex\vert\kern-0.25ex\vert^p_{p}}{\epsilon^p} \leq d\exp\left(-\frac{\eta}{\mathrm{e}}(\frac{\epsilon r^{\ell}}{2c_1'(bt)^{\ell+1}})^{1/\eta}\right) = \delta. \tag{8.37}$$

Where the factor of dimension d came from a trace in Schatten p-norm. Choose r such that $p_r = \left(\frac{\epsilon r^{\ell}}{2c_1(bt)^{\ell+1}}\right)^{1/\eta}\frac{1}{\mathrm{e}} \geq \max(2, \log(d/\delta)/\eta)$, which is

$$r \geq \Omega\left(b\sqrt{\log(d/\delta)}t \cdot \left(\frac{c_1 b\sqrt{\log(d/\delta)}t}{\epsilon}\right)^{\frac{1}{\ell}}\right). \tag{8.38}$$

Both Constraint VIII.1.1 and Constraint VIII.1.2 are satisfied by the same (8.35). Hence,

$$r = \Omega\Bigg[\left\lVert \boldsymbol{H} \right\rVert_{(1),2}\sqrt{n+\log(1/\delta)}t \cdot \left(\frac{\left\lVert \boldsymbol{H} \right\rVert^2_{(0),2}\sqrt{n+\log(1/\delta)}t}{\left\lVert \boldsymbol{H} \right\rVert_{(1),2}\,\epsilon}\right)^{\frac{1}{\ell}} + \frac{\epsilon\left\lVert \boldsymbol{H} \right\rVert^2_{(1),2}}{\left\lVert \boldsymbol{H} \right\rVert^2_{(0),2}}\ln^{(k-1)(\ell+1)}\left(\frac{\left\lVert \boldsymbol{H} \right\rVert_{(0),1}}{\left\lVert \boldsymbol{H} \right\rVert_{(0),2}}\right)\Bigg]. \tag{8.39}$$

Again, the first term dominates.

□

IX. COUNTING LOWER BOUNDS

We give an argument for gate complexity lower bounds for k-local Hamiltonians. For a particular unitary $e^{\mathrm{i}\boldsymbol{H}t}$, it appears stubbornly hard to rule out whether a shorter circuit exists. Fortunately, for a set of unitaries, lower bounds do exist by a counting argument. In the 1d spatially local model, nt is known to be the upper and lower bounds of gate complexity:

Fact IX.1 (upper bounds, analog to digital [8, 48])**.** *For every piece-wise constant Hamiltonian*

$$\boldsymbol{H}([T, T+1]) = \boldsymbol{H}_i(T), \left\lVert \boldsymbol{H}_i \right\rVert \leq 1, \tag{9.1}$$

product formula approximates it well using $e^{(\tilde{\mathcal{O}}(nt))}$ *gates*

Fact IX.2 (Lower bounds, digital to analog [48]). *In the family of piece-wise constant Hamiltonian, there exists $\tilde{\mathcal{O}}(nt)$ different instances of Boolean circuits, and hence require a circuit of size $\tilde{\Omega}(nt)$.*

Now, for k-local random Hamiltonians, we have shown $e^{-\mathrm{i}\boldsymbol{H}t}$ can be simulated with gate complexity of $\mathcal{O}(n^k \left\|\boldsymbol{H}\right\|_{(1),2} t)$ for a fixed input state. Is the factor of the number of Hamiltonian terms $\Gamma = n^4$ a feature or a bug? We conjecture it is the former.

