

# Quantum simulation via randomized product formulas: Low gate complexity with accuracy guarantees

Chi-Fang Chen,<sup>1,\*</sup> Hsin-Yuan (Robert) Huang,<sup>2,3,\*</sup> Richard Kueng,<sup>2,3,4</sup> and Joel A. Tropp<sup>3</sup>

<sup>1</sup>*Department of Physics, Caltech, Pasadena, CA, USA*

<sup>2</sup>*Institute for Quantum Information and Matter, Caltech, Pasadena, CA, USA*

<sup>3</sup>*Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA*

<sup>4</sup>*Institute for Integrated Circuits, Johannes Kepler University Linz, Austria*

(Dated: September 2, 2020)

Quantum simulation has wide applications in quantum chemistry and physics. Recently, scientists have begun exploring the use of randomized methods for accelerating quantum simulation. Among them, a simple and powerful technique, called qDRIFT, is known to generate random product formulas for which the *average* quantum channel approximates the ideal evolution. This work provides a comprehensive analysis of a *single realization* of the random product formula produced by qDRIFT. The main results prove that a typical realization of the randomized product formula approximates the ideal unitary evolution up to a small diamond-norm error. The gate complexity is independent of the number of terms in the Hamiltonian, but it depends on the system size and the sum of the interaction strengths in the Hamiltonian. Remarkably, the same random evolution starting from an arbitrary, but fixed, input state yields a much shorter circuit suitable for that input state. If the observable is also fixed, the same random evolution provides an even shorter product formula. The proofs depend on concentration inequalities for vector and matrix martingales. Numerical experiments verify the theoretical predictions.

## I. INTRODUCTION

Simulating complex quantum systems is one of the most promising applications for quantum computers. This task has many applications, such as developing new pharmaceuticals, catalysts, and materials [5, 15, 25], as well as solving linear algebra problems [3, 18, 32]. Finding the most efficient quantum simulation algorithm has been a prospering research field for decades; for example, see [22, 34]. Recently, there have been unprecedented advances in Hamiltonian simulation techniques both in theory and in practice [6, 7, 9, 10, 21, 23, 28]. Among these techniques, *product formulas* [34], which are also known as *Trotterization* or *splitting methods*, have undergone a renaissance [11]. These techniques are simple and intuitive, yet they are very competitive, even when compared with more sophisticated methods [11]. The purpose of this paper is to advance our understanding of randomized methods for constructing product formulas.

Consider a quantum many-body Hamiltonian with  $L$  terms:  $H = \sum_{k=1}^L h_k$ . A product formula with  $N$  gates is a sequence of  $N$  consecutive short-time evolutions by individual terms, chosen to approximate the ideal quantum evolution  $U$ :

$$U = e^{-itH} \approx e^{-it_N h_{k(N)}} \dots e^{-it_1 h_{k(1)}}, \quad (1)$$

where  $t_1, \dots, t_N$  are the short time intervals and  $k(1), \dots, k(N)$  identify which term from the Hamiltonian is applied at each step. A general technique for constructing a short-time evolution is the Lie–Suzuki–Trotter formula [33]. The first-order Suzuki approximation takes the form

$$U \approx \left( \exp(-i(tL/N)h_L) \dots \exp(-i(tL/N)h_1) \right)^{N/L}.$$

To approximate the target unitary  $U$  up to accuracy  $\epsilon$ , a total gate count  $N = \mathcal{O}(L\lambda^2 t^2 / \epsilon)$  is sufficient [11, Section 3]. Here,  $t$  is the simulation time,  $L$  is the number of terms, and  $\lambda = \sum_k \|h_k\|$  summarizes the interaction strengths within  $H$ . The unadorned norm  $\|\cdot\|$  is the spectral norm. More intricate product expansions, such as high-order Suzuki formulas, can yield gate complexities of order  $N = \mathcal{O}(L(\lambda t)^{1+o(1)} / \epsilon^{o(1)})$ . An explicit dependence on the number  $L$  of terms in the Hamiltonian seems unavoidable in general, given how these formulas are constructed and analyzed.

Recently, researchers started using randomization to improve the performance of product formulas [8, 10, 27]. This paper focuses on the qDRIFT algorithm, which was proposed by Campbell in [8].

---

\*These authors contributed equally.

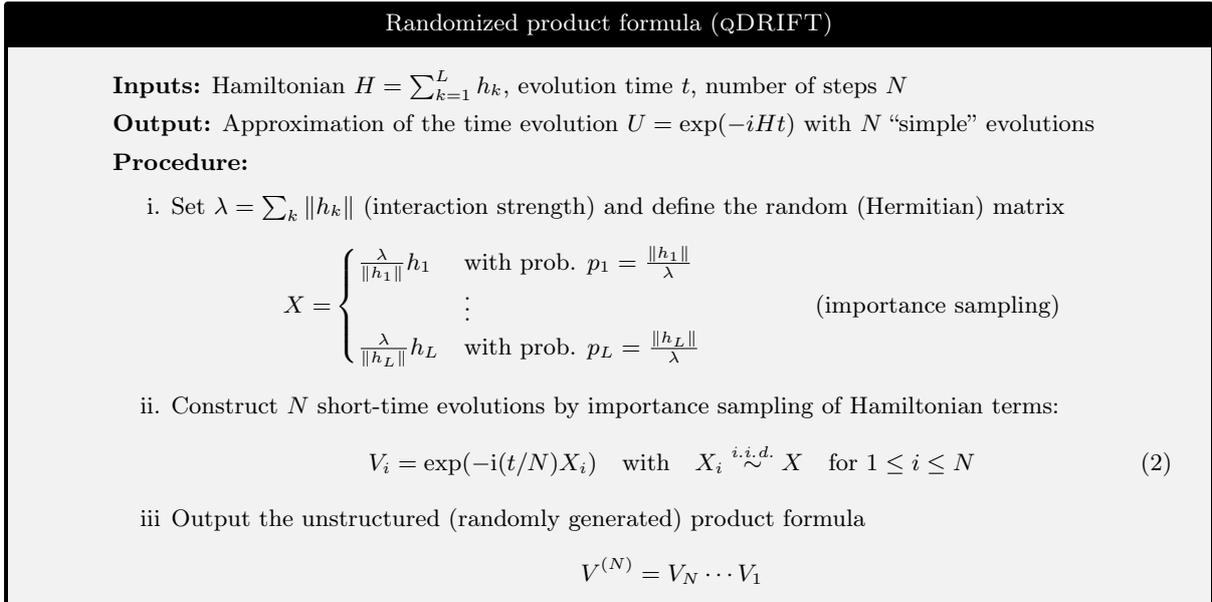


Figure 1: qDRIFT [8]: An importance sampling procedure for constructing product formulas.

This procedure uses a single, randomly selected, Hamiltonian term to approximate the whole quantum simulation up to time  $t/N$ . Subsequent time steps are approximated in an analogous fashion. The importance sampling distribution over Hamiltonian terms is designed to provide accurate approximations in expectation:  $\mathbb{E}V_i \approx U^{1/N}$ .

Arguably, this is the simplest randomized procedure for generating product formulas. It approximates a target unitary  $U$  by a product  $V_N \cdots V_1$  of random unitaries. Each  $V_i$  corresponds to a short-time evolution based on a single term  $h_K$  from the Hamiltonian. The index  $K$  is selected randomly, according to an importance sampling distribution  $(p_1, \dots, p_L)$ , constructed so that

$$\mathbb{E}[V_i] \approx \exp(-i(t/N)\mathbb{E}[h_K/p_K]) = U^{1/N}.$$

By independence,

$$\mathbb{E}[V_N \cdots V_1] = \mathbb{E}[V_N] \cdots \mathbb{E}[V_1] \approx (U^{1/N})^N = U.$$

We refer to Figure 1 for a summary of this procedure.

The main technical result in [8] establishes an error bound for the average channel of the randomized product formula (2). Operationally, Campbell considers a black box that applies a new random product  $V_N \cdots V_1$  of unitaries every time it is invoked. The black box forms a completely positive trace-preserving (CPTP) map given as the average of the possible product formulas  $\mathcal{V}^{(N)}(X) = \mathbb{E}[V_N \cdots V_1 X V_1^\dagger \cdots V_N^\dagger]$ . The ideal unitary also forms a CPTP map given as  $\mathcal{U}(X) = U X U^\dagger$ . Campbell proves that the channels  $\mathcal{V}^{(N)}$  and  $\mathcal{U}$  are  $\epsilon$ -close in diamond distance, provided that the gate count obeys  $N \geq \Omega(\lambda^2 t^2 / \epsilon)$ .

It is surprising that invoking randomness allows for a gate count that is independent of the number  $L$  of terms in the Hamiltonian. This property contrasts sharply with traditional product formulas. Indeed, for a sufficiently small gate count,  $V_N \cdots V_1$  may be too short to include every term in the Hamiltonian. Should not such a product formula deviate substantially from the ideal evolution?

In this work, we study the performance of a *random instance* of the product formula  $V_N \cdots V_1$ . Mathematically, a random product formula may be viewed as a random walk on the unitary group. We start at the identity  $\mathbb{I}$  and take small, random steps  $V_i$  in succession. Considering the worst-case behavior over all possible input states, we confirm that the averaged behavior, obtained in [8], is not at all representative for typical instances. We can exhibit a Hamiltonian where, for short evolution times, each realization of the random product formula is maximally distant from the ideal evolution. Even so, the averaged channel (i.e., Campbell’s black box) approximates the ideal evolution very well (Section III A). This behavior is reminiscent of an unbiased random walk in one dimension: although the average position remains at the origin, most trajectories end up at a point far from the origin.

Viewed from this perspective, it may seem surprising that concentration around the ideal evolution does occur as the number of time steps increases. We establish strong concentration bounds for three different use cases that are visually summarized in Figure 2. Let  $n$  denote the number of system constituents, e.g.

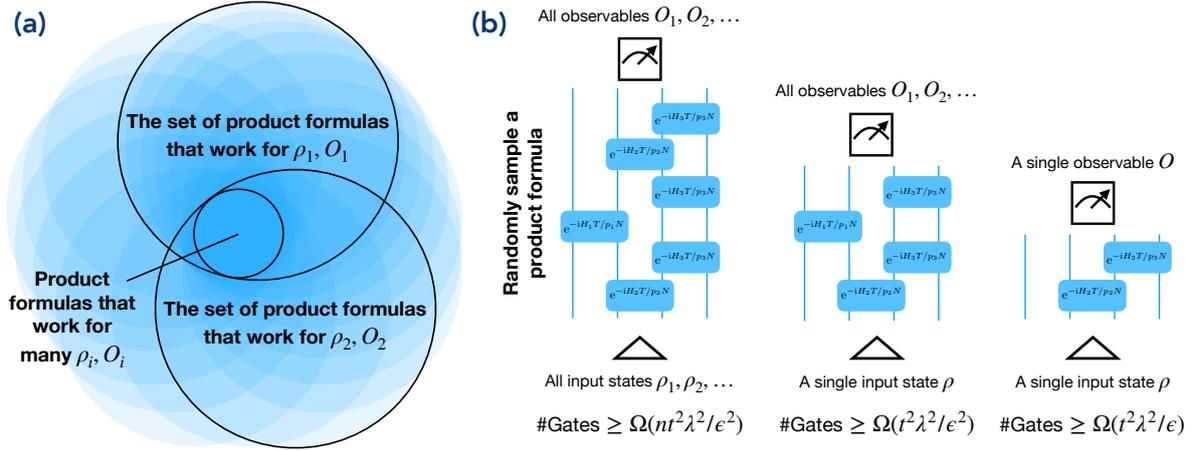


Figure 2: An illustration of the main results proved in this work.

(a): The space of all product formulas with a fixed number  $N$  of gates and can be sampled from the randomized procedure. Some of the product formulas work well for input state  $\rho_1$  and measurement observable  $O_1$ , some work for  $\rho_2, O_2$ , etc. The intersection of product formulas that work for all input states and observables is much smaller. Hence, the probability of sampling a product formula that lies in the intersection is smaller. To enlarge the intersecting region, more gates are required. The quantitative bounds for the probability of sampling a product formula within the intersection is given in (23).

(b): A pictorial summary of the main results. To sample a product formula that works for all  $n$ -qudit input states and observables *with high probability* (left), the number of gates is larger than sampling a product formula that works for a single, yet arbitrary, input state (center). This gate count can be further reduced by restricting attention to a single, yet arbitrary, observable (right).

qubits. If the gate count  $N$  obeys

$$N \geq \Omega(nt^2\lambda^2/\epsilon^2)$$

then, with high probability, a single realization of the random product formula approximates the ideal target unitary up to accuracy  $\epsilon$  in an appropriate norm. For comparison, recall that a high-order Suzuki formula involves  $N = \mathcal{O}(L(\lambda t)^{1+o(1)}/\epsilon^{o(1)})$  gates. When the simulation time  $t$  is bounded and  $L$  is very large, such as in long-range interacting systems or the SYK model [24, 30, 31], a randomly sampled product formula provides a more efficient simulation than a Suzuki formula.<sup>1</sup>

In contrast with Lie–Suzuki–Trotter formulas, randomized product formulas are *unstructured* because they are drawn from a complicated probability distribution. Nevertheless, by the probabilistic method, our analysis establishes the *existence* of product formulas whose gate count  $N$  is independent of the number  $L$  of terms in the Hamiltonian and depends only on the system size  $n$ .

In practice, we often wish to evolve a fixed input quantum state  $\rho$ , which may be arbitrary and unknown. This change in the problem statement has profound implications for randomized quantum simulation. With high probability, a random product formula with

$$N \geq \Omega(t^2\lambda^2/\epsilon^2)$$

terms suffices to achieve an  $\epsilon$ -approximation  $V_N \cdots V_1 \rho V_1^\dagger \cdots V_N^\dagger$  of the ideal time-evolved state  $U \rho U^\dagger$  with respect to trace distance. This result implies that each input state has a set of product formulas that are  $n$  times shorter than a “general-purpose” product formula that works for all input states simultaneously.

If we merely want to estimate the expectation value of a fixed observable with a fixed input state, the typical gate complexity  $N$  can be further improved to

$$N \geq \Omega(t^2\lambda^2/\epsilon).$$

<sup>1</sup> While a sample from the qDRIFT procedure provides performance guarantee on rather flexible choices of models, there is a specialized quantum algorithm for simulating the SYK models using much fewer gates [4].

Although the set of effective product formulas depends on the choice of state and observable, the formulas are all produced by the same randomized procedure without exploiting any knowledge of the particular input state or measurement observable. This is an elegant feature of randomized quantum simulation.

**Roadmap** The rest of this paper is organized as follows. Section II presents the main theoretical contributions to this work in detail. These are further supported and illustrated by numerical experiments presented in Section II D. Section III contains two instructive examples, as well as a non-technical illustration of the underlying proof idea. Details and rigorous arguments are provided in the subsequent Section IV. Finally, Section V establishes asymptotic tightness for time-evolving two simple (commuting) Hamiltonians.

## II. MAIN RESULTS

This section gives rigorous results for the error incurred by a randomly sampled product formula  $V_N \cdots V_1$ , as compared with the ideal unitary evolution operator  $U = \exp(-iHt)$ . The results depend on the distance measure and the particular setup, which we discuss separately in the following subsections.

