Composite Quantum Simulations

Matthew Hagan[1] and Nathan Wiebe[2,3,4]

[1]Department of Physics, University of Toronto, Toronto ON, Canada

[2]Department of Computer Science, University of Toronto, Toronto ON, Canada

[3]Pacific Northwest National Laboratory, Richland Wa, USA

[4]Canadian Institute for Advanced Study, Toronto ON, Canada

In this paper we provide a framework for combining multiple quantum simulation methods, such as Trotter-Suzuki formulas and QDrift into a single Composite channel that builds upon older coalescing ideas for reducing gate counts. The central idea behind our approach is to use a partitioning scheme that allocates a Hamiltonian term to the Trotter or QDrift part of a channel within the simulation. This allows us to simulate small but numerous terms using QDrift while simulating the larger terms using a high-order Trotter-Suzuki formula. We prove rigorous bounds on the diamond distance between the Composite channel and the ideal simulation channel and show under what conditions the cost of implementing the Composite channel is asymptotically upper bounded by the methods that comprise it for both probabilistic partitioning of terms and deterministic partitioning. Finally, we discuss strategies for determining partitioning schemes as well as methods for incorporating different simulation methods within the same framework.

1 Introduction

The simulation of quantum systems remains one of the most compelling applications for future digital quantum computers [1, 2, 3, 4, 5, 6]. As such, there are a plethora of algorithm options for compiling a unitary evolution operator $U(t) = e^{-iHt}$ to circuit gates [7, 8, 9, 10, 11, 12, 13, 14]. Some of the simplest such algorithms are product formulas in which each term in a Hamiltonian $H = \sum_i h_i H_i$ is implemented as $e^{iH_i t}$. A product formula is then a particular sequence of these gates that approximates the overall operator $U(t)$. Two of the most well known product formula include Trotter-Suzuki Formulas [8, 15, 10, 16] and the QDrift protocol in which terms are sampled randomly [14, 17]. These two approaches are perhaps the most popular ancilla-free simulation methods yet discovered.

One of the main drawbacks of Trotter-Suzuki formulas is that each term in the Hamiltonian has to be included in the product formula regardless of the magnitude of the term. This leads to a circuit with a depth that scales at least linearly with the number of terms in H, typically denoted L. QDrift avoids this by randomly choosing which term to implement next in the product formula according to an importance sampling scheme in which higher weight terms have larger probabilities. The downside to QDrift is that it has the same asymptotic scaling with t/ϵ as a first-order Trotter formula, meaning it is outperformed at large t/ϵ by even a second-order Trotter formula.

In this paper we present a framework for combining simulation channels in a way that allows one to flexibly interpolate the gate cost tradeoffs between the individual channels. The primary example we study is the composition of Trotter-Suzuki and QDrift channels. This is motivated in some part as an effort to extend randomized compilers to include conditional probabilities and in some part to encapsulate progress in chemistry simulations of dropping small weight terms or shuffling terms around different time steps [18]. This latter concept was first developed with the idea of "coalescing" terms into "buckets" by Wecker et al. [18] and further explored by Poulin et al. [19]. They showed that grouping terms of similar sizes together to be skipped during certain Trotter steps led to negligible increases in error and reduced gate counts by about a factor of 10. Similar

improvements are also seen in the randomized setting of [20]. In this work we extend on these ideas by placing a specific set of terms into a Trotter partition and the rest in a QDrift partition. This simple division can then be studied analytically and we are able to provide sufficient conditions on asymptotic improvements over completely Trotter or completely QDrift channels. Although we are not able to develop the idea of conditional samples in QDrift protocols, our procedure can be viewed as a specific subset of what a generic Markovian QDrift would look like. We briefly mention these generalizations in Section 7.

Recent approaches have sought to use the advantages of randomized compilation as a subset of an overall simulation, such as the hybridized scheme for interaction picture simulations [21]. What separates these two works is that our approach offers a more flexible approach for generic time-independent simulation problems whereas the hybridized schemes are specifically tailored to taking advantage of the time dependence introduced by moving to an interaction picture. As such, the hybridized approach achieves asymptotic advantages when the size of the interaction picture term dominates the overall Hamiltonian. This typically occurs in instances in which the size of an operator is unbounded, which can occur in lattice field theory simulations or constrained systems. The way the hybridized scheme in [21] works is via a "vertical" stacking of simulation channels, for example one channel to handle the Interaction Picture rotations and then other channels on top of this to simulate the time-dependence it generates on the remaining Hamiltonian terms. Our work instead remains in the Schrodinger time evolution picture and we perform a "horizontal" stacking of simulation techniques. By horizontal we mean for a given simulation time we split the Hamiltonian up into (potentially) disjoint partitions and simulate each partition for the full simulation time but with different techniques, such as Trotter or QDrift. These techniques allow us to achieve asymptotic improvements over either method for a loose set of assumptions.

There are two other simulation techniques that have been proposed recently that have a similar interpolation behavior between QDrift and Trotter channels. The first of these methods is the SparSto, or Stochastic Sparsification, technique by Ouyang, White, and Campbell [22]. The procedure [22] randomly sparsifies the Hamiltonian and performs a randomly ordered first-order Trotter formula on the sampled Hamiltonian. They construct these probabilities such that the expected Hamiltonian is equal to the Hamiltonian being simulated. They then fix the expected number of oracle queries of the form $e^{iH_i t'}$ and give diamond distance bounds on the resulting channel error. The claim for interpolation between Trotter and QDrift is that one can fix the expected number of gates to be 1 for each time step, in which case the sparsification mimics QDrift, whereas if no sparsification is performed then the channel is simply implementing Trotter. They show that this allows for one to have reduced simulation error up to an order of magnitude on numerically studied systems as compared to Trotter or QDrift. One downside to these techniques is that the number of gates applied is a random variable, so making gate cost comparisons is rather difficult especially considering that no tail bounds on high gate cost sampled channels are provided. In [22] they prefer to fix the expected gate cost and analyze the resulting diamond norm error. In contrast, our procedures directly implement both QDrift and Trotter channels and have a fixed, deterministic gate cost.

The second method of note with both QDrift and Trotter behavior is that of Jin and Li [23]. They develop an analysis of the variance of a unitary consisting of a first-order Trotter sequence followed by a QDrift channel. They focus on bounding the Mean Squared Error (MSE) of the resulting channel and use a simple partition of the Hamiltonian terms based on spectral norm. Their partitioning scheme places all terms below some cutoff into the first-order Trotter sequence and all terms above the cutoff into the QDrift channel. Their main results show an interpolation of the MSE between 0 when the partitioning matches a solely Trotter channel and matching upper bounds for QDrift when all terms are randomly sampled. This work goes beyond the results from Jin and Li by providing an analysis of the diamond distance between an ideal evolution and our implemented channel, which is more useful analytically than the MSE, as well as providing upper bounds on the number of gates needed in an implementation to meet this diamond distance. In addition our work remains independent of specific partitioning schemes as much as possible and instead places restrictions on which partitions achieve improvements. In the interest of practicality we do show methods for partitioning that can be useful in both the first-order and higher-order Trotter cases. Specifically for higher-order Trotter formulas we give a probabilistic partitioning scheme that is easily computable and matches gate cost upper bounds in the extreme limits as our

probabilities saturate the QDrift and Trotter limits.

The rest of the paper is organized as follows. We first provide a brief summary of the main results in Section 2. After reviewing known results and notation in Section 3, we explore methods for creating Composite channels using First-Order Trotter Formulas with QDrift in Section 4 as a warmup. This is broken down into three parts in which we find the gate cost for an arbitrary partition, we then give a method for producing a good partitioning, and then we analyze conditions in which a Composite channel can beat either first-order Trotter or QDrift channels. In Section 5 we then extend this framework to more general higher-order Trotter Formulas. This section mirrors the organization of the first-order Trotter section, namely we find the cost of an arbitrary partition, we give a method for producing a partition efficiently, and then we analyze when one could see improvements over the constituent channels. Finally, in Section 7 we discuss extensions to this model that allow a flexible interpolation between various types of product formulas that could be leveraged numerically.

2 Main Results

In this section we summarize the gate cost performance of a higher-order Composite channel, the probabilistic partitioning scheme developed, and the conditions needed for a partitioning scheme to satisfy in order to expect asymptotic improvements over QDrift or Trotter. These results are motivated and proved throughout Section 5. We do not state the first-order Composite channel results here as they are more specific and do not achieve as strong asymptotic improvements as the higher-order channels.

Our first theorem, presented in Section 5.1, gives an upper bound on the number of queries to oracles of the form $e^{iH_i t'}$ to implement a Composite channel with desired error. We provide a conceptually simple packaging of this bound in terms of the number of oracle queries that would be needed to perform this same simulation if either Trotter or QDrift had been used alone.

Theorem 2.1 (Gate Cost for Higher-Order Composite Channel). *Given a time t, error bound ϵ, partitioned Hamiltonian $H = A + B$, and let $\widetilde{\mathcal{U}}^{(2k)}$ denote the higher-order Composite channel to approximate the exact unitary evolution $\mathcal{U}(t)$. By using r iterations of $\widetilde{\mathcal{U}}^{(2k)}(t/r)$ we can satisfy the error requirement $\left\lVert \mathcal{U}(t) - \widetilde{\mathcal{U}}^{(2k)}(t/r)^{\circ r} \right\rVert_\diamond \leq \epsilon$ by using at most the following number of operator exponentials*

$$
\begin{aligned}
&C_{comp}(A, B, t, \epsilon, 2k) \\
&\leq \Upsilon(\Upsilon L_A + N_B) \left\lceil \frac{(\Upsilon t)^{1+1/2k} 4^{1/2k}}{\epsilon^{1/2k}} \left(\frac{\Upsilon \alpha_{comm}(A, 2k) + \alpha_{comm}(\{A, B\}, 2k)}{2k+1} \right)^{1/2k} + \frac{4\Upsilon \lambda_B^2 t^2}{N_B \epsilon} \right\rceil.
\end{aligned}
\tag{1}
$$

By making the definition $q_B := \frac{\alpha_{comm}(B, 2k)}{\alpha_{comm}(H, 2k)}$ and utilizing the upper bounds from Theorems 2 and 4, where C_{Trott} and C_{QD} below are upper bounds on the number of operator exponentials required in the Trotter and QDrift channels, we can write the cost upper bound as

$$
C_{comp} \leq \Upsilon(\Upsilon L_A + N_B) \left\lceil C_{Trott}(H, t, \epsilon, 2k) \frac{(1 - q_B)^{1/2k}}{\Upsilon^{1-1/2k} L} + C_{QD}(H, t, \epsilon) \frac{\Upsilon}{N_B} \frac{\lambda_B^2}{\lambda^2} \right\rceil.
\tag{2}
$$

The next result, presented in Section 5.2, gives an easily computable probabilistic partitioning scheme for an arbitrary Hamiltonian. It is based around a probability for each term to end up in either the QDrift partition or the Trotter partition that can be viewed as an importance sampling routine on the inverse spectral norms $\frac{1}{h_i}$. This distribution was motivated by upper bounding the expected QDrift error, which is expressed in terms of λ_B, in terms of the Trotter error as intuitively Trotter formulas of higher-orders are more accurate than QDrift channels at smaller times. One feature of note from this lemma is that it introduces a *lower* bound on the number of QDrift samples which is required for the probabilities to remain nonnegative.

Lemma 2.1 (Probabilistic Partitioning Scheme). *For a composite simulation of H for time t and error ϵ, let p_i denote the probability of placing term $h_i H_i$ into the Trotter partition of Composite channel. we have that choosing*

$$1 - p_i = \min\left\{ \frac{\lambda}{h_i L} \left(\sqrt{N_B \left(\frac{\epsilon}{\lambda t}\right)^{1-1/2k} \left(\frac{2k+\Upsilon}{2k+1}\right)^{1/2k} \frac{\Upsilon^{1/2k}}{2^{1-1/k}}} - 1 \right), 1 \right\} =: \min\left\{ \frac{\chi}{h_i}, 1 \right\}, \quad (3)$$

along with choosing the number of QDrift samples N_B to satisfy

$$N_B \geq \left(\frac{\lambda t}{\epsilon}\right)^{1-1/2k} \left(\frac{2k+1}{2k+\Upsilon}\right)^{1/2k} \frac{2^{1-1/k}}{\Upsilon^{1/2k}} \quad (4)$$

guarantees the following:

1. *$p_i \in [0, 1]$,*

2. *the expectation value of the coefficients in the QDrift partition satisfies*

$$\frac{\mathbb{E}[\lambda_B]}{\lambda} \leq \frac{1}{2} \sqrt{\left(\frac{4k+2\Upsilon}{2k+1}\right)^{1/2k} (2\Upsilon)^{1+1/2k}} \sqrt{N_B \left(\frac{\epsilon}{\lambda t}\right)^{1-1/2k}}.$$

This bound follows from a rigorous interpretation of making the QDrift and Trotter errors approximately equivalent.

Our final result, proved in Section 5.3.1, of significant importance are the conditions on the parameters of a given partition to provide asymptotic cost improvements of a Composite channel over it's constituent channels, Trotter or QDrift. This theorem is relevant when one is considering a family of Hamiltonians and has a provided partitioning scheme for each. We then study how these partitions can give rise to asymptotically superior simulation techniques over Trotter and QDrift as the number of terms in the Hamiltonian is taken to infinity. Although the situation of an infinite family of Hamiltonians with a provided partitioning scheme may not arise in practice, analyzing the performance of these families provides intuition as for when a Composite channel should be able to yield significant savings.

Theorem 2.2 (Conditions for Composite Channel Improvements). *Assume a Hamiltonian H along with a partitioning scheme to generate A and B that varies with L. For a simulation time t with a desired diamond distance error at most ϵ, let $\beta > 0$ be a number such that $C_{QD} = C_{Trott}^{\beta}$. There exist asymptotic regimes for the parameters $L_A, \lambda_B,$ and N_B such that*

$$C_{Comp} \in o(\min\{C_{Trott}, C_{QD}\}),$$

outlined below for the cases when $C_{QD} \geq C_{Trott}$ ($\beta \geq 1$) and $C_{QD} < C_{Trott}$ ($0 < \beta < 1$).

For the case when $\beta > 1$, indicating Trotter uses fewer queries than QDrift, if the parameters λ_B, L_A, and number of QDrift samples N_B satisfy the following

1. *$L_A (1 - q_B)^{1/2k} \in o(L)$ where $q_B = \frac{\alpha_{comm}(B, 2k)}{\alpha_{comm}(H, 2k)}$,*

2. *$\lambda_B \in o\left(\lambda^{1/\beta} \left(\frac{\sqrt{\epsilon}}{t}\right)^{1-1/\beta} \right)$,*

3. *$N_B \in \Omega(L_A)$ and $N_B \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$,*

then we have that $C_{comp} \in o(C_{Trott})$.

If instead $0 < \beta < 1$, indicating $C_{QD} < C_{Trott}$, and the parameters λ_B, L_A and N_B satisfy the following

1. *the total number of terms in the Trotter partition satisfies*

$$L_A \in o\left(L^{1/\beta} \left(\frac{\epsilon^{1-1/\beta}}{t^{(2k+1)(1-1/\beta)}} \frac{\alpha_{comm}^{1/\beta}(H)}{\alpha_{comm}(A) + \alpha_{comm}(\{A, B\})} \right)^{1/2k} \right), \quad (5)$$

2. $\lambda_B \in o(\lambda)$,

3. and $N_B \in \Theta(L_A)$,

then we have that $C_{comp} = o(C_{QD})$. Note that for $\beta = 1$ exactly, the conditions on λ_B and L_A are the same for both cases: $L_A \in o(L)$ and $\lambda_B \in o(\lambda)$. The conditions for N_B are satisfied by $N_B \in \Theta(L_A)$. If these conditions are satisfied for all ranges of β then $C_{comp} \in o\left(\min\left\{C_{QD}, C_{Trott}\right\}\right)$.

Conditions for Advantage for Simulations with Composite Channels

To help give an idea of when a Composite channel would be most effective we briefly and informally discuss the intuition provided throughout the paper of when one should expect to see cost savings from a Composite channel. The most straightforward tool to build intuition is the situation in which t and ϵ are such that $C_{Trott}(H, t, \epsilon) = C_{QD}(H, t, \epsilon)$. We will refer to this particular ratio of t and ϵ as the cost crossover time. In this setting, all one has to do is find a partitioning scheme such that the number of Trotter terms is much smaller than the total number of terms ($L_A \in o(L)$) and that the spectral norm of the remaining terms is negligible compared to the overall sum ($\lambda_B \in o(\lambda)$). Note that this last expression can be rewritten: $\lambda - \lambda_A \in o(\lambda) \implies 1 - \lambda_A/\lambda \in o(1)$. Combining these pieces of information tells us that the Composite framework should provide the best improvements whenever a vanishingly small number of terms contain almost all of the "spectral weight" of the Hamiltonian and have negligible commutator structure.

When considering the task of how to partition a given Hamiltonian we can unfortunately not offer more insight beyond the intuition provided above. This is likely to be a very difficult problem that will have to take advantage of domain specific knowledge in regards to the provided Hamiltonian. A possible starting point to constructing partitions could be gleaned from the effectiveness of our provided probabilistic partitioning scheme when applied to an exponentially decaying Hamiltonian. This probabilistic scheme depends solely on the spectral norm of each term, so a useful starting point for constructing deterministic partitions could be to pick a cutoff weight in which stronger terms are assigned to Trotter and lighter ones to QDrift. If a cutoff can be found that is small compared to the total norm λ and only has a small percentage of terms (roughly $\log_2(L)/L$ would align with Theorem 15), then this should be enough to see significant improvements in cost for simulation times near the cost crossover time. We note that many "toy" chemistry and material science models, such as Jellium [24] for interacting electrons and Hydrogen chains [1] for molecules, exhibit these kinds of strong decays in spectral norms.

Another useful consideration is the question of when a given simulation is likely to *not* see significant savings from a composite approach. When considering Hamiltonian norms and commutator structure the worst case scenario is one in which each term has equal spectral norm and there is a cyclic commutator behavior (e.g. angular momentum $[J_i, J_j] = \epsilon_{ijk} J_k$) that is the same magnitude at any order. We also note that the ability to find useful partitions depends heavily on the simulation time to error ratio t/ϵ. As this ratio tends to either 0 or ∞ the ability to find an economical partitioning vanishes. For example if one needs a very accurate simulation ($\epsilon \to 0$) then any terms that are put in a QDrift partition will require too many samples to meet the lower error budget compared to just putting the term in a higher-order Trotter formula. At the other extreme in which one has a higher error or very short time t, then any error savings by putting terms into a Trotter partition are likely to waste gates when sampling these terms with QDrift would suffice.

