

Compilation by stochastic Hamiltonian sparsification

Yingkai Ouyang¹, David R. White¹, and Earl T. Campbell^{1,2}

¹Department of Physics and Astronomy, University of Sheffield, Sheffield, UK

²Riverlane, Cambridge, UK

Simulation of quantum chemistry is expected to be a principal application of quantum computing. In quantum simulation, a complicated Hamiltonian describing the dynamics of a quantum system is decomposed into its constituent terms, where the effect of each term during time-evolution is individually computed. For many physical systems, the Hamiltonian has a large number of terms, constraining the scalability of established simulation methods. To address this limitation we introduce a new scheme that approximates the actual Hamiltonian with a sparser Hamiltonian containing fewer terms. By stochastically sparsifying weaker Hamiltonian terms, we benefit from a quadratic suppression of errors relative to deterministic approaches. Relying on optimality conditions from convex optimisation theory, we derive an appropriate probability distribution for the weaker Hamiltonian terms, and compare its error bounds with other probability ansatzes for some electronic structure Hamiltonians. Tuning the sparsity of our approximate Hamiltonians allows our scheme to interpolate between two recent random compilers: qDRIFT and randomized first order Trotter. Our scheme is thus an algorithm that combines the strengths of randomised Trotterisation with the efficiency of qDRIFT, and for intermediate gate budgets, outperforms both of these prior methods.

1 Introduction

Chemistry simulation is expected to be a principal application of quantum computing, revealing the properties of chemical bonds and interactions by simulating classically intractable systems, with applications in pharmaceuticals, material science, and industrial chemical manufacture [1, 29]. For example, efficient simulation of the chemical cluster FeMoCo [5, 9, 35] may allow more efficient nitrogen fixation, improving the manufacture of fertilizers. The difficulty of directly solving the Schrödinger equation of chemically interesting problems renders even moderately com-

plex systems classically intractable. Using quantum mechanics to simulate quantum systems [25] will provide unprecedented detail in the solution to chemical problems, enabling us to better understand the dynamics of highly entangled systems, predicting their properties and chemical reactions [1, 17]. For example, simulation can be used in phase estimation to extract the eigenspectrum of a Hamiltonian [22, 33, 44], and by sufficiently understanding its energy spectra we can accurately predict chemical reaction rates [23].

Given a Hamiltonian H expressed as the sum of multi-qubit Pauli matrices, solution of the Schrödinger equation requires the calculation of its exponentiation. Quantum simulation techniques are distinguished by the way they map the chemical Hamiltonian to an effective Hamiltonian on qubits [3, 4, 20, 31, 36], and subsequent mapping of the exponentiation of this effective Hamiltonian to a computation. One well-established method is to apply the Trotter-Suzuki formula [38–40] to reduce this larger exponentiation to a product of Pauli exponentials. Each Pauli exponential can then be interpreted as a single quantum gate. To date, many variants of Trotter-Suzuki methods have been studied in the context of quantum simulation of chemical systems [2, 21, 37].

Trotterisation is an attractive approach to quantum simulation because of its simplicity. However, a significant limitation of Trotterisation is that the number of quantum gates required scales linearly with the number of terms in the Hamiltonian, which may grow very large [42, 44]. Campbell [12] observed the problematic scaling of the Trotter-Suzuki decomposition and introduces qDRIFT, a stochastic approach that samples terms from the Hamiltonian, trading accuracy for computational cost; importantly, qDRIFT scales independently of the number of terms. We build on Campbell’s approach, introducing an algorithm that combines the strengths of standard Trotterisation with the efficiency of qDRIFT.

Our approach approximates the target Hamiltonian via sparsification, yielding a Hamiltonian with fewer terms; the Trotterised computation then requires far fewer gates per Trotter step. Although sparsification introduces a new approximation error, the reduction in gate count means we can apply more Trotter steps within a fixed budget. Our key insight is that sparsification can be performed stochastically instead of deterministically, and that this leads to improved per-

Yingkai Ouyang: y.ouyang@sheffield.ac.uk
 David R. White: d.r.white@sheffield.ac.uk
 Earl T. Campbell: earltcampbell@gmail.com

formance. By defining our approximate Hamiltonian to be a random variable with an expectation value equal to the actual Hamiltonian, we benefit from a quadratic suppression of errors relative to deterministic methods; this behaviour is also seen in other stochastic compilers [11, 12, 15, 16]. We refer to our combination of first order Trotterisation and stochastically sparsified Hamiltonians as SPARSTO.