### A. Error bound in diamond distance: Worst-case error over all input states and observables

The diamond distance is a standard distance measure for quantum channels. To compare two unitaries  $U_1$  and  $U_2$ , the diamond distance is equivalent to

$$\begin{aligned} \text{dist}_\diamond(U_1, U_2) &= \max_{|\psi\rangle: \text{state}} \left\| U_1 |\psi\rangle\langle\psi| U_1^\dagger - U_2 |\psi\rangle\langle\psi| U_2^\dagger \right\|_1 \\ &= \max_{|\psi\rangle: \text{state}} \max_{O: O^\dagger=O, \|O\| \leq 1} \left| \langle O \rangle_{U_1|\psi} - \langle O \rangle_{U_2|\psi} \right|, \end{aligned}$$

where  $\|\cdot\|_1$  is the trace norm and  $\langle O \rangle_{|\phi\rangle} = \langle \phi | O | \phi \rangle$  is the expectation value of an observable  $O$  for the quantum state  $|\phi\rangle$ . Operationally, this means that the expectation value of  $O$  evaluated on the output state would differ at most by the diamond distance between  $U_1, U_2$  for any input quantum state  $|\psi\rangle$  and any observable  $O$  with eigenvalues in  $[-1, 1]$ .

Theorem 1 bounds the gate complexity that suffices to guarantee that the randomly sampled product formula  $V_N \cdots V_1$  is close to the ideal evolution  $\exp(-itH)$  in this *worst-case* error metric. The complementary Theorem 2 gives an error bound for a given number  $N$  of gates.

**Theorem 1** (qDRIFT: Gate complexity for small diamond distance). *Consider an  $n$ -qubit Hamiltonian  $H = \sum_i h_i$  with  $\lambda = \sum_i \|h_i\|$ . Draw a randomized product formula  $V_N \cdots V_1$  from (2) with gate count*

$$N \geq \Omega \left( (n + \log(1/\delta)) t^2 \lambda^2 / \epsilon^2 \right). \quad (3)$$

With probability at least  $1 - \delta$ , the diamond distance error satisfies

$$\max_{|\psi\rangle: \text{state}} \max_{O: O^\dagger=O, \|O\| \leq 1} \left| \langle O \rangle_{V_N \cdots V_1 |\psi} - \langle O \rangle_{\exp(-itH) |\psi} \right| < \epsilon.$$

**Theorem 2** (qDRIFT: Error bound in diamond distance). *Consider an  $n$ -qubit Hamiltonian  $H = \sum_i h_i$  with  $\lambda = \sum_i \|h_i\|$ . A randomized product formula  $V_N \cdots V_1$ , drawn from (2), has expected diamond-distance error*

$$\mathbb{E} \left[ \max_{|\psi\rangle: \text{state}} \max_{O: O^\dagger=O, \|O\| \leq 1} \left| \langle O \rangle_{V_N \cdots V_1 |\psi} - \langle O \rangle_{\exp(-itH) |\psi} \right| \right] \lesssim \sqrt{\frac{nt^2 \lambda^2}{N}}.$$

The symbol  $\lesssim$  applies in the large- $N$  regime and suppresses constants. The proof sketch is presented in Section III C, and the detailed proof is given in Section IV A.

For comparison, the error bounds for the average quantum channel, established in [8], imply that

$$\mathbb{E} \left[ \max_{|\psi\rangle: \text{state}} \max_{O: O^\dagger=O, \|O\| \leq 1} \left| \mathbb{E}_{V_1, \dots, V_N} [\langle O \rangle_{V_N \cdots V_1 |\psi}] - \langle O \rangle_{\exp(-itH) |\psi} \right| \right] \lesssim \frac{t^2 \lambda^2}{N}.$$

As we can see, the error bound of the average over all product formulas is smaller than the error for an individual random product formula. To understand the discrepancy, it is valuable to think about a

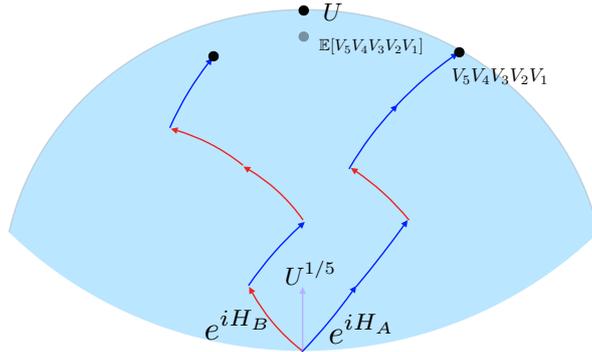


Figure 3: *Illustration of concentration effects for random walks (and their averages) on the unitary group.* The expectation  $\mathbb{E}[V_N \cdots V_1]$  of a random product formula is not unitary, but it may be very close to the ideal evolution. A sampled random product formula  $V_N \cdots V_1$  is unitary, but its distance from the ideal evolution is about  $\mathcal{O}(\sqrt{nt^2\lambda^2/N})$ . The average of the random product formulas results in an error of  $\mathcal{O}(t^2\lambda^2/N)$ .

randomly sampled product formula as a random walk on the group of  $2^n \times 2^n$  unitary matrices. Figure 3 indicates why a single realization of the random walk  $V_N \cdots V_1$  has much greater error than the average of the random walks.

In the error bound  $\mathcal{O}(\sqrt{nt^2\lambda^2/N})$ , the square root reflects the statistical nature of the fluctuations in the random walk around its expectation. The diamond norm requires us to control the behavior of the random product formula when applied to every  $2^n$ -dimensional input state. Remarkably, we only pay for the *logarithm* of the dimension, which coincides with the number  $n$  of qubits. This feature of the bound emerges naturally from the proof, which is based on concentration for matrix martingales. Similar proof techniques could potentially be useful for controlling stochastic errors in other quantum computing applications.

### B. Faster simulation for a fixed input state

In practice, it is common to perform the quantum simulation starting from a particular input state. The distinction with the previous setting is that the product formula only needs to work for one (arbitrary and possibly unknown) input state, not for all states simultaneously. The next theorem asserts that much shorter product formulas suffice in the easier setting.

**Theorem 3** (qDRIFT: Gate complexity for fixed input). *Consider an  $n$ -qubit Hamiltonian  $H = \sum_i h_i$  with  $\lambda = \sum_i \|h_i\|$  and any input quantum state  $|\psi\rangle$ . Draw a randomized product formula  $V_N \cdots V_1$  from (2) with the number of gates*

$$N \geq \Omega(t^2\lambda^2 \log(1/\delta)/\epsilon^2). \quad (4)$$

With probability at least  $1 - \delta$ , the output state  $V_N \cdots V_1 |\psi\rangle$  satisfies

$$\max_{O: O^\dagger=O, \|O\| \leq 1} |\langle O \rangle_{V_N \cdots V_1 |\psi\rangle} - \langle O \rangle_{\exp(-itH) |\psi\rangle}| < \epsilon,$$

where  $\langle O \rangle_{|\psi\rangle} = \langle \psi | O | \psi \rangle$ . This is equivalent to the output state  $V_N \cdots V_1 |\psi\rangle$  being  $\epsilon$ -close to the ideal output state  $\exp(-itH) |\psi\rangle$  in trace distance.

Theorem 3 yields an  $n$ -fold improvement over the gate count from Theorem 1. So, for a 100-qubit system, focusing on a single input state leads to a 100 $\times$  reduction in gate complexity over a simulation that works for all input states. The product formulas that work for a fixed input state may vary with the choice of state, but they are all generated using the same qDRIFT procedure. As a consequence, we can also construct short product formulas that work for a moderate number of input states by increasing the gate complexity slightly.

The proof of Theorem 3 is similar in spirit to the proof of Theorem 1. We construct a random walk from the (fixed) starting state  $|\psi\rangle$ . We show that, with high probability, the output state is close to the ideal output state  $U |\psi\rangle$ . However, to remove the system size dependence that results from matrix concentration inequalities, we instead analyze the random walk using a geometric tool, called uniform smoothness [17]. The details appear in Section IV B.

### C. Even faster simulation for fixed input states and fixed observables

We have seen that restricting attention to a fixed (but arbitrary) input state can yield considerable gate count improvements for random product formulas. It should not come as a surprise that additional savings are possible if one restricts attention to predicting a finite collection of outcome observables only (instead of demanding an accurate approximation for all observables).

In this section, we consider the task of accurately estimating  $M$  expectation values  $\langle O_j \rangle_{\exp(-itH)|\psi_j\rangle}$  associated with the time evolutions of (potentially distinct) input states  $|\psi_j\rangle$ . Estimating expectation values necessitates quantum measurements. In turn, the probabilistic nature of the measurement outcomes requires multiple samples (and, thus, repetitions of the time evolution procedure) for each pair of input and observable.

This unavoidable bottleneck motivates the following procedure: We first sample  $R$  random product formulas:  $V_N^{(1)} \dots V_1^{(1)}, \dots, V_N^{(R)} \dots V_1^{(R)}$ . For the  $j$ th pair of input state  $|\psi_j\rangle$  and observable  $O_j$ , we perform  $R$  measurement<sup>2</sup> repetitions:

$$\begin{aligned} \text{repetition 1 : } & |\psi_j\rangle \rightarrow V_N^{(1)} \dots V_1^{(1)} |\psi_j\rangle \rightarrow \text{Measure } O_j \text{ to get } \hat{o}_j^{(1)}. \\ & \vdots \\ \text{repetition } R : & |\psi_j\rangle \rightarrow V_N^{(R)} \dots V_1^{(R)} |\psi_j\rangle \rightarrow \text{Measure } O_j \text{ to get } \hat{o}_j^{(R)}. \end{aligned}$$

Subsequently, we approximate each target expectation value by the corresponding empirical average:

$$\langle O_j \rangle_{\exp(-itH)|\psi_j\rangle} \approx \frac{1}{R} \sum_{r=1}^R \hat{o}_j^{(r)} =: \hat{o}_j. \quad (5)$$

Naïvely applying Theorem 3 would require gate count  $N \geq \Omega(t^2 \lambda^2 / \epsilon^2)$  and  $R \geq \log(M) / \epsilon^2$  repetitions<sup>3</sup> to ensure that all  $M$  estimators  $\hat{o}_1, \dots, \hat{o}_M$  have additive error less than  $\epsilon$ . However, the fact that we are only measuring a particular set of observables  $O_1, \dots, O_M$  allows us to reduce the gate complexity  $N$  to  $\Omega(t^2 \lambda^2 / \epsilon)$ .

**Theorem 4** (qDRIFT: Gate complexity for fixed inputs and observables). *Consider a Hamiltonian  $H = \sum_i h_i$  with  $\lambda = \sum_i \|h_i\|$  and  $M$  (arbitrary) input-observable pairs  $(|\psi_j\rangle, O_j)$ ,  $1 \leq j \leq M$ . Set*

$$N \geq \Omega(t^2 \lambda^2 / \epsilon) \text{ and } R \geq \Omega(\log(M/\delta) / \epsilon^2). \quad (6)$$

*Draw  $R$  randomized product formulas of length  $N$  according to the qDRIFT procedure (2). Then, with probability at least  $1 - \delta$ , the corresponding empirical averages  $\hat{o}_j$  defined in Eq. (5) obey*

$$|\hat{o}_j - \langle O_j \rangle_{\exp(-itH)|\psi_j\rangle}| < \epsilon \quad \text{for all } 1 \leq j \leq M.$$

The gate complexity achieved by this result exactly reproduces prior results [8] regarding the concentration of the average channel formed by random product formulas. This is not a coincidence, as Theorem 4 may be viewed as a nontrivial consequence of the existing result. We refer to Section IV C for a derivation.

This estimation procedure is inspired by [19], and it has a similar flavor. The paper [19] shows that order  $\log(M)$  random measurements of a quantum system allow for the accurate estimation of  $M$  (arbitrary) observables. In stark contrast, a deterministic/structured product formula must make explicit use of the particular input state and measurement observable to guarantee its accuracy. We refer to Section III A for an example that illustrates this discrepancy.

### D. Numerical experiments

In this section, we perform numerical experiments for simulating a simple Heisenberg model on a one-dimensional chain with a randomly sampled product formula. For  $n$  qubits,  $H = \sum_{i=1}^{n-1} X_i X_{i+1} + Y_i Y_{i+1} +$

<sup>2</sup> We perform single shot measurement for each repetition rather than estimate the average of the observable  $O_j$ . This is equivalent to measuring in the eigenbasis of  $O_j$  to obtain a single eigenvalue.

<sup>3</sup> The factor  $1/\epsilon^2$  in  $R$  comes from the statistical noise in Monte Carlo averaging in the measurements. The factor  $\log(M)$  in  $R$  comes from union bound to ensure that all  $M$  estimates are accurate.

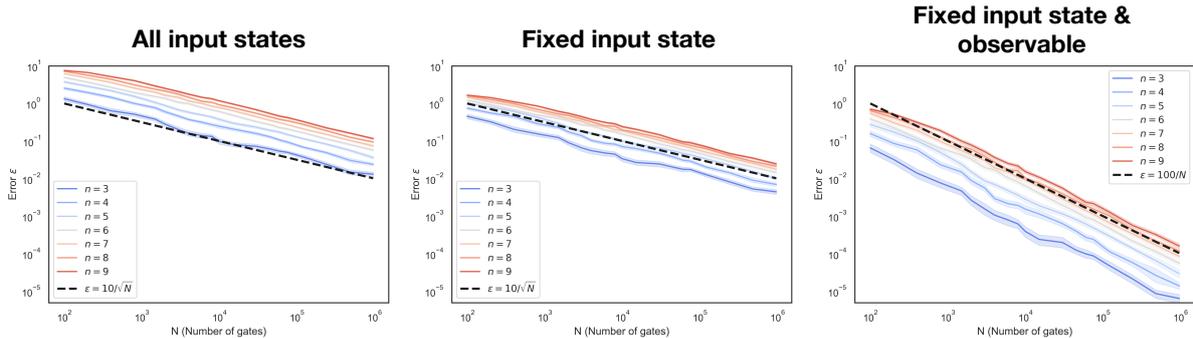


Figure 4: Numerical experiments for simulating 1D Heisenberg model under different gate count  $N$ . In *All input states* (left), we consider  $2\|U - V_N \dots V_1\|$ , which considers the error over all input states and observables. In *Fixed input state* (center), we consider the error in trace distance for the output state of a random input state. The input state is chosen to be the tensor product of single-qubit Haar-random states. In *Fixed input state & observable* (right), we consider the error in the expectation value of an observable for the output state of a random input product state. The observable is chosen to be the fidelity with the true output state. The error  $\epsilon$  for both *All input state* and *Fixed input state* are roughly  $\epsilon \propto 1/\sqrt{N}$ . The error for *Fixed input state & observable* is roughly  $\epsilon \propto 1/N$ . The shaded regions are the standard deviation over 50 independent runs.

$Z_i Z_{i+1}$  and we view this as a sum of  $3(n-1)$  simple terms. The interaction strength is  $\lambda = 3(n-1)$  and we consider constant time evolution  $t = 1$ . The numerical experiments for the error under various setups using different gate count  $N$  is given in Figure 4. First and foremost, we found that the error is proportional to  $1/\sqrt{N}$  when we consider the error over all input states (left) and a fixed input state (center). If we also fix the observable (right), the error decay rate improves to order  $1/N$ . This is in accordance with the theoretical predictions presented in the previous sections.

It is also worthwhile to briefly discuss error dependence on system size  $n$ . There, we already see considerable improvements when focusing on fixed input states (center) instead of all possible input states (left). In accordance with Theorem 3, we can use much shorter gate sequences.