3 Preliminaries

In this section we will first introduce the necessary notation we will use and then state known results about Trotter-Suzuki formulas and QDrift channels. We work exclusively with time-independent Hamiltonians H in a 2^n dimensional Hilbert space $\mathcal{H}$. We also assume that H consists of L terms $H = \sum_{i=1}^{L} h_i H_i$ where h_i represents the spectral norm of the term, H_i is a Hermitian operator on $\mathcal{H}$, and $||H_i|| = 1$. Note without loss of generality we can always assume $h_i \geq 0$, as we can always absorb the phase into the operator H_i itself. We use $||M||$ to refer to the spectral norm, or the magnitude of the largest singular value of M. We use λ to refer to the sum of h_i, namely

$\lambda = \sum_i h_i$. We will also use subscripts on lambda, such as λ_A to refer to sums of subsets of the terms of H. For example, if $H = 1H_1 + 2H_2 + 3H_3$ and $G = 1H_1 + 2H_2$, then $\lambda = 6$ and $\lambda_G = 3$.

We use $U(t)$ to refer to the unitary operator e^{iHt} and $\mathcal{U}(t)$ to refer to the channel $U(t)\rho U(t)^\dagger$. We will be particularly concerned with simulations of subsets of the terms of H, which we denote as follows. We typically work with a partition of H into two matrices $H = A + B$, and we let $A = \sum_i a_i A_i$ and $B = \sum_j b_j B_j$, where we have simply relabeled the relevant h_i and H_i into a's, b's, A's, and B's. This allows us to define the exact unitary time evolution operators $U_A(t) = e^{iAt}$ and channels $\mathcal{U}_A(t) = U_A(t)\rho U_A(t)^\dagger$, similarly defined for B. As we will be working with approximations to these channels, any operator or channel with a tilde represents an "implemented" channel, for example a first-order Trotter operator for A would look like $\widetilde{U_A}(t) = e^{ia_1 A_1 t} \ldots e^{ia_L A_L t}$. We avoid using $\mathcal{E}$ to represent an approximation or product formula as $\mathcal{E}$ will be used for error channels.

Although much of the literature for Trotter-Suzuki formulas is written in terms of unitary operators $U = e^{iHt}$ acting on state vectors $|\psi\rangle$ for our purposes it will prove most natural to consider a product formula as a channel $\mathcal{U} = e^{iHt}\rho e^{-iHt}$ acting on a density matrix ρ. After reviewing known results on unitary constructions of Trotter-Suzuki formulas we give a straightforward extension of these bounds to channels.

3.1 Trotter-Suzuki Formulas

Definition 1 (Trotter-Suzuki Decomposition [25])**.** *Given a Hamiltonian H, let $U_{TS}^{(1)}(\rho; t)$ denote the first-order Trotter-Suzuki time evolution operator, which is defined as*

$$U_{TS}^{(1)}(t) := e^{ih_L H_L t} \ldots e^{ih_1 H_1 t} = \prod_{i=1}^{L} e^{ih_i H_i t}. \tag{6}$$

Note that the ordering of the factors in the notation $\prod_{i=1}^{L}$ is defined to start from the rightmost operator and end at the leftmost. Following this we can define the second-order Trotter-Suzuki time evolution operator as

$$U_{TS}^{(2)}(\rho; t) := e^{ih_1 H_1 \frac{t}{2}} \ldots e^{ih_L H_L \frac{t}{2}} e^{ih_L H_L \frac{t}{2}} \ldots e^{ih_1 H_1 \frac{t}{2}} \tag{7}$$

$$= \prod_{i=L}^{1} e^{ih_i H_i \frac{t}{2}} \prod_{j=1}^{L} e^{ih_j H_j \frac{t}{2}}. \tag{8}$$

This formula serves as the base case for the higher-order formulas which can be written as

$$U_{TS}^{(2k)}(t) := U_{TS}^{(2k-2)}(u_k t)^2 \cdot U_{TS}^{(2k-2)}((1 - 4u_k)t) \cdot U_{TS}^{(2k-2)}(u_k t)^2, \tag{9}$$

where $u_k := 1/\left(4 - 4^{1/(2k-1)}\right)$. In addition we define $\Upsilon := 2 \cdot 5^{k-1}$ as the number of "stages" in the higher-order product formula. We can now introduce the time evolution channels as

$$\mathcal{U}_{TS}^{(2k)}(\rho; t) := U_{TS}^{(2k)}(t)\rho U_{TS}^{(2k)\dagger}(t), \tag{10}$$

where for consistency we use the calligraphic $\mathcal{U}_{TS}^{(2k)}$ to represent the applied channels.

Before we introduce the cost and error scaling for Trotter-Suzuki formulas we will make use of the following notation that captures information about the commutator structure of a Hamiltonian or subset of terms from a Hamiltonian. First used in Childs et. al [16] the sum of norms of commutators α_{comm} (represented as $\widetilde{\alpha}_{comm}$ in [16]) is given by

$$\alpha_{comm}(H, 2k) := \sum_{\gamma_i \in \{1, \ldots, L\}} \left(\prod h_{\gamma_i}\right) \left\| [H_{\gamma_{2k+1}}, [H_{\gamma_{2k}}, \ldots [H_{\gamma_2}, H_{\gamma_1}] \ldots]] \right\|_\infty. \tag{11}$$

Another variation we will make is the restriction of α_{comm} to subsets of H, for example if we can form two subsets A, B of H such that $H = A + B$ then we can write the following

$$\alpha_{comm}(A, 2k) = \sum_{\gamma_i \in \{1, \ldots L\}} \left(\prod a_{\gamma_i}\right) \left\| [A_{\gamma_{2k+1}}, [A_{\gamma_{2k}}, \ldots [A_{\gamma_2}, A_{\gamma_1}] \ldots]] \right\|_\infty. \tag{12}$$

We can then define the commutator structure between the two subsets as all nested commutators that contain *at least* one term from both A and B. This then allows for the expression

$$\alpha_{comm}(\{A, B\}, 2k) = \alpha_{comm}(H, 2k) - \alpha_{comm}(A, 2k) - \alpha_{comm}(B, 2k), \tag{13}$$

as any nested commutator with only terms consisting of A matrices is contained in $\alpha_{comm}(A, 2k)$ and similarly for B.

Now we can state a summary of the performance of Trotter-Suzuki formulas as proved in [16].

Theorem 2 (Trotter-Suzuki Formulas). *Given a Hamiltonian H, time t, and error bound ϵ, a $2k^{th}$-order Trotter-Suzuki channel as defined in 1 satisfies $\left\|\mathcal{U}(t) - \mathcal{U}_{TS}^{(2k)}(t)\right\|_\diamond < \epsilon$ and uses the following number of gates of the form $e^{iH_i t'}$*

$$C_{Trott}(H, t, \epsilon, 2k) = \Upsilon L r \leq \Upsilon L \left\lceil \frac{(\Upsilon t)^{1+1/2k}}{\epsilon^{1/2k}} \left(\frac{4\alpha_{comm}(H, 2k)^{1/2k}}{2k+1} \right) \right\rceil. \tag{14}$$

Similarly, a first-order Trotter-Suzuki formula has the following cost

$$C_{Trott}(H, t, \epsilon, 1) = L r \leq L \left\lceil \frac{t^2}{2\epsilon} \sum_{i,j} h_i h_j \left\| [H_i, H_j] \right\|_\infty \right\rceil \tag{15}$$

Proof. We first upper bound the diamond distance between our implementation and the ideal time evolution channels as follows.

$$\left\|\mathcal{U}(t) - \mathcal{U}_{TS}^{(2k)}(t)\right\|_\diamond := \left\|\left(\mathcal{U}(t) - \mathcal{U}_{TS}^{(2k)}(t)\right) \otimes \mathbb{1}\right\|_1 \tag{16}$$

$$= \max_{\rho: \|\rho\|_1 \leq 1} \left\|\left(e^{iHt} \otimes \mathbb{1}\right) \rho \left(e^{-iHt} \otimes \mathbb{1}\right) - \left(U_{TS}^{(2k)}(t) \otimes \mathbb{1}\right) \rho \left(U_{TS}^{(2k)\dagger}(t) \otimes \mathbb{1}\right)\right\|_1 \tag{17}$$

$$\leq \max_{\rho: \|\rho\|_1 \leq 1} \left\|\left(e^{iHt} \otimes \mathbb{1}\right) \rho \left(e^{-iHt} \otimes \mathbb{1}\right) - \left(e^{iHt} \otimes \mathbb{1}\right) \rho \left(U_{TS}^{(2k)\dagger}(t) \otimes \mathbb{1}\right)\right\|_1 \tag{18}$$

$$+ \max_{\rho: \|\rho\|_1 \leq 1} \left\|\left(e^{iHt} \otimes \mathbb{1}\right) \rho \left(U_{TS}^{(2k)\dagger}(t) \otimes \mathbb{1}\right) - \left(U_{TS}^{(2k)}(t) \otimes \mathbb{1}\right) \rho \left(U_{TS}^{(2k)\dagger}(t) \otimes \mathbb{1}\right)\right\|_1$$

$$= \max_{\rho: \|\rho\|_1 \leq 1} \left\|\rho \left(e^{-iHt} - U_{TS}^{(2k)\dagger}(t)\right) \otimes \mathbb{1}\right\|_1 + \max_{\rho: \|\rho\|_1 \leq 1} \left\|\left(e^{iHt} - U_{TS}^{(2k)}(t)\right) \otimes \mathbb{1}\rho\right\|_1 \tag{19}$$

$$\leq 2 \left\|e^{iHt} - U_{TS}^{(2k)}(t)\right\|_\infty \max_{\rho: \|\rho\|_1 \leq 1} \|\rho\|_1 \tag{20}$$

$$= 2 \left\|e^{iHt} - U_{TS}^{(2k)}(t)\right\|_\infty. \tag{21}$$

We can then make use of Eq. (189) and Theorem 10 from [16], which provides the following bound

$$\left\|e^{iHt/r} - U_{TS}^{(2k)}(t/r)\right\|_\infty \leq \frac{2\alpha_{comm}(H, 2k)}{2k+1} \left(\frac{\Upsilon t}{r}\right)^{2k+1}. \tag{22}$$

We note that this equation differs from Eq. (189) in [16] due to the different α_{comm} used. The denominator of $(2k+1)!$ is replaced by $2k+1$ due to a factor of $(2k)!$ from upper bounds on the α_{comm} used in [16]. Note that this also leads to the extra factors of Υ^{2k}, as opposed to just Υ in Eq. (189) in [16].

For the first-order formula we will use the following upper bound which follows from an application of the triangle inequality to Eq. (143) from [16]

$$\left\|e^{iHt/r} - U_{TS}^{(1)}(t/r)\right\|_\infty \leq \frac{t^2}{2r^2} \sum_{i,j} h_i h_j \left\|[H_i, H_j]\right\|_\infty \tag{23}$$

Combining Eqs. (21) and (22), along with the inequality $\|X^{\circ r} - Y^{\circ r}\|_\diamond \leq r \|X - Y\|_\diamond$, yields

$$\left\|\mathcal{U}(t) - \mathcal{U}_{TS}^{(2k)}(t/r)^{\circ r}\right\|_\diamond \leq r \left\|\mathcal{U}(t/r) - \mathcal{U}_{TS}^{(2k)}(t/r)\right\|_\diamond \tag{24}$$

$$\leq \frac{4r\alpha_{comm}(H, 2k)}{2k+1} \left(\frac{\Upsilon t}{r}\right)^{2k+1}. \tag{25}$$

We then can require the inequality in Eq. (25) to be less than ϵ and solve for r, yielding

$$r > \frac{(\Upsilon t)^{1+1/2k}}{\epsilon^{1/2k}} \left(\frac{4\alpha_{comm}(H,2k)}{2k+1} \right)^{1/2k}. \tag{26}$$

By taking the ceiling of the RHS of (26) and plugging the result into $C_{Trott}(H,t,\epsilon) = \Upsilon L r$ yields the expression in the statement. Similar results hold for the first-order case. $\qquad\square$

3.2 QDrift

We now shift our attention to the other main product formula we will make use of, that is QDrift. Introduced by Campbell in [14], the main premise of QDrift is that one randomly picks a term H_i from the overall set of terms according to the ratio of spectral norms h_i/λ and then apply the exponential gate $e^{iH_i\tau}$, for some $\tau \propto t$. This is summarized in the following definition.

Definition 3 (QDrift Channel). *Let $p_i = \frac{h_i}{\lambda}$, where $\lambda = \sum_i h_i$, represent a probability distribution over terms in a Hamiltonian $H = \sum_i h_i H_i$. We define the QDrift channel for a single sample from this distribution as*

$$\mathcal{U}_{QD}(t) : \rho \mapsto \sum_i p_i e^{iH_i\lambda t} \rho e^{-iH_i\lambda t}. \tag{27}$$

Below we restate the main results from [14], in which multiple independent samples of the above channel are studied, with only minor modifications to the allowable range of ϵ.

Theorem 4 (QDrift). *Given a Hamiltonian H, time t, and error bound ϵ one can approximate the ideal unitary dynamics of $\mathcal{U}(t)$ by taking N i.i.d samples of the QDrift channel from Definition 3. To meet the error bound ϵ, namely $\left\| \mathcal{U}(t) - \mathcal{U}_{QD}(t/N)^{\circ N} \right\|_\diamond < \epsilon$, it suffices to choose $N = \frac{4t^2\lambda^2}{\epsilon}$ if we restrict allowed values of ϵ to within the range $(0, \lambda t \ln(2)/2)$. This gives the cost of the channel, or the number of operator exponentials of the form $e^{iH_i t'}$ that the channel requires, as*

$$C_{QD}(H,t,\epsilon) \le \frac{4\lambda^2 t^2}{\epsilon}. \tag{28}$$

Our minor modification follows from the proof of the following expression

$$\left\| \mathcal{U}(t) - \mathcal{U}_{QD}(t/N)^{\circ N} \right\|_\diamond \le \frac{2\lambda^2 t^2}{N} e^{2\lambda t/N}, \tag{29}$$

which was given in [14]. We upper bound the coefficient $e^{2\lambda t/N} \le 2$ by using $N = 4\lambda^2 t^2/\epsilon$ and restricting $\epsilon \in (0, \lambda t \ln(2)/2)$.

4 First-Order Trotter with QDrift

The most straightforward Composite channel to analyze is combining a first-order Trotter formula with a QDrift channel. We proceed in four steps. First, we assume a partitioning $H = A + B$ which allows us to determine the diamond distance error scaling of the Composite channel. Next, we formulate an upper bound on the number of exponential gates of the form $e^{iH_i t}$ needed to achieve this error. Following this, we use the derived cost function to determine a useful partitioning scheme for determining whether a term in a given Hamiltonian H should end up in the Trotter channel or the QDrift channel. Finally, we give an instance in which a Composite channel can offer asymptotic improvements over either a purely Trotter or QDrift channel.

4.1 Query Complexity

To analyze the error of our Composite channel we need to first reduce the overall time evolution channel $\rho \mapsto e^{-iHt} \rho e^{+iHt}$ into the simpler pieces that we can analyze with our Trotter and QDrift results. Assuming a partitioning $H = A + B$, where A consists of terms that we would like to

simulate with Trotter and B has the terms we would like to sample from with QDrift. We now introduce the "outer-loop" error $E_{\{A,B\}}$ induced by this partitioning, which is as follows

$$e^{-iHt}\rho e^{+iHt} = e^{-iBt}e^{-iAt}\rho e^{+iAt}e^{+iBt} + E_{\{A,B\}}(t). \tag{30}$$

We use the phrase "outer-loop" as this decomposition is done before any simulation channels are implemented.

The rest of the error analysis is captured in the following lemma.

Theorem 5 (First-Order Composite Channel). *Given a time t, error bound ϵ, and a partitioned Hamiltonian $H = A + B$ one can construct a first-order composite simulation channel $\widetilde{\mathcal{U}}(t)$ that approximates the ideal channel $\mathcal{U}(t)$ within a diamond distance ϵ as follows. Let $\widetilde{\mathcal{U}}(t) = \widetilde{\mathcal{U}_B}(t) \circ \widetilde{\mathcal{U}_A}(t)$ represent the composition of a Trotter-Suzuki channel $\widetilde{\mathcal{U}_A}$ to simulate A and a QDrift channel $\widetilde{\mathcal{U}_B}$ for B. By repeating $\widetilde{\mathcal{U}}(t/r)$ for r iterations, the diamond distance bound $\left\lVert \mathcal{U}(t) - \widetilde{\mathcal{U}}(t/r)^{\circ r} \right\rVert_\diamond < \epsilon$ can be acheived by using no more than*

$$C_{Comp}(A, B, t, \epsilon) = (L_A + N_B)r \tag{31}$$

$$= (L_A + N_B) \left\lceil \frac{t^2}{\epsilon} \left(\sum_{i,j} a_i a_j \lVert[A_i, A_j]\rVert_\infty + \sum_{i,j} a_i b_j \lVert[A_i, B_j]\rVert_\infty + \frac{4\lambda_B^2}{N_B} \right) \right\rceil \tag{32}$$

gates of the form $e^{iH_i t'}$.

