Concentration for Random Product Formulas

Chi-Fang Chen,[1,*,†] Hsin-Yuan Huang,[2,3,†] Richard Kueng,[2,3,4] and Joel A. Tropp[3]

[1]*Department of Physics, Caltech, Pasadena, California, USA*
[2]*Institute for Quantum Information and Matter, Caltech, Pasadena, California, USA*
[3]*Department of Computing and Mathematical Sciences, Caltech, Pasadena, California, USA*
[4]*Institute for Integrated Circuits, Johannes Kepler University Linz, Austria*

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. QDRIFT achieves a gate count that does not explicitly depend on the number of terms in the Hamiltonian, which contrasts with Suzuki formulas. This work aims to understand the origin of this speedup by comprehensively analyzing 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 already 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. In contrast, in deterministic settings, such an improvement usually requires initial state knowledge. The proofs depend on concentration inequalities for vector and matrix martingales, and the framework is applicable to other randomized product formulas. Our bounds are saturated by certain commuting Hamiltonians.

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 [1–3], as well as solving linear algebra problems [4–6]. The task of digital quantum (dynamics) simulation can be phrased as a compiling problem: approximate a given unitary, say a Hamiltonian evolution $U = e^{-iHt}$, by a product of "simple" unitaries g_k:

$$U = e^{-iHt} \approx V = g_1 \cdots g_N. \tag{1}$$

Such a decomposition into elementary gates should obey two conditions. (i) It should accurately approximate the target unitary. In this work, we require that the error in

_____
*chifang@caltech.edu
†These authors contributed equally.

operator norm [7] satisfies $\|U - V\| \le \epsilon$ for some specified accuracy parameter ϵ. Moreover, (ii) the decomposition should cost as little as possible. Several cost functions make sense in this context, but we focus on the gate complexity, i.e., the number N of simple gates [8] on the right-hand side of Eq. (1).

The earliest compilation procedures for quantum simulation were based on *product formulas* [9,10], also known as *Trotterization*, or *splitting methods*. They depend on the idea of approximating the matrix exponential of a sum by a product of matrix exponentials.

We review one such construction below in Eq. (2). Subsequently, alternative principles have led to the development of other quantum simulation algorithms. These include linear combination of unitaries [11], quantum signal processing [12], and qubitization [13]. By and large, these algorithms rely on more powerful quantum-computing primitives to yield improved performance in accuracy and cost. We refer to Ref. [14] for a systematic review.

Despite these advanced simulation techniques, product formulas have recently undergone a renaissance [12,15–19]. They are not only simple and (comparatively) easy to implement on near-term devices, but they

also remain very competitive [19] provided that they incorporate information about the structure of the problem, such as initial state knowledge [20], locality [21], or the commutator structure of the Hamiltonian [19]. The purpose of this paper is to explore the randomized aspects of constructing product formulas.

We begin by reviewing the first-order Lie-Trotter formula, which is later contrasted with a randomized variant (QDRIFT). To that end, consider a quantum many-body Hamiltonian $H = \sum_{j=1}^{L} h_j$ composed of L simple terms h_j. The first-order formula cycles through each term in the Hamiltonian

$$U \approx \{\exp[-i(tL/N)h_L] \cdots \exp[-i(tL/N)h_1]\}^{N/L}. \quad (2)$$

To approximate the target unitary U up to accuracy ϵ in operator norm, a total gate count $N = \mathcal{O}(L\lambda^2 t^2/\epsilon)$ suffices [22] [19, Sec. 3]. In this expression, t is the simulation time, L is the number of terms, and $\lambda = \sum_j \|h_j\|$ summarizes the interaction strengths within H. The main idea is to cancel out the leading-order term in the Taylor expansion by including each of the L terms. Owing to this construction, the factor L remains in higher-order Suzuki formulas where the gate count $N = \mathcal{O}[L(\lambda t)^{1+o(1)}/\epsilon^{o(1)}]$,

TABLE I. *Error-bound comparison* QDRIFT *versus p th order Suzuki formulas* [19]. We consider systems with spatially local interaction in d dimension, power-law interaction in d dimension, and all-to-all k-local interaction. Mutiplicative overhead depending only on $p, k, \alpha, d, \log(n)$ are dropped, and practically we may think of $p = 4$ or 6. The t^2-dependence sets a time scale before which QDRIFT yields an advantage. For geometrically local Hamiltonians, the Suzuki formulas are effectively tight [21] [Exploiting locality, Ref. [21] further removes the $o(1)$ dependence], while QDRIFT already performs poorly after very short times [$t \sim \mathcal{O}(1/n)$]. However, once L gets very large, such as in k-local models, there are physically relevant time scales $t = \mathcal{O}[n^{(k-3)/2}]$ where QDRIFT becomes advantageous. Note that normalization conventions do affect the gate count substantially. For systems with random coefficients (e.g., SYK-like), one typically fixes $\sum_j \|h_j\|^2 = n$ [27,29]. But other conventions, like fixing $\lambda = \sum_j \|h_j\| = n$, are also widespread [30]. We refer to Appendix A for detailed calculations.

qDRIFT	pth order Suzuki [19]	Systems
$\lambda^2 t^2/\epsilon$	$L\|\|H\|\|_1 \lambda^{1/p} t^{1+1/p}/\epsilon^{1/p}$	General
$n^2 t^2/\epsilon$	$n^{1+(1/p)} t^{1+1/p} \epsilon^{1/p}$	Nearest neighbor
		Power law $(1/r^\alpha)$
$n^2 t^2/\epsilon$	$(nt)^{1+1/p+d/\alpha-d}/\epsilon^{d/\alpha-d+1/p}$	$2d \leq \alpha$
$n^2 t^2/\epsilon$	$n^{2+(1/p)} t^{1+1/p}/\epsilon^{1/p}$	$d \leq \alpha \leq 2d$
$n^{4-2(\alpha/d)} t^2/\epsilon$	$n^{3-\alpha/d+(2-\alpha/d)/p} t^{1+1/p}/\epsilon^{1/p}$	$0 \leq \alpha \leq d$
		k local
$n^{k+1} t^2/\epsilon$	$n^{3k-1/2+(k+1/2p)}$	$\|h_j\| = \mathcal{O}\left(\sqrt{1/n^{k-1}}\right)$
	$t^{1+1/p}/\epsilon^{1/p}$	
$n^2 t^2/\epsilon$	$n^{k+(1/p)} t^{1+1/p}/\epsilon^{1/p}$	$\|h_j\| = \mathcal{O}\left(1/n^{k-1}\right)$

even though the time dependence becomes nearly optimal. We refer to Table I for a sharper gate count incorporating commutators. Recently, researchers started using randomization to improve the performance of product formulas [17,18,23,24]. Campbell [17] introduced the QDRIFT algorithm, which approximates the target evolution $U = e^{-iHt}$ by a quantum channel that results from averaging products $V_N \cdots V_1$ of random unitaries. Each V_k 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 $(p_1, \ldots, p_L)$, constructed to match the leading order of time step

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

This approximation is achieved by averaging over a single (random) unitary. [In contrast, the first-order Suzuki formula (2) must cycle through all terms, which incurs an extra L-factor.] Independence among the V_k's then ensures

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

Campbell proved that the averaged Hamiltonian approximates the true Hamiltonian when the gate count satisfies

$$N = \mathcal{O}(\lambda^2 t^2/\epsilon). \quad (3)$$

Observe that the gate count is *independent* of the number of terms L in the Hamiltonian. A summary of this procedure is as follows.

Algorithm 0.1: (QDRIFT)

Inputs: Hamiltonian $H = \sum_{j=1}^{L} h_j$ with interaction strength $\lambda = \sum_j \|h_j\|$, evolution time t, and number of steps N.

At each t/N interval: evolve a random term in Hamiltonian

$$V_k = \exp[-i(t/N)X_k] \quad (4)$$

according to its importance

$$X_k \overset{\text{IID}}{\sim} X = \begin{cases} \frac{\lambda}{\|h_1\|} h_1 & \text{with prob. } p_1 = \frac{\|h_1\|}{\lambda} \\ \vdots & \\ \frac{\lambda}{\|h_L\|} h_L & \text{with prob. } p_L = \frac{\|h_L\|}{\lambda} \end{cases}.$$

Output: the unstructured (randomly generated) product formula

$$V^{(N)} = V_N \cdots V_1.$$

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 average of all possible product formulas is $\mathcal{V}^{(N)}(X) = \mathbb{E}[V_N \cdots V_1 X V_1^\dagger \cdots V_N^\dagger]$ and forms a completely positive trace-preserving (CPTP) map. The ideal unitary also forms a CPTP map given by $\mathcal{U}(X) = UXU^\dagger$. Campbell proves that Eq. (3) ensures that the two CPTP maps are ϵ close in diamond distance.

It is interesting to compare the gate count, Eq. (3), achieved by QDRIFT with very recent and powerful results for (deterministic) product formulas [19,21], see Table I. Broadly speaking, deterministic product formulas can achieve gate counts that are linear in time t and the number L of terms. The QDRIFT gate count, on the other hand, is independent of L, but quadratic in t. This implies that QDRIFT should be favored for short-time simulations rather than long-time simulations. For systems with many terms in the Hamiltonian ($L \gg 1$), such as in quantum chemistry and the Sachdev-Ye-Kitaev (SYK) model [25–28], there are indeed physically relevant time scales for which QDRIFT outperforms Suzuki formulas; see Table I.

To summarize, there are interesting quantum-simulation problems where the QDRIFT gate count, Eq. (3), compares favorably with very recent and powerful results about high-order Suzuki formulas. This advantage is a consequence of *randomization*. But we may ask whether the benefit arises from the random choice of an individual product formula or whether it is due to the averaging of many product formulas together. Which aspect produces the speedup?

In this work, we analyze why random product formulas are efficient. To do so, we study the performance of a *random instance* of the product formula $V_N \cdots V_1$. By contrasting this instance with the average channel $\mathcal{V}^{(N)}(X)$, we deduce that random sampling alone allows us to avoid the dependency on the number L of terms in the Hamiltonian; it is not necessary to average many different product formulas. Our results establish strong concentration bounds for two cases, depicted in Fig. 1.

First, let n denote the number of system constituents, e.g., 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 ϵ in operator norm.

By the probabilistic method, this result also establishes, for the first time, the *existence* of product formulas whose gate count N is independent of the number L of terms in the Hamiltonian but depends on the system size n instead. On the other hand, we cannot easily verify the quality of any given instance.

In practice, we often wish to evolve a fixed input quantum state ρ, 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

$$N \geq \Omega(t^2\lambda^2/\epsilon^2) \tag{5}$$

terms suffices to achieve an ϵ 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
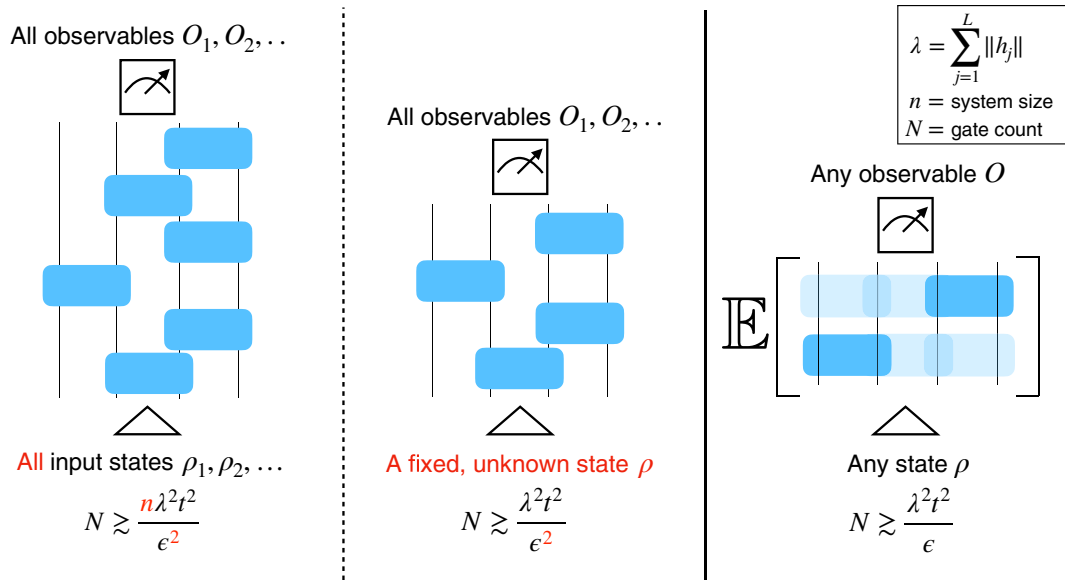

FIG. 1. Pictorial summary of the main results. (Left) To sample a product formula that works for all *n*-qubit input states and observables *with high probability*, the number of gates is larger than sampling a product formula that works for a fixed, yet arbitrary, input state (center). Resampling fresh product formulas every time (right) produces an average channel that requires even fewer gates; this is the original QDRIFT guarantee [17].

trace distance. Roughly speaking, 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. 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. Remarkably, this procedure does not exploit any knowledge of the particular input state.

Note that the gate count required for the original QDRIFT protocol, Eq. (3), and Eq. (5) differ only in the scaling with accuracy: order $1/\epsilon$ for the full QDRIFT channel versus order $1/\epsilon^2$ for a single random product formula with fixed input [31]. This discrepancy is reminiscent of the *mixing lemma* by Hastings [32] and Campbell [33]: mixing unitaries yields a "quadratic" improvement in accuracy. See Lemma 3.1 below for a statement.

In this work, we analyze several classes of randomized product formulas, including QDRIFT as a particular example. The underlying theory is far more general because it relies on powerful concentration results for matrix and vector martingales. The approach yields strong concentration results for any product of random unitary matrices, such as the ones that arise from randomized compiling [32,33]. We are confident that these ideas are applicable to other problems that involve functions of random matrices, such as Refs. [34,35].

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 Sec. D. Section III contains two instructive examples, as well as a nontechnical illustration of the underlying proof idea. We introduce the related background for martingales in Appendix B. Details and rigorous arguments are provided in Appendix C. The final appendix section (Appendix D) 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

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

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 $[-1, 1]$.

Our first main result 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.

