Composite Qdrift-product formulas for quantum and classical simulations in real and imaginary time

Matthew Pocrnic [1], Matthew Hagan [1], Juan Carrasquilla [1,2,3], Dvira Segal [4,1] and Nathan Wiebe [5,6,7]

[1]*Department of Physics, 60 Saint George St., University of Toronto, Toronto, Ontario, Canada M5S 1A7*

[2]*Vector Institute, MaRS Centre, Toronto, Ontario, Canada M5G 1M1*

[3]*Department of Physics and Astronomy, University of Waterloo, Ontario, Canada N2L 3G1*

[4]*Department of Chemistry and Centre for Quantum Information and Quantum Control,*
University of Toronto, 80 Saint George St., Toronto, Ontario, Canada M5S 3H6

[5]*Department of Computer Science, University of Toronto, Toronto Ontario, Canada*

[6]*Pacific Northwest National Laboratory, Richland, 99354 Washington, USA*

[7]*Canadian Institute for Advanced Research, Toronto, Ontario, Canada*

Recent study has shown that it can be advantageous to implement a composite channel that partitions the Hamiltonian H for a given simulation problem into subsets A and B such that $H = A + B$, where the terms in A are simulated with a Trotter-Suzuki channel and the B terms are randomly sampled via the Qdrift algorithm. Here we extend Qdrift and composite product formulas to imaginary time, formulating candidate classical algorithms for quantum Monte Carlo calculations. We upper bound the induced Schatten-$1 \to 1$ norm on both imaginary-time Qdrift and composite channels. Another recent result demonstrated that simulations of lattice Hamiltonians containing geometrically local interactions can be improved using a Lieb-Robinson argument to decompose H into subsets that contain only terms supported on that subset of the lattice. Here, we provide a quantum algorithm by unifying this result with the composite approach into "local composite channels" and we upper bound the diamond distance. We provide exact numerical simulations of algorithmic cost by counting the number of gates of the form $e^{-iH_j t}$ and $e^{-H_j \beta}$ to meet a certain error tolerance ϵ. In doing so, we optimize the partitioning into sets A and B using gradient boosted tree models from machine learning. These numerical studies are important given that product formulas have been historically known to outperform analytic upper bounds. We show constant factor advantages for a variety of interesting Hamiltonians, the maximum of which is a $\approx$20-fold speedup that occurs in the simulation of Jellium.

I. INTRODUCTION

Quantum simulations of quantum systems, suggested by Feynman in 1982 [1], provides what is perhaps the most natural application for quantum computers with expected exponential advantages over the best classical algorithms [2–4]. The quantum simulation problem can be summarized as follows: Given a Hamiltonian H, a time t, and an error tolerance ϵ, we wish to implement, on a quantum computer, an approximate function f to the *time evolution operator e^{-iHt}* such that $||e^{-iHt} - f(iHt)||_\kappa \leqslant \epsilon$ for some distance norm κ. The significance of this operator comes from quantum mechanics where e^{-iHt} is the solution to the Schrödinger equation with a time-independent Hamiltonian, $i\partial_t |\Psi\rangle = H|\Psi\rangle$ in units where $\hbar = 1$. Motivation for quantum simulation is drawn from the fact that it is a necessary subroutine for phase estimation, which allows one to learn the eigenvalues of H [5,6]. This is of particular interest in quantum chemistry and materials [7–11] where Hamiltonians are large and analytic solutions to

the eigenvalue problem are not known. In addition, quantum simulation also allows for the study of dynamical quantum observables [12]. Classically, implementing the time-evolution operator is a difficult problem due to how rapidly the matrices grow with respect to the size of the system of interest. As well, quantum effects such as interference are hard to simulate classically. This problem has received extensive attention over the last few decades, and many different solutions have been proposed. These solutions can in some sense be broken down into two families; those built on product formulas including Trotter-Suzuki [2,13,14] formulas and Qdrift [11,15–17] and those implementing more intricate Hamiltonian transformations such as linear combinations of unitaries [18,19], and qubitization [20–22]. Put simply, if we have a Hamiltonian that can be expressed as a sum of terms $H = \sum_j H_j$ with a time evolution operator $e^{-it \sum_j H_j}$, a product formula is an approximation to $e^{-iHt} \approx \prod_j e^{-iH_j t}$, which is not exact in general due to nonzero commutators in the Taylor series. The crucial difference between Trotter and Qdrift formulas is that the former directly and deterministically implements a product of propagators, while the latter provides the means for constructing stochastic formulas via importance sampling. This consequentially leads to very different and somewhat complementary error behavior that suggests these approaches

can be effectively combined [23]. This has recently led to composite Qdrift-Trotter simulation algorithms [24].

In the absence of fault-tolerant quantum computers, powerful methods have also been discovered to study the properties of H on classical computers, particularly those of the Quantum Monte Carlo type. The field of Quantum Monte Carlo encompasses a wide variety of algorithms, which broadly speaking involve approximating the properties of quantum systems by means of pseudorandom number generation and sampling. These methods are effective in calculating necessary quantities such as observables and integrals, and they are parallel by construction, making them a natural choice for use in high performance computing clusters. These methods have been deployed across a wide variety of fields including nuclear [25], impurity [26], many-body [27], and solid state physics [28]. In the aforementioned studies, multiple Monte Carlo algorithms are discussed. Our study here most naturally applies to the path integral Monte Carlo approach [29,30], particularly the discrete-time case [31–33] with time-independent Hamiltonians, although we believe these methods may be generalized to apply in the time-dependent case. Here, the propagator of interest is often the time-evolution operator, which has been Wick transformed to imaginary time $e^{-\beta H}$: $it \to \beta$. This is implemented both as a projection method to study the ground state of H, and to study the partition function $Z = \mathrm{Tr}\, e^{-\beta H}$ from statistical physics where β is the inverse temperature of the system. Using path integral Monte Carlo we can write the partition function or expectation values of other observables as a sum over paths, and then use sampling methods to approximate the integrals (see Sec. III D). Trotter-Suzuki formulas found some of their first applications in Monte Carlo calculations prior to becoming a popular ancilla-free quantum simulation algorithm. In this paper, we take the reverse approach and consider how composite Trotter-Qdrift simulation strategies [24] from quantum computing carry over to imaginary time. We also extend these composite strategies to account for locality of H based on Lieb-Robinson bounds [34] in terms of their application for quantum simulation (in real time). Much effort has also been dedicated towards improving the error scaling and resource requirements on Trotter-Suzuki formulas [35–37] and Qdrift [38–40]; however, for our purposes we consider only the original formulations.

Within this paper, we provide three contributions. First, we show that Qdrift and composite Qdrift-Trotter algorithms [24] can be extended to imaginary time with the same asymptotic scaling as in real time. The basis of these algorithms is a partitioning of the Hamiltonian such that $H = A + B$, where we simulate the subset A with Trotter-Suzuki formulas and B with Qdrift, to take advantage of the strengths of each. This promotes them to a candidate classical algorithm for quantum Monte Carlo calculations. In doing so, we provide bounds on Qdrift and Trotter-Suzuki formulas in imaginary time. Next, we show that composite algorithms can be applied to geometrically local lattices, by first decomposing the lattice into local subsets [34], and then partitioning the Hamiltonian terms supported on said subsets into localized Qdrift and Trotter simulations. In comparison to standard composite simulations, the local composite simulations recover the form of the original error bounds with an additional error term that is exponentially small. Finally, we provide a library to numerically evaluate the performance of composite channels in both real and imaginary time, as well as the performance of the local composite quantum algorithm in real time. Here, we also numerically introduce optimization schemes for conducting this partitioning based on heuristics and intuition that comes from the error bounds of each respective algorithm. Rather than computing bounds, we evaluate the exact exponential gate cost (the number of exponentials $e^{-iH_j t'}$ with time slice t' that appear in the product formula) at simulation time t for implementing an approximation to the time-evolution operator. Equivalent counts are performed in imaginary time for $e^{-\beta H}$ at inverse temperature β. This is done for a series of interesting Hamiltonians, which are drawn from quantum chemistry and spin systems, the former of which are generated using OpenFermion [41]. The importance of the numerical analysis lies in the historic observation that product type formulas can greatly exceed expectations, often significantly outperforming the tightest error bounds given the intricacy of commutator based formulas [42]. We find significant constant factor advantages using our composite approaches, with the maximum being a factor of 18.8 for a six-site Jellium Hamiltonian. We make the observation that composite methods seem to work better for Hamiltonians containing a large number of terms, and for those with sharply peaked spectral norm distributions.

The paper is organized as follows: We begin with detailed preliminary material in Sec. II that introduces the algorithms in question as well as some details regarding how they are analyzed analytically and numerically. Here we also provide some background on the Lieb-Robinson bound and define geometrically local Hamiltonians for which they apply. In Sec. III, we show how to extend composite channels to imaginary time and we provide bounds on the error for said channels. In doing so, we show that the Qdrift algorithm holds in imaginary time and that the same ϵ scaling can be achieved, suggesting it can also be used effectively for classical computation. We also provide a quantum algorithm that allows us to include locality as means for partitioning in the "local composite" channel, and show that doing so introduces exponentially small errors in Sec. IV. In Sec. V, we outline the details of our numerical simulations and describe the general features of our library. We then provide plots of the gate cost, or the number of exponential operators needed to meet a chosen error tolerance. We conclude in Sec. VI with a discussion and consider open questions.

II. PRELIMINARIES

In this section, we introduce the general physical setup, notation, and simulation approaches that will be used throughout the paper. We then review and summarize the algorithmic cost of Trotter-Suzuki and Qdrift product formulas of real-time time evolution, as well as their hybrid approach, which was introduced in Ref. [24].

A. General notation and assumptions

Before introducing the simulation algorithms, it is important to outline the notation that will be used throughout the paper, as well as the assumptions that remain relevant for calculations and proofs. The following assumptions will

remain valid unless otherwise stated. We assume that Hamiltonian operators H are time independent, bounded, and that they live in a finite 2^n dimensional Hilbert space $\mathcal{H}$. We further assume that these operators can be written as a finite sum of terms, $H = \sum_{i=1}^{L} H_i$ with each $H_i \in \mathbb{C}^{2^n \times 2^n}$. Later, in Sec. IV, we will discuss algorithms for geometrically local Hamiltonians in which case each term in this sum will only be supported on a finite subset of $\mathcal{H}$; we save those details for later. Note that we can normalize the terms in the above sum by factoring out their respective spectral norms (Schatten infinity norm on matrices), $h_i = ||H_i||$ such that $H = \sum_{i}^{L} h_i H_i$. We also impose $h_i \geqslant 0$ by pulling any phase factors into H_i; this can always be done without loss of generality. The Schatten p-norms will be a useful concept throughout, and they are defined as $||H||_p = (\sum_i s_i^p)^{\frac{1}{p}}$, where s_i are the singular values of the matrix H. From this definition, we observe that the Schatten infinity norm coincides with the spectral norm. Other norms will be discussed throughout the paper. Wherever a norm $|| \cdot ||$ is lacking a subscript, it will be assumed to be the Schatten infinity norm. All other norms will be subscripted appropriately. In what follows, Hamiltonian operators are taken to be time independent, and they generate dynamics through the time-evolution operator, $U = e^{-iHt}$.

Throughout, we make use of the density matrix formalism to describe mixed quantum states, with quantum channels being the object that evolve states, rather than the evolution operator e^{-iHt}. By quantum channels we mean a mapping $\mathcal{A}(\rho) = \sum_i E_i \rho E_i^\dagger$ that is completely positive and preserves the trace $\mathrm{Tr}(\rho)$. For this to be true, the operators E_i must satisfy the condition $\sum_i E_i E_i^\dagger = \mathbb{1}$. These conditions are required so that information regarding the channel operation is preserved and not lost to an external environment. For example, consider performing a measurement on the output density matrix $\mathcal{A}(\rho)$. If we find that $\mathrm{Tr}(\mathcal{A}(\rho)) \leqslant 1$, we conclude that the mapping does not provide a complete description of the processes involved, since there must be some other process(es) and therefore measurement outcomes that occur with nonzero probability [6]. For easy distinction, we reserve calligraphic fonts for channels $\mathcal{A}$ while operators A are written in the standard way. Given a density matrix $\rho(t_0)$ that describes the state at some initial time t_0, the time-evolution channel is then defined as

$$\mathcal{U}(\rho(t_0), t) := e^{-iH(t-t_0)} \rho(t_0) e^{iH(t-t_0)} \rightarrow \rho(t). \quad (1)$$

When not specifically defining evolutionary channels, we will often use $t_0 = 0$ and write $\rho(t = 0) = \rho$ to be the initial state. If time evolution is broken into m time intervals $\{\Delta t_1, \Delta t_2, \ldots, \Delta t_m\}$, the state vector evolution looks like $e^{-iH \Delta t_m} \ldots e^{-iH \Delta t_2} e^{-iH \Delta t_1} |\psi\rangle$. However, in the channel description we do not take products of channels, rather we apply a channel multiple times in a channel composition, which we denote with $\circ$. For example, $\mathcal{U}(t_{i+1}) \circ \mathcal{U}(t_i)(\rho) = e^{-iH t_{i+1}} e^{-iH t_i} \rho e^{iH t_i} e^{iH t_{i+1}}$. Note that in this expression t_i corresponds to the time interval during which the channel acts.

A significant portion of this paper is dedicated to the investigation of imaginary-time channels $\mathcal{U}(\rho, \beta) = \frac{1}{\mathcal{N}} e^{-\beta H} \rho e^{-\beta H}$ and the algorithms that closely approximate them. Upon first glance, the notion of imaginary time seems to violate the trace preserving property of a quantum chan-
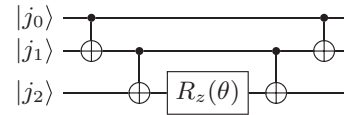

FIG. 1. A simple quantum circuit showing an implementation of an exponentiated product of Pauli spin operators.

nel, so we introduce $\mathcal{N}$ as a normalization factor. To avoid confusion, we use β throughout these discussions instead of t, motivated by statistical mechanics where β is considered an inverse temperature. In these cases, we are not considering a dynamical evolution of the system in the physical sense, but are rather interested in calculating properties of the Hamiltonian H, such as its partition function $Z = \mathrm{Tr} e^{-\beta H}$ or its ground state. This is because time evolution of a quantum system is isomorphic to the cooling of a statistical ensemble via a Wick rotation. We wish to make clear that all discussions pertaining to algorithms in imaginary time are not being proposed for *digital quantum simulation*, but as *classical algorithms* for use in calculations like quantum Monte Carlo. For a study of imaginary-time evolution on a quantum computer, see Ref. [43].

B. Algorithmic cost and input model

Given that we are dealing with channels containing only product formulas, a natural cost model is to count the number of exponential gates of the form $e^{-iH_j t'}$, required to achieve some error ϵ in a simulation. Here, H_i is a specific term chosen from a sum of Hamiltonian terms and t' is an appropriately chosen time slice of the total simulation time t; the following sections on the algorithms of interest explain how to choose these time slices to meet a desired simulation accuracy ϵ. Throughout, the aforementioned cost will be denoted C and often subscripted with the algorithm of interest. This cost model is convenient analytically as the algorithms we explore contain an exact amount of iterations or samples of terms that can be easily equated to a count of exponential gates. As input, we assume that the user has access to the decomposed Hamiltonian $H = \sum_j h_j H_j$ and not just the matrix version of H. In fact, this is the way that Hamiltonians are naturally written for spin systems and second-quantized chemistry anyway. From a computer science perspective, we could more formally assume we have query access to an oracle O_H that provides the exponentiated Hamiltonian terms: $O_H |010 \ldots\rangle \otimes |\psi\rangle = |010 \ldots\rangle \otimes e^{-iH_2 t'} |\psi\rangle$. In this case, we think of C as the total number of queries to O_H. As compilation procedures from O_H queries to gates are typically efficient, but application dependent, we perform our cost analysis on these queries as opposed to individual gate counts. In the case of common Pauli and fermionic Hamiltonians, circuits to implement said exponentials are well known and relatively simple to analyze [44]. For example, it is common in Heisenberg-like models to have interactions that are a product of Pauli operators such as $\sigma_i^\nu \sigma_{i+1}^\nu \sigma_{i+2}^\nu$. The time evolution of these operators $e^{-it(\sigma_i^\nu \sigma_{i+1}^\nu \sigma_{i+2}^\nu)}$ can be simply implemented using rotation gates $R_\nu(\theta) = exp(-i\frac{\theta}{2}\sigma_\nu)$. In fact, using the circuit in Fig. 1, we can implement the time evolution for any product of Pauli operators using just $R_z(\theta)$. This circuit effectively implements

$\exp(-i\frac{\theta}{2}\sigma_i^z\sigma_{i+1}^z\sigma_{i+2}^z)$, and we can build a product of any length by continuing this pattern. Switching this to a product of σ_x or σ_y is straightforward given that $R_x = \hat{H}R_z\hat{H}$ and $R_y = T^2R_z(T^2)^\dagger$. We can then adjust the simulation time t through our choice of angle θ. Fermionic Hamiltonians follow in a similar way, given that we can map them to a Pauli Hamiltonian using the Jordan-Wigner transformation, which will be outlined in Sec. V A 1. The exampled circuit also reinforces the need for product formula approximations since while we can simply implement the above-mentioned products of operators, implementing $e^{-it(\sigma^x+\sigma^z)}$ is highly nontrivial, and we approximate it as $e^{-it\sigma^x}e^{-it\sigma^z}$ as we detail in the following sections. In our cost model, this approximation would be counted as two gates. Numerically, counting exponential gates is also a convenient cost model since it allows for straightforward counting of gates as they are applied.

C. Trotter-Suzuki formulas

Given an operator that can be written as a finite sum of terms $A = \sum_i^L A_i$, the simplest approximation for the corresponding operator exponential is then $e^{\sum_i^L A_i} \approx \prod_i^L e^{A_i}$. While this holds as an equality for exponential functions for variables, it does not in general hold for operator exponentials since individual terms of A do not necessarily commute. This can be easily seen by considering the Taylor series of a simple exponential approximation $e^{(A+B)x} \approx e^{Ax}e^{Bx}$ to second order in x,

$$e^{(A+B)x} - e^{Ax}e^{Bx}$$

$$= \left[\mathbb{1} + Ax + Bx + \frac{(A+B)^2x^2}{2!} + \cdots\right]$$
$$- \left[\left(\mathbb{1} + Ax + \frac{A^2x^2}{2!} + \cdots\right)\left(\mathbb{1} + Bx + \frac{B^2x^2}{2!} + \cdots\right)\right] \tag{2}$$

$$= \frac{(AB+BA)x^2}{2} - ABx^2 + \cdots \tag{3}$$

$$= \frac{[B,A]x^2}{2} + \cdots \tag{4}$$

An error clearly arises due to a nonvanishing commutator, meaning the approximation is exact for commuting algebras. However, the nature of this error opens up the possibility for improving this approximation. If one can symmetrize this formula, the second-order commutator will vanish. In reference to the previous example, a perhaps obvious symmetrization is $e^{\frac{Ax}{2}}e^{Bx}e^{\frac{Ax}{2}}$, which annuls the error term $\frac{[B,A]x^2}{2}$, leaving only error terms in $\mathcal{O}(x^3)$. More generally, for an operator $A = \sum_i A_i$, the second-order Trotter-Suzuki decomposition is given by

$$S_2(Ax) = \prod_i^L e^{\frac{A_ix}{2}} \prod_L^i e^{\frac{A_ix}{2}}, \tag{5}$$

where the flipping of indices in the two products is indicative of reversing the ordering of products. One can readily imagine that it is then possible to continue this strategy and build higher-order formulas and cancel out arbitrarily high commutator errors in the product formula approximation. Indeed,

we can build these highly nontrivial decompositions in the following way [45,46]:

Definition II.1 (Trotter-Suzuki product formula). Given a linear operator A acting on a finite-dimensional Hilbert space that can be represented as a finite sum of linear operators $A = \sum_i^L A_i$, a Trotter-Suzuki product formula approximation $S_{2k}(Ax)$ for A of order $2k$ ($k > 1$) can be constructed in the following recursive fashion:

$$S_{2k}(Ax) := (S_{2k-2}(Axs_{2k}))^2$$
$$\times S_{2k-2}((1-4s_{2k})Ax)(S_{2k-2}(Axs_{2k}))^2, \tag{6}$$

where $s_{2k} = \frac{1}{4-4^{\frac{1}{2k-1}}}$.

Hence, in order to obtain a desired product formula of order $2k$, we start with the second-order formula $S_2(Ax)$ given in Eq. (5) and recursively build the higher-order formulas. To reiterate, the advantage of doing so is that for a product formula of order $2k$, the error of the approximation is $\in \mathcal{O}(t^{2k+1})$ with t here as the time interval [46].

When the general operator e^{Ax} is replaced by the Hamiltonian time-evolution operator e^{-iHt}, where $H = \sum_i^L H_i$, it becomes obvious how product formulas can be used to evolve quantum states $|\psi\rangle$. However, we also wish to generalize this to evolve probabilistic mixtures of states, which are represented by density matrices $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$. A Trotter-Suzuki evolution formula for a channel can be written as follows:

Definition II.2 (Trotter-Suzuki channel). Given a Hamiltonian H, a density matrix ρ, times t and t_0 ($t > t_0 \geqslant 0$), and order 2k, then a Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho(t_0), \Delta t)$ performs the operation $\mathcal{T}^{2k}(\rho(t_0), \Delta t) \rightarrow \rho(t)$ and can be defined as

$$\mathcal{T}^{2k}(\rho(t_0), t) := S_{2k}(-iH\Delta t)\rho S_{2k}(-iH\Delta t)^\dagger \tag{7}$$

where $S_{2k}(-iHt)$ represent the product formulas from Eq. (5) and $\Delta t = t - t_0$.

Since the error of this approximation accumulates considerably for long iteration times, we introduce the iteration parameter r, and apply the channel r times with time intervals t/r. As a result, longer time simulations become more expensive in the number of iterations r required to manage the error that accumulates to do the simulation time t. The composition of r Trotter-Suzuki channels is straightforward, and applying r channels each for time t/r can be written as

$$\mathcal{T}^{2k}(\rho, t)^{\circ r} = \prod_{j=1}^r S_{2k}\left(-iH\frac{t}{r}\right)\rho \prod_{j=1}^r S_{2k}\left(-iH\frac{t}{r}\right)^\dagger. \tag{8}$$

The cost of this algorithm for first-order decomposition is simply rL where L is the number of Hamiltonian terms. For order $2k$, the cost is ΥLr where $\Upsilon = 2 \times 5^{k-1}$.

Looking at cost formulas for different orders $2k$, we desire information about r, specifically, the minimum value of r required to meet some error tolerance ϵ where $||\mathcal{T}^{2k}(\rho, t/r)^{\circ r} - \mathcal{U}(\rho, t)||_\kappa \leqslant \epsilon$ for some norm κ. In Ref. [24], the diamond norm is selected, which is the completely bounded trace norm between channels tensored with the environment. The diamond norm is defined as $||\mathcal{U}(t) - \mathcal{V}(t)||_\diamond := \sup_{\rho:||\rho||\leqslant 1} ||(\mathcal{U}(\rho, t) - \mathcal{V}(\rho, t)) \otimes \mathbb{1}||_1$. This is a good analytical norm to choose, given that it maintains the interpretation of

distinguishability between states from the trace distance (see Sec. B 1 b). However, it is a bound on channels (maximized over possible input states ρ) and thus yields the interpretation of distinguishability between quantum channels. This norm is also subadditive and submultiplicative, which makes it convenient to work with. The procedure in Ref. [24] upper bounds this norm, sets the value equal to ϵ, and solves the expression for r to arrive at the following first- and $2k$th-order costs:

$$C_{TS}^1(H, t, \epsilon, 1) = Lr \leqslant L \left\lceil \frac{t^2}{2\epsilon} \sum_{i,j} h_i h_j ||[H_i, H_j]||_\infty \right\rceil, \quad (9)$$

$$C_{TS}^{2k}(H, t, \epsilon, 2k) = \Upsilon Lr$$

$$\leqslant \Upsilon L \left\lceil \frac{(\Upsilon t)^{1+1/2k}}{\epsilon^{1/2k}} \left(\frac{4\alpha_{\text{comm}}(H, 2k)^{1/2k}}{2k+1} \right) \right\rceil, \quad (10)$$

where C, to reiterate, is the number of operator exponentials that need be applied rather than an exact gate count where gates are drawn from some universal gate set. Here, α_{comm} has the following definition:

$$\alpha_{\text{comm}}(H, 2k) := \sum_{\gamma_i \in \{1, \dots L\}} \left(\prod h_{\gamma_i} \right)$$

$$\times \left|\left|[H_{\gamma_{2k+1}}, [H_{\gamma_{2k}}, \dots [H_{\gamma_2}, H_{\gamma_1}] \dots]]\right|\right|_\infty. \quad (11)$$

This is a nested commutator error that arises from $2k$ symmetrizations in the product formula. The main takeaway from this formula is the structure of the Hamiltonian that the simulation cost depends on. Trotter-Suzuki channels clearly depend on the number of terms in the Hamiltonian L and on the commutator structure of the Hamiltonian. The simulation cost does not directly depend on the *magnitude* of the terms $||H_i||$, only indirectly through their commutators. This is important to keep in mind when examining the next algorithm and its cost dependence.

D. Qdrift random compiler

The Qdrift algorithm was discovered by Campbell [15]. It generates random product formulas allowing for the evolution of a system to *stochastically* drift (through the Hilbert space) towards the true evolution with high probability. Using the introduced quantum channel notation, the algorithm works as follows:

Definition II.3 (Qdrift channel). Given a Hamiltonian $H = \sum_{i=1}^L h_i H_i$, density matrix ρ, times t and t_0 ($t > t_0 \geqslant 0$), and let $p_i = \frac{h_i}{\lambda}$ with $\lambda = \sum_{i=1}^L h_i$ be a probability distribution from which Hamiltonian terms are drawn, then a Qdrift channel $\mathcal{Q}(\rho(t_0), \Delta t) \to \rho(t)$ can be defined as

$$\mathcal{Q}(\rho(t_0), t) := \sum_{i=1}^L p_i e^{-iH_i \lambda \Delta t} \rho e^{iH_i \lambda \Delta t}, \quad (12)$$

where $\Delta t = t - t_0$.

This channel mixes unitaries with a single sample from the distribution p_i. Similarly to the Trotter-Suzuki approach, the accuracy of this approach improves with the number of samples N that go into our "random product formula". To write the channel for multiple samples, it is useful to think of sampling a vector $\mathbf{j} = \{j_1, j_2, \dots j_N\}$ of length N, i.i.d from the distribution $P_\mathbf{j} = \lambda^{-N} \prod_{k=1}^N h_{j_k}$. This corresponds to implementing the unitary $\mathbf{V}_\mathbf{j} = \prod_{k=1}^N e^{-iH_{j_k} \tau}$ with $\tau = t\lambda/N$. With this in mind, we can neatly write down the expression for the channel with arbitrary samples N as

$$\mathcal{Q}(\rho, t)^{\circ N} = \sum_\mathbf{j} P_\mathbf{j} \mathbf{V}_\mathbf{j} \rho \mathbf{V}_\mathbf{j}^\dagger, \quad (13)$$

where the sum is performed over all possible vectors $\mathbf{j}$ of length N. In terms of preparing these gates with a quantum circuit, this algorithm can be thought of as a linear combination of unitaries under classical control. Similarly to the Trotter-Suzuki formula, we are also interested in the algorithmic cost of Qdrift. A diamond distance upper bound is provided in Ref. [15], which is then used via the same procedure to provide a cost function in Ref. [24], by using $N = 4\lambda^2 t^2/\epsilon$ and restricting $\epsilon \in (0, \lambda t \ln(2)/2)$,

$$C_{QD}(H, t, \epsilon) = \frac{4\lambda^2 t^2}{\epsilon}. \quad (14)$$

The important take home message here is the difference in the structure of the algorithmic cost between the Trotter-Suzuki formula [Eq. (9)] and Qdrift [Eq. (14)]. In Qdrift, the cost does not depend on the number of Hamiltonian terms L, or the commutator structure. In contrast, it depends on the size of the terms in the sum of the spectral norms λ. It is important to keep this in mind when constructing composite channels.