In many systems, including electronic structure Hamiltonians, we empirically observe power-law distributions of term strengths, which is promising for sparsification since weaker terms are clear candidates for truncation. In a stochastic compiler, it is natural to relate the probability of a term being truncated from the Hamiltonian with the magnitude of its strength. One of our main technical results is a rigorous upper bound on the error of SPARSTO for an arbitrary probability distribution where the terms are sampled independently.

To obtain the best possible upper-bound we need to select the best probability distribution. In our analysis, we place the most important terms, with the largest strengths, inside an *active* set so that they always appear in the sparsified Hamiltonian. Random sparsification is instead applied only to a tail of weaker Hamiltonian terms that we label the *inactive* set. We rely on optimality conditions in convex optimisation theory in order to derive a probability distribution over the inactive set, which we call the “linear ansatz”. We numerically optimise and analyse the performance of our error bounds for some electronic structure Hamiltonians. For low gate budgets SPARSTO behaves similarly to qDRIFT, and for larger gate budgets it exactly reproduces randomized first order Trotter; as such it interpolates between these approaches. However, for intermediate gate budgets, which coincide with parameter regimes of practical interest, our new sparsification method outperforms both methods by around an order of magnitude. We emphasize numerical optimisation is always performed at the level of upper bounds and not by considering empirical performance of small simulatable systems. Though SPARSTO uses first order Trotter, sparsification of Hamiltonians could also be naturally combined with higher order Trotter schemes to yield higher order randomized compilers.

When comparing with other Trotter methods we can just count the number of gates of the form e^{-isH_j} that are unitaries generated by easily accessible Hamiltonians H_j . However, to fairly compare against post-Trotter methods [7, 9, 13, 26, 27], we would also need to assess the resource cost of extra ancilla and select and prepare gadgets that are not built using e^{-isH_j} . This makes direct comparison with post-Trotter methods an involved task and sensitive to the cost model. However, post-Trotter methods have better asymptotic scaling and for problems of interest they have often been found to have a con-

siderable advantage over Trotterisation methods, and it is likely that this advantage persists over SPARSTO. However, the value of this work is two-fold. Firstly, we improve the best known Trotter methods. Secondly, we have demonstrated the value of randomly sparsifying Hamiltonians as a technique that might later be incorporated into post-Trotter methods. Indeed, Berry’s perspective article [6] also highlighted that potential application of randomization to post-Trotter protocols is a promising future research direction.

2 Trotterisation

Consider a time-independent Hamiltonian that admits a decomposition $H = \sum_{j=1}^L h_j P_j$. While P_j can often be considered to be multi-qubit Pauli matrices, we make no such assumption, and for us P_j are matrices with singular value at most 1. Without loss of generality, the corresponding coefficients h_j can then be positive. Solving the Schrödinger equation $|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle$ allows us to model the continuous evolution of the state $|\psi(t)\rangle$ over time. Over a short time period s , the first order Trotter-Suzuki decomposition approximates the exponential e^{-iHt} with a product of exponentials given by [38]

$$\prod_{j=1}^L e^{-isH_j} = e^{-isH} + \mathcal{O}(s^2), \quad (1)$$

where $H_j = h_j P_j$. We call this the “vanilla Trotterisation” scheme, as there are many variants to this approach [14, 15, 39, 40], including ones that propose to deterministically coalesce Hamiltonian terms [34, 43]. We assume that each e^{-isH_j} , which we call a gate, can be efficiently implemented on a target quantum computer. To simulate e^{-itH} , we approximate e^{-isH} repeatedly r times, such that $t = rs$. The number of gates G required by a simulation is thus the principal measure of its computational cost. Since each Trotterisation of e^{-isH} involves L gates, vanilla Trotterisation requires $G = rL$ gates, and can become potentially computationally expensive when L is large. In particular, it is known that an effective Hamiltonian for the electronic structure problem for a system with N modes typically has $L = \mathcal{O}(N^4)$ terms [1].