Theorem 1 (qDRIFT: gate complexity for small diamond distance)**:** *Consider an n-qubit Hamiltonian $H = \sum_j h_j$ with $\lambda = \sum_j \|h_j\|$. Draw a randomized product formula $V_N \cdots V_1$ from Eq. (4) with gate count*

$$
N \geq \Omega\{[n + \log(1/\delta)]t^2\lambda^2/\epsilon^2\}. \tag{6}
$$

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

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

To keep notation as simple as possible, we formulate this result for pure input states $|\psi\rangle$. Convexity readily allows for extending the bound to mixed input states $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ as well. A complementary result bounds the expected approximation error in terms of gate count N.

Corollary 1.1 (qDRIFT: error bound in diamond distance)**:** *Consider an n-qubit Hamiltonian $H = \sum_j h_j$ with $\lambda = \sum_j \|h_j\|$. A randomized product formula $V_N \cdots V_1$, drawn from Eq. (4), has expected diamond-distance error*

$$
\mathbb{E}\left[\max_{|\psi\rangle:\text{ state}} \max_{O:\|O\|\leq 1} \left| \langle O\rangle_{V_N\cdots V_1|\psi\rangle} - \langle O\rangle_{\exp(-itH)|\psi\rangle} \right| \right]
$$

$$
\lesssim \sqrt{\frac{nt^2\lambda^2}{N}}.
$$

The symbol $\lesssim$ applies in the large-N regime and suppresses constants. The proof sketch is presented in Sec. C, and the detailed proof is given in Sec. 1.
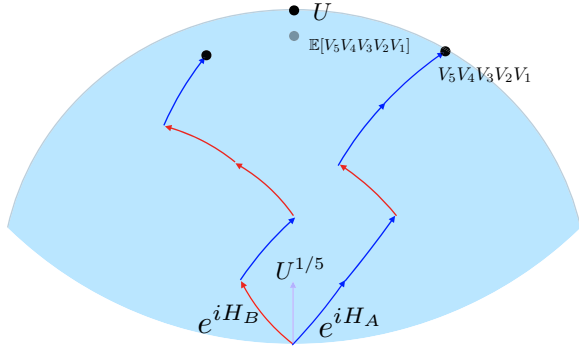
FIG. 2.   *Illustration of concentration effects for random walks (and their averages) on the unitary group.* The expectation $\mathbb{E}[V_N \cdots V_1]$ of a random product formula is not unitary, but it may be very close to the ideal evolution. A sampled random product formula $V_N \cdots V_1$ is 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 $\mathcal{O}\left(t^2\lambda^2/N\right)$.

For comparison, the error bounds for the average quantum channel, established in Ref. [17], imply that

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

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 randomly sampled product formula as a random walk on the group of $2^n \times 2^n$ unitary matrices. Figure 2 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 pay only for the *logarithm* of the dimension, which coincides with the number n of qubits. We discuss an intuitive example, where $\log(2^n)$ naturally arises from a union bound in Sec. III B. Formally, this prefactor is a general feature of concentration inequality for matrix martingales (Sec. B). 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 needs

only 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 2 (qDRIFT: gate complexity for fixed input)**:** *Consider an n-qubit Hamiltonian $H = \sum_j h_j$ with $\lambda = \sum_j \|h_j\|$ and any input quantum state $|\psi\rangle$. Draw a randomized product formula $V_N \cdots V_1$ from Eq. (4) with the number of gates*

$$N \geq \Omega[t^2\lambda^2 \log(1/\delta)/\epsilon^2]. \tag{7}$$

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

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

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 ϵ close to the ideal output state $\exp(-itH)|\psi\rangle$ in trace distance.

Again, convexity allows us to extend this bound to a (fixed) mixed state $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ as well. And, a complementary result bounds the expected approximation error in terms of gate count N.

Corollary 2.1 (qDRIFT: error bound in trace distance)**:** *Consider an n-qubit Hamiltonian $H = \sum_j h_j$ with $\lambda = \sum_j \|h_j\|$. A randomized product formula $V_N \cdots V_1$, drawn from Eq. (4), has expected trace-distance error for any fixed input*

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

Theorem 2 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× 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.

The probabilistic origin of this improvement is in stark contrast with deterministic constructions, where additional structure, such as low energy input states [20], is required to further reduce the gate count. Here, we make only one assumption that is much weaker: the state must be fixed and cannot depend on the concrete gate sequence being sampled.

We can even construct short product formulas that work for a moderate number of (arbitrary) input states by increasing the gate complexity slightly and invoking a union bound argument.

The proof of Theorem 2 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, what marks the difference from the unitary results (Theorem 1) is that *vector* concentration inequalities suffice. More precisely, we analyze the vector random walk using a geometric tool, called uniform smoothness [36]. The details appear in Sec. 2. The resulting *n*-fold improvement can also be understood as a result of switching the order of quantifiers, see Sec. III B for an explicit example.

C. Extension to general products of random unitaries

So far, we have presented our results exclusively for QDRIFT. But the underlying concentration analysis readily extends to more general random walks on the unitary group. Let $V_N \cdots V_1 \in U(2^n)$ be a random product designed to approximate a target unitary $U = U_N \cdots U_1$. We need two properties.

(i) *Causality:* for $1 \le k \le N$ the random selection of V_k can depend only on previous choices for $V_1, \ldots, V_{k-1}$:

$$\Pr[V_k|V_N \ldots V_{k+1}V_{k-1}, \ldots V_1] = \Pr[V_k|V_{k-1}, \ldots, V_1].$$
(8)

(ii) *Accurate approximation:* each realization of V_k (and their conditional expectation) must be close to the ideal unitary U_k. More precisely, let $a_k, b_k > 0$ be constants such that

$$\|V_k - \mathbb{E}_{k-1}V_k\| \le a_k \quad \text{and} \quad \|\mathbb{E}_{k-1}V_k - U_k\| \le b_k,$$
$$\text{where} \quad \mathbb{E}_{k-1}V_k := \mathbb{E}[V_k|V_{k-1}, \ldots, V_1]$$
(9)

almost surely for all $1 \le k \le N$. All concentration results from this work, as well as the main result from Ref. [17], do readily extend to random products that do obey these two properties.

Theorem 3 (general concentration bounds for products of random unitaries): *Let $V = V_N \cdot V_1 \in U(2^n)$ be a random product that approximates a target product $U = U_N \cdots U_1$ in a causal [Eq. (8)] and small-step [Eq. (9)] fashion. Then, the associated unitary channels $\mathcal{V}(X) = VXV^\dagger$ and $\mathcal{U}(X) = UXU^\dagger$ obey*

$$\|\mathcal{U} - \mathcal{V}\|_\diamond \le 2 \sum_{k=1}^{N} a_k \qquad \text{(worst case)},$$

$$\mathbb{E}\|\mathcal{U} - \mathcal{V}\|_\diamond \lesssim \sqrt{n \sum_{k=1}^{N} a_k^2 + 2 \sum_{k=1}^{N} b_k} \qquad \text{(typical case)},$$

$$\mathbb{E}\|\mathcal{U}(\rho) - \mathcal{V}(\rho)\|_1 \lesssim \sqrt{\sum_{k=1}^{N} a_k^2 + 2 \sum_{k=1}^{N} b_k} \qquad \text{(fixed input)},$$

$$\|\mathcal{U} - \mathbb{E}\mathcal{V}\|_\diamond \le 2 \sum_{k=1}^{N} b_k \qquad \text{(average case)},$$

where $\lesssim$ suppressed absolute constants.

1. Concentration of randomly permuted Suzuki formulas

So far, we have only introduced the first-order Lie-Trotter formulas (2). The second-order formula is typically called the Suzuki formula. For some short time $\tau > 0$, define

$$S_2(\tau) := \prod_{j=1}^{L} \exp[-i(\tau/2)h_j] \prod_{j=L}^{1} \exp[-i(\tau/2)h_j].$$

Higher-order formulas, also named after Suzuki, are constructed recursively from $S_2(\tau)$:

$$S_{2p}(\tau) := S_{2p-2}(q_p\tau)^2 S_{2p-2}[(1 - 4q_p)\tau]S_{2p-2}(q_p\tau)^2,$$

where $q_p := 1/(4 - 4^{1/(2p-1)})$ [10]. We can combine r rounds of $2p$th-order Suzuki formulas with time $\tau = t/r$ each to approximate a desired quantum evolution.

This yields

$$U = e^{-itH} = e^{-i(t/r)H} \cdots e^{-i(t/r)H}$$
$$\approx S_{2p}(t/r) \cdots S_{2p}(t/r) = S_{2p}(t/r)^r.$$

For ϵ fixed, we choose an appropriate number of rounds r and obtain a gate count $N \approx rL$ that scales as

$$N_{\text{det}} \gtrsim t\Lambda L^2 \left(\frac{t\Lambda L}{\epsilon} \right)^{1/2p} \quad \text{with} \quad \Lambda = \max_j \|h_j\|, \quad (10)$$

simple gates $e^{-i(t/r)h_j}$ to ensure worst-case approximation error $\|\mathcal{U} - \mathcal{V}\|_\diamond \le \epsilon$ [18]. Note that the $2p$th-order Suzuki formula $S_{2p}(\tau)$ does not specify a preferred order in which the terms in Hamiltonian h_j should appear. Childs, Ostrander, and Su observed that randomly permuting the order of Hamiltonian terms within each block $S_{2p}(t/r)$ can suppress the approximation error for low-order Suzuki formulas [18] considerably. More precisely, we reshuffle the ordering in each Suzuki block by applying uniformly random permutations $\pi_1, \ldots, \pi_r$ to the term indices ($h_j \mapsto h_{\pi_1(j)}$ for the first Suzuki block, etc.). Denote the resulting randomized Suzuki formula by $V_{2p}^{\pi_r \cdots \pi_1} = S_{2p}^{\pi_r}(t/r) \cdots S_{2p}^{\pi_1}(t/r)$. Childs, Ostrander, and Su proved that a total gate count

$$N_{\text{avg}} \gtrsim t\Lambda L^2 \left(\frac{t\Lambda}{\epsilon} \right)^{1/2p} \quad (11)$$

ensures that the average channel (averaged over all possible permutations) obeys $\|\mathcal{U} - \mathbb{E}_{\pi_r,\ldots,\pi_1} \mathcal{V}_{2p}^{\pi_r \cdots \pi_1}\|_\diamond \le \epsilon$.

We note in passing that Eq. (11) is tighter than the original bound presented in Ref. [18]. This slight improvement follows from replacing the original mixing lemma [32,33] in the proof of Childs *et al.* by a stronger statement derived in this work (Lemma 3.1 below).

Applying Theorem 3 to the problem at hand supplies strong concentration around this expected behavior.

Corollary 3.1 (concentration of randomly permuted Suzuki formulas): *Consider a n-qubit Hamiltonian $H = \sum_j h_j$ with $\Lambda = \max_j \|h_j\|$, an associated time-t evolution $U = e^{-itH}$ and a Suzuki order $2p$. Then, a total gate count of*

$$N_{\text{typ}} \gtrsim t\Lambda L^2 \left(\frac{n\Lambda L t}{\epsilon^2} \right)^{1/(4p+1)} + t\Lambda L^2 \left(\frac{t\Lambda}{\epsilon} \right)^{1/2p},$$

ensures that a randomly permuted Suzuki formula with r terms obeys $\mathbb{E}_{\pi_1,\ldots,\pi_r} \|\mathcal{U} - \mathcal{V}_{2p}^{\pi_r \cdots \pi_1} \cdots S_{2p}^{\pi_1}(t/r)\| \le \epsilon$ (typical case). For a fixed (but arbitrary) input state, the gate

count can be further reduced to

$$N_{\text{fix}} \gtrsim t\Lambda L^2 \left(\frac{\Lambda L t}{\epsilon^2} \right)^{1/(4p+1)} + t\Lambda L^2 \left(\frac{t\Lambda}{\epsilon} \right)^{1/2p}.$$

For simplicity, we omit the logarithmic dependence on failure probability δ.

In contrast to qDRFIT, the parameters n, ϵ, L parameters now all raised to the $1/(4p + 1)$st power, and the different randomized settings yield very much the same gate count $t\Lambda^2 L$ for large p. This is in accordance with observations provided in Ref. [18].

For illustration, we choose to directly import older results, Eq. (10), to compare with the averaged channel bounds, Eq. (11). However, Eq. (10) can possibly be sharpened to scale with commutator [19]. Future work remains to carry out commutator analysis for the averaged channel, Eq. (11), and then swiftly upgrade Corollary 3.1.

We refer to Appendix C (Sec. 3) for detailed statements and proofs.

2. Compiling

Beyond Hamiltonian simulation, random product also help suppressing error in compiling. The task of compiling turns a perfect circuit diagram into a sequence of executable gates. However, since the gate set is discrete (or imperfect), the compiler can return only an approximate circuit

$$\|U_N \cdots U_1 - V_N \cdots V_1\| \le \epsilon, \quad (12)$$

where each V_k will be further decomposed into universal gates (such as using the Solovay-Kiteav theorem, see e.g., Ref. [37]) up to some individual accuracy $\|U_k - V_k\| \le \eta$. In the worst case, the local error $\|U_k - V_k\| \le \eta$ accumulates linearly and we must conclude

$$\|U - V\| \le N\eta, \quad (13)$$

by means of the triangle inequality. This means that individual accuracy $\eta = \epsilon/N$ is required to ensure an overall approximation error of (at most) ϵ. This in turn, requires more gates for each approximation, because η-approximating V_k requires a gate count [38] proportional to $\log^c(\eta)$.

Randomized compiling [32,33] addresses precisely this issue. It uses random gate synthesis to avoid that individual approximation errors add up "coherently" over the entire compilation. The resulting "incoherent" error accumulation can be much more favorable and the mixing lemma [32,33] makes this intuition precise. In the following, we present a sharpened version of this statement that seems to be novel.

Lemma 3.1 ([Mixing lemma (sharpened)]): *Let U_k be a fixed unitary and V_k a random approximation thereof that obeys $\|U_k - \mathbb{E}V_k\| \le b$, for some $b > 0$. Then, the associated channels obey*

$$\|\mathcal{U}_k - \mathbb{E}\mathcal{V}_k\|_\diamond \le 2b.$$

We refer to Appendix C (Sec. C1a) for a self-contained proof.