E. Composite simulation formulas

Equipped with the Trotter-Suzuki and Qdrift channel formulas, we wish to hybridize these approaches to form a *composite* time evolution channel. The eventual goal is to do so in such a way that we can minimize the "side effects" of each algorithm while taking advantage of the strengths of each approach. Following Ref. [24], we define the composite channel as follows. We consider the Hamiltonian $H = A + B$, where $A = \sum_{j=1}^{L_A} A_j$ and $B = \sum_{j=1}^{L_B} B_j$. Other terms of interest, such as the number of summands in a set L are written with a subscript A to indicate that they belong to a set of Hamiltonian terms that will be simulated by a Trotter-Suzuki channel, and the same goes for the B terms, for a Qdrift channel. The composite channel takes the form

$$\mathcal{X}^{2k}(\rho, t) = \mathcal{Q}_B(t) \circ \mathcal{T}_A^{2k}(t)(\rho), \quad (15)$$

where the channel subscripts indicate the subset of the Hamiltonian that it is simulating. Note that there is substantial freedom in the construction of this channel. The first obvious freedom is the partitioning into the sets A and B. Section V B is devoted to this task. The next choice lies in the Trotter-Suzuki order $2k$. Further, we can construct an outer loop as in Ref. [24], which uses a Trotter-Suzuki style symmetrization of the channels themselves. We further discuss this in later sections where we will resort to the notation $\mathcal{X}^{2k,2g}$ for $2k$ *inner,* $2g$ *outer* orders respectively. If only a single superscript is given, it is understood to be the inner order with the outer order fixed at 1 as in Eq (15). For example, a composite channel with outer order $2g = 2$ is written as

$$\mathcal{X}^{2k,2}(\rho, t) = \mathcal{Q}_B(t/2) \circ \mathcal{T}_A^{2k}(t) \circ \mathcal{Q}_B(t/2)(\rho). \quad (16)$$

Such an "outer loop" with a respective "outer order" was introduced as a strategy to make more detailed analysis of simulation costs. The outer loop can be thought of as a product formula for composite channels, symmetrizing the channels and allocating the respective time slice, while the inner Trotter order does so on the operator exponentials as before. However, when introducing the outer order in Ref. [24] it was enforced to match the same order as the inner factorization. In this paper, one of our goals is to numerically evaluate the exact costs of these composite algorithms, and this gives us the freedom to independently set the inner and outer orders. However, an actual application of the outer loop is situational as it requires specific knowledge about the commutator structure for it to be of use.

We now restate the first-order cost result for composite channels from Ref. [24], where the A terms are placed in a Trotter channel and the B terms are placed in Qdrift,

$$C_{\text{comp}}(A, B, t, \epsilon) = (L_A + N_B)r$$

$$= (L_A + N_B) \left\lceil \frac{t^2}{\epsilon} \left(\frac{1}{2} \sum_{i,j} a_i a_j ||[A_i, A_j]||_\infty \right. \right.$$

$$\left. \left. + \frac{1}{2} \sum_{i,j} a_i b_j ||[A_i, B_j]||_\infty + \frac{4\lambda_B^2}{N_B} \right) \right\rceil. \tag{17}$$

We only restate the first-order Trotter-Suzuki result due to the fact that this is the expression from which we will draw most of intuition to build our heuristics for partitioning. As well, the second-order cost formula is of a similar structure with extra α_{comm} and Υ terms, but it is based on channels with matching inner order and outer order, which our numerical work does not follow in general. Inspecting Eq. (17), it clearly inherits the structure from Trotter in its dependence on the commutators and number of terms L_A in the set A (first term), and on Qdrift in the sum of the spectral norms λ_B and number of samples N_B of the terms in the set B (third term). The second term adds to the cost due to the hybridization of the Trotter and Qdrift channels. Clearly, in order to gain as much advantage as possible using this algorithm over the noncomposite approaches, we desire that the Hamiltonian of interest has a structure that we can exploit via a good choice of partitioning. An interesting structure might be of an H that contains terms of largely varying magnitude, such that there might be a sharp contrast between the small and large terms, rather than their spectral norms following, say, a Gaussian distribution. Having small magnitude terms that largely outnumber the large-magnitude terms is also likely desirable. This understanding naturally leads into the later topic of how to choose the partitioning. This task can be motivated by the structure of H and its commutators, if known *a priori*. For example, inspecting Eq. (17), if we have a small number of large commuting terms and numerous small non-commuting terms, it seems natural to place the large terms into the Trotter channel, but sample the smaller magnitude terms with Qdrift. As we build higher orders of the composite channels, we can also expect a certain time scaling trade-off between the algorithms. Higher-order Trotter-Suzuki

formulas will asymptotically outperform Qdrift at long times; however, the overhead that comes with building these formulas will surely lead to them being outperformed at short times by taking a small number of Qdrift samples to approximate ρ. In this spirit, with a good partitioning, we expect there to exist a large intermediate t region in which composite channels outperform their constituents. The question of choosing effective partitioning will be further addressed in Sec. V B.

F. Local lattice Hamiltonians

Given the observation of commutator-dependent error in previous sections, an interesting system to study are geometrically local Hamiltonians. These systems are defined on a lattice, and they have only local interactions, such as nearest-neighbour terms. It turns out, via a Lieb-Robinson bound (Sec. II G), that commutators of time evolved local observables fall off exponentially with the separation distance in the lattice. This means that guarantees can be made on the rate at which information propagates in these lattice models, which directly implies that the dynamics of disjoint subsets are uncorrelated for some finite time. In the context of information processing, bounded information propagation suggests certain algorithmic approximations like a product formula can be made much more effective than in the naive case where locality is not taken into account. This motivates the search for a local product decomposition to simulate the time evolution of local systems. In fact, computational advantages have been demonstrated for local Hamiltonians in Ref. [34]. We follow this study in defining geometrically local Hamiltonians as

Definition II.4. Given a D-dimensional lattice $\Gamma \in \mathbb{Z}^D$, a local Hamiltonian can be written as

$$H = \sum_{X \subseteq \Gamma} H_X, \tag{18}$$

where H_X is only supported on the subset X that contains only immediately adjacent lattice points, and it acts as $\mathbb{1}$ on $\mathbb{Z}^D \in \Gamma \backslash X$. Equivalently, $H_X = 0$ if $diam(X) > 1$, where $diam(X) = \max_{x,x' \in X} dist(x, x')$ and $dist$ is the graph distance between the lattice indices. Each term in the Hamiltonian is also normalized such that $||H_X|| \leqslant 1$.

In our analysis below, we will explicitly state when we are referring to a local Hamiltonian. These Hamiltonians can be used to describe a wide variety of physical systems, perhaps most famously is the quantum Heisenberg model, which will be investigated numerically in Sec. V. If locality is not explicitly mentioned then the Hamiltonian is of the aforementioned more general structure.

G. Lieb-Robinson bound

A Lieb-Robinson bound is essentially a bound on the speed at which information propagates through a quantum system that has interactions governed by a local Hamiltonian. This result is particularly interesting due to the fact that it holds for nonrelativistic quantum systems, meaning that in no part of the system is the finite speed of light c enforced. Instead, the locality of the system's interactions, as well as the geometry of the lattice leads to the emergence of a Lieb-Robinson velocity v_{LR}, which bounds the speed of causality. The central idea behind this bound is that the commutator between the time evolution of an operator A_X and another operator B_Y, such that

the operators are supported on disjoint sets $X \cap Y = \varnothing$, have an exponentially small commutator. More formally, a general Lieb-Robinson bound is often given as [47,48]

$$||[e^{-iH_\Gamma t} A e^{iH_\Gamma t}, B]|| \leqslant 2||A||||B||F e^{-\mu(dist(X,Y) - v_{LR}|t|)}, \quad (19)$$

where μ is a constant that depends on the lattice and F depends on the cardinality of the sets X and Y. It is also important to note how the distance between sets is defined, and it appears in the exponent above. The distance between the sets X and Y is $dist(X, Y) = \min_{x \in X, y \in Y} dist(x, y)$, where $dist(x, y)$ is the distance between elements in the set; in a lattice this measure is just the absolute value of the difference of their indices. From a simulation viewpoint, and specifically regarding our Trotter-Suzuki product formulas, we know that the error in these approximations depend on commutators of terms. Therefore, carefully-chosen product decompositions may have exponentially small error. This understanding motivated the algorithm in Ref. [34], and in Sec. IV we combine this algorithm with the composite framework.

III. IMAGINARY-TIME CHANNELS ALGORITHMIC ANALYSIS

Within this section, we show that the composite algorithms proposed by Hagan and Wiebe in Ref. [24], for the purposes of digital real-time quantum simulation, can be applied onto the *imaginary-time* case with similar error bounds and asymptotics. As mentioned previously, the algorithms in this section are formulated with the intention of calculating properties of the Hamiltonian on *classical* computers. These algorithms can therefore be considered as quantum-inspired classical algorithms, a topic unrelated to the problem of quantum imaginary-time evolution [43]. For comparison, quantum imaginary-time evolution, sometimes abbreviated QITE, is a method for conducting imaginary-time evolution on a quantum computer. The key idea for QITE algorithms is that unitary evolution by an effective Hamiltonian can approximate the action of imaginary-time evolution [Eq. (21) below]. Given that we start and end on a normalized state, the success of the approach relies on k-local Hamiltonians with strongly bounded spatial correlation lengths. Rather than focusing on this problem, our paper extends recent developments in the theory of quantum algorithms to imaginary time in the classical setting; we are not limited by conditions of unitary evolution with this restriction to classical algorithms. Within this section, we introduce the imaginary-time Qdrift, Trotter-Suzuki, and the composite channels, and we bound their distance norm with respect to the ideal unitary evolution.

We first introduce the notion of an imaginary-time evolution channel, given in the definition below, and demonstrate how it can be viewed as a ground-state preparation, or "cooling" procedure.

Definition III.1 (Imaginary-time evolution channel).
Given a Hamiltonian H and input state ρ, we define the imaginary time evolved state as the action of the following map:

$$\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}\mathcal{U}(\rho, \beta)} = \frac{e^{-\beta H} \rho e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H} \rho e^{-\beta H})}. \quad (20)$$

This can be seen as an imaginary-time evolution channel most straightforwardly when considering quantum state vectors as opposed to density matrix. The time-independent Schödinger equation tells us $|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle$. If we take this expression and perform a Wick rotation that sends $t \to -i\beta$ we get that $|\psi(\beta)\rangle = e^{-\beta H}|\psi(0)\rangle$. If we consider the action of this matrix $e^{-\beta H}$ on the density matrix of the state $|\psi(0)\rangle$ we get $e^{-\beta H}|\psi(0)\rangle\langle\psi(0)|e^{-\beta H}$ due to the fact that $e^{-\beta H}$ is a Hermitian operator. Relaxing the input from a pure state to a density matrix and normalizing by the trace yields the channel provided in Definition III.1.

Often when imaginary time is discussed in quantum mechanics the state vector representation is used, in which case the imaginary-time exponential operator has the property

$$\lim_{\beta \to \infty} \frac{e^{-\beta H}|\psi_i\rangle}{||e^{-\beta H}|\psi_i\rangle||_2} = |\psi\rangle_{GS}. \quad (21)$$

This clearly only holds whenever our initial state $|\psi\rangle_i$ has nonzero overlap with the ground state $|\psi\rangle_{GS}$, which is expressed as $\langle\psi_i|\psi_{GS}\rangle \neq 0$. In the density matrix picture, we have the following equivalent property:

$$\lim_{\beta \to \infty} \frac{e^{-\beta H} \rho_i e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H} \rho_i e^{-\beta H})} = \rho_{GS}, \quad (22)$$

such that $\mathrm{supp}(\rho_i) \cap \mathrm{supp}(\rho_{GS}) \neq \varnothing$. This property can be seen by expanding ρ into an arbitrary mixed state in the energy eigenbasis. Now that the maps of interest have been made clear, we can consider algorithms that closely approximate them and analyze their performance analytically by bounding their error. An important detail is the choice of distance norm to quantify this error. In comparison to real-time composite algorithms, the aforementioned diamond norm is not a good choice for imaginary time, given that this norm is about distinguishing quantum operations on a system and ancillary space, whereas here we are simply performing classical calculations. For this reason, we bound the following induced Schatten $p \to q$ norm on a map Φ: $||\Phi||_{p \to q} = \max_{||\rho||_q = 1} ||\Phi(\rho)||_p$. Here, Φ will be the difference between the ideal map and that which is generated by the composite algorithm, and we will investigate the case of the induced trace norm where $p = q = 1$. Despite the fact that the maps being investigated in this section are nonlinear, due to the trace operation in the denominator, the outputs of the maps are Schatten-class linear operators, and thus the induced Schatten norm is well defined. Further, by analyzing $1 \to 1$ norms for two density matrices, such as $\max_{||\rho||_1 = 1} ||\rho - \sigma||_1 \leqslant \epsilon$, we can guarantee from properties of the trace distance that the measurement statistics for ρ and σ will deviate by at most ϵ.

A. Imaginary-time Qdrift

Introduced in Sec. II D, Qdrift was introduced as a quantum simulation algorithm that, in real-time simulations, has a unique property in that its error does not depend on the number Hamiltonian summands, nor on the commutators between them, only on the total size of these operators. To the best of our knowledge, this algorithm has not been applied to imaginary-time calculations. We recover all the characteristic properties of the Qdrift channel in imaginary time with renormalizing trace operations in the following Theorem.

Theorem III.1 (Imaginary-time Qdrift). Given a Hamiltonian $H = \sum_j h_j H_j$ with $\lambda = \sum_j h_j$, imaginary time β, number of samples N, and a density matrix initial state ρ, then the induced Schatten $1 \to 1$ norm of the difference between the imaginary-time Qdrift channel and exact imaginary-time evolutionary channel has the following bound:

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}\,\mathcal{U}(\rho, \beta)} - \left(\frac{\mathcal{Q}(\rho, \beta/N)}{\mathrm{Tr}\,\mathcal{Q}(\rho, \beta/N)} \right)^{\circ N} \right\|_{1 \to 1}$$

$$\leqslant \frac{4\beta^2 \lambda^2}{N} \frac{e^{2\lambda\beta/N}}{2 - e^{2\beta\lambda/N}} + 2 \frac{\frac{4\beta^4\lambda^4}{N^3} e^{6\lambda\beta/N}}{1 - \frac{2\beta^2\lambda^2}{N^2} e^{2\lambda\beta/N}} + 2 \frac{2\beta^2\lambda^2}{N} e^{4\lambda\beta/N}, \tag{23}$$

given that $|e^{2\beta\lambda/N} - 1| < 1$, which is satisfied if $N > \frac{2\beta\lambda}{\ln 2}$. Further, If the constraint $\frac{\lambda}{N} \leqslant 0.01$ is satisfied, then the bound simplifies to

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}\,\mathcal{U}(\rho, \beta)} - \left(\frac{\mathcal{Q}(\rho, \beta/N)}{\mathrm{Tr}\,\mathcal{Q}(\rho, \beta/N)} \right)^{\circ N} \right\|_{1 \to 1} \leqslant \frac{C\beta^2 \lambda^2}{N}, \tag{24}$$

where the constant $C \approx 29.71747$.

For a proof of this theorem see Appendix A. The proof consists of repeated applications of the triangle inequality and submultiplicative property of the diamond norm, as well as manipulations with the geometric series and a tail bound on a Taylor series. The geometric series arises when one uses a Taylor expansion on the terms in the denominator and takes the trace of the zeroth-order term, which is just $\mathrm{Tr}\,\rho = 1$. A constant theme throughout the analytics in this paper are multiple exponential factors that arise due to the techniques used to bound trace terms. Multiple techniques are used to accomplish this throughout, and the exponential factors are unavoidable in imaginary time, given that the exponential operator is no longer unitary and has an exponentially decaying norm. For clearer interpretation of the bounds we use linearization techniques to either show that they disappear under some constraints when proving an inequality, or showing they become $\mathcal{O}(1)$ in some limit for an asymptotic bound. The important takeaway from this theorem is that the Qdrift error scales essentially the same was in imaginary time as it does in real time, thus making it a candidate algorithm to be applied in quantum Monte Carlo calculations. As well, this promise on the ratio $\frac{\lambda}{N}$ is reasonable to make as this algorithm excels in the regime of small terms.

B. Imaginary-time Trotter-Suzuki

Different from Qdrift, work has already been done to provide bounds on nonunitary product formulas [14]. However, these bounds are given in terms of operator norms rather than channel distances, as well as they are given without renormalizing trace operations. However, these existing results are integral in proving the theorem below.

Theorem III.2 (Imaginary-time Trotter-Suzuki channels). Given a Hamiltonian $H = \sum_j h_j H_j$ with $\lambda = \sum_j h_j$, imaginary time β, time steps r, and a density matrix initial state ρ, then the induced Schatten $1 \to 1$ norm of the difference between the imaginary-time Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho, \beta)$ and the exact imaginary-time evolutionary

$\mathcal{U}(\rho, \beta)$ channel has the following bound:

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))} \right)^{\circ r} \right\|_{1 \to 1}$$

$$\leqslant 2\Upsilon^{2k+1} \frac{\alpha_{\mathrm{comm}}(H, 2k)}{(2k+1)!} \frac{\beta^{2k+1}}{r^{2k}} e^{4\Upsilon\beta\lambda/r} (e^{4\beta\lambda/r} e^{2\beta\|H\|/r} + e^{2\beta\|H\|/r}), \tag{25}$$

which yields the following asymptotic bound:

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))} \right)^{\circ r} \right\|_{1 \to 1}$$

$$\in \mathcal{O}\left(\Upsilon^{2k+1} \frac{\alpha_{\mathrm{comm}}(H, 2k)}{(2k+1)!} \frac{\beta^{2k+1}}{r^{2k}} \right). \tag{26}$$

Here Υ is the number of stages of the product formula, $2k$ is the order of the product formula, α_{comm} is defined as in Eq. (11), and λ is the sum of the spectral norms of the Hamiltonian summands. Here $\mathcal{O}(\cdot)$ is understood in the infinite limit of its arguments.

For a proof of this theorem, see Appendix A. The proof follows similar strategies to that of imaginary-time Qdrift; however, one cannot simply expand a geometric series in the denominator here given that we have a product instead of a sum of terms. So we instead bound the trace terms using a combination of von Neumann's trace inequality and Weyl's inequality regarding the singular values of matrices. Similar to the Qdrift case, we obtain a similar looking bound to that obtained in [24] for the real-time case, with some additional exponential factors as expected. Considering asymptotics, the bounds behave the same in both cases where we again see the dependence of the error on a nested commutator sum.

C. Imaginary-time composite channels

Now equipped with error bounds on both Qdrift and Trotter-Suzuki channels in imaginary time, we can proceed with the analysis of channels composed of the two. Given the difficulty of analyzing the renormalizing operations in the trace terms, we will again make use of asymptotic notation as was done in the Trotter-Suzuki analysis in the previous section. We will once again write a partitioning of the Hamiltonian as $H = A + B$, where $A = \sum_{i=1}^{L_A} a_i A_i$ are the terms placed in the Trotter-Suzuki channel and $B = \sum_{i=1}^{L_B} b_i B_i$ are placed in the Qdrift channel.

Theorem III.3 (Imaginary-time composite channels). Given a partitioned Hamiltonian $H = \sum_j a_j A_j + \sum_i b_i B_i$ with $\lambda_A = \sum_j a_j$, $\lambda_B = \sum_j b_j$, imaginary time β, time steps r, Qdrift samples N_B, and a density matrix initial state ρ, then the induced Schatten $1 \to 1$ norm of the difference between the imaginary-time composite channel $\mathcal{X}^{2k}(\rho, \beta)$ of order $2k$ and the exact imaginary-time evolutionary channel $\mathcal{U}(\rho, \beta)$

has the following bound:

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}\mathcal{U}(\rho, \beta)} - \frac{\mathcal{X}^{2k}(\rho, \beta)}{\mathrm{Tr}\mathcal{X}^{2k}(\rho, \beta)} \right\|_{1 \to 1}$$

$$\in \mathcal{O}\left(\frac{\beta^2 \lambda_B^2}{N_B r} + \Upsilon^{2k+1} \frac{\alpha_{\mathrm{comm}}(A, 2k)}{(2k+1)!} \frac{\beta^{2k+1}}{r^{2k}} + ||[A, B]|| \frac{\beta^2}{r} \right) \tag{27}$$

given that $r \geqslant \Upsilon \beta \lambda$. Here, all parameters in this bound correspond to the same quantities in Theorems III.1 and III.2, where the subscripts A and B indicate their belonging corresponding set.

For a proof of this bound see Appendix A. The condition on r is not overly restrictive, nor arbitrary. In general, we do not expect to simulate the imaginary-time evolution with r sublinear in either the imaginary time β or the norms of the Hamiltonian λ, and Υ is an artefact of using higher-order Trotter-Suzuki formulas. This result yields three error terms that we might expect. It is a sum of the Qdrift and Trotter error, plus the error that arises from the initial partitioning of the Hamiltonian into two subsets. Therefore, it becomes apparent that if we have information about the terms in the Hamiltonian, we can potentially exploit the form of this error bound by choosing a good partition. By good partition, we mean one that is close to the optimal one that minimizes the error. We later show in Sec. V B how to possibly go about choosing a partition using intuition from physics, as well as provide a machine learning regressor to find the optimal partition. The later is more for a proof of principle rather than something you would run in practice due to the large overhead.

D. Examples and implementation

Within this section we illustrate an implementation of the imaginary-time composite channel with an example. Approximating the imaginary-time propagator $e^{-\beta H}$, which is a function of a Hamiltonian H that is hard to exponentiate in general, can be achieved through the algorithm written in pseudocode below. This algorithm returns a list of propagators (of which their product approximates $e^{-\beta H}$) that are instead easily diagonalizable.

Algorithm 1. Pseudocode for implementation of an imaginary-time composite simulation using a high-order Trotter formula and a partitioning heuristic to divide the Hamiltonian terms between the two channels.

Input. Simulation parameters: Trotter-Suzuki inner-order $2k$, Qdrift samples N_B, iterations r, and imaginary time β. A list of Hamiltonian terms $H = \sum_j h_j H_j$, two deterministic classical functions: PARTITION(H), which returns two lists of Hamiltonian terms $A = \sum_q a_q A_q$ and $B = \sum_p b_p B_p$, and TROTTER($A, 2k, \beta, r$) that returns a Trotter-Suzuki product formula of order $2k$ with time slice $\frac{\beta}{r}$. We also require a classical oracle function SAMPLE(), which returns a value j from the probability distribution $p_j = \frac{b_j}{\sum_l b_l}$.

Output. A *composite imaginary-time product formula* in the form of a list of propagators, with *inner-order* $2k$ and *outer-order* 1, which closely approximates the imaginary-time propagator $e^{-\beta H}$.

Function *CompositeListCompilation*($H, \beta, r, 2k, N_B$):

  $A, B \leftarrow$ PARTITION(H);
  $m \leftarrow 1$;
  $\mathcal{X} \leftarrow \{ \}$(empty ordered list);
  $\mathcal{T} \leftarrow$ TROTTER($A, 2k, \beta, r$);
  while $m \leq r$ **do**
    $\mathcal{Q} \leftarrow \{ \}$(empty ordered list);
    $\lambda_B \leftarrow \sum_l b_l$;
    $j \leftarrow 1$;
    while $j \leq N_B$ **do**
      $j \leftarrow$ SAMPLE();
      append $e^{-\beta \lambda_B B_j / N_B r}$ to $\mathcal{Q}$;
      $j \leftarrow j + 1$;
    end
    append $\mathcal{T}$ to $\mathcal{X}$;
    append $\mathcal{Q}$ to $\mathcal{X}$;
    $m \leftarrow m + 1$;
  end
  return $\mathcal{X}$;

The algorithm above is written for illustrative purposes and simply constructs a composite product formula in the form of an ordered list. This can then be used as a subroutine for other algorithms in computational physics, such as those from the field of quantum Monte Carlo (QMC). For example, in statistical physics calculating the partition function Z is of

central interest. Given this returned list, which for clarity, is not a channel, we can approximate Z as

$$Z = \text{Tr}(e^{-\beta H})$$
$$= \text{Tr}\left(\left(e^{-\frac{\beta}{r}A_1} \dots e^{-\frac{\beta}{r}A_{L_A}} e^{-\frac{\beta\lambda_B}{rN_B}B_1} \dots e^{-\frac{\beta\lambda_B}{rN_B}B_{N_B}}\right)^r\right) + \epsilon, \quad (28)$$

where ϵ is bounded by a set of functions implied by Theorem III.3. Next, we insert the resolution of the identity between each propagator such that the trace disappears

$$Z = \sum_{j_1, j_2, \dots j_n} \langle j_n | e^{-\frac{\beta}{r}A_1} | j_1 \rangle \dots \langle j_{L_A-1} | e^{-\frac{\beta}{r}A_{L_A}} | j_{L_A} \rangle$$
$$\times \langle j_{L_A} |^{-\frac{\beta\lambda_B}{rN_B}B_1} | j_{L_A+1} \rangle \dots \langle j_{n-1} | e^{-\frac{\beta\lambda_B}{rN_B}B_{N_B}} | j_n \rangle + \epsilon, \quad (29)$$

given that $n = |\mathcal{X}|r$ or the number of terms in the list multiplied by the number of iterations. Note that the indices are written such that there are $L_A r$ Trotter terms and $N_B r$ Qdrift terms, with the difference that the subscripts indexing the Qdrift terms are dummy indices, as these propagators are importance sampled at random, as shown in the pseudocode. Now, with the above expression, we can either diagonalize these terms and Monte Carlo sample this sum, or impose a classical mapping that allows us to apply some of the well known QMC algorithms. These classical mappings depend on the system in play, for example, if we are interested in Heisenberg or XXZ chains, it is possible to map the model onto a higher-dimensional Ising model and apply the method of world lines [49]. The motivation for using this more complicated propagator list, as opposed to the standard Trotterization, is that a propagator with a lower error requires a smaller r to achieve the desired precision, and can therefore reduce the path length of the Monte Carlo path integral required to approximate Z above. Our numerics show that this framework can achieve a substantially lower error in the right conditions via attaining a smaller C in multiple cost plots in Sec. V D 2, therefore, highlighting the potential applications of this algorithm to improve QMC simulations.

In this example, it was chosen to show arbitrary inner-order and outer-order 1, given that the goal is simplicity. Recalling that outer order refers to the symmetrization of how channels are applied, as previously outlined in Eq. (16), this pseudocode can be simply generalized to higher outer orders by also requiring an additional subroutine that symmetrizes $\mathcal{T}$ and $\mathcal{Q}$, and adjusts their time-slices accordingly. We also remark that the classical function PARTITION(H) is presented here somewhat as a black box. This is done for generality, given that the user is free to use a method of their choosing. In Sec. V B, we present an explicit algorithm for this function, which requires a single additional input ω_c called the *chop threshold*, and simply involves a looping *if statement* that sorts Hamiltonian summands into A or B depending on whether their spectral norm is greater or less than ω_c.

IV. LOCAL COMPOSITE CHANNELS FOR REAL-TIME EVOLUTION

We introduce here an algorithm for *real-time evolution*, which utilizes the *local* structure of the Hamiltonian and is *composite*. It combines intuition from both the Trotter-Qdrift composite channels as discussed in Sec. II E and what we refer to as the Lieb-Robinson or block decomposition. This decomposition procedure is motivated by the Lieb-Robinson bound, first introduced in Sec. II G, and provides the means for constructing a product formula such that the propagators are supported on finite subsets of a lattice with exponentially small error (introduced in Ref. [34]). Our main result is Eq. (39), which bounds the error of the local composite channel. In this section we accomplish the following goals:

(i) We introduce product formulas motivated by the locality of H, summarizing the main result of Ref. [34], and explain how to unify this approach with composite product formulas into *local composite channels*.