### III. INSTRUCTIVE EXAMPLES AND PROOF IDEA

#### A. Comparison between stochastic averages of product formulas and concrete instances

This section considers an extremely simple Hamiltonian to pinpoint important differences between averaging random product formulas (that is, Campbell's black box) and concrete instances of product formulas. The example Hamiltonian is a 1-local non-interacting Hamiltonian with a Pauli- $Z$  operator acting on each qubit:

$$H = \frac{1}{n} \sum_{k=1}^n Z_k \quad \text{where} \quad Z_k = \underbrace{\mathbb{I} \otimes \dots \otimes \mathbb{I}}_{(k-1)\text{-times}} \otimes Z \otimes \underbrace{\mathbb{I} \otimes \dots \otimes \mathbb{I}}_{(n-k)\text{-times}} \quad \text{for } 1 \leq k \leq n \quad (7)$$

The relevant parameters are  $L = n$  (number of terms),  $\lambda = \frac{1}{n} \sum_{k=1}^n \|Z_k\| = 1$  (interaction strength) and we fix the evolution time to  $t = \pi$ .

Stochastic averages of random product formulas can accurately approximate the associated unitary evolution  $U = \exp(-i\pi H)$  after only a few iterations. The following observation is an immediate consequence of Campbell's main result [8], see also Proposition 1 below.

**Corollary 4.1.** *Fix a target accuracy  $\epsilon$  and set  $N = t^2 \lambda^2 / \epsilon \approx 10/\epsilon$ . Then,  $N$  successive applications of the QDRIFT single-step average  $\mathcal{V}(X) = \frac{1}{n} \sum_k \exp(-i\frac{\pi}{N} Z_k) \otimes \mathbb{I}^{(else)} X \exp(i\frac{\pi}{N} Z_k) \otimes \mathbb{I}^{(else)}$  (Campbell's black box) approximate the target unitary channel  $\mathcal{U}(X) = UXU^\dagger$  up to accuracy  $\epsilon$  in diamond distance. In particular,  $\frac{1}{2} \|\mathcal{V}^{(N)}(|\psi\rangle\langle\psi|) - U|\psi\rangle\langle\psi|U^\dagger\|_1 \leq \epsilon$  for all input states  $|\psi\rangle\langle\psi|$ .*

This assertion seems remarkably strong. In particular, the sequence length  $N$  does not depend on the number of qubits  $n$ . Once  $n$  is sufficiently large it becomes impossible for concrete product formulas to achieve comparable results. The problem is that the sequence length  $N$  is too small to address all  $n$  qubits. This necessarily leads to substantial discrepancies between the simulated time evolution  $V_N \dots V_1$  and the actual target  $U$ , see Figure 5 for an illustration.

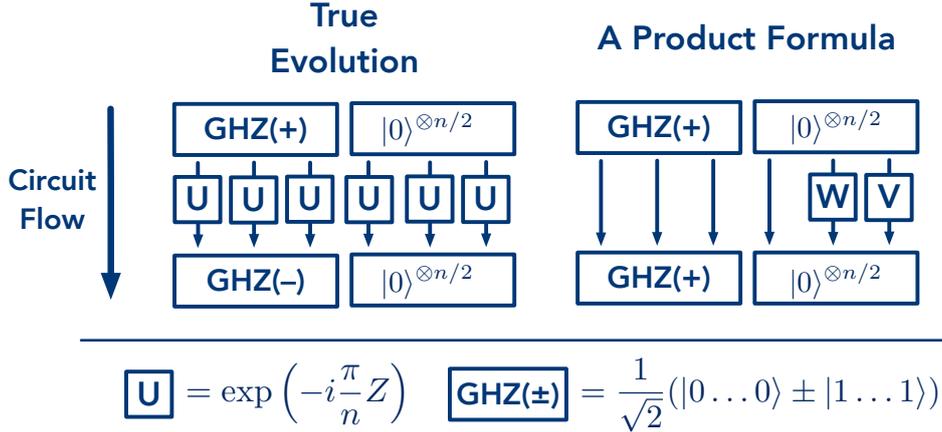


Figure 5: *Illustration of the worst-case input for a product formula simulating evolution of a simple Hamiltonian.* The Hamiltonian single-site Hamiltonian  $H = \frac{1}{n} \sum_{k=1}^n Z_k$  produces a time evolution that factorizes into single qubit unitaries  $U$  (left). A product formula with fewer than  $n/2$  single-site terms (right) is too small to address all qubits; at least  $n/2$  of them must remain untouched. These errors accumulate for a GHZ-state comprised of these untouched qubits. If  $n$  is large, even small evolution times ( $U = \exp(-i\frac{\pi}{n}Z)$ ) can accumulate and lead to a maximal approximation error ( $\langle \text{GHZ}(+), \text{GHZ}(-) \rangle = 0$ ).

**Lemma 1.** *Assume that  $n$  is an even number. It is impossible to accurately approximate the time evolution  $U$  defined in Eq. (7) with fewer than  $n/2$  elementary gates of the form  $V_i = \exp(-i\frac{\pi}{N}Z_{k(i)}) \otimes \mathbb{I}^{(else)}$ . More precisely, for each product formula  $V = V_N \cdots V_1$ , there exists an input state  $|\psi\rangle\langle\psi|$  such that  $\frac{1}{2} \|V|\psi\rangle\langle\psi|V^\dagger - U|\psi\rangle\langle\psi|U^\dagger\|_1 = 1$ .*

*Proof.* All terms in the Hamiltonian (7) commute. Hence, the associated target evolution factorizes nicely into tensor products:  $U = \exp(-i\pi H) = \exp(-i\frac{\pi}{n}Z_1) \otimes \cdots \otimes \exp(-i\frac{\pi}{n}Z_n)$ . Up to a global phase, each single-qubit unitary affects the computational basis in the following fashion:  $\exp(-i\frac{\pi}{n}Z)|0\rangle = |0\rangle$  and  $\exp(-i\frac{\pi}{n}Z)|1\rangle = \exp(i\frac{2\pi}{n})|1\rangle$ . These small phase shifts can add up for states that are in superposition. Consider the tensor product of a GHZ state on  $n/2$  qubits with the all-zeroes state on the remaining half:  $|\tilde{\psi}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n/2} + |1\rangle^{\otimes n/2}) \otimes |0\rangle^{\otimes n/2}$ . Then,

$$U|\tilde{\psi}\rangle = \exp(-i\frac{2\pi}{n}Z)^{\otimes n}|\tilde{\psi}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (\exp(i\frac{2\pi}{n}))^{n/2}|1\rangle) \otimes |0\rangle^{\otimes n/2} = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n/2} - |1\rangle^{\otimes n/2}) \otimes |0\rangle^{\otimes n/2}$$

and we can easily check that input and output are orthogonal to each other:  $\frac{1}{2} \|U|\tilde{\psi}\rangle\langle\tilde{\psi}|U^\dagger - |\tilde{\psi}\rangle\langle\tilde{\psi}|\|_1 = 1$ . These features do not change if we permute the qubits in the input state  $|\tilde{\psi}\rangle$ . Any combination of a GHZ state on one half of the qubits with computational  $|0\rangle$ -states on the remaining ones obeys the same orthogonality relation. We can use this freedom to construct a worst-case input  $|\psi\rangle$  for a fixed product formula  $V = V_N \cdots V_1$  comprised of fewer than  $n/2$  single-qubit gates. Simply initialize the (at most)  $n/2$  qubits on which the product formula acts nontrivially in the computational 0-state and hide the GHZ component among the remaining qubits. By construction, the product formula  $V$  does not affect this input state at all. This is a worst case, because the target unitary  $U$  does rotate the hidden GHZ component into an orthogonal configuration:  $\|U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1 = \|U|\psi\rangle\langle\psi|U^\dagger - |\psi\rangle\langle\psi|\|_1 = 1$ .  $\square$

This negative statement highlights that the gate count of (worst case) accurate product formulas must in general depend on the number of qubits and justifies the appearance of  $n$  in Theorem 1. Note, however, that Lemma 1 is contingent on identifying a worst-case input state for a fixed (and known) product formula. If the input state is fixed, the situation can change dramatically. For instance, we could use explicit knowledge of the input to construct a product formula that accurately approximates its time evolution. Identifying an optimal product formula seems like a daunting task, but randomness can help. Theorem 3 asserts that a collection of  $N \gtrsim \pi^2/\epsilon^2$  randomly selected single-qubit gates approximate the time evolution (7) of any fixed input state  $|\psi\rangle$  up to accuracy  $\epsilon$  in trace distance. While this gate count is considerably larger than the one put forth in Observation 4.1, it is still independent of the number of qubits. What is more, this assertion applies with high probability to *any* fixed input state. This capitalizes on another advantage of generating unstructured product formulas according to a randomized procedure: it is extremely difficult to fool a randomized compiling procedure with an already fixed input.

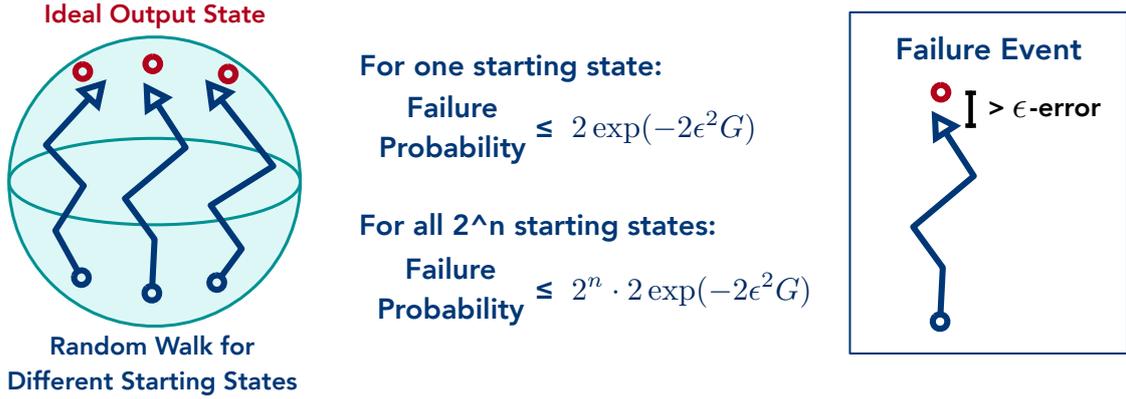


Figure 6: An illustration of the probabilistic proof for the commuting Hamiltonian given in Equation (8). We consider all the  $2^n$  computational basis states as the starting state. The probability for one of the starting state to incur at least an error  $\epsilon$  is exponentially smaller than the probability for the maximum of the  $2^n$  starting states to incur error  $> \epsilon$ . However the failure probability is exponential suppressed by increasing the gate count  $G$ . Hence one only need to set  $G = n/\epsilon^2$ .

### B. Instructive concentration argument for a simple Hamiltonian

This section provides intuition for the concentration effects that ultimately imply Theorem 1 by means of another example Hamiltonian that is composed of (commuting) Pauli-Z terms only:

$$H = \frac{1}{2^n} \sum_{\mathbf{p} \in \{0,1\}^n} \alpha_{\mathbf{p}} Z_{\mathbf{p}} \quad \text{where} \quad Z_{\mathbf{p}} = Z_{(p_1, \dots, p_n)} = Z^{p_1} \otimes \dots \otimes Z^{p_n} \quad (8)$$

(with the convention that  $Z^0 = \mathbb{I}$ ) and  $\alpha_{\mathbf{p}} \in \{-1, 1\}$ . That is, the Hamiltonian is a sum of  $2^n$  signed Pauli strings that are comprised of  $Z$  and  $\mathbb{I}$ , as well as a global sign. A high-order Suzuki formula would require a gate complexity of  $\mathcal{O}(L) = \mathcal{O}(2^n)$ . In contrast, Theorem 1 yields a gate complexity of  $\mathcal{O}(n/\epsilon^2)$ . This is an exponential improvement in terms of system size.

The physical intuition is that all the terms in the Hamiltonian act on the same system with  $n$  qubits (a  $2^n$ -dimensional Hilbert space), so their actions must overlap with one another. To see this effect more clearly, let us write down the unitary evolution  $\exp(-iH)$  in the computational basis  $|\mathbf{b}\rangle$  with multi-index  $\mathbf{b} = (b_1, \dots, b_n) \in \{0, 1\}^n$ . Note that all terms in the Hamiltonian (8) are diagonal in the computational basis. This implies

$$\exp(-iH) |\mathbf{b}\rangle = \exp\left(-i \frac{1}{2^n} \sum_{\mathbf{p} \in \{0,1\}^n} \alpha_{\mathbf{p}} Z_{\mathbf{p}}\right) |\mathbf{b}\rangle = \exp\left(-i \frac{1}{2^n} \sum_{\mathbf{p} \in \{0,1\}^n} c_{\mathbf{p}}(\mathbf{b})\right) |\mathbf{b}\rangle := e^{-iS(\mathbf{b})} |\mathbf{b}\rangle, \quad (9)$$

where  $c_{\mathbf{p}}(\mathbf{b}) = \alpha_{\mathbf{p}} \langle \mathbf{b} | Z_{\mathbf{p}} | \mathbf{b} \rangle \in \{-1, 1\}$ . When we select  $N$  random terms from the Hamiltonian (with replacement), the constructed product formula would be

$$\exp\left(-i \frac{1}{N} \alpha_{\mathbf{p}_N} Z_{\mathbf{p}_N}\right) \dots \exp\left(-i \frac{1}{N} \alpha_{\mathbf{p}_1} Z_{\mathbf{p}_1}\right) |\mathbf{b}\rangle = \exp\left(-i \frac{1}{N} \sum_k c_{\mathbf{p}_k}(\mathbf{b})\right) |\mathbf{b}\rangle := e^{-i\hat{S}(\mathbf{b})} |\mathbf{b}\rangle. \quad (10)$$

By the intuition from central limit theorem (or Hoeffding's inequality to be rigorous),  $\hat{S}(\mathbf{b}) = \frac{1}{N} \sum_k c_{\mathbf{p}_k}(\mathbf{b})$  should concentrate around  $S(\mathbf{b}) = \frac{1}{2^n} \sum_{\mathbf{p} \in \{0,1\}^n} c_{\mathbf{p}}(\mathbf{b})$  with standard deviation  $1/\sqrt{N}$  and an exponentially decaying tail. An illustration and some facts can be found in Figure 6. When  $N = 1/\epsilon^2$ , the probability of  $|\hat{S}(\mathbf{b}) - S(\mathbf{b})| > \epsilon$  would be at most  $1/e$ . And when  $N = n/\epsilon^2$ , the probability becomes exponentially suppressed to  $1/e^n$ . By a union bound,  $|\hat{S}(\mathbf{b}) - S(\mathbf{b})| \leq \epsilon$  for all  $2^n$  computational basis states  $|\mathbf{b}\rangle$  with probability at least  $1 - 2^n/e^n$ . This demonstrates that a random product formula can accurately simulate  $\exp(-iH)$  up to error  $\epsilon$  with only  $N = n/\epsilon^2$  gates, albeit in the simplest example (commuting Hamiltonian). The powerful tool of matrix concentration for matrix martingales allows us to prove the same statement for any (non-commuting) many-body Hamiltonian.

We will return to this example Hamiltonian in Section V to show that this more general analysis yields an essentially optimal parameter dependence: dimension dependence that is tight: the scaling  $N \geq \Omega(nt^2\lambda^2/\epsilon^2)$  in Theorem 1 is unavoidable in general.