Hypothesis IX.1. *Simulation of a typical sample of random k-local (SYK normalization) Hamiltonian for time t requires $\tilde{\Omega}(n^k t)$ gates.*

We present a supportive early time argument. Consider the random Hamiltonian drawn randomly

$$\boldsymbol{H}_{k,n} := \sum_{i_1<\ldots<i_k\le n} \boldsymbol{H}_{i_1\cdots i_k} = \sum_{i_1<\ldots<i_k\le n} J_{i_1\cdots i_k}\boldsymbol{Z}_{i_1\cdots i_k}, \tag{9.2}$$

where the k-local matrices are o.n. $\mathrm{Tr}(\boldsymbol{H}_{i_1\cdots i_k}\boldsymbol{H}_{i'_1\cdots i'_k}) = \delta_{\boldsymbol{i},\boldsymbol{i}'}D$, and D is the dimension of Hilbert space. The number of terms is $\Gamma = \binom{n}{k} = \mathcal{O}(n^k)$. The coefficients are i.i.d Gaussian[16] with variance

$$\mathbb{E}[J_{\boldsymbol{i}}^2] = \frac{J^2(k-1)!}{kn^{k-1}} = \mathcal{O}(\frac{1}{n^{k-1}}). \tag{9.3}$$

The guidance question is, how many different Hamiltonians are there? More precisely, we aim to control the size of epsilon net $N(\epsilon)$ via collision probability. Draw N i.i.d sample from the random Hamiltonian and any pair collide with some probability we can upper bound $\Pr(\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_\infty < \epsilon)$. Then, take a union bound over the chance that pair of random samples collide

$$\Pr\left(\exists \boldsymbol{H},\boldsymbol{H}' : \left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_\infty < \epsilon\right) \le \binom{N}{2}\Pr(\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_\infty < \epsilon). \tag{9.4}$$

So long as RHS < 1, then there must exist an epsilon-net of size N, i.e., $N(\epsilon)$ can be as large as

$$N(\epsilon) = \lfloor\sqrt{2/\Pr(\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_\infty < \epsilon)}\rfloor. \tag{9.5}$$

To bound the RHS, we reduce to controlling the 2-norm

$$\Pr(\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_\infty < \epsilon) \le \Pr(\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_2 < \epsilon\sqrt{D}), \tag{9.6}$$

where the dimension of Hilbert space D is not dangerous and will be canceled. The 2-norm calculation is a scalar concentration bound

$$\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_2^2 = \left\|\sum_{\boldsymbol{i}}(J_{\boldsymbol{i}}-J'_{\boldsymbol{i}})\boldsymbol{H}_{\boldsymbol{i}}\right\|_2^2 = \sum_{\boldsymbol{i}}(J_{\boldsymbol{i}}-J'_{\boldsymbol{i}})^2 D \simeq 2\sum_{\boldsymbol{i}} J_{\boldsymbol{i}}^2 D. \tag{9.7}$$

In the last line we use that two i.i.d Gaussians sums to another Gaussian. We can use Bernsteins' inequality for variables $x_{\boldsymbol{i}} := J_{\boldsymbol{i}}^2$,

$$\Pr\left(|\sum_{\boldsymbol{i}} x_{\boldsymbol{i}} - \mathbb{E}\sum_{\boldsymbol{i}} x_{\boldsymbol{i}}| \ge \delta\right) \le 2\exp\left(\frac{-\delta^2/2}{v+\Gamma\delta/3}\right) \tag{9.8}$$

For our parameters,

$$\begin{aligned}\Pr(\left\|\boldsymbol{H}-\boldsymbol{H}'\right\|_2 < \epsilon\sqrt{D}) \le \Pr(\sum_{\boldsymbol{i}} J_{\boldsymbol{i}}^2 \le \epsilon^2/2) &\le \Pr\left(|\sum_{\boldsymbol{i}} J_{\boldsymbol{i}}^2 - \Omega(n)| \ge \Omega(n) - \epsilon^2/2\right)\\ &\lesssim \exp(-\Omega(n^k)) = \exp(-\Omega(\Gamma)).\end{aligned}$$

where we plugged in $\mathbb{E}\sum_{\boldsymbol{i}} x_{\boldsymbol{i}} = \mathcal{O}(n)$, $\delta = n - \epsilon/2 = \mathcal{O}(n)$, $\Gamma = \mathcal{O}(1/n^{k-1})$, and $v = \mathcal{O}(n^k/n^{2(k-1)}) = \mathcal{O}(1/n^{k-2})$. To translate this into estimates of circuit low bound, consider random unitary evolutions up to a short (say $t_* \sim \theta(1)$, [17]) .