(ii) We generalize the bound from Ref. [34] to a bound on quantum channels in Theorem IV.1, and then extend this to a diamond norm bound in Lemma IV.1.

(iii) We prove a diamond norm error bound on local composite channels in Theorem IV.2 using a simple property regarding channel composition proven in Lemma IV.2.

The basic idea behind the method is as follows. We decompose the Hamiltonian evolution into local blocks and build composite channels out of the operators supported on these local blocks. Each block is thus assigned a local partition $\{A, B\}$ and a sample number N_B. The local block decomposition goes as follows: We start with some Hamiltonian $H = \sum_X H_X$ defined on some lattice Γ with $X \subseteq \Gamma$ and by breaking up the lattice into two regions $A \cup B = \Gamma$. Now we take a smaller region of the lattice $Y \subseteq A$ and use the result to approximate the time evolution operator in the following way,

$$e^{-iHt} \approx e^{-iH_A t}(e^{-iH_Y t})^{\dagger} e^{-iH_{Y \cup B} t}. \quad (30)$$

Note that this breakup is distinct from Trotterization in that we are decomposing the operator into a product formula of which the operator exponentials still contain sums of operators. However, the operator sums now only contain terms in each local "block". We can recurse this process to make m blocks of near-equal size. As an illustration, the second round of decomposition would look like the following. Suppose the region of the lattice $A = G \cup K$, and we pick another small region $Z \in G$, then this further allows us to break down the larger region A from the original definition of the lattice set $\Gamma = A \cup B$. Writing out the full formula now gives

$$e^{-iHt} \approx e^{-iH_G t}(e^{-iH_Z t})^{\dagger} e^{-iH_{Z \cup K} t}(e^{-iH_Y t})^{\dagger} e^{-iH_{Y \cup B} t}, \quad (31)$$

and since the set $G = A \setminus K$, we have successfully further locally decomposed the Hamiltonian in this product formula. Next, we can decide which simulation method to employ to approximate and implement the block-evolution operators, which in our case is the composite product formula. Based on prior intuition, we expect this approach to perform well on disordered models yet with local interactions, such that there is a structure for the local composite channels to exploit. An example might be a Hamiltonian with a sharply peaked spectral norm on each block (as is the case with weak coupling). However, by adding Qdrift sampling to the local framework, with the speed limit v_{LR} in mind we are "fuzzing" the light cone in the quantum circuit evolution of the system. For example, in Fig. 2 we show a circuit that illustrates the Trotter time evolution of a local Hamiltonian where
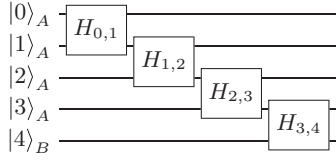
FIG. 2. Brickwork circuit diagram to highlight the emergent light cones from nearest-neighbor interactions. The notion of "fuzzing" the light cone occurs when we use Qdrift to sample terms along these light-cone boundaries in the circuit.

the $H_{i,j}$ terms only connect neighboring qubits within and between each disjoint subset A and B of the one-dimensional lattice.

The light-cone analogy is now made clear: A term $H_{0,1}$ influences the most adjacent qubit in B only once three "layers" of gates have been applied (intermittent gates cannot improve this speed). This behavior is analogous to the Lieb-Robinson velocity that upper bounds the propagation of information flow in the lattice. We note though that in a local composite channel, some of the gates are being randomly sampled in the Qdrift channel, thus the evolution of the light cone is not exactly reproduced since not all terms are necessarily implemented due to sampling. Keep in mind that local composite channel does include more complicated subsets of gates, some of which evolve boundary region terms like $H_{3,4}$ backwards in time; the circuit above is presented simply for conceptual purposes.

Towards bounding the error of a local composite channel, we build on the result from Ref. [34], which concerns with bounds on time evolution operators, and reformulate it as a bound on channels. The motivation for the reformulation of the bound into a channel norm is so that Qdrift could be incorporated into this algorithm. This would later allow us to bound the overall error. Since Qdrift produces mixed states, we need to work with the density matrix formalism.

Theorem IV.1. Given a time t and a local Hamiltonian $H = \sum_X H_X$ that generates a time-evolution unitary U and an evolutionary channel $\mathcal{U}$, $\exists\,(\mu > 0)\,|$ for any disjoint regions A, B, C we have the following operator norm bound:

$$||\mathcal{U}_{A\cup B}\mathcal{U}_B^\dagger\mathcal{U}_{B\cup C} - \mathcal{U}_{A\cup B\cup C}|| \in \mathcal{O}(e^{-\mu dist(A,C)}) \sum_{X:bd(AB,C)} ||H_X||,$$
$$(32)$$

where the evolutionary channels contain only Hamiltonians terms supported on their subscripted sets of the lattice, and X:bd(AB,C) indicates $X \subseteq A \cup B \cup C$ and $X \not\subseteq A \cup B$ and $X \not\subseteq C$.

Proof. We first restate the main result from Ref. [34] on time evolution operators,

$$||U_{A\cup B}U_B^\dagger U_{B\cup C} - U_{A\cup B\cup C}|| \in \mathcal{O}(e^{-\mu dist(A,C)}) \sum_{X:bd(AB,C)} ||H_X||.$$
$$(33)$$

Now, write $V := U_{A\cup B}U_B^\dagger U_{B\cup C}$ and $U := U_{A\cup B\cup C}$ and consider the following induced channel infinity norm

$$\max_{\rho:||\rho||_1 \leqslant 1} ||V\rho V^\dagger - U\rho U^\dagger||$$

$$= \max_{\rho:||\rho||_1 \leqslant 1} ||V\rho V^\dagger - U\rho V^\dagger + U\rho V^\dagger - U\rho U^\dagger||$$

$$= \max_{\rho:||\rho||_1 \leqslant 1} ||(V - U)\rho V^\dagger + U\rho(V^\dagger - U^\dagger)||$$

$$\leqslant \max_{\rho:||\rho||_1 \leqslant 1} ||(V - U)\rho V^\dagger|| + \max_{\rho:||\rho||_1 \leqslant 1} ||U\rho(V - U)^\dagger||$$

$$\leqslant 2||V - U||.$$
$$(34)$$

In the last line we used the fact that the infinity norm of a density matrix is upper bounded by 1, and the unitary invariance of the Schatten norms. Inserting the original definitions for V and U into $\max_{\rho:||\rho||_1 \leqslant 1} ||\mathcal{V} - \mathcal{U}|| \leqslant 2||V - U||$, where $\mathcal{V}$ is just $V\rho V^\dagger$, and similarly $\mathcal{U} = U\rho U^\dagger$, we successfully upper bounded the desired channel norm using the above result with the constant factor disappearing due to $\mathcal{O}$ notation. ∎

In the previous theorem (34), a bound was proven for only one round of decomposition, but note that when defining V, the proof holds regardless of how many unitary blockings the operator is composed of. In Ref. [34], it was stated that the error in Theorem IV.1 for D rounds of decomposition in a D-dimensional lattice is $\in \mathcal{O}(e^{-\mu l}DL^D/l)$ for a lattice with chain length L and blocks with overlap l. The overlap for the block in the example from Eq. (30) would be the diameter of the set Y, which can be written $diam(Y) = \max_{y,y' \in Y} dist(y, y')$; the latter distance is again the distance between lattice indices.

The bound (34) we have proven above is still not of the form of that given in Qdrift. The Qdrift bound utilizes the diamond distance, which is a completely bounded trace norm or Schatten 1-norm defined as $||\mathcal{V} - \mathcal{U}||_\diamond = \max_{\rho:||\rho||_1 \leqslant 1} ||(\mathcal{V} - \mathcal{U}) \otimes \mathbb{1}||_1$. While this is an upper bound on the infinity norm in general, it is not immediately clear how tight this bound may be as factors of dimensionality of the system may come into play, which we wish to avoid. This can be made clear by a simple example: Consider an $N \times N$ identity matrix $\mathbb{1}_N$. This matrix has $||\mathbb{1}_N||_1 = N$ while $||\mathbb{1}_N||_\infty = 1$, where N is a factor of dimensionality. Therefore, if possible, we would also like to convert this result (34) to a diamond norm. Building on Ref. [24], we accomplish this in the following lemma:

Lemma IV.1. Given a time t and a local Hamiltonian $H = \sum_X H_X$ that generates a time evolution unitary U and an evolutionary channel $\mathcal{U}$, $\exists\,(\mu > 0)\,|$ for any disjoint regions A, B, C we have the following diamond distance bound:

$$||\mathcal{U}_{A\cup B}\mathcal{U}_B^\dagger\mathcal{U}_{B\cup C} - \mathcal{U}_{A\cup B\cup C}||_\diamond$$

$$\in \mathcal{O}(e^{-\mu dist(A,C)}) \sum_{X:bd(AB,C)} ||H_X||. \quad (35)$$

Proof. We use Eqs. (11)–(17) from Ref. [24] replacing their Trotter channel with the local block channel $\mathcal{V} = \mathcal{U}_{A\cup B}\mathcal{U}_B^\dagger\mathcal{U}_{B\cup C}$ and set $\mathcal{U} = \mathcal{U}_{A\cup B\cup C}$ with V and U representing

the unitaries that induce each channel,

$$||\mathcal{U}(t) - \mathcal{V}(t)||_\diamond := ||(\mathcal{U}(t) - \mathcal{V}(t)) \otimes \mathbb{1}||_1$$

$$= \max_{\rho:||\rho||_1 \leqslant 1} ||U(t) \otimes \mathbb{1} \rho U(t)^\dagger \otimes \mathbb{1} - V(t) \otimes \mathbb{1} \rho V^\dagger(t) \otimes \mathbb{1}||_1$$

$$\leqslant \max_{\rho:||\rho||_1 \leqslant 1} ||U(t) \otimes \mathbb{1} \rho U(t)^\dagger \otimes \mathbb{1} - U(t) \otimes \mathbb{1} \rho V^\dagger(t) \otimes \mathbb{1}||_1$$

$$+ \max_{\rho:||\rho||_1 \leqslant 1} ||U(t) \otimes \mathbb{1} \rho V^\dagger(t) - V(t) \otimes \mathbb{1} \rho V^\dagger(t) \otimes \mathbb{1}||_1$$

$$= \max_{\rho:||\rho||_1 \leqslant 1} ||\rho(U(t) - V^\dagger(t)) \otimes \mathbb{1}||_1 + \max_{\rho:||\rho||_1 \leqslant 1} ||(U(t) - V(t)) \otimes \mathbb{1} \rho||_1$$

$$\leqslant 2||U(t) - V(t)||_\infty \max_{\rho:||\rho||_1 \leqslant 1} ||\rho||_1$$

$$= 2||U(t) - V(t)||_\infty. \tag{36}$$

To obtain the third line from the bottom we apply Eq. 1.175 from [50]. This can similarly be obtained using the submultiplicativity of the Schatten p norms and then applying Hölder's inequality. In either case, we have a term like $||U(t) \otimes \mathbb{1}||$, however, the Schatten infinity norm is independent of the tensor product, and is equal to 1 for any unitary. In the last line, we inserted the Theorem IV.1 to complete the proof. We observe the same bound on our Diamond distance as appeared in the induced infinity norm with the constant factor 2 once again absorbed into $\mathcal{O}$. ∎

Before investigating this channel numerically, we provide a bound on the diamond norm between this local composite channel and the *exact* evolutionary channel. Along with the fact that the diamond norm is subadditive and submultiplicative, we will make use of the following two properties:

$$||\mathcal{A}^{\circ r} - \mathcal{B}^{\circ r}||_\diamond \leqslant r||\mathcal{A} - \mathcal{B}||_\diamond, \quad ||\mathcal{A}\mathcal{B} - \mathcal{C}\mathcal{D}||_\diamond \leqslant ||\mathcal{A} - \mathcal{C}||_\diamond + ||\mathcal{B} - \mathcal{D}||_\diamond,$$

where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ represent channels.

These properties were shown in Ref. [50]. For our purposes, we require a more general property that holds for the composition of k channels $\mathcal{A}_i$. We provide a simple proof in the following lemma:

Lemma IV.2. Given a composition of k quantum channels $\mathcal{A}_i$, and another composition of k quantum channels $\widetilde{\mathcal{A}}_i$, the diamond norm of the difference between these two channels has the bound

$$\left|\left|\bigcirc_{i=1}^k \mathcal{A}_i - \bigcirc_{i=1}^k \mathcal{B}_i\right|\right|_\diamond \leqslant \sum_{i=1}^k ||\mathcal{A}_i - \mathcal{B}_i||_\diamond. \tag{37}$$

Proof. The proof is achieved by induction and it uses the fact that $||\mathcal{A}_i||_\diamond \leqslant 1$ for any quantum channel, which is due to the fact that the spectral norm is unitary invariant and density matrices have a maximum eigenvalue of 1. Starting from the left and suppressing the composition operation $\circ$ above, we have

$$||\mathcal{A}_k\mathcal{A}_{k-1}\ldots\mathcal{A}_1 - \mathcal{B}_k\mathcal{B}_{k-1}\ldots\mathcal{B}_1||_\diamond = ||\mathcal{A}_k\mathcal{A}_{k-1}\ldots\mathcal{A}_1 - \mathcal{B}_k\mathcal{A}_{k-1}\mathcal{A}_{k-2}\ldots\mathcal{A}_1 + \mathcal{B}_k\mathcal{A}_{k-1}\mathcal{A}_{k-2}\ldots\mathcal{A}_1 - \mathcal{B}_k\mathcal{B}_{k-1}\ldots\mathcal{B}_1||_\diamond$$

$$= ||(\mathcal{A}_k - \mathcal{B}_k)\mathcal{A}_{k-1}\ldots\mathcal{A}_1 + \mathcal{B}_k(\mathcal{A}_{k-1}\mathcal{A}_{k-2}\ldots\mathcal{A}_1 - \mathcal{B}_{k-1}\mathcal{B}_{k-2}\ldots\mathcal{B}_1)||_\diamond$$

$$\leqslant ||\mathcal{A}_k - \mathcal{B}_k||_\diamond||\mathcal{A}_{k-1}\ldots\mathcal{A}_1||_\diamond + ||\mathcal{B}_k||_\diamond||\mathcal{A}_{k-1}\mathcal{A}_{k-2}\ldots\mathcal{A}_1 - \mathcal{B}_{k-1}\mathcal{B}_{k-2}\ldots\mathcal{B}_1||_\diamond$$

$$\leqslant ||\mathcal{A}_k - \mathcal{B}_k||_\diamond + ||\mathcal{A}_{k-1}\mathcal{A}_{k-2}\ldots\mathcal{A}_1 - \mathcal{B}_{k-1}\mathcal{B}_{k-2}\ldots\mathcal{B}_1||_\diamond. \tag{38}$$

Repeating this procedure, we decompose the rightmost term into a sum of two channel differences by induction, allowing one to prove the original identity in Eq. (37). ∎

Next, we utilize Theorem IV.1, where we converted the spectral norm result on local block unitaries to a bound on a local decomposition channel. It is now made clear why this was done; we require a channel description to describe the Qdrift algorithm and relate their bounds. Using Lemma IV.2, we proceed with the initial goal of bounding the diamond norm between the local composite channel and the exact evolutionary channel. Here, for compactness we adopt the notation of writing blocked composite local channels as

$\mathcal{X}_{Y_i}$ and "exact" local channels (prior to hybridization into Trotter and Qdrift) $\mathcal{U}_{Y_i}$. Here, Y_i indicates a set of lattice points $Y_i \subseteq \Gamma$ such that each Y_i is constructed via the algorithm laid out earlier in this section. Therefore, only certain Hamiltonian interaction terms are supported on the lattice subsets Y_i. Additionally, let $\mathcal{U}_{Y_i}$ contain terms that have also undergone the appropriate conjugation (given that the intersections of some blocks are evolved backwards in time). Finally, let each set $Y_i \in Y$ such that $Y = \{Y_1, Y_2, \ldots, Y_m\}$ and m is the number of blockings constructed by the Lieb-Robinson decomposition. This provides convenient notation for the following theorem:

Theorem IV.2 (Local composite channels). Given a time t, local channel iterations r, and a local Hamiltonian H_Γ on a D-dimensional lattice Γ, let $\mathcal{X}_Y(t)$ be a local-composite channel with m local blocks Y_i, and partition $\{A, B\}_Y$ on each block, then the diamond norm of the difference between the local composite channel and the evolutionary channel has the following bound:

$$
\left|\left| \bigcirc_Y (\mathcal{X}_Y(t/r))^{\circ r} - \mathcal{U}(t) \right|\right|_\diamond
$$

$$
\leqslant m \max_Y \left|\left| (\mathcal{X}_Y(t/r))^{\circ r} - \mathcal{U}_Y(t) \right|\right|_\diamond + \epsilon_{LR}
$$

$$
= \frac{t^2 m}{2r} \max_Y \left(\left(\sum_{i,j} a_{Y_i} a_{Y_j} ||[A_{Y_i}, A_{Y_j}]|| + \sum_{i,j} a_{Y_i} b_{Y_j} ||[A_{Y_i}, B_{Y_j}]|| \right) + \frac{4\lambda_{B_Y}^2 t^2}{N_{B_Y} r} \right) + \mathcal{O}(e^{-\mu l} D L^D / l), \tag{39}
$$

where A_Y, B_Y are once again the terms of H_Y in Trotter and Qdrift with their respective spectral norms a_Y, b_Y, N_B is the number of Qdrift samples and λ is the sum of the spectral norms of the terms in Qdrift. It should also be stressed that r is a local channel iteration, which means that we do not iterate the channel that is a composition of blocks but rather each individual local composite channel built out of the block terms. The reason for this is that the error in the local decomposition is independent of time for time independent H with $t \in \mathcal{O}(1)$. We also require $m \in \mathcal{O}((L/l)^D)$ [34].

Proof. We begin with the norm we wish to bound, and then apply the above identities to format the norm such that known bounds from [24,34] can be applied,

$$
\left|\left| \bigcirc_{i=1}^m (\mathcal{X}_Y(t/r))^{\circ r} - \mathcal{U}(t) \right|\right|_\diamond
$$

$$
= \left|\left| \bigcirc_{i=1}^m (\mathcal{X}_{Y_i}(t/r))^{\circ r} - \bigcirc_{i=1}^m \mathcal{U}_{Y_i}(t) + \epsilon_{LR} \right|\right|_\diamond \tag{40}
$$

$$
\leqslant \sum_{i=1}^m \left|\left| (\mathcal{X}_{Y_i}(t/r))^{\circ r} - \mathcal{U}_{Y_i}(t) \right|\right|_\diamond + \epsilon_{LR} \tag{41}
$$

$$
\leqslant m \max_Y \left|\left| (\mathcal{X}_Y(t/r))^{\circ r} - \mathcal{U}_Y(t) \right|\right|_\diamond + \epsilon_{LR}. \tag{42}
$$

Here the Lieb-Robinson error ϵ_{LR} is used to represent the error incurred from block localizing the exact evolution channel in Eq. (40). Then by inserting the bound given in [24], Eq. (31), we arrive at the original expression. It then immediately follows that we can bound local composite channels of any order $2k$ by inserting the result of [24], Eq. (58); however, we omit this as it follows trivially from above. ∎

The takeaway from this is that the block-local composite channel is bounded by an error that is very similar to that of the general composite channel asides from the addition of an extra term that is asymptotically exponentially small. However, while we expect to see improvement from this localization scheme, getting more out of this bound does not seem feasible due to the amount of freedom in this algorithm as almost every parameter is dependent on the "worst" subset of the set Y.

V. NUMERICAL SIMULATIONS

In this section we highlight the details of configuring our numerical investigations by defining Hamiltonians of interest. We then discuss partitioning strategies and choice of an error measure before presenting numerical results. The choice of error measure usually does not require a dedicated subsection. However, when considering composite algorithms, there is a subtle feature in that not all algorithms may be treated on equal footing by the same error measure, which can lead to inconsistencies. For our purposes, the trace distance is the most useful measure of choice. Rather than computing upper bounds, the numerical results we provide here are exact calculations of the cost C, or the precise counts on the number of quantum gates with the form $e^{-i\frac{t}{r} \otimes_j \sigma_j^v}$ required to meet some desired precision ϵ. We write gates in this form to highlight that the Hamiltonian summands in our simulations are always written as a tensor product of Pauli operators, allowing for a nice parallel to the well-known rotation gates in the context of quantum computing. For imaginary time, although this form still holds in our numerics, the same analogy does not hold, as we are simply constructing $e^{-\frac{\beta}{r} H_j}$ on a classical computer, which can be of somewhat arbitrary form.

In order to calculate C, we have constructed a library to compile any desired product formula simulation, given a list of Hamiltonian terms, partition, Qdrift samples, simulation time, and desired precision. These simulator objects are handled by external functions that can partition the simulators, calculate errors and exact costs, or approximate simulation cost via Monte Carlo methods and more. The library is built on NumPy, but also contains conversion functions to load Hamiltonians generated by quantum chemistry packages OpenFermion [41] and PySCF [51] to simulate systems of interest. It also contains methods for geometrically local simulations to compute blockings using the required set logic.

A. Hamiltonians of interest

Within this section, we introduce the Hamiltonians in our numerical investigation of composite algorithms, as well as briefly outline methods used for their generation.

1. Electronic structure Hamiltonians in second quantization

The electronic structure problem is perhaps one of the most famous classically intractable problems that has vast applications in quantum chemistry. In order to write down these Hamiltonians, it is first necessary to introduce fermionic creation and annihilation operators. Fermions are particles with half-integer spin that obey Fermi-Dirac statistics, meaning they obey the following anticommutation relations:

$$
\{a_m, a_n^\dagger\} := a_m a_n^\dagger + a_n^\dagger a_m = \delta_{mn}, \tag{43}
$$

$$
\{a_m, a_n\} = \{a_m^\dagger, a_n^\dagger\} = 0. \tag{44}
$$

We work in Fock space where the subscripts of operators indicate the excitation number or the atomic orbital of an

electron. Molecular electronic structure Hamiltonians then take the form

$$H = \sum_{mn} h_{mn} a_m a_n^\dagger + \sum_{pqrs} h_{pqrs} a_p a_q a_r^\dagger a_s^\dagger. \qquad (45)$$

Here the first term represents single excitations and the second term keeps track of double excitations or "hopping" amongst orbitals. The coefficients h are molecular integrals that depend on the basis of choice to describe the molecule [44]. Further, more can be done to aid in the implementation of this problem on a quantum computer. The Jordan-Wigner transformation provides a one-to-one mapping between the fermionic and spin operators. This will allow us to write down the time-evolution operator in terms of the universal rotation gates (and CNOT gates). To understand the transformation, first observe

$$a^\dagger = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \frac{\sigma^x - i\sigma^y}{2} := \sigma^-, \qquad (46)$$

$$a = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{\sigma^x + i\sigma^y}{2} := \sigma^+. \qquad (47)$$

Now to build in the desired commutation relations and generalize this to a Hilbert space for N qubits, or the tensor (Kronecker) product of N two-dimensional Hilbert spaces $\mathscr{H} = \bigotimes_{i=1}^N \mathscr{H}_i$,

$$a_n^\dagger \Leftrightarrow \mathbb{1}^{\otimes n-1} \otimes \sigma^- \otimes (\sigma^z)^{\otimes N-n-1}, \qquad (48)$$

$$a_n \Leftrightarrow \mathbb{1}^{\otimes n-1} \otimes \sigma^+ \otimes (\sigma^z)^{\otimes N-n-1}. \qquad (49)$$

In order to build these Hamiltonians numerically, we use the OpenFermion [41] and PySCF [51] packages for quantum chemistry. OpenFermion is a library that allows for the easy manipulation of fermionic operators that arise in quantum chemistry, as well as it interfaces with a variety of electronic structure packages that perform molecular integrals in the basis of choice to generate Hamiltonians in the form of Eq. (45). Further, OpenFermion also has the Jordan Wigner transform built in, allowing one to construct this Hamiltonian in the Pauli basis. PySCF was our electronic structure package of choice to compute molecular integrals.

Given the form of Eq. (45), we observe that our Hilbert space needs to be truncated. An active space calculation does exactly this; the Hamiltonian is written in a space such that only so many orbitals are "active" or such that an electron can be excited to occupy active orbitals. We generate all of our electronic structure Hamiltonians in the minimal basis where we use a the number of qubits equal to the total period of the molecule. For our numerical investigation, we provide a function to generate chains of hydrogen atoms given a very simple input; bond length and number of atoms. The function uses PySCF to compute the molecular integrals, and then uses the data to build the Hamiltonian in an active space implied by the minimal basis, and using a minimal spin configuration.

2. Jellium uniform electron gas

Jellium is a model of a uniform electron gas that captures the interactions between delocalized electrons in a solid with uniformly distributed positive potentials serving as nuclei. It is not only a system of interest in materials science, but also as a benchmark system in quantum simulation. More compact

representations of this Hamiltonian have been proposed as a candidate for experimental simulation on near-term hardware [9]. The system Hamiltonian has a closed form representation and does not require any additional molecular integrals to construct,

$$H = \frac{1}{2} \sum_{p,\sigma} k_p^2 a_{p,\sigma}^\dagger a_{p,\sigma} - \frac{4\pi}{\Omega} \sum_{p \neq q, j, \sigma} \left(\zeta_j \frac{e^{ik_{q-p} \cdot R_j}}{k_{p-q}^2} \right) a_{p,\sigma}^\dagger a_{q,\sigma}$$

$$+ \frac{2\pi}{\Omega} \sum_{(p,\sigma) \neq (q,\sigma'), \nu \neq 0} \frac{a_{p,\sigma}^\dagger a_{q,\sigma'}^\dagger a_{q+\nu,\sigma'} a_{p-\nu,\sigma}}{k_\nu^2}, \qquad (50)$$

where the jth nuclei has position R_j and atomic number ζ_j, and $k_\nu = \frac{2\pi\nu}{\Omega^{1/3}}$ with cell volume Ω and σ containing both up and down spins. For the derivation of this Hamiltonian see Appendix B of [9]. Conveniently, OpenFermion also provides simple functions to quickly generate this Hamiltonian, and we do so in the momentum plane wave basis with periodic boundary conditions. We elect not to use the more compact plane wave dual basis representation presented in [9], due to the fact that we are using this Hamiltonian as a benchmark, rather than studying the outputs of the simulation. For the composite simulation, Jellium provides many Hamiltonian terms and a very sharply peaked distribution (see Fig. 3 for a system of size equal to that of the spin models we study. Given that system size is more of a limiting factor than term number in our numerical study, this presents an opportunity to see how a composite channel performs on a system with greater L. To limit the system size we also use a spinless model, and then perform the Jordan-Wigner transformation on the second quantized Hamiltonian to represent our Hamiltonian as a sum of Pauli operators. This Hamiltonian is constructed with the necessary transformations using OpenFermion [41].

3. Graph Hamiltonian model

The Hamiltonian we explore here involves a spin chain imposed on a lattice $\Gamma \in \mathbb{Z}^D$ with a graph distance metric $dist(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|_1$ where $\mathbf{u}$ and $\mathbf{v}$ are vectorized coordinates on the graph with dimension equal to D. For our investigation we only examine lattices with $D = 1$, given that for a fixed number of sites, this gives the most sharply peaked spectral distribution than any other D,

$$H = \sum_{i>j} e^{-dist(i,j)} \alpha_{ij} \sigma_i^x \sigma_j^x + \sum_k \beta_k \sigma_k^z. \qquad (51)$$

This system is similar to the quantum transverse field Ising model but with interactions that fall of exponentially with graph distance. The coefficients α_{ij} and β_k are site dependant coupling constants that allow for the introduction of more disorder and/or structure in the Hamiltonian. To add some disorder to the model, we draw these coefficients pseudorandomly from a Gaussian distribution with mean 0 and variance 1.