### C. Proof idea for Theorem 1 and 2

This section sketches the main ideas and tools required to establish Theorem 1 and Theorem 2. The other results follow from more elementary arguments that are similar in spirit. Detailed arguments and rigorous statements are provided in Section IV below.

Consider an  $n$ -qubit Hamiltonian  $H = \sum_{i=1}^L h_i$  and an evolution time  $t$ . The associated unitary evolution defines a (unitary) channel on  $n$ -qubit states:

$$\mathcal{U}(|\psi\rangle\langle\psi|) = U|\psi\rangle\langle\psi|U^\dagger \quad \text{where} \quad U = \exp(-itH) = \exp\left(-it \sum_{k=1}^L h_k\right).$$

Fix a number of steps  $N$  and set  $\lambda = \sum_{k=1}^L \|h_k\|$ . The task is to accurately approximate the target unitary  $U$  by a product formula, i.e., the composition of  $N$  *simple* unitary evolutions:

$$\mathcal{V}^{(N)}(|\psi\rangle\langle\psi|) = \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(|\psi\rangle\langle\psi|) = V_N \dots V_1 |\psi\rangle\langle\psi| V_1^\dagger \dots V_N^\dagger.$$

We quantify the difference between  $\mathcal{V}^{(N)}$  and  $\mathcal{U}$  in *diamond distance*. That is, the worst case approximation error over all possible input states  $\rho$  in the presence of an unaffected quantum memory. Let  $\mathcal{E}, \mathcal{F}$  be two quantum channels, and let  $\mathcal{I}(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|$  denote the identity channel on an equally large ancilla system. The diamond distance between  $\mathcal{E}$  and  $\mathcal{F}$  is defined as

$$\frac{1}{2} \|\mathcal{E} - \mathcal{F}\|_\diamond = \frac{1}{2} \max_{|\psi\rangle\langle\psi|} \|\mathcal{E} \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \mathcal{F} \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1, \quad (11)$$

where the maximization ranges over all pure<sup>4</sup> input states  $|\psi\rangle\langle\psi|$  and  $\|\cdot\|_1$  denotes the trace norm. First, we relate the diamond distance (11) between the channels  $\mathcal{V}^{(N)}$  and  $\mathcal{U}$ , which are both unitary, to an operator norm distance of the associated matrices:

$$\frac{1}{2} \|\mathcal{V}^{(N)} - \mathcal{U}\|_\diamond = \frac{1}{2} \max_{|\psi\rangle\langle\psi|} \|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}^{(N)}(|\psi\rangle\langle\psi|)\|_1 \leq \|V_N \dots V_1 - U\|. \quad (\text{Lemma 2}) \quad (12)$$

This relation exploits the fact that stabilization (i.e., tensoring with the identity channel) is not necessary for computing the diamond distance of two unitary channels [40, Thm. 3.55].

Now, we can deal with the i.i.d. random matrices  $V_N, \dots, V_1$  in the more familiar operator norm. Add and subtract the expected product  $\mathbb{E}[V_N \dots V_1] = \mathbb{E}[V_N] \dots \mathbb{E}[V_1] = (\mathbb{E}V)^N$  to decompose the operator-norm difference into two qualitatively different contributions:

$$\|V_N \dots V_1 - U\| \leq \underbrace{\|(\mathbb{E}V)^N - U\|}_{\text{deterministic bias}} + \underbrace{\|V_N \dots V_1 - \mathbb{E}[V_N \dots V_1]\|}_{\text{random fluctuation}}. \quad (13)$$

These two contributions can be analyzed separately:

- i. *Deterministic bias*: Most product formulas arise from first decomposing the target unitary into a sequence of many small steps:  $U = (U^{1/N})^N$ , where  $U^{1/N} = \exp(-i(t/N)H)$  is close to the identity matrix. This allows for approximating  $U^{1/N}$  by another process that is easier to implement. The random importance sampling model (2) over individual Hamiltonian terms is designed to achieve this goal. The average approximation error scales inverse quadratically in the number of steps:  $\|(\mathbb{E}V) - U^{1/N}\| \leq t^2 \lambda^2 / N^2$ ; see Lemma 3 below. While small, this expected error does constitute a bias that is present in each of the  $N$  approximation steps. It can, and in general will, accumulate across different time steps:

$$\|\mathbb{E}[V_N \dots V_1] - U\| = \|(\mathbb{E}V)^N - (U^{1/N})^N\| \leq N \|(\mathbb{E}V) - U^{1/N}\| \leq \frac{t^2 \lambda^2}{N}. \quad (\text{Lemma 4}) \quad (14)$$

The first inequality is obtained from a telescoping sum. This upper bound diminishes as the number of steps  $N$  increases. For  $\epsilon > 0$ ,

$$N \geq \frac{2t^2 \lambda^2}{\epsilon} \quad \text{ensures} \quad \|(\mathbb{E}V)^N - U\| \leq \frac{\epsilon}{2}. \quad (15)$$

---

<sup>4</sup> Convexity ensures that the worst-case discrepancy is attained at a pure state  $|\psi\rangle\langle\psi|$ . Hence, it is not necessary to consider mixed states  $\rho$  in this definition. We refer to [40] for details.

- ii. *Random fluctuation:* We also need to control the deviation of a product of i.i.d. unitaries  $V_N \cdots V_1$  from its expectation  $\mathbb{E}[V_N \cdots V_1] = (\mathbb{E}V)^N$  in operator norm. In order to achieve this goal, we introduce a random process  $\{B_k : k = 0, \dots, N\}$  that interpolates between the extreme cases we need to compare:

$$B_k = (\mathbb{E}V)^{N-k} V_k \cdots V_1 \quad \text{such that} \quad B_0 = (\mathbb{E}V)^N \quad \text{and} \quad B_N = V_N \cdots V_1.$$

Note that adjacent elements of this process only differ in a single term:  $B_k$  arises from  $B_{k-1}$  by replacing  $\mathbb{E}V$  at position  $k$  (counted from the right) by a random realization  $V_k$  of  $V$ . The discrepancies are small:  $\|V_k - (\mathbb{E}V)\| \leq 2t\lambda/N$ , because each realization of  $V$  is very close to the identity matrix. Moreover, the entire process is causal in the sense that the current iterate  $B_k$  only depends on realizations in the past. These desirable properties endow this problem reformulation with the flavor of a random walk.

More formally, such a process forms a *matrix-valued martingale*. Powerful tail bounds for matrix-valued martingales are available in the literature [26, 35]. Adapting these results to the task at hand yields the bound

$$\Pr [\|V_N \cdots V_1 - (\mathbb{E}V)^N\| \geq \epsilon/2] \leq 2d \exp\left(-\frac{N\epsilon^2}{44t^2\lambda^2}\right). \quad (\text{Proposition 2}) \quad (16)$$

In words, the product  $V_N \cdots V_1$  will concentrate around its expectation once  $N$  is sufficiently large. Similar to more conventional random walks on integer lattices, the error is subgaussian with variance proportional to  $N \cdot (\lambda^2 t^2 / N^2)$ . There is an extra dimensional factor  $d = 2^n$  that arises because the martingale is matrix-valued; this is the origin of the factor  $n$  in the gate count  $N$ . For error parameters  $\epsilon, \delta \in (0, 1)$ ,

$$N \geq 44 \frac{t^2 \lambda^2}{\epsilon^2} \log(2d/\delta) \quad \text{implies} \quad \|V_N \cdots V_1 - (\mathbb{E}V)^N\| \leq \frac{\epsilon}{2} \quad \text{with probability} > 1 - \delta. \quad (17)$$

Theorem 1 can be derived by combining the previous results. We instantiate the bound (15) for deterministic bias and the bound (17) for random fluctuation and insert them into the problem reformulation (13). This provides a high probability error bound in the diamond distance when  $N \geq \Omega(nt^2\lambda^2/\epsilon^2)$ .

Theorem 2 is derived using similar ideas. Start with the problem reformulation (13) and use the fact that the bias bound (15) is deterministic and not affected by taking expectation values. Integrating the tail bound (16) over  $\epsilon$  produces a bound on the expected size of random fluctuations:

$$\mathbb{E}\|V_N \cdots V_1 - (\mathbb{E}V)^N\| \lesssim \sqrt{\frac{t^2 \lambda^2}{N} \log_2(d)}.$$

The symbol  $\lesssim$  suppresses a modest multiplicative constant, and we refer to Section IV A 4 for details.

## IV. TECHNICAL DETAILS AND PROOFS

### A. Proof of Theorem 1 and 2: Approximation error under the worst-possible input

The proofs of Theorem 1 and 2 were sketched in Section III. This section contains the details. In Section IV A 1, we first relate the diamond distance to the operator norm. This allows us to work with the operator norm, which is mathematically simpler. Then we bound the two error contributions arising from the deterministic bias (in Section IV A 2), as well as random fluctuations (in Section IV A 3). Finally, we combine the two bounds to obtain a convergence guarantee for randomly sampled product formulas. This is the content of Section IV A 4.

#### 1. Conversion from diamond distance into operator norm

The diamond distance is a rather intricate object. Although it can be phrased implicitly as a semidefinite program, analytical formulas are rare and far between. A notable exception is the diamond distance between two unitary channels, which is completely understood [1, Sec. 5.3]. The first part of the following statement is a direct consequence of this characterization. The second part is based on more recent insights [40, Thm. 3.56]. We provide elementary, self-contained proofs for convenience.

**Lemma 2.** Let  $\mathcal{U}(\rho) = U\rho U^\dagger$  and  $\mathcal{V}(\rho) = V\rho V^\dagger$  be unitary channels. Then,  $\frac{1}{2}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \leq \|(U - V)|\psi\rangle\|_{\ell_2}$  for any pure state  $|\psi\rangle$ . In turn,  $\frac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond \leq \|U - V\|$ . The latter relation generalizes to averages of random unitary channels:  $\frac{1}{2}\|\mathcal{U} - \mathbb{E}[\mathcal{V}]\|_\diamond \leq \|U - \mathbb{E}[V]\|$ .

*Proof.* Fix an input  $|\psi\rangle$  and denote the output state vectors by  $|u\rangle = U|\psi\rangle$  and  $|v\rangle = V|\psi\rangle$ , respectively. Normalization ensures that these state vectors obey  $|\langle u, v\rangle| \leq 1$ , as well as  $\| |u\rangle - |v\rangle \|_{\ell_2} = \sqrt{2(1 - \operatorname{Re}(\langle u, v\rangle))}$ . Apply the Fuchs–van de Graaf relations [40, Theorem 3.33] to convert the output trace distance into a (pure) output fidelity:

$$\frac{1}{2}\| |u\rangle\langle u| - |v\rangle\langle v| \|_1 = \sqrt{1 - |\langle u, v\rangle|^2} = \sqrt{(1 + |\langle u, v\rangle|)(1 - |\langle u, v\rangle|)} \leq \| |u\rangle - |v\rangle \|_{\ell_2}.$$

The diamond distance bound then is a direct consequence of this relation. Use the fact that stabilization is not necessary for computing the diamond distance of two unitary channels to conclude

$$\frac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond = \max_{|\psi\rangle\langle\psi|} \frac{1}{2}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \leq \max_{|\psi\rangle} \|(U - V)|\psi\rangle\|_{\ell_2} = \|U - V\|.$$

Here, we have also used the definition of the operator norm. In order to handle expectation values, we need an additional argument. Let  $(p_k, V_k)$  be an ensemble of unitaries with weights  $p_k \geq 0$  that obey  $\sum_k p_k = 1$ . Then, Cauchy–Schwarz asserts

$$|\langle\psi|U^\dagger\mathbb{E}[V]|\psi\rangle|^2 = \left| \sum_k \sqrt{p_k}\sqrt{p_k}\langle\psi|U^\dagger V_k|\psi\rangle \right|^2 \leq \left( \sum_k p_k \right) \sum_k p_k |\langle\psi|U^\dagger V_k|\psi\rangle|^2 = \sum_k p_k |\langle\psi|U^\dagger V_k|\psi\rangle|^2,$$

for any unitary  $U$  and state  $|\psi\rangle$ . Combined with Fuchs–van de Graaf, this observation delivers

$$\frac{1}{2}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathbb{E}[\mathcal{V}(|\psi\rangle\langle\psi|)]\|_1 \leq \left(1 - \sum_k p_k |\langle\psi|U^\dagger V_k|\psi\rangle|^2\right)^{1/2} \leq \left(1 - |\langle\psi|U^\dagger\mathbb{E}[V]|\psi\rangle|^2\right)^{1/2}$$

for any pure input state  $|\psi\rangle$ . This is enough to conclude  $\frac{1}{2}\|U|\psi\rangle\langle\psi|U^\dagger - \mathbb{E}[V|\psi\rangle\langle\psi|V^\dagger]\|_1 \leq \|(U - \mathbb{E}[V])|\psi\rangle\|_{\ell_2}$ , much as before. We emphasize that this relation is true for any fixed unitary  $U$  and any unitary ensemble  $(p_k, V_k)$ . This flexibility is essential to deduce the diamond distance bound, because  $\mathbb{E}[\mathcal{V}]$  is not unitary and stabilization must be taken into account:

$$\begin{aligned} \|\mathcal{U} - \mathbb{E}[\mathcal{V}]\|_\diamond &= \max_{|\psi\rangle\langle\psi|} \frac{1}{2}\|\mathcal{U} \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \mathbb{E}[\mathcal{V} \otimes \mathcal{I}(|\psi\rangle\langle\psi|)]\|_1 \\ &= \max_{|\psi\rangle\langle\psi|} \frac{1}{2}\|(U \otimes \mathbb{I})|\psi\rangle\langle\psi|(U \otimes \mathbb{I})^\dagger - \mathbb{E}[(V \otimes \mathbb{I})|\psi\rangle\langle\psi|(V \otimes \mathbb{I})^\dagger]\|_1 \\ &\leq \max_{|\psi\rangle} \|(U \otimes \mathbb{I} - \mathbb{E}[V \otimes \mathbb{I}])|\psi\rangle\|_{\ell_2} = \|(U - \mathbb{E}[V]) \otimes \mathbb{I}\| = \|U - \mathbb{E}[V]\|. \end{aligned}$$

This is what we needed to show.  $\square$

## 2. Controlling the deterministic bias

Next, we establish a bound on the deterministic bias between the averaged channel and the ideal unitary evolution.

**Proposition 1.** Consider the i.i.d. unitary product constructed by the QDRIFT protocol (2) for simulating  $U = \exp(-itH)$ . Define the total strength  $\lambda = \sum_k \|h_k\|$  and the evolution time  $t$ . Then

$$\|U - \mathbb{E}[V_N \cdots V_1]\| \leq \frac{t^2\lambda^2}{N}.$$

Note that Lemma 2 allows for converting this statement into a diamond distance bound for the associated channels:

$$\frac{1}{2}\|\mathcal{U} - \mathbb{E}[\mathcal{V}_N \circ \cdots \circ \mathcal{V}_1]\|_\diamond \leq \|U - \mathbb{E}[V_N \cdots V_1]\| \leq \frac{t^2\lambda^2}{N}. \quad (18)$$

This is a slight improvement over the main technical result regarding QDRIFT [8]. Indeed, Campbell labels the total average QDRIFT channel  $\mathcal{E} = \mathbb{E}[\mathcal{V}_N \circ \cdots \circ \mathcal{V}_1]$ , and he establishes that  $\frac{1}{2}\|\mathcal{U} - \mathcal{E}\|_\diamond \leq (t^2\lambda^2/N)e^{2t\lambda/N}$  in [8, Eq. (B12)]. Both assertions become very similar in the large  $N$  limit, but (18) is always tighter and the discrepancy can be quite pronounced for small and intermediate values of  $N$ .