Proof. We will need to make use of the following minor result

$$\lVert X^{\circ r} - Y^{\circ r} \rVert_\diamond \leq r \lVert X - Y \rVert_\diamond, \tag{33}$$

where X and Y are channels. This follows straightforwardly from subadditivity of the diamond norm with respect to composition of channels. Now starting with the outer-loop decomposition mentioned above we can reduce the overall channel distance to a per-iteration distance as follows

$$\left\lVert \mathcal{U}(t) - \widetilde{\mathcal{U}}(t/r)^{\circ r} \right\rVert_\diamond \leq \left\lVert \left(\mathcal{U}_B(t/r) \circ \mathcal{U}_A(t/r) + E_{\{A,B\}}(t/r) \right)^{\circ r} - \left(\widetilde{\mathcal{U}_A}(t/r) \circ \widetilde{\mathcal{U}_B}(t/r) \right)^{\circ r} \right\rVert_\diamond \tag{34}$$

$$\leq r \left(\left\lVert \mathcal{U}_A(t/r) - \widetilde{\mathcal{U}_A}(t/r) \right\rVert_\diamond + \left\lVert \mathcal{U}_B(t/r) - \widetilde{\mathcal{U}_B}(t/r) \right\rVert_\diamond + \left\lVert E_{\{A,B\}}(t/r) \right\rVert_\diamond \right). \tag{35}$$

This is now in a form where we can use the results from Section 3 for Trotter formulas and QDrift channels. We use the QDrift results from Theorem 4 that $\left\lVert \mathcal{U}_B(t/r) - \widetilde{\mathcal{U}_B}(t/r) \right\rVert_\diamond \leq \frac{4\lambda_B^2 t^2}{N_B r^2}$ and Eqs (16) - (21) from Theorem 2 to reduce $\left\lVert \mathcal{U}_A(t/r) - \widetilde{\mathcal{U}_A}(t/r) \right\rVert_\diamond \leq 2 \left\lVert e^{iAt/r} - U_{TS}^{(1)}(t/r) \right\rVert_\infty$. The last term we need to bound is the outer-loop error

$$\left\lVert E_{\{A,B\}}(t/r) \right\rVert_\diamond = \lVert \mathcal{U}(t/r) - \mathcal{U}_B(t/r) \circ \mathcal{U}_A(t/r) \rVert_\diamond \tag{36}$$

$$\leq 2 \left\lVert e^{iH(t/r)} - e^{iB(t/r)}e^{iA(t/r)} \right\rVert_\infty \tag{37}$$

$$\leq \frac{t^2}{r^2} \sum_{i,j} a_i b_j \lVert[A_i, B_j]\rVert_\infty \tag{38}$$

Plugging in Theorems 4 and 2 into Eq. 35 yields

$$\frac{1}{r} \left\lVert \mathcal{U}(t) - \widetilde{\mathcal{U}}(t/r)^r \right\rVert_\diamond \leq \left(\frac{t}{r} \right)^2 \left(\sum_{i,j} a_i a_j \lVert[A_i, A_j]\rVert_\infty + \sum_{i,j} a_i b_j \lVert[A_i, B_j]\rVert_\infty \right) + \frac{4\lambda_B^2 t^2}{N_B r^2} \leq \frac{\epsilon}{r}, \tag{39}$$

where N_B represents the number of samples used by QDrift to simulate e^{iBt}. It is straightforward to solve for r that satisfies the inequality in Eq (39) to plug into the expression $C_{Comp}(A, B, t, \epsilon) = (L_A + N_B)r$ which yields the theorem statement. $\square$

We now use a relaxation of the first-order Composite channel cost from Eq. (32) in which we allow for non-integer values, which is given as

$$\widetilde{C}_{Comp}(A, B, t, \epsilon) := (L_A + N_B)\frac{t^2}{\epsilon}\left(\sum_{i,j} a_i a_j \left\|[A_i, A_j]\right\|_\infty + \sum_{i,j} a_i b_j \left\|[A_i, B_j]\right\|_\infty + \frac{4\lambda_B^2}{N_B}\right). \quad (40)$$

One unspecified quantity in the above expression is N_B which is specifically left as a user-defined parameter. This means we can optimize the non-integer cost $\widetilde{C}_{Comp}$ with respect to N_B, which is done in the following lemma.

Lemma 6. *Let $\widetilde{C}_{Comp}(t, \epsilon, A, B)$ denote the non-integer cost of a first-order Composite channel approximation to $\mathcal{U}(t)$. Then $\widetilde{C}_{Comp}$ can be optimized with respect to N_B when*

$$N_B = \sqrt{\frac{4\lambda_B^2 L_A}{\left(\sum_{i,j} a_i a_j \left\|[A_i, A_j]\right\|_\infty + \sum_{i,j} a_i b_j \left\|[A_i, B_j]\right\|_\infty\right)}}, \quad (41)$$

note that this expression is only defined if $\left\|[A_i, A_j]\right\|_\infty > 0$ or $\left\|[A_i, B_j]\right\|_\infty > 0$ for at least one A_i or B_j.

Proof. The result follows from basic calculus with the additional assumption that we will treat the above cost upper bound as exact

$$\frac{\partial \widetilde{C}_{comp}}{\partial N_B} = \frac{t^2}{\epsilon}\left[\sum_{i,j} a_i a_j \left\|[A_i, A_j]\right\| + \sum_{i,j} a_i b_j \left\|[A_i, B_j]\right\| - \frac{4\lambda_B^2 L_A}{N_B^2}\right]. \quad (42)$$

Setting the above equal to zero and solving for N_B yields the stated value. The second derivative can be shown as

$$\frac{\partial^2 \widetilde{C}_{comp}}{\partial N_B^2} = \frac{4t^2\lambda_B^2 L_A}{\epsilon N_B^3} \geq 0, \quad (43)$$

which indicates the optima found is the minimal cost with respect to N_B. $\qquad\square$

4.2 Hamiltonian Partitioning

Now that we have upper bounded the number of operator exponentials needed for a Composite channel to satisfy $\left\|\mathcal{U}(t) - \widetilde{\mathcal{U}}(t)\right\|_\diamond < \epsilon$ with a predetermined partition we move on to the question of how to decide a partition. There are many different ways one could determine a partitioning, for example by using a greedy algorithm or a spectral norm based decider, and here we propose a new method that is based on our derived cost function. Our method allows one to take into account information about the commutation structure between terms and spectral norm information to compute a cost function gradient that can be minimized in a gradient descent approach. We also show an analytic minima of this gradient that allows for a greedy approach.

The first step we have to take is to determine how to parametrize our cost function $\widetilde{C}_{Comp}$. We introduce new parameters w_i which represent a weighting of each term H_i between the Trotter and QDrift partitions. Starting with our Hamiltonian $H = \sum_i h_i H_i$ we rewrite each term as a parametrized sum, $h_i H_i \mapsto w_i h_i H_i + (1 - w_i)h_i H_i$. Then we place all terms $w_i h_i H_i$ in the Trotter partition $A = \sum_i w_i h_i H_i$ and all the terms $(1 - w_i)h_i H_i$ into the QDrift partition $B = \sum_i (1 - w_i)h_i H_i$. Now instead of determining the discrete placement of each term into A or B we only need to determine an appropriate weighting w_i of each term between the two partitions. As we would like the coefficients to remain positive after this remapping we require $w_i \in [0, 1]$. We will first work out the gradient of the cost with respect to each weight and then discuss its behavior.

First we write the non-integer cost function of the weighted partitioned Composite channel as

$$\widetilde{C}_{comp} = (L_A + N_B)\frac{t^2}{\epsilon}\left(\sum_{i,j} w_i w_j h_i h_j \left\|[H_i, H_j]\right\| + \sum_{i,j} w_i(1 - w_j)h_i h_j \left\|[H_i, H_j]\right\| + \frac{4\left(\sum_i(1 - w_i)h_i\right)^2}{N_B}\right), \quad (44)$$

note that we are leaving N_B as a user-defined integer and not the optimized value as found before. Now we can easily take the derivative of Eq. (44) with respect to the mth weight w_m, which is

$$\frac{\partial \widetilde{C}_{comp}}{\partial w_m} = (L_A + N_B)\frac{t^2}{\epsilon}\left(h_m \sum_j h_j \left|\left|[H_j, H_m]\right|\right|_\infty - \frac{8h_m \sum_i (1 - w_i)h_i}{N_B}\right). \tag{45}$$

This is enough information to perform a gradient descent to find a optima from an initial partitioning. However, it is relatively easy to find the exact optima for a single Hamiltonian term with respect to the other weightings. We can set Eq. (45) equal to 0 and solve for w_m which yields

$$\frac{\partial \widetilde{C}_{comp}}{\partial w_m} = 0 \implies w_m = 1 - \sum_{i \neq m} \frac{h_i}{h_m}\left(\frac{\left|\left|[H_i, H_m]\right|\right|_\infty}{8} - (1 - w_i)\right). \tag{46}$$

There are a few pieces of intuition we can gather from these expressions. First, if a term H_m commutes with every other term in the Hamiltonian then $[H_i, H_m] = 0$ and $w_m = 1 + \sum_{i \neq m} h_i/h_m(1 - w_i)$, which is always greater than 1. Since we restrict our weights to $[0, 1]$ this implies that the mth term should always be fully placed in the Trotter channel. The other piece of intuition is that smaller terms are pushed more towards the QDrift side of the partitioning. This can be seen from Eq. (46) while considering the limit as $h_m \to 0$. If we assume that $\left|\left|[H_i, H_m]\right|\right| \geq (1 - w_i)$ on average, then the expression becomes $w_m \to -\infty$ in this limit, which we stop at 0.

One major drawback to the above expressions is the dependency of each optimal weight w_m on every other weight w_i. As there does not seem to be a clear basis in which to decouple these weights, this means that Eq. (46) can only be used to update individual weights given an initialization. This is the same situation as the greedy approach as discussed above, but we note that our expression gives us some intuition as for why which weights or partitionings are chosen.

4.3 Comparison with Trotter and QDrift

Now that we have analyzed the cost and given a partitioning scheme we would like to know under what conditions this Composite channel can lead to comparable errors with lower gate cost. Instead of aiming to show that a Composite channel will outperform either first-order Trotter or QDrift for arbitrary Hamiltonians we instead illustrate a concrete setting in which we achieve guaranteed asymptotic improvements. In later sections we are able to show more generic conditions on which asymptotic improvements can be obtained for higher-order formulas.

The final case we consider for the first-order Trotter Composite channel is designed to take full advantage of this richer commutator structure of the Composite channel over first-order Trotter. Consider a Hamiltonian H that has a partitioning into A and B such that the following two conditions hold

1. The number of non-zero commutators between terms in A scales with the square root of L_A. Mathematically,

$$\left|\{(i, j) : \left|\left|[A_i, A_j]\right|\right| \neq 0\}\right| \coloneqq N_{nz}^2 \in o(L_A). \tag{47}$$

2. The strength of the B terms, $\lambda_B = \sum_i b_i$, is asymptotically less than the the maximum commutator norm divided by the number of terms in A

$$\lambda_B^2 \leq a_{max}^2 \left(N_{nz}^4 / L_A^2\right), \tag{48}$$

   where $a_{max} \coloneqq \max_i a_i$.

3. The number of terms in the A partition is vanishingly small compared to the total number of terms $L_A \in o(L)$.

Next we can use the optimal N_B value from Lemma 6 and (48) to show that

$$N_B^{-1} \in O\left(\frac{1}{\lambda_B}\sqrt{(N_{nz}^2 a_{max}^2 + L_A a_{max} \lambda_B)}\right) = O\left(\frac{N_{nz} a_{max}}{\lambda_B}\right). \tag{49}$$

Similarly we have

$$N_B \in O\left(\frac{\lambda_B \sqrt{L_A}}{a_{max} N_{nz}}\right). \tag{50}$$

Thus Theorem 2 shows that the number of exponentials needed to perform the simulation is in

$$C_{Comp} \in O\left(\frac{t^2}{\epsilon}\left(L_A + \frac{\lambda_B \sqrt{L_A}}{a_{max} N_{nz}}\right)\left(a_{max}^2 N_{nz}^2 + L_A a_{max} \lambda_B + \frac{\lambda_B N_{n_z} a_{max}}{\sqrt{L_A}}\right)\right)$$

$$\in O\left(\frac{t^2 L_A}{\epsilon}\left(a_{max}^2 N_{nz}^2\right)\right) \tag{51}$$

$$\in o\left(\frac{t^2}{\epsilon} L_A^2 a_{max}^2\right). \tag{52}$$

Were we to use the lowest order Trotter formula for this simulation, the cost would be

$$C_{trot} \in O\left(\frac{t^2}{\epsilon}\left(L N_{nz}^2 a_{max}^2\right)\right) \tag{53}$$

$$\in o\left(\frac{t^2}{\epsilon} L L_A a_{max}^2\right) \tag{54}$$

$$\subseteq \omega(C_{comp}). \tag{55}$$

In contrast the cost for QDrift is

$$C_{QD} = O\left(\frac{t^2}{\epsilon}\left(L^2 a_{max}^2\right)\right) \subseteq \omega(C_{comp}). \tag{56}$$

This shows that there exist circumstances where the cost of the Composite channel scales better than either of the two methods that compose it.

5 Higher-Order Trotter Formulas

We now move on from first-order Trotter formulas to arbitrary higher-order Trotter formulas. To analyze this case there are a few distinct differences with the first-order channels. The first is that we now have a choice for what order formula we would like to use for the outer-loop decomposition of $\widetilde{\mathcal{U}}$. For example, a first-order decomposition would be $\widetilde{\mathcal{U}}(t) = \widetilde{\mathcal{U}}_B(t) \circ \widetilde{\mathcal{U}}_A(t)$ and a second-order decomposition would be $\widetilde{\mathcal{U}}(t) = \widetilde{\mathcal{U}}_A(t/2) \circ \widetilde{\mathcal{U}}_B(t/2) \circ \widetilde{\mathcal{U}}_B(t/2) \circ \widetilde{\mathcal{U}}_A(t/2)$. In general, we can choose any order formula we like but it is analytically convenient to match the innermost Trotter formula. The next difference is that the time scaling between QDrift, Trotter, and the outer-loop errors could all be of different orders in t/r which leads to a non-analytically solvable polynomial in r. The last issue that we address is that the commutator structure is no longer quadratic with respect to the Hamiltonian spectral norms, so we cannot follow the term weighting partitioning scheme from the first-order case. We will follow the same organizational structure as the first-order case and first set up our definitions and bound the diamond distance error, then compute the number of $e^{iH_i t}$ queries, followed by developing a partitioning scheme, and finally discuss the cost comparisons between our Composite channel and its constituents.

5.1 Query Complexity

We first need to determine an error equation for the Composite channel which we will then use to bound the number of iterations needed. In the first order Trotter formula we simply used the following overall evolution $\widetilde{\mathcal{U}} = \widetilde{\mathcal{U}}_B \circ \widetilde{\mathcal{U}}_A$, but this is not sufficient for the higher-order case. We now introduce a generalization of this which mimics the Trotter formula recursion.

Definition 7 (Higher-Order Outer-Loop). *Let $\mathcal{U}(t) = \mathcal{U}_B(t) \circ \mathcal{U}_A(t)$ denote the first-order outer-loop decomposition. We define the second-order outer-loop decomposition as*

$$\mathcal{U}^{(2)}(t) \coloneqq \mathcal{U}_A(t/2) \circ \mathcal{U}_B(t/2) \circ \mathcal{U}_B(t/2) \circ \mathcal{U}_A(t/2). \tag{57}$$

This forms the base case for the recursive strategy for higher-order outer-loops defined as

$$\mathcal{U}^{(2k)}(t) := \mathcal{U}^{(2k-2)}(u_k t)^2 \circ \mathcal{U}^{(2k-2)}((1 - 4u_k)t) \circ \mathcal{U}^{(2k-2)}(u_k t)^2, \tag{58}$$

where $u_k := 1/\left(4 - 4^{1/(2k-1)}\right)$ and $\Upsilon := 2 \cdot 5^{k-1}$. Note that we use the same recursive strategy to define approximations to the overall time evolution channel where we put tildes on each of the implemented channels.

To analyze the overall error we need to break down the overall channel into individual channels that we have known results for. We specifically use the approach of using the same decomposition order for the outer-loop that we use for the innermost Trotter formula.

Lemma 8. *Let $\mathcal{U}$ denote the exact unitary time evolution channel and $\widetilde{\mathcal{U}}_{2k}$ denote an implemented product formula according to Definition 7 for a partitioning of H into A, B. Then we have the following diamond distance upper bound*

$$\left\|\mathcal{U}(t) - \widetilde{\mathcal{U}}^{(2k)}(t)\right\|_\diamond \le \Upsilon \left\|\mathcal{U}_A(t) - \widetilde{\mathcal{U}_A}(t)\right\|_\diamond + \Upsilon \left\|\mathcal{U}_B(t) - \widetilde{\mathcal{U}_B}(t)\right\|_\diamond + 2\left\|e^{iHt} - U_{TS}^{(2k)}(\{A, B\}, t)\right\|. \tag{59}$$

Proof. The proof follows from repeated applications of the triangle inequality as well as subadditivity of the diamond norm with respect to channel composition

$$\left\|\mathcal{U}(t) - \widetilde{\mathcal{U}}^{(2k)}(t)\right\|_\diamond \le \left\|\mathcal{U}(t) - \mathcal{U}^{(2k)}(t)\right\|_\diamond + \left\|\mathcal{U}^{(2k)}(t) - \widetilde{\mathcal{U}}^{(2k)}(t)\right\|_\diamond \tag{60}$$

$$\le 2\left\|e^{iHt} - U_{TS}^{(2k)}(\{A, B\}, t)\right\| + \left\|\mathcal{U}^{(2k)}(t) - \widetilde{\mathcal{U}}^{(2k)}(t)\right\|_\diamond \tag{61}$$

$$= 2\left\|e^{iHt} - U_{TS}^{(2k)}(\{A, B\}, t)\right\|$$
$$+ \left\|\mathcal{U}_B(t_\Upsilon) \circ \mathcal{U}_A(t_\Upsilon) \circ \ldots \mathcal{U}_B(t_1) \circ \mathcal{U}_A(t_1) - \widetilde{\mathcal{U}_B}(t_\Upsilon) \circ \widetilde{\mathcal{U}_A}(t_\Upsilon) \circ \ldots \widetilde{\mathcal{U}_B}(t_1) \circ \widetilde{\mathcal{U}_A}(t_1)\right\|_\diamond \tag{62}$$

$$\le 2\left\|e^{iHt} - U_{TS}^{(2k)}(\{A, B\}, t)\right\| + \max_{t_i} \Upsilon \left(\left\|\mathcal{U}_A(t_i) - \widetilde{\mathcal{U}_A}(t_i)\right\|_\diamond + \left\|\mathcal{U}_B(t_i) - \widetilde{\mathcal{U}_B}(t_i)\right\|_\diamond\right). \tag{63}$$

Note that each t_i is a constant multiple of t, at each layer in the recursive formula t picks up either a factor of $1 - 4u_k$, u_k or $1/2$. Since $u_k \le 1/2$ we can say that $|1 - 4u_k| \le 1$. This implies that we can upper bound each time interval as $t_i \le t$, which is sufficient for our purposes. Plugging this in to the previous equation yields the expression in the statement. □

Now that we have derived a basis for the Composite channel error we can provide an upper bound on the number of operator exponentials needed to accurately approximate the ideal time evolution channel.