Back to compiling, $\|\mathbb{E}V_k - U_k\|$ can as small as $\mathcal{O}(\eta^2)$ [33] by mixing appropriate synthesis protocols for V_k. This yields an improved bound for the average channel,

$$\|\mathcal{U} - \mathbb{E}\mathcal{V}\|_\diamond \le \mathcal{O}(N\eta^2). \tag{14}$$

In words, the individual accuracy needed is suppressed quadratically to $\eta = \Omega(\sqrt{\epsilon/N})$. In Ref. [33], Campbell pointed out that this square-root improvement may turn into a multiplicative overhead reduction, depending on the gate synthesis efficiency $\log^c(\sqrt{\eta}) = \log^c(\eta)/2^c$. Using Theorem 3, we can immediately bound the performance of randomized compiling *without mixing*.

Corollary 3.2 (randomized compiling without mixing)): *Suppose that we wish to approximate $U = U_N \cdots U_1$ by a gate collection $V = V_N \cdots V_1$ such that each V_k is random and obeys $\|U_k - V_k\| \le \eta$ almost surely. Then,*

$$\mathbb{E}\|U - V\| \lesssim \sqrt{nN}\,\eta.$$

What is more, the $\sqrt{n}$ factor on the rhs disappears if we restrict attention to a fixed input state.

This translates to individual accuracy $\eta = \mathcal{O}(\epsilon/\sqrt{nN})$, and $\eta = \mathcal{O}(\epsilon/\sqrt{N})$, respectively. Both results interpolate between the worst case, Eq. (13) ("coherent" error accumulation), and best case, Eq. (14) ("incoherent" error accumulation).

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, $H = 1/n - 1 \sum_{i=1}^{n-1} X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1}$ and we view this as a sum of $3(n-1)$ simple terms. The normalization keeps the interaction strength $\lambda = 3$ as a constant and we consider a constant time evolution $t = 2$. The numerical experiments for the error under various setups using different gate count N and different system sizes n are shown in Fig. 3. We can see that the error ϵ when we consider all input states scales as $\sqrt{n}$. In contrast, the error ϵ stays roughly the same when we consider only a single input state. This is in accordance with the theoretical predictions presented in Theorems 1 and 2.

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 noninteracting Hamiltonian with
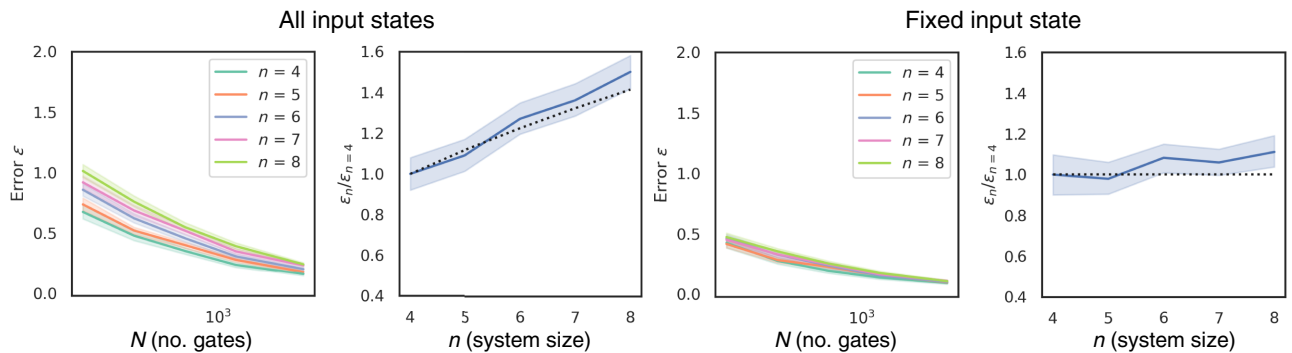

FIG. 3. *Numerical experiments for simulating 1D Heisenberg model under different gate count N. In* All input states *(left), we consider $\epsilon = \|U - V_N \ldots V_1\|_\infty$, which considers the error over all input states and observables. In* Fixed input state *(right), we consider $\epsilon = \|U|\psi\rangle - V_N \ldots V_1 |\psi\rangle\|_{\ell_2}$, which considers the error over all observables. The input state $|\psi\rangle$ is chosen to be the tensor product of single-qubit Haar-random states. For both* All input state *and* Fixed input state, *we give an additional plot showing how the error ϵ increases as the system size n increases for a fixed number of gates $N = 160$. The y-axis is normalized using the average error for system size $n = 4$ over 50 independent runs. Bounds in Theorems 1 and 2 show that the relative error $\epsilon_n/\epsilon_{n=4}$ scales as $\sqrt{n/4}$ for* All input state *and stays as constant 1 for* Fixed input state, *which are shown as the dotted lines. The shaded regions are the standard deviation over 50 independent runs.*

a Pauli-Z operator acting on each qubit:

$$H = \frac{1}{n} \sum_{k=1}^{n} Z_k, \quad \text{where}$$

$$Z_k = \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{(k-1) \text{ times}} \otimes Z \otimes \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{(n-k) \text{ times}} \tag{15}$$

for $1 \le k \le n$. The relevant parameters are $L = n$ (number of terms), $\lambda = 1/n \sum_{k=1}^{n} \|Z_k\| = 1$ (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 [17], see also Proposition 3.2 below.

Corollary 3.3: *Fix a target accuracy ϵ and set $N = t^2 \lambda^2 / \epsilon \approx 10/\epsilon$. Then, N successive applications of the QDRIFT single-step average $\mathcal{V}(\rho) = 1/n \sum_k \exp[-i(\pi/N) Z_k] \otimes \mathbb{I}^{(else)} \rho \exp[i(\pi/N)Z_k] \otimes \mathbb{I}^{(else)}$ (Campbell's black box) approximate the target unitary channel $\mathcal{U}(\rho) = U\rho U^\dagger$ up to accuracy ϵ in diamond distance. In particular, $\frac{1}{2} \|\mathcal{V}^{(N)}(|\psi\rangle\langle\psi|) - U|\psi\rangle\langle\psi|U^\dagger\|_1 \le \epsilon$ 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 $V_N \cdots V_1$ and the actual target U, see Fig. 4 for an illustration.
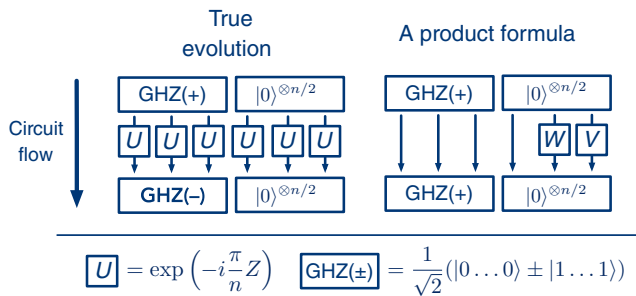

FIG. 4. *Illustration of the worst-case input for a product formula simulating evolution of a simple Hamiltonian.* The Hamiltonian $H = 1/n \sum_{k=1}^{n} Z_k$ produces a time evolution that factorizes into single qubit unitaries U (left). A product formula with fewer than $n/2$ single-site terms (right) 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 $\{U = \exp[-i(\pi/n)Z]\}$ can accumulate and lead to a maximal approximation error $[\langle \text{GHZ}(+), \text{GHZ}(-) \rangle = 0]$.

Lemma 3.2: *Assume that n is an even number. It is impossible to accurately approximate the time evolution U defined in Eq. (15) with $N < n/2$ elementary gates of the form $V_i = \exp[-i(\pi/N)Z_{k(i)}] \otimes \mathbb{I}^{(else)}$. 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 Hamiltonian (15) commute. Hence, the associated target evolution factorizes nicely into tensor products: $U = \exp(-i\pi H) = \exp[-i(\pi/n)Z_1] \otimes \cdots \otimes \exp[-i(\pi/n)Z_n]$. Up to a global phase, each single-qubit unitary affects the computational basis in the following fashion: $\exp[-i(\pi/n)Z]|0\rangle = |0\rangle$ and $\exp[-i(\pi/n)Z]|1\rangle = \exp(i\frac{2\pi}{n})|1\rangle$. These small phase shifts can add up for states that are in superposition. Consider the tensor product of a Greenberger-Horne-Zeilinger (GHZ) state on $n/2$ qubits with the all-zeroes state on the remaining half: $|\tilde\psi\rangle = 1/\sqrt{2}(|0\rangle^{\otimes n/2} + |1\rangle^{\otimes n/2}) \otimes |0\rangle^{\otimes n/2}$. Then,

$$\begin{aligned}
U|\tilde\psi\rangle &= \exp(-i\frac{2\pi}{n}Z)^{\otimes n}|\tilde\psi\rangle \\
&= \tfrac{1}{\sqrt{2}}(|0\rangle + [\exp(i\tfrac{2\pi}{n})]^{n/2}|1\rangle) \otimes |0\rangle^{\otimes n/2} \\
&= \tfrac{1}{\sqrt{2}}(|0\rangle^{\otimes n/2} - |1\rangle^{\otimes n/2}) \otimes |0\rangle^{\otimes n/2}
\end{aligned}$$

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: $\|U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1 = \|U|\psi\rangle\langle\psi|U^\dagger - |\psi\rangle\langle\psi|\|_1 = 1$. ∎

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 3.2 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 2 asserts

that a collection of $N \gtrsim \pi^2/\epsilon^2$ randomly selected single-qubit gates approximate the time evolution, Eq. (15), of any fixed input state $|\psi\rangle$ up to accuracy ϵ in trace distance. While this gate count is considerably larger than the one put forth in Corollary 3.3, 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.

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:

$$H = \frac{1}{2^n} \sum_{\mathbf{p}\in\{0,1\}^n} \alpha_{\mathbf{p}} Z_{\mathbf{p}} \quad \text{where}$$

$$Z_{\mathbf{p}} = Z_{(p_1,\dots,p_n)} = Z^{p_1} \otimes \cdots \otimes Z^{p_n} \quad (16)$$

(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 $\mathcal{O}(L) = \mathcal{O}(2^n)$ by putting down every term. In contrast, by random sampling, Theorem 1 yields a gate complexity of $\mathcal{O}(n/\epsilon^2)$ [Theorem 2 yields $\mathcal{O}(1/\epsilon^2)$]. This is an exponential improvement in 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, and can be efficiently estimated by random sampling. 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 Hamiltonian (16) are diagonal in the computational basis. This implies

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

$$= \exp\left(-i\frac{1}{2^n}\sum_{\mathbf{p}\in\{0,1\}^n}c_{\mathbf{p}}(\mathbf{b})\right)|\mathbf{b}\rangle$$

$$:= e^{-iS(\mathbf{b})}|\mathbf{b}\rangle, \quad (17)$$

where $c_{\mathbf{p}}(\mathbf{b}) = \alpha_{\mathbf{p}}\langle\mathbf{b}|Z_{\mathbf{p}}|\mathbf{b}\rangle \in \{-1,1\}$. When we instead subsample an effective Hamiltonian $\hat{H}$ comprised of N
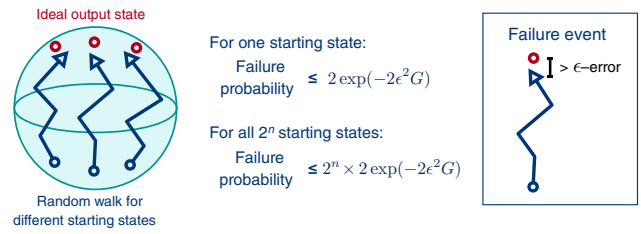

FIG. 5. *Illustration of the probabilistic proof for the commuting Hamiltonian given in Eq. (16).* We consider all the 2^n computational basis states as the starting state. The probability for one of the starting state to incur at least an error ϵ is exponentially smaller than the probability for the maximum of the 2^n starting states to incur error $> \epsilon$. However the failure probability is exponential suppressed by increasing the gate count N. Hence one need only to set $N = n/\epsilon^2$.

randomly selected terms, the constructed product formula would be

$$\exp(-i\hat{H}) = \exp\left(-i\frac{1}{N}\alpha_{\mathbf{p}_N}Z_{\mathbf{p}_N}\right)\cdots\exp\left(-i\frac{1}{N}\alpha_{\mathbf{p}_1}Z_{\mathbf{p}_1}\right)|\mathbf{b}\rangle$$

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

By Hoeffding's inequality, $\hat{S}(\mathbf{b}) = 1/N\sum_k c_{\mathbf{p}_k}(\mathbf{b})$ should concentrate around $S(\mathbf{b}) = 2^{-n}\sum_{\mathbf{p}\in\{0,1\}^n}c_{\mathbf{p}}(\mathbf{b})$ with standard deviation $1/\sqrt{N}$ and an exponentially decaying tail (think Monte Carlo). An illustration and some helpful facts can be found in Fig. 5.

Let us now fix an (arbitrary) n-qubit basis state. Then, the probability of an ϵ-deviation (or larger), i.e., $|\hat{S}(b) - S(b)| > \epsilon$, can be bounded by $1/e$ if we choose $N = 1/\epsilon^2$. However, if we wish the effective Hamiltonian to work *for all* basis states simultaneously, we would choose a larger $N = n/\epsilon^2$ to ensure that the deviation probability is further suppressed exponentially to $1/e^n$. By a union bound, $|\hat{S}(b) - S(b)| \leq \epsilon$ for all 2^n computational basis states $|\mathbf{b}\rangle$ with probability at least $1 - 2^n/e^n$. Albeit in the simplest example (commuting Hamiltonian), this demonstrates that a random product formula $\exp(-i\hat{H})$ can accurately simulate $\exp(-iH)$ up to error ϵ with only $N = n/\epsilon^2$ gates, which further reduces to $\mathcal{O}(1/\epsilon^2)$ for an arbitrary basis state. The powerful tool of concentration for matrix (vector) martingales allows us to prove the same statement for any (noncommuting) many-body Hamiltonian.

We return to this example Hamiltonian in Appendix D 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 Theorems 1 and 2

This section sketches the main ideas and tools required to establish Theorems 1 and 2. The other results follow from more elementary arguments. Detailed arguments and rigorous statements are provided in Appendix C below.