4. Heisenberg Model

The Heisenberg model describes a quantum spin system in a magnetic field with nearest-neighbour interactions. The
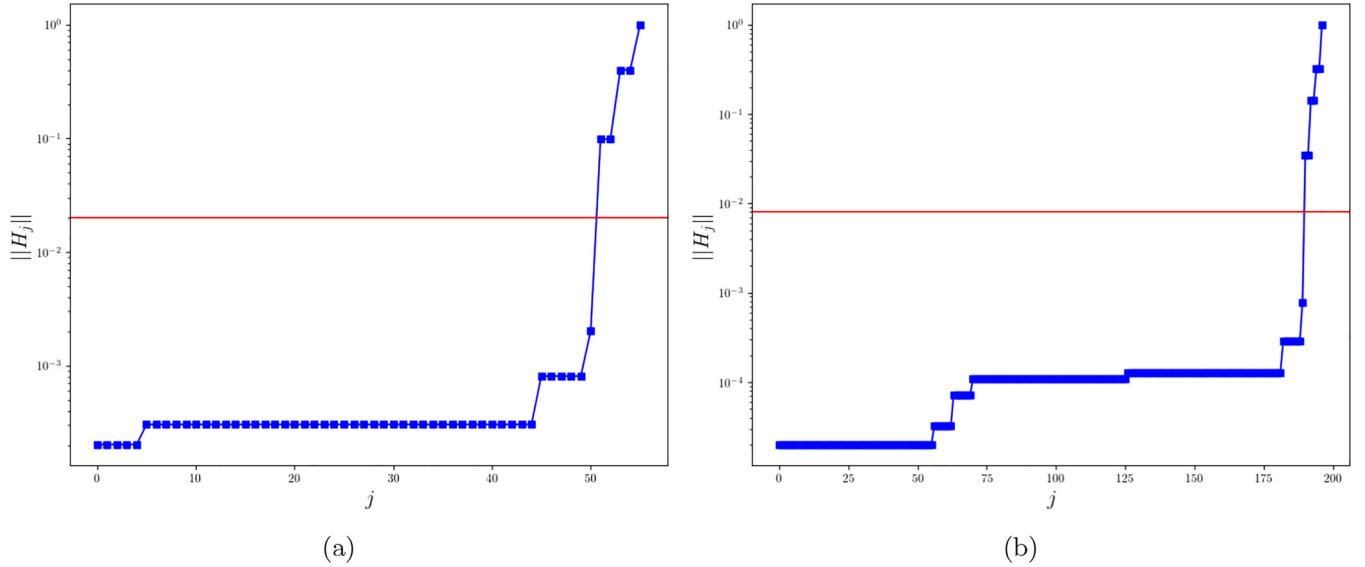
FIG. 3. Jellium spectral norm distribution. Semi-log plots of the sorted normalized spectral norms vs Hamiltonian index for 5 and 7 site Jellium models in (a) and (b) respectively. The plots show the increases in number of terms as well as how the distributions become increasingly sharply peaked. In red we provide a potential choice of ω_c for the partitioning heuristic.

Hamiltonian takes the form

$$H = \sum_j \left(J_x \sigma_j^x \sigma_{j+1}^x + J_y \sigma_j^y \sigma_{j+1}^y + J_z \sigma_j^z \sigma_{j+1}^z \right) + \sum_i B_z \sigma_i^z. \tag{52}$$

Here B_z is the strength of the magnetic field in the z direction and $J_{\{x,y,z\}}$ are coupling constants. Given the intuition of our composite channel, we expect this model to take advantage of our algorithm when the coupling constants largely differ in magnitude such that partitioning into Trotter and Qdrift takes advantage of more Hamiltonian structure. Furthermore, introducing site-dependant coupling constants or writing down a highly disordered spin system could further add structure that the algorithm can take advantage of. A Hamiltonian of this nature would look something like the following:

$$H = \sum_j \left(J_x^{(j)} \sigma_j^x \sigma_{j+1}^x + J_y^{(j)} \sigma_j^y \sigma_{j+1}^y + J_z^{(j)} \sigma_j^z \sigma_{j+1}^z \right)$$
$$+ \sum_i B_z \sigma_i^z. \tag{53}$$

This Hamiltonian appears somewhat contrived for extracting the performance of the local composite channel. However, this system closely resembles the Edwards-Anderson model of a spin glass, a system of interest in condensed matter physics. In an attempt to create a sharp distribution, we simply sample $J_v^{(j)}$ from an exponential distribution with a scale parameter of 0.1.

B. Partitioning schemes

The main difficulty with deploying the composite simulation framework concerns finding a good partitioning. In introducing composite simulations, Hagan and Wiebe suggested partitioning schemes derived on the basis of optimizing analytic cost functions both in deterministic and probabilistic settings [24]. Here, we take a different approach involving the exact calculation of the simulation error, and an optimization routine that, given convergence, finds the optimal partitioning and gate count with respect to a chosen error tolerance and simulation time. This approach is used to answer the question regarding the best savings one can hope to achieve when deploying composite methods to simulate a specific Hamiltonian. We are not proposing this as a preprocessing routine (for a study of this nature see Ref. [52]), as it has complexity greater than that of the simulation itself, which is trivial as the optimization involves solving the simulation problem recursively. In addition, we also arrive at simple heuristics that can be used to partition certain Hamiltonians with little overhead, which we do propose as a strategy for using a composite approach.

Chop is the partition that we introduce in this paper. The idea is based on the heuristic of placing a few terms with larger spectral norms into Trotter-Suzuki channels and numerous small terms into Qdrift, assuming that the Hamiltonian presents this structure. We start by sorting the terms by their spectral weights and introducing a "chop threshold" $\omega_c \in [0, \max_i h_i]$. This scale will determine the partition such that if a term has spectral norm $h_i \geqslant \omega_c$ then $H_i \to A$ if $h_i < \omega_c$ then $H_i \to B$. Now we can express the error tolerance ϵ as a function of channel iterations r, with partitioning chop ω_c and a sample number N_b that will be chosen in an optimization routine, and for a fixed initial state and time,

$$||(\mathcal{X}^{2k})^{\circ r}(\rho, t/r, \omega_c, N_b) - \mathcal{U}(\rho, t)||_1 = \epsilon(r). \tag{54}$$

By fixing an error tolerance for ϵ, the exact cost of the simulation becomes a black box function with no closed form expression,

$$C_{\text{comp}} = f(\epsilon_{\text{thresh}}, r, \omega_c, N_b). \tag{55}$$

This is the cost function we wish to minimize. However, we cannot do that by conventional methods such as with direct gradient descent. Also, with no strong intuition for a choice of N_B, if we wish to optimize this parameter we have to deal

with integer optimization as well. The iterations r, however, while an integer, does not require optimization, but rather emits a search problem. If we allow an optimizer to pick initial random values for N_b and ω_c from a fixed interval, then we must find the value r required to meet the error threshold ϵ_{thresh}, which will ultimately be determined by the optimizer's choice of the other two parameters. To complete this, we perform an exponential search on r until we find some r where $\epsilon(r) \geqslant \epsilon_{\text{thresh}}$ and set this as an upper bound on r. We then perform a binary search to find the smallest value of r required to meet this condition and count the number of gates in the channel. This is a very expensive function given that we are precisely building the composite channel, applying it to the density matrix initial state in the problem, and counting the gates applied in each iteration of the search. The expensive nature emerges due to the sheer number of matrix multiplications required in performing this task, not in the search for r, which is nearly optimal. Note the importance of using the trace distance in this approach as it guarantees monotonicity of $\epsilon(r)$, which makes the search possible. This is not so in the framework of sampling the quantum infidelity, as finding the cost here would require other statistical methods (see Appendix B).

Now a glaring question left unanswered is the choice of an optimizer. We implement the gradient-boosted regression trees (GBRT) algorithm included in Sci-Kit Optimize [53]. This algorithm is specifically-designed to handle the optimization of very expensive functions. It is also convenient for our purposes given that it can handle both integer and real optimization parameters simultaneously. At a high level, the algorithm works by using a series of decision trees with an associated loss function. The decision trees perform regression to fit the input function and are iteratively generated based on the minimization of the loss function via gradient descent. This optimizer and cost function (55) can then be easily generalized to the local composite channels where now we have an N_b and ω_c for each blocking. As the number of local blocks grows, the optimization routine will need to take a larger number of input parameters in this prescription. However, the size of the system becomes classically intractable long before we would consider using this many local blocks, so this is far from a concern.

In some cases, models may exhibit a partitioning that is somewhat canonical and can lead to excellent performance of composite methods. This occurs when we have a Hamiltonian that fits naturally into the intuition behind the algorithm, such that we have a set A containing large terms with small commutators and a set B with small terms that are known not to commute in general. We are, therefore, proposing to use the chop partition but by choosing the chop threshold ω_c based on physical intuition regarding the Hamiltonian, rather than some expensive optimization routine. A perfect example of such a system is a Heisenberg model with weak coupling. In this case, looking at Eq. (52), we would set the chop threshold $\omega_c = \max\{J_x, J_y, J_z\}$, which implies we simulate the interactions with Qdrift $\{J_\nu \sigma_j^\nu \sigma_{j+1}^\nu\}_{\nu=x,y,z} \rightarrow B$, and simulate the site energy terms with Trotter-Suzuki $\{B_z \sigma_j^z\} \rightarrow A$. In this way, the terms in the set A all commute with each other, whereas the terms in the set B are guaranteed to have a small spectral norm. We bring numerical evidence that this provides

computational advantages in the sections below. In general, any system with perturbative interactions may benefit from this framework, given that the commutators within the system are small, as they will avail this canonical partitioning. In cases where the partitioning is not as obvious, as is the case with H_3 and Jellium, we can achieve similar advantages by choosing $\omega_c = \max \frac{d||H_j||}{dj} \mathbf{s.t.}||B|| \geqslant ||A||$, meaning we sweep an ordered list of the Hamiltonian spectral norms and track the largest difference between terms, chopping the list where this occurs, given $j \geqslant \frac{L}{2}$. The final condition is just to ensure the majority of terms are simulated by Qdrift. We also use this strategy throughout Sec. V D and show advantages.

C. Error measures

In this investigation, there is some arbitrariness in the error measure one can choose in order to quantify the performance of a simulation channel. In order to evaluate the resources required by an algorithm, one must evaluate the number of gates required to meet a certain ϵ, which is calculated by said error measure. In the literature, this ϵ is often quantified by the diamond distance utilized in previous sections. However, while analytically convenient, for any reasonably-sized system, computing this quantity becomes computationally expensive. While it is possible to evaluate it efficiently, this requires finding the solution of a semidefinite program, which is much less efficient than using some other error measures. In addition, since we are not constrained to analytically solvable expressions or closed form equations with our numerical methods, we can optimize this cost in terms of some partition scheme. This is the idea behind the optimal chop partition, and doing so requires frequent computations of ϵ. With this in mind, the error of our algorithm should be a quantity that we can compute in a reasonable amount of time while also being a fair error measure. The criteria for "fairness" comes within the error measures treating each algorithm that comprises the composite channel on an equal footing. For example, if we are to optimize the partitioning with respect to the gate cost (which is dependent on ϵ), then if Trotter is more performant with respect to Qdrift in one error measure than in another, then our composite optimizer will favor Trotter, which will be reflected in the partition. As a result the total cost of the composite channel will be skewed by the error measure used.

We consider the infidelity and trace distance as possible measures of ϵ. For definitions and an additional discussion regarding the scaling and complexity of computing these quantities, see Appendix B. Analytics are required in order to answer the question of which measure might provide a more fair comparison. More specifically, we can ask the question of how the error measures will scale with respect to the total simulation time t, and then test these results numerically. In terms of the infidelity, we provide the following Theorems:

Proposition V.1 (Qdrift infidelity time scaling). Given a Qdrift channel $\mathcal{Q}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, {}^t/_N)$, for a density matrix ρ, time t/N, then the infidelity between the outputs of the channels $1 - \mathscr{F}(\mathcal{Q}(\rho, t), \mathcal{U}(\rho, t/N)) \in \mathcal{O}(t^2)$.

Proposition V.2 (Trotter-Suzuki infidelity time scaling). Given a Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, t)$, for a density matrix pure state

ρ and time t, then the infidelity between the outputs of the channels $1 - \mathscr{F}(\mathcal{T}^{2k}(\rho, t)\mathcal{U}(\rho, t)) \in \mathcal{O}((t^{2k+1})^2)$.

The proofs of these propositions are included in Appendix B. Here, we obtain the nontrivial result that the infidelity is squared when considering Trotter-Suzuki formulas. This would lead to our optimized chop algorithm to heavily favor this channel over Qdrift, and for this reason, we consider it an "unfair" error measure. On the other hand, if we consider how the trace distance scales with simulation time in both algorithms, we obtain the following Theorems:

Proposition V.3 (Qdrift trace distance time scaling). Given a Qdrift channel $\mathcal{Q}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, t/N)$, for an arbitrary density matrix ρ and time t/N, then the trace distance between the outputs of the channels $\mathscr{T}(\mathcal{Q}(\rho, t), \mathcal{U}(\rho, t/N)) \in \mathcal{O}(t^2)$.

Proposition V.4 (Trotter-Suzuki trace distance time scaling). Given a Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, t)$, for an arbitrary density matrix ρ and time t, then the trace distance between the outputs of the channels $\mathscr{T}(\mathcal{T}^{2k}(\rho, t), \mathcal{U}(\rho, t)) \in \mathcal{O}(t^{2k+1})$.

Here, we see that no such squaring occurs, and the expected time-scaling is obtained. For this reason, we compute the entire density matrix and ϵ using the trace distance in all of our numerical simulations. Proofs of the above theorems, as well as further discussions can again be found in Appendix B.

D. Performance results

In this section, we first numerically analyze the real-time quantum algorithm given by Hagan and Wiebe [24] and then show equivalent numerical calculations of the imaginary-time classical case. To accomplish this, we provide cost plots in which we provide the minimum C, or the number of rotation gates to achieve a desired simulation accuracy ϵ (calculated by the trace distance) for each point in time t or β. To reiterate, here we exactly compute entire evolution channels with a random initial state ρ sampled from the unit hyper-sphere, and directly apply and count gates. We conclude this section with a brief discussion about numerical studies for the local composite simulation algorithms. In these plots, we study variants of the composite channel and display results with the aforementioned notation with the addition of a tilde over the channel if the partition and N_B have been optimized with GBRT. For example, a composite channel with inner-order 2 and outer-order 1 with an optimized partition and number of Qdrift samples N_B is written like so $\widetilde{\mathcal{X}}^{2,1}$. Before presenting all of the results, for ease of reference, we find it useful to remind the reader of all relevant notations by summarizing them in Table I.

Throughout this section, we normalize $||H|| = 1$ and run simulations for times $t \in (0, \frac{3\pi}{2}]$ so as to ensure the system undergoes nontrivial dynamics without overlapping the phase. This is done due to the fact that Trotter formulas have a periodic error for $||H||t \geqslant 2\pi$, and running simulations in this range would lead to the optimizer finding the "good points", where the error happens to be small, which would provide a very low cost simulation and a sharp drop in the cost trend. We also report the cost advantages achieved on *crossover points*, which are values of $t = t'$ such that $C_{QD}(t') = C_{TS}(t')$. We

TABLE I. Summary of notation used in the numerical analysis, consistent with previous sections. The channels do not indicate whether we are working in real or imaginary time, however, that will be clear based on the subsections in which plots appear. This will also be the case for local channels simulating Hamiltonians defined on graphs.

Notation	Description
C	Algorithmic cost defined in Sec. II B
t	Total simulation time
β	Imaginary time/inverse temperature
r	Number of channel iterations/time-steps
N_B	Number of Qdrift samples
$\mathcal{T}^{2k}$	A Trotter-Suzuki channel of order $2k$
$\mathcal{Q}$	A Qdrift channel
$\mathcal{X}^{2k,2g}$	A composite channel with inner-order $2k$ and outer-order $2g$
$\widetilde{\mathcal{X}}^{2k,2g}$	A composite channel with partition and N_B optimized by the GBRT algorithm
$\mathcal{X}^{2k}_{l=j}$	A local composite channel with inner-order $2k$ and block length $l = j$

denote the composite channel *crossover advantage* as

$$\xi := C_{QD}(t')/C_{\mathcal{X}}(t') = C_{TS}(t')/C_{\mathcal{X}}(t'). \tag{56}$$

As we are unable to compute these times exactly we use interpolation methods to report values of ξ. This is motivated by the fact that analytics suggest this to be the point of greatest advantage for a composite channel [24]. This is intuitive, especially given higher-order Trotter-Suzuki formulas, which are known to asymptotically outperform Qdrift for large t, whilst Qdrift is dominant in the small t limit, suggesting a region where their strengths can be combined. Therefore, studying ξ is interesting in that it (given that our optimization scheme is convergent) represents the maximum advantage of using composite simulation algorithms. The value t' at which ξ appears may also suggest time scales for which composite algorithms are most advantageous. However, this analysis does not provide methods for finding t' or predicting ξ *a priori* as this is a very hard problem. The significance in t' here lies in the fact that we always observe it to fall within the numerical domain of $t||H|| \in (0, \frac{3\pi}{2}]$, which coincides with the domain suitable for quantum phase estimation, given we are implementing the real-time algorithm. For imaginary time, these values can be viewed as relevant energy-time scales in which these formulas bring a computational benefit, rather than some physically interesting regime. Summarized in Table II are our calculated crossover advantages.

1. Real-time composite quantum simulations

Beginning with hydrogen chains, our results are highlighted in Fig. 4. This plot reveals two interesting features: For long-time simulations, heuristics can be found that essentially match the performance of an optimized formula simply by inspecting the distribution of the norms of individual summands $||H_j||$, and with the optimized routine, we find a significant improvement at the crossover point with $\xi = 2.3$. These plots begin flat for most of the simulation channels, which for

TABLE II. Summary of gate cost improvements observed via the *crossover advantage* ξ defined in Eq. (56) (contingent on optimization convergence). We observe that savings tend to somewhat improve as the number of terms increases (within the same model), with the exception of Jellium 7 where optimizer struggles with partitioning due to the number of terms. This is evident in the lack of monotonicity of $C(\widetilde{\mathcal{X}}^1)$ in Fig. 6. The most significant savings are seen for the Jellium models. Even in cases where the number of terms are comparable to other models, larger advantages are persistent in Jellium. This further establishes the spectral norm distribution as one of the most important indicators of performance in the composite framework.

Hamiltonian	ξ	Num. terms	Time
Hydrogen-3	2.3	62	Real
5 site Jellium	9.2	56	Real
6 site Jellium	18.8	94	Real
7 site Jellium	10.4	197	Real
7 spin Graph	4.1	49	Real
8 spin Graph	3.9	64	Real
8 spin Heisenberg	3.1	29	Imag.
Hydrogen-3	2.3	62	Imag.
6 site Jellium	18.8	94	Imag.

the most part, indicate that one application of the channel achieves the desired ϵ for multiple sequential simulations at small times. This is expected, and is especially common with Trotter-Suzuki formulas, given that with one iteration they apply at least L gates depending on the order, while Qdrift provides the option of sampling single gates.

Figure 4 is interesting given that both the heuristic and optimized channels provide significant advantages at the crossover point, as well as the heuristic partition seems to match the asymptotic performance of the optimized channel. This demonstrates that optimization subroutines are not re-
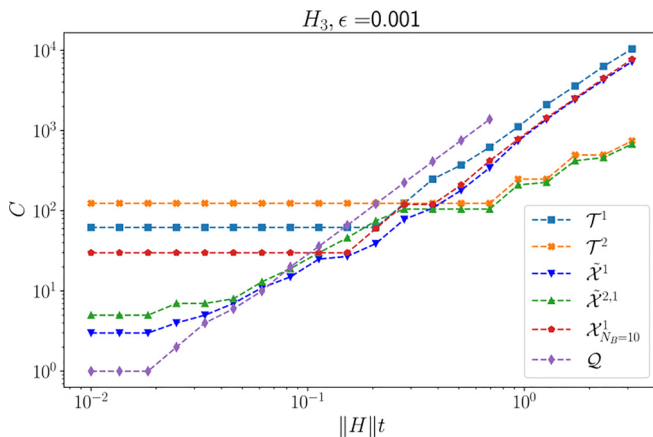

FIG. 4. H_3 cost plot simulation (real time). Log-log gate cost plot for the H_3 Hamiltonian generated with OpenFermion using three-dimensional Gaussians in a minimal basis. The bond distance is chosen to be that which minimizes the energy surface of H_2, which is ≈ 0.8 angstrom. We achieve a crossover advantage of $\xi = 2.3$, as well as remaining constant factor advantages at long times. For other plot notations see Table I.

quired to gain advantages in this framework. Furthermore, the second inner-order composite channel also shows a consistent advantage over the second-order Trotter channel.

To gain insight into effective choices of partition and Qdrift samples, we also plot the optimized N_B values and the ratio of Trotter terms to total terms $|A|/|H|$ with time in Fig. 5. Here, we find choices that somewhat agree with our prior intuition. For short times, places almost all terms into Qdrift, and slowly increases N_B. As t increases, more terms are placed into Trotter with the partition bouncing around in the regime where Trotter and Qdrift have similar costs, which is also expected. The composite channel $\widetilde{\mathcal{X}}^{2,1}$ essentially places all terms into the Trotter simulator, given the favorable asymptotic performance of higher-order Trotter formulas over Qdrift, while $\widetilde{\mathcal{X}}^1$ finds a balance between the two at long times, likely due to their equivalent t scaling. The most interesting behavior is that of N_B at mid to long times. Here, N_B peaks near the crossover point and then falls off as Trotter t scaling becomes dominant in $\widetilde{\mathcal{X}}^{2,1}$. However, for the $\widetilde{\mathcal{X}}^1$ channel, N_B experiences somewhat of a revival after the peak, and stabilizes at 15, which is about 24% of the terms. We use this percentage to motivate future heuristic choices of N_B in our investigation of Jellium in Fig. 6, which turns out to work quite well.

When it comes to the simulation of Jellium, we find some of the most significant performance improvements within this section, including an order of magnitude cost difference at the crossover point (see Fig. 6). Specifically, in the case of six-site Jellium, the Trotter and Qdrift cost at the crossover point is approximately 100 gates, versus the composite channel, which achieves the same precision ϵ with only about seven gates. Here, it is also shown that one can find an adequate partition that leads to advantages at longer times without the need for any optimization. This is also the only model whereby the optimization routine struggles to find optimal partitions in the neighbourhood of the crossover point. This leads to the Jellium 7 model inheriting a smaller ξ than what is likely achievable.

The final system we investigate in this section is that of the graph toy model with exponentially decaying interactions, which is a beyond nearest neighbor model. In Fig. 7, we study this model for chains of length 7 and 8, and find essentially identical behavior. When moving from 7 to 8 spins, we only add 15 more terms to the Hamiltonian, which is clearly not enough to see any significant advantages. In fact, the crossover advantage is slightly smaller for the bigger model, but this could also be due to the optimizer not fully converging.

2. Imaginary-time composite channels

This section contains cost plots of simulations of the same aforementioned Hamiltonians, but with imaginary-time propagators, where we present cost plots of the most interesting Hamiltonians from the previous section.