The proof of Proposition 1 is based on an extension of the numerical bounds  $|e^{ix} - 1| \leq |x|$  and  $|e^{ix} - ix - 1| \leq x^2/2$  for all  $x \in \mathbb{R}$  to Hermitian matrices.

**Fact 1.** Let  $X$  be Hermitian. Then we have the zero-order bound  $\|\exp(iX) - \mathbb{I}\| \leq \|X\|$  and the first-order bound  $\|\exp(iX) - iX - \mathbb{I}\| \leq \frac{1}{2}\|X\|^2$ .

These observations can be converted into accurate operator-norm bounds for the expected error of individual QDRIFT steps.

**Lemma 3.** Fix a Hamiltonian  $H = \sum_{l=1}^L H_l$  and parameters  $N, t$ . Set  $U^{1/N} = \exp(-i(t/N)H)$  and  $\lambda = \sum_{l=1}^L \|h_l\|$ . Then, the random matrix  $V$  defined in (2) obeys

$$\|V - (\mathbb{E}V)\| \leq \frac{2t\lambda}{N} \text{ (almost surely)} \quad \text{and} \quad \|(\mathbb{E}V) - U^{1/N}\| \leq \frac{t^2\lambda^2}{N^2}$$

*Proof.* Streamline the notation from Figure 1 by absorbing the scaling factor  $(t/N)$  into the random Hermitian matrix  $X$ . In particular,  $V = \exp(-iX)$ ,  $\mathbb{E}V = \mathbb{E}[\exp(-iX)]$ ,  $U^{1/N} = \exp(-i\mathbb{E}[X])$  and  $\|X\| = (t\lambda)/N$  almost surely. Observe that

$$\|V - (\mathbb{E}V)\| \leq \|\exp(-iX) - \mathbb{I}\| + \|\mathbb{I} - \mathbb{E}[\exp(-iX)]\| \leq \|\exp(-iX) - \mathbb{I}\| + \mathbb{E}\|\mathbb{I} - \exp(-iX)\|,$$

where the last inequality is Jensen's. Fact 1 and uniform normalization ( $\|X\| = (t\lambda)/N$ ) then imply  $\|\exp(-iX) - \mathbb{I}\| \leq \|X\| = (t\lambda)/N$  for any instance of the random matrix  $X$ . This uniform bound also covers the expected norm difference and we conclude  $\|V - (\mathbb{E}V)\| \leq 2t\lambda/N$ . The (tighter) second claim can be derived in a similar fashion. A combination of Jensen's inequality, Fact 1, and normalization delivers

$$\begin{aligned} \|(\mathbb{E}V) - U^{1/N}\| &= \|\mathbb{E}[\exp(-iX)] - \mathbb{I} + iX\| + \|\mathbb{I} - i\mathbb{E}[X] - \exp(-i\mathbb{E}[X])\| \\ &\leq \mathbb{E}\|\exp(-iX) - \mathbb{I} + iX\| + \|\exp(-i\mathbb{E}[X]) - \mathbb{I} + i\mathbb{E}[X]\| \\ &\leq \frac{1}{2}\mathbb{E}\|X\|^2 + \frac{1}{2}\|\mathbb{E}[X]\|^2 \leq \mathbb{E}\|X\|^2 = (t\lambda/N)^2. \end{aligned}$$

This is the advertised result.  $\square$

We also need a statement regarding error accumulation over several applications of similar, but not identical, linear operators. It is a rather intuitive consequence of operator norm sub-multiplicativity and the triangle inequality. See [33] for related results.

**Lemma 4.** Let  $\mathbb{E}V$  and  $U^{1/N}$  be matrices with bounded operator norm:  $\|\mathbb{E}V\| \leq 1$  and  $\|U^{1/N}\| \leq 1$ . Then

$$\|(\mathbb{E}V)^N - (U^{1/N})^N\| \leq N\|(\mathbb{E}V) - U^{1/N}\|.$$

*Proof.* The triangle inequality and sub-multiplicativity imply

$$\|A_1A_2 - B_1B_2\| = \|(A_1 - B_1)A_2 + B_1(A_2 - B_2)\| \leq \|A_2\|\|A_1 - B_1\| + \|B_1\|\|A_2 - B_2\|$$

for any matrix quadruple  $A_1, A_2, B_1, B_2$  with compatible dimensions. Use the assumed operator norm bounds to iteratively apply this relation and deduce the statement:

$$\begin{aligned} \|(\mathbb{E}V)^N - (U^{1/N})^N\| &= \|(\mathbb{E}V)(\mathbb{E}V)^{N-1} - U^{1/N}(U^{1/N})^{N-1}\| \\ &\leq \|(\mathbb{E}V) - U^{1/N}\| + \|(\mathbb{E}V)^{N-1} - (U^{1/N})^{N-1}\| \leq \dots \leq N\|(\mathbb{E}V) - U^{1/N}\|. \end{aligned}$$

This is the stated result.  $\square$

*Proof of Proposition 1.* The main result of this section immediately follows from combining Lemma 4 and Lemma 3. Decompose  $U = \exp(-itH)$  into  $N$  steps  $U^{1/N} = \exp(-i(t/N)H)$  and conclude

$$\|U - (\mathbb{E}V)^N\| = \|(U^{1/N})^N - (\mathbb{E}V)^N\| \leq N\|U^{1/N} - (\mathbb{E}V)\| \leq \frac{t^2\lambda^2}{N}.$$

This is what we had to show.  $\square$

### 3. Controlling random fluctuations

In the previous subsection we have essentially recapitulated the state of the art regarding QDRIFT: the algorithm provides an accurate approximation in expectation over all possible random choices (deterministic bias). In this section, things start to get interesting. We want to show that a single realization of QDRIFT is likely to provide a good approximation, provided that the number of steps  $N$  is sufficiently large. In order to achieve this goal, we need to show that concrete fluctuations around the (accurate) expected behavior remain small:

$$V_N \cdots V_1 \approx \mathbb{E}[V_N \cdots V_1] = (\mathbb{E}V)^N \quad \text{with high probability.} \quad (19)$$

In words, we need to show that a product of i.i.d. random matrices concentrates sharply around its expectation value. This is an interesting and nontrivial problem, even in the (asymptotic) large  $N$ -limit. While sharp concentration bounds for sums of i.i.d. random matrices have been available for more than a decade now [2, 36], our understanding of concentration for random matrix products is more limited; see [17] and references therein. There is a lot of math literature on random walks on Lie groups, but the focus is usually on asymptotic convergence and the machinery is different; see [38] and references therein. The small-step regime has seen less development, although there are some asymptotic bounds [39].

Fortunately, the QDRIFT construction has several appealing features: the random unitaries  $V_N, \dots, V_1$  are i.i.d. unit-norm matrices that are close to the identity matrix ( $\|V - \mathbb{I}\| \leq t\lambda/N$  almost surely) and close to their expectation ( $\|V - (\mathbb{E}V)\| \leq 2t\lambda/N$  almost surely). These properties allow us to use the matrix martingale formalism to derive a strong, nonasymptotic result on the quality of the approximation.

**Proposition 2** (QDRIFT: Spectral norm concentration). *Consider a Hamiltonian  $H = \sum_{i=1}^L H_i$  with interaction strength  $\lambda = \sum_{i=1}^L \|h_i\|$ , and fix parameters  $N, t$ . Suppose that  $V_N, \dots, V_1$  are i.i.d. instances of the random unitary  $d \times d$  matrix  $V$  constructed by the QDRIFT protocol (2). Then*

$$\Pr [\|V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1]\| \geq \epsilon/2] \leq 2d \exp\left(-\frac{N\epsilon^2}{44t^2\lambda^2}\right) \quad \text{for any } \epsilon \in [0, 4t\lambda].$$

In particular,  $N \geq (44t^2\lambda^2/\epsilon^2) \log(2d/\delta)$  implies that  $\|V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1]\| \leq \epsilon/2$  with probability at least  $1 - \delta$ .

This statement provides a strong tail bound for random fluctuations in the small-error regime  $\epsilon \leq 4t\lambda$ . As  $N$  increases, the probability of incurring (at least) error  $\epsilon/2$  diminishes exponentially. For  $\epsilon > 4t\lambda$ , we have instead a subexponential tail bound:  $\Pr [\|V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1]\| \geq \tau] \leq 2d \exp(-N\epsilon/6t\lambda)$ . We refer to (21) for a unified statement that covers both regimes.

The proof technique deserves some exposition, as it is rather general and may be of independent interest. For fixed  $N$ , we interpolate between both sides of Rel. (19) by means of a random process  $\{B_k : k = 0, \dots, N\}$ :

$$B_k = (\mathbb{E}V)^{N-k} V_k \cdots V_1 \quad \text{where} \quad B_0 = (\mathbb{E}V)^N \quad \text{and} \quad B_N = V_N \cdots V_1.$$

The increments of this random process are certainly not independent. For instance,  $B_k$  depends on the (random) choice of  $V_k$  and *all* previous choices  $V_{k-1}, \dots, V_1$ . This suggests that the random process  $\{B_k\}$  may resemble a random walk in matrix space. The following observations support this intuition:

1. *Causality*: Each  $B_k$  is completely determined by the information we have collected up to step  $k$ . That is, the (random) choices of  $V_k, \dots, V_1$ .
2. *Status quo*: Conditioned on previous choices, the expectation of  $B_{k+1}$  equals  $B_k$ : for  $1 \leq k \leq N$

$$\mathbb{E}[B_{k+1} | V_k \cdots V_1] = (\mathbb{E}V)^{N-(k+1)} \mathbb{E}_{V_{k+1}} [V_{k+1}] V_k \cdots V_1 = (\mathbb{E}V)^{N-k} V_k \cdots V_1 = B_k. \quad (20)$$

This feature underscores similarities to an unbiased random walk. On average, “tomorrow” ( $B_{k+1}$ ) is the same as “today” ( $B_k$ ).

An (integrable) random process  $\{B_k : 0 \leq k \leq N\}$  with these two properties is called a *martingale*. The martingale in question is *matrix-valued* and also *bounded*:

$$\mathbb{E}\|B_k\| \leq \|\mathbb{E}V\|^{N-k} \mathbb{E}\|V_k \cdots V_1\| \leq 1 \quad \text{for each } k = 1, \dots, N.$$

This bounded matrix martingale interpolates between  $B_0 = (\mathbb{E}V)^N$ , the deterministic expectation value, and  $B_N = V_N \cdots V_1$ , a product of i.i.d. random matrices:

$$B_N - B_0 = \sum_{k=1}^N (B_k - B_{k-1}) =: \sum_{k=1}^N C_k.$$

We have introduced the elements  $C_k := B_k - B_{k-1}$  of the *difference sequence*. The martingale condition (20) suggests that this difference sequence may control the fluctuations within the random process  $\{B_k\}$ . The following, rather crude, concentration inequality suffices to make this intuition precise.

**Fact 2** (Matrix Freedman). *Let  $\{B_k : k = 0, \dots, N\}$  be a bounded matrix martingale in  $\mathbb{M}_d$ . Assume that the associated difference sequence  $C_k = B_k - B_{k-1}$  obeys  $\|C_k\| \leq R$  almost surely. Then*

$$\Pr[\|B_N - B_0\| \geq \tau] \leq 2d \exp\left(\frac{-\tau^2/2}{NR^2 + R\tau/3}\right) \quad \text{for any } \tau > 0.$$

This statement is a consequence of more general and fine-grained matrix martingale bounds, most notably [35, Corollary 1.3] and also [16, Theorem 11].

*Proof of Proposition 2.* We have already established that the random process  $B_k = (\mathbb{E}V)^{N-k} V_k \cdots V_1$  forms a bounded matrix martingale that interpolates between  $B_0 = \mathbb{E}[V_N \cdots V_1] = (\mathbb{E}V)^N$  and  $B_N = V_N \cdots V_1$ . The elements of the associated difference sequence are

$$C_k = B_k - B_{k-1} = (\mathbb{E}V)^{N-k} (V_k - (\mathbb{E}V_k)) V_{k-1} \cdots V_1 \quad \text{with } k = 1, \dots, N.$$

They are readily bounded. Recall that  $V_k = \exp(-iX_l)$  for some Hermitian matrices  $X_l = \frac{t}{N} \frac{\lambda}{\|h_l\|} h_l$  with index  $1 \leq l \leq L$ . Boundedness ( $\|\mathbb{E}V\|, \|V_k\| \leq 1$ ) and Fact 1 ( $\|\exp(-iX) - \mathbb{I}\| \leq \|X\|$  for  $X$  Hermitian) ensure

$$\begin{aligned} \|C_k\| &\leq \|\mathbb{E}V\|^{N-k} \|V_k - (\mathbb{E}V_k)\| \|V_{k-1} \cdots V_1\| \leq \|V_k - (\mathbb{E}V_k)\| = \|V_k - \mathbb{I} - \mathbb{E}[V_k - \mathbb{I}]\| \\ &\leq \|V_k - \mathbb{I}\| + \|\mathbb{E}[V_k - \mathbb{I}]\| \leq 2 \max_l \|\exp(-iX_l) - \mathbb{I}\| \leq 2 \max_l \|X_l\| = \frac{2t\lambda}{N} \end{aligned}$$

almost surely. Set  $R = 2t\lambda/N$ , and invoke Fact 2 to conclude that

$$\Pr[\|B_N - B_0\| \geq \tau] \leq 2d \exp\left(-\frac{N\tau^2}{8(t\lambda)^2 + 4(t\lambda)\tau/3}\right). \quad (21)$$

The statement follows from bounding the somewhat complicated exponential by either  $\exp(-3\tau^2/(8NR^2))$  for  $\tau \leq 2\lambda t$  or by  $\exp(-3\tau^2/(8R))$  for  $\tau \geq 2\lambda t$ . Last, we substitute  $\tau = \epsilon/2$ .  $\square$

In fact, the same proof works for more general small-step random walks on the unitary group:

**Proposition 3** (Random walk on unitary group). *Let  $\{U_1, U_2, \dots, U_N\} \subset U(d)$  independent, random unitary matrices. Suppose that*

$$\sum_{j=1}^N \|(U_j - \mathbb{E}U_j)(U_j^\dagger - \mathbb{E}U_j^\dagger)\| \leq \sigma^2 \quad \text{and} \quad \|U_j - \mathbb{E}U_j\| \leq B. \quad (22)$$

*Then the product satisfies a concentration inequality:*

$$\Pr(\|U_N \cdots U_1 - \mathbb{E}[U_N \cdots U_1]\| \geq \epsilon) \leq 2d \exp\left(\frac{-\epsilon^2}{\sigma^2 + B\epsilon/3}\right). \quad (23)$$

There are several recent independent papers that also use matrix martingale tools to study products of random matrices that are close to the identity. The work [17] addresses the problem using uniform smoothness tools. The paper [20] uses the matrix Freedman inequality; their proof is quite similar to ours. In contrast, we are interested in unitary products, which allows for additional simplifications. For more background on matrix martingales, see [12, 17, 26, 29, 35].

## 4. A bound for expected error

In the previous subsection, we established that a sufficiently long QDRIFT random product formula concentrates sharply around its expectation. We can translate this statement into a bound on the expected fluctuation around the true evolution.