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

$$\mathcal{U}(|\psi\rangle\langle\psi|) = U|\psi\rangle\langle\psi|U^\dagger \quad \text{where}$$

$$U = \exp\left(-itH\right) = \exp\left(-it\sum_{j=1}^{L} h_j\right).$$

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

$$\mathcal{V}^{(N)}(|\psi\rangle\langle\psi|) = \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1(|\psi\rangle\langle\psi|)$$

$$= V_N \cdots V_1 |\psi\rangle\langle\psi| V_1^\dagger \cdots V_N^\dagger.$$

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 ρ 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

$$\frac{1}{2}\|\mathcal{E} - \mathcal{F}\|_\diamond = \frac{1}{2} \max_{|\psi\rangle\langle\psi|} \|\mathcal{E} \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \mathcal{F} \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1, \tag{18}$$

where the maximization ranges over all pure [39] input states $|\psi\rangle\langle\psi|$ and $\|\cdot\|_1$ denotes the trace norm. First, we relate the diamond distance, Eq. (18), between the channels $\mathcal{V}^{(N)}$ and $\mathcal{U}$, which are both unitary, to an operator-norm distance of the associated matrices:

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

$$\leq \|V_N \cdots V_1 - U\|, \tag{19}$$

see Lemma 3.4 below. 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, Theorem 3.55].

Now, we can deal with the independent and identically distributed (IID) random matrices $V_N, \ldots, V_1$ in the more familiar operator norm. Add and subtract the expected product $\mathbb{E}[V_N \cdots V_1] = \mathbb{E}[V_N] \cdots \mathbb{E}[V_1] =$

$(\mathbb{E}V)^N$ to decompose the operator-norm difference into two qualitatively different contributions:

$$\|V_N \cdots V_1 - U\|$$

$$\leq \underbrace{\|(\mathbb{E}V)^N - U\|}_{\text{deterministic bias}} + \underbrace{\|V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1]\|}_{\text{random fluctuation}}. \tag{20}$$

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, Eq. (4), 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 3.5 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:

$$\|\mathbb{E}[V_N \cdots V_1] - U\| = \|(\mathbb{E}V)^N - (U^{1/N})^N\|$$

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

$$\leq \frac{t^2\lambda^2}{N}, \tag{21}$$

see Lemma 3.6 below. The first inequality is obtained from a telescoping sum. This upper bound diminishes as the number of steps N increases. For $\epsilon > 0$,

$$N \geq \frac{2t^2\lambda^2}{\epsilon} \quad \text{ensures} \quad \|(\mathbb{E}V)^N - U\| \leq \frac{\epsilon}{2}. \tag{22}$$

ii. *Random fluctuation:* we also need to control the deviation of a product of IID unitaries $V_N \cdots V_1$ from its expectation $\mathbb{E}[V_N \cdots V_1] = (\mathbb{E}V)^N$ in operator norm. We achieve this by applying modern matrix martingale tools, which may be of independent interest. In short (see Appendix B for a more thorough introduction), a *martingale* is a random process $\{B_k : k = 0, \ldots, N\}$ such that the distribution of B_k depends only on past elements $B_{k-1}, \ldots, B_1$ ("causality") and also $\mathbb{E}[B_k|B_{k-1} \cdots B_1] = B_{k-1}$ ("on average, tomorrow is the same as today"). To control fluctuations, we introduce a martingale that interpolates between the extreme cases we need to compare:

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

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

Note that adjacent elements of this process differ only 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, and thus $\mathbb{E}_{k-1}B_k = B_{k-1}$. Moreover, the current iterate B_k depends only on realizations in the past and the interpolation process $\{B_k : k = 0, \ldots, N\}$ forms a martingale comprised of $d \times d$ matrices, where $d = 2^n$ is the Hilbert-space dimension. The discrepancies are small: $\|V_k - (\mathbb{E}V)\| \leq 2t\lambda/N$, because each realization of V is also very close to the identity matrix. Powerful tail bounds for matrix-valued martingales (which are relatively modern compared to their scalar counterparts) are available in the literature [41,42]. Adapting these results to the task at hand yields the bound

$$\Pr\left[\|V_N \cdots V_1 - (\mathbb{E}V)^N\| \geq \epsilon/2\right] \tag{23}$$

$$\leq 2d \exp\left(-\frac{N\epsilon^2}{44t^2\lambda^2}\right), \tag{24}$$

see Proposition 3.3 below. 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 sub-Gaussian with variance proportional to $N \times (\lambda^2 t^2/N^2)$. There is an extra dimensional factor $d = 2^n$ that arises because the martingale is matrix valued. It converts to the number of qubits $n = \log(d)$ in the gate count N. For error parameters $\epsilon, \delta \in (0, 1)$, a gate count obeying

$$N \geq 44\frac{t^2\lambda^2}{\epsilon^2}\log(2d/\delta) \quad \text{implies}$$

$$\|V_N \cdots V_1 - (\mathbb{E}V)^N\| \leq \frac{\epsilon}{2} \text{ with probability } > 1 - \delta. \tag{25}$$

Theorem 1 can be derived by combining the bound, Eq. (22), for the deterministic bias with the bound, Eq. (25), for random fluctuations. This provides a high probability error bound in the diamond distance when $N \geq \Omega(nt^2\lambda^2/\epsilon^2)$.

Corollary 1.1 is derived by integrating the tail bounds of Theorem 1. This produces

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

where the symbol $\lesssim$ suppresses a modest multiplicative constant. We refer to Appendix C (Sec. d) for details.

With fixed input (Theorem 2), we use convexity to restrict our attention to a fixed, pure input state $|\psi\rangle$. We then need to compare $U|\psi\rangle$ with $V_N \cdots V_1|\psi\rangle$, which can be achieved by constructing an interpolating random process that describes a *vector*-valued martingale. We then

call another concentration inequality (Lemma 3.7)

$$\Pr\left[\|(V_N \cdots V_1 - \mathbb{E}[V]^N)|\psi\rangle\|_{\ell_2} > \epsilon\right] \leq \exp\left(\frac{-\epsilon^2 N}{8et^2\lambda^2}\right),$$

which, in contrast, does not contain a dimensional factor. This is the reason why Theorem 2 and Corollary 2.1 do not depend on system size at all.

IV. DISCUSSION AND OUTLOOK

This work shows interesting characteristics of randomization that might help to further improve quantum simulation. *(a)* By studying typical unitary instances of QDRFIT, we have shown that L independence (the number of terms in Hamiltonian) of QDRIFT attributes to randomly sampling terms; mixing different realizations is not essential. *(b)* Gate complexities can be reduced substantially by restricting attention to a particular input state or/and target observable. Simple randomized compilation procedures, like QDRIFT, do not utilize extra information. We believe that more specialized algorithms might be able to exploit additional structure. *(c)* We have shown that channel averages can be much closer to the ideal evolution than any individual product formula, however, in the case of QDRIFT it is only a quadratic improvement $1/\epsilon^2 \to 1/\epsilon$. This can be seen as a manifestation of Jensen's inequality for convex functions (distance to ideal evolution) and we may also see such behavior in other quantum simulation algorithms. To our knowledge, we also provide the first matrix concentration analysis for a randomly sampled product formulas and—more generally—Hamiltonian simulation. Similar proof techniques readily apply to any random product formula, e.g., randomly permuted Suzuki formulas [18], and in fact any (causal) random unitary product, e.g., randomized compiling [32,33].

Beyond random products, we expect the developed matrix concentration tools to be useful for controlling stochastic errors in other quantum-computing applications, as well as analyzing properties of random Hamiltonians [34].

APPENDIX A: CALCULATIONS FOR TABLE I

Here, we provide detailed calculations for the numbers presented in Table I. Recall from Eq. (3) that QDRIFT achieves a gate count proportional to $\lambda^2 t^2/\epsilon$, where $\lambda = \sum_j \|h_j\|$ is the 1-norm. For pth-order product formulas, we quote the state-of-the-art Trotter error analysis from Ref. [19] with gate count

$$N = \mathcal{O}\left(\|H\|_1 \lambda^{1/p} \frac{t^{1+1/p}}{\epsilon^{1/p}}\right). \qquad (A1)$$

And for a k-local Hamiltonian, the induced 1-norm (using notation from Ref. [19]) $\|H\|_1$ is defined as

$$\|H\|_1 := \max_\ell \max_{j_\ell} \sum_{j_1 \cdots j_{\ell-1} j_{\ell+1} \cdots j_k} \|h_{j_1,\ldots,j_\ell,\ldots,j_k}\|, \quad (A2)$$

which is smaller than the 1-norm λ. Here, we also use the convention that $\mathcal{O}(\cdot)$ absorbs constants that may depend on k, p, α, d, Poly $\log(n)$ as long as they are independent of system size n, time t, and the Hamiltonian H. For Suzuki formulas, values like $p = 4$ or $p = 6$ are fairly standard. For larger p, the constant overheads get prohibitively expensive.

Nearest-neighbor interaction. On an integer lattice $[0, R]^d$, the nearest-neighbor interacting Hamiltonian reads

$$H = \sum_{\mathbf{r} \in [0,R]^d} \sum_{|\mathbf{r}'-\mathbf{r}|=1} h_{\mathbf{r},\mathbf{r}'} \quad \text{with} \quad \|h_{\mathbf{r},\mathbf{r}'}\| = \mathcal{O}(1). \quad (A3)$$

This, in turn, implies

$$\lambda = \mathcal{O}(R^d) = \mathcal{O}(n) \quad \text{and} \qquad (A4)$$

$$\|H\|_1 = \mathcal{O}(1). \qquad (A5)$$

Power-law interaction. On an integer lattice $[0, R]^d$, the power-law interacting Hamiltonian is 2-local and reads

$$H = \sum_{\mathbf{r},\mathbf{r}' \in [0,R]^d} h_{\mathbf{r},\mathbf{r}'} \quad \text{with} \quad \|h_{\mathbf{r},\mathbf{r}'}\| = \mathcal{O}\left(\frac{1}{|\mathbf{r}'-\mathbf{r}|^\alpha}\right). \qquad (A6)$$

This, in turn, implies

$$\|H\|_1 = \begin{cases} \mathcal{O}(1) & \text{if } \alpha > d \\ \mathcal{O}[\log(R)] = \mathcal{O}[\log(n)] & \text{if } \alpha = d \\ \mathcal{O}(R^{d-\alpha}) = \mathcal{O}(n^{1-\alpha/d}) & \text{if } \alpha < d \end{cases} \quad \text{and}$$
$$(A7)$$

$$\lambda = \mathcal{O}(n\|H\|_1). \qquad (A8)$$

Note that the $\|H\|_1$ is essentially the single-site energy integral. The total number of terms is $L = \mathcal{O}(n^2)$. For $\alpha >$

$2d$, as in Ref. [19] one can sacrifice time dependence and truncate terms with $|\mathbf{r}' - \mathbf{r}| > \ell$, where $\ell = \mathcal{O}(nt/\epsilon)^{1/(\alpha-d)}$ is tuned to match the Trotter error with the truncation error. In this case,

$$L = \mathcal{O}(n\ell^d) = \mathcal{O}\left[n\left(\frac{nt}{\epsilon}\right)^{d/(\alpha-d)}\right], \qquad (A9)$$

$$\|H\|_1 = \mathcal{O}(1) \quad (\alpha > 2d), \qquad (A10)$$

$$\lambda = \mathcal{O}(n\|H\|_1) = \mathcal{O}(n). \qquad (A11)$$

k-local Hamiltonians (**SYK-like normalization**). Consider an all-to-all interacting k-local Hamiltonian [27]

$$H := \sum_{j_1 < \cdots < j_k \leq n} h_{j_1 \cdots j_k} = \sum_{j_1 < \cdots < j_k \leq n} g_{j_1 \cdots j_k} Z_{j_1 \cdots j_k}$$

$$\text{with normalization} \quad \|Z\| = \sqrt{\frac{J^2(k-1)!}{kn^{k-1}}}, \quad (A12)$$

and $g_{j_1 \cdots j_k}$ are random couplings sampled independently from the standard Gaussian distribution, $Z_{j_1 \cdots j_k}$ are deterministic k-local operators, and J is some energy scale. The number of terms is $L = \binom{n}{k} = \mathcal{O}(n^k)$ and the normalization is chosen such that

$$\sum_{j_1 < \cdots < j_k \leq n} \|Z_{j_1 \cdots j_k}\|^2 = \mathcal{O}(n). \qquad (A13)$$

However, let us drop $g_{j_1 \cdots j_k}$ and not bother with the randomness. For the cautious reader, we should call a union bound that $\Pr(\max_{j_1 \cdots j_k} |g_{j_1 \cdots j_k}| \geq \epsilon) \leq Le^{-\epsilon^2}$, i.e., we at most lose an extra factor of $\mathcal{O}[\sqrt{\log(L)}] = \mathcal{O}[\sqrt{\log(n)}]$. Therefore, up to a $\sqrt{\log(n)}$ overhead,

$$\|H\|_1 = \sum_{j_2, \ldots, j_k} \|h_{j_1, \ldots, j_k}\| = \mathcal{O}(n^{k-1} \times n^{-(k-1)/2})$$

$$= \mathcal{O}(n^{(k-1)/2}), \qquad (A14)$$

$$\lambda = \mathcal{O}(n^k \times n^{-(k-1)/2}) = \mathcal{O}(n^{(k+1)/2}) \qquad (A15)$$

in the calculation we fix j_1 due to symmetry.

k-local Hamiltonians (**1-norm normalization**). In some other settings, one may choose [30]

$$H := \sum_{j_1 < \cdots < j_k \leq n} h_{j_1 \cdots j_k} = \sum_{j_1 < \cdots < j_k \leq n} Z_{j_1 \cdots j_k}$$

$$\text{with normalization} \quad \|Z\| = \mathcal{O}(1/n^{k-1}), \quad (A16)$$

and without randomly sampled coefficients. The normalization is such that the 1-norm is extensive

$$\sum_{j_1 < \cdots < j_k \leq n} \|Z_{j_1 \cdots j_k}\| = \mathcal{O}(n). \qquad (A17)$$

This in turn implies

$$\|H\|_1 = \sum_{j_2,\dots,j_k} \|h_{j_1,\dots,j_k}\| = \mathcal{O}(n^{k-1} \times n^{-(k-1)}) = \mathcal{O}(1),$$
(A18)

$$\lambda = \mathcal{O}(n).$$
(A19)

APPENDIX B: MATRIX AND VECTOR-VALUED MARTINGALES.