Theorem 2.1 (Gate Cost for Higher-Order Composite Channel)**.** *Given a time t, error bound ϵ, partitioned Hamiltonian $H = A + B$, and let $\widetilde{\mathcal{U}}^{(2k)}$ denote the higher-order Composite channel to approximate the exact unitary evolution $\mathcal{U}(t)$. By using r iterations of $\widetilde{\mathcal{U}}^{(2k)}(t/r)$ we can satisfy the error requirement $\left\|\mathcal{U}(t) - \widetilde{\mathcal{U}}^{(2k)}(t/r)^{\circ r}\right\|_\diamond \le \epsilon$ by using at most the following number of operator exponentials*

$$C_{comp}(A, B, t, \epsilon, 2k)$$
$$\le \Upsilon(\Upsilon L_A + N_B)\left\lceil \frac{(\Upsilon t)^{1+1/2k} 4^{1/2k}}{\epsilon^{1/2k}}\left(\frac{\Upsilon\alpha_{comm}(A, 2k) + \alpha_{comm}(\{A, B\}, 2k)}{2k + 1}\right)^{1/2k} + \frac{4\Upsilon\lambda_B^2 t^2}{N_B\epsilon}\right\rceil. \tag{1}$$

By making the definition $q_B := \frac{\alpha_{comm}(B, 2k)}{\alpha_{comm}(H, 2k)}$ and utilizing the upper bounds from Theorems 2 and 4, where C_{Trott} and C_{QD} below are upper bounds on the number of operator exponentials required in the Trotter and QDrift channels, we can write the cost upper bound as

$$C_{comp} \le \Upsilon(\Upsilon L_A + N_B)\left\lceil C_{Trott}(H, t, \epsilon, 2k)\frac{(1 - q_B)^{1/2k}}{\Upsilon^{1-1/2k}L} + C_{QD}(H, t, \epsilon)\frac{\Upsilon}{N_B}\frac{\lambda_B^2}{\lambda^2}\right\rceil. \tag{2}$$

Proof. We first note that by using prior arguments, namely Eq. (33) and Lemma 8, it is sufficient to show that $\left\|\mathcal{U}(t/r) - \widetilde{\mathcal{U}}(t/r)\right\|_\diamond \leq \epsilon/r$ to satisfy the total diamond distance error bound of ϵ. Using Lemma 8 as well as the Trotter and QDrift errors from Eqs (25) and (29)

$$\left\|\mathcal{U}(t/r) - \widetilde{\mathcal{U}}(t/r)\right\|_\diamond \leq \Upsilon \left\|\mathcal{U}_A(t/r) - \widetilde{\mathcal{U}_A}(t/r)\right\|_\diamond + \Upsilon \left\|\mathcal{U}_B(t/r) - \widetilde{\mathcal{U}_B}(t/r)\right\|_\diamond + 2 \left\|e^{iHt/r} - U_{TS}^{(2k)}(\{A,B\}, t/r)\right\|$$
$$(64)$$

$$\leq 2\Upsilon \left\|e^{iAt/r} - U_{TS}^{(2k)}(A, t/r)\right\| + 2 \left\|e^{iHt/r} - U_{TS}^{(2k)}(\{A,B\}, t/r)\right\| + \left(\frac{t}{r}\right)^2 \frac{4\Upsilon\lambda_B^2}{N_B}$$
$$(65)$$

$$\leq \left(\frac{t}{r}\right)^{2k+1} \frac{4\Upsilon^{2k+1}}{2k+1} \left(\Upsilon\alpha_{comm}(A, 2k) + \alpha_{comm}(\{A,B\}, 2k)\right) + \left(\frac{t}{r}\right)^2 \frac{4\Upsilon\lambda_B^2}{N_B}.$$
$$(66)$$

It will prove useful for brevity to define the following quantities

$$P(t) := t^{2k+1} \frac{4\Upsilon^{2k+1}}{2k+1} \left(\Upsilon\alpha_{comm}(A, 2k) + \alpha_{comm}(\{A,B\}, 2k)\right) \tag{67}$$

$$Q(t) := t^2 \frac{4\Upsilon\lambda_B^2}{N_B} \tag{68}$$

$$\tag{69}$$

where P represents the "product formula" error and Q captures the QDrift error. We can then use results from Theorems 2 and 4, as well as the upper bound $\alpha_{comm}(\{A,B\}, 2k) \leq \Upsilon\alpha_{comm}(\{A,B\}, 2k)$ to write the following expressions

$$\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} \leq C_{Trott}(H, t, \epsilon) \frac{(1-q_B)^{1/2k}}{\Upsilon^{1-1/2k}L} \tag{70}$$

$$\frac{Q(t)}{\epsilon} = C_{QD}(H, t, \epsilon) \frac{\Upsilon\lambda_B^2}{\lambda^2 N_B}. \tag{71}$$

This gives our error as $\left\|\mathcal{U}(t/r) - \widetilde{\mathcal{U}}(t/r)\right\|_\diamond \leq \frac{P(t)}{r^{2k+1}} + \frac{Q(t)}{r^2}$. We now pivot to finding a good value for r that satisfies this inequality. Since there are no generic analytic solutions to polynomials of the form $ax^n + bx^2 = c$ for an arbitrary positive integer n we have to resort to lower bounds on r. In other words, we would like to have a computable lower bound $r_{min} < r$ such that the following inequalities are satisfied

$$\frac{P(t)}{r^{2k+1}} + \frac{Q(t)}{r^2} \leq \frac{P(t)}{r_{min}^{2k+1}} + \frac{Q(t)}{r_{min}^2} \leq \frac{\epsilon}{r} \leq \frac{\epsilon}{r_{min}}. \tag{72}$$

We can make use of the above by finding expressions relating r_{min} to Q and P. The first expression we will make use of is

$$\frac{P(t)}{r^{2k+1}} + \frac{Q(t)}{r^2} \leq \frac{P(t)}{r^2 r_{min}^{2k-1}} + \frac{Q(t)}{r^2} \leq \frac{\epsilon}{r} \tag{73}$$

$$\frac{1}{\epsilon} \left(\frac{P(t)}{r_{min}^{2k-1}} + Q(t)\right) \leq r, \tag{74}$$

which reduces our task to finding a bound on r_{min} using just P alone. This is feasible if we revisit Eq. 72 and use the assumption that $Q(t) \geq 0$ for all possible inputs

$$\frac{P(t)}{r_{min}^{2k+1}} + \frac{Q(t)}{r_{min}^2} \leq \frac{\epsilon}{r_{min}} \tag{75}$$

$$\frac{P(t)}{r_{min}^{2k+1}} \leq \frac{\epsilon}{r_{min}} \tag{76}$$

$$\left(\frac{P(t)}{\epsilon}\right)^{1/2k} \leq r_{min}. \tag{77}$$

Plugging 77 into 74 yields

$$\left(\frac{P(t)}{\epsilon}\right)^{1/2k} + \frac{Q(t)}{\epsilon} \leq r. \tag{78}$$

Now that we have a lower bound for r this gives us an expression for the query cost as

$$C_{comp}(A, B, t, \epsilon) = \Upsilon(\Upsilon L_A + N_B)r \tag{79}$$

$$= \Upsilon(\Upsilon L_A + N_B)\left\lceil \frac{P(t)^{1/2k}}{\epsilon^{1/2k}} + \frac{Q(t)}{\epsilon} \right\rceil, \tag{80}$$

plugging in equations (67) and (67), along with their simplifications in terms of C_{Trott} and C_{QD} yields the expressions in the statement. $\qquad\square$

5.2 Probabilistic Partitioning

As mentioned in Section 4.2 there are multiple ways to go about determining a partition. In this section we develop a novel probabilistic approach that can be used to compute expectation values of necessary parameters in our composite cost function. Importantly the probabilities are computable in time $\Theta(L)$ and require only simple constants to be evaluated. We first discuss why we cannot simply adapt the methods from Section 4.2 before developing our approach.

The main difficulty with adopting our weighting approach from before is the combinatorial landmine $\frac{\partial \alpha_{comm}(\{A,B\},2k)}{\partial w_m}$. When considering $\frac{\partial \alpha_{comm}(\{A,B\},2k)}{\partial w_m}$ note that α_{comm} has $2k+1$ factors of variously indexed w_γ with nested commutators. Taking the derivative with respect to one of the innermost commutator terms leaves for a sum over all possibilities in where this term could be placed and over the remaining operators in the nested commutator. For feasibility one would have to upper bound α_{comm} with spectral 1-norms, such as $\alpha_{comm}(A, 2k) \in O\left(\lambda_A^{2k+1}\right)$, which we do following this discussion. Unfortunately, these upper bounds do not help analytically as $\alpha_{comm}(\{A,B\}, 2k) \in O\left(\sum_l \lambda_A^l \lambda_B^{2k+1-l}\right)$. The resulting derivative is a polynomial over all possible powers of w_i and $(1-w_i)$, which as we have seen before does not have analytically solvable roots in general. This however could be used as the basis for a numeric approach, where one could compute these gradients as a subroutine in an optimization scheme, but it is not useful for our discussion.

The approach we consider is by reinterpreting the weights w_i from Section 4.2 as probabilities p_i for each term to end up in Trotter or QDrift. This means that the expected Hamiltonian we simulate with Trotter is $\mathbb{E}[A] = \sum_i p_i h_i H_i$ and the expected QDrift partition is $\mathbb{E}[B] = \sum_i (1-p_i) h_i H_i$. We also introduce the indicators variables I_i^A which is 1 if the ith term ends up in Trotter and 0 if it is in QDrift. Similarly we can define $I_i^B = 1 - I_i^A$. One main benefit is that this now gives us a probability over all possible partitions, which allows us to compute expectation values for quantities such as C_{comp}, P, and Q. We remark that computing the expected value of the cost $\mathbb{E}[C_{comp}(A, B)]$, which is our main priority, is different than computing the cost of the expected partition $C_{comp}(\mathbb{E}[A], \mathbb{E}[B])$. The expected partition is computationally no different than the weighting scheme mentioned above so we instead compute the expectation of costs over partitions, which is clearly defined and computationally tractable.

The first task we have is to find a useful distribution for each of the p_i's. To do so we start with our cost function and introduce a heuristic that will allow us to find computable values for p_i. The first objects we introduce bounds for are the α_{comm} terms, which will be of additional use later on.

Lemma 9. Let $\alpha_{comm}(\{A,B\}, 2k), \alpha_{comm}(A, 2k)$ be defined as in Eq. (11). Then it holds that

$$\alpha_{comm}(\{A,B\}, 2k) \leq 2^{2k} \sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l} \tag{81}$$

$$\alpha_{comm}(A, 2k) \leq 2^{2k} \lambda_A^{2k+1}. \tag{82}$$

These inequalities can be used to upper bound $P(t)$ as

$$P(t) \leq t^{2k+1} \frac{2^{2k+2}\Upsilon^{2k+1}}{2k+1} \left(\Upsilon \lambda_A^{2k+1} + \sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l} \right) \tag{83}$$

$$\leq 2\frac{2k+\Upsilon}{2k+1}(2\Upsilon\lambda t)^{2k+1} =: P_{max}(t), \tag{84}$$

where we introduce the upper bound $P_{max}(t)$ for P which will be of use later.

Proof. The bounds on the α_{comm} factors are straightforward and computed as follows from the triangle inequality and submultiplicativity of the spectral norm. First we compute the commutator for just the A terms

$$\alpha_{comm}(A, 2k) \leq \sum_{\gamma_{2k+1}=1}^{L_A} \cdots \sum_{\gamma_1=1}^{L_A} ||[A_{\gamma_{2k+1}}, \ldots, [A_{\gamma_2}, A_{\gamma_1}]\ldots]|| \tag{85}$$

$$\leq \sum_{\gamma_{2k+1}=1}^{L_A} \cdots \sum_{\gamma_1=1}^{L_A} 2^{2k} ||A_{\gamma_1}|| \ldots ||A_{\gamma_{2k+1}}|| \tag{86}$$

$$= 2^{2k} \lambda_A^{2k+1}, \tag{87}$$

which is easily generalized to the commutator structure between the A and B terms as

$$\alpha_{comm}(\{A, B\}, 2k) \leq \sum_{l=1}^{2k} \sum_{i_1,\ldots,i_l=1}^{L_A} \sum_{i_{l+1},\ldots,i_{2k+1}=1}^{L_B} ||[A_{i_1}, [A_{i_2}, \ldots, [B_{i_{2k}}, B_{i_{2k+1}}]\ldots]|| \tag{88}$$

$$\leq \sum_{l=1}^{2k} \sum_{i_1,\ldots,i_l=1}^{L_A} \sum_{i_{l+1},\ldots,i_{2k+1}=1}^{L_B} 2^{2k} ||A_{i_1}|| ||A_{i_2}|| \ldots ||B_{i_{2k}}|| ||B_{i_{2k+1}}|| \tag{89}$$

$$\leq 2^{2k} \sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l}, \tag{90}$$

and we note that at least one power of λ_A and λ_B must be present in each term as there must be a minimum of one term from A and a minimum of one term from B in the original nested commutator. We can then use the simple bounds $\lambda_A \leq \lambda$ and $\lambda_B \leq \lambda$ to get partition independent bounds. Plugging these into the definition of $P(t)$ from Eq. (67) yields the bounds in the statement. $\square$

Returning to the original task of finding useful probabilities p_i, we now discuss what heuristics we can introduce to help us towards this goal. The cost of a Composite channel was computed as $C_{comp} \leq (\Upsilon L_A + N_B) \left(\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} + \frac{Q(t)}{\epsilon} \right)$, which we can use to give an upper bound on the expected cost as

$$\mathbb{E}\left[C_{comp}(A, B, t, \epsilon, 2k)\right] \leq \mathbb{E}\left[(\Upsilon L_A + N_B)\left(\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} + \frac{Q(t)}{\epsilon}\right)\right] \tag{91}$$

$$\leq (\Upsilon L + N_B)\left(\frac{\mathbb{E}\left[P(t)^{1/2k}\right]}{\epsilon^{1/2k}} + \frac{\mathbb{E}\left[Q(t)\right]}{\epsilon}\right), \tag{92}$$

where we upper bounded L_A with L and the number of gates performed during the QDrift channel as N_B. This latter point is a bit more subtle than it first appears, as there could be a non-zero probability of having 0 or 1 terms in B, which would mean no matter how many QDrift samples N_B we take we only need to apply either 0 or 1 exponential gate to implement them.

This expression makes clear that the Composite channel cost is a balancing act between higher-order product scaling and QDrift scaling. The heuristic we introduce is that we would like these expected quantities to be somewhat of the same magnitude. This is motivated by the observation that if one simulation method has much higher cost than another method we can simply start our partitioning off completely in the smaller cost method. We can then shift probability mass to the

higher cost channel until the two contributions to the Composite channel are comparable. By making $\mathbb{E}\left[P(t)^{1/2k}\right]\epsilon^{-1/2k} \approx \mathbb{E}\left[Q(t)\right]\epsilon^{-1}$ rigorous we will get useful expressions for the probabilities p_i. This is done in the following lemma.

Lemma 2.1 (Probabilistic Partitioning Scheme). *For a composite simulation of H for time t and error ϵ, let p_i denote the probability of placing term $h_i H_i$ into the Trotter partition of Composite channel. we have that choosing*

$$1 - p_i = \min\left\{\frac{\lambda}{h_i L}\left(\sqrt{N_B\left(\frac{\epsilon}{\lambda t}\right)^{1-1/2k}\left(\frac{2k+\Upsilon}{2k+1}\right)^{1/2k}\frac{\Upsilon^{1/2k}}{2^{1-1/k}}} - 1\right), 1\right\} =: \min\left\{\frac{\chi}{h_i}, 1\right\}, \quad (3)$$

along with choosing the number of QDrift samples N_B to satisfy

$$N_B \geq \left(\frac{\lambda t}{\epsilon}\right)^{1-1/2k}\left(\frac{2k+1}{2k+\Upsilon}\right)^{1/2k}\frac{2^{1-1/k}}{\Upsilon^{1/2k}} \quad (4)$$

guarantees the following:

1. *$p_i \in [0,1]$,*

2. *the expectation value of the coefficients in the QDrift partition satisfies*

$$\frac{\mathbb{E}\left[\lambda_B\right]}{\lambda} \leq \frac{1}{2}\sqrt{\left(\frac{4k+2\Upsilon}{2k+1}\right)^{1/2k}(2\Upsilon)^{1+1/2k}}\sqrt{N_B\left(\frac{\epsilon}{\lambda t}\right)^{1-1/2k}}.$$

This bound follows from a rigorous interpretation of making the QDrift and Trotter errors approximately equivalent.