**Proposition 4** (QDRIFT: Expected diamond norm error). *Consider an  $n$ -qubit Hamiltonian  $H = \sum_{i=1}^L h_i$  with total size  $\lambda = \sum_{i=1}^L \|h_i\|$ . Fix parameters  $N, t$ , and assume that  $N \geq n$ . Set  $\mathcal{U} = UXU^\dagger$  with  $U = \exp(-itH)$ , and suppose that  $\mathcal{V}_N, \dots, \mathcal{V}_1 \sim \mathcal{V}$  are i.i.d. realizations of the QDRIFT protocol. That is,  $\mathcal{V}(X) = VXV^\dagger$ , where  $V$  is defined by (2). Then*

$$\mathbb{E}\left[\frac{1}{2}\|\mathcal{U} - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1\|_\diamond\right] \leq \frac{t^2\lambda^2}{N} + C\frac{nt\lambda}{N} + C\sqrt{\frac{nt^2\lambda^2}{N}} \approx C\sqrt{\frac{nt^2\lambda^2}{N}}, \quad (24)$$

where  $C > 0$  is a (modest) numerical constant. The symbol  $\approx$  denotes an accurate approximation in the large- $N$  regime.

It is instructive to compare this assertion to the original QDRIFT result [8] and the improvement in (18):

$$\frac{1}{2}\|\mathcal{U} - \mathbb{E}[\mathcal{V}_N \circ \dots \circ \mathcal{V}_1]\|_\diamond \leq \frac{t^2\lambda^2}{N}.$$

Note that the expectation over all possible realizations of all  $N$  unitary channels appears inside the diamond distance. This implies that QDRIFT performs well on average over many random realizations, provided that the number  $N$  of steps exceeds  $t^2\lambda^2/\epsilon$ . In contrast, (24) has the expectation outside the diamond distance.

Our result gives a much stronger conclusion: An *individual* realization of the randomized QDRIFT protocol does not deviate much from the target evolution, for any input states and observables, with very high probability. The price for such an improvement is a larger number of steps that depends on the system size. For  $n$  qubits, the gate complexity  $N \geq Cnt^2\lambda^2/\epsilon^2$  is sufficient to ensure  $\epsilon$ -closeness on average. The quadratic scaling in the accuracy parameter  $\epsilon$  is necessary (for large  $N$ ) because of the central limit theorem for martingales. The appearance of the number  $n$  of qubits is a consequence of measuring closeness in diamond distance. To obtain

$$\epsilon \geq \mathbb{E}\left[\frac{1}{2}\|\mathcal{U} - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1\|_\diamond\right] = \mathbb{E}\left[\frac{1}{2}\max_{\rho \text{ state}} \|\mathcal{U}_\rho - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1\right],$$

we need the random product formula to behave for all possible  $n$ -qubit input states  $\rho$  simultaneously. If we restrict our attention to any fixed input state, we can obtain a gate complexity that does not depend on  $n$ . This is the topic of the next section.

*Proof of Proposition 4.* First, we relate the expected diamond distance to an expected operator norm distance and split it up into deterministic bias and (expected) fluctuations:

$$\mathbb{E}\left[\frac{1}{2}\|\mathcal{U} - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1\|_\diamond\right] \leq \|U - (\mathbb{E}V)^N\| + \mathbb{E}\|V_N \cdots V_1 - (\mathbb{E}V)^N\|$$

The first term is deterministic and controlled by Proposition 1:  $\|U - (\mathbb{E}V)^N\| \leq t^2\lambda^2/N$ . The second term can be bounded by integrating the tail bound in Proposition 2, or rather the tighter bound presented (21); see [36, Remark 6.5]. Set  $d = 2^n$  to conclude

$$\begin{aligned} \mathbb{E}\|V_N \cdots V_1 - (\mathbb{E}V)^N\| &= \int_0^\infty \Pr[\|V_N \cdots V_1 - (\mathbb{E}V)^N\| \geq \tau] d\tau \\ &\leq \int_0^\infty 2 \times 2^n \exp\left(-\frac{\tau^2/2}{4\lambda^2 t^2 + 2\lambda t\tau/3}\right) d\tau \\ &\leq 2C \max\left\{\sqrt{\frac{nt^2\lambda^2}{N}}, \frac{nt\lambda}{N}\right\} \leq C\left(\frac{nt\lambda}{N} + \sqrt{\frac{nt^2\lambda^2}{N}}\right), \end{aligned}$$

where  $2C$  is a constant. □

### B. Proof of Theorem 3: Approximation error under a single arbitrary input

Proposition 4 asserts that a single, random realization of the QDRIFT protocol (2) accurately approximates a unitary target evolution with respect to the diamond norm:

$$\mathbb{E}\left[\frac{1}{2}\|\mathcal{U} - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1\|_{\diamond}\right] = \mathbb{E}\left[\frac{1}{2} \max_{\rho \text{ state}} \|\mathcal{U}(\rho) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1\right] \lesssim C \sqrt{\frac{nt^2\lambda^2}{N}}.$$

Here,  $\lesssim$  denotes an accurate approximation of the true bound in the large  $N$  regime. This bound scales linearly in the (qubit) system size  $n$ . The dependence on  $n$  should not come as a surprise, since the diamond norm produces a very stringent worst-case distance measure. As emphasized by the above reformulation, the approximation must be accurate even when we optimize to find the worst possible input state  $\rho$ .

In Hamiltonian simulation, demanding such a stringent worst-case promise may be excessive. In most practical applications, the input state  $\rho$  is fixed and simple, e.g., a product state. In this more practical setting, we can obtain a gate complexity  $N$  that does not depend on the system size  $n$ . The main result of this section asserts

$$\max_{\rho \text{ state}} \mathbb{E}\left[\frac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1\right] \leq C \sqrt{\frac{t^2\lambda^2}{N}}.$$

In other words, fixing an arbitrary input state  $\rho$  helps a lot. A total number of  $N = 4(t\lambda/\epsilon)^2$  steps ensures that QDRIFT produces an  $\epsilon$ -accurate output state, with respect to trace distance.

The proof is similar in spirit to the argument behind Proposition 4. We construct a vector-valued martingale that describes the evolution of the state. We control the behavior of this martingale using the uniform smoothness of the  $L_q(\ell_2)$  norm. This argument is inspired by the work [17] on concentration of random matrix products.

#### 1. Approximation error for a fixed state

In this section, we state and prove our main technical result on the action of the QDRIFT protocol on a fixed input state.

**Proposition 5** (QDRIFT: Action on a fixed state). *Consider a Hamiltonian  $H$  with total strength  $\lambda$  and evolution time  $t$ . Let  $\mathcal{V}_1, \dots, \mathcal{V}_N$  be the i.i.d. random unitary evolution operators constructed by the QDRIFT protocol (2). For  $N \geq (t\lambda)^2$ ,*

$$\max_{\rho \text{ state}} \mathbb{E}\left[\frac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1\right] \leq 4\sqrt{\frac{t^2\lambda^2}{N}}.$$

Moreover, for  $\epsilon > 0$ ,

$$\max_{\rho \text{ state}} \Pr\left[\frac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1 > \epsilon\right] \leq \exp\left(\frac{-\epsilon^2 N}{32et^2\lambda^2}\right).$$

*Proof.* First, we reduce the problem to a question about pure states. For any  $q \geq 2$ , Markov's inequality implies that

$$\Pr\left[\frac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1 > \epsilon\right] \leq \epsilon^{-q} \mathbb{E}\left[2^{-q}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(\rho)\|_1^q\right]. \quad (25)$$

The right-hand side of this equation is a convex function of the state  $\rho$ . Thus, the maximum of the right-hand side over all states is attained at a pure state. As a consequence, we can establish both claims in the proposition by limiting our attention to an (unknown) pure state  $\rho = |\psi\rangle\langle\psi|$  that does not depend on the random unitaries  $\mathcal{V}_i$ .

Next, we convert the trace distance of the output states into a Euclidean distance on the state vectors themselves. The power  $q \geq 2$  will remain fixed until the last step of the argument. Lemma 2 implies

$$\begin{aligned} (\mathbb{E}[2^{-q}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}_N \circ \dots \circ \mathcal{V}_1(|\psi\rangle\langle\psi|)\|_1^q])^{1/q} &\leq (\mathbb{E}\|(V_N \dots V_1 - U)|\psi\rangle\|_{\ell_2}^q)^{1/q} \\ &\leq 2 \max\left\{\underbrace{\|\mathbb{E}[V_N \dots V_1] - U\|_{\ell_2}}_{\text{deterministic bias } |\psi_{\text{bias}}\rangle}, \underbrace{\left(\mathbb{E}\|(V_N \dots V_1 - \mathbb{E}[V_N \dots V_1])|\psi\rangle\|_{\ell_2}^q\right)^{1/q}}_{\text{random fluctuation } |\psi_{\text{rand}}\rangle}\right\}. \end{aligned} \quad (26)$$

The last bound follows from the triangle inequality and the fact  $(a + b)^q \leq 2^q \max\{a^q, b^q\}$  for  $a, b \geq 0$ .

We have split up the difference into two components, a deterministic bias and a random fluctuation. To control the deterministic bias, we simply apply Proposition 1:

$$\|(\mathbb{E}[V_N \cdots V_1] - U)|\psi\rangle\|_{\ell_2} = \|((\mathbb{E}V)^N - U)|\psi\rangle\|_{\ell_2} \leq \|(\mathbb{E}V)^N - U\| \leq \frac{(t\lambda)^2}{N}. \quad (27)$$

We will see that the bias is always negligible in comparison with the fluctuation. To control the second term, we need the following lemma.

**Lemma 5.** *Let  $V_N, \dots, V_1$  be i.i.d. unitaries that implement the QDRIFT protocol (2) with parameters  $t$  and  $\lambda$ . Then, for any  $q \geq 2$ ,*

$$(\mathbb{E}\|(V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1])|\psi\rangle\|_{\ell_2}^q)^{1/q} \leq 2\sqrt{\frac{(q-1)(t\lambda)^2}{N}}.$$

We will establish this lemma below. The basic idea behind the proof is to express the random vector using a martingale sequence.

Introduce the inequalities from (27) and Lemma 5 into the bound (26). We obtain

$$(\mathbb{E}[2^{-q}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1(|\psi\rangle\langle\psi|)\|_1^q])^{1/q} \leq 4\sqrt{\frac{(q-1)(t\lambda)^2}{N}}. \quad (28)$$

We have used the assumption that  $N \geq (t\lambda)^2$  to see that the second branch of the maximum always dominates the first.

We may now complete the proof. To obtain the expectation bound, we set  $q = 2$  in (28) and apply Lyapunov's inequality. To obtain the probability bound, we combine (25) and (28) to arrive at

$$\Pr\left[\frac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1(\rho)\|_1 > \epsilon\right] \leq \left(\frac{16q(t\lambda)^2}{\epsilon^2 N}\right)^{q/2}.$$

Select  $q = (\epsilon^2 N)/(16t^2\lambda^2)$  to obtain the stated result. The resulting probability bound is vacuous unless  $q \geq 2$ .  $\square$

## 2. Proof of Lemma 5

In this section, we establish the bound on the size of the fluctuations. The main ingredient in the argument is a powerful method for exploiting the orthogonality of the martingale difference sequence.

**Fact 3** (Subquadratic averages). *Let  $x, y \in \mathbb{C}^d$  be two random vectors that obey  $\mathbb{E}[y|x] = 0$ . Then, for any  $q \geq 2$ ,*

$$(\mathbb{E}\|x + y\|_{\ell_2}^q)^{2/q} \leq (\mathbb{E}\|x\|_{\ell_2}^q)^{2/q} + (q-1)(\mathbb{E}\|y\|_{\ell_2}^q)^{2/q}.$$

Fact 3 is a consequence of Bonami's inequality [14, Cor. 13.1.1] and some standard arguments; see [17, Sec. 3]. Geometrically, this result expresses the uniform smoothness of the space  $L_q(\ell_2)$ .

*Proof of Lemma 5.* Fix a vector  $|\psi\rangle$ , and introduce a sequence of random vectors:  $|\psi_k\rangle = \prod_{i=1}^k V_i|\psi\rangle$  for  $1 \leq k \leq N$ . As a consequence,  $(V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1])|\psi\rangle = |\psi_N\rangle - \mathbb{E}[|\psi_N\rangle]$ . We can recast this difference as a sum of two random vectors that are conditionally orthogonal in expectation:

$$\mathbb{E}\| |\psi_N\rangle - \mathbb{E}[|\psi_N\rangle] \|_{\ell_2}^q = \mathbb{E}\|(V_N - \mathbb{E}[V_N])|\psi_{N-1}\rangle + \mathbb{E}[V_N](|\psi_{N-1}\rangle - \mathbb{E}[|\psi_{N-1}\rangle])\|_{\ell_2}^q =: \mathbb{E}\|y + x\|_{\ell_2}^q.$$

Indeed,  $\mathbb{E}[y|x] = \mathbb{E}[V_N - (\mathbb{E}V_N)]|\psi_{N-1}\rangle = 0$ . We can apply Fact 3 to split up the contributions:

$$\begin{aligned} (\mathbb{E}\| |\psi_N\rangle - \mathbb{E}[|\psi_N\rangle] \|_{\ell_2}^q)^{2/q} &\leq (q-1)(\mathbb{E}\|(V_N - (\mathbb{E}V_N))|\psi_{N-1}\rangle\|_{\ell_2}^q)^{2/q} \\ &\quad + (\mathbb{E}\|\mathbb{E}[V_N](|\psi_{N-1}\rangle - \mathbb{E}[|\psi_{N-1}\rangle])\|_{\ell_2}^q)^{2/q} \\ &\leq (q-1)(\mathbb{E}\|V_N - \mathbb{E}[V_N]\|_{\ell_2}^q)^{2/q} + (\mathbb{E}\| |\psi_{N-1}\rangle - \mathbb{E}[|\psi_{N-1}\rangle] \|_{\ell_2}^q)^{2/q}. \end{aligned}$$

We can now iterate this argument to conclude that

$$(\mathbb{E}\|\psi_N - \mathbb{E}[\psi_N]\|_{\ell_2}^q)^{2/q} \leq (q-1) \sum_{k=1}^N (\mathbb{E}\|V_k - \mathbb{E}[V_k]\|_q^q)^{2/q}.$$

Invoke Lemma 3, using the properties of the random unitaries constructed by qDRIFT:

$$(\mathbb{E}\|V - \mathbb{E}[V]\|_q^q)^{2/q} \leq \left(2 \frac{t\lambda}{N}\right)^2.$$

Combine the last two displays to reach the stated result.  $\square$

Lemma 5 allows us to control the probability of having a large random fluctuation. We will see that the probability would decay exponentially, nearly the same as Proposition 2, though without a factor of dimension  $d$ .