Let $X_1, \dots, X_N$ be IID random variables. Then, the strong law of large numbers (LLN) implies that the sample mean $\hat{\mu} = (1/N) \sum_{k=1}^{N} X_k$ converges to the actual mean $\mu = \mathbb{E}X_k$ ($\hat{\mu} \approx \mu$ as $N \to \infty$). In fact, even for finite N, the sample average is likely to be close to the expectation. This behavior is called *concentration*. It turns out that concentration is a surprisingly generic phenomenon, and it kicks in earlier than one might expect. Asymptotically large sample sizes ($N \to \infty$), which are essential for the law of large numbers and the central limit theorem, are not required establish concentration. An example is Bernstein's inequality for sums of bounded random numbers.

Fact 3.1 (Bernstein's inequality): *Let $X_1, \dots, X_N$ be independent random variables that obey $\mathbb{E}X_i = 0$ and $|X_i| \leq R$ almost surely. Then, for $\epsilon > 0$,*

$$\Pr\left[\left|\sum_{k=1}^{N} X_k\right| \geq \epsilon\right] \leq 2 \exp\left(\frac{\epsilon^2/2}{v^2 + R\epsilon/3}\right) \quad where$$

$$v = \sum_{k=1}^{N} \mathbb{E}X_k^2.$$
(B1)

Bernstein's inequality equips random sums with a LLN-type concentration behavior already for nonasymptotic sample sizes $N \gtrsim \max\{v^2/\epsilon^2, R/\epsilon\}$. However, it requires strong assumptions on the random variables involved. They need to be independent, bounded scalars. In fact, random processes may concentrate under much weaker conditions.

Martingales form a richer family of processes that capture more realistic random processes and that still enjoy powerful concentration inequalities. Let us offer a minimal technical introduction following Tropp [42]. Consider a filtration of the master sigma algebra $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \cdots \subset \mathcal{F}_k \subset \cdots \mathcal{F}$, where we denote the conditional expectation with respect to $\mathcal{F}_k$ by the symbol $\mathbb{E}_k$. A *martingale* is a sequence $\{B_0, B_1, B_2, \dots\}$ of random variables that satisfies

$$\sigma(B_k) \subset \mathcal{F}_k \qquad \text{(causality)},$$
(B2)

$$\mathbb{E}_{k-1}B_k := \mathbb{E}[B_k|B_{k-1}, \dots, B_0] = B_{k-1} \quad \text{(status quo)}.$$
(B3)

Heuristically, we can think of k as a "time" index, and $\mathcal{F}_k$ contains all possible events that are determined by the history up to time k. The present, also known as B_k may depend on the past (i.e., $B_0, \dots, B_{k-1}$), but it cannot depend on the future ("causality"). The second condition formalizes the requirement that, on average, tomorrow (B_k) is the same as today (B_{k-1}).

A martingale sequence may be composed of scalars, e.g., a one-dimensional random walk, but we can also consider vector- or matrix-valued martingales. Analyzing concentration for vector- and matrix-valued martingales is an ongoing field of research; for example, see Refs. [36,42,43], but many powerful concentration inequalities already exist. In this work, we use the matrix generalization of Freedman's inequality (due to one of the authors). Let $\mathbb{M}_{d \times d}$ denote the space of complex-valued $d \times d$ matrices.

Fact 3.2 (Matrix Freedman [42, Corollary 1.3]): *Let $\{B_k : k = 0, \dots, \ell, \dots\} \subset \mathbb{M}_{d \times d}$ be a matrix martingale. Assume that the associated difference sequence $C_k = B_k - B_{k-1}$ obeys $\|C_k\| \leq R$ almost surely. Define the random variable*

$$w_\ell := \max\left(\left\|\sum_{k=1}^{\ell} \mathbb{E}_{k-1} C_k^\dagger C_k\right\|, \left\|\sum_{k=1}^{\ell} \mathbb{E}_{k-1} C_k C_k^\dagger\right\|\right).$$
(B4)

Then, for any $\tau > 0$

$$\Pr[\exists \ell : \|B_\ell - B_0\| \geq \tau, w_\ell \leq v] \leq 2d \exp\left(\frac{-\tau^2/2}{v + R\tau/3}\right).$$

This bound exponentially suppresses the probability of the following undesirable event: the conditional variance w is small while the actual deviation is large. The intricacy of this bound is that the conditional variance w_ℓ is itself a random variable. However, for this work we use a weaker but more transparent version.

Corollary 3.4: *Let $\{B_k : k = 0, \dots, N\} \subset \mathbb{M}_{d \times d}$ be a matrix martingale. Assume that the associated difference sequence $C_k = B_k - B_{k-1}$ obeys $\|C_k\| \leq R$ and its conditional variance obeys $\|\sum_{k=1}^{N} \mathbb{E}_{k-1} C_k C_k^\dagger\| \leq v$ almost surely. Then*

$$\Pr[\|B_N - B_0\| \geq \tau]$$

$$\leq 2d \exp\left(\frac{-\tau^2/2}{v + R\tau/3}\right) \quad for\ any\ \tau > 0.$$

To arrive at this, ignore the events for $\ell < N$ and use that $w_\ell \leq v$ holds almost surely. This simplified matrix Freedman inequality closely resembles Bernstein's inequality. Actually, Fact 3.1 can be viewed as a special case of Corollary 3.4 where $d = 1$ and $B_k = \sum_{k'=0}^{k} X_k$. But for

matrix-valued martingales, the tail bound must depend on the dimension d.

Recapitulating the proof of the general matrix Freedman inequality would go beyond the scope of this work. It makes full use of Lieb's concavity theorem, stopping times, and Burkholder functions. There is a slightly weaker result that follows from the Golden-Thompson inequality [44, Theorem 1.2]; see also [45, Theorem 11]. Similar concentration inequalities are valid for martingales on the complex vector space $\mathbb{C}^d$. Remarkably, these results are independent of the ambient dimension d.

Fact 3.3: *Let* $\{B_k : k = 0, \dots, N\} \subset \mathbb{C}^d$ *be a vector martingale taking values in the inner product space* $\|\cdot\|_{\ell_2}$. *Assume that the associated difference sequence* $C_k = B_k - B_{k-1}$ *obeys* $\sum_{k=1}^N \mathbb{E}_{k-1}\|C_k\|_{\ell_2}^2 \le v$ *and* $\|C_k\|_{\ell_2} \le R$ *almost surely. Then, for any* $\tau > 0$

$$\Pr\left[\|B_N - B_0\|_{\ell_2} \ge \tau\right] \le 2\exp\left(\frac{-\tau^2/2}{v + R\tau/3}\right).$$

This is simplified version of Ref. [43, Theorem 3.3]. To prove the above version, start with Ref. [43, Theorem 3.1] for general martingales on Banach spaces. In our case, vectors are equipped with the standard ℓ_2 norm, set the smoothness constant $D = 1$ (in the notation of Ref. [43, Theorem 3.1]), and optimize for a λ like in Bernstein's inequality.

Our concentration analysis of random product formulas with fixed inputs relies on another tool: *uniform smoothness* of the underlying space [36]. Given a vector martingale, uniform smoothness supplies a recursive and local bound on moment growth. Decomposing $B_k = C_k + B_{k-1}$, all we need is $\mathbb{E}[C_k|B_{k-1}] = 0$, to apply the following result.

Proposition 3.1 (Uniform smoothness)**:** *Let* $x, y \in \mathbb{C}^d$ *be two random vectors that obey* $\mathbb{E}[y|x] = 0$. *When* $q \ge 2$

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

A slightly weaker version of this classic fact follows from a short argument.

Proof. Start with Bonami's inequality [46, Cor. 13.1.1] for normed vector spaces (denoting $\|\cdot\|_{\ell_2}$ as $\|\cdot\|_2$):

$$\left(\frac{\|x + y\|_2^q + \|x - y\|_2^q}{2}\right)^{2/q}$$
$$\le \frac{1}{2}\left(\|x + \sqrt{q-1}\,y\|_2^2 + \|x - \sqrt{q-1}\,y\|_2^2\right)$$
$$= \|x\|_2^2 + (q - 1)\|y\|_2^2. \tag{B5}$$

This formula can be converted to the desired statement (Proposition 3.1) via elementary manipulations. We follow

Ref. [36, Corollary 4.2]: take expectation and use triangle inequality for $L_{q/2}$ norm

$$\frac{\mathbb{E}\|x + y\|_2^q + \mathbb{E}\|x - y\|_2^q}{2}$$
$$\le \mathbb{E}(\|x\|_2^2 + (q - 1)\|y\|_2^2)^{q/2}$$
$$\le \left((\mathbb{E}\|x\|_2^q)^{2/q} + (q - 1)(\mathbb{E}\|y\|_2^q)^{2/q}\right)^{q/2}. \tag{B6}$$

Next, following Ref. [36, Proposition 4.3], by Jensen's in the first and Lyapunov's in the second inequality,

$$\frac{(\mathbb{E}\|x + y\|_2^q)^{2/q} + (\mathbb{E}\|x\|_2^q)^{2/q}}{2}$$
$$\le \frac{(\mathbb{E}\|x + y\|_2^q)^{2/q} + (\mathbb{E}\|x - y\|_2^q)^{2/q}}{2} \tag{B7}$$
$$\le \left(\frac{\mathbb{E}\|x + y\|_2^q + \mathbb{E}\|x - y\|_2^q}{2}\right)^{2/q}$$
$$\le (\mathbb{E}\|x\|_2^q)^{2/q} + (q - 1)(\mathbb{E}\|y\|_2^q)^{2/q}. \tag{B8}$$

By rearranging terms, we obtain the result with the constant $2(q - 1)$, which is off by a factor of 2. The advertised constant (Proposition 3.1) can be obtained via a more involved trick [36, Lemma A.1]. Geometrically, this result expresses the uniform smoothness of the space $(\mathbb{E}\|\cdot\|_{\ell_2}^q)^{1/q}$. ∎

We refer to Ref. [36] for further exposition on uniform smoothness for general Schatten p norm. Recently, this method has been applied to dynamic properties of random Hamiltonians [34]. For the task at hand, we can use Markov's inequality to convert such bounds on moment growth into a strong tail bound, similar to Fact 3.3. This is the content of Lemma 3.7 below.

APPENDIX C: TECHNICAL DETAILS AND PROOFS

1. Proof of Theorem 1 and Corollary 1.1: approximation error under the worst-possible input

The proofs of Theorem 1 and Corollary 1.1 were sketched in Sec. III. This section contains the details. In Sec. C1a, 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 Sec. C1b), as well as random fluctuations (in Sec. c). Finally, we combine the two bounds to obtain a convergence guarantee for randomly sampled product formulas. This is the content of Sec. C1d.

a. 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. It is, however, sometimes possible to relate diamond distances to other figures of merit that are easier to access. The mixing lemma by Campbell [33] and Hastings [32] is one very useful example.

Lemma 3.3 (Mixing lemma [32,33]): *Let U be a fixed unitary and V be a random approximation thereof that obeys $\|V - U\| \leq a$ almost surely. Then, the associated channels $\mathcal{U}(X) = UXU^\dagger$ and $\mathcal{V}(X) = VXV^\dagger$ obey*

$$\tfrac{1}{2}\|\mathbb{E}\mathcal{V} - \mathcal{U}\|_\diamond \leq \|\mathbb{E}V - U\| + \tfrac{1}{2}a^2.$$

More recent insights, in particular, Ref. [40, Theorem 3.56], allow for sharpening this bound. In particular, we can completely remove the $a^2/2$ term on the rhs.

Lemma 3.4: *Let $\mathcal{U}(\rho) = U\rho U^\dagger$ and $\mathcal{V}(\rho) = V\rho V^\dagger$ be unitary channels. Then, $\tfrac{1}{2}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}(|\psi\rangle\langle\psi|)\|_1 \leq \|(U - V)|\psi\rangle\|_{\ell_2}$ for any pure state $|\psi\rangle$. In turn, $\tfrac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond \leq \|U - V\|$. The latter relation generalizes to averages of random unitary channels:*

$$\tfrac{1}{2}\|\mathcal{U} - \mathbb{E}[\mathcal{V}]\|_\diamond \leq \|U - \mathbb{E}[V]\|.$$

The first claim follows from the fact that stabilization is not required to compute the diamond distance between two unitary channels [47, Sec. 5.3]. The second claim improves the mixing lemma.

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 $|\langle u, v\rangle| \leq 1$. Apply the Fuchs-van de Graaf relations [40, Theorem 3.33] to convert the output trace distance into a (pure) output fidelity and conclude

$$\begin{aligned}
\tfrac{1}{2}\||u\rangle\langle u| - |v\rangle\langle v|\|_1 &= \sqrt{1 - |\langle u, v\rangle|^2} \\
&= \sqrt{(1 + |\langle u, v\rangle|)(1 - |\langle u, v\rangle|)} \\
&\leq \sqrt{2[1 - \mathrm{Re}(\langle u, v\rangle)]} = \||u\rangle - |v\rangle\|_{\ell_2}.
\end{aligned}$$

The first 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 [47, Sec. 5.3] to conclude

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

Here, we also use 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

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

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

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

for any pure input state $|\psi\rangle$. This is enough to conclude $\tfrac{1}{2}\|U|\psi\rangle\langle\psi|U^\dagger - \mathbb{E}[V|\psi\rangle\langle\psi|V^\dagger]\|_1 \leq \|(U - \mathbb{E}[V])|\psi\rangle\|_{\ell_2}$, 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:

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

This is what we need to show.   ∎

b. Controlling the deterministic bias

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

Proposition 3.2: *Consider the IID unitary product constructed by the* QDRIFT *protocol, Eq. (4), for simulating* $U = \exp(-itH)$ *for evolution time t. Define the total strength* $\lambda = \sum_j \|h_j\|$. *Then*

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

Note that the improved mixing lemma, Lemma 3.4 above, allows for converting this statement into a diamond distance bound for the associated channels:

$$\tfrac{1}{2}\|\mathcal{U} - \mathbb{E}[\mathcal{V}_N \circ \cdots \circ \mathcal{V}_1]\|_\diamond \leq \|U - \mathbb{E}[V_N \cdots V_1]\| \leq \frac{t^2\lambda^2}{N}. \tag{C1}$$

This is a slight improvement over the main existing technical result regarding QDRIFT [17]. 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 $\tfrac{1}{2}\|\mathcal{U} - \mathcal{E}\|_\diamond \leq (t^2\lambda^2/N)e^{2t\lambda/N}$ in Ref. [17, Eq. (B12)]. Both assertions become very similar in the large N limit, but Eq. (C1) is always tighter and the discrepancy can be quite pronounced for small and intermediate values of N.