Proof. We begin by making rigorous our notion that $\mathbb{E}\left[P(t)^{1/2k}\right]\epsilon^{-1/2k} \approx \mathbb{E}\left[Q(t)\right]\epsilon^{-1}$. To do so we will equate an upper bound for $\mathbb{E}\left[P(t)\right]$ with a lower bound for $\mathbb{E}\left[Q\right]$. As we know that $\mathbb{E}\left[P(t)^{1/2k}\right] \leq P_{max}(t)^{1/2k}$ from Lemma 9, we only need to lower bound $\mathbb{E}\left[Q\right]$. This is rather straightforward

$$\mathbb{E}\left[Q(t)\right] = \frac{4\Upsilon\mathbb{E}\left[\lambda_B^2\right]t^2}{N_B} \geq \frac{4\mathbb{E}\left[\lambda_B\right]^2 t^2}{N_B}, \quad (93)$$

which is given from the definition of Q along with Jensen's Inequality and convexity of $f(x) = x^2$. To enforce our heuristic of approximately equal we set the lower bound on $\mathbb{E}\left[Q\right]/\epsilon$ to be less than the upper bound on $\mathbb{E}\left[P^{1/2k}\right]/\epsilon^{1/2k}$. This is straightforward as

$$\frac{4\Upsilon\mathbb{E}\left[\lambda_B\right]^2 t^2}{N_B\epsilon} \leq \frac{P_{max}(t)^{1/2k}}{\epsilon^{1/2k}} \quad (94)$$

$$\mathbb{E}\left[\lambda_B\right] \leq \frac{1}{2t}\sqrt{\frac{P_{max}(t)^{1/2k}N_B\epsilon^{1-1/2k}}{\Upsilon}}. \quad (95)$$

As λ_B is simply the sum of the spectral norms for each term in the B channel, we can write it as $\sum_i \mathbb{E}\left[I_i^B h_i\right]$. By plugging in the expectation of I_i^B and P_{max} from Eq. (84) into the above we arrive at the following

$$\sum_i (1-p_i)h_i \leq \lambda\sqrt{N_B\left(\frac{\epsilon}{\lambda t}\right)^{1-1/2k}\left(\frac{2k+\Upsilon}{2k+1}\right)^{1/2k}\frac{\Upsilon^{1/2k}}{2^{1-1/k}}}. \quad (96)$$

It is now be apparent that we would like $1 - p_i$ to be proportional to the RHS of the above. There are a few adjustments that need to be made for consistency, such as introducing a minimum, dividing by L and subtracting a factor of λ, and including these gives us our final definition as

$$1 - p_i := \min\left\{\frac{\lambda}{h_i L}\left(\sqrt{N_B\left(\frac{\epsilon}{\lambda t}\right)^{1-1/2k}\left(\frac{2k+\Upsilon}{2k+1}\right)^{1/2k}\frac{\Upsilon^{1/2k}}{2^{1-1/k}}} - 1\right), 1\right\}. \quad (97)$$

For convenience, we define

$$\chi := \frac{\lambda}{L} \left(\sqrt{N_B \left(\frac{\epsilon}{\lambda t} \right)^{1-1/2k} \left(\frac{2k+\Upsilon}{2k+1} \right)^{1/2k} \frac{\Upsilon^{1/2k}}{2^{1-1/k}}} - 1 \right) \tag{98}$$

such that $1 - p_i = \min\left\{ \frac{\chi}{h_i}, 1 \right\}$. We will also define

$$\mathcal{S} := \{ i : 1 - p_i < 1 \} \tag{99}$$

which consists of the indices of terms that have a non-zero probability of being in the Trotter channel. The complement of $\mathcal{S}$, denoted $\mathcal{S}^C$, contains the remaining indices of $\{1, 2, \ldots, L\}$ $\mathcal{S}$ that are guaranteed to be placed in the QDrift channel. We now show that the definition of our probabilities leads to the correct bound in (96)

$$\sum_i (1 - p_i) h_i = \sum_{i \in \mathcal{S}} \frac{\chi}{h_i} h_i + \sum_{i \in \mathcal{S}^C} h_i \tag{100}$$

$$= \chi |\mathcal{S}| + \lambda_{\mathcal{S}^C} \tag{101}$$

$$= \frac{\lambda |\mathcal{S}|}{L} \left(\frac{1}{2} \left(\frac{4k+2\Upsilon}{2k+1} \right)^{1/4k} \sqrt{N_B 2^{1+1/2k} \Upsilon^{1/2k} \left(\frac{\epsilon}{\lambda t} \right)^{1-1/2k}} - 1 \right) + \lambda_{\mathcal{S}^C} \tag{102}$$

$$\leq \frac{\lambda}{2} \left(\frac{4k+2\Upsilon}{2k+1} \right)^{1/4k} \sqrt{N_B 2^{1+1/2k} \Upsilon^{1/2k} \left(\frac{\epsilon}{\lambda t} \right)^{1-1/2k}} + (\lambda_{\mathcal{S}^C} - \lambda) \tag{103}$$

$$\leq \lambda \sqrt{N_B \left(\frac{\epsilon}{\lambda t} \right)^{1-1/2k} \left(\frac{2k+\Upsilon}{2k+1} \right)^{1/2k} \frac{\Upsilon^{1/2k}}{2^{1-1/k}}} \tag{104}$$

as desired. In the step from (102) to (103) we assumed that the innermost term is positive. This is necessary for our probabilities to be greater than 0, but is not necessarily true as $\frac{\epsilon}{\lambda t} \to 0$ for fixed N_B. Therefore by requiring $1 - p_i > 0$ we introduce the following lower bound on N_B

$$N_B \geq \left(\frac{\lambda t}{\epsilon} \right)^{1-1/2k} \left(\frac{2k+1}{2k+\Upsilon} \right)^{1/2k} \frac{2^{1-1/k}}{\Upsilon^{1/2k}}. \tag{105}$$

We have therefore satisfied both guarantees as outlined in the statement of the Theorem, which completes the proof. $\qquad\square$

There are a few comments to be made about the behavior of some of the quantities introduced in the above lemma. First we look at the lower bound on N_B, which scales as $\Theta \left(\left(\frac{\lambda t}{\epsilon} \right)^{1-1/2k} \right)$. If we assume that t, ϵ are independent of L then this overall scaling is sublinear with respect to L as $\lambda \leq \max_i h_i L$, which indicates that by moving a term from Trotter to QDrift we do not automatically lose out in gate cost. If t, ϵ are dependent on L then we cannot make the same guarantee. Second, if we parametrize N_B to be within a constant factor of this lower bound $N_B = (1+c)^2 \left(\frac{\lambda t}{\epsilon} \right)^{1-1/2k} \left(\frac{2k+1}{2k+\Upsilon} \right)^{1/2k} \frac{2^{1-1/k}}{\Upsilon^{1/2k}}$, we can then simplify the expression for χ as

$$\chi = \frac{\lambda}{L} \left(\sqrt{N_B \left(\frac{\epsilon}{\lambda t} \right)^{1-1/2k} \left(\frac{2k+\Upsilon}{2k+1} \right)^{1/2k} \frac{\Upsilon^{1/2k}}{2^{1-1/k}}} - 1 \right) \tag{106}$$

$$= c \frac{\lambda}{L}. \tag{107}$$

This is a nice simplification that we will use later and shows χ can be thought of as an "average strength" of the overall Hamiltonian. This then gives the intuition that our probability definitions are somewhat analogous to an inverse importance sampling procedure with respect to the spectral norms h_i. We note that as $c \to \infty$, $1 - p_i = \min\left\{ \frac{c\lambda}{L h_i}, 1 \right\} \to 1$, which implies that $p_i \to 0$. This means that as we increase the number of QDrift samples the probability distribution will put more

probability mass into the QDrift partition. Intuitively, the resulting partition takes advantage of having a low-error QDrift simulation by placing more terms into it's partition. Contrast this behavior with the opposite limit, $c \to 0$. In this case we have a very noisy QDrift partition, due to the QDrift error scaling as $1/N_B$, and we see that the distribution in this case properly places more probability mass into the Trotter partition.

5.3 Comparison with Trotter and QDrift

In this section we shift our focus to analyzing when a two term Composite channel can outperform a simulation of just a Trotter or QDrift channel. The first result we show gives asymptotic bounds on certain quantities that result from a partitioning, such as λ_B, that yield asymptotic improvements for the overall Composite channel query cost over either Trotter or QDrift. After analyzing the asymptotic cost for a predetermined partition we then investigate when the probabilistic partitioning scheme introduced above can yield gate cost improvements. We find that in the extremal cases for the probabilities (i.e $p_i \to 0$ or $p_i \to 1$ for all i) our resulting Composite channel exactly matches the QDrift and second-order Trotter costs, with minor constant factors for higher-order Trotter channels. Finally, in the case where a Hamiltonian has spectral norms that decay exponentially, i.e. $h_i = 2^{-i}$, in expectation our partitioning meets the requirements for the asymptotic improvements shown in Theorem 2.2 under the condition that t/ϵ is such that $C_{Trott} = C_{QD}$.

5.3.1 Deterministic Partitioning Improvements

We first investigate the asymptotic bounds that a partition must satisfy in order to offer asymptotic improvements over either Trotter or QDrift.

Theorem 2.2 (Conditions for Composite Channel Improvements). *Assume a Hamiltonian H along with a partitioning scheme to generate A and B that varies with L. For a simulation time t with a desired diamond distance error at most ϵ, let $\beta > 0$ be a number such that $C_{QD} = C_{Trott}^{\beta}$. There exist asymptotic regimes for the parameters $L_A, \lambda_B,$ and N_B such that*

$$C_{Comp} \in o(\min\{C_{Trott}, C_{QD}\}),$$

outlined below for the cases when $C_{QD} \geq C_{Trott}$ ($\beta \geq 1$) and $C_{QD} < C_{Trott}$ ($0 < \beta < 1$).
For the case when $\beta > 1$, indicating Trotter uses fewer queries than QDrift, if the parameters $\lambda_B, L_A,$ and number of QDrift samples N_B satisfy the following

1. *$L_A(1 - q_B)^{1/2k} \in o(L)$ where $q_B = \frac{\alpha_{comm}(B, 2k)}{\alpha_{comm}(H, 2k)}$,*

2. *$\lambda_B \in o\left(\lambda^{1/\beta}\left(\frac{\sqrt{\epsilon}}{t}\right)^{1-1/\beta}\right),$*

3. *$N_B \in \Omega(L_A)$ and $N_B \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right),$*

then we have that $C_{comp} \in o(C_{Trott})$.
If instead $0 < \beta < 1$, indicating $C_{QD} < C_{Trott}$, and the parameters λ_B, L_A and N_B satisfy the following

1. *the total number of terms in the Trotter partition satisfies*

$$L_A \in o\left(L^{1/\beta}\left(\frac{\epsilon^{1-1/\beta}}{t^{(2k+1)(1-1/\beta)}} \frac{\alpha_{comm}^{1/\beta}(H)}{\alpha_{comm}(A) + \alpha_{comm}(\{A, B\})}\right)^{1/2k}\right), \quad (5)$$

2. *$\lambda_B \in o(\lambda),$*

3. *and $N_B \in \Theta(L_A),$*

then we have that $C_{comp} = o(C_{QD})$. *Note that for* $\beta = 1$ *exactly, the conditions on* λ_B *and* L_A *are the same for both cases:* $L_A \in o(L)$ *and* $\lambda_B \in o(\lambda)$. *The conditions for* N_B *are satisfied by* $N_B \in \Theta(L_A)$. *If these conditions are satisfied for all ranges of* β *then* $C_{comp} \in o(\min\{C_{QD}, C_{Trott}\})$.

Proof. We start with the expression for the Composite channel cost from Theorem 2.1

$$C_{comp} \leq \Upsilon(\Upsilon L_A + N_B) \left[C_{Trott}(H, t, \epsilon, 2k) \frac{(1-q_B)^{1/2k}}{\Upsilon^{1-1/2k}L} + C_{QD}(H, t, \epsilon) \frac{\Upsilon}{N_B} \frac{\lambda_B^2}{\lambda^2} \right], \qquad (108)$$

where $1 - q_B$ quantifies the contribution of the A partition to the overall nested commutator structure $\alpha_{comm}(H, 2k)$. The most straightforward way to proceed is to split this expression based on our two cases, $\beta > 1$ and $0 < \beta < 1$.

We first examine the case when $\beta > 1$, which implies $C_{Trott} < C_{QD}$. Our composite cost expression then can be written as

$$C_{comp} \leq C_{Trott} \left(\Upsilon^{1+1/2k}(1-q_B)^{1/2k}\frac{L_A}{L} + \Upsilon^3 \frac{\lambda_B^2}{\lambda^2}\frac{L_A}{N_B}C_{QD}^{1-1/\beta} + \Upsilon^{1/2k}(1-q_B)^{1/2k}\frac{N_B}{L} + \Upsilon^2\frac{\lambda_B^2}{\lambda^2}C_{QD}^{1-1/\beta} \right).$$
$$(109)$$

Our goal is to show that if each of the terms above are in $o(1)$, then the sum is in $o(1)$ which implies $C_{comp} \in o(C_{Trott})$. Starting with the simplest term

$$\Upsilon^{1+1/2k}\frac{(1-q_B)^{1/2k}L_A}{L} \in o(1), \qquad (110)$$

which holds when $L_A(1-q_B)^{1/2k} \in o(L)$. The second term we analyze is the QDrift only term

$$\Upsilon^2\frac{\lambda_B^2}{\lambda^2}C_{QD}^{1-1/\beta} = \left(\frac{t^2}{\epsilon}\right)^{1-1/\beta}\frac{\lambda_B^2}{\lambda^{1/\beta}} \in o(1), \qquad (111)$$

where we used the assumption that $\lambda_B \in o\left(\lambda^{1/2\beta}\left(\frac{\sqrt{\epsilon}}{t}\right)^{1-1/\beta}\right)$ to reduce the expression to $o(1)$. We now move on to the remaining terms involving N_B. The first we can simplify is

$$\Upsilon^{1/2k}(1-q_B)^{1/2k}\frac{N_B}{L} \in o(1), \qquad (112)$$

by the assumption that $N_B(1-q_B)^{1/2k} \in o(L)$. The second term involving N_B is

$$\Upsilon^3\frac{C_{QD}^{1-1/\beta}\lambda_B^2}{\lambda^2}\frac{L_A}{N_B} \in o(1), \qquad (113)$$

where we used the fact that $\frac{C_{QD}^{1-1/\beta}\lambda_B^2}{\lambda^2} \in o(1)$ and the assumption that $N_B \in \Omega(L_A)$. Given that all four terms in the cost function expression 108 are $o(1)$, we have shown that $C_{comp} \in o(C_{Trott})$ for $\beta > 1$.

We can now move on to the case when $0 < \beta < 1$, which essentially repeats the above logic. In this situation the cost expression from Eq (108) reduces to

$$C_{comp} \leq C_{QD} \left(\Upsilon^{1+1/2k}\frac{(1-q_B)^{1/2k}C_{Trott}^{1-1/\beta}L_A}{L} + \Upsilon^3\frac{\lambda_B^2}{\lambda^2}\frac{L_A}{N_B} + \Upsilon^{1/2k}(1-q_B)^{1/2k}\frac{N_B}{L}C_{Trott}^{1-1/\beta} + \Upsilon^2\frac{\lambda_B^2}{\lambda^2} \right).$$
$$(114)$$

Starting with the rightmost term we have

$$\Upsilon^2\frac{\lambda_B^2}{\lambda^2} \in o(1), \qquad (115)$$

which is guaranteed by the assumption $\lambda_B \in o(\lambda)$. We then look at the only other term that does not include a factor of N_B, which reduces to

$$\Upsilon^{1+1/2k}(1-q_B)^{1/2k}\frac{L_A}{L}C_{Trott}^{1-1/\beta} = \Upsilon^{2+1/2k}\left(\frac{t^{1+1/2k}}{\epsilon^{1/2k}}\right)^{1-1/\beta}\frac{L_A}{L^{1/\beta}}\left(\frac{\alpha_{comm}(A, 2k) + \alpha_{comm}(\{A, B\}, 2k)}{\alpha_{comm}^{1/\beta}(H, 2k)}\right)^{1/2k}.$$
$$(116)$$

This can be shown to be in $o(1)$ given Assumption 1 for $0 < \beta < 1$. We can utilize the fact that this term is vanishing to reduce one of the other terms involving C_{Trott} as

$$\Upsilon^{1/2k} N_B \frac{(1 - q_B)^{1/2k} C_{Trott}^{1-1/\beta}}{L} = \frac{N_B}{L_A} \cdot o(1) \in o(1), \tag{117}$$

where we use the assumption that $N_B \in O(L_A)$ for the last step. The last remaining term is rather straightforward

$$\Upsilon^2 \frac{\lambda_B^2}{\lambda^2} \frac{L_A}{N_B} = \frac{L_A}{N_B} \cdot o(1) \in o(1), \tag{118}$$

where we used $N_B \in \Omega(L_A)$ as well as $\lambda_B \in o(\lambda)$. Taken together, we have shown that $C_{comp} = C_{QD} \cdot o(1)$, implying $C_{comp} \in o(C_{QD})$ for $0 < \beta < 1$.

This combined with the above proof for $\beta > 1$ implies that $C_{comp} \in o(\min\{C_{QD}, C_{Trott}\})$ if the assumptions in the statement are met, completing the proof. $\square$

	$C_{QD} > C_{Trott}$	$C_{QD} < C_{Trott}$
$L_A \in$	$o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$	$o\left(L^{1/\beta}\left(\frac{\epsilon^{1-1/\beta}}{t^{(2k+1)(1-1/\beta)}} \frac{\alpha_{comm}^{1/\beta}(H)}{\alpha_{comm}(A) + \alpha_{comm}(\{A,B\})}\right)^{1/2k}\right)$
$\lambda_B \in$	$o\left(\lambda^{1/\beta}\left(\frac{\sqrt{\epsilon}}{t}\right)^{1-1/\beta}\right)$	$o(\lambda)$
(Lower Bound) $N_B \in$	$\Omega(L_A)$	$\Omega(L_A)$
(Upper Bound) $N_B \in$	$o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$	$O(L_A)$

Table 1: Summary of asymptotic requirements for parameters of interest when $C_{QD} = C_{Trott}^{\beta}$ to yield $C_{Comp} \in o(\min\{C_{QD}, C_{Trott}\})$.

Now that we have given bounds on a partitioning that can yield asymptotic improvements, we turn to the problem of finding a good value of N_B that can satisfy these necessary assumptions. We first show a straightforward generalization of the first-order optimal value of N_B to the higher-order Trotter case.

Lemma 10. *Given a Hamiltonian H with fixed partitions A and B, the optimal number of QDrift samples N_B is given as*

$$N_B = 2\lambda_B \sqrt{\left(\frac{\Upsilon t}{\epsilon}\right)^{1-1/2k} L_A \left(\frac{2k+1}{\Upsilon \alpha_{comm}(A) + \alpha_{comm}(\{A,B\})}\right)^{1/2k}} \tag{119}$$

Proof. This is a rather straightforward result which follows the logic of the first-order results in section 4.3 so we will simply show some of the intermediate steps. Starting with the upper bound on C_{Comp}

$$C \leq \Upsilon(\Upsilon L_A + N_B)\left(\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} + \frac{Q(t)}{\epsilon}\right), \tag{120}$$

we then require that the upper bound be an optima with respect to N_B, namely that $\frac{\partial}{\partial N_B} C_{Comp} = 0$. Computing the derivative using the expressions for $Q(t)$ and $P(t)$, given in Eqs. (68) and (67) respectively, we find

$$0 = \Upsilon\left(\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} + \frac{Q(t)}{\epsilon}\right) - \Upsilon(\Upsilon L_A + N_B)\frac{Q(t)}{N_B^2} \tag{121}$$

$$\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} N_B^2 = \Upsilon L_A \frac{4\lambda_B^2 t^2}{\epsilon} \tag{122}$$

$$N_B = 2\lambda_B \sqrt{\frac{\Upsilon L_A t^2}{\epsilon^{1-1/2k} P(t)^{1/2k}}}. \tag{123}$$

Plugging in the expression for $P(t)$ yields the stated result. $\square$

Now that we have given an expression for an optimal value of N_B given a partitioning, we briefly show that this N_B value satisfies the assumptions from Theorem 2.2 for the case when $C_{QD} = C_{Trott}$. As mentioned in the theorem statement, for this situation we note that $N_B \in \omega(L_A)$ and $N_B \in o(L)$ is sufficient to satisfy the requirements. The equal cost condition yields the equation

$$C_{QD} = C_{Trott} \tag{124}$$

$$\left(\frac{t}{\epsilon}\right)^{1-1/2k} = \frac{L\alpha_{comm}^{1/2k}(H, 2k)}{\lambda^2} \cdot O(1). \tag{125}$$

This then simplifies the optimal value for N_B as

$$N_B = 2\lambda_B \sqrt{\left(\frac{\Upsilon t}{\epsilon}\right)^{1-1/2k} L_A \left(\frac{2k+1}{\Upsilon \alpha_{comm}(A, 2k) + \alpha_{comm}(\{A, B\}, 2k)}\right)^{1/2k}} \tag{126}$$

$$= \frac{\lambda_B}{\lambda}\sqrt{L_A L}(1 - q_B)^{1/2k} \cdot O(1). \tag{127}$$

This last expression is clearly seen to lie in both $\omega(L_A)$ and $o(L)$ if $\lambda_B \in o(\lambda)$ and $L_A \in o(L)$. This shows that at times t and errors ϵ where QDrift and Trotter have equal query cost the optimal value of N_B leads to a Composite channel that has asymptotic improvements over either of the channels it composes.