### C. Proof of Theorem 4: Approximation error under fixed set of input states and observables

This result is derived from (18), which improves on Campbell's original technical result [8] for the error in the average channel. An immediate consequence of (18) is the following important statement:

$$N \geq 4t^2\lambda^2/\epsilon \quad \text{implies} \quad \|\mathcal{U} - \mathbb{E}[\mathcal{V}_N \cdots \mathcal{V}_1]\|_\diamond \leq \epsilon/2.$$

That is, the diamond distance error in the average channel is negligible when the number  $N$  of gates is sufficiently large. Given  $M$  arbitrary input-observable pairs  $(|\psi_1\rangle, O_1), \dots, (|\psi_M\rangle, O_M)$ , the diamond distance bound ensures that

$$\left| \langle \psi_j | U^\dagger O_j U | \psi_j \rangle - \mathbb{E} \left[ \langle \psi_j | V_1^\dagger \cdots V_N^\dagger O_j V_N \cdots V_1 | \psi_j \rangle \right] \right| \leq \epsilon/2, \quad \text{for all } 1 \leq j \leq M. \quad (29)$$

Thus, it suffices to approximate each  $\mathbb{E}[\langle \psi_j | V_1^\dagger \cdots V_N^\dagger O_j V_N \cdots V_1 | \psi_j \rangle]$  up to accuracy  $\epsilon/2$  in order to get an  $\epsilon$ -approximation of the original target  $\langle \psi_j | U^\dagger O_j U | \psi_j \rangle$ . This can be achieved by executing the following procedure. We first sample  $R$  random product formulas,  $V_N^{(1)} \cdots V_1^{(1)}, \dots, V_N^{(R)} \cdots V_1^{(R)}$  using the qDRIFT procedure. For each input-observable pair  $(|\psi_j\rangle, O_j)$ , we perform  $R$  measurement repetitions:

$$\begin{aligned} \text{repetition 1 : } & |\psi_j\rangle \rightarrow V_N^{(1)} \cdots V_1^{(1)} |\psi_j\rangle \rightarrow \text{measure } O_j \text{ to get } \hat{o}_j^{(1)}. \\ & \vdots \\ \text{repetition } R : & |\psi_j\rangle \rightarrow V_N^{(R)} \cdots V_1^{(R)} |\psi_j\rangle \rightarrow \text{measure } O_j \text{ to get } \hat{o}_j^{(R)}. \end{aligned}$$

Single-shot measurements for each repetition are sufficient. It is not necessary to estimate the full expectation value associated with observable  $O_j$ . We then average the  $R$  measurement outcomes to obtain the empirical estimates

$$\hat{o}_j = \frac{1}{R} \sum_{r=1}^R \hat{o}_j^{(r)} \quad \text{for each } j = 1, \dots, M.$$

We assume that the observable  $O_j$  has eigenvalues bounded between  $[-1, 1]$ . Hence,  $\hat{o}_j^{(r)} \in [-1, 1]$ . By construction, the expectation value of the individual measurement outcome  $\hat{o}_j^{(r)}$  is given by

$$\mathbb{E}[\hat{o}_j^{(r)}] = \mathbb{E} \left[ \langle \psi_j | V_1^\dagger \cdots V_N^\dagger O_j V_N \cdots V_1 | \psi_j \rangle \right].$$

Hoeffding's inequality tells us that the average of the  $R$  outcomes concentrates around the expectation:

$$\mathbb{P} \left( \left| \hat{o}_j - \mathbb{E} \left[ \langle \psi_j | V_1^\dagger \cdots V_N^\dagger O_j V_N \cdots V_1 | \psi_j \rangle \right] \right| \geq \epsilon \right) \leq 2 \exp \left( -\frac{R\epsilon^2}{2} \right).$$

Fixing an error parameter  $\delta \in [0, 1]$  and setting  $R \geq 16 \log(M/\delta)/\epsilon^2$  ensures that

$$\mathbb{P} \left( \left| \hat{o}_j - \mathbb{E} \left[ \langle \psi_j | V_1^\dagger \cdots V_N^\dagger O_j V_N \cdots V_1 | \psi_j \rangle \right] \right| \geq \frac{\epsilon}{2} \right) \leq \frac{\delta}{M} \quad \text{for each } 1 \leq j \leq M.$$

By a union bound over the  $M$  pairs of input state and observable, we obtain

$$\mathbb{P} \left( \max_{1 \leq j \leq M} \left| \hat{\delta}_j - \mathbb{E} \left[ \langle \psi_j | V_1^\dagger \cdots V_N^\dagger O_j V_N \cdots V_1 | \psi_j \rangle \right] \right| \geq \frac{\epsilon}{2} \right) \leq \delta.$$

Taking the complement, we learn that, with probability at least  $1 - \delta$ , each of the  $M$  estimates  $\hat{\delta}_j$  with  $1 \leq j \leq M$  will be  $\epsilon/2$ -close to its expected value. In view of (29), we obtain

$$|\hat{\delta}_j - \langle \psi_j | U^\dagger O_j U | \psi_j \rangle| < \epsilon \quad \text{for all } 1 \leq j \leq M$$

with probability at least  $1 - \delta$ . This concludes the proof of Theorem 4.

## V. ASYMPTOTIC TIGHTNESS

It is natural to wonder whether the bound (24) is tight for some Hamiltonian, i.e., whether  $N = \Omega(n\lambda^2 t^2 / \epsilon^2)$  is also necessary to achieve concentration. More precisely, we want to understand whether the dependences on system size  $n = \log_2(d)$ , evolution time  $t$  and interaction strength  $\lambda$  are also necessary to control the typical deviation of the unitary random walk we considered.

In the context of matrix concentration inequalities, this question has been thoroughly addressed [36]. The answer is affirmative for sums of bounded matrices: concentration inequalities are tight and saturated for collections of *commuting* matrices. Although in this work we consider products of random matrices, we are still using a telescoping sum in the small step regime and expect an analogy.

This observation motivates us to look at artificial Hamiltonians whose associated unitary evolution saturates the upper bounds put forth in this work. The cases we can handle lie at the two extremes: either the sum of single-site Pauli  $Z$ s or the sum of all  $2^n$  many-body Pauli  $Z$ s. We will see the presence of the system size factor  $n = \log_2(d)$  at both extremes, so one may believe the same to hold for the intermediate  $q$ -local cases. However, this factor arises for very different reasons. It arises in the single-site case, because the operator norm completely factorizes into  $n$  constituents (one for each term). For Hamiltonians that encompass all  $2^n$  many-body  $Z$ s, it comes from the fact that diagonal entries are nearly independent, so the union bound we used in Section III B is tight. Independence of entries requires the presence of all many-body terms, and does not extend to the few-body case.

The multivariate central limit Theorem will be crucial for analyzing both cases, as it greatly simplifies the analysis in large  $N$  limit.

**Fact 4** (CLT for the multinomial distribution). *The multinomial distribution  $\mathbf{m} = (m_1, \dots, m_K) \sim \text{Mult}(N, (1/K, \dots, 1/K))$  (roll a fair  $K$ -sided dice  $N$  times) obeys a central limit theorem (CLT):*

$$\frac{1}{\sqrt{N}}(\mathbf{m} - \mathbb{E}\mathbf{m}) \sim \mathcal{N}(0, \Sigma) \quad \text{almost surely as } N \rightarrow \infty.$$

The covariance matrix is  $\Sigma = \frac{1}{K}(\mathbb{I} - \frac{1}{K}J)$ , where  $J$  denotes the  $K \times K$  matrix of ones.

### A. Sum of single site Pauli-Z operators

This example demonstrates the saturation of our martingale bounds for single site Hamiltonians that factorize completely. To this end, we revisit a variant of the  $n$ -qubit example Hamiltonian discussed in Section III A:

$$H = \sum_{k=1}^n Z_k \quad \text{where} \quad Z_k = \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{(k-1)\text{-times}} \otimes Z \otimes \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{(n-k)\text{-times}} \quad \text{for } 1 \leq k \leq n. \quad (30)$$

**Proposition 6.** *Suppose that we wish to obtain an  $N$ -term approximation of the time evolution  $U = \exp(-itH)$  associated with the  $n$ -qubit Hamiltonian (30) for evolution time  $t$ . In the large  $N$  limit (CLT), the QDRIFT approximation (2) incurs an operator norm error that matches the (upper) bound from Theorem 1 up to a constant factor:*

$$\mathbb{E} \|U - V_N \cdots V_1\| \geq \sqrt{\frac{2}{\pi}} \sqrt{(n-1) \frac{(t\lambda)^2}{N}} - \frac{1}{2}(n-1) \frac{(t\lambda)^2}{N}.$$

We have chosen to state this result directly in terms of operator norm deviation. A conversion into diamond distance is also possible:  $\frac{1}{2}\|U - V\|_\diamond \geq \frac{1}{2}\|U - V\|$  for any pair of unitary channels. This conversion rule readily follows from the geometric characterization of  $\frac{1}{2}\|U - V\|_\diamond$  provided in Ref. [1].

*Proof of Proposition 6.* Each of the  $n$  terms in the Hamiltonian (30) has unit operator norm ( $\|Z_k\| = 1$ ) and the strength is  $\lambda = \sum_{k=1}^n \|Z_k\| = n$ . For fixed  $N$  and  $t$ , each short-time approximation (2) has the form  $V_i = \exp(-i\frac{t\lambda}{N}Z_{k(i)})$ , where each  $k(i)$  is an index chosen uniformly from the set  $\{1, \dots, n\}$  (multinomial distribution). Since all  $Z_k$ s commute, we can rewrite the entire product formula as

$$V_N \cdots V_1 = \exp\left(-i\frac{t\lambda}{N} \sum_{i=1}^N Z_{k(i)}\right) = \exp\left(-i\frac{t\lambda}{N} \sum_{k=1}^n m_k Z_k\right).$$

Here, we have introduced the count statistics  $m_k$  for each site label  $k$  – that is the number of times location  $k$  has been selected throughout  $N$  independent selection rounds – to rearrange the sum. This count statistics obeys  $\bar{m}_k = \mathbb{E}m_k = N/n = N/\lambda$  for each  $1 \leq k \leq n$ . We can use this observation to re-express the target unitary  $U$  in a compatible fashion:

$$U = \exp\left(-it \sum_{k=1}^n Z_k\right) = \exp\left(-i\frac{t\lambda}{N} \sum_{k=1}^n \bar{m}_k Z_k\right).$$

Unitary invariance then implies that the operator norm difference between both unitaries becomes

$$\|V_N \cdots V_1 - U\| = \left\| \exp\left(-i\frac{t\lambda}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\sqrt{N}} (m_k - \bar{m}_k) Z_k\right) - \mathbb{I} \right\|. \quad (31)$$

This is a promising starting point. The multinomial CLT (Fact 4) ensures that the  $n$  centered and normalized random variables  $s_k = \frac{1}{\sqrt{N}}(m_k - \bar{m}_k)$  approach the coefficients of a Gaussian vector  $s \in \mathbb{R}^n$  with covariance matrix  $\Sigma = \frac{1}{n}(\mathbb{I} - \frac{1}{n}J)$ . This, in particular implies  $\mathbb{E}s_k = 0$  and  $\mathbb{E}s_k^2 = \frac{1}{n}(1 - \frac{1}{n}) = \sigma^2$  for all  $1 \leq k \leq n$ . We can capitalize on this observation by simplifying (31) via a second-order Taylor expansion. Set  $X = -\frac{t\lambda}{\sqrt{N}} \sum_k s_k Z_k$  for brevity and apply Fact 1 to obtain

$$\|V_N \cdots V_1 - U\| = \|\exp(iX) - \mathbb{I}\| \geq \|iX\| - \|\exp(iX) - iX - \mathbb{I}\| \geq \|X\| - \frac{1}{2}\|X\|^2.$$

This relation is preserved under expectations and we obtain

$$\mathbb{E}\|V_N \cdots V_1 - U\| \geq \frac{t\lambda}{\sqrt{N}} \mathbb{E}\left\| \sum_k s_k Z_k \right\| - \frac{1}{2} \left(\frac{t\lambda}{\sqrt{N}}\right)^2 \mathbb{E}\left\| \sum_k s_k Z_k \right\|^2.$$

Let us focus on the leading order term first. The particular structure of the Hamiltonian (30) – each  $Z_k$  is the tensor product of a single Pauli-Z matrix at location  $k$  with  $(n-1)$  identity matrices – ensures that the operator norm factorizes nicely. Use  $\|X \otimes \mathbb{I} + \mathbb{I} \otimes Y\| = \|X\| + \|Y\|$  iteratively to conclude

$$\mathbb{E}\left\| \sum_k s_k Z_k \right\| = \mathbb{E} \sum_{k=1}^n \|s_k Z_k\| = \sum_{k=1}^n |s_k| \stackrel{N \rightarrow \infty}{\approx} n \sqrt{\frac{2}{\pi} \frac{1}{n} (1 - \frac{1}{n})} = \sqrt{\frac{2}{\pi} (n-1)},$$

because the CLT asserts that each  $|s_k|$  approaches a half-normal random variable with  $\sigma^2 = \frac{1}{n}(1 - \frac{1}{n})$ .

To bound the quadratic term, we combine the factorization trick from above with a well-known relation among  $\ell_p$ -norms in  $\mathbb{R}^n$ :

$$\mathbb{E}\left\| \sum_{k=1}^n s_k Z_k \right\|^2 = \mathbb{E} \left( \sum_{k=1}^n |s_k| \right)^2 = \mathbb{E} \|s\|_{\ell_1}^2 \leq n \mathbb{E} \|s\|_{\ell_2}^2 = n \sum_{k=1}^n \mathbb{E}s_k^2 = n^2 \sigma^2 = (n-1).$$

No CLT is required for this argument. Inserting both bounds into Eq. (31) completes the argument.  $\square$

## B. Sum of many-body Pauli-Z operators

Let us revisit the example Hamiltonian from Sec. IIIB, albeit without additional sign factors. Recall the multi-indices  $\mathbf{p} = (p_1, \dots, p_n) \in \{0, 1\}^n$  and set

$$H = \sum_{\mathbf{p} \in \{0,1\}^n} Z_{\mathbf{p}} = \sum_{\mathbf{p} \in \{0,1\}^n} Z^{p_1} \otimes \cdots \otimes Z^{p_n}, \quad (32)$$

where we use the conventions  $Z^1 = Z$  and  $Z^0 = \mathbb{I}$ . This Hamiltonian is not local. All constituents commute and have the same operator norm:  $\|Z_{\mathbf{p}}\| = 1$  for all  $\mathbf{p} \in \{0, 1\}^n$ . This in turn implies that the strength parameter  $\lambda = \sum_{\mathbf{p}} \|Z_{\mathbf{p}}\| = 2^n$  equals the Hilbert space dimension. It is also worthwhile to point out that each term is diagonal in the computational basis  $|\mathbf{b}\rangle = |b_1, \dots, b_n\rangle$  with  $\mathbf{b} = (b_1, \dots, b_n) \in \{0, 1\}^n$ . Overlaps of the Hamiltonian terms with computational basis states are given by

$$\langle \mathbf{b} | Z_{\mathbf{p}} | \mathbf{b} \rangle = (-1)^{\langle \mathbf{b}, \mathbf{p} \rangle} = (-1)^{\sum_i b_i p_i} \in \{\pm 1\}. \quad (33)$$

The following claim highlights that our findings are tight for asymptotically large step sizes  $N$ . This complements the example upper bound derived in Sec. III B, as well as Theorem 1.

**Proposition 7.** *Suppose that we wish to obtain an  $N$ -term approximation of the time evolution  $U = \exp(-itH)$  associated with the  $n$ -qubit Hamiltonian (32) for evolution time  $t$ . In the large  $N$  limit (CLT) the QDRIFT approximation (2) incurs an operator norm error that matches the (upper) bound from Theorem 1 up to a constant factor:*

$$\mathbb{E} \|U - V_N \cdots V_1\| \geq \frac{1}{2} \sqrt{n \frac{(t\lambda)^2}{N}} - 2 \left(n + \frac{1}{2}\right) \frac{(t\lambda)^2}{N}$$

The conversion rule  $\frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_{\diamond} \geq \frac{1}{2} \|U - V\|$  (for unitary channels) once more allows for addressing the expected diamond distance as well.