The proof of Proposition 3.2 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 3.4: *Let X be Hermitian. Then we have the zero-order bound* $\|\exp(iX) - \mathbb{I}\| \leq \|X\|$ *and the first-order bound* $\|\exp(iX) - iX - \mathbb{I}\| \leq \tfrac{1}{2}\|X\|^2$.

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

Lemma 3.5: *Fix a Hamiltonian* $H = \sum_{j=1}^L h_j$ *and parameters* N, t. *Set* $U^{1/N} = \exp[-i(t/N)H]$ *and* $\lambda = \sum_{j=1}^L \|h_j\|$. *Then, the random matrix V defined in Eq. (4) obeys*

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

Proof. Streamline the notation from QDRIFT(Algorithm 0.1) 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 3.4 and uniform normalization 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 3.4, and uniform normalization delivers

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

This is the advertised result. ∎

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 submultiplicativity and the triangle inequality. See Ref. [48] for related results.

Lemma 3.6: *Let* $\mathbb{E}V$ *and* $U^{1/N}$ *be matrices with bounded operator norm, i.e.,* $\|\mathbb{E}V\| \leq 1$ *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 submultiplicativity 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:

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

This is the stated result. ∎

Proof of Proposition 3.2. The main result of this section immediately follows from combining Lemmas 3.6 and 3.5. 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. ∎

c. 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:

$$V_N \cdots V_1 \approx \mathbb{E}[V_N \cdots V_1]$$

$$= (\mathbb{E}V)^N \quad \text{with high probability.} \quad \text{(C2)}$$

In words, we need to show that a product of IID 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 IID random matrices have been available for more than a decade now [49,50], our understanding of concentration for random matrix products is more limited; see Ref. [36] 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 Ref. [51] and references therein. The small-step regime has seen less development, although there are some asymptotic bounds [52].

Fortunately, the QDRIFT construction has several appealing features: the random unitaries $V_N, \ldots, V_1$ are IID 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 3.3 (QDRIFT: operator-norm concentration)**:** *Consider a Hamiltonian $H = \sum_{j=1}^{L} h_j$ with interaction strength $\lambda = \sum_{j=1}^{L} \|h_j\|$, and fix parameters N, t. Suppose that $V_N, \ldots, V_1$ are IID instances of the random unitary $d \times d$ matrix V constructed by the QDRIFT protocol,*

Eq. (4). Then

$$\Pr\left[\|V_N \cdots V_1 - \mathbb{E}(V_N \cdots V_1)\| \geq \epsilon/2\right]$$

$$\leq 2d \exp\left(-\frac{N\epsilon^2}{44t^2\lambda^2}\right) \quad \text{for any } \epsilon \in [0, 4t\lambda].$$

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: $\Pr\left[\|V_n \cdots V_1 - \mathbb{E}(V_N \cdots V_1)\| \geq \tau\right] \leq 2d \exp(-N\epsilon/6t\lambda)$. We refer to Eq. (C4) 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. It heavily utilizes the martingale concentration tools introduced in Appendix B. For fixed N, we interpolate between both sides of Eq. (C2) by means of a random process $\{B_k : k = 0, \ldots, N\}$:

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

$$B_0 = (\mathbb{E}V)^N \text{ and } B_N = V_N \cdots V_1.$$

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 $V_{k-1}, \ldots, V_1$. Fortunately, we can recognize it as a (matrix-valued) martingale satisfying the defining properties:

1. *Causality:* each B_k is completely determined by the information we have collected up to step k. That is, the (random) choices of $V_k, \ldots, V_1$.

2. *Status quo:* conditioned on previous choices, the expectation of B_{k+1} equals B_k: for $1 \leq k \leq N$

$$\mathbb{E}\left[B_{k+1}|V_k \cdots V_1\right]$$

$$= (\mathbb{E}V)^{N-(k+1)}\mathbb{E}_{V_{k+1}}\left[V_{k+1}\right]V_k \cdots V_1$$

$$= (\mathbb{E}V)^{N-k}V_k \cdots V_1 = B_k. \quad \text{(C3)}$$

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

With this matrix martingale reformulation at hand, we can prove Proposition 3.3 using a concentration inequality for matrix martingales, namely Corollary 3.4.

Proof of Proposition 3.3. We have already established that the random process $B_k = (\mathbb{E}V)^{N-k}V_k \cdots V_1$ forms a matrix

martingale that interpolates between $B_0 = \mathbb{E}[V_N \cdots V_1] = (\mathbb{E}V)^N$ and $V_N = V_N \cdots V_1$. The elements of the associated difference sequence are

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

Recall that $V_k = \exp(-iX_j)$ for some Hermitian matrices $X_j = t/N(\lambda/\|h_j\|)h_j$ with index $1 \le j \le L$. Boundedness ($\|\mathbb{E}V\|, \|V_k\| \le 1$), Fact 3.4 [$\|\exp(-iX) - \mathbb{I}\| \le \|X\|$ for X Hermitian], and Lemma 3.5 ensure

$$\|C_k\| \le \|\mathbb{E}V\|^{N-k}\|V_k - (\mathbb{E}V_k)\|\|V_{k-1} \cdots V_1\|$$
$$\le \|V_k - (\mathbb{E}V_k)\| = \frac{2t\lambda}{N}$$

almost surely. Set $R = 2t\lambda/N$, and invoke Corollary 3.4 to conclude that

$$\Pr[\|B_N - B_0\| \ge \tau] \le 2d \exp\left(-\frac{N\tau^2}{8(t\lambda)^2 + 4(t\lambda)\tau/3}\right). \tag{C4}$$

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

In fact, the same proof works for any adapted small-step random walks on the unitary group. Such a generalization results in Theorem 3 and we refer to Appendix 3 for details.

d. A bound for expected errors

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 3.4 (QDRIFT: expected diamond-norm error): *Consider an n-qubit Hamiltonian $H = \sum_{j=1}^{L} h_j$ with total strength $\lambda = \sum_{j=1}^{L} \|h_j\|$. Fix parameters $N, t,$ and assume that $N \ge n$. Set $\mathcal{U} = UXU^\dagger$ with $U = \exp(-itH),$ and suppose that $\mathcal{V}_N, \ldots, \mathcal{V}_1 \sim \mathcal{V}$ are IID realizations of the QDRIFT protocol. That is, $\mathcal{V}(X) = VXV^\dagger,$ where V is defined by Eq. (4). Then*

$$\mathbb{E}\left[\tfrac{1}{2}\|\mathcal{U} - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1\|_\diamond\right]$$
$$\le \frac{t^2\lambda^2}{N} + C\frac{nt\lambda}{N} + C\sqrt{\frac{nt^2\lambda^2}{N}} \approx C\sqrt{\frac{nt^2\lambda^2}{N}}, \tag{C5}$$

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 [17] and the improvement in Eq. (C1):

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

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, Eq. (C5) 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 $N \ge Cnt^2\lambda^2/\epsilon^2$ is sufficient to ensure ϵ closeness on average. The quadratic scaling in the accuracy parameter ϵ 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

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

we need the random product formula to behave for all possible n-qubit input states ρ 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 3.4. First, we relate the expected diamond distance to an expected operator-norm distance and split it up into deterministic bias and (expected) fluctuations:

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

The first term is deterministic and controlled by Proposition 3.2: $\|U - (\mathbb{E}V)^N\| \le t^2\lambda^2/N$. The second term can be bounded by integrating the tail bound in Proposition 3.3, or rather the tighter bound presented, Eq. (C4); see Ref. [50, Remark 6.5]. For n qubits, we have $d = 2^n$

and conclude

$$\mathbb{E}\|V_N \cdots V_1 - (\mathbb{E}V)^N\|$$

$$= \int_0^\infty \Pr\left[\|V_N \cdots V_1 - (\mathbb{E}V)^N\| \geq \tau\right] d\tau$$

$$\leq \int_0^\infty \min\left[1, 2 \times 2^n \exp\left(-\frac{N\tau^2/2}{4\lambda^2 t^2 + 2\lambda t\tau/3}\right)\right] d\tau$$

$$\leq C\max\left\{\sqrt{\frac{nt^2\lambda^2}{N}}, \frac{nt\lambda}{N}\right\} \leq 2C\left(\frac{nt\lambda}{N} + \sqrt{\frac{nt^2\lambda^2}{N}}\right).$$

Here, we have evaluated the integral by two segments: the integrand is at most 1 until $\tau \sim \max(\sqrt{nt^2\lambda^2/N}, nt\lambda/N)$; for larger τ, the expression decays exponentially and integrating it produces only a contribution of order $\mathcal{O}[\max(\sqrt{t^2\lambda^2/N}, t\lambda/N)]$. Also, $2C$ absorbs all constants. ∎

2. Proof of Theorem 2: approximation error under a single input

Proposition 3.4 asserts that a single, random realization of the QDRIFT protocol, Eq. (4), accurately approximates a unitary target evolution with respect to the diamond norm:

$$\mathbb{E}\left[\tfrac{1}{2}\|\mathcal{U} - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1\|_\diamond\right]$$

$$= \mathbb{E}\left[\tfrac{1}{2}\max_{\rho \text{ state}} \|\mathcal{U}(\rho) - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_N(\rho)\|_1\right]$$

$$\lesssim C\sqrt{\frac{nt^2\lambda^2}{N}}.$$

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 ρ.

In Hamiltonian simulation, demanding such a stringent worst-case promise may be excessive. In most practical applications, the input state ρ 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

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

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

The proof is similar in spirit to the argument behind Proposition 3.4. We construct a vector 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 [36] on concentration of random matrix products.

a. 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 3.5 (QDRIFT: action on a fixed state): *Consider a Hamiltonian $H = \sum_{j=1}^L h_j$ with total strength $\lambda = \sum_{j=1}^L \|h_j\|$. Fix evolution time t and a gate count N that obeys $N \geq (t\lambda)^2$. Let $\mathcal{V}_1, \ldots, \mathcal{V}_N$ be the IID random unitary evolution operators constructed by the QDRIFT protocol, Eq. (4). Then,*

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

Moreover, for $\epsilon > 0$,

$$\max_{\rho \text{ state}} \Pr\left[\tfrac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \cdots \mathcal{V}_1(\rho)\|_1 > \epsilon\right]$$

$$\leq \exp\left(\frac{-\epsilon^2 N}{32et^2\lambda^2}\right).$$

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

$$\Pr\left[\tfrac{1}{2}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \cdots \mathcal{V}_1(\rho)\|_1 > \epsilon\right]$$

$$\leq \epsilon^{-q}\mathbb{E}\left[2^{-q}\|\mathcal{U}(\rho) - \mathcal{V}_N \circ \cdots \mathcal{V}_1(\rho)\|_1^q\right]. \quad \text{(C6)}$$

The right-hand side of this equation is a convex function of the state ρ. Thus, the maximum 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 unitary channels $\mathcal{V}_k(X) = VXV^\dagger$.

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 3.4 implies

$$\{\mathbb{E}[2^{-q}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1(|\psi\rangle\langle\psi|)\|_1^q]\}^{1/q}$$

$$\leq \left[\mathbb{E}\|(V_N \cdots V_1 - U)|\psi\rangle\|_{\ell_2}^q\right]^{1/q}$$

$$\leq 2 \max \left\{ \left\|\underbrace{(\mathbb{E}[V_N \cdots V_1] - U)|\psi\rangle}_{\text{deterministic bias}|\psi_{\text{bias}}\rangle}\right\|_{\ell_2}, \right.$$

$$\left. \{\mathbb{E}\left\|\underbrace{(V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1])|\psi\rangle}_{\text{random fluctuation }|\psi_{\text{rand}}\rangle}\right\|_{\ell_2}^q\}^{1/q}\right\}. \quad \text{(C7)}$$

The last bound follows from the triangle inequality and $(a + b)^q \leq 2^q \max\{a^q, b^q\}$ 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 3.2:

$$\|(\mathbb{E}[V_N \cdots V_1] - U)|\psi\rangle\|_{\ell_2}$$

$$= \|[(\mathbb{E}V)^N - U]|\psi\rangle\|_{\ell_2} \leq \|(\mathbb{E}V)^N - U\| \leq \frac{(t\lambda)^2}{N}. \quad \text{(C8)}$$

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

Lemma 3.7: *Let $V_N, \ldots, V_1$ be IID unitaries that implement the* QDRIFT *protocol, Eq. (4), with parameters t and λ. 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 establish this lemma below. The basic idea behind the proof is to express the random vector using a martingale sequence, similar to the matrix case. We could have already call vector martingale tail bounds (Fact 3.3) to arrive at the desired results. However, we demonstrate the same results via Markov's inequality and moment bounds (uniform smoothness, Proposition 3.1), which we introduce in Appendix B.

Introduce the inequalities from Eq. (C8) and Lemma 3.7 into the bound, Eq. (C7). We obtain

$$\{\mathbb{E}[2^{-q}\|\mathcal{U}(|\psi\rangle\langle\psi|) - \mathcal{V}_N \circ \cdots \circ \mathcal{V}_1(|\psi\rangle\langle\psi|)\|_1^q]\}^{1/q}$$

$$\leq 4\sqrt{\frac{(q-1)(t\lambda)^2}{N}}. \quad \text{(C9)}$$

We use 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 Eq. (C9) and apply Lyapunov's inequality. To obtain the probability bound, we combine Eqs. (C6) and (C9) to arrive at

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

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

b. Proof of Lemma 3.7

In this section, we establish the bound on the size of the fluctuations.