$$\Pr(\left\|e^{\mathrm{i}\boldsymbol{H}t_*} - e^{\mathrm{i}\boldsymbol{H}'t_*}\right\|_\infty \le \epsilon) \overset{?}{\lesssim} \Pr(\left\|(\boldsymbol{H}-\boldsymbol{H}')t^*\right\|_\infty < \epsilon) \le \exp(\Omega(-\Gamma)). \tag{9.9}$$

Unfortunately, there is still a missing step from the Hamiltonian to the unitary; the second inequality is rigorous. If this line holds, then a circuit of size $\Omega(\Gamma) = \Omega(n^k)$ is needed, matching our Trotter bounds for non-random and random Hamiltonians (Theorem III.1, Theorem VIII.1) at early times $t = \mathcal{O}(1)$.

---

[16] The argument generalizes to sub-Gaussian coefficients, e.g., bounded coefficients.
[17] The argument also works by considering the scrambling time [26, 49, 50] for k-local models $t_* \sim \Omega(\log(n))$

Appendix A: Truncating the Hamiltonian

When the Hamiltonian has many weak terms and fewer strong terms, directly calling Trotter costs a considerable price for the number of terms Γ. An common fix is to simulate a truncated Hamiltonian with much fewer terms [3, 8]

$$\boldsymbol{H} = \boldsymbol{H}_{tr} + \delta \boldsymbol{H}. \tag{A1}$$

When the Hamiltonian is k-local (including the Fermionic cases and/or in low particle number subspaces), Hypercontractivity (Proposition II.2.4) quickly applies to the p-norm of the truncated part $\delta \boldsymbol{H}$.

For example, consider the power-law interacting Hamiltonian on a d-dimensional cube, with a total number of sites n

$$\boldsymbol{H} = \sum_{x,y} \boldsymbol{H}_{xy}, \left\|\boldsymbol{H}_{xy}\right\| \le \frac{1}{|x-y|^{\alpha}}, \tag{A2}$$

and $\alpha < d$. We truncate the Hamiltonian for distance $|x-y|$ larger than a tunable cut-off ℓ. Then

$$\left\| \mathrm{e}^{i\boldsymbol{H}t} - \boldsymbol{S}(t/r)^r \right\|_p = \left\| \mathrm{e}^{i\boldsymbol{H}t} - \mathrm{e}^{i\boldsymbol{H}_{tr}t} \right\|_p + \left\| \mathrm{e}^{i\boldsymbol{H}_{tr}t} - \boldsymbol{S}(t/r)^r \right\|_p \tag{A3}$$

$$\le t \left\| \delta \boldsymbol{H} \right\|_p + \left\| \mathrm{e}^{i\boldsymbol{H}_{tr}t} - \boldsymbol{S}(t/r)^r \right\|_p . \tag{A4}$$

Set $\ell = \theta((\frac{nt^2}{\epsilon})^{\frac{1}{2\alpha-d}})$ and drop all polynomial factors of p, then

$$t \left\| \delta \boldsymbol{H} \right\|_p = t\sqrt{n\ell^{d-2\alpha}} \lesssim \epsilon, \tag{A5}$$

and the gate complexity for typical input states would be (dropping dependence of failure probability Poly $\ln(\delta)$)

$$G = \Omega(n\ell^d \left\| \boldsymbol{H}_{tr} \right\|_{(1),2}) = \Omega(nt \cdot (\frac{nt^2}{\epsilon})^{\frac{d}{2\alpha-d}}), \tag{A6}$$

which is slightly better than [8].

---

[1] Ryan Babbush, Nathan Wiebe, Jarrod McClean, James McClain, Hartmut Neven, and Garnet Kin-Lic Chan, "Low-depth quantum simulation of materials," Phys. Rev. X **8**, 011044 (2018).

[2] Ian D. Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush, "Quantum simulation of electronic structure with linear depth and connectivity," Phys. Rev. Lett. **120**, 110501 (2018).

[3] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer, "Quantum computing enhanced computational catalysis," Physical Review Research **3** (2021), 10.1103/physrevresearch.3.033055.