5.3.2 Probabilistic Partitioning Performance

We now focus on showing that a Composite channel with a probabilistic partitioning scheme according to Lemma 2.1 can lead to gate cost savings compared to either Trotter or QDrift channels alone. The main expression we will be working with is

$$\mathbb{E}\left[C_{comp}(H, t, \epsilon, N_B)\right] \leq \mathbb{E}\left[(L_A + N_B)\left(\frac{P(t)^{1/2k}}{\epsilon^{1/2k}} + \frac{Q(t)}{\epsilon}\right)\right] \tag{128}$$

$$\leq \sqrt{\mathbb{E}\left[L_A^2\right]\frac{\mathbb{E}\left[P(t)^{1/k}\right]}{\epsilon^{1/k}}} + \sqrt{\mathbb{E}\left[N_B^2\right]\frac{\mathbb{E}\left[P(t)^{1/k}\right]}{\epsilon^{1/k}}}$$

$$+ \sqrt{\mathbb{E}\left[L_A^2\right]\frac{\mathbb{E}\left[Q(t)^2\right]}{\epsilon^2}} + \sqrt{\mathbb{E}\left[N_B^2\right]\frac{\mathbb{E}\left[Q(t)^2\right]}{\epsilon^2}} \tag{129}$$

$$\leq \left(\sqrt{\mathbb{E}\left[L_A^2\right]} + \sqrt{\mathbb{E}\left[N_B^2\right]}\right)\left(\frac{\mathbb{E}\left[P(t)^{1/2k}\right]}{\epsilon^{1/2k}} + \frac{\sqrt{\mathbb{E}\left[Q(t)^2\right]}}{\epsilon}\right), \tag{130}$$

which follows from repeated applications of the Cauchy-Schwarz inequality and Jensen's inequality for $f(x) = x^{1/k}$ for positive x. The approach we take is to first give useful upper bounds on this expression and then show that these bounds saturate to the costs for Trotter and QDrift as we shift our probability mass towards either 0 or 1 for each term. Finally, we investigate a specific Hamiltonian that in expectation satisfies the assumptions for asymptotic improvements from Theorem 2.2.

In this section we will present some lemmas that provide useful upper bounds to the quantities used in the above expected cost expression. Since these proofs are relatively straightforward applications of the definitions given above we leave the proofs to appendix A.

Lemma 11. *Let L_A denote the number of terms simulated with a higher-order Trotter formula in a Composite channel with the assumptions as defined in section 3. Using a probabilistic partitioning scheme as defined in 2.1, then the second moment of L_A obeys*

$$\mathbb{E}(L_A^2) \leq |\mathcal{S}|^2 - |\mathcal{S}|\sum_{i \in \mathcal{S}}\frac{\chi}{h_i},$$

where χ is given in 98 and $\mathcal{S}$ is the set of p_i such that $1 - p_i < 1$ as given in 99.

This result makes intuitive sense, as L_A can be no larger than $|\mathcal{S}|$ as this is the set of terms that have non-zero probability of ending up in the Trotter partition. This expression then adjusts this upper bound by $|\mathcal{S}| \sum_{i \in \mathcal{S}} (1 - p_i)$, which can be thought of as the expected number of terms that will end up being placed in QDrift from $\mathcal{S}$.

We know move on to bounding the expected contribution from the QDrift channel due to the error, namely $Q(t)$.

Lemma 12. *Let Q denote the contribution to the error of a Composite channel, where Q is defined in 68, with the standard assumptions from Section 3 and a probabilistic partitioning scheme as outlined in 2.1. The following upper bounds hold on the first two moments for Q*

$$\mathbb{E}\left[Q(t)\right] \leq \frac{4t^2}{N_B} \left(\chi \lambda_{\mathcal{S}} + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2 \right) \tag{131}$$

$$\mathbb{E}\left[Q(t)^2\right] \leq \frac{4t^2\lambda^2}{N_B} \mathbb{E}\left[Q(t)\right] \leq \frac{16t^4\lambda^2}{N_B^2} \left(\chi \lambda_{\mathcal{S}} + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2 \right) \tag{132}$$

The main intuition to be gained from here is that as $|\mathcal{S}| \to 0$ the dominant contribution comes from $\lambda_{\mathcal{S}^C}$ which is to be expected as this is the set of terms that are guaranteed to be placed in QDrift.

We now move on to the last term we need to bound, which is P.

Lemma 13. *Let $P(t)$ denote the product formula error scaling as defined in 67 for a Composite channel with the standard assumptions as defined in 3. Using a probabilistic partitioning scheme as defined in 2.1 then the following upper bounds hold for the expected value of $P(t)$*

$$\mathbb{E}\left[P(t)\right] \leq \frac{(2\Upsilon)^{2+2k}}{2k+1} t^{2k+1} \lambda^{2k} \left(\lambda_{\mathcal{S}} - \chi|\mathcal{S}| \right). \tag{133}$$

Even though this upper bound on P is not tight at all we can still capture interesting edge case behavior. Note that as $\mathcal{S} \to \varnothing$ this indicates that all of our terms are in the set $\mathcal{S}^C$, meaning the probability they are in the Product Formula channel tends to 0, which is reflected as $\lambda_{\mathcal{S}} \to 0$ and $|\mathcal{S}| \to 0$ which implies that $\mathbb{E}\left[P(t)\right] \to 0$ appropriately. We will investigate the other regime in which $1 - p_i \to 0$ shortly.

Now that we have workable upper bounds we would like to make some direct comparisons between a Composite channel and both Trotter and QDrift. However before showing that this expression can outperform known upper bounds on Trotter or QDrift, we first would like to perform some consistency checks. What we need to show is that as we shift the probability mass over our partitions, the Composite channel cost should tend towards Trotter or QDrift, based on which direction we shift the mass. In other words, as $p_i \to 0$ for all i, we would like $\mathbb{E}\left[C_{Comp}\right] \to C_{QD}$ and as $p_i \to 1$ we would like $\mathbb{E}\left[C_{Comp}\right] \to C_{Trott}$. In the following Theorem we verify this intuition precisely for the QDrift regime and second-order Trotter formulas. For higher-order formulas we are able to show that as $p_i \to 1$ the Composite channel cost is at most 1.12 times worse than the Trotter cost.

Theorem 14 (Expected Cost reduction to Trotter and QDrift in proper limits). *Assume as inputs a Hamiltonian H, time t, and error ϵ. Let p_i denote the probability of assigning term H_i of H to the Trotter partition of a composite simulation channel, given by Lemma 2.1. Then as we vary the number of QDrift samples N_B the expected cost of the Composite channel, $\mathbb{E}\left[C_{Comp}\right]$, saturates towards the Trotter and QDrift gate costs in the respective limits for N_B. Specifically, as N_B approaches its lower bound in Eq. (105) then $\mathbb{E}\left[C_{Comp}\right] \to \left(\Upsilon^{1/2k}/2^{1-1/2k}\right) \cdot C_{Trott} \leq 1.12 C_{Trott}$ and as $N_B \to \infty$ then $\mathbb{E}\left[C_{Comp}\right] \to C_{QD}$ exactly. Note that for $k = 1$, or a second-order Trotter formula, $\Upsilon^{1/2k}/2^{1-1/2k} = 1$.*

Proof. The first limit we tackle is the QDrift regime. Dropping the specific probabilities that were defined in Lemma 2.1 we consider adjusting each probability $p_i \to 0$ for all i. We have three pieces we then need to upper bound: P, Q, and L_A. We can ignore $\sqrt{\mathbb{E}\left[N_B^2\right]}$ as there is no probability we do not have an empty or near-empty QDrift channel. The first observation is that as each $p_i \to 0$ and as we set more and more probabilities to $1 - p_i = 1$ then the size of $|\mathcal{S}|$ will decrease

accordingly. We consider the case where $|\mathcal{S}| \to 0$ as this represents the QDrift only channel. This limit then implies that $\lambda_\mathcal{S} \to 0$ and $\lambda_{\mathcal{S}^C} \to \lambda$. This allows us to vastly simplify many of our bounds

$$\mathbb{E}\left[L_A^2\right] \leq |\mathcal{S}|^2 - |\mathcal{S}| \sum_{i \in \mathcal{S}} \frac{\chi}{h_i} \to 0 \tag{134}$$

$$\mathbb{E}\left[P(t)^{1/2k}\right] \frac{(2\Upsilon)^{1+1/k}}{(2k+1)^{1/2k}} \lambda(\lambda_\mathcal{S} - \chi|\mathcal{S}|)^{1/2k} \to 0 \tag{135}$$

$$\mathbb{E}\left[Q(t)^2\right] \leq \frac{16t^4\lambda^2}{N_B^2}\left(\chi\lambda_\mathcal{S} + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2\right) \to \frac{16t^4\lambda^4}{N_B^2}. \tag{136}$$

By plugging these into the cost expression from Eq. 130 we get the following limit

$$\mathbb{E}\left[C_{comp}(H, t, \epsilon, N_B)|p_i \to 0\ \forall i\right] \leq \left(\sqrt{\mathbb{E}\left[L_A^2\right]} + \sqrt{\mathbb{E}\left[N_B^2\right]}\right)\left(\frac{\mathbb{E}\left[P(t)^{1/2k}\right]}{\epsilon^{1/2k}} + \frac{\sqrt{\mathbb{E}\left[Q(t)^2\right]}}{\epsilon}\right) \tag{137}$$

$$\to \frac{4t^2\lambda^2}{\epsilon}, \tag{138}$$

which exactly matches the upper bounds on the cost of a QDrift only channel. We remind the reader of the extraneous factor of 2 that we incur from restricting $\epsilon \in (0, \ln(2)\lambda t)$ as opposed to Campbell's original result [14].

Next we look at the limit in which $p_i \to 1$ and we are performing a Trotter only channel. By using the probability distributions from Lemma 2.1 it is most straightforward to consider this limit by using a parametrized value for N_B

$$N_B(c) = (1+c)^2\left(\frac{\lambda t}{\epsilon}\right)^{1-1/2k}\frac{1}{4(2\Upsilon)^{1+1/2k}}\left(\frac{2k+1}{4k+2\Upsilon}\right)^{1/2k}, \tag{139}$$

which was shown to lead to $\chi = c\frac{\lambda}{L}$. For probabilities which are in $\mathcal{S}$, we can say $p_i = 1 - c\frac{\lambda}{Lh_i}$. This then implies that by sending $c \to 0$ we appropriately have $\chi \to 0$ and $p_i \to 1$. We can use this to compute the limiting upper bound on $\mathbb{E}\left[L_A^2\right]$ as

$$\mathbb{E}\left[L_A^2\right] \leq |\mathcal{S}|^2 - |\mathcal{S}| \sum_i \chi/h_i \to L^2. \tag{140}$$

The contribution due to the QDrift error is also straightforward in the limiting case as $c \to 0$

$$\mathbb{E}\left[Q(t)^2\right] \leq \frac{16t^4\lambda^2}{N_B^2}\left(\chi\lambda_\mathcal{S} + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2\right) \to 0, \tag{141}$$

as $\chi \to 0$ and $\lambda_{\mathcal{S}^C} \to 0$. The remaining simple upper bound is for P which is as follows

$$\mathbb{E}\left[P(t)^{1/2k}\right] \leq \frac{(2\Upsilon)^{1+1/k}}{(2k+1)^{1/2k}}t^{1+1/2k}\lambda(\lambda_\mathcal{S} - \chi|\mathcal{S}|)^{1/2k} \to \frac{(2\Upsilon)^{1+1/k}}{(2k+1)^{1/2k}}(\lambda t)^{1+1/2k}. \tag{142}$$

The last quantity we have to bound is $\sqrt{\mathbb{E}\left[N_B^2\right]}$, which up until now has been harmlessly upper bounded by N_B. However we now have a very high probability, tending to 1, of not having any terms in our QDrift channel. This means that the expected number of exponential gates performed during the QDrift sampling procedures should tend to 0. We make this rigorous through a simple definition of expectation

$$\mathbb{E}\left[N_B^2\right] = \Pr\left[|B| = 0\right] \cdot 0 + \Pr\left[|B| = 1\right] \cdot 1 + \Pr\left[|B| > 1\right] \cdot N_B^2 \tag{143}$$

$$= \sum_i \prod_{j \neq i}(1-p_i)p_j + (1 - \Pr\left[|B| = 0\right] - \prod_{j \neq i}(1-p_i)p_j) \cdot N_B^2 \tag{144}$$

$$\to 0 + (1 - \prod_i p_i) \cdot N_B^2 \tag{145}$$

$$\to 0. \tag{146}$$

Intuitively, the above is simply a result of $\Pr\left[|B| = 0\right] \to 1$ as $p_i \to 1$, resulting in all other terms tending to 0. This gives the final upper bound on our expected cost as

$$\mathbb{E}\left[C_{comp}(H, t, \epsilon, N_B)|p_i \to 1 \; \forall i\right] \leq \left(\sqrt{\mathbb{E}\left[L_A^2\right]} + \sqrt{\mathbb{E}\left[N_B^2\right]}\right) \left(\frac{\mathbb{E}\left[P(t)^{1/2k}\right]}{\epsilon^{1/2k}} + \frac{\sqrt{\mathbb{E}\left[Q(t)^2\right]}}{\epsilon}\right) \quad (147)$$

$$\to \frac{(L\lambda t)^{1+1/2k}}{\epsilon^{1/2k}} \frac{(2\Upsilon)^{1+1/k}}{(2k+1)^{1/2k}}. \quad (148)$$

We note that by using similar upper bounds on the factor of α_{comm} for the Trotter gate cost yields the following

$$C_{Trott} \leq \frac{Lt^{1+1/2k}}{\epsilon^{1/2k}} \alpha_{comm}^{1/2k}(H, 2k) \frac{\Upsilon^{1+1/2k} 4^{1/2k}}{(2k+1)^{1/2k}} \quad (149)$$

$$= \frac{L(\lambda t)^{1+1/2k}}{\epsilon^{1/2k}} \frac{\Upsilon^{1+1/2k} 2^{2+1/2k}}{(2k+1)^{1/2k}}. \quad (150)$$

The simple ratio $\mathbb{E}\left[C_{comp}\right]/C_{trott} = \Upsilon^{1/2k} 2^{-1+1/2k} = 2^{-1+1/k} 5^{1/2-1/2k}$ shows that for $k = 1$ we exactly match the Trotter cost. For higher-orders this constant factor is negligible and is upper bounded by 1.12. $\qquad\square$

5.3.3 Exponentially Decaying Hamiltonian Terms

Now that we have shown that the probabilistic partitioning scheme leads to a composite cost that saturates to C_{Trott} and C_{QD} in the appropriate regime, we will analyze a situation in the middle of these two limits. This is rather difficult to do in generality, so we will restrict our attention to Hamiltonians that have exponentially decaying spectral norms for each term, i.e. $h_i = 2^{-i}$ when sorted by spectral norm, and at the crossover time t such that $C_{QD} = C_{Trott}$. For this analysis we would like to show that the parameters of the partitioning, namely L_A, N_B, and λ_B, satisfy the constraints for asymptotic improvements given by Theorem 2.2. At the crossover time these constraints are simply $L_A \in o(L)$, $\lambda_B \in o(\lambda)$ and $N_B \in \Theta(L_A)$. We prove that in expectation these parameters will satisfy these constraints, then by using simple tail bounds we show that the probability the parameters fall into an asymptotically bad regime is vanishingly small.

Theorem 15. *Given a Hamiltonian $H = \sum_i h_i H_i$ with exponentially decaying spectral norms $h_i = 2^{-i}$, a small commutator structure $\alpha_{comm} \in \Theta\left(\frac{\log_2 L}{L}\right)$, time t, and error ϵ such that $C_{Trott}(H, t, \epsilon) = C_{QD}(H, t, \epsilon)$, then the probabilistic partitioning scheme given by Lemma 2.1 yields a partition that satisfies the asymptotic requirements of Theorem 2.2 with high probability.*

Proof. The outline of the proof is to first provide a parametrization of N_B that simplifies the calculation, then use this to bound expectations of L_A and λ_B, then finally to use tail bounds to show that the expectation values are sufficient to work with. We first resolve a conundrum involving N_B. As we have seen in Section 5.3.2, the value of N_B essentially determines the partitioning, however in Theorem 2.2 we see that N_B has to satisfy certain asymptotic requirements that depend on the partitioning! To get around this we introduce the following parametrization

$$N_B(c) = (1 + 2^{-c})^2 \left(\frac{\lambda t}{\epsilon}\right)^{1-1/2k} \left(\frac{2k+1}{2k+\Upsilon}\right)^{1/2k} \frac{2^{1-1/k}}{\Upsilon^{1/2k}}, \quad (151)$$

where $c \in \Theta(1)$ is a constant and can be negative. This specific form of $(1 + 2^{-c})^2$ leads to many simplifications, the first of which is $\chi = 2^{-c}\lambda L^{-1}$, following the definition of χ from Eq. (3). This allows us to compute the size of the sampling set of indices $|\mathcal{S}|$ as we just need to determine which j leads to $1 - p_j < 1$. This is shown as

$$1 - p_j < 1 \quad (152)$$

$$\frac{\chi}{h_j} < 1 \quad (153)$$

$$2^j < 2^c \frac{L}{\lambda} \quad (154)$$

$$j < c + \log_2 L - \log_2 \lambda. \quad (155)$$

We next compute $\lambda = \sum_i h_i = \sum_i 2^{-i} = 1 - 2^{-L}$ to yield a value for $\mathbb{E}[L_A]$. Since $\log(1-x) \in O(x)$ for $x \to 0$, it is easy to see that $|S|$ is asymptotically small compared to L:

$$|S| = \lfloor j \rfloor \le c + \log_2(L) + O\left(2^{-L}\right) \in \Theta\left(\log_2 L\right).$$

This shows that $\mathbb{E}[L_A]$ satisfies it's asymptotic requirements:

$$\mathbb{E}[L_A] = \sum_{i \in \mathcal{S}} \mathbb{E}[I_i] \tag{156}$$

$$= \sum_{i \in \mathcal{S}} p_i \tag{157}$$

$$= \sum_{i \in \mathcal{S}} 1 - (1 - p_i) \tag{158}$$

$$= |\mathcal{S}| - \sum_{i \in \mathcal{S}} \frac{\chi}{h_i} \tag{159}$$

$$= |\mathcal{S}| - \frac{2^{-c}\lambda}{L} \sum_{i \in \mathcal{S}} 2^{-i} \tag{160}$$

$$\in \Theta(|\mathcal{S}|) \tag{161}$$

$$= \Theta(\log_2 L) \subset o(L). \tag{162}$$

The next task is to show that L_A does not deviate from $\mathbb{E}[L_A]$ in an asymptotically significant way with high probability. As L_A is the sum of Bernoulli random variables we can use the multiplicative Chernoff bound

$$\Pr\left[|L_A - \mathbb{E}[L_A]| > \delta \mathbb{E}[L_A]\right] \le 2 e^{-\delta^2 \mathbb{E}[L_A]/3} \tag{163}$$

$$\in \Theta\left(\frac{1}{L^{\delta^2}}\right). \tag{164}$$

We see that this vanishes even for constant deviations $\delta \in O(1)$ of L_A from $\mathbb{E}[L_A]$, implying that any asymptotically significant deviations, such as $\delta \in O(L^a)$, vanish even quicker.