For simulations of the Heisenberg model, we find similar advantages to those in real time. In Fig. 8, we see that our proposed heuristic leads to an advantage over Trotter-Suzuki and Qdrift in the regions of interest. What is different about this plot is that the optimizer finds the same N_B and partitioning for all short times. This is an artefact of both the Hamiltonian and how the optimizer is programmed. Since all the splitting
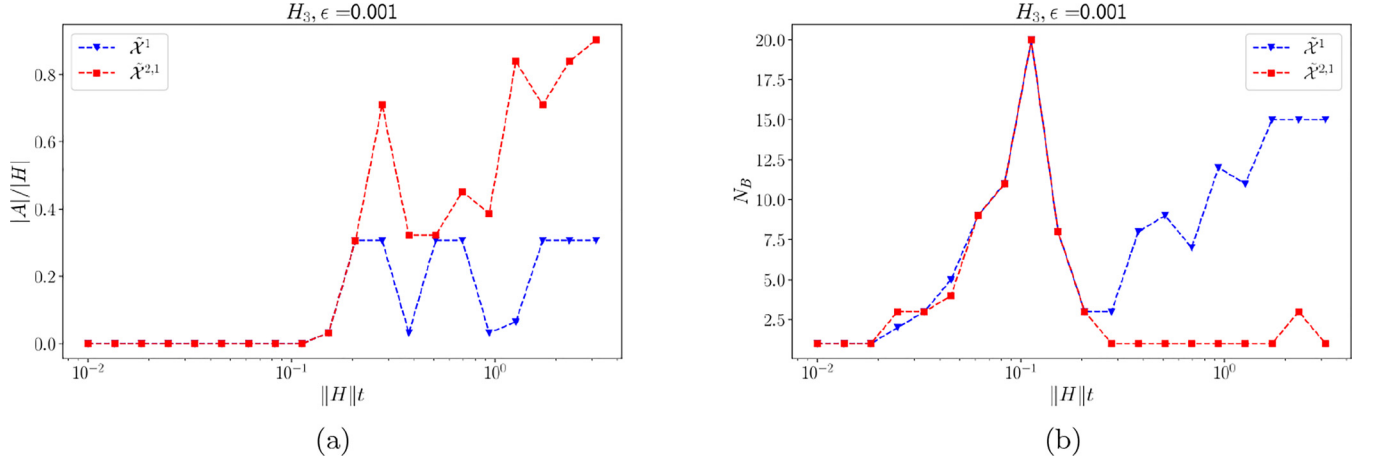
FIG. 5. Optimized H_3 simulation parameters. Semi-log plots of parameters obtained by the GBRT optimization routine for the $\widetilde{\mathcal{X}}^1$ and $\widetilde{\mathcal{X}}^{2,1}$ real-time channels. These parameter choices correspond to the H_3 simulation in Fig. 4. In (a) we plot the cardinality of the A set over the total number of terms, as a function of time. These values are dictated by GBRT optimized value of ω_c. In (b) we present the equivalent plot with N_B. For other plot notations see Table I.
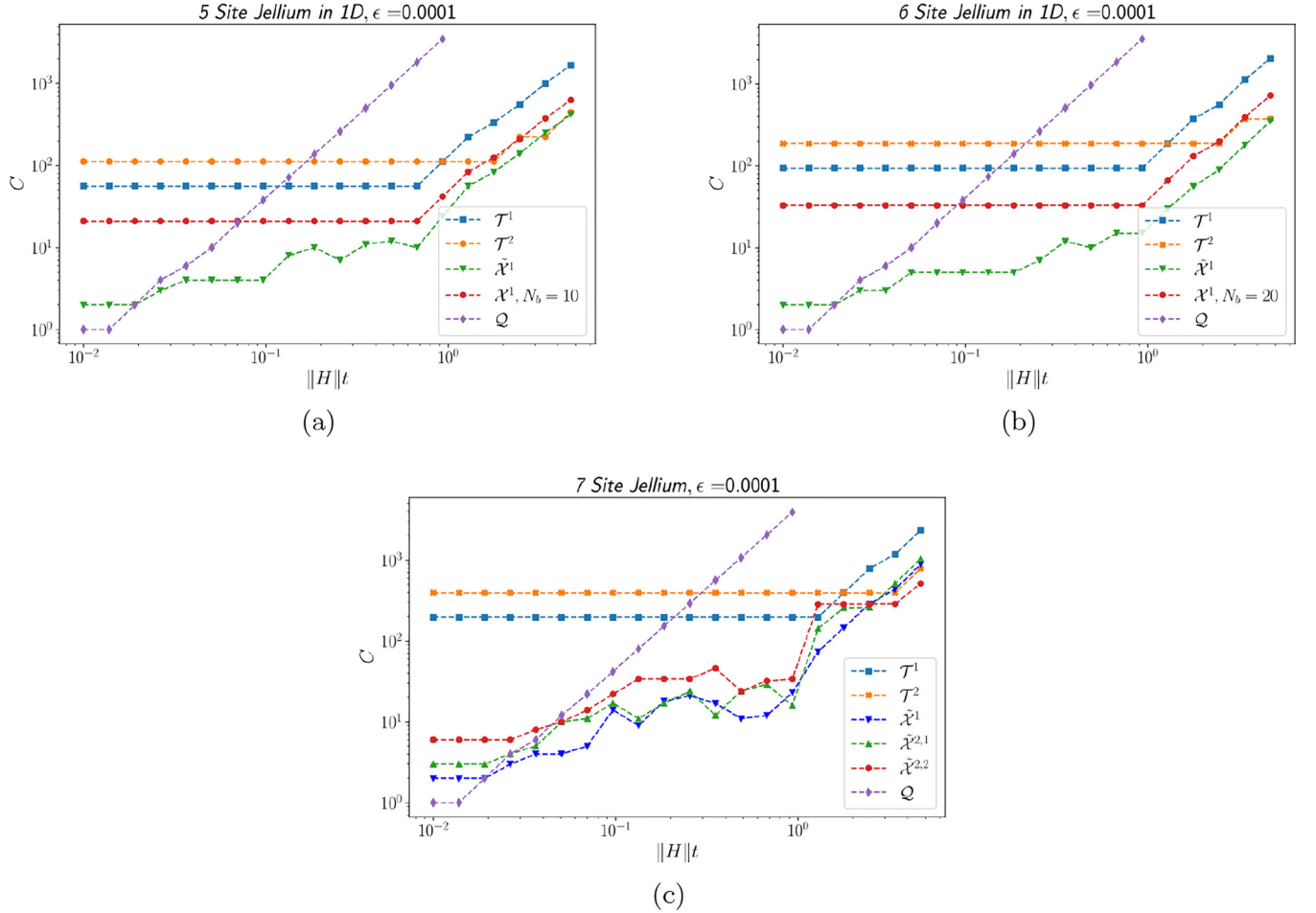
FIG. 6. Jellium simulation cost plots (real time). Log-log cost plots of quantum simulations of Jellium with five, six, and seven sites in (a), (b), and (c) respectively. In (a) and (b) we have in red, a chop heuristic where no optimization overhead is used. The distribution of Hamiltonian terms is chopped immediately before $\max \frac{d\|H_j\|}{dj}$, and approximately $\frac{1}{5}L$ terms are sampled. This heuristic works quite well, although it is outperformed by the optimized version, especially at short times. In (a) we achieve $\xi = 9.2$. In (b) we achieve an impressive advantage of $\xi = 18.8$, the largest of all our real-time results. In (c), we perform a similar analysis of the seven-site model with some higher-order composite channels, but find that the optimizer has increased difficulty with larger numbers of Hamiltonian terms. Here $\xi = 10.4$, but through inspecting some neighboring points of the crossover region, it surely has the potentially to be much larger. For other plot notations see Table I.
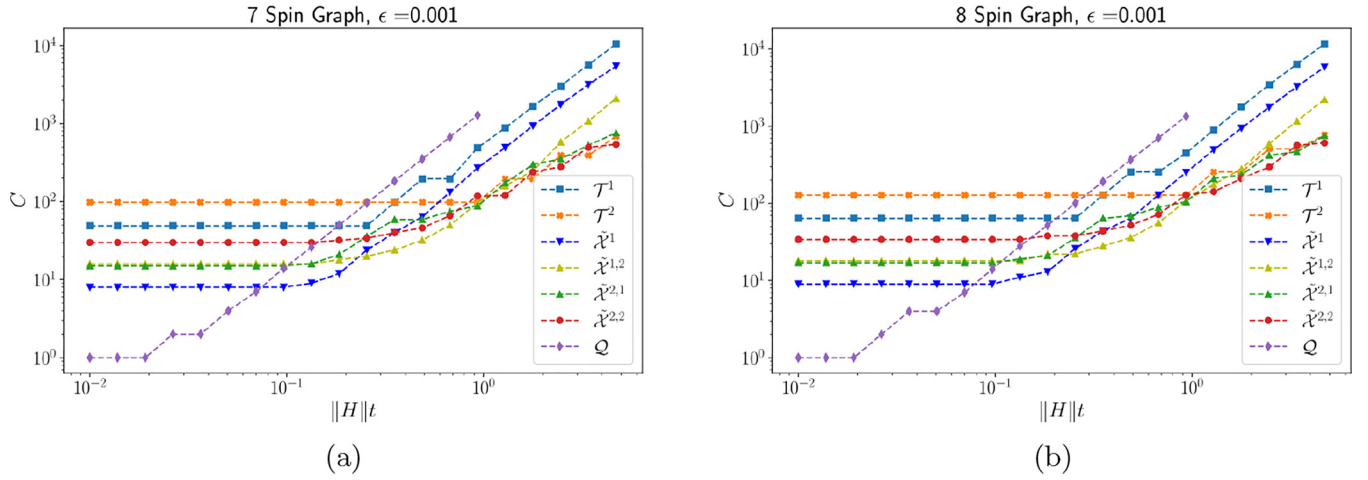
FIG. 7. Toy spin graph model cost plot. In (a) the we have a cost plot for the seven-spin model where we obtain a crossover advantage of $\xi = 4.1$, which is fairly significant. The plot has multiple regions where different composite channels are optimal. In (b) we have the eight-spin model where we establish $\xi = 3.9$. This advantage, as well as the channel performance is almost identical to (a). For other plot notations see Table I.

(single site) terms have equal spectral norm, the optimizer is placed in an all or nothing scenario, as choosing $\omega_c < 1$ immediately places all terms into Qdrift. Given that after receiving the cost, the program then solves a search problem to find the minimal r to achieve ϵ precision, it is rare that it finds the ideal conditions to build a pure Qdrift channel with $r = 1$. However, significant savings are still achieved at the crossover point, and over $\mathcal{T}^1$ and $\mathcal{T}^2$ at large β. Once again, the validity of heuristic partitions are shown, specifically in the first-order composite channel, which exactly matches its optimized version at large β.

For Hydrogen chains, we obtain strikingly similar results and compared with those in real time, as seen in Fig. 9. We once again obtain a significant crossover advantage, as well as constant factor advantages at large β, or low temperature. Heuristics are also shown to continue to hold in their effectiveness, in this case, from the crossover point and onward.

For Jellium, we choose to investigate the system size with the best-behaved optimizer, as well as the largest ξ, which occurs for the six-site model. In imaginary time, we once again reproduce a significant advantage, shown in Fig. 10. As in the case with H_3, this plot is quite similar to the real-time case in Fig. 6. However, here at large β the composite channel seems to do better in imaginary time given that even the first-order composite channel (with optimization) outperforms the second-order Trotter channel. This happens in the final point of the plot where $\mathcal{T}^2$ is no longer in the "flat-regime". While this is very interesting, it is unclear analytically why this occurs, and we would likely not expect this trend to continue asymptotically.

Overall, this section nicely complements some of the analytics in Sec. III both by reinforcing the fact that composite quantum channels allow for similar advantages in both real and imaginary time, as well as through calculation of exact
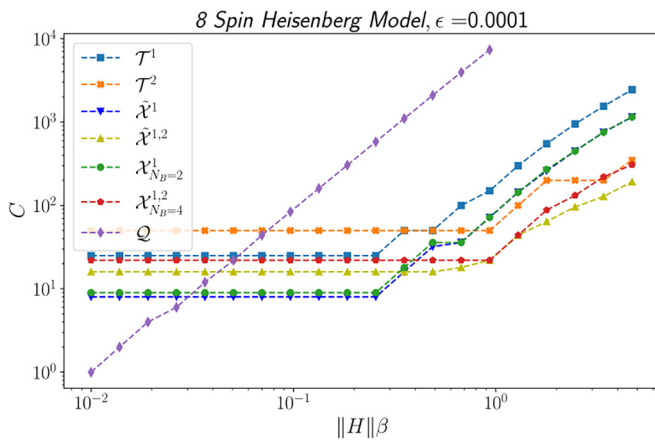

FIG. 8. Eight-spin Heisenberg model cost plot. In this imaginary-time simulation we establish $\xi = 3.1$, as well as maintain advantages at large β. We also find that our chosen heuristics are essentially optimal at large β, with the green and dark blue lines overlapping. For other plot notations see Table I.
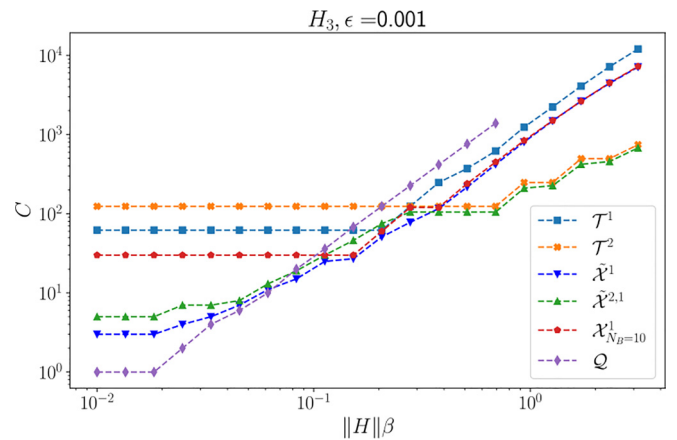
FIG. 9. H_3 simulation cost plot (imaginary time). The same parameters are used to build the three-atom Hydrogen chain as we done in real time. We achieve incredibly similar results, and again recover the real-time result of $\xi = 2.3$ that was achieved in Fig. 4. For other plot notations see Table I.
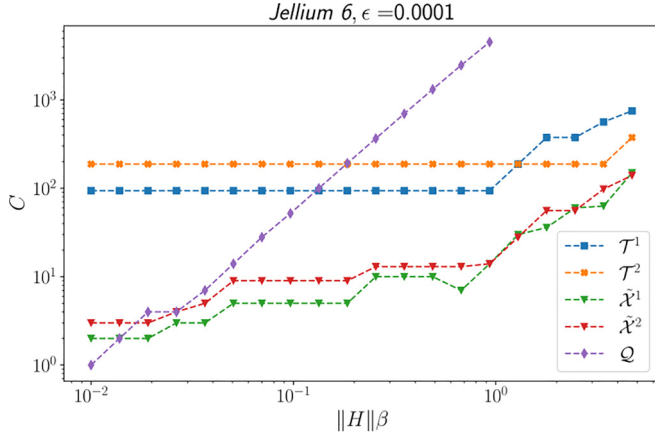
FIG. 10. Six-site Jellium simulation cost plot (imaginary time): Here we recover the crossover advantage of $\xi = 18.8$ from the real-time simulation in Fig. 6, which is also the largest advantage achieved in our imaginary-time simulations. We additionally achieve advantages over second-order Trotter at large β. For other plot notations see Table I.
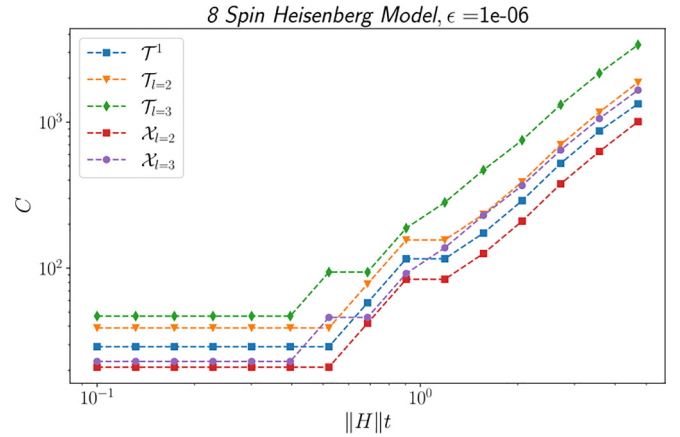

FIG. 11. Eight-spin Heisenberg model simulated by localized and standard channels: This cost plot compares Trotter to its localized versions, with the subscript l indicating the overlap of the boundary region in the block-local simulation. Channels with no l subscript are the standard simulations from prior sections. As expected, we observe that the composite channel is most efficient; however, given that the standard Trotter algorithm outperforms the localized Trotter channel, we can conclude that we are not in the regime where locality is providing advantages. The same heuristics were used for the composite channel as in Fig. 8, with $N_B = (4, 1, 4)$ on lattice subsets (A, Y, B). For other plot notations see Table I.

constant factor advantages. In other words, this section provides convincing evidence on the applicability of composite formulas to classical imaginary-time Monte Carlo algorithms.

3. Local composite quantum channels

Distinct from the previous two algorithms, with the local composite channels we do not immediately expect to see significant simulation advantages for the small systems we can compute. Recall that this algorithm makes use of the Lieb-Robinson velocity v_{LR} that limits the propagation of information and thus correlations in a local lattice model. In our numerics, our lattices contain $\leqslant 8$ sites, meaning even for small v_{LR}, the lattice can still become quickly entangled. In this section, it is important to understand where the observed advantages are originating, whether they are from the local decomposition, or something else. For example, the results in prior sections already suggest that composite channels can outperform Trotter and Qdrift channels in certain regimes. If a block decomposition is introduced, we pay a small gate cost to break the simulation into subsets (as gates on the boundary are applied more than once), but the advantages of the composite simulation are almost guaranteed to outweigh this cost. Thus we wish to find a regime in which we can perform calculations of exact costs with local composite channels that explicitly gain advantages via block decomposition. Otherwise, we will observe essentially the same behavior as before, but with slightly smaller constant factors. There are two ways to go about achieving this; one is to add more sites to the model, which quickly becomes computationally intractable with standard methods. The second strategy is to decrease the coupling between sites in the lattice, which naturally decreases v_{LR}. This is the strategy we utilize. In reference to Eq. (52) we perform our cost plots on Heisenberg models with eight sites with $B_z = 1$, and $J_v^{(j)}$ sampled from an exponential distribution with a scale parameter (serving as a coupling constant) of 0.00005. To allow for a fair simulation, we then choose $\epsilon = 0.000001$, such that statistically, 98% of terms will be

greater than ϵ, which can be seen from a simple integration of the PDF. Results of this simulation are shown in Fig. 11.

Given the gap between $\mathcal{T}^1$ and $\mathcal{T}_{l=2}$ in this very weak-coupling regime, it is unclear whether our methods (exact gate counts) provide the means for investigating advantages gained by locality. In Ref. [34], via computations of bounds, numerics did not show advantages (in the form of T gates) until approximately 100 sites were included in a more strongly-coupled model with $J_v^{(j)} \in [-1, 1]$ sampled i.i.d., so our results with far fewer spins are not unexpected. However, we still theoretically expect to see advantages in the limit where system sizes are large, and we can take advantage of the Lieb-Robinson bound.

VI. DISCUSSION AND SUMMARY

The main contributions of this paper are the extension of the composite channel to imaginary time and to local Hamiltonians, and the construction of a composite simulation library for numerically evaluating algorithmic performance. The imaginary-time bounds provided advocate for the potential use of Qdrift and composite Qdrift-product formulas in classical simulations where they can potentially improve the efficiency/accuracy of quantum Monte Carlo simulations. This is possible via shortening the path length of discrete time path integrals in these calculations given that our algorithms can approximate $e^{-\beta H}$ with fewer propagators than Trotter-Suzuki formulas in most cases. Our results also highlight that depending on the choice of β, one may also perform an imaginary-time simulation using Qdrift, particularly for high-temperature physics. It is an open question as to whether these methods can be used to improve the overall accuracy before one runs into the sign problem. Nevertheless, these

methods are interesting as one can use physical information about the Hamiltonian to choose a partitioning that possibly grants a computational advantage.

Furthermore, the extension of the composite channel to local Hamiltonians provides an interesting approach to lattice simulations in which there is *a priori* knowledge about the lattice structure. The diamond distance bound on this algorithm is significant in that the cost of establishing local composite blocks seemingly minimal in terms of asymptotics, making it a potentially useful algorithm. However, it seems more powerful numerical methods may be required to explore the regimes in which this approach shines, given the size of the systems likely required. An interesting future trade-off to be explored here is the size of the blocking l, and the size of the commutator sums in Theorem IV.2. Furthermore, one can investigate optimizations of partitions and Qdrift samples on lattice subsets. Despite our library containing these methods, we did not employ them as they can lead to misleading figures. Here, one must tread cautiously, as adding more blocks with parameterized weights and Qdrift samples can lead to an optimizer finding lower costs, similar to how adding hidden units to a neural network can lead to the over-fitting of a data set in machine learning. In this scenario, it may be possible to find localized composite channels that outperform the standard channels, simply because the increase in parameters allowed us to find a more tailored partitioning.

One of the difficulties with the techniques introduced in Ref. [24] is finding good parameters to yield cost improvements. Our numeric results seem to suggest that even heuristic approaches for parameters, such as N_B, can yield at least factors of 2 reductions for small Hamiltonians, as seen in Fig. 4. The difficulties of multiple parameter optimization becomes more of an issue when moving to the local simulation framework. With m blocks there are at least $2m$ parameters that need to be determined with a chop partitioning scheme. However, simple heuristics still lead to performance advantages over localized Trotter-Suzuki formulas in Fig. 11. The fact that said heuristics provide advantages with only an eight-spin model suggests that cost savings should scale favorably with increased lattice sizes.

In Sec. V C, a series of short but significant proofs were also provided to remind the reader to be cautious in examining error within hybridized algorithms. In our case, the infidelity treated Qdrift differently than Trotter by a square, and using it would have skewed the partitions and provided inaccurate costs on the composite channel.

Finally, the library built for this project is easy to implement. It is also flexible, and it reliably evaluates the exact number of operator of exponentials $e^{-iH_j t}$ required to execute on a quantum computer (or $e^{-iH_j \beta}$ for classical) to time evolve a system within a given error tolerance. To reiterate, the reason this is important is because the cost of these algorithms is highly dependent on the Hamiltonian, specifically in the number and size of the terms, and their commutators. Additionally, the costs and errors discussed in this paper are derived from analytic upper bounds on the diamond distance and induced Schatten $1 \rightarrow 1$ norm, which are often bounded with repeated applications of the triangle inequality and using max operations, and can therefore be loose. These exact numerics can capture the true costs and errors of these algorithms, and

advantages can motivate further studies into these approaches. This library is open source and available on GitHub [54].

APPENDIX A: PROOFS OF IMAGINARY-TIME BOUNDS

1. Imaginary-time Qdrift

Theorem III.1 (Imaginary-time Qdrift). Given a Hamiltonian $H = \sum_j h_j H_j$ with $\lambda = \sum_j h_j$, imaginary time β, number of samples N, and a density matrix initial state ρ, then the induced Schatten $1 \rightarrow 1$ norm of the difference between the imaginary-time Qdrift channel and exact imaginary-time evolutionary channel has the bound

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\text{Tr}\,\mathcal{U}(\rho, \beta)} - \left(\frac{\mathcal{Q}(\rho, \beta/N)}{\text{Tr}\,\mathcal{Q}(\rho, \beta/N)} \right)^{\circ N} \right\|_{1 \rightarrow 1}$$

$$\leqslant \frac{4\beta^2 \lambda^2}{N} \frac{e^{2\lambda\beta/N}}{2 - e^{2\beta\lambda/N}} + 2\frac{\frac{4\beta^4 \lambda^4}{N^3} e^{6\lambda\beta/N}}{1 - \frac{2\beta^2 \lambda^2}{N^2} e^{2\lambda\beta/N}} + 2\frac{2\beta^2 \lambda^2}{N} e^{4\lambda\beta/N},$$

$$\tag{23}$$

given that $|e^{2\beta\lambda/N} - 1| < 1$, which is satisfied if $N > \frac{2\beta\lambda}{\ln 2}$. Further, If the constraint $\frac{\lambda}{N} \leqslant 0.01$ is satisfied, then the bound simplifies to

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\text{Tr}\,\mathcal{U}(\rho, \beta)} - \left(\frac{\mathcal{Q}(\rho, \beta/N)}{\text{Tr}\,\mathcal{Q}(\rho, \beta/N)} \right)^{\circ N} \right\|_{1 \rightarrow 1} \leqslant \frac{C\beta^2 \lambda^2}{N}, \quad \tag{24}$$

where the constant $C \approx 29.71747$.

Proof. Following the Supplemental Material of Ref. [15], we will make use of the Liouvillian super-operator formalism for the imaginary-time channel. In imaginary time we have

$$\mathcal{L}(\rho) = \{H, \rho\} = H\rho + \rho H, \tag{A1}$$

which generates the following channel:

$$\frac{e^{-\beta \mathcal{L}(\rho)}}{\text{Tr}\,e^{-\beta \mathcal{L}(\rho)}} = \frac{e^{-\beta H} \rho e^{-\beta H}}{\text{Tr}(e^{-\beta H} \rho e^{-\beta H})}. \tag{A2}$$

Throughout this section, whenever $\mathcal{L}$ appears, it is implied that we mean $\mathcal{L}(\rho)$, we use the former for more compact notation. The difference between $\mathcal{L}$ defined here and that in

[15] is that the commutator becomes an anticommutator, and the i vanishes due to the nature of the imaginary-time channel. Equivalence can be seen via the Taylor series expansion of $e^{-\beta\mathcal{L}(\rho)}$, where $\mathcal{L}^n$ means n applications of the super-operator: $\{H, \{H, \ldots, \{H, \{H, \rho\}\} \ldots\}\}$. From [15], this super-operator yields the following properties:

$$||\mathcal{L}||_{1\to1} \leqslant 2||H|| \leqslant 2\lambda. \tag{A3}$$

Similarly, if we define $\mathcal{L}_j$ that generate unitaries under each H_j as in [15], and given that the Hamiltonian terms are normalized by h_j the following also holds:

$$||\mathcal{L}_j||_{1\to1} \leqslant 2||H_j|| = 2. \tag{A4}$$

We are now ready to proceed with the proof. To avoid cumbersome notation we will write $\mathcal{L}(\rho)$ simply as $\mathcal{L}$. Let us first examine the induced 1-norm of the difference between exact channel $\mathcal{U}_N$ that acts for imaginary time β/N and the Qdrift channel $\mathcal{Q}$, and the generalize this for N compositions of the Qdrift channel representing N samples,

$$\left|\left|\frac{\mathcal{U}_N}{\text{Tr}\mathcal{U}_N} - \frac{\mathcal{Q}}{\text{Tr}\mathcal{Q}}\right|\right|_{1\to1} = \left|\left|\frac{e^{-\beta/N\mathcal{L}}}{\text{Tr}e^{-\beta/N\mathcal{L}}} - \frac{\sum_i \frac{h_i}{\lambda}e^{-\beta\lambda/N\mathcal{L}_i}}{\text{Tr}\left(\sum_j \frac{h_j}{\lambda}e^{-\beta\lambda/N\mathcal{L}_j}\right)}\right|\right|_{1\to1} \tag{A5}$$

$$= \left|\left|\frac{e^{-\beta/N\mathcal{L}}}{1 + \text{Tr}\left(\sum_{n=1}^{\infty}\frac{(-\beta\mathcal{L})^n}{N^n n!}\right)} - \frac{\sum_i \frac{h_i}{\lambda}e^{-\beta\lambda/N\mathcal{L}_i}}{1 + \text{Tr}\left(\sum_j \frac{h_j}{\lambda}\sum_{m=1}^{\infty}\frac{(-\beta\mathcal{L}_j)^m}{N^m m!}\right)}\right|\right|_{1\to1} \tag{A6}$$

$$\leqslant \left|\left|\frac{e^{-\beta/N\mathcal{L}}}{1 + \text{Tr}\left(\sum_{n=1}^{\infty}\frac{(-\beta\mathcal{L})^n}{N^n n!}\right)} - \frac{\sum_i \frac{h_i}{\lambda}e^{-\beta\lambda/N\mathcal{L}_i}}{1 + \text{Tr}\left(\sum_{n=1}^{\infty}\frac{(-\beta\mathcal{L})^n}{N^n n!}\right)}\right|\right|_{1\to1}$$

$$+ \left|\left|\frac{\sum_i \frac{h_i}{\lambda}e^{-\beta\lambda/N\mathcal{L}_i}}{1 + \text{Tr}\left(\sum_{n=1}^{\infty}\frac{(-\beta\mathcal{L})^n}{N^n n!}\right)} - \frac{\sum_i \frac{h_i}{\lambda}e^{-\beta\lambda/N\mathcal{L}_i}}{1 + \text{Tr}\left(\sum_j \frac{h_j}{\lambda}\sum_{m=1}^{\infty}\frac{(-\beta\mathcal{L}_j)^m}{N^m m!}\right)}\right|\right|_{1\to1} \tag{A7}$$

$$:= a + b. \tag{A8}$$

Now we shall proceed by dealing with each of these terms individually. The denominator of the a can be expanded via its geometric series,

$$a = \left|\left|\left(\sum_{k=0}^{\infty}\left(-\text{Tr}\left(\sum_{n=1}^{\infty}\frac{(-\beta\mathcal{L})^n}{N^n n!}\right)\right)^k\right)\left(e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\mathcal{L}_i/N}\right)\right|\right|_{1\to1} \tag{A9}$$

$$= \left|\sum_{k=0}^{\infty}\left(-\text{Tr}\left(\sum_{n=1}^{\infty}\frac{(-\beta\mathcal{L})^n}{N^n n!}\right)\right)^k\right|\left|\left|e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\lambda\mathcal{L}_i/N}\right|\right|_{1\to1}. \tag{A10}$$

We will first focus on the leftmost factor. To bound this term we will apply the triangle inequality to apply the norm to each term in the sum. We will also use the fact that the absolute value of the trace is trivially upper bounded by the induced 1-norm: $|\text{Tr}\mathcal{L}(\rho)| \leqslant ||\mathcal{L}||_{1\to1}$. Here it is understood that this absolute value is maximized over all $\rho : ||\rho|| = 1$. We then use the submultiplicative property $||\mathcal{L}^n||_{1\to1} \leqslant ||\mathcal{L}||_{1\to1}^n$,

$$a \leqslant \sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty}\frac{\beta^n|\text{Tr}(\mathcal{L})^n|}{N^n n!}\right)^k\left|\left|e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\lambda\mathcal{L}_i/N}\right|\right|_{1\to1} \tag{A11}$$

$$\leqslant \sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty}\frac{\beta^n||\mathcal{L}||_{1\to1}^n}{N^n n!}\right)^k\left|\left|e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\lambda\mathcal{L}_i/N}\right|\right|_{1\to1} \tag{A12}$$

$$\leqslant \sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty}\frac{\beta^n 2^n \lambda^n}{N^n n!}\right)^k\left|\left|e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\lambda\mathcal{L}_i/N}\right|\right|_{1\to1} \tag{A13}$$

$$= \sum_{k=0}^{\infty}(e^{2\lambda\beta/N} - 1)^k\left|\left|e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\lambda\mathcal{L}_i/N}\right|\right|_{1\to1} \tag{A14}$$

$$= \frac{||e^{-\beta\mathcal{L}/N} - \sum_i \frac{h_i}{\lambda}e^{-\beta\lambda\mathcal{L}_i/N}||_{1\to1}}{2 - e^{2\beta\lambda/N}}. \tag{A15}$$

Where we require $|e^{2\beta\lambda/N} - 1| < 1$. Next, inspecting the numerator, this is of the form bounded in the Appendix of [15], therefore, the following holds trivially:

$$a \leqslant \frac{4\beta^2\lambda^2}{N^2} \frac{e^{2\lambda\beta/N}}{2 - e^{2\beta\lambda/N}}. \tag{A16}$$