*Proof of Proposition 7.* Each of the  $2^n$  terms in the Hamiltonian (32) has unit operator norm ( $\|Z_{\mathbf{p}}\| = 1$  for all  $\mathbf{p} \in \{0, 1\}^n$ ) and the strength is  $\lambda = \sum_{\mathbf{p}} \|Z_{\mathbf{p}}\| = 2^n$ . For fixed  $N$  and  $t$ , each short-time approximation (2) has the form  $V_i = \exp(-i \frac{t\lambda}{N} Z_{\mathbf{p}(i)})$ , where  $\mathbf{p}(i)$  is a string chosen uniformly at random from all  $2^n$  possibilities (multinomial distribution). Since all  $Z_{\mathbf{p}}$ s commute, we can rephrase and simplify the expected operator norm difference in a fashion analogous to the proof of Proposition 6:

$$\mathbb{E} \|V_N \cdots V_1 - U\| \geq \frac{t\lambda}{\sqrt{N}} \mathbb{E} \left\| \sum_{\mathbf{p}} s_{\mathbf{p}} Z_{\mathbf{p}} \right\| - \frac{1}{2} \left( \frac{t\lambda}{\sqrt{N}} \right)^2 \mathbb{E} \left\| \sum_{\mathbf{p}} s_{\mathbf{p}} Z_{\mathbf{p}} \right\|^2. \quad (34)$$

Here,  $s_{\mathbf{p}} = \frac{1}{\sqrt{N}}(m_{\mathbf{p}} - \bar{m}_{\mathbf{p}})$  is the centered and normalized variant of the count statistics  $m_{\mathbf{p}}$  associated with bit string  $\mathbf{p} \in \{0, 1\}^n$  – that is the number of times the Hamiltonian term  $Z_{\mathbf{p}}$  has been selected throughout  $N$  independent selection rounds. The multinomial CLT (Fact 4) asserts that the  $2^n$  centered and normalized random variables  $s_{\mathbf{p}} = \frac{1}{\sqrt{N}}(m_{\mathbf{p}} - m_{\mathbf{p}})$  approach distinct coefficients of a  $2^n$ -dimensional Gaussian vector with covariance matrix  $\Sigma = \frac{1}{2^n} (\mathbb{I} - \frac{1}{2^n} \mathbf{J}) = \frac{1}{2^n} (\mathbb{I} - |\mathbf{1}\rangle\langle\mathbf{1}|)$ , where  $|\mathbf{1}\rangle = \frac{1}{2^n} \sum_{\mathbf{b} \in \{0, 1\}^n} |\mathbf{b}\rangle$  (the normalized all-ones vector in  $\mathbb{R}^{2^n}$ ). In contrast to before, the individual contributions to this operator norm don't factor nicely anymore. Establishing tight bounds requires additional analysis.

Let us focus on the (leading) first-order term for now. All matrix summands in the expression commute and are diagonal in the computational basis  $|\mathbf{b}\rangle$  with  $\mathbf{b} = (b_1, \dots, b_n) \in \{0, 1\}^n$ . This ensures that the operator norm is attained at a computational basis state:

$$\left\| \sum_{\mathbf{p}} Z_{\mathbf{p}} s_{\mathbf{p}} \right\| = \max_{\mathbf{b} \in \{0, 1\}^n} \left| \sum_{\mathbf{p}} s_{\mathbf{p}} \langle \mathbf{b} | Z_{\mathbf{p}} | \mathbf{b} \rangle \right| = \max_{\mathbf{b} \in \{0, 1\}^n} \left| \sum_{\mathbf{p}} (-1)^{\langle \mathbf{b}, \mathbf{p} \rangle} s_{\mathbf{p}} \right|,$$

where the last equation is due to Rel. (33). This expression is proportional to the largest entry (in modulus) of the Walsh-Hadamard transform of the  $2^n$ -dimensional vector  $s$  with entries  $s_{\mathbf{p}}$  for  $\mathbf{p} \in \{0, 1\}^n$ . More precisely,

$$\max_{\mathbf{b} \in \{0, 1\}^n} \left| \sum_{\mathbf{p}} (-1)^{\langle \mathbf{b}, \mathbf{p} \rangle} s_{\mathbf{p}} \right| = 2^{n/2} \|\text{Had}^{\otimes n} s\|_{\ell_{\infty}} =: 2^{n/2} \|\hat{s}\|_{\ell_{\infty}} \quad \text{where} \quad \text{Had} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

We emphasize that the Walsh-Hadamard transform is an orthogonal transformation, which also applies to the limiting covariance matrix of  $\hat{s} = \text{Had}^{\otimes n} s$  (CLT):

$$\hat{\Sigma} = \frac{1}{2^n} \text{Had}^{\otimes n} (\mathbb{I} - |\mathbf{1}\rangle\langle\mathbf{1}|) \text{Had}^{\otimes n} = \frac{1}{2^n} (\mathbb{I} - |0, \dots, 0\rangle\langle 0, \dots, 0|).$$

Hence, the CLT asserts that the transformed vector  $\hat{s}$  approaches a standard Gaussian vector with  $2^n - 1$  degrees of freedom:  $\hat{s} = (0, g_2, \dots, g_{2^n})^T$  with  $g_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2^{-n})$  (one degree of freedom is erased by the

normalization constraint  $\sum_{\mathbf{p}} m_{\mathbf{p}} = N$  of the count statistics). The bound on the expected leading order contribution now follows from invoking the well-known fact that the expected maximum of  $K$  standard Gaussian random variables with equal variance  $\sigma^2$  is lower-bounded by  $0.265\sqrt{\log(K)\sigma^2}$ , see e.g. [13, Proposition 8.1]:

$$\begin{aligned} \frac{t\lambda}{\sqrt{N}} \mathbb{E} \left\| \sum_{\mathbf{p}} s_{\mathbf{p}} Z_{\mathbf{p}} \right\| &= \frac{t\lambda}{\sqrt{N}} 2^{n/2} \mathbb{E} \|\hat{s}\|_{\ell_{\infty}} \stackrel{N \rightarrow \infty}{\simeq} \frac{t\lambda}{\sqrt{N}} 2^{n/2} \mathbb{E} \max_{2 \leq i \leq 2^n} |g_i| \\ &\geq 0.625 \frac{t\lambda}{\sqrt{N}} 2^{n/2} \sqrt{\log(2^n - 1)2^{-n}} \geq \frac{1}{2} \sqrt{n \frac{(t\lambda)^2}{N}}. \end{aligned}$$

Here, we have used the numerical bound  $0.625\sqrt{\log(2^n - 1)/n} \geq 0.5$  which is valid for any  $n \geq 3$  (for  $n = 2$  the ratio is slightly smaller). This completes the argument for the leading term.

Moving on to the quadratic term in Eq. (34), we employ a similar strategy. Observe

$$\begin{aligned} \mathbb{E} \left\| \sum_{\mathbf{p}} Z_{\mathbf{p}} \right\|^2 &= \mathbb{E} \left\| \sum_{\mathbf{p}, \mathbf{p}'} s_{\mathbf{p}} s_{\mathbf{p}'} Z_{\mathbf{p}} Z_{\mathbf{p}'} \right\|^2 = \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} \left| \sum_{\mathbf{p}, \mathbf{p}'} s_{\mathbf{p}} s_{\mathbf{p}'} \langle \mathbf{b} | Z_{\mathbf{p}} Z_{\mathbf{p}'} | \mathbf{b} \rangle \right|^2 \\ &= \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} \left| \sum_{\mathbf{p}} (-1)^{\langle \mathbf{b}, \mathbf{p} \rangle} s_{\mathbf{p}} \sum_{\mathbf{p}'} (-1)^{\langle \mathbf{b}, \mathbf{p}' \rangle} s_{\mathbf{p}'} \right|^2, \end{aligned}$$

where the last equation follows from combining Eq. (33) with the appealing group structure of the  $Z_{\mathbf{p}}$ 's:  $Z_{\mathbf{p}} Z_{\mathbf{p}'} = Z_{\mathbf{p} \oplus \mathbf{p}'}$ , where  $\oplus$  denotes entry-wise addition modulo 2 (the set of all  $Z_{\mathbf{p}}$ 's form a maximal stabilizer group). We can now recognize two independent Walsh-Hadamard transforms of the  $2^n$ -dimensional vector  $s$  in this expression:

$$\mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} \left| \sum_{\mathbf{p}} (-1)^{\langle \mathbf{b}, \mathbf{p} \rangle} s_{\mathbf{p}} \sum_{\mathbf{p}'} (-1)^{\langle \mathbf{b}, \mathbf{p}' \rangle} s_{\mathbf{p}'} \right|^2 = 2^n \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} |\hat{s}_{\mathbf{b}}|^2$$

We already know from the CLT that the  $2^n$ -dimensional Walsh-Hadamard transform of  $s$  approaches a standard Gaussian vector:  $\hat{s} = (0, g_2, \dots, g_{2^n})^T$  with  $g_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2^{-n})$ . In the large  $N$  limit (CLT), the r.h.s. of the above display becomes an expected maximum of  $K = 2^n - 1$  squares of i.i.d. Gaussian variables with mean zero and variance  $\sigma^2 = 2^{-n}$ . Such expected maxima can be bounded using standard arguments, see e.g. [37, Lemma 5.1]:  $\mathbb{E} \max_{1 \leq i \leq K} |g_i|^2 \leq 4\sigma^2 \log(\sqrt{2}K)$  (the constants are chosen based on simplicity, not tightness). This allows us to conclude

$$\mathbb{E} \left\| \sum_{\mathbf{p}} Z_{\mathbf{p}} \right\|^2 = 2^n \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} |\hat{s}_{\mathbf{b}}|^2 \stackrel{N \rightarrow \infty}{\simeq} 2^n \mathbb{E} \max_{2 \leq i \leq N} |g_i|^2 \leq 2^n 4\sigma^2 \log(\sqrt{2}(2^n - 1)) \leq 4(n + 1/2).$$

Inserting linear and quadratic bound into Eq. (34) completes the argument.  $\square$

#### Acknowledgments:

The authors want to thank John Preskill and Yuan Su for valuable inputs and inspiring discussions. Earl Campbell and Nathan Wiebe provided insightful comments, as well as encouraging feedback. CC is thankful for Physics TA Relief Fellowship at Caltech. HH is supported by the Kortschak Scholars Program. RK acknowledges funding from ONR Award N00014-17-1-2146 and ARO Award W911NF121054). JAT gratefully acknowledges funding from the ONR Awards N00014-17-1-2146 and N00014-18-1-2363 and from NSF Award 1952777.

- 
- [1] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, page 20–30, New York, NY, USA, 1998. Association for Computing Machinery.
- [2] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

- [3] D. An and L. Lin. Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm. *arXiv preprint arXiv:1909.05500*, 2019.
- [4] R. Babbush, D. W. Berry, and H. Neven. Quantum simulation of the sachdev-ye-kitaev model by asymmetric qubitization. *Physical Review A*, 99(4), Apr 2019.
- [5] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan. Low-depth quantum simulation of materials. *Physical Review X*, 8(1):011044, 2018.
- [6] D. W. Berry, A. M. Childs, and R. Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 792–809. IEEE, 2015.
- [7] D. W. Berry, A. M. Childs, Y. Su, X. Wang, and N. Wiebe. Time-dependent hamiltonian simulation with  $l_1$ -norm scaling. *Quantum*, 4:254, 2020.
- [8] E. Campbell. Random compiler for fast hamiltonian simulation. *Physical review letters*, 123(7):070503, 2019.
- [9] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, 115(38):9456–9461, 2018.
- [10] A. M. Childs, A. Ostrander, and Y. Su. Faster quantum simulation by randomization. *Quantum*, 3:182, 2019.
- [11] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu. A theory of trotter error. *arXiv preprint arXiv:1912.08854*, 2019.
- [12] D. Christofides and K. Markström. Expansion properties of random cayley graphs and vertex transitive graphs via matrix martingales. *Random Structures & Algorithms*, 32(1):88–100, 2008.
- [13] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [14] D. J. H. Garling. *Inequalities: A Journey into Linear Analysis*. Cambridge University Press, 2007.
- [15] I. M. Georgescu, S. Ashhab, and F. Nori. Quantum simulation. *Reviews of Modern Physics*, 86(1):153, 2014.
- [16] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.
- [17] D. Huang, J. Niles-Weed, J. A. Tropp, and R. Ward. Matrix concentration for products. *arXiv preprint arXiv:2003.05437*, 2020.
- [18] H.-Y. Huang, K. Bharti, and P. Rebentrost. Near-term quantum algorithms for linear systems of equations. *arXiv preprint arXiv:1909.07344*, 2019.
- [19] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 2020.
- [20] T. Kathuria, S. Mukherjee, and N. Srivastava. On concentration inequalities for random matrix products. *arXiv preprint arXiv:2003.06319*, 2020.
- [21] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush. Quantum simulation of electronic structure with linear depth and connectivity. *Physical review letters*, 120(11):110501, 2018.
- [22] S. Lloyd. Universal quantum simulators. *Science*, pages 1073–1078, 1996.
- [23] G. H. Low and I. L. Chuang. Optimal hamiltonian simulation by quantum signal processing. *Physical review letters*, 118(1):010501, 2017.
- [24] J. Maldacena and D. Stanford. Remarks on the sachdev-ye-kitaev model. *Physical Review D*, 94(10):106002, 2016.
- [25] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, 92(1):015003, 2020.
- [26] R. I. Oliveira. The spectrum of random k-lifts of large graphs (with possibly large k), 2009.
- [27] Y. Ouyang, D. R. White, and E. T. Campbell. Compilation by stochastic hamiltonian sparsification. *Quantum*, 4:235, 2020.
- [28] P. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, et al. Scalable quantum simulation of molecular energies. *Physical Review X*, 6(3):031007, 2016.
- [29] G. Pisier and Q. Xu. Random series in the real interpolation spaces between the spaces  $vp$ . In J. Lindenstrauss and V. D. Milman, editors, *Geometrical Aspects of Functional Analysis*, pages 185–209, Berlin, Heidelberg, 1987. Springer Berlin Heidelberg.
- [30] J. Polchinski and V. Rosenhaus. The spectrum in the sachdev-ye-kitaev model. *Journal of High Energy Physics*, 2016(4):1, 2016.
- [31] S. Sachdev and J. Ye. Gapless spin-fluid ground state in a random quantum heisenberg magnet. *Physical review letters*, 70(21):3339, 1993.
- [32] Y. Subaşı, R. D. Somma, and D. Orsucci. Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing. *Physical review letters*, 122(6):060504, 2019.
- [33] M. Suzuki. Decomposition formulas of exponential operators and lie exponentials with some applications to quantum mechanics and statistical physics. *Journal of mathematical physics*, 26(4):601–612, 1985.
- [34] M. Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *Journal of Mathematical Physics*, 32(2):400–407, 1991.
- [35] J. Tropp et al. Freedman’s inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.
- [36] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathe-*

- matics*, 12(4):389–434, 2012.
- [37] R. van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.
  - [38] P. P. Varjú. Random walks in compact groups, 2012.
  - [39] R. Versendaal. Large deviations for random walks on lie groups, 2019.
  - [40] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.