Proof of Lemma 3.7. 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 uniform smoothness (Proposition 3.1) to split up the contributions:

$$(\mathbb{E}\||\psi_N\rangle - \mathbb{E}[|\psi_N\rangle]\|_{\ell_2}^q)^{2/q}$$

$$\leq (q-1)\{\mathbb{E}\|[V_N - (\mathbb{E}V_N)]|\psi_{N-1}\rangle\|_{\ell_2}^q\}^{2/q}$$

$$+ \{\mathbb{E}\|(\mathbb{E}V)(|\psi_{N-1}\rangle - \mathbb{E}[|\psi_{N-1}\rangle])\|_{\ell_2}^q\}^{2/q}$$

$$\leq (q-1)(\mathbb{E}\|V_N - \mathbb{E}[V_N]\|^q)^{2/q} + (\mathbb{E}\||\psi_{N-1}\rangle$$

$$- \mathbb{E}[|\psi_{N-1}\rangle]\|_{\ell_2}^q)^{2/q}.$$

We can now iterate this argument to conclude that

$$(\mathbb{E}\||\psi_N\rangle - \mathbb{E}[|\psi_N\rangle]\|_{\ell_2}^q)^{2/q}$$

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

$$= (q-1)N(\mathbb{E}\|V - \mathbb{E}[V]\|^q)^{2/q}.$$

Invoke Lemma 3.5, using the properties of the random unitaries constructed by QDRIFT:

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

Combine the last two displays to reach the stated result. ∎

3. Proof of Theorem 3 and Corollary 3.1

The proof techniques for establishing Theorems 1 and 2 are remarkably general and we condense them into Theorem 3. Let us reiterate the premise. Consider approximating a target unitary product $U = U_N \cdots U_1$ by a random unitary $V = V_N \cdots V_1$ such that $\{V_k\}$ satisfy:
(i) *Causality:* for $1 \leq k \leq N$ the random selection of V_k can depend only on previous choices for $V_1, \ldots, V_{k-1}$:

$$\Pr[V_k | V_N, \ldots, V_{k+1}, V_{k-1}, \ldots, V_1] = \Pr[V_k | V_{k-1}, \ldots, V_1].$$

(ii) *Accurate approximation:* each realization of V_k (and their conditional expectation) must be close to the ideal unitary U_k. More precisely, let $R, b_k > 0$ be constants such that

$$\|V_k - \mathbb{E}_{k-1} V_k\| \leq R \quad \text{and} \quad \|\mathbb{E}_{k-1} V_k - U_k\| \leq b_k, \text{ where}$$
$$\mathbb{E}_{k-1} V_k := \mathbb{E}[V_k | V_{k-1}, \ldots, V_1],$$

almost surely for all $1 \leq k \leq N$. Note this is more general then we need to prove Theorem 3; this is to take into account the cases when the conditional variance may be much smaller than NR^2.

Proof of Theorem 3. Recall the decomposition of approximation error into a deterministic bias and a random fluctuation:

$$\|V_N \cdots V_1 - U_N \cdots U_1\|$$
$$\leq \underbrace{\|(\mathbb{E}V)^N - U_N \cdots U_1\|}_{\text{deterministic bias}}$$
$$+ \underbrace{\|V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1]\|}_{\text{random fluctuation}}.$$

The deterministic bias can once again be controlled by a telescoping sum,

$$\|(\mathbb{E}V)^N - U\| \leq \sum_{k=1}^{N} b_k. \tag{C10}$$

Note that this also controls the performance of channel $\|\mathbb{E}\mathcal{V} - \mathcal{U}\|_\diamond$ by Lemma 3.4. For the *random fluctuations*, tweaking the proofs of Theorems 1 and 2 implies the following general result. ∎

Proposition 3.6 (adapted random walk on the unitary group; with and without fixed input state): *Let* $\{V_1, V_2, \ldots, V_N\} \subset U(d)$ *an adapted (causal) random*

unitary matrices that obey

$$\sum_{k=1}^{N} \mathbb{E}_{k-1} \|(V_k - \mathbb{E}V_k)(V_k^\dagger - \mathbb{E}V_k^\dagger)\| \leq \sigma^2 \quad \text{and}$$

$$\|V_k - \mathbb{E}V_k\| \leq R \text{ (almost surely)} \tag{C11}$$

for some constants $v, R \geq 0$. *Then, the product of N random unitaries satisfies a concentration inequality:*

$$\Pr(\|V_N \cdots V_1 - \mathbb{E}[V_N \cdots V_1]\| \geq \epsilon)$$
$$\leq 2d \exp\left(\frac{-\epsilon^2/2}{v + R\epsilon/3}\right) \quad \text{for } \epsilon > 0. \tag{C12}$$

For a fixed, but arbitrary, input state ρ, concentration is independent of the ambient dimension d:

$$\max_{\rho \text{ state}} \Pr\{\tfrac{1}{2}\|\mathbb{E}[\mathcal{V}_N \circ \cdots \mathcal{V}_1(\rho)] - \mathcal{V}_N \circ \cdots \mathcal{V}_1(\rho)\|_1 > \epsilon\}$$
$$\leq 2 \exp\left(\frac{-\epsilon^2/2}{v + R\epsilon/3}\right) \quad \text{for } \epsilon > 0.$$

For a fixed input state, we would call the vector Freedman inequality (Fact 3.3) instead of uniform smoothness (Proposition 3.1) in Theorem 2. 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 [36] addresses the problem using uniform smoothness tools. The paper [53] 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 Refs. [36,41,42,54].

It is instructive to illustrate these improvements by example. In QDRIFT, all steps have a uniform bound R, but in the fully general statement the variance v can differ from the crude uniform bound NR^2. In such a regime, the subexponential tail of size $e^{-3\epsilon/2R}$ can start playing a role.

Lastly, for illustration in Theorem 3 we give loose estimates on the variance to avoid complication with the heavy tail effects. Plugging in the parameters $v = ra^2, R = 2a$ as $\|V - \mathbb{E}V\| \leq \mathbb{E}\|V - V'\| \leq 2\|U - V\| = 2a$ translates to a typical fluctuation $\epsilon^2 \sim nra^2$ (and $\epsilon^2 \sim ra^2$ for fixed input). We conclude the proof by combining the bound for the deterministic bias and random fluctuation.

Proof of Corollary 3.1. Consider randomly permuting the $2k$th-order Suzuki formulas as $V_k = S_{2k}^\sigma(t/r)$ with uniform probability $1/L!$ Then by direct calculation [18,48]:

$$\|V - U\| \leq O\left(\frac{(t\Lambda L)^{2k+1}}{r^{2k+1}}\right) = a, \tag{C13}$$

$$\|\mathbb{E}V - U\| \leq O(t\Lambda/r)^{2k+1} L^{2k} = b \tag{C14}$$

by Theorem 3,

$$\epsilon_{\text{det}} \leq ra = O\left(\frac{(t\Lambda L)^{2k+1}}{r^{2k}}\right)$$

$$\epsilon_{\text{typ}} = O(\sqrt{n}ra)$$
$$\qquad + 2rb = O\left[\sqrt{n}\frac{(t\Lambda L)^{2k+1}}{r^{2k+1/2}} + (t\Lambda)^{2k+1}\left(\frac{L}{r}\right)^{2k}\right]$$

$$\epsilon_{\text{fix}} = O(\sqrt{r}a)$$
$$\qquad + 2rb = O\left[\frac{(t\Lambda L)^{2k+1}}{r^{2k+1/2}} + (t\Lambda)^{2k+1}\left(\frac{L}{r}\right)^{2k}\right]$$

$$\epsilon_{\text{avg}} \leq 2rb = O\left[(t\Lambda)^{2k+1}\left(\frac{L}{r}\right)^{2k}\right],$$

which translates to the sufficient gate counts $N = rL$

$$N_{\text{det}} \gtrsim t\Lambda L^2 \left(\frac{t\Lambda L}{\epsilon}\right)^{1/2k},$$

$$N_{\text{typ}} \gtrsim t\Lambda L^2 \left(\frac{n\Lambda Lt}{\epsilon^2}\right)^{1/(4k+1)} + t\Lambda L^2 \left(\frac{t\Lambda}{\epsilon}\right)^{1/2k},$$

$$N_{\text{fix}} \gtrsim t\Lambda L^2 \left(\frac{\Lambda Lt}{\epsilon^2}\right)^{1/(4k+1)} + t\Lambda L^2 \left(\frac{t\Lambda}{\epsilon}\right)^{1/2k},$$

$$N_{\text{avg}} \gtrsim t\Lambda L^2 \left(\frac{t\Lambda}{\epsilon}\right)^{1/2k}. \qquad \blacksquare$$

APPENDIX D: ASYMPTOTIC TIGHTNESS

It is natural to wonder whether the bound (C5) is tight for some Hamiltonian $H = \sum_j h_j$, 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 $\lambda = \sum_j \|h_j\|$ 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 [55, Section 4.1.2]. 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 Zs or the sum of all 2^n many-body Pauli Zs. We 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 Zs, it comes from the fact that diagonal entries are nearly independent, so the union bound we used in Sec. 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 3.5 (CLT for the multinomial distribution)**:** *The multivariate distribution* $\mathbf{m} = (m_1, \ldots, m_K) \sim \text{Mult}[N, (1/K, \ldots, 1/K)]$ *(roll a fair K-sided dice N times) obeys a central limit theorem (CLT):*

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

The covariance matrix is $\Sigma = 1/K \left[\mathbb{I} - (1/K)J\right]$*, where J denotes the $K \times K$ matrix of ones.*

1. 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 Sec. III A:

$$H = \sum_{k=1}^n Z_k \quad \text{where} \quad Z_k = \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{(k-1) \text{ times}} \otimes Z$$
$$\otimes \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{(n-k) \text{ times}} \quad \text{for } 1 \leq k \leq n. \qquad \text{(D1)}$$

Proposition 3.7: *Suppose that we wish to obtain an N-term approximation of the time evolution $U = \exp(-itH)$ associated with the n-qubit Hamiltonian (D1) for evolution time t. In the large N limit (CLT), the QDRIFT approximation, Eq. (4), incurs an operator-norm error that matches the (upper) bound from Corollary 1.1 (and indirectly 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 choose to state this result directly in terms of operator-norm deviation. A conversion into diamond distance is also possible: $\frac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond \geq \frac{1}{2}\|U - V\|$ for any pair of unitary channels. This conversion rule readily follows from the geometric characterization of $\frac{1}{2}\|\mathcal{U} - \mathcal{V}\|_\diamond$ provided in Ref. [47].

Proof of Proposition 3.7. Each of the n terms in the Hamiltonian (D1) has unit operator norm ($\|Z_k\| = 1$) and the total strength is $\lambda = \sum_{k=1}^{n} \|Z_k\| = n$. For fixed N and t, each short-time approximation, Eq. (4), has the form $V_k = \exp\left[-i(t n/N)Z_{k(i)}\right]$, where each $k(i)$ is an index chosen uniformly from the set $\{1, \ldots, n\}$ (multinomial distribution). Since all Z_ks 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 introduce 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 $1 \le k \le n$. 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

$$\|V_N \cdots V_1 - U\| = \left\|\exp\left(-i\frac{t\lambda}{\sqrt{N}}\sum_{k=1}^{N}\frac{m_k - \bar{m}_k}{\sqrt{N}}Z_k\right) - \mathbb{I}\right\|.$$
(D2)

This is a promising starting point. The multinomial CLT (Fact 3.5) ensures that the n centered and normalized random variables $s_k = (m_k - \bar{m}_k)/\sqrt{N}$ approach the coefficients of a Gaussian vector $s \in \mathbb{R}^n$ with covariance matrix $\Sigma = 1/n\left(\mathbb{I} - \frac{1}{n}J\right)$. This, in particular, implies $\mathbb{E}s_k = 0$ and $\mathbb{E}s_k^2 = 1/n(1 - 1/n) = \sigma^2$ for all $1 \le k \le n$. We can capitalize on this observation by simplifying Eq. (D2) via a second-order Taylor expansion. Set $X = -t\lambda/\sqrt{N}\sum_k s_k Z_k$ for brevity and apply Fact 3.4 to obtain

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

This relation is preserved under expectations and we obtain

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

Let us focus on the leading-order term first. The particular structure of the Hamiltonian (D1)—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

$$\mathbb{E}\left\|\sum_k s_k Z_k\right\| = \mathbb{E}\sum_{k=1}^{n}\|s_k Z_k\|$$

$$= \sum_{k=1}^{n}|s_k| \overset{N\to\infty}{=} n\sqrt{\frac{2}{\pi}\frac{1}{n}\left(1 - \frac{1}{n}\right)}$$

$$= \sqrt{\frac{2}{\pi}(n-1)},$$

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

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

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

$$= n\sum_{k=1}^{n}\mathbb{E}s_k^2 = n^2\sigma^2 = (n-1).$$

No CLT is required for this argument. Inserting both bounds into Eq. (D2) completes the argument. ∎

2. Sum of many-body Pauli-Z operators

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

$$H = \sum_{\mathbf{p}\in\{0,1\}^n} Z_{\mathbf{p}} = \sum_{\mathbf{p}\in\{0,1\}^n} Z^{p_1} \otimes \cdots \otimes Z^{p_n}, \quad \text{(D3)}$$

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 total strength $\lambda = \sum_{\mathbf{p}}\|Z_{\mathbf{p}}\| = 2^n$ equals the Hilbert-space dimension. It is also worthwhile to point out that each term is diagonal in the computational basis $|\mathbf{b}\rangle = |b_1, \ldots, b_n\rangle$ with $\mathbf{b} = (b_1, \ldots, b_n) \in \{0, 1\}^n$. Overlaps of the Hamiltonian terms with computational basis states are given by

$$\langle\mathbf{b}|Z_{\mathbf{p}}|\mathbf{b}\rangle = (-1)^{\langle\mathbf{b},\mathbf{p}\rangle} = (-1)^{\sum_i b_i p_i} \in \{\pm 1\}. \quad \text{(D4)}$$

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 3.8: *Suppose that we wish to obtain an N-term approximation of the time evolution $U = \exp(-itH)$ associated with the n-qubit Hamiltonian (D3) for evolution time t. In the large N limit (CLT) the QDRIFT approximation, Eq. (4), incurs an operator-norm error that matches the (upper) bound from Corollary 1.1 (and indirectly Theorem 1) up to a constant factor:*

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

The conversion rule $\|\mathcal{U} - \mathcal{V}\|_\diamond \geq \|U - V\|$ (for unitary channels) [47] once more allows for addressing the expected diamond distance as well.