Next, let us proceed to bound the *b* term

$$b = \left\| \frac{\sum_i \frac{h_i}{\lambda} e^{-\beta\mathcal{L}_i}}{1 + \mathrm{Tr}\left(\sum_{n=1}^{\infty} \frac{(-\beta\mathcal{L})^n}{n!}\right)} - \frac{\sum_i \frac{h_i}{\lambda} e^{-\beta\mathcal{L}_i}}{1 + \mathrm{Tr}\left(\sum_j \frac{h_j}{\lambda} \sum_{m=1}^{\infty} \frac{(-\beta\mathcal{L}_j)^m}{m!}\right)} \right\|_{1\to 1} \tag{A17}$$

$$= \left\| \sum_i \frac{h_i}{\lambda} e^{-\beta\mathcal{L}_i} \right\|_{1\to 1} \left| \sum_{k=0}^{\infty} \left(-\mathrm{Tr}\left(\sum_{n=1}^{\infty} \frac{(-\beta\mathcal{L})^n}{n!}\right)\right)^k - \sum_{l=0}^{\infty} \left(-\mathrm{Tr}\left(\sum_j \frac{h_j}{\lambda} \sum_{n=1}^{\infty} \frac{(-\beta\mathcal{L}_j)^n}{n!}\right)\right)^l \right| \tag{A18}$$

$$\leqslant e^{2\beta\lambda/N} \left| \sum_{k=0}^{\infty} \left(-\mathrm{Tr}\left(\sum_{n=1}^{\infty} \frac{(-\beta\mathcal{L})^n}{n!}\right)\right)^k - \sum_{l=0}^{\infty} \left(-\mathrm{Tr}\left(\sum_j \frac{h_j}{\lambda} \sum_{n=1}^{\infty} \frac{(-\beta\mathcal{L}_j)^n}{n!}\right)\right)^l \right|. \tag{A19}$$

Now, carefully expanding the double sums up to second order,

$$= e^{2\beta\lambda/N} \left| \sum_{k=2}^{\infty} \left(-\mathrm{Tr}\left(\sum_{n=2}^{\infty} \frac{(-\beta\mathcal{L})^n}{n!}\right)\right)^k - \sum_{l=2}^{\infty} \left(-\mathrm{Tr}\left(\sum_j \frac{h_j}{\lambda} \sum_{n=2}^{\infty} \frac{(-\beta\mathcal{L}_j)^n}{n!}\right)\right)^l + \mathrm{Tr}\sum_j \frac{h_j}{\lambda} \sum_{m=2}^{\infty} \frac{(-\beta\lambda\mathcal{L}_j)^m}{N^m m!} \right.$$

$$\left. - \mathrm{Tr}\sum_{n=2}^{\infty} \frac{(-\beta\lambda\mathcal{L})^n}{N^n m!} + \mathrm{Tr}\frac{\beta\mathcal{L}}{N} - \mathrm{Tr}\sum_j \frac{h_j}{\lambda} \frac{\beta\lambda\mathcal{L}_j}{N} \right|. \tag{A20}$$

Using the fact that $\sum_j h_j \mathcal{L}_j = \mathcal{L}$, the final two terms cancel and we have

$$= e^{2\beta\lambda/N} \left| \sum_{k=2}^{\infty} \left(-\mathrm{Tr}\left(\sum_{n=2}^{\infty} \frac{(-\beta\mathcal{L})^n}{N^n n!}\right)\right)^k - \sum_{l=2}^{\infty} \left(-\mathrm{Tr}\left(\sum_j \frac{h_j}{\lambda} \sum_{n=2}^{\infty} \frac{(-\beta\mathcal{L}_j)^n}{N^n n!}\right)\right)^l \right.$$

$$\left. + \mathrm{Tr}\sum_j \frac{h_j}{\lambda} \sum_{m=2}^{\infty} \frac{(-\beta\lambda\mathcal{L}_j)^m}{N^m m!} - \mathrm{Tr}\sum_{n=2}^{\infty} \frac{(-\beta\lambda\mathcal{L})^n}{N^n m!} \right| \tag{A21}$$

$$\leqslant e^{2\beta\lambda/N} \left(\left| \sum_{k=2}^{\infty} \left(-\mathrm{Tr}\left(\sum_{n=2}^{\infty} \frac{(-\beta\mathcal{L})^n}{N^n n!}\right)\right)^k \right| + \left| \sum_{l=2}^{\infty} \left(-\mathrm{Tr}\left(\sum_j \frac{h_j}{\lambda} \sum_{n=2}^{\infty} \frac{(-\beta\mathcal{L}_j)^n}{N^n n!}\right)\right)^l \right| \right.$$

$$\left. + \left| \mathrm{Tr}\sum_j \frac{h_j}{\lambda} \sum_{m=2}^{\infty} \frac{(-\beta\lambda\mathcal{L}_j)^m}{N^m m!} \right| + \left| \mathrm{Tr}\sum_{n=2}^{\infty} \frac{(-\beta\lambda\mathcal{L})^n}{N^n m!} \right| \right) \tag{A22}$$

$$\leqslant e^{2\beta\lambda/N} \left(\sum_{k=2}^{\infty} \left(\sum_{n=2}^{\infty} \frac{\beta^n |\mathrm{Tr}\mathcal{L}^n|}{N^n n!}\right)^k + \sum_{l=2}^{\infty} \left(\sum_j \frac{h_j}{\lambda} \sum_{n=2}^{\infty} \frac{\beta^n |\mathrm{Tr}\mathcal{L}^n|}{N^n n!}\right)^l + \sum_j \frac{h_j}{\lambda} \sum_{m=2}^{\infty} \frac{\beta^m \lambda^m |\mathrm{Tr}\mathcal{L}_j^n|}{N^m m!} + \sum_{n=2}^{\infty} \frac{\beta^n \lambda^n |\mathrm{Tr}\mathcal{L}^n|}{N^n m!} \right). \tag{A23}$$

Now we once again use the property $|\mathrm{Tr}\mathcal{L}(\rho)| \leqslant \|\mathcal{L}(\rho)\|_{1\to 1}$,

$$\leqslant e^{2\beta\lambda/N} \left(\sum_{k=2}^{\infty} \left(\sum_{n=2}^{\infty} \frac{\beta^n \|\mathcal{L}\|_{1\to 1}^n}{N^n n!}\right)^k + \sum_{l=2}^{\infty} \left(\sum_j \frac{h_j}{\lambda} \sum_{n=2}^{\infty} \frac{\beta^n \|\mathcal{L}_j\|_{1\to 1}^n}{N^n n!}\right)^l + \sum_j \frac{h_j}{\lambda} \sum_{m=2}^{\infty} \frac{\beta^m \lambda^m \|\mathcal{L}_j\|_{1\to 1}^m}{N^m m!} + \sum_{n=2}^{\infty} \frac{\beta^n \lambda^n \|\mathcal{L}\|_{1\to 1}^n}{N^n m!} \right) \tag{A24}$$

$$\leqslant e^{2\beta\lambda/N}\left(\sum_{k=2}^{\infty}\left(\sum_{n=2}^{\infty}\frac{2^n\beta^n\lambda^n}{N^n n!}\right)^k + \sum_{l=2}^{\infty}\left(\sum_j\frac{h_j}{\lambda}\sum_{n=2}^{\infty}\frac{2^n\beta^n\lambda^n}{N^n n!}\right)^l + \sum_j\frac{h_j}{\lambda}\sum_{m=2}^{\infty}\frac{2^m\beta^m\lambda^m}{N^m m!} + \sum_{n=2}^{\infty}\frac{2^n\beta^n\lambda^n}{N^n m!}\right) \tag{A25}$$

$$\leqslant e^{2\beta\lambda/N}\left(2\sum_{k=2}^{\infty}\left(\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}\right)^k + 2\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}\right) \tag{A26}$$

$$= 2\frac{\frac{4\beta^4\lambda^4}{N^4}e^{6\lambda\beta/N}}{1-\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}} + 2\frac{2\beta^2\lambda^2}{N^2}e^{4\lambda\beta/N}. \tag{A27}$$

Where in the last three lines we used both the tail bound on an exponential sum; $\sum_{n=2}^{\infty}\frac{x^n}{n!} \leqslant \frac{x^2}{2}e^x$ and the tail-sum of the geometric series; $\sum_{n=2}^{\infty} x^n = \frac{x^2}{1-x}$. The latter of the two requires that $|x| < 1$ for convergence, so we therefore, require $|\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}| < 1$, which is not a strong requirement given we are interested in large N. Now combining the results for a and b we obtain the final result for a channel that simulates a fraction $1/N$ of the dynamics,

$$||\mathcal{U}_N - \Lambda_{QD}||_{1\rightarrow 1} \leqslant a + b \tag{A28}$$

$$\leqslant \frac{4\beta^2\lambda^2}{N^2}\frac{e^{2\lambda\beta/N}}{2-e^{2\beta\lambda/N}} + 2\frac{\frac{4\beta^4\lambda^4}{N^4}e^{6\lambda\beta/N}}{1-\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}} + 2\frac{2\beta^2\lambda^2}{N^2}e^{4\lambda\beta/N}. \tag{A29}$$

Using the fact that the norm is subadditive under composition

$$\left|\left|\mathcal{U}_N^{\circ N} - \Lambda_{QD}^{\circ N}\right|\right|_{1\rightarrow 1} \leqslant N||\mathcal{U}_N - \Lambda_{QD}||_{1\rightarrow 1}, \tag{A30}$$

we are left with the following result for the full-time channel:

$$\left|\left|\mathcal{U} - \Lambda_{QD}^{\circ N}\right|\right|_{1\rightarrow 1} \leqslant \frac{4\beta^2\lambda^2}{N}\frac{e^{2\lambda\beta/N}}{2-e^{2\beta\lambda/N}} + 2\frac{\frac{4\beta^4\lambda^4}{N^3}e^{6\lambda\beta/N}}{1-\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}} + 2\frac{2\beta^2\lambda^2}{N}e^{4\lambda\beta/N}. \tag{A31}$$

Next, we wish to linearize this bound for a clearer interpretation of the error. We can do so by utilizing the conditions required to hold in the proof

$$|e^{2\beta\lambda/N} - 1| < 1, \tag{A32}$$

$$\left|\frac{2\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}\right| < 1. \tag{A33}$$

The first condition (A32) gives $e^{2\beta\lambda/N} \in (0, 2)$, which implies $\frac{2\beta\lambda}{N} \in (-\infty, \ln 2)$, and given that we require this quantity necessarily be positive we have $\frac{2\beta\lambda}{N} \in [0, \ln 2)$. Now, our second condition (A33), given the positivity constraint, further implies $\frac{x^2}{2}e^x \in [0, 1)$ with $x = 2\beta\lambda/N$. Using max value obtained from Eq. (A32), that $x < \ln 2$, Eq. (A33) reduces to $\frac{x^2}{2}e^x \leqslant 2\ln 2 \in [0, 1)$. Given that Eq. (A33) is a product of two monotonic functions in the domain $(0, \infty)$, any value of x less than the maximum obtained in the first condition also satisfies the second. Therefore, going forward we focus only on satisfying the constraint in Eq. (A32).

Going forward in linearizing Eq. (A31) we will take λ and N to be constant and consider variable β, and continue by upper bounding each term. Beginning with the first term we have

$$\frac{4\beta^2\lambda^2}{N}\frac{e^{2\lambda\beta/N}}{2-e^{2\beta\lambda/N}} = \frac{2\beta^2\lambda^2}{N}\frac{1}{y-\frac{1}{2}}, \quad \textbf{s.t.}\ y(\beta) = e^{-2\lambda\beta/N} \text{ and } \frac{2}{3} \leqslant y(\beta) \leqslant 1. \tag{A34}$$

Now we rewrite the aforementioned constraint $1 \leqslant e^{2\lambda\beta/N} < 2$, to be $1 \leqslant e^{2\lambda\beta/N} \leqslant \frac{3}{2}$ to prevent one from approaching a divergence in the proof arbitrarily closely. Utilizing the following property:

$$f(b) = f(a) + \int_a^b \frac{df}{dy}dy \leqslant f(a) + (b-a)\max_{y\in[a,b]}\left|\frac{df(y)}{dy}\right|, \tag{A35}$$

with $f(\beta) = \frac{1}{y(\beta)-\frac{1}{2}}$ Eq. (A34) is upper bounded in the following way:

$$\frac{4\beta^2\lambda^2}{N}\frac{4\beta^2\lambda^2}{N}e^{4\beta\lambda/N} \leqslant \frac{4\beta^2\lambda^2}{N}\left(3 + 12\frac{\lambda}{N}\right). \tag{A36}$$

Moving on to the third term in Eq. (A31), we have

$$\frac{4\beta^2\lambda^2}{N}e^{4\beta\lambda/N} = \frac{4\beta^2\lambda^2}{N}y^2, \quad \textbf{s.t.} \ \ y = e^{2\beta\lambda/N} \text{ and } 1 \leqslant y(\beta) \leqslant \frac{3}{2}, \tag{A37}$$

and using the same integral bound we obtain

$$\frac{4\beta^2\lambda^2}{N}e^{4\beta\lambda/N} \leqslant \frac{4\beta^2\lambda^2}{N}\left(1 + 3\frac{\lambda}{N}\right). \tag{A38}$$

Finally, we are left to bound the second term in Eq. (A31), which we can write in the following way:

$$\left(\frac{4\beta^2\lambda^2}{N}e^{4\lambda\beta/N}\right)\frac{\frac{4\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}}{2 - \frac{4\beta^2\lambda^2}{N^2}e^{2\lambda\beta/N}} = 2\left(\frac{4\beta^2\lambda^2}{N}e^{4\lambda\beta/N}\right)\frac{y\ln y}{1 - y\ln y}, \quad \textbf{s.t.} \ \ y(\beta) = e^{2\lambda\beta/N} \text{ and } 1 \leqslant y(\beta) \leqslant \frac{3}{2}. \tag{A39}$$

The function $f(y) = \frac{y\ln y}{1 - y\ln y}$ is monotonic over the constrained domain and can be upper bounded by inserting the largest value that $\ln y$ achieves. Also, notice the term in parentheses is precisely the third term in Eq. (A31) that we bounded above. Using these two observations,

$$2\left(\frac{4\beta^2\lambda^2}{N}e^{4\lambda\beta/N}\right)\frac{y\ln y}{1 - y\ln y} \leqslant \frac{4\beta^2\lambda^2}{N}2\left(1 + 3\frac{\lambda}{N}\right)\frac{y\ln\frac{3}{2}}{1 - y\ln\frac{3}{2}} \tag{A40}$$

$$\leqslant \frac{4\beta^2\lambda^2}{N}2\ln\frac{3}{2}\left(1 + 3\frac{\lambda}{N}\right)\left(\frac{1}{\frac{2}{3} - \ln\frac{3}{2}} + \frac{2\lambda}{3N}\frac{1}{\left(\frac{2}{3} - \ln\frac{3}{2}\right)^2}\right), \tag{A41}$$

where in the last step we once again used the integral identity following the same procedure as we did for the first term. Now that we have linearized each term, we combine all the results to obtain

$$\left|\left|\mathcal{U} - \Lambda_{QD}^{\circ N}\right|\right|_{1\to 1} \leqslant \frac{\beta^2\lambda^2}{N}\left(a + b\frac{\lambda}{N} + c\frac{\lambda^2}{N^2}\right) \tag{A42}$$

where $a \approx 28.41845$, $b \approx 128.95110$, and $c \approx 95.08717$. If we make the promise that $\frac{\lambda}{N} \leqslant 0.01$, the bound simplifies to

$$\left|\left|\mathcal{U} - \Lambda_{QD}^{\circ N}\right|\right|_{1\to 1} \leqslant \frac{C\beta^2\lambda^2}{N} \tag{A43}$$

where $C \approx 29.71747$, completing the proof. ∎

This condition is not very restrictive on the imaginary time β. Considering the condition given in Eq. (A32), we have $\beta < \frac{\ln 2N}{2\lambda} \approx 34.657$, so $\beta < 10\pi$, meaning that given this promise on the ratio $\frac{\lambda}{N}$, the bound holds for a simulation time that allows the phase to oscillate up to five times if we were dealing with real time.

2. Imaginary-time Trotter-Suzuki formulas

Following the procedure in [24], we must first convert the existing bound to a bound on channels and then bound the error with the renormalizing operations.

Theorem III.2 (Imaginary-time Trotter-Suzuki channels). Given a Hamiltonian $H = \sum_j h_j H_j$ with $\lambda = \sum_j h_j$, imaginary time β, time-steps r, and a density matrix initial state ρ, then the induced Schatten $1 \to 1$ norm of the difference between the imaginary-time Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho, \beta)$ and the exact imaginary-time evolutionary $\mathcal{U}(\rho, \beta)$ channel has the following bound:

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\text{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\text{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))}\right)^{\circ r}\right|\right|_{1\to 1} \leqslant 2\Upsilon^{2k+1}\frac{\alpha_{\text{comm}}(H, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}e^{4\Upsilon\beta\lambda/r}(e^{4\beta\lambda/r}e^{2\beta||H||/r} + e^{2\beta||H||/r}), \tag{25}$$

which yields the following asymptotic bound:

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\text{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\text{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))}\right)^{\circ r}\right|\right|_{1\to 1} \in \mathcal{O}\left(\Upsilon^{2k+1}\frac{\alpha_{\text{comm}}(H, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}\right). \tag{26}$$

Here Υ is the number of stages of the product formula, $2k$ is the order of the product formula, α_{comm} is defined as in Eq. (11), and λ is the sum of the spectral norms of the Hamiltonian summands. Here $\mathcal{O}(\cdot)$ is understood in the infinite limit of its arguments.

Proof. First consider the induced 1-norm of the difference between the channels after a single iteration or time-step r. As in the previous proof, absolute values containing a channel operation are understood to be maximized over all inputs $\rho : ||\rho||_1 = 1$,

$$\left|\left| \frac{\mathcal{U}(\rho, \beta/r, \beta/r)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta/r, \beta/r))} - \frac{\mathcal{T}^{2k}(\rho, \beta/r, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r, \beta/r))} \right|\right|_{1\to 1}$$

$$\leqslant \left|\left| \frac{\mathcal{U}(\rho, \beta/r) - \mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}\mathcal{U}(\rho, \beta/r)} \right|\right|_{1\to 1} + \left|\left| \mathcal{T}^{2k}(\rho, \beta/r)\left(\frac{1}{\mathrm{Tr}\mathcal{T}^{2k}(\rho, \beta/r)} - \frac{1}{\mathrm{Tr}\mathcal{U}(\rho, \beta/r)} \right) \right|\right|_{1\to 1} \tag{A44}$$

$$= \left| \frac{1}{\mathrm{Tr}\mathcal{U}(\rho, \beta/r)} \right| ||\mathcal{U}(\rho, \beta/r) - \mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1} + ||\mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1} \left| \frac{\mathrm{Tr}(\mathcal{U}(\rho, \beta/r) - \mathcal{T}^{2k}(\rho, \beta/r))}{\mathrm{Tr}(\mathcal{U}(\rho, \beta/r))\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))} \right| \tag{A45}$$

$$\leqslant \left| \frac{1}{\mathrm{Tr}\mathcal{U}(\rho, \beta/r)} \right| ||\mathcal{U}(\rho, \beta/r) - \mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1} + ||\mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1}||\mathcal{U}(\rho, \beta/r)$$

$$- \mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1} \left| \frac{1}{\mathrm{Tr}(\mathcal{U}(\rho, \beta/r))\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))} \right| \tag{A46}$$

$$= ||\mathcal{U}(\rho, \beta/r) - \mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1} \left(\frac{||\mathcal{T}^{2k}(\rho, \beta/r)||_{1\to 1}}{|\mathrm{Tr}(\mathcal{U}(\rho, \beta/r))\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))|} + \frac{1}{|\mathrm{Tr}\mathcal{U}(\rho, \beta/r)|} \right), \tag{A47}$$

where in the second line we simply cross multiply the traces of the outputs of the channels (since they are just numbers), and in third line we use the fact that the absolute value of the trace is trivially upper bounded by the diamond norm. Going forward we need to lower bound the traces in the denominator, which can be done using the well-known von Neumann trace inequality, $\sum_{i=1}^{N} \sigma_i^A \sigma_{N-i+1}^B \leqslant \mathrm{Tr}(AB)$ where $\sigma_i^{(A,B)}$ represent the singular values of the matrix A or B respectfully, and they are ordered such that $\sigma_i \leqslant \sigma_{i+1}$. The absolute values in the denominator also need not be written as these traces are positive by definition. For the exact evolution channel we have

$$\mathrm{Tr}\mathcal{U}(\rho, \beta/r) = \mathrm{Tr}(e^{-\beta H}\rho e^{-\beta/r H}) \tag{A48}$$

$$= \mathrm{Tr}(e^{-2\beta/r H}\rho) \tag{A49}$$

$$\geqslant \sum_i \sigma_i(e^{-2\beta/r H})\sigma_{N-i+1}(\rho) \tag{A50}$$

$$\geqslant \sigma_{\min}(e^{-2\beta/r H})\mathrm{Tr}(\rho) \tag{A51}$$

$$\geqslant e^{-2\beta/r||H||}, \tag{A52}$$

where we use the fact that the trace of ρ is 1, along with the fact that ρ is positive semidefinite and hermitian, such that the sum of its singular values is equal to its trace. Next, we follow a similar procedure in dealing with the trace of the output of the Trotter-Suzuki channel,

$$\mathrm{Tr}\mathcal{T}^{2k}(\rho, \beta/r) = \mathrm{Tr}(S_{2k}(\beta/r)\rho S_{2k}^\dagger(\beta/r)) \tag{A53}$$

$$\geqslant \sigma_{\min}(S_{2k}^\dagger(\beta/r)S_{2k}(\beta/r))\mathrm{Tr}\rho \tag{A54}$$

$$= \sigma_{\min}(S_{2k}^\dagger(\beta/r)S_{2k}(\beta/r)). \tag{A55}$$

In order to understand how this term behaves with respect to β/r and H, we require more careful analysis. To do this it will be useful to introduce some more convenient notation. Just like an arbitrary product formula can be defined in the following way (following the notation of [14]):

$$S_{2k}(\beta/r) = \prod_{\upsilon=1}^{\Upsilon} \prod_{l=1}^{L} e^{-\beta/r s_{\upsilon,l} H_{\pi_{\upsilon(l)}}}. \tag{A56}$$

Where s controls the prefactor that modifies the time step and π controls the ordering of the H summands in each stage of the formula. We utilize the following property of singular values: given two matrices $A, B \in \mathbb{C}$ the following holds $\sigma_n(A)\sigma_i(B) \leqslant \sigma_i(AB) \leqslant \sigma_1(A)\sigma_i(B)$, given that singular values are ordered such that $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n$. Also, note that since the individual evolution operators are no longer anti-Hermitian, coupled with the fact that Trotter-Suzuki formulas are symmetric, $S_{2k}^\dagger = S_{2k}$,

$$\mathrm{Tr}\mathcal{T}^{2k}(\rho, \beta/r) \geqslant \sigma_{\min}(S_{2k}^2(\beta/r)) \tag{A57}$$

$$\geqslant \sigma_{\min}^2 \left(\prod_{\upsilon=1}^{\Upsilon} \prod_{l=1}^{L} e^{-\beta/r s_{\upsilon,l} H_{\pi_{\upsilon(l)}}} \right) \tag{A58}$$

$$\geqslant \left(\prod_{\upsilon=1}^{\Upsilon} \prod_{l=1}^{L} \sigma_{\min}(e^{-\beta/r s_{\upsilon,l} H_{\pi_{\upsilon(l)}}}) \right)^2 \tag{A59}$$

$$\geqslant \left(\prod_{\upsilon=1}^{\Upsilon} \prod_{l=1}^{L} e^{-\beta/r s_{\upsilon,l} ||H_{\upsilon(l)}||} \right)^2 \tag{A60}$$

$$= \left(e^{\sum_{\upsilon=1}^{\Upsilon} \sum_{l=1}^{L} -\beta/r s_{\upsilon,l} ||H_{\upsilon(l)}||} \right)^2 \tag{A61}$$

$$= e^{-2\beta/r \lambda}, \tag{A62}$$

where each inequality step is a repeated application of the aforementioned property. In the last step, we complete the using the property of Trotter-Suzuki formulas that $\sum_{\upsilon,l} s_{\upsilon,l} = 1$ and the definition $\sum_{l} ||H_l|| = \lambda$.