The vanilla Trotterisation scheme approximates e^{-itH} with a simulation error that is at most $\epsilon_{\text{van}} \lesssim \lambda^2 t^2 / 2r = L\lambda^2 t^2 / 2G$ [38], where $\lambda = \|\mathbf{h}\|_1$, $\mathbf{h} = (h_1, \dots, h_L)$. In contrast to Trotterisation, the qDRIFT method introduced by Campbell has a computational cost that is independent of L ; its gate count is $\mathcal{O}(\lambda^2 t^2 / \epsilon)$. qDRIFT simulates an ideal unitary process by a Markovian evolution, sampling a sequence of Pauli gates from a predetermined distribution; each exponentiation in the resulting circuit is given the same weight τ such that the distribution alone determines the outcome of the calculation. The probability

p_j of choosing a given $e^{i\tau H_j}$ as the next gate in a computation is weighted by the corresponding interaction strength: $p_j = h_j/\lambda$, ensuring that the stochastic process of repeated sampling drifts stochastically towards the target unitary. The number of gates is set at a fixed computational budget G representing the number of primitive gates, and gives approximation error $\epsilon \lesssim 4\lambda^2 t^2/G$ ¹.

Whilst Trotter-Suzuki decompositions have worse computational complexity than qDRIFT in the number of terms, their computational cost scales better with respect to t and ϵ . To exploit this trade-off, we introduce a new approach SPARSTO, which interpolates between qDRIFT and the higher-order Trotter-Suzuki decompositions whilst also building and improving on the analysis of randomised simulation in Ref. [15].

Like qDRIFT, SPARSTO approximates a unitary evolution with a probabilistic ensemble of unitary evolutions instead of direct Trotterisation — although higher-order Trotterisation can subsequently be applied. This stochastic scheme is in the spirit of related work [8, 14, 15] and its merits lie in the ability to use mixtures of unitary operators to approximate a unitary operator [11, 16]; intuitively stochastic methods avoid systematic noise.

3 SPARSTO Analysis

SPARSTO uses a random Hamiltonian \hat{H} and crucial to our analysis is that the expectation value is equal to the system Hamiltonian $\mathbb{E}(\hat{H}) = H$. To reduce gate counts, we would like \hat{H} to have far fewer terms than H and to be a good approximation, or at least to do this with high probability. Rather than considering arbitrary probability distributions we consider the term-wise independent sampling where \hat{H} contains the term $h_j P_j/p_j$ with probability p_j and with probability $1 - p_j$ this term is dropped. This ensures $\mathbb{E}(\hat{H}) = H$ and that the expected number of terms is $\mu = \sum_{j=1}^L p_j$.

Next, we review Lindblad’s formalism of unitary maps [24], where such maps are generated by exponentiating Liouville operators. We let $\mathcal{L}_j = h_j \mathcal{P}_j$ so that \mathcal{P}_j is a Liouville operator that maps ρ to $-i(P_j \rho - \rho P_j)$. Clearly then, $\mathcal{L}_j(\rho) = -i(H_j \rho - \rho H_j)$. For any positive number s , Liouville operators generate unitary evolutions in the sense that $e^{s\mathcal{L}_j}(\rho) = e^{-iH_j s} \rho e^{iH_j s}$. The ideal evolution operator can be written in terms of the Liouville operator $\mathcal{L} = \sum_{j=1}^L \mathcal{L}_j$, because $e^{s\mathcal{L}}(\rho) = e^{-iH s} \rho e^{iH s}$. Using a vanilla Trotterisation analogous to that given in (1), a first order approximation of $e^{s\mathcal{L}}$ is given

¹In [12], the simulation error used is the diamond distance, which differs from the diamond norm of the difference of channels by a factor of 2. This explains why the bound in [12] is approximately at most $2\lambda^2 t^2/G$ instead of $4\lambda^2 t^2/G$

by $\mathcal{T}_{s,\rightarrow} = \prod_{j=1}^L e^{s\mathcal{L}_j}$. Given no *a priori* reason to simulate \mathcal{L}_j in any particular order, it is natural to consider permutations of this operator sequence. A second order approximation of the Taylor expansion can be made by mixing the above Trotterisation $\mathcal{T}_{s,\rightarrow}$ with $\mathcal{T}_{s,\leftarrow} = \prod_{j=L}^1 e^{s\mathcal{L}_j}$, where the arrows denote the ordering over the term index j . The uniformly mixed operation $\frac{1}{2}(\mathcal{T}_{s,\rightarrow} + \mathcal{T}_{s,\leftarrow})$, which is a randomised first order Trotterisation scheme that we denote as R1OTROTT, approximates $e^{s\mathcal{L}}$ with error that is third order in sL [15, Theorem 1]. More generally, Trotterisation can be further improved by completely randomising the order in which the gates are performed [15, 42]. To approximate $e^{t\mathcal{L}}$ for a fixed time t , we approximate $e^{s\mathcal{L}}$ for a total of r times, where $s = t/r$. By taking the number of repeats r to be large, the simulation time s can become small, and (1) holds to a good approximation.