Proof of Proposition 3.8. Each of the 2^n terms in the Hamiltonian (D3) 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, Eq. (4), has the form $V_k = \exp[-i(t\lambda/N)Z_{\mathbf{p}(i)}]$, 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 3.7:

$$\mathbb{E}\|V_N \cdots V_1 - U\|$$
$$\geq \frac{t\lambda}{\sqrt{N}}\mathbb{E}\left\|\sum_{\mathbf{p}} s_{\mathbf{p}} Z_{\mathbf{p}}\right\| - \frac{1}{2}\left(\frac{t\lambda}{\sqrt{N}}\right)^2 \mathbb{E}\left\|\sum_{\mathbf{p}} s_{\mathbf{p}} Z_{\mathbf{p}}\right\|^2.$$
$$(D5)$$

Here, $s_{\mathbf{p}} = (m_{\mathbf{p}} - \bar{m}_{\mathbf{p}})/\sqrt{N}$ 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 3.5) asserts that the 2^n centered and normalized random variables $s_{\mathbf{p}}$ approach distinct coefficients of a 2^n-dimensional Gaussian vector with covariance matrix $\Sigma = 1/2^n\left(\mathbb{I} - \frac{1}{2^n}J\right) = \frac{1}{2^n}(\mathbb{I} - |\mathbf{1}\rangle\langle\mathbf{1}|)$, where $|\mathbf{1}\rangle = 1/2^n \sum_{\mathbf{b}\in\{0,1\}^n} |\mathbf{b}\rangle$ (the normalized all-ones vector in $\mathbb{R}^{2^n}$). In contrast to before, the individual contributions to this operator norm do not 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 $\mathbf{b} = (b_1, \ldots, b_n) \in \{0,1\}^n$. 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 Eq. (D4). 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,

$$\max_{\mathbf{b}\in\{0,1\}^n}\left|\sum_{\mathbf{p}}(-1)^{\langle\mathbf{b},\mathbf{p}\rangle} s_{\mathbf{p}}\right| = 2^{n/2}\left\|\mathrm{Had}^{\otimes n} s\right\|_{\ell_\infty}$$

$$=: 2^{n/2}\|\hat{s}\|_{\ell_\infty} \quad \text{where} \quad \mathrm{Had} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

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

$$\hat{\Sigma} = \frac{1}{2^n}\mathrm{Had}^{\otimes n}\left(\mathbb{I} - |\mathbf{1}\rangle\langle\mathbf{1}|\right)\mathrm{Had}^{\otimes n}$$

$$= \frac{1}{2^n}\left(\mathbb{I} - |0, \ldots, 0\rangle\langle0, \ldots, 0|\right).$$

Hence, the CLT asserts that the transformed vector $\hat{s}$ approaches a standard Gaussian vector with $2^n - 1$ degrees of freedom: $\hat{s} = (0, g_2, \ldots, g_{2^n})^T$ with $g_i \overset{\mathrm{IID}}{\sim} \mathcal{N}(0, 2^{-n})$ (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 σ^2 is lower bounded by $0.265\sqrt{\log(K)\sigma^2}$, see, e.g., Ref. [56, Proposition 8.1]:

$$\frac{t\lambda}{\sqrt{N}}\mathbb{E}\left\|\sum_{\mathbf{p}} s_{\mathbf{p}} Z_{\mathbf{p}}\right\|$$

$$= \frac{t\lambda}{\sqrt{N}}2^{n/2}\mathbb{E}\|\hat{s}\|_{\ell_\infty} \overset{N\to\infty}{=} \frac{t\lambda}{\sqrt{N}}2^{n/2}\mathbb{E}\max_{2\leq i\leq 2^n}|g_i|$$

$$\geq 0.625\frac{t\lambda}{\sqrt{N}}2^{n/2}\sqrt{\log(2^n - 1)2^{-n}} \geq \frac{1}{2}\sqrt{n\frac{(t\lambda)^2}{N}}.$$

Here, we use 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 in Eq. (D5).

Moving on to the quadratic term in Eq. (D5), we employ a similar strategy. Observe

$$
\mathbb{E} \left\| \sum_{\mathbf{p}} Z_{\mathbf{p}} \right\|^2 = \mathbb{E} \left\| \sum_{\mathbf{p},\mathbf{p}'} s_{\mathbf{p}} s_{\mathbf{p}'} Z_{\mathbf{p}} Z_{\mathbf{p}'} \right\|
$$

$$
= \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} \left| \sum_{\mathbf{p},\mathbf{p}'} s_{\mathbf{p}} s_{\mathbf{p}'} \langle \mathbf{b} | Z_{\mathbf{p}} Z_{\mathbf{p}'} | \mathbf{b} \rangle \right|
$$

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

where the last equation follows from combining Eq. (D4) 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:

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

$$
= 2^n \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} \left| \hat{s}_{\mathbf{b}} \right|^2.
$$

We already know from the CLT that the 2^n-dimensional Walsh-Hadamard transform of s approaches a standard Gaussian vector: $\hat{s} = (0, g_2, \ldots, g_{2^n})^T$ with $g_i \overset{\text{IID}}{\sim} \mathcal{N}(0, 2^{-n})$. In the large N limit (CLT), the rhs of the above display becomes an expected maximum of $K = 2^n - 1$ squares of IID Gaussian variables with mean zero and variance $\sigma^2 = 2^{-n}$. Such expected maxima can be bounded using standard arguments, see, e.g., Ref. [57, Lemma 5.1]: $\mathbb{E} \max_{1 \le i \le K} |g_i|^2 \le 4\sigma^2 \log(\sqrt{2}K)$ (the constants are chosen based on simplicity, not tightness). This allows us to conclude

$$
\mathbb{E} \left\| \sum_{\mathbf{p}} Z_{\mathbf{p}} \right\|^2 = 2^n \mathbb{E} \max_{\mathbf{b} \in \{0,1\}^n} \left| \hat{s}_{\mathbf{b}} \right|^2 \overset{N \to \infty}{=} 2^n \mathbb{E} \max_{2 \le i \le N} |g_i|^2
$$

$$
\le 2^n 4\sigma^2 \log[\sqrt{2}(2^n - 1)] \le 4(n + 1/2).
$$

Inserting linear and quadratic bound into Eq. (D5) completes the argument. ∎

[1] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, Low-Depth Quantum Simulation of Materials, Phys. Rev. X **8**, 011044 (2018).

[2] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Rev. Mod. Phys. **86**, 153 (2014).

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

[4] D. An and L. Lin, Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm. *preprint* arXiv:1909.05500, 2019.

[5] H.-Y. Huang, K. Bharti, and P. Rebentrost, Near-term quantum algorithms for linear systems of equations. *preprint* arXiv:1909.07344, 2019.

[6] Y. Subaşi, R. D. Somma, and D. Orsucci, Quantum Algorithms for Systems of Linear Equations Inspired by Adiabatic Quantum Computing, Phys. Rev. Lett. **122**, 060504 (2019).

[7] For measuring time-evolved observables tr$[\rho(t)O]$, the operator norm suffices; for quantum phase estimation of ground-state energies things can get more complicated, though. There, the estimates may depend on the details and vary orders of magnitude [58] and the assessment of real costs is an on-going research direction that is beyond the scope of this work.

[8] In this work, we count $e^{-ih_j t}$ as a single gate, as in Refs. [16,18,19]. Strictly speaking, in a fault-tolerant quantum computer, we would further decompose each $e^{-ih_j t}$ into universal 2-qubit gates. According to the Solovay-Kitaev theorem [37] this incurs at most a constant (multiplicative) overhead, but the exact cost depends on the type of hardware. Note we do not discuss the query complexity that appears in LCU approaches [14].

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

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

[11] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, Quantum Inf. Comput. **12**, 901 (2012).

[12] G. H. Low and I. L. Chuang, Optimal Hamiltonian Simulation by Quantum Signal Processing, Phys. Rev. Lett. **118**, 010501 (2017).

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

[14] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, A grand unification of quantum algorithms. *preprint* arXiv:arXiv:2105.02859, 2021.

[15] D. W. Berry, A. M. Childs, and R. Kothari, in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015*, p. 792. IEEE Computer Soc., Los Alamitos, CA, 2015.

[16] 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).

[17] E. Campbell, Random Compiler for Fast Hamiltonian Simulation, Phys. Rev. Lett. **123**, 070503 (2019).

[18] A. M. Childs, A. Ostrander, and Y. Su, Faster quantum simulation by randomization, Quantum **3**, 182 (2019).

[19] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter Error with Commutator Scaling, Phys. Rev. X **11**, 011020 (2021).

[20] B. Sahinoglu and R. D. Somma, Hamiltonian simulation in the low energy subspace. *preprint* arXiv:2006.02660, 2020.

[21] J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, Quantum algorithm for simulating real time evolution of lattice hamiltonians. *SIAM J. on Comput.*, p. FOCS18–250–FOCS18–284 2021.

[22] To finally compile into universal gates, run a standard gate synthesis algorithm (such as Solovay-Kiteav) for each simple term. This yields another multiplicative factor on the gate count.

[23] P. K. Faehrmann, M. Steudtner, R. Kueng, M. Kieferova, and J. Eisert, Randomizing multi-product formulas for improved Hamiltonian simulation. *preprint* arXiv:2101.07808, 2021.

[24] Y. Ouyang, D. R. White, and E. T. Campbell, Compilation by stochastic hamiltonian sparsification, Quantum **4**, 235 (2020).

[25] J. Maldacena and D. Stanford, Remarks on the sachdev-ye-Kitaev model, Phys. Rev. D **94**, 106002 (2016).

[26] J. Polchinski and V. Rosenhaus, The spectrum in the sachdev-ye-Kitaev model, J. High Energy Phys. **2016**, 001 (2016).front matter+24

[27] S. Sachdev and J. Ye, Gapless Spin-Fluid Ground State in a Random Quantum Heisenberg Magnet, Phys. Rev. Lett. **70**, 3339 (1993).

[28] While QDRIFT provides performance guarantee on rather flexible choices of models, there is a specialized quantum algorithm for simulating the SYK models using much fewer gates [59].

[29] C.-F. Chen and A. Lucas, Operator growth bounds from graph theory, Commun. Math. Phys. **385**, 1273 (2021).

[30] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Towards the fast scrambling conjecture, J. High Energy Phys. **2013**, 1 (2013).

[31] An anonymous referee pointed out a potential connection with stochastic differential equation solvers in the Euler-Maruyama scheme. There, given a sample X_{sample} produced by the solver, there are also two ways to measure the error: strong error $\mathbb{E}|X - X_{\text{sample}}|$ and weak error $|\mathbb{E}X - \mathbb{E}X_{\text{sample}}|$. Interestingly, the convergence order is $1/2$ and 1, respectively, seemingly resembling the discussion here on QDRIFT.

[32] M. B. Hastings, Turning gate synthesis errors into incoherent errors, arXiv:1612.01011 (2016).

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

[34] C.-F. Chen, Concentration of OTOC and Lieb-Robinson velocity in random Hamiltonians, arXiv:2103.09186 (2021).

[35] C.-F. Chen and A. Lucas, Optimal Frobenius light cone in spin chains with power-law interactions, arXiv:2105.09960 (2021).

[36] D. Huang, J. Niles-Weed, J. A. Tropp, and R. Ward, Matrix concentration for products. *preprint* arXiv:2003.05437, 2020.

[37] C. M. Dawson and M. A. Nielsen, The solovay-kitaev algorithm, Quantum Info. Comput. **6**, 81 (2006).

[38] For the Solovay-Kiteav theorem, $3 \leq c \leq 4$ [37] and it is also known that $c = 1$ is the best possible exponent [60].

[39] 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 ρ in this definition. We refer to Ref. [40] for details.

[40] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[41] R. I. Oliveira, The spectrum of random k-lifts of large graphs (with possibly large k), J. Comb. **1**, 285 (2010).

[42] J. A. Tropp, Freedman's inequality for matrix martingales, Electron Commun. Probab. **16**, 262 (2011).

[43] I. Pinelis, Correction: "optimum bounds for the distributions of martingales in banach spaces" [Ann. Probab. **22**, 1679 (1994); MR1331198 (96b:60010)], Ann. Probab. **27**, 2119 (1999).

[44] R. I. Oliveira, Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges, arXiv:0911.0600 (2010).

[45] D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inform. Theory **57**, 1548 (2011).

[46] D. J. H. Garling, *Inequalities: a Journey Into Linear Analysis* (Cambridge University Press, Cambridge, 2007).

[47] D. Aharonov, A. Kitaev, and N. Nisan, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, p. 20, New York, NY, USA, 1998. Association for Computing Machinery.

[48] M. Suzuki, Decomposition formulas of exponential operators and lie exponentials with some applications to quantum mechanics and statistical physics, J. Math. Phys. **26**, 601 (1985).

[49] R. Ahlswede and A. Winter, Strong converse for identification via quantum channels, IEEE Trans. Inform. Theory **48**, 569 (2002).

[50] J. A. Tropp, User-friendly tail bounds for sums of random matrices, Found Comput. Math. **12**, 389 (2012).

[51] P. P. Varjú, Random walks in compact groups, Doc. Math. **18**, 1137 (2013).

[52] R. Versendaal, Large deviations for random walks on Lie groups, arXiv:1909.05065 (2019).

[53] T. Kathuria, S. Mukherjee, and N. Srivastava, On concentration inequalities for random matrix products. *preprint* arXiv:2003.06319 (2020).

[54] D. Christofides and K. Markström, Expansion properties of random cayley graphs and vertex transitive graphs via matrix martingales, Random Struct. Algorithms **32**, 88 (2008).

[55] J. A. Tropp, An introduction to matrix concentration inequalities, Foundations and Trends® in Machine Learning, **8**, 1–230 (2015).

[56] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis* (Birkhäuser, Basel, 2013).

[57] R. van Handel, Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.

[58] J. Lee, D. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction. *preprint* arXiv:arXiv:2011.03494, 2020.

[59] R. Babbush, D. W. Berry, and H. Neven, Quantum simulation of the sachdev-ye-kitaev model by asymmetric qubitization, Phys. Rev. A **99**, 040301 (2019).

[60] A. W. Harrow, B. Recht, and I. L. Chuang, Efficient discrete approximations of quantum gates, J. Math. Phys. **43**, 4445 (2002).