[4] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan, "Quantum computational chemistry," Rev. Mod. Phys. **92**, 015003 (2020).

[5] Christopher Chamberland, Kyungjoo Noh, Patricio Arrangoiz-Arriola, Earl T. Campbell, Connor T. Hann, Joseph Iverson, Harald Putterman, Thomas C. Bohdanowicz, Steven T. Flammia, Andrew Keller, Gil Refael, John Preskill, Liang Jiang, Amir H. Safavi-Naeini, Oskar Painter, and Fernando G. S. L. Brandão, "Building a fault-tolerant quantum computer using concatenated cat codes," (2020), arXiv:2012.04108 [quant-ph].

[6] Guang Hao Low and Isaac L. Chuang, "Hamiltonian simulation by qubitization," Quantum **3**, 163 (2019).

[7] Earl Campbell, "Random compiler for fast hamiltonian simulation," Phys. Rev. Lett. **123**, 070503 (2019).

[8] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu, "Theory of trotter error with commutator scaling," Physical Review X **11** (2021), 10.1103/physrevx.11.011020.

[9] Masuo Suzuki, "General theory of fractal path integrals with applications to many-body theories and statistical physics," J. Math. Phys. **32**, 400–407 (1991).

[10] Seth Lloyd, "Universal quantum simulators," Science **273**, 1073–1078 (1996), https://science.sciencemag.org/content/273/5278/1073.full.p

[11] Burak Sahinoglu and Rolando D Somma, "Hamiltonian simulation in the low energy subspace," preprint arXiv:2006.02660 (2020).

[12] Jeongwan Haah, Matthew B. Hastings, Robin Kothari, and Guang Hao Low, "Quantum algorithm for simulating real time evolution of lattice hamiltonians," SIAM J. on Comput. , FOCS18–250–FOCS18–284 (2021).

[13] M. B. Hastings, "Turning gate synthesis errors into incoherent errors," (2016), arXiv:1612.01011 [quant-ph].

[14] Earl Campbell, "Shorter gate sequences for quantum computing by mixing unitaries," Phys. Rev. A **95**, 042306 (2017).

[15] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta, "Error mitigation for short-depth quantum circuits," Physical Review Letters **119** (2017), 10.1103/physrevlett.119.180509.

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

[17] Subir Sachdev and Jinwu Ye, "Gapless spin-fluid ground state in a random quantum heisenberg magnet," Physical Review Letters **70**, 3339–3342 (1993).

[18] Yuan Su, Hsin-Yuan Huang, and Earl T. Campbell, "Nearly tight trotterization of interacting electrons," Quantum **5**, 495 (2021).