We also need to deal with the most important term, which is the actual induced 1-norm of the difference between the channels. To do this, we will massage this term into a form where we can apply a result from [14]. To do this we follow the same procedure from Sec. II G,

$$||\mathcal{U}(\beta) - \mathcal{T}^{2k}(\beta)||_{1\to1} := \max_{\rho:||\rho||=1} ||\mathcal{U}(\beta) - \mathcal{T}^{2k}(\beta)||_1 \tag{A63}$$

$$= \max_{\rho:||\rho||_1=1} ||e^{-H\beta} \rho e^{-H\beta} - S_{2k}(\beta) \rho S_{2k}(\beta)||_1 \tag{A64}$$

$$\leqslant \max_{\rho:||\rho||_1=1} ||e^{-H\beta} \rho e^{-H\beta} - e^{-H\beta} \rho S_{2k}(\beta)||_1 \tag{A65}$$

$$+ \max_{\rho:||\rho||_1\leqslant1} ||e^{-H\beta} \rho S_{2k}(\beta) - S_{2k}(\beta) \rho S_{2k}(\beta)||_1 \tag{A66}$$

$$= \max_{\rho:||\rho||_1=1} ||e^{-\beta H} \rho(e^{-H\beta} - S_{2k}(\beta))||_1 + \max_{\rho:||\rho||_1=1} ||(e^{-H\beta} - S_{2k}(\beta)) \rho S_{2k}(\beta)||_1 \tag{A67}$$

$$\leqslant (||e^{-\beta H}||_\infty + ||S_{2k}(\beta)||_\infty) ||e^{-H\beta} - S_{2k}(\beta)||_\infty \max_{\rho:||\rho||_1\leqslant1} ||\rho||_1 \tag{A68}$$

$$= 2e^{\beta\lambda} ||e^{-H\beta} - S_{2k}(\beta)||_\infty \tag{A69}$$

$$= 2e^{\beta\lambda} ||U(\beta) - S_{2k}(\beta)||_\infty. \tag{A70}$$

Next we apply the following bound from Appendix E in [14],

$$||U(\beta/r) - S_{2k}(\beta/r)||_\infty \leqslant 2\Upsilon^{2k+1} \frac{\alpha_{\text{comm}}(H, 2k)}{(2k+1)!} \frac{\beta^{2k+1}}{r^{2k+1}} e^{4\frac{\beta}{r}\Upsilon\lambda} \tag{A71}$$

where Υ and λ are again the number of stages of the product formula and the sum of the spectral norms of the Hamiltonian summands respectively. Now we are simply left with a single term to bound in the diamond norm of the Trotter-Suzuki channel. We can similarly define the Trotter-Suzuki channel as a composition of multiple stages of channels, each containing compositions of evolution channels corresponding to each Hamiltonian summand. We write this in the following way:

$$\mathcal{T}^{2k}(\beta/r) \equiv \bigcirc_{\upsilon=1}^{\Upsilon} \bigcirc_{l=1}^{L} e^{-\frac{\beta}{r} s_{\upsilon,l} \mathcal{L}_{\pi_{\upsilon(l)}}}, \tag{A72}$$

where we once again have the Liouvillian generator for imaginary time $\mathcal{L}(\rho) = \{H, \rho\}$, and all other machinery is defined in the same was as above. In this way it is more natural to proceed in using the submultiplicative property of the norm,

$$||\mathcal{T}^{2k}(\beta/r)||_{1\to1} \leqslant \prod_{\upsilon=1}^{\Upsilon} \prod_{l=1}^{L} ||e^{-\frac{\beta}{r} s_{\upsilon,l} \mathcal{L}_{\pi_{\upsilon(l)}}}||_{1\to1} \tag{A73}$$

$$\leqslant \prod_{\upsilon=1}^{\Upsilon} \prod_{l=1}^{L} e^{2\frac{\beta}{r} s_{\upsilon,l} ||H_{\pi_{\upsilon(l)}}||} \tag{A74}$$

$$= e^{\sum_{\upsilon=1}^{\Upsilon} \sum_{l=1}^{L} 2\frac{\beta}{r} s_{\upsilon,l} ||H_{\upsilon(l)}||} \tag{A75}$$

$$= e^{2\frac{\beta}{r}\lambda}. \tag{A76}$$

In the second line we use the Taylor expansion of the exponential of the Liouvillian to apply the induced $1 \to 1$ norm to the exponent as seen in the Qdrift proof. We make use of the aforementioned fact that $||\mathcal{L}_l||_{1\to1} \leqslant 2h_l$ where h_l is the spectral norm of the H_l Hamiltonian term. Now, combining all of our results we are left with the following:

$$\left\| \frac{\mathcal{U}(\rho, \beta)}{\text{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\text{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))} \right)^{\circ r} \right\|_{1\to1} \leqslant 4\Upsilon^{2k+1} \frac{\alpha_{\text{comm}}(H, 2k)}{(2k+1)!} \frac{\beta^{2k+1}}{r^{2k+1}} e^{4\frac{\beta}{r}\Upsilon\lambda} e^{\beta\lambda/r} \left(e^{4\frac{\beta}{r}\lambda} e^{2\frac{\beta}{r}||H||} + e^{2\frac{\beta}{r}||H||} \right) \tag{A77}$$

and we wish to compare the exact evolution to the channel after r iterations. In other words, we use the property $||X^{\circ r} - Y^{\circ r}||_{1\to 1} \leqslant r||X - Y||_{1\to 1}$, proven earlier (for the diamond norm, but it also holds here) in Eq. (37), to obtain

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))}\right)^{\circ r}\right|\right|_{1\to 1} \leqslant 4\Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(H, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}e^{4\frac{\beta}{r}\Upsilon\lambda}e^{\beta\lambda/r}\left(e^{4\frac{\beta}{r}\lambda}e^{2\frac{\beta}{r}||H||} + e^{2\frac{\beta}{r}||H||}\right). \tag{A78}$$

Similar to the Qdrift case, we wish to eliminate the exponential factors to give a more straightforward interpretation of scaling of the error bound. However, in this case the proof did not impose restrictions on the size of these exponentials. Therefore, we instead upper bound each exponential term in a common way and then make on observation about the scaling of r, to show that the exponential factors are $\mathcal{O}(1)$. We begin by using the fact that $||H|| \leqslant \lambda$, and $e^{\beta\lambda/r} \leqslant e^{\Upsilon\beta\lambda/r}$ to obtain

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))}\right)^{\circ r}\right|\right|_{1\to 1} \leqslant 4\Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(H, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}\left(e^{\frac{11\Upsilon\beta\lambda}{r}} + e^{\frac{7\Upsilon\beta\lambda}{r}}\right). \tag{A79}$$

Now, let us posit that $r \in \Omega(\Upsilon^a\beta^b\lambda^c) \mid a, b, c \geqslant 1$. Focusing on the terms in parentheses above, the terms in the exponents are positive by definition, so we have $(e^{10\Upsilon\beta\lambda/r} + e^{6\Upsilon\beta\lambda/r}) \geqslant 1$. If the posited scaling holds, then $(e^{11\Upsilon\beta\lambda/r} + e^{7\Upsilon\beta\lambda/r}) \in \mathcal{O}(1)$ such that this term is asymptotically upper bounded by some constant C. In fact, if a, b, or $c > 1$, then $\mathrm{C} = 1$ in the infinite limit. From [24], we use the fact that $\alpha_{\mathrm{comm}}(H, 2k) \leqslant 2^{2k}\lambda^{2k+1}$ to verify this intuition,

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))}\right)^{\circ r}\right|\right|_{1\to 1} \leqslant 4^{2k+1}\Upsilon^{2k+1}\frac{\lambda^{2k+1}}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}C < \epsilon, \tag{A80}$$

where solving for r in terms of ϵ gives

$$(4\Upsilon\lambda\beta)^{1+\frac{1}{2k}}\left(\frac{C}{\epsilon(2k+1)!}\right)^{\frac{1}{2k}} < r. \tag{A81}$$

Asymptotically, this gives the following asymptotic bound:

$$r \in \Omega\left(\frac{\Upsilon^{1+\frac{1}{2k}}\lambda^{1+\frac{1}{2k}}\beta^{1+\frac{1}{2k}}}{\epsilon^{\frac{1}{2k}}}\right). \tag{A82}$$

Here, we enforce $\epsilon \leqslant 1$, which then confirms the initial intuition about the asymptotic scaling of r, such that that $r \in \Omega(\Upsilon^a\beta^b\lambda^c)$, where $a = b = c = 1 + 1/2k$, and the exponential term vanishes in the infinite limit, yielding

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}(\mathcal{U}(\rho, \beta))} - \left(\frac{\mathcal{T}^{2k}(\rho, \beta/r)}{\mathrm{Tr}(\mathcal{T}^{2k}(\rho, \beta/r))}\right)^{\circ r}\right|\right|_{1\to 1} \in \mathcal{O}\left(\Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(H, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}\right), \tag{A83}$$

thus completing the proof. ∎

3. Imaginary-time composite channels

Theorem III.3 (Imaginary-time composite channels). Given a partitioned Hamiltonian $H = \sum_j a_j A_j + \sum_i b_i B_i$ with $\lambda_A = \sum_j a_j$, $\lambda_B = \sum_j b_j$, imaginary time β, time steps r, Qdrift samples N_B, and a density matrix initial state ρ, then the induced Schatten $1 \to 1$ norm of the difference between the imaginary-time composite channel $\mathcal{X}^{2k}(\rho, \beta)$ of order $2k$ and the exact imaginary-time evolutionary channel $\mathcal{U}(\rho, \beta)$ has the following bound:

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}\mathcal{U}(\rho, \beta)} - \frac{\mathcal{X}^{2k}(\rho, \beta)}{\mathrm{Tr}\mathcal{X}^{2k}(\rho, \beta)}\right|\right|_{1\to 1} \in \mathcal{O}\left(\frac{\beta^2\lambda_B^2}{N_B r} + \Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(A, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}} + ||[A, B]||\frac{\beta^2}{r}\right) \tag{27}$$

given that $r \geqslant \Upsilon\beta\lambda$. Here, all parameters in this bound correspond to the same quantities in Theorems III.1 and III.2, where the subscripts A and B indicate their belonging corresponding set.

Proof. We begin by expanding the imaginary-time composite channel $\mathcal{X}(\rho, \beta)$ in terms of familiar quantities that have been previously bounded. That being said, this proof relies almost entirely on Theorems III.1 and III.2. Throughout, notation is used such that $\mathcal{U}(\rho, \beta)_A$ represents an exact evolution channel under the Hamiltonian A from the the theorem statement. It is implied throughout that terms in A are simulated by the Trotter-Suzuki channel, while terms in B are simulated by Qdrift. Therefore, we write $\mathcal{T}_A^{2k}(\rho, \beta/r)$ and $\mathcal{Q}_B(\rho, \beta/r)$ as Trotter and Qdrift channels generated by the Hamiltonians A and B for clearer notation, and once again we initial consider a time-step β/r. Expanding the composite channel we have

$$\left|\left|\frac{\mathcal{U}(\rho, \beta/r)}{\mathrm{Tr}\mathcal{U}(\rho, \beta/r)} - \frac{\mathcal{X}^{2k}(\rho, \beta/r)}{\mathrm{Tr}\mathcal{X}^{2k}(\rho, \beta/r)}\right|\right|_{1\to 1} = \left|\left|\frac{\mathcal{U}_B(\rho, \beta/r) \circ \mathcal{U}_A(\rho, \beta/r) + E_{A,B}(\beta/r)}{\mathrm{Tr}\mathcal{U}(\rho, \beta/r)} - \frac{\mathcal{Q}_B(\rho, \beta/r) \circ \mathcal{T}_A^{2k}(\rho, \beta/r)}{\mathrm{Tr}\mathcal{Q}_B(\rho, \beta/r) \circ \mathcal{T}_A^{2k}(\rho, \beta/r)}\right|\right|_{1\to 1}, \tag{A84}$$

where $E_{A,B}(\beta/r)$ is an error term that comes from partitioning the exact evolution $\mathcal{U}$ into two evolution channels over A and B. We leave the denominator un-expanded for convenience. As before, we can add and subtract a convenient term to give us terms with common denominators that are more straightforward to bound, and then apply the triangle inequality,

$$= \left|\left|\frac{\mathcal{U}_B(\rho,\beta/r) \circ \mathcal{U}_A(\rho,\beta/r) - \mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}{\mathrm{Tr}\mathcal{U}(\rho,\beta/r)}\right|\right|_{1\to1} + \left|\left|\frac{E_{A,B}(\beta/r)}{\mathrm{Tr}\mathcal{U}(\rho,\beta/r)}\right|\right|_{1\to1}$$
$$+ \left|\left|\frac{\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}{\mathrm{Tr}\mathcal{U}(\rho,\beta/r)} - \frac{\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}{\mathrm{Tr}\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}\right|\right|_{1\to1}. \tag{A85}$$

Now we proceed by bounding each of these terms individually. Starting with the first term

$$\left|\left|\frac{\mathcal{U}_B(\rho,\beta/r) \circ \mathcal{U}_A(\rho,\beta/r) - \mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}{\mathrm{Tr}\mathcal{U}(\rho,\beta/r)}\right|\right|_{1\to1}$$

$$\leqslant \frac{||\mathcal{U}_B(\rho,\beta/r) - \mathcal{Q}_B(\rho,\beta/r)||_{1\to1} + \left|\left|\mathcal{U}_A(\rho,\beta/r) - \mathcal{T}_A^{2k}(\rho,\beta/r)\right|\right|_{1\to1}}{\mathrm{Tr}\mathcal{U}(\rho,\beta/r)} \tag{A86}$$

$$\leqslant \left(||\mathcal{U}_B(\rho,\beta/r) - \mathcal{Q}_B(\rho,\beta/r)||_{1\to1} + \left|\left|\mathcal{U}_A(\rho,\beta/r) - \mathcal{T}_A^{2k}(\rho,\beta/r)\right|\right|_{1\to1}\right)e^{2||H||\beta/r}, \tag{A87}$$

where in the second line we use the trace bound found in the proof of Theorem III.2. Each of the diamond norms remaining above have been bounded in Theorems III.1 and III.2, and we will plug in said results at the end of the proof. Proceeding to the second term we have

$$\left|\left|\frac{\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}{\mathrm{Tr}\mathcal{U}(\rho,\beta/r)} - \frac{\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}{\mathrm{Tr}\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)}\right|\right|_{1\to1} \tag{A88}$$

$$\leqslant ||\mathcal{Q}_B(\rho,\beta/r)||_{1\to1}\left|\left|\mathcal{T}_A^{2k}(\rho,\beta/r)\right|\right|_{1\to1}\left|\frac{\mathrm{Tr}\big(\mathcal{U}(\rho,\beta/r) - \mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)}{\mathrm{Tr}(\mathcal{U}(\rho,\beta/r))\mathrm{Tr}\big(\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)}\right| \tag{A89}$$

$$= ||\mathcal{Q}_B(\rho,\beta/r)||_{1\to1}\left|\left|\mathcal{T}_A^{2k}(\rho,\beta/r)\right|\right|_{1\to1}\frac{\left|\mathrm{Tr}\big(\mathcal{U}_B(\rho,\beta/r) \circ \mathcal{U}_A(\rho,\beta/r) + E_{A,B}(\beta/r) - \mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)\right|}{\mathrm{Tr}(\mathcal{U}(\rho,\beta/r))\mathrm{Tr}\big(\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)} \tag{A90}$$

$$\leqslant e^{\frac{2\beta\lambda_B}{r}}e^{\frac{2\beta\lambda_A}{r}}e^{\frac{2\beta||H||}{r}}\frac{\left|\left|\mathcal{U}_B(\rho,\beta/r) \circ \mathcal{U}_A(\rho,\beta/r) + E_{A,B}(\beta/r) - \mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\right|\right|_{1\to1}}{\mathrm{Tr}\big(\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)} \tag{A91}$$

$$\leqslant e^{\frac{2\beta\lambda}{r}}e^{\frac{2\beta||H||}{r}}\left(\frac{||\mathcal{U}_B(\rho,\beta/r) - \mathcal{Q}_B(\rho,\beta/r)||_{1\to1} + \left|\left|\mathcal{U}_A(\rho,\beta/r) - \mathcal{T}_A^{2k}(\rho,\beta/r)\right|\right|_{1\to1}}{\mathrm{Tr}\big(\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)} + \frac{||E_{A,B}(\beta/r)||_{1\to1}}{\mathrm{Tr}\big(\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big)}\right), \tag{A92}$$

where in the last line we once again apply the triangle inequality, and also observe that $\lambda_A + \lambda_B = \lambda$. Now we only have two unfamiliar terms left to bound, the partitioning error $E_{A,B}(\beta/r)$ and the trace of the composition of the Qdrift and Trotter channels. To analyze the trace, we expand both of the channels into their operator form

$$\mathrm{Tr}\big(\mathcal{Q}_B(\rho,\beta/r) \circ \mathcal{T}_A^{2k}(\rho,\beta/r)\big) = \mathrm{Tr}\left(\sum_{\mathbf{j}} P_{\mathbf{j}}\mathbf{V}_{\mathbf{j}}S_{2k}\rho S_{2k}\mathbf{V}_{\mathbf{j}}^{\dagger}\right) \tag{A93}$$

$$= \sum_{\mathbf{j}} P_{\mathbf{j}}\mathrm{Tr}(\mathbf{V}_{\mathbf{j}}S_{2k}\rho S_{2k}\mathbf{V}_{\mathbf{j}}^{\dagger}) \tag{A94}$$

$$= \sum_{\mathbf{j}} P_{\mathbf{j}}\mathrm{Tr}(S_{2k}\mathbf{V}_{\mathbf{j}}^{\dagger}\mathbf{V}_{\mathbf{j}}S_{2k}\rho) \tag{A95}$$

$$\geqslant \sum_{\mathbf{j}} P_{\mathbf{j}}\sigma_{\min}(S_{2k}\mathbf{V}_{\mathbf{j}}^{\dagger}\mathbf{V}_{\mathbf{j}}S_{2k})\mathrm{Tr}\rho \tag{A96}$$

$$\geqslant \sigma_{\min}^2(S_{2k})\sum_{\mathbf{j}} P_{\mathbf{j}}\sigma_{\min}^2(\mathbf{V}_{\mathbf{j}}) \tag{A97}$$

$$\geqslant e^{-\frac{2\beta\lambda_A}{r}}\sum_{\mathbf{j}} P_{\mathbf{j}}\sigma_{\min}^2(\mathbf{V}_{\mathbf{j}}). \tag{A98}$$

Now recall the definition of the product formula from the multisample Qdrift channel defined in Definition II.3 $\mathbf{V}_{\mathbf{j}} = \prod_{k=1}^N e^{-iH_{j_k}\tau}$ with $\tau = \beta\lambda/N$. To reiterate, the index $\mathbf{j}$ is a vector of length N, and the index set contains every possible length-N permutation

of the L_B Hamiltonian terms in the Qdrift channel. Since we already have exponential terms in our bound, we can use a naive lower bound here, which coincides with the most likely case of drawing the largest term N times. This gives

$$\geqslant e^{-\frac{2\beta\lambda_A}{r}} \sum_{\mathbf{j}} P_{\mathbf{j}} \prod_{k=1}^{N} \sigma_{\min}^2 \left(e^{-B_{j_k}\frac{\beta\lambda_B}{Nr}}\right) \tag{A99}$$

$$\geqslant e^{-\frac{2\beta\lambda_A}{r}} e^{-\frac{2||B||\beta\lambda_B}{r}} \tag{A100}$$

$$= e^{-\frac{2\beta\lambda}{r}}, \tag{A101}$$

where in the last line $||B|| = 1$ based by definition of Qdrift, and $\lambda_A + \lambda_B = \lambda$.

Now combining all of our results we have the following:

$$\left(||\mathcal{U}_B(\rho, \beta/r) - \mathcal{Q}_B(\rho, \beta/r)||_{1\to 1} + \left||\mathcal{U}_A(\rho, \beta/r) - \mathcal{T}_A^{2k}(\rho, \beta/r)\right||_{1\to 1} + ||E_{A,B}(\beta/r)||_{1\to 1}\right) e^{\frac{2\beta||H||}{r}} \left(1 + e^{\frac{4\beta\lambda}{r}}\right). \tag{A102}$$

This term is a sum of the errors of each respective algorithm, with the addition of the partitioning error, multiplied by some exponential terms that come from lower bounding traces, which have been common throughout. Therefore, this is exactly what we might expect a priori. Going forward we wish to represent this error for the entire simulation rather than just a single time step, and then show that the exponential terms become irrelevant in the limit of interest. For the full time simulation we have

$$\left|\left|\left(\frac{\mathcal{U}(\rho, \beta/r)}{\mathrm{Tr}\,\mathcal{U}(\rho, \beta/r)}\right)^{\circ r} - \left(\frac{\mathcal{X}^{2k}(\rho, \beta/r)}{\mathrm{Tr}\,\mathcal{X}^{2k}(\rho, \beta/r)}\right)^{\circ r}\right|\right|_{1\to 1} \leqslant r\left|\left|\frac{\mathcal{U}(\rho, \beta/r)}{\mathrm{Tr}\,\mathcal{U}(\rho, \beta/r)} - \frac{\mathcal{X}^{2k}(\rho, \beta/r)}{\mathrm{Tr}\,\mathcal{X}^{2k}(\rho, \beta/r)}\right|\right|_{1\to 1}, \tag{A103}$$

where the quantity on the right is that which has been bounded. Inserting the imaginary-time Qdrift bound from Theorem III.1, and the single step Trotter bound from Eq. (A77) we have

$$\left(\frac{C\beta^2\lambda_B^2}{N_B r} + 2\Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(A, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}} e^{4\beta/r\Upsilon\lambda_A}(e^{4\beta/r\lambda_A}e^{2\beta/r||A||} + e^{2\beta/r||A||}) + r||E_{A,B}(\beta/r)||_{1\to 1}\right) e^{\frac{2\beta||H||}{r}} \left(1 + e^{\frac{4\beta\lambda}{r}}\right). \tag{A104}$$

Next we deal with the partitioning error $||E_{A,B}(\beta/r)||_{1\to 1} = ||\mathcal{U}(\rho, \beta/r) - \mathcal{U}_B(\rho, \beta/r) \circ \mathcal{U}_A(\rho, \beta/r)||_{1\to 1}$. We know from Theorem III.2 that $||\mathcal{U}(\beta, \rho) - \mathcal{U}_B(\rho, \beta/r) \circ \mathcal{U}_A(\rho, \beta/r)||_{1\to 1} \leqslant 2e^{\frac{\beta\lambda}{r}}||\mathcal{U}(\beta, \rho) - \mathcal{U}_B(\rho, \beta/r) \circ \mathcal{U}_A(\rho, \beta/r)||_{\infty}$, then once again using the bound from Appendix E of [14], we have

$$\left(\frac{C\beta^2\lambda_B^2}{N_B r} + 2\Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(A, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}} e^{4\beta/r\Upsilon\lambda_A}(e^{4\beta/r\lambda_A}e^{2\beta/r||A||} + e^{2\beta/r||A||}) + ||[A, B]||e^{\frac{\beta\lambda}{r}}\frac{2\beta^2}{r}e^{\frac{4\beta\lambda}{r}}\right) e^{\frac{2\beta||H||}{r}} \left(1 + e^{\frac{4\beta\lambda}{r}}\right). \tag{A105}$$

We are now left with an equation nonlinear in r, which makes it more difficult to make asymptotic arguments similar to those in made Theorem III.2. Since, we wish to eliminate the exponentials from this bound, we first upper bound all exponentials by $e^{n\Upsilon\beta\lambda/r}$ where we keep the coefficients n for each of them. This allows us to collect the exponentials and simplify the expression to the following:

$$\left(\frac{C\beta^2\lambda_B^2}{N_B r} + 2\Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(A, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}}(e^{10\Upsilon\beta\lambda/r} + e^{6\Upsilon\beta\lambda/r}) + ||[A, B]||\frac{2\beta^2}{r}e^{5\Upsilon\beta\lambda/r}\right)(e^{2\Upsilon\beta\lambda/r} + e^{6\Upsilon\beta\lambda/r}). \tag{A106}$$

Now, we introduce the constraint that $r \geqslant \Upsilon\beta\lambda$. Imposing this constraint over r allows all of the exponentials to be upper bounded by a constant such that in the limit that all other parameters approach infinity, this bound becomes the following:

$$\left|\left|\frac{\mathcal{U}(\rho, \beta)}{\mathrm{Tr}\,\mathcal{U}(\rho, \beta)} - \frac{\mathcal{X}^{2k}(\rho, \beta)}{\mathrm{Tr}\,\mathcal{X}^{2k}(\rho, \beta)}\right|\right|_{1\to 1} \in \mathcal{O}\left(\frac{\beta^2\lambda_B^2}{N_B r} + \Upsilon^{2k+1}\frac{\alpha_{\mathrm{comm}}(A, 2k)}{(2k+1)!}\frac{\beta^{2k+1}}{r^{2k}} + ||[A, B]||\frac{\beta^2}{r}\right), \tag{A107}$$

thus completing the proof. ∎

APPENDIX B: NUMERICAL ERROR ANALYSIS

1. Types of Error Measures

Here, we propose two different error measures and discuss the advantages associated with each.

a. Infidelity of quantum states

The fidelity is a quantity that serves as a measure of the closeness between two quantum states, of which can be either pure or mixed states and is summarized in the following definition.

Definition B.1 (Fidelity). Let $\rho, \sigma \in \mathbb{C}^{2^n \times 2^n}$ be density matrices that represent quantum states with equal Hilbert space dimension. The fidelity between the states is then defined as

$$\mathscr{F}(\rho\sigma) := \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]^2 \tag{B1}$$

where $\mathscr{F}(\rho, \sigma) \in [0, 1]$ and $\mathscr{F}(\rho, \sigma) = \mathscr{F}(\sigma, \rho)$.

The infidelity is then simply defined as $1 - \mathscr{F}$ and represents a measure of distance between two states with the latter properties in the definition still holding. The computational advantage of using infidelity over some other measures is that

it is relatively cheap to compute, especially for large systems. This does not appear obvious on first glance given the need to compute two matrix square roots, which often involves diagonalization. However, we can take advantage of a nice property, $\mathscr{F} = \mathrm{Tr}(\rho, \sigma)$ if one of ρ or σ is pure. A simple proof is provided in the following lemma.

Lemma 1 (Pure-nonpure fidelity property). Given a two density matrices ρ and σ where at least one of which is pure, then the expression for the fidelity simplifies to $\mathscr{F}(\rho, \sigma) = \mathrm{Tr}(\rho\sigma)$.

Proof.

$$\mathscr{F}(\rho, \sigma) := \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\,\sigma\,\sqrt{\rho}}\right]^2 \tag{B2}$$

$$= \mathrm{Tr}\left[\sqrt{\sqrt{|\psi\rangle\langle\psi|}\,\sigma\,\sqrt{|\psi\rangle\langle\psi|}}\right]^2 \tag{B3}$$

$$= \mathrm{Tr}\left[\sqrt{|\psi\rangle\langle\psi|\sigma|\psi\rangle\langle\psi|}\right]^2 \tag{B4}$$

$$= \langle\psi|\sigma|\psi\rangle\mathrm{Tr}(|\psi\rangle\langle\psi|) \tag{B5}$$

$$= \mathrm{Tr}(\rho\sigma). \tag{B6}$$

Here we choose ρ to be a pure state $|\psi\rangle\langle\psi|$, which is a projector and thus $\rho = \rho^2 = \sqrt{\rho}$. We use these properties along with the fact that the trace of a pure state is equal to one to complete the proof. Note that $\mathscr{F}$ is also known to be symmetric $\mathscr{F}(\rho, \sigma) = \mathscr{F}(\sigma, \rho)$ but we do not show this here. This implies the statement is true for either ρ or σ pure. ∎

The important part of this lemma is that we only require one of the density matrices to be a pure state. To evaluate the error of a composite channel, we will be taking the infidelity between channel outputs, and due to the nature of Qdrift the channel will always output a mixed state. However, the exact evolution channel providing the means for comparison only produces pure states given pure state input. With these definitions and properties we can then present the following strategy: Reformulate the fidelity expression such that it can be written as an expectation value that only requires matrix vector multiplication to compute, and then use Monte Carlo sampling to approximate this quantity. This is shown via the following derivation:

$$\mathscr{F}(\mathcal{U}\,(\rho, t), \mathcal{X}^{2k}(\rho, t))$$

$$= \mathrm{Tr}(e^{-iHt}|\psi\rangle\langle\psi|e^{iHt}e^{-iAt}e^{-iBt}|\psi\rangle\langle\psi|e^{iBt}e^{iAt}) \tag{B7}$$

$$= \mathrm{Tr}(e^{-iHt}|\psi\rangle\langle\psi|e^{iHt}S_{2k}(-iAt)$$
$$\times \sum_j P_j e^{-iB_j\tau}|\psi\rangle\langle\psi|e^{iB_jt}S_{2k}(-iAt)^\dagger) \tag{B8}$$

$$= \sum_j P_j\mathrm{Tr}(\langle\psi|e^{iB_jt}S_{2k}(-iAt)^\dagger e^{-iHt}|\psi\rangle$$
$$\times \langle\psi|e^{-iHt}S_{2k}(-iAt)e^{-iB_j\tau}|\psi\rangle) \tag{B9}$$

$$= \sum_j P_j|\langle\psi|e^{iB_j\tau}S_{2k}(-iAt)^\dagger e^{-iHt}|\psi\rangle|^2 \tag{B10}$$

$$= \mathbb{E}[|\langle\psi|e^{iB_j\tau}S_{2k}(-iAt)^\dagger e^{-iHt}|\psi\rangle|^2] \tag{B11}$$

$$\approx \frac{1}{N}\sum_j^N |\langle\psi|e^{iB_j\tau}S_{2k}(-iAt)^\dagger e^{-iHt}|\psi\rangle|^2. \tag{B12}$$

Now when numerically evaluating the error, only matrix-vector multiplication is needed as opposed to matrix-matrix before. Here we no longer need to build exact density matrices outputted by the composite channel, and the step involving the matrix square root has been eliminated. However, the trade-off is that we have to use Monte Carlo sampling to estimate this expectation value. Given that the naive methods for matrix-vector and matrix-matrix multiplication scale like $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$ respectively, for small systems the overhead of Monte Carlo sampling may lead to a loss of any computational advantage. However, for large systems where $n^3 \gg n^2$, this sampling method is quite advantageous. It is also important to note that sampling makes Qdrift channel much cheaper to study as well. For a Qdrift channel with L terms and N samples, building the exact channel yields additional matrix multiplication overhead $\mathcal{O}(NL)$ (not considering arithmetic overhead). While this does not seem prohibitive, it does become expensive when we have a Hamiltonian with many terms that we wish to sample many times for a high degree of accuracy. It should be noted, however, that this is not as bad as it initially seems given that there are L^N terms Qdrift channel sum. The numerical routine that computes this sum was optimized to achieve this, by using the fact that channel samples are i.i.d. With intermediate summation to produce a new input ρ after each sample vector, we do not need to compute each term in this sum individually,

$$\sum_j p_j e^{-iH_jt}\rho_i e^{iH_jt} \to \rho_{i+1} \tag{B13}$$

repeating this procedure N times until $\rho_i \to \rho_N$, which constructs the exact Qdrift channel output with $\mathcal{O}(NL)$ matrix multiplications. We mention this to highlight the advantages of Monte Carlo sampling where it can be done reliably, as well as to highlight how Qdrift is numerically computed exactly in our paper (which will be made necessary by results in the following section).

b. Trace distance over density matrices

The next error measure we wish to investigate is the trace distance, which is equivalent to the Schatten 1-norm. The trace distance is a metric over the space of density matrices that can be defined in the following way:

Definition B.2 (Trace distance). Let $\rho, \sigma \in \mathbb{C}^{2^n \times 2^n}$ be density matrices that represent quantum states with equal Hilbert space dimension. The trace distance between the states is then defined as

$$\mathscr{T}(\rho, \sigma) := ||\rho - \sigma||_1$$

$$= \mathrm{Tr}\left[\sqrt{(\rho-\sigma)(\rho-\sigma)^\dagger}\right], \tag{B14}$$

where $\mathscr{T}(\rho, \sigma) \in [0, 1]$.