In SPARSTO, we stochastically sparsify the Hamiltonian using some term-wise independent probability distribution and then apply one step of randomized first order Trotter. For each of the r Trotter steps a fresh stochastic Hamiltonian is sampled. The random Hamiltonian \hat{H} induces a random Liouville operator $\hat{\mathcal{L}}$ in the natural way $\hat{\mathcal{L}}(\rho) = i(\hat{H}\rho - \rho\hat{H})$. Similarly, we have random terms $\hat{\mathcal{L}}_j$ such that

$$\hat{\mathcal{L}}_j = \begin{cases} \mathcal{L}_j/p_j & \text{with probability } p_j \\ 0 & \text{with probability } 1 - p_j \end{cases} \quad (2)$$

Here, $\hat{\mathcal{L}}_j$ approximates the ideal Liouville operator \mathcal{L}_j in the sense that $\mathbb{E}(\hat{\mathcal{L}}_j) = \mathcal{L}_j$. Given a sampled \hat{H} or $\hat{\mathcal{L}}$, we also randomize the order of the gates in each Trotter step and so introduce the randomized operators of forward and reverse Trotter steps

$$\hat{\mathcal{T}}_{s,\rightarrow} = \prod_{j=1}^L e^{s\hat{\mathcal{L}}_j} \quad (3)$$

$$\hat{\mathcal{T}}_{s,\leftarrow} = \prod_{j=L}^1 e^{s\hat{\mathcal{L}}_j}. \quad (4)$$

A single step of SPARSTO is then described by

$$\hat{\mathcal{E}}_s = \frac{1}{2} \left(\hat{\mathcal{T}}_{s,\rightarrow} + \hat{\mathcal{T}}_{s,\leftarrow} \right) \quad (5)$$

which has μ gates on average. To approximate $e^{t\mathcal{L}}$, SPARSTO simulates $\hat{\mathcal{E}}_s$ independently and sequentially r times. By fixing the expected number of gates G of SPARSTO to be constant, we require in the first $\lfloor G/\mu \rfloor$ repeats to have $s = \mu t/G$ and in the final repeat to have $s = t - \lfloor G/\mu \rfloor$. The total number of repeats is then $r = \lceil G/\mu \rceil$. We do not consider completely randomising the gate orders in $\hat{\mathcal{E}}_s$ because it renders our subsequent analysis overly complicated.

We quantify the maximum error of SPARSTO using the diamond norm [19], which when evaluated on the difference between quantum channels, quantifies

their distinguishability. For us, we quantify the distinguishability between the average of r repeats of $\hat{\mathcal{E}}_s$ and the ideal channel $e^{t\mathcal{L}}$ with the error

$$\|\mathbb{E}(\hat{\mathcal{E}}_s^r) - e^{t\mathcal{L}}\|_\diamond, \quad (6)$$

where $\|\cdot\|_\diamond$ denotes the diamond norm.

Our main result is an analytic upper bound on $\|\hat{\mathcal{E}}_s - e^{s\mathcal{L}}\|_\diamond$ that we denote as ϵ , which is expressed in terms of the 1-norm $\|\cdot\|_1$.

Theorem 1. *Using SPARSTO with vector of probabilities $\mathbf{p} = (p_1, \dots, p_L)$, vector of Hamiltonian coefficients $\mathbf{h} = (h_1, \dots, h_L)$, $L \geq 3$, and expected number of gates G , the error of simulating $e^{t\mathcal{L}}$ with SPARSTO is at most ϵ where*

$$\epsilon = \frac{2t^2\mu}{G} \|\mathbf{u}\|_1 + \frac{4t^3\mu^2}{3G^2} K + \mathcal{O}\left(\frac{t^4\mu^3}{G^3}\right),$$

with $K = (\|\mathbf{v}\|_1 + \lambda\|\mathbf{w}\|_1 + 4\lambda^3/3)$, $\lambda = \|\mathbf{h}\|_1$ and $\mu = \|\mathbf{p}\|_1$. Moreover, \mathbf{u} , \mathbf