We now move on to bounding $\mathbb{E}[\lambda_B]$. The first thing we need to do is provide an upper bound on $|\mathcal{S}^C|$, which is done via a lower bound on $|\mathcal{S}|$. For this we can use the simple bound $\lfloor x \rfloor \ge x - 1$. This allows us to compute the required sums for $\mathbb{E}[\lambda_B]$:

$$\mathbb{E}[\lambda_B] = \sum_{i \in \mathcal{S}} (1 - p_i) h_i + \sum_{i \in \mathcal{S}^C} h_i \tag{165}$$

$$\le \chi|\mathcal{S}| + \sum_{i=c+\log_2 L - \log_2 \lambda - 1}^{L} 2^{-i} \tag{166}$$

$$\le \frac{2^{-c}\lambda}{L}\left(c + \log_2 L - \log_2 \lambda\right) + 2^{1-c-\log_2(L)+\log_2(\lambda)+1} - 2^{-L} \tag{167}$$

$$= \frac{2^{-c}\lambda}{L}(c + \log_2 L - \log_2 \lambda) + \frac{2^{2-c}\lambda}{L} - 2^{-L} \tag{168}$$

$$\le \frac{2^{-c}\lambda}{L}(4 + c + \log_2 L - \log_2 \lambda) \tag{169}$$

$$\in O\left(\frac{(1 - 2^{-L})(4 + c + \log_2 L + \sum_{k=1}^{\infty} 2^{-kL}/k)}{L}\right) \tag{170}$$

$$\subseteq \tilde{O}\left(L^{-1}\right) \tag{171}$$

$$\subseteq o(1). \tag{172}$$

For the tail bound necessary on λ_B we need to show that with high probability it will be in $o(1)$. For this a straightforward application of Markov's inequality to bound the probability that λ_B is

greater than some constant suffices:

$$\Pr\left[\lambda_B \geq x\right] \leq \frac{\mathbb{E}\left[\lambda_B\right]}{x} \tag{173}$$

$$\leq \frac{2^{-c}\lambda}{xL}(4 + c + \log_2 L - \log_2 \lambda) \tag{174}$$

$$\in \widetilde{O}\left(\frac{1}{xL}\right), \tag{175}$$

which can be clearly seen to approach 1 as $L \to \infty$ for any constant value $x \in \Theta(1)$. Note that Markov's inequality is applicable due to our assumption that $h_i \geq 0$ for all i.

The last expression we need to satisfy is $N_B \in \Theta(L_A)$. To do so we will use the bounds on t, ϵ to find a useful expression for N_B

$$C_{QD} = C_{Trott} \tag{176}$$

$$\frac{4\lambda^2 t^2}{\epsilon} = \Upsilon^{2+1/2k} L \frac{t^{1+1/2k}}{\epsilon^{1/2k}} \alpha_{comm}(H, 2k)^{1/2k}(4/2k + 1)^{1/2k} \tag{177}$$

$$\left(\frac{\lambda t}{\epsilon}\right)^{1-1/2k} = \frac{\Upsilon^{2+1/2k}}{4^{1-1/2k}(2k+1)^{1/2k}} \frac{L\alpha_{comm}(H, 2k)^{1/2k}}{\lambda^{1/2k}}. \tag{178}$$

Now we plug this in to the parametrization of N_B given in Eq. (151)

$$N_B = (1 + 2^{-c})^2 \left(\frac{2k+1}{2k+\Upsilon}\right)^{1/2k} \frac{2^{1-1/k}}{\Upsilon^{1/2k}} \frac{\Upsilon^{2+1/2k}}{4^{1-1/2k}(2k+1)^{1/2k}} \frac{L\alpha_{comm}(H, 2k)^{1/2k}}{\lambda^{1/2k}} \tag{179}$$

$$\in \Theta\left(\frac{L\alpha_{comm}(H, 2k)^{1/2k}}{\lambda^{1/2k}}\right). \tag{180}$$

Since $\lambda \to 1$, we can drop it from the denominator in the asymptotic limit. It is then straightforward to see that if

$$\alpha_{comm}(H, 2k)^{1/2k} \in \Theta\left(\frac{\log_2 L}{L}\right),$$

then $N_B \in \Theta(\log_2 L) = \Theta(L_A)$, meaning that N_B satisfies its asymptotic requirements from Theorem 2.2. This completes the proof. $\qquad\square$

6 General Composite Channels

Now that we have gone over the details of how to apply a Composite simulation for the particular case of Trotter and QDrift, let us now consider how one could generalize this exact same idea to any combination of known simulation methods. This approach, we will see, provides a broader perspective within which LCU, multiproduct, Trotter and QDrift can be seen to be special cases of the larger framework.

Claim 16. *Assume a Hilbert space of the form $\mathcal{H}_A \otimes \mathcal{H}_{sys}$ where $\mathcal{H}_A$ is the Hilbert space of an ancillary system. Specifically, let N_t represent the number of different individual exponentials that are called in a single step and let W_j and V_j be for each $j \in 1, \ldots, N_t$ be unitary operations acting on disjoint subspace $\mathcal{H}_{A,j}$ such that $\bigcup_j \mathcal{H}_{A,j} = \mathcal{H}_A$. Next let $\mathrm{Sel}_j|p\rangle_j|\psi\rangle_{sys} = |p\rangle_j e^{-iH_j t}|\psi\rangle_{sys}$ select a particular time evolution based on the index stored in the jth register then any Composite channel consisting of randomized Trotter formulas, deterministic Trotter formulas, multiproduct formulas and QDrift can be thought of as special cases of the following template*

$$\mathcal{U}_{comp} : |0\rangle\!\langle 0| \otimes \rho \mapsto \mathrm{Tr}_A\left(\prod_{j=1}^{N_t}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1})\right)$$

Proof. First let us show that a segment of QDrift can be thought of as a special case of this template. In this case let us take $V_j = \mathbb{1}$ and W_j to be a unitary such that for a Hamiltonian of the form $H_B = \sum_{k=0}^{L-1} h_k H_k$ with $\sum_k h_k = \lambda_B$ and $h_k \geq 0$, then

$$W_j|0\rangle = \sum_{k=0}^{N-1} \sqrt{\frac{h_k}{\lambda_B}}|k\rangle. \tag{181}$$

We then have from some elementary algebra that

$$\mathrm{Tr}_{A_j}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1}) = \sum_k \frac{h_k}{\lambda_B}e^{-i\lambda_B H_k t}\rho e^{i\lambda_B H_k t} = \mathcal{U}_{QD}(H_B, t) \circ \rho. \tag{182}$$

As each Hilbert space $\mathcal{A}_j$ is assumed to be disjoint, we then have that this also holds for the multi-segment case as well. Specifically, let us assume that for some positive integer Q that the channel maps

$$|0\rangle\langle0| \otimes \rho \mapsto \mathrm{Tr}_A\left(\prod_{j=1}^{Q}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1})\right) = \mathcal{U}_{QD}(H_B, t)^{\circ Q} \circ \rho \tag{183}$$

We then have using the fact that each of the V_j and W_j act on disjoint Hilbert spaces that

$$\mathrm{Tr}_A\left(\prod_{j=1}^{Q+1}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1})\right)$$

$$= \mathrm{Tr}_{A_{Q+1}}\mathrm{Tr}_{\bigcup_{q=1}^{Q} A_q}\left(\prod_{j=1}^{Q+1}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1})\right)$$

$$= \mathrm{Tr}_{A_{Q+1}}\Bigg((V_{Q+1} \otimes \mathbb{1})\mathrm{Sel}_{Q+1}(W_{Q+1} \otimes \mathbb{1})$$

$$\times \mathrm{Tr}_{\bigcup_{q=1}^{Q} A_q}\left(\prod_{j=1}^{Q}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1})\right)$$

$$\times (W_{Q+1}^\dagger \otimes \mathbb{1})\mathrm{Sel}_{Q+1}^\dagger(V_{Q+1}^\dagger \otimes \mathbb{1})\Bigg) \tag{184}$$

We then have from (183) and (184) that

$$\mathrm{Tr}_A\left(\prod_{j=1}^{Q+1}(V_j \otimes \mathbb{1})\mathrm{Sel}_j(W_j \otimes \mathbb{1})\rho(W_j^\dagger \otimes \mathbb{1})\mathrm{Sel}_j^\dagger(V_j^\dagger \otimes \mathbb{1})\right) = \mathcal{U}_{QD}(H_B, t)^{\circ Q+1} \circ \rho, \tag{185}$$

and thus the claim holds for all positive integer Q by induction since we already demonstrated the $Q = 1$ case.

Next we wish to show that any randomized Trotter formula can be thought of as a special case of this framework. While we have solely focused on deterministic Trotter formulas in this paper, we generalize to randomly ordered formulas in this section, as developed in [10], because the formalism required is the same as the deterministic case. We define the first-order Trotter channel given by a permutation $\sigma \in S_L$, where S_L is the symmetric group on L elements, as

$$\mathcal{U}_\sigma^{(1)}(H, t) : \rho \mapsto \prod_{i=1}^{L} e^{iH_{\sigma(i)}t}\rho\prod_{i=L}^{1} e^{-iH_{\sigma(i)}t}. \tag{186}$$

The first-order Randomized Trotter-Suzuki channel is then given as $\frac{1}{L!}\sum_{\sigma \in S_L}\mathcal{U}_\sigma^{(1)} \circ \rho$. Our goal is to show that this is clearly encodable within the framework developed here.

We start by noting that for a single iteration a single ancilla register is sufficient. For a Hamiltonian H_A we then take $V_{RTS} = \mathbb{1}$, $W_{RTS}|0\rangle = \sum_{\sigma \in S_{L_A}} \frac{1}{\sqrt{L_A!}}|\sigma\rangle$, where $|\sigma\rangle$ can be an arbitrary binary encoding of an index of S_{L_A}. Then define the action of the Select operator as $\text{Sel}_\sigma(|\sigma\rangle\langle\sigma| \otimes \rho)\text{Sel}_\sigma^\dagger = |\sigma\rangle\langle\sigma| \otimes \mathcal{U}_\sigma^{(1)} \circ \rho$. This then gives the overall action of the channel as

$$\text{Tr}_A\left((V_{RTS} \otimes \mathbb{1})\text{Sel}_\sigma(W_{RTS} \otimes \mathbb{1})|0\rangle\langle 0| \otimes \rho(W_{RTS}^\dagger \otimes \mathbb{1})\text{Sel}_\sigma^\dagger(V_{RTS}^\dagger \otimes \mathbb{1})\right) = \frac{1}{L_A!}\sum_{\sigma \in S_{L_A}}\mathcal{U}_\sigma^{(1)}(H_A, t) \circ \rho, \tag{187}$$

which matches the first-order randomized Trotter-Suzuki formula as given in [10]. It is a straightforward process to extend this mapping to higher-order randomized formulas.

The final case that fits within this framework is the multi-product formula. The primary difference between this case and the probabilistic Trotter formula is that a quantum superposition of the Trotter formulas is used [26, 12, 27]. Specifically, this method finds a set of N coefficients, c_σ, such that

$$\|\sum_{\sigma=1}^N c_\sigma(\prod_{q=1}^{L_A}\exp(-i[H_A]_q t/k_\sigma))^{k_\sigma} - e^{-iH_A t}\| \in O(t^{2N+1}) \tag{188}$$

To implement multiproduct formulas within our formalism, we define

$$W_{MPF}|0\rangle := \sum_{\sigma=1}^N \sqrt{|c_\sigma|}/\sqrt{\sum_j |c_\sigma|}|\sigma\rangle,$$

with $V_{MPF}^\dagger = W_{MPF}$. The coefficients c are chosen such that they are the solutions to the Vandermonde system of equations for some sequence of positive integers k_j [26]

$$\begin{bmatrix} 1 & k_1^{-1} & k_1^{-2} & \cdots & k_1^{-N+1} \\ 1 & k_2^{-1} & k_2^{-2} & \cdots & k_2^{-N+1} \\ \vdots & & & \ddots & \vdots \\ 1 & k_N^{-1} & k_N^{-2} & \cdots & k_N^{-N+1} \end{bmatrix}^T \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{189}$$

We take $\text{Sel}_\sigma|\sigma\rangle|\psi\rangle = |\sigma\rangle U_{\text{TS}}(H_A, t)|\psi\rangle = |\sigma\rangle(\prod_{q=1}^{L_A}\exp(-i[H_A]_q/k_\sigma t))^{k_\sigma}|\psi\rangle$. This encapsulates the select operation needed in [15, 12]. The LCU Lemma [26, 9] then implies that this channel will map,

$$(\langle 0|\otimes\mathbb{1})\Lambda(|0\rangle\langle 0|\otimes\rho)(|0\rangle\otimes\mathbb{1}) = \kappa^{-2}\sum_{\sigma'=1}^N\sum_{\sigma=1}^N c_\sigma(\prod_{q=1}^{L_A}\exp(-i[H_A]_q t/k_\sigma))^{k_\sigma}\rho c_{\sigma'}(\prod_{q=1}^{L_A}\exp(-i[H_A]_q t/k_\sigma))^{-k_\sigma}, \tag{190}$$

where κ^{-1} is a constant needed to block encode the formula within a larger unitary matrix. Thus multiproduct formulas can also be considered as special cases of a Composite channel.

Finally, following the exact same arguments used above since the registers used to block encode each of these individual channels are taken to be disjoint, any composition of the channels applied in this manner forms a Composite channel as we defined it above. This proves our claim. □

Note that above we do not discuss the complications of using oblivious amplitude amplification to prevent the success probability of multiproduct formulas shrinking exponentially with the number of segments [28]. This generalization is straightforward though and simply requires wrapping the evolution within a larger unitary transformation that resembles Grover's search. Further, it is also clear that the exact same ideas used for multi-product formulas could be used to implement a linear combinations of unitaries approach such as [9] or [11]. For simplicity, we choose our language above to focus on the case of composite Trotter-like channels but this approach also falls within our framework. [29] Qubitization is somewhat of an awkward fit within this approach [13], as it involves a sequence of rotations as opposed to simple linear expressions, and so it remains the one major simulation method that falls outside of the generalization of the Composite channel that we present above. We hope that this provides a compact framework for future analysis of composite simulations with more involved simulation techniques.

7 Discussion

This work has presented a new framework to analyze and design quantum simulation algorithms that have a compositional character. In particular, we examine mixing Trotter-Suzuki methods and QDrift and find evidence that the whole is greater than the sum of its parts. The central new concept behind this work is the idea of term partitioning, both in a deterministic as well as a randomized setting, which allows us to selectively apply a simulation algorithms to the parts of the Hamiltonian that benefit from high accuracy while relegating less sensitive parts of the Hamiltonian to a low-accuracy approximation.

We show for low order formulas that in cases where there are a small number of large terms that are (mostly) mutually commuting and a large number of smaller terms in the Hamiltonian, then a deterministic partitioning exists such that the lowest order Trotter formula combined with QDrift provides an asymptotic advantage over either model. For higher-order Trotter formulas we provide a set of asymptotic regimes for parameters of a predetermined partition, namely the number of terms in the Trotter partition and the sum of the spectral norms for the QDrift partition, that yield asymptotic improvements over Trotter or QDrift individually. Moreover we provide a method for producing partitions probabilistically that is easy to compute and exactly matches the operator exponential cost for second-order Trotter formulas and QDrift in the relevant limits.

To showcase how each of these results can be used to analyze a Hamiltonian we specifically examined the case where a given H has exponentially decaying spectral norms ($h_i = 2^{-i}$) and a small commutator structure ($\alpha_{comm}(H; 2k) \in \Theta(\log_2 L/L)$). We show that by using the probabilistic scheme provided that we can produce partitions that satisfy the asymptotic requirements in Theorem 2.2 with probability approaching 1 as $L \to \infty$. We note that although this example Hamiltonian is contrived for the purposes of demonstrating our techniques there are many Hamiltonians, such as electronic structure Hamiltonians, that are dominated by a handful of large terms with many more small terms. Commutator structures are highly dependent on specific Hamiltonians but our techniques give an asymptotic scaling that leads to *provable* advantage. This gives strong evidence that optimized partitions that take advantage of information about the Hamiltonian could lead to significant empirical improvements.

Looking forward, while this framework shows a way that we can think about combining several disparate Hamiltonian simulation methods, this contribution is by no means the end of this line of inquiry. Firstly, this work was inspired to no small extent by the coalescing scheme presented in [19] wherein a factor of 10 improvement was found numerically by decreasing the frequency with which low importance Hamiltonian terms are applied when looking at chemical simulations. Our framework provides a much more general way of thinking about such partitioning and further numerical work would be useful to understand the impact that these ideas will have on systems of interest, such as chemical compounds or lattice gauge theories.

More broad approaches within our framework can also be considered specifically, multiple orders of Trotter formulas could be considered and further randomization over the ordering of terms can also be performed easily within the randomization part of our algorithm. Incorporating further gradations in the accuracy of the underlying formula could lead to further practical improvements and pave the way for more highly optimized compilers. In systems where commutator or spectral norm information about subsets of the Hamiltonian is known these techniques could bring significant circuit depth reductions for simulations on quantum computers.