Given that ρ and σ are necessarily Hermitian, the trace distance simplifies to $\mathscr{T}(\rho, \sigma) = \sum_i |\lambda_i|^2$ where λ_i are eigenvalues of the matrix $(\rho - \sigma)$. There are no simplifications in our approach to computing this numerically. Here, we must compute the exact density matrix output of the composite channel (a necessarily mixed state outputted by the Qdrift channel component) and the output of the unitary channel. Exact diagonalization is performed to obtain the eigenvalues, and the sum above is then computed. Providing a rigorous

argument for crossover points in the compute time between the fidelity and trace distance (as a function of N, L) is beyond the scope of this paper; however, given the arguments from the previous section our intuition is that for small systems the trace distance will be computable in a reasonable amount of time whereas Monte Carlo sampling the infidelity will become more advantageous as the system size grows.

2. Comparison of error measures

In the following section we provide short proofs regarding how the error of each of the Qdrift and Trotter-Suzuki algo-

rithms scale with time, both in the infidelity and trace distance framework. It is important to note that we are specifically using the Schatten 1-norm over density matrices and not an induced 1-norm over channels. Therefore, we are computing this quantity using the outputs of the two channels, and the same goes for the infidelity, which is only defined on the density matrices that are the outputs of the channels in question. We begin with the infidelity.

Proposition V.1 (Qdrift infidelity time scaling). Given a Qdrift channel $\mathcal{Q}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, t/N)$, for a density matrix ρ, time t/N, then the infidelity between the outputs of the channels $1 - \mathscr{F}(\mathcal{Q}(\rho,t), \mathcal{U}(\rho, t/N)) \in \mathcal{O}(t^2)$.

Proof.

$$1 - \mathscr{F}(\mathcal{Q}(\rho,t), \mathcal{U}(\rho, t/N)) := 1 - \mathrm{Tr}(\mathcal{Q}(\rho,t)\mathcal{U}(\rho, t/N)) \tag{B15}$$

$$= 1 - \mathrm{Tr}\left(\sum_j P_j e^{-iH_j\tau}|\psi\rangle\langle\psi|e^{iH_j\tau}e^{-iHt/N}|\psi\rangle\langle\psi|e^{iHt/N}\right) \tag{B16}$$

$$= 1 - \sum_j P_j \langle\psi|e^{iHt/N}e^{-iH_j\tau}|\psi\rangle\langle\psi|e^{iH_j\tau}e^{-iHt/N}|\psi\rangle \tag{B17}$$

$$= 1 - \sum_j P_j |\langle\psi|\left(\mathbb{1} - iHt/N - \frac{H^2t^2/N^2}{2!} + \cdots\right)\left(\mathbb{1} + iH_j\tau - \frac{H_j^2\tau^2}{2!} + \cdots\right)|\psi\rangle|^2 \tag{B18}$$

$$= 1 - \sum_j P_j |\langle\psi|\left(\mathbb{1} - iHt/N + iH_j\tau + HH_j t\tau/N - \frac{H^2t^2/N^2}{2!} - \frac{H_j^2\tau^2}{2!} + \cdots\right)|\psi\rangle|^2 \tag{B19}$$

$$= 1 - \sum_j P_j |\langle\psi|\left(\mathbb{1} - iHt/N + iH_j t\lambda/N + HH_j t^2\lambda/N^2\right. \tag{B20}$$

$$\left. - \frac{H^2t^2/N^2}{2!} - \frac{H_j^2t^2\lambda^2/N^2}{2!} + \cdots\right)|\psi\rangle|^2 \tag{B21}$$

$$= 1 - \sum_j P_j(1 - 2\langle\psi|H|\psi\rangle\langle\psi|H_j|\psi\rangle t^2\lambda/N^2 + \langle\psi|H|\psi\rangle^2 t^2/N^2 + \langle\psi|H_j|\psi\rangle^2 t^2\lambda^2/N^2 + ct^2\cdots) \tag{B22}$$

$$= t^2 \sum_j P_j(2\langle\psi|H|\psi\rangle\langle\psi|H_j|\psi\rangle\lambda/N^2 - \langle\psi|H|\psi\rangle^2/N^2 - \langle\psi|H_j|\psi\rangle^2\lambda^2/N^2 - c\cdots) \tag{B23}$$

$$\leqslant t^2 \sum_j P_j(|2\langle\psi|H|\psi\rangle\langle\psi|H_j|\psi\rangle|\lambda/N^2 + |\langle\psi|H|\psi\rangle^2|/N^2 + |\langle\psi|H_j|\psi\rangle^2|\lambda^2/N^2 + |c|\cdots) \in \mathcal{O}(t^2). \tag{B24}$$

We use the cyclic property of the trace and the fact that ρ is a pure state $|\psi\rangle\langle\psi|$ and then replace the operator exponentials with their respective Taylor series expansions, keeping terms only up to t^2 throughout. In the final step, the modulus is taken and the triangle inequality applied to upper bound the sum. This is done for the purpose of ensuring the signs of the terms are "well behaved" such that $\mathcal{O}()$ notation can be aptly applied. As well, c is used for brevity where $c = \langle\psi|(HH_j\lambda/N^2 - \frac{H^2}{2!}/N^2 - \frac{H_j^2\lambda^2/N^2}{2!})|\psi\rangle + \text{H.c.}$ Here the notation $\mathcal{O}(t^2)$ is understood in the limit as $t \to 0$. ∎

Given the diamond distance bound in [15], the $\mathcal{O}(t^2)$ infidelity scaling is not surprising. Not only this, but recall from Eq. (B18), where the largest order in commutator error from the product formula was quadratic, and given that Qdrift applies no symmetrization strategy, we did not expect this to disappear. We would like to remind the reader that the shining feature of Qdrift is that the cost is independent of the number of terms in the Hamiltonian. Next, we examine how a Trotter-Suzuki channel behaves in the infidelity measure.

Proposition V.2 (Trotter-Suzuki infidelity time scaling). Given a Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, t)$, for a density matrix pure state ρ and time t, then the infidelity between the outputs of the channels $1 - \mathscr{F}(\mathcal{T}^{2k}(\rho, t)\mathcal{U}(\rho, t)) \in \mathcal{O}((t^{2k+1})^2)$.

Proof.

$$1 - \mathscr{F}(\mathcal{T}^{2k}(\rho, t), \mathcal{U}(\rho, t)) := 1 - \mathrm{Tr}(\mathcal{T}^{2k}(\rho, t)\mathcal{U}(\rho, t)) \tag{B25}$$

$$= 1 - \mathrm{Tr}(S_{2k}(-iHt)|\psi\rangle\langle\psi|S_{2k}(-iHt)^\dagger e^{-iHt}|\psi\rangle\langle\psi|e^{iHt}) \tag{B26}$$

$$= 1 - \langle\psi|e^{iHt}S_{2k}(-iHt)|\psi\rangle\langle\psi|S_{2k}(-iHt)^\dagger U(t)|\psi\rangle \tag{B27}$$

$$= 1 - |\langle\psi|e^{iHt}S_{2k}(-iHt)|\psi\rangle|^2 \tag{B28}$$

$$= 1 - |\langle\psi|(S_{2k}(-iHt)^\dagger \pm it^{2k+1}R_{2k+1} \mp it^{2k+3}R_{2k+3} + \cdots)S_{2k}(-iHt)|\psi\rangle|^2 \tag{B29}$$

$$= 1 - |\langle\psi|\mathbb{1} + \pm it^{2k+1}R_{2k+1}S_{2k}(-iHt)|\psi\rangle \mp it^{2k+3}R_{2k+3}S_{2k}(-iHt)|\psi\rangle + \cdots|^2 \tag{B30}$$

$$= 1 - |1 \pm it^{2k+1}\langle\psi|R_{2k+1}S_{2k}(-iHt)|\psi\rangle \mp it^{2k+3}\langle\psi|R_{2k+3}S_{2k}(-iHt)|\psi\rangle \pm \cdots|^2 \tag{B31}$$

$$\leqslant 1 - |1|^2 + (t^{2k+1})^2|\langle\psi|R_{2k+1}S_{2k}(-iHt)|\psi\rangle|^2 + \cdots \tag{B32}$$

$$= (t^{2k+1})^2|\langle\psi|R_{2k+1}S_{2k}(-iHt)|\psi\rangle|^2 + \cdots \in \mathcal{O}((t^{2k+1})^2). \tag{B33}$$

We again use the cyclic property of the trace and the fact that ρ is a pure state $|\psi\rangle\langle\psi|$. The evolution operator e^{-iHt} is rewritten in terms of the product formula and corrections at significant orders. R_n are remainder operators of order n of which we keep to order $2k + 3$. In the second last step, we use the triangle inequality and ignore terms of higher order. Keeping higher-order terms is unnecessary in that we cannot use them to generate terms of order $\leqslant (t^{2k+1})^2$. Once again, $\mathcal{O}((t^{2k+1})^2)$ is understood in the limit as $t \to 0$. ∎

This theorem provides a peculiar result. The expected error in the output of a Trotter-Suzuki channel from Sec. III B is squared by the nature of the infidelity calculation. Given that $\epsilon^2 \leqslant \epsilon \leqslant 1$, error squaring can lead to a much lower gate cost, and given that Trotter experiences this phenomenon while Qdrift does not, a numerical optimizer may heavily favor Trotter and the potential advantages from the composite approach will likely fade. This is the main issue with choosing infidelity as an error measure whilst hybridizing algorithms with different error scaling. The following two theorems show that the trace distance suffers no such side effect:

Proposition V.3 (Qdrift trace distance time scaling). Given a Qdrift channel $\mathcal{Q}(\rho, t)$ and the standard evolutionary channel $\mathcal{U}(\rho, {}^t/_N)$, for an arbitrary density matrix ρ and time t/N, then the trace distance between the outputs of the channels $\mathscr{T}(\mathcal{Q}(\rho, t), \mathcal{U}(\rho, t/N)) \in \mathcal{O}(t^2)$.

Proof.

$$\mathscr{T}(\mathcal{Q}(\rho, t), \mathcal{U}(\rho, {}^t/_N)) := \mathrm{Tr}\left[\sqrt{(\mathcal{Q}(\rho, t) - \mathcal{U}(\rho, {}^t/_N))(\mathcal{Q}(\rho, t) - \mathcal{U}(\rho, t/N))^\dagger}\right] \tag{B34}$$

$$= \mathrm{Tr}\left[\sqrt{(\mathcal{Q}(\rho, t) - \mathcal{U}(\rho, {}^t/_N))^2}\right] \tag{B35}$$

$$= \mathrm{Tr}\left[\sqrt{\left(\left(\sum_j P_j e^{-iH_j\tau}\rho e^{iH_j\tau}\right) - (e^{-iHt/N}\rho e^{iHt/N})\right)^2}\right] \tag{B36}$$

$$= \mathrm{Tr}\left[\sqrt{\left(\sum_j P_j(\mathbb{1} - iH_j\tau - H_j^2\tau^2 + \cdots)\rho(\mathbb{1} + iH_j\tau - H_j^2\tau^2 + \cdots) - (e^{iHt/N}\rho e^{-iHt/N})\right)^2}\right] \tag{B37}$$

$$= \mathrm{Tr}\left[\sqrt{\left(\sum_j P_j(\rho + i\lambda t/N[H_j, \rho] + \lambda^2 t^2/N^2(H_j\rho H_j - i\{H_j, \rho\}) + \cdots) - (e^{iHt/N}\rho e^{-iHt/N})\right)^2}\right] \tag{B38}$$

$$= \mathrm{Tr}\left[\sqrt{\left(\left(\rho + \sum_j \frac{h_j}{\lambda}(i\lambda t/N[H_j,\rho] + \lambda^2 t^2/N^2(H_j\rho H_j - i\{H_j,\rho\}) + \cdots) - (e^{iHt/N}\rho e^{-iHt/N})\right)^2}\right]$$

(B39)

$$= \mathrm{Tr}\left[\sqrt{\left(\rho + it/N[H,\rho] + \sum_j \frac{h_j}{\lambda}\left(\lambda^2 t^2/N^2\left(H_j\rho H_j - i\{H_j^2,\rho\}\right) + \cdots\right) - (e^{iHt/N}\rho e^{-iHt/N})\right)^2}\right]$$

(B40)

$$= \mathrm{Tr}\left[\sqrt{\sum_j \frac{h_j}{\lambda}\left(\lambda^2 t^2/N^2\left(H_j\rho H_j - i\{H_j^2,\rho\}\cdots\right) + (t^2(H\rho H + i\{H^2,\rho\}) + \cdots)\right)^2}\right]$$

(B41)

$$= t^2\mathrm{Tr}\left[\sqrt{(R(t))^2}\right] \in \mathcal{O}(t^2).$$

(B42)

Here we expand the Qdrift channel with the Taylor series of the exponential operators and show that the terms of this series cancel with that of the Taylor series of $\mathcal{U}_N$ up to order t^2. To achieve this we simply use the definitions of $P_j = h_j/\lambda$ and $\sum_j H_j = H$. In the last line $R(t)$ is considered a remainder operator, which is only a function of polynomials t^n, $n \geqslant 0$. The notation $\mathcal{O}(t^2)$ is understood in the limit as $t \to 0$.

Proposition V.4 (Trotter-Suzuki trace distance time scaling). Given a Trotter-Suzuki channel $\mathcal{T}^{2k}(\rho,t)$ and the standard evolutionary channel $\mathcal{U}(\rho,t)$, for an arbitrary density matrix ρ and time t, then the trace distance between the outputs of the channels $\mathscr{T}(\mathcal{T}^{2k}(\rho,t),\mathcal{U}(\rho,t)) \in \mathcal{O}(t^{2k+1})$.

Proof.

$$\mathscr{T}(\mathcal{T}^{2k}(\rho,t),\mathcal{U}(\rho,t))$$

$$:= \mathrm{Tr}\left[\sqrt{(\mathcal{T}^{2k}(\rho,t) - \mathcal{U}(\rho,t))(\mathcal{T}^{2k}(\rho,t) - \mathcal{U}(\rho,t))^\dagger}\right]$$

(B43)

$$= \mathrm{Tr}\left[\sqrt{(\mathcal{T}^{2k}(\rho,t) - \mathcal{U}(\rho,t))^2}\right]$$

(B44)

$$= \mathrm{Tr}\left[\sqrt{((S_{2k}(-iHt)\rho S_{2k}(-iHt)^\dagger) - (e^{-iHt}\rho e^{iHt}))^2}\right]$$

(B45)

$$= \mathrm{Tr}\left[\sqrt{((S_{2k}(-iHt)\rho S_{2k}(-iHt)^\dagger) - ((S_{2k}(-iHt)^\dagger \pm it^{2k+1}R_{2k+1} + \cdots)\rho(S_{2k}(-iHt)^\dagger \mp it^{2k+1}R_{2k+1}^\dagger + \cdots))^2)}\right]$$

(B46)

$$= \mathrm{Tr}\left[\sqrt{((\mp it^{2k+1}S_{2k}(-iHt)\rho R_{2k+1}^\dagger \pm it^{2k+1}R_{2k+1}S_{2k}(-iHt)^\dagger + \cdots))^2}\right]$$

(B47)

$$= t^{2k+1}\mathrm{Tr}\left[\sqrt{(R(t))^2}\right] \in \mathcal{O}(t^{2k+1}).$$

(B48)

Here expand the exponential operators in the evolutionary channel $\mathcal{U}$ in terms of their error terms for some general $2k$th-order Trotter-Suzuki channel, where R_n are remainder operators of order n. As the algebra proceeds we only keep terms of order $\leqslant t^{2k+1}$ throughout. Keeping higher-order terms is unnecessary in that we cannot use them to generate terms of order $\leqslant t^{2k+1}$. In the last line $R(t)$ is considered a general remainder operator, which is only a function of polynomials t^n, $n \geqslant 0$. Once again, $\mathcal{O}(t^{2k+1})$ is understood in the limit as $t \to 0$. ∎

(a) QDrift Infidelity

(b) Trotter Infidelity

(c) TS Order $2k = 2$ Infidelity

(d) QDrift Trace Distance

(e) Trotter Trace Distance

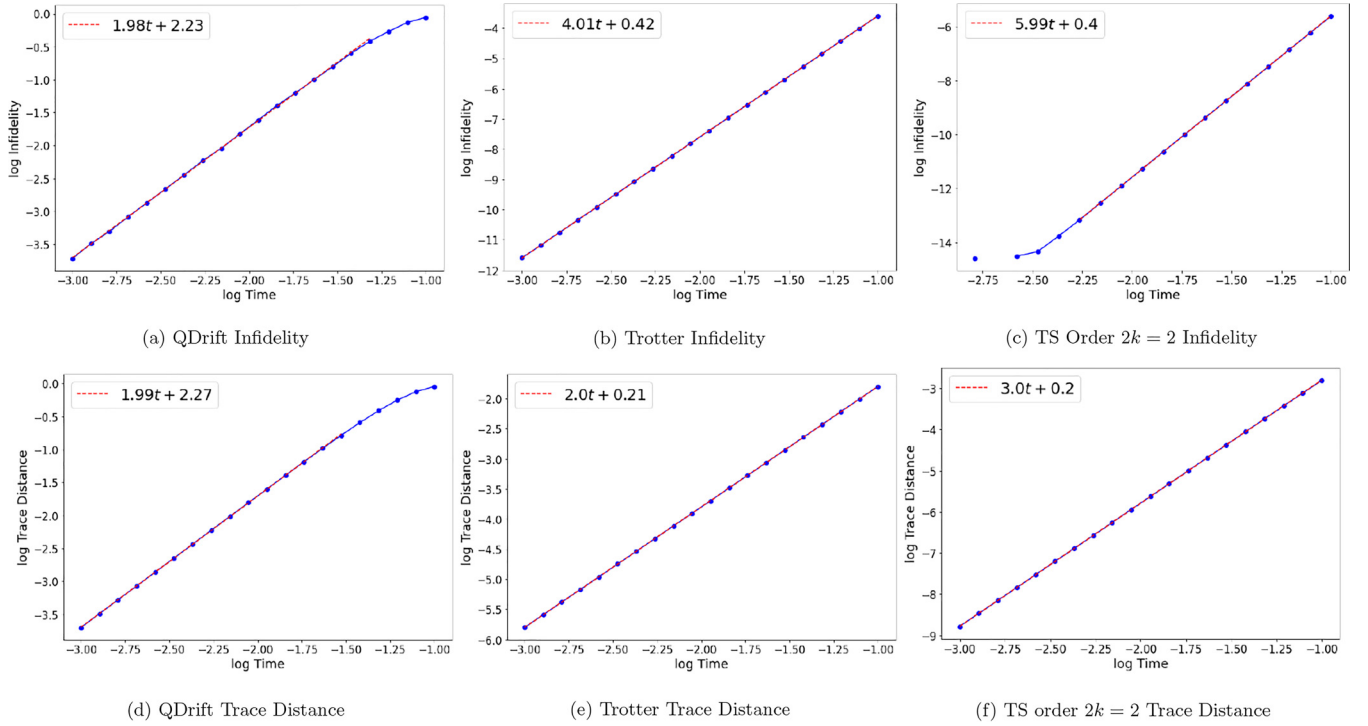(f) TS order $2k = 2$ Trace Distance

FIG. 12. Numerical plots of log infidelity and trace distance scaling with log time for both Trotter-Suzuki and Qdrift channels. Plots are all constructed using one iteration or sample respectively. Note that the slope of the plots follow the analytical predictions exactly, with the infidelity squaring the expected Trotter error in (b) and (c). Note that in (c) we run into machine precision error at small times and for long times in (a) and (d) Qdrift no longer converges. The plots were compute using 100 Monte Carlo samples for Qdrift infidelity, and exact density matrices elsewhere for a graph Hamiltonian with five spins.
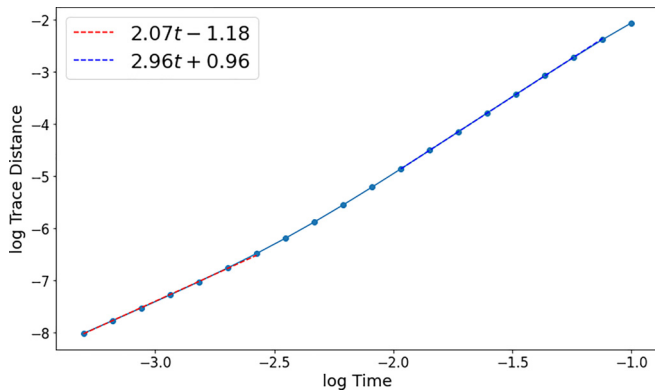

FIG. 13. Trace distance time scaling for a composite channel in inner order $2k = 2$. This plot emits an interesting structure in that the time dependence of the trace distance goes through a "phase transition" between Qdrift (red) and second-order Trotter-Suzuki (blue) dominated regions. Given that Qdrift excels in short time simulations and whereas the higher-order Trotter-Suzuki dominates at longer simulation times this is not entirely unexpected.

There are no surprises with the trace distance error as it yields the expected time scaling from the results of Appendix B 1 b. Numerical investigations also support our proofs (see Fig. 12). Therefore, while it may be more expensive to compute for larger systems, this makes it favourable to work with as it provides a more "fair" measure, treating the algorithms on more equal footing. Thus, it is expected that the composite channel will appear more performant using this metric. In addition, using this framework requires no sampling, which means that monotonicity of the cost will be guaranteed with respect to channel iterations. This will prove useful in the cases in which we choose to optimize over possible partitions with respect to the cost. Now, out of interest, we also wish to numerically investigate the trace distance of the composite channel. Analytically, this is likely a messy problem given the partition, but intuitively we expect the scaling to be some linear combination $\alpha t^2 + \beta t^3 + \gamma$ with the slope of the log plot being between 2 and 3 (see Fig. 13).

[1] R. P. Feynman, Simulating physics with computers, in *Feynman and Computation* (CRC Press, Boca Raton, FL, 2018), pp. 133–153.

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

[3] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, Exponential algorithmic speedup by a quantum walk, in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 59–68 (2003).

[4] R. Babbush, D. W. Berry, R. Kothari, R. D. Somma, and N. Wiebe, Exponential quantum speedup in simulating coupled classical oscillators, Phys. Rev. X **13**, 041041 (2023).

[5] A. Yu Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.

[6] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2002).

[7] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proc. Natl. Acad. Sci. USA **114**, 7555 (2017).

[8] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chem. Rev. **120**, 12685 (2020).

[9] 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).

[10] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, Phys. Rev. Res. **3**, 033055 (2021).

[11] D. W. Berry, A. M. Childs, Y. Su, X. Wang, and N. Wiebe, Time-dependent Hamiltonian simulation with l1-norm scaling, Quantum **4**, 254 (2020).

[12] S. P. Jordan, K. S. M. Lee, and J. Preskill, Quantum algorithms for quantum field theories, Science **336**, 1130 (2012).

[13] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, Efficient quantum algorithms for simulating sparse Hamiltonians, Commun. Math. Phys. **270**, 359 (2007).

[14] 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).

[15] E. Campbell, Random compiler for fast Hamiltonian simulation, Phys. Rev. Lett. **123**, 070503 (2019).

[16] K. Wan, M. Berta, and E. T. Campbell, Randomized quantum algorithm for statistical phase estimation, Phys. Rev. Lett. **129**, 030503 (2022).

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

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

[19] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Simulating Hamiltonian dynamics with a truncated taylor series, Phys. Rev. Lett. **114**, 090502 (2015).

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

[21] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 193–204 (2019).

[22] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, Grand unification of quantum algorithms, PRX Quantum **2**, 040203 (2021).

[23] S. Jin and X. Li, A partially random trotter algorithm for quantum Hamiltonian simulations, arXiv:2109.07987.

[24] M. Hagan and N. Wiebe, Composite quantum simulations, Quantum **7**, 1181 (2023).

[25] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Quantum Monte Carlo methods for nuclear physics, Rev. Mod. Phys. **87**, 1067 (2015).

[26] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, Rev. Mod. Phys. **83**, 349 (2011).

[27] W. von der Linden, A quantum Monte Carlo approach to many-body physics, Phys. Rep. **220**, 53 (1992).

[28] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[29] Y. Yan and D. Blume, Path integral Monte Carlo ground state approach: Formalism, implementation, and applications, J. Phys. B: At., Mol. Opt. Phys. **50**, 223001 (2017).

[30] M. Boninsegni, N. V. Prokof'Ev, and B. V. Svistunov, Worm algorithm and diagrammatic Monte Carlo: A new approach to continuous-space path integral Monte Carlo simulations, Phys. Rev. E **74**, 036701 (2006).

[31] J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Monte Carlo simulations of one-dimensional fermion systems, Phys. Rev. B **26**, 5033 (1982).

[32] D. M. Ceperley, Path integrals in the theory of condensed helium, Rev. Mod. Phys. **67**, 279 (1995).

[33] C. Chakravarty, M. C. Gordillo, and D. M. Ceperley, A comparison of the efficiency of Fourier- and discrete time-path integral Monte Carlo, J. Chem. Phys. **109**, 2123 (1998).

[34] J. Haah, M. B. Hastings, R. Kothari, and Guang H. Low, Quantum algorithm for simulating real time evolution of lattice Hamiltonians, SIAM J. Comput. **52**, 6 (2023).

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

[36] D. Wecker, B. Bauer, B. K. Clark, M. B. Hastings, and M. Troyer, Gate-count estimates for performing quantum chemistry on small quantum computers, Phys. Rev. A **90**, 022305 (2014).

[37] D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doherty, and M. Troyer, The Trotter step size required for accurate quantum simulation of quantum chemistry, Quantum Inf. Comput. **15**, 361 (2014).

[38] O. Kiss, M. Grossi, and A. Roggero, Importance sampling for stochastic quantum simulations, Quantum **7**, 977 (2023).

[39] C.-F. Chen, H.-Y. Huang, R. Kueng, and J. A. Tropp, Concentration for random product formulas, PRX Quantum **2**, 040305 (2021).

[40] K. Nakaji, M. Bagherimehrab, and A. Aspuru-Guzik, qSWIFT: High-order randomized compiler for Hamiltonian simulation, arXiv:2302.14811.

[41] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby *et al.*, Openfermion: The electronic structure package for quantum computers, Quantum Sci. Technol. **5**, 034014 (2020).

[42] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, Proc. Natl. Acad. Sci. USA **115**, 9456 (2018).

[43] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. GSL Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer

using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).

[44] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Simulation of electronic structure Hamiltonians using quantum computers, Mol. Phys. **109**, 735 (2011).

[45] M. Suzuki, Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations, Phys. Lett. A **146**, 319 (1990).

[46] N. Hatano and M. Suzuki, Finding exponential product formulas of higher orders, in *Quantum Annealing and Other Optimization Methods* (Springer, Berlin, 2005), pp. 37–68.

[47] B. Nachtergaele and R. Sims, Lieb-Robinson bounds and the exponential clustering theorem, Commun. Math. Phys. **265**, 119 (2006).

[48] B. Nachtergaele and R. Sims, Much ado about something: Why Lieb-Robinson bounds are useful, arXiv:1102.0835.

[49] M. Suzuki, Quantum Monte Carlo methods—Recent developments, Physica A **194**, 432 (1993).

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

[51] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma *et al.*, Pyscf: The python-based simulations of chemistry framework, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1340 (2018).

[52] C. Mc Keever and M. Lubasch, Classically optimized Hamiltonian simulation, Phys. Rev. Res. **5**, 023146 (2023).

[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, Scikit-learn: Machine learning in python, J. Machine Learning Res. **12**, 2825 (2011).

[54] M. Pocrnic and M. Hagan, Repository containing the library used to produce all data in this paper, https://github.com/mattp97/Trotter-Qdrift-Simulation.