[19] Ryan Babbush, Dominic W. Berry, and Hartmut Neven, "Quantum simulation of the sachdev-ye-kitaev model by asymmetric qubitization," Phys. Rev. A **99**, 040301 (2019).
[20] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush, "Even more efficient quantum computations of chemistry through tensor hypercontraction," PRX Quantum **2** (2021), 10.1103/prxquantum.2.030305.
[21] David J. Garling, "Inequalities: A journey into linear analysis," (2007).
[22] Nicole Tomczak-Jaegermann, "The moduli of smoothness and convexity and the rademacher averages of the trace classes $s_p(1 \leq p \leq \infty)$," Studia Mathematica **50**, 163–182 (1974).
[23] ASSAF NAOR, "On the banach-space-valued azuma inequality and small-set isoperimetry of alon–roichman graphs," Combinatorics, Probability and Computing **21**, 623–634 (2012).
[24] De Huang, Jonathan Niles-Weed, Joel A Tropp, and Rachel Ward, "Matrix concentration for products," arXiv preprint arXiv:2003.05437 (2020).
[25] Chi-Fang Chen and Andrew Lucas, "Optimal frobenius light cone in spin chains with power-law interactions," (2021), arXiv:2105.09960 [quant-ph].
[26] Chi-Fang Chen, "Concentration of otoc and lieb-robinson velocity in random hamiltonians," (2021), arXiv:2103.09186 [quant-ph].
[27] Chi-Fang Chen, Hsin-Yuan Huang, Richard Kueng, and Joel A. Tropp, "Quantum simulation via randomized product formulas: Low gate complexity with accuracy guarantees," (2020), arXiv:2008.11751 [quant-ph].
[28] Ashley Montanaro and Tobias J. Osborne, "Quantum boolean functions," (2010), arXiv:0810.2435 [quant-ph].
[29] Ashley Montanaro, "Some applications of hypercontractive inequalities in quantum information theory," Journal of Mathematical Physics **53**, 122206 (2012).
[30] Christopher King, "Hypercontractivity for semigroups of unital qubit channels," (2012), arXiv:1210.8412 [quant-ph].
[31] Mark M. Wilde, "Preface to the second edition," Quantum Information Theory , xi–xii.
[32] Salman Beigi, "Sandwiched rényi divergence satisfies data processing inequality," Journal of Mathematical Physics **54**, 122202 (2013).
[33] Eric A. Carlen and Elliott H. Lieb, "A minkowski type trace inequality and strong subadditivity of quantum entropy ii: Convexity and concavity," Letters in Mathematical Physics **83**, 107–126 (2008).
[34] Eric A. Carlen and Elliott H. Lieb, "Optimal hypercontractivity for fermi fields and related non-commutative integration inequalities," Communications in Mathematical Physics **155**, 27–46 (1993).
[35] Joel A. Tropp, "Freedman's inequality for matrix martingales," (2011), arXiv:1101.3039 [math.PR].
[36] Françoise Lust-Piquard, "Inégalites de Khintchine dans $C_p(1 < p < \infty)$. (Khinchin inequalities in C_p $(1 < p < \infty)$)," C. R. Acad. Sci., Paris, Sér. I **303**, 289–292 (1986).
[37] Françoise Lust-Piquard and Gilles Pisier, "Non commutative khintchine and paley inequalities," Arkiv för Matematik **29**, 241–260 (1991).
[38] Gilles Pisier and Quanhua Xu, "Non-commutative martingale inequalities," Communications in Mathematical Physics **189**, 667–698 (1997).
[39] Roberto Imbuzeiro Oliveira, "Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges," (2010), arXiv:0911.0600 [math.CO].
[40] Marius Junge and Qiang Zeng, "Noncommutative martingale deviation and poincarétype inequalities with applications," Probability Theory and Related Fields **161**, 449–507 (2015).
[41] Keith Ball, Eric A. Carlen, and Elliott H. Lieb, "Sharp uniform convexity and smoothness inequalities for trace norms," Inventiones mathematicae **115**, 463–482 (1994).
[42] Jeong Han Kim and Van H. Vu, "Concentration of multivariate polynomials and its applications," Combinatorica **20**, 417–434 (2000).
[43] Rafal Latala, "Estimates of moments and tails of gaussian chaoses," The Annals of Probability **34** (2006), 10.1214/009117906000000421.
[44] Warren Schudy and Maxim Sviridenko, "Concentration and moment inequalities for polynomials of independent random variables," (2012), arXiv:1104.4997 [math.PR].
[45] Ryan O'Donnell, "Analysis of boolean functions," (2021), arXiv:2105.10386 [cs.DM].
[46] Svante Janson, *Gaussian Hilbert Spaces*, Cambridge Tracts in Mathematics (Cambridge University Press, 1997).
[47] Joel A. Tropp, "An introduction to matrix concentration inequalities," (2015), arXiv:1501.01571 [math.PR].
[48] Jeongwan Haah, Matthew B. Hastings, Robin Kothari, and Guang Hao Low, "Quantum algorithm for simulating real time evolution of lattice hamiltonians," (2020), arXiv:1801.03922 [quant-ph].
[49] Nima Lashkari, Douglas Stanford, Matthew Hastings, Tobias Osborne, and Patrick Hayden, "Towards the fast scrambling conjecture," Journal of High Energy Physics **2013** (2013), 10.1007/jhep04(2013)022.
[50] Yasuhiro Sekino and L Susskind, "Fast scramblers," Journal of High Energy Physics **2008**, 065–065 (2008).