Acknowledgements

References

[1] James D Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. "Simulation of electronic structure hamiltonians using quantum computers". Molecular Physics **109**, 735–750 (2011).

url: https://doi.org/10.1080/00268976.2011.552441.

[2] Stephen P Jordan, Keith SM Lee, and John Preskill. "Quantum algorithms for quantum field theories". Science **336**, 1130–1133 (2012). url: https://doi.org/10.1126/science.1217069.

[3] Markus Reiher, Nathan Wiebe, Krysta M Svore, Dave Wecker, and Matthias Troyer. "Elucidating reaction mechanisms on quantum computers". Proceedings of the national academy of sciences **114**, 7555–7560 (2017). url: https://doi.org/10.1073/pnas.1619152114.

[4] Ryan Babbush, Dominic W. Berry, and Hartmut Neven. "Quantum simulation of the sachdev-ye-kitaev model by asymmetric qubitization". Phys. Rev. A **99**, 040301 (2019).

[5] Yuan Su, Dominic W. Berry, Nathan Wiebe, Nicholas Rubin, and Ryan Babbush. "Fault-tolerant quantum simulations of chemistry in first quantization". PRX Quantum **2**, 040332 (2021).

[6] Thomas E. O'Brien, Michael Streif, Nicholas C. Rubin, Raffaele Santagati, Yuan Su, William J. Huggins, Joshua J. Goings, Nikolaj Moll, Elica Kyoseva, Matthias Degroote, Christofer S. Tautermann, Joonho Lee, Dominic W. Berry, Nathan Wiebe, and Ryan Babbush. "Efficient quantum computation of molecular forces and other energy gradients". Phys. Rev. Res. **4**, 043210 (2022).

[7] Dorit Aharonov and Amnon Ta-Shma. "Adiabatic quantum state generation and statistical zero knowledge". In Proceedings of the thirty-fifth annual ACM symposium on Theory of computing. Pages 20–29. (2003). url: https://doi.org/10.1145/780542.780546.

[8] Dominic W Berry, Graeme Ahokas, Richard Cleve, and Barry C Sanders. "Efficient quantum algorithms for simulating sparse hamiltonians". Communications in Mathematical Physics **270**, 359–371 (2007). url: https://doi.org/10.1007/s00220-006-0150-x.

[9] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. "Simulating hamiltonian dynamics with a truncated taylor series". Phys. Rev. Lett. **114**, 090502 (2015).

[10] Andrew M. Childs, Aaron Ostrander, and Yuan Su. "Faster quantum simulation by randomization". Quantum **3**, 182 (2019).

[11] Guang Hao Low and Isaac L. Chuang. "Hamiltonian Simulation by Qubitization". Quantum **3**, 163 (2019).

[12] Guang Hao Low, Vadym Kliuchnikov, and Nathan Wiebe. "Well-conditioned multiproduct hamiltonian simulation" (2019). url: https://doi.org/10.48550/arXiv.1907.11679.

[13] Guang Hao Low and Nathan Wiebe. "Hamiltonian simulation in the interaction picture" (2019). arXiv:1805.00675.

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

[15] Nathan Wiebe, Dominic Berry, Peter Høyer, and Barry C Sanders. "Higher order decompositions of ordered operator exponentials". Journal of Physics A: Mathematical and Theoretical **43**, 065203 (2010).

[16] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. "Theory of trotter error with commutator scaling". Phys. Rev. X **11**, 011020 (2021).

[17] Dominic W. Berry, Andrew M. Childs, Yuan Su, Xin Wang, and Nathan Wiebe. "Time-dependent Hamiltonian simulation with L^1-norm scaling". Quantum **4**, 254 (2020).

[18] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. "Gate-count estimates for performing quantum chemistry on small quantum computers". Physical Review A**90** (2014).

[19] David Poulin, Matthew B Hastings, Dave Wecker, Nathan Wiebe, Andrew C Doherty, and Matthias Troyer. "The trotter step size required for accurate quantum simulation of quantum chemistry" (2014). url: https://doi.org/10.48550/arXiv.1406.4920.

[20] Ian D Kivlichan, Christopher E Granade, and Nathan Wiebe. "Phase estimation with randomized hamiltonians" (2019). arXiv:1907.10070.

[21] Abhishek Rajput, Alessandro Roggero, and Nathan Wiebe. "Hybridized Methods for Quantum Simulation in the Interaction Picture". Quantum **6**, 780 (2022).

[22] Yingkai Ouyang, David R. White, and Earl T. Campbell. "Compilation by stochastic hamiltonian sparsification". Quantum **4**, 235 (2020).

[23] Shi Jin and Xiantao Li. "A partially random trotter algorithm for quantum hamiltonian simulations" (2021). url: https://doi.org/10.48550/arXiv.2109.07987.

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

[25] Masuo Suzuki. "Fractal decomposition of exponential operators with applications to many-body theories and monte carlo simulations". Physics Letters A **146**, 319–323 (1990).

[26] Andrew M Childs and Nathan Wiebe. "Hamiltonian simulation using linear combinations of unitary operations" (2012). url: https://doi.org/10.26421/QIC12.11-12.

[27] Paul K Faehrmann, Mark Steudtner, Richard Kueng, Maria Kieferova, and Jens Eisert. "Randomizing multi-product formulas for improved hamiltonian simulation" (2021). url: https://ui.adsabs.harvard.edu/link_gateway/2022Quant...6..806F/doi:10.48550/arXiv.2101.07808.

[28] Dominic W. Berry, Andrew M. Childs, and Robin Kothari. "Hamiltonian simulation with nearly optimal dependence on all parameters". In 2015 IEEE 56th Annual Symposium on Foundations of Computer Science. Pages 792–809. (2015).

[29] Chi-Fang Chen, Hsin-Yuan Huang, Richard Kueng, and Joel A. Tropp. "Concentration for random product formulas". PRX Quantum**2** (2021).

A   Moment Bounds for Higher-Order Formulas

We now return to proving the moment bounds from Section 5.3.1.

Lemma 11. *Let L_A denote the number of terms simulated with a higher-order Trotter formula in a Composite channel with the assumptions as defined in section 3. Using a probabilistic partitioning scheme as defined in 2.1, then the second moment of L_A obeys*

$$\mathbb{E}(L_A^2) \leq |\mathcal{S}|^2 - |\mathcal{S}| \sum_{i \in \mathcal{S}} \frac{\chi}{h_i},$$

where χ is given in 98 and $\mathcal{S}$ is the set of p_i such that $1 - p_i < 1$ as given in 99.

Proof. Given that the simplest definition of our probabilities is for $1 - p_i$ we will try to work with expressions for the B channel as much as possible. It is easy to convert between the two as

$$\mathbb{E}\left[L_A^2\right] = \mathbb{E}\left[(L - L_B)^2\right] = L^2 - 2L\mathbb{E}\left[L_B\right] + \mathbb{E}\left[L_B^2\right]. \tag{191}$$

The expectation value of L_B then follows from plugging in the definitions

$$\mathbb{E}\left[L_B\right] = \sum_i \mathbb{E}\left[I_i^B\right] = \sum_i 1 - p_i = \chi \sum_{i \in \mathcal{S}} \frac{1}{h_i} + |\mathcal{S}^C|. \tag{192}$$

Now we find a relatively simple upper bound for $\mathbb{E}\left[L_B^2\right]$ if use the two facts that I_i^B and I_j^B

are independent for $i \neq j$ and that $\left(I_i^B\right)^2 = I_i^B$

$$\mathbb{E}\left[L_B^2\right] = \mathbb{E}\left[\left(\sum_i I_j^B\right)^2\right] \tag{193}$$

$$= \sum_i \mathbb{E}\left[I_i^B\right] + \sum_{i \neq j} \mathbb{E}\left[I_i^B\right]\mathbb{E}\left[I_j^B\right] \tag{194}$$

$$= \sum_i (1 - p_i) + \left(\sum_i (1 - p_i)\right)^2 - \sum_i (1 - p_i)^2 \tag{195}$$

$$= \chi \sum_{i \in \mathcal{S}} \frac{1}{h_i} + |\mathcal{S}^C| + \left(\chi \sum_{i \in \mathcal{S}} \frac{1}{h_i} + |\mathcal{S}^C|\right)^2 - \sum_{i \in \mathcal{S}} \frac{\chi^2}{h_i^2} - |\mathcal{S}^C|. \tag{196}$$

Combining equations (192) and (196) we get the following expression for an upper bound on L_A^2

$$\mathbb{E}\left[L_A^2\right] = L^2 - 2L(\chi \sum_{i \in \mathcal{S}} \frac{1}{h_i} + |\mathcal{S}^C|) + \chi \sum_{i \in \mathcal{S}} \frac{1}{h_i} + \left(\chi \sum_{i \in \mathcal{S}} \frac{1}{h_i} + |\mathcal{S}^C|\right)^2 - \sum_{i \in \mathcal{S}} \frac{\chi^2}{h_i^2} \tag{197}$$

$$= L^2 - 2L|\mathcal{S}^C| + |\mathcal{S}^C|^2 + \left(-2L + 1 + 2|\mathcal{S}^C|\right)\sum_{i \in \mathcal{S}} \frac{\chi}{h_i} + \left(\sum_{i \in \mathcal{S}} \frac{\chi}{h_i}\right)^2 - \sum_{i \in \mathcal{S}} \frac{\chi^2}{h_i^2} \tag{198}$$

$$= \left(L - |\mathcal{S}^C|\right)^2 + (1 - 2|\mathcal{S}|)\sum_{i \in \mathcal{S}} \frac{\chi}{h_i} + \sum_{i \neq j \in \mathcal{S}^2} \frac{\chi}{h_i}\frac{\chi}{h_j} \tag{199}$$

$$\leq |\mathcal{S}|^2 + (1 - 2|\mathcal{S}|)\sum_{i \in \mathcal{S}} \frac{\chi}{h_i} + (|\mathcal{S}| - 1)\sum_{i \in \mathcal{S}} \frac{\chi}{h_i} \tag{200}$$

$$= |\mathcal{S}|^2 - |\mathcal{S}| \sum_{i \in \mathcal{S}} \frac{\chi}{h_i}. \tag{201}$$

Note that we only used the following inequality $\frac{\chi}{h_i} = 1 - p_i \leq 1$ for $i \in \mathcal{S}$. $\qquad\square$

Lemma 12. *Let Q denote the contribution to the error of a Composite channel, where Q is defined in 68, with the standard assumptions from Section 3 and a probabilistic partitioning scheme as outlined in 2.1. The following upper bounds hold on the first two moments for Q*

$$\mathbb{E}\left[Q(t)\right] \leq \frac{4t^2}{N_B}\left(\chi \lambda_{\mathcal{S}} + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2\right) \tag{131}$$

$$\mathbb{E}\left[Q(t)^2\right] \leq \frac{4t^2\lambda^2}{N_B}\mathbb{E}\left[Q(t)\right] \leq \frac{16t^4\lambda^2}{N_B^2}\left(\chi \lambda_{\mathcal{S}} + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2\right) \tag{132}$$

Proof. $\mathbb{E}\left[Q(t)\right]$ can be computed very similarly to L_A above, since $\mathbb{E}\left[Q(t)\right] \propto \mathbb{E}\left[\lambda_B^2\right]$. This second moment for λ_B mostly follows from the definitions but we first will get an easier expression involving the indicator variables

$$\mathbb{E}\left[\lambda_B^2\right] = \mathbb{E}\left[\left(\sum_i h_i I_i^B\right)^2\right] \tag{202}$$

$$= \sum_i h_i^2 \mathbb{E}\left[I_i^B\right] + \sum_{i \neq j} h_i h_j \mathbb{E}\left[I_i^B I_j^B\right] \tag{203}$$

$$= \sum_i h_i^2 \mathbb{E}\left[I_i^B\right] + \sum_{i \neq j} h_i \mathbb{E}\left[I_i^B\right] h_j \mathbb{E}\left[I_j^B\right] \tag{204}$$

$$= \sum_i h_i^2 \mathbb{E}\left[I_i^B\right] + \left(\sum_i h_i \mathbb{E}\left[I_i^B\right]\right)^2 - \sum_j h_j^2 \mathbb{E}\left[I_j^B\right]^2, \tag{205}$$

where we used the fact that I_i^B is independent from I_j^B for all $i \neq j$. Now we can utilize our probability distributions as $\mathbb{E}\left[I_i^B\right] = 1 - p_i$, which yields

$$\mathbb{E}\left[\lambda_B^2\right] = \sum_{i \in \mathcal{S}} h_i \chi + \sum_{i \in \mathcal{S}^C} h_i^2 + (\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C})^2 - \sum_{j \in \mathcal{S}} h_j^2(1 - p_j)^2 - \sum_{j \in \mathcal{S}^C} h_j^2(1 - p_j)^2 \tag{206}$$

$$= \lambda_{\mathcal{S}}\chi + \sum_{i \in \mathcal{S}^C} h_i^2 + \chi^2|\mathcal{S}|^2 + 2\chi|\mathcal{S}|\lambda_{\mathcal{S}^C} + \lambda_{\mathcal{S}^C}^2 - \chi^2|\mathcal{S}| - \sum_{j \in \mathcal{S}^C} h_j^2 \tag{207}$$

$$\leq \chi^2|\mathcal{S}|^2 + \chi\left(\lambda_{\mathcal{S}} + 2\lambda_{\mathcal{S}^C}|\mathcal{S}|\right) + \lambda_{\mathcal{S}^C}^2 \tag{208}$$

$$= \chi\lambda_{\mathcal{S}} + \left(\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C}\right)^2. \tag{209}$$

The only inequality comes from dropping the correction term $\chi^2|\mathcal{S}|$, which is subleading to $\chi^2|\mathcal{S}|^2$.

Our expression for a upper bound on $\mathbb{E}\left[Q(t)\right]$ is then proportional to the above expression (209)

$$\mathbb{E}\left[Q(t)\right] \leq \frac{2t^2}{N_B}\left(\chi\lambda_{\mathcal{S}} + \left(\chi|\mathcal{S}| + \lambda_{\mathcal{S}^C}\right)^2\right). \tag{210}$$

To compute $\mathbb{E}\left[Q(t)^2\right]$ we can reduce it to our prior results. Since $\mathbb{E}\left[Q(t)^2\right] = \frac{4t^4}{N_B^2}\mathbb{E}\left[\lambda_B^4\right]$, we will use the following upper bound

$$\mathbb{E}\left[\lambda_B^4\right] = \left(\sum_i h_i I_i^B\right)^4 \leq \left(\sum_i h_i 1\right)^2 \left(\sum_i h_i I_i^B\right)^2 = \lambda^2 \mathbb{E}\left[\lambda_B^2\right]. \tag{211}$$

This means we can re-use the above computation as

$$\mathbb{E}\left[Q(t)^2\right] \leq \frac{2t^2\lambda^2}{N_B}\mathbb{E}\left[Q(t)\right]. \tag{212}$$

$\square$

Lemma 13. *Let $P(t)$ denote the product formula error scaling as defined in 67 for a Composite channel with the standard assumptions as defined in 3. Using a probabilistic partitioning scheme as defined in 2.1 then the following upper bounds hold for the expected value of $P(t)$*

$$\mathbb{E}\left[P(t)\right] \leq \frac{(2\Upsilon)^{2+2k}}{2k+1} t^{2k+1}\lambda^{2k}\left(\lambda_{\mathcal{S}} - \chi|\mathcal{S}|\right). \tag{133}$$

Proof. To simplify this we will use intermediate steps from the calculation of $P_{\max}(t)$ that bound the α_{comm} factors, namely equations (87), (90) which are repeated below

$$\alpha_{comm}(A, 2k) \leq 2^{2k}\lambda_A^{2k+1} \tag{213}$$

$$\alpha_{comm}(\{A, B\}, 2k) \leq 2^{2k}\sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l} \tag{214}$$

$$P(t) = \frac{2^2(t\Upsilon)^{2k+1}}{2k+1}\left(\Upsilon\alpha_{comm}(A, 2k) + \alpha_{comm}(\{A, B\}, 2k)\right) \tag{215}$$

We will use these to compute a useful upper bound on $\mathbb{E}\left[P(t)\right]$. Since our random variables are more easily described for the I_i^B variables, we will convert all powers of λ_A into functions of L and λ_B as well as upper bound both by λ_A and λ_B by λ. This results in an upper bound on $\mathbb{E}\left[P(t)\right]$ as

$$\mathbb{E}\left[P(t)\right] \leq \frac{2^{2+2k}(t\Upsilon)^{2k+1}}{2k+1}\mathbb{E}\left[\Upsilon\lambda_A^{2k+1} + \sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l}\right]. \tag{216}$$

We will simplify the expectation value using the facts that $\lambda_A, \lambda_B \leq \lambda$ and that in $\sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l}$ each term has at least one factor of $\lambda_A \lambda_B$. This results in the following simplifications

$$\mathbb{E}\left[\Upsilon \lambda_A^{2k+1} + \sum_{l=1}^{2k} \lambda_A^l \lambda_B^{2k+1-l}\right] \leq \lambda^{2k-1} \mathbb{E}\left[\Upsilon \lambda_A^2 + (2k)\lambda_A \lambda_B\right] \tag{217}$$

$$\leq \Upsilon \lambda^{2k-1} \mathbb{E}\left[\lambda_A^2 + \lambda_A \lambda_B\right] \tag{218}$$

$$= \Upsilon \lambda^{2k-1} \mathbb{E}\left[\lambda_A(\lambda_A + \lambda_B)\right] \tag{219}$$

$$= \Upsilon \lambda^{2k} \left(\lambda - \mathbb{E}\left[\lambda_B\right]\right) \tag{220}$$

$$= \Upsilon \lambda^{2k} \left(\lambda_\mathcal{S} - \chi|\mathcal{S}|\right), \tag{221}$$

where we used an exact expression for $\mathbb{E}\left[\lambda_B\right]$ that is a straightforward computation in addition to the fact that $2k \leq \Upsilon$ for $k \geq 1$. Combining the above expressions (221) and (216) our final expression for an upper bound on $\mathbb{E}\left[P(t)\right]$ is

$$\mathbb{E}\left[P(t)\right] \leq \frac{(2\Upsilon)^{2+2k}}{2k+1} t^{2k+1} \lambda^{2k} \left(\lambda_\mathcal{S} - \chi|\mathcal{S}|\right) \tag{222}$$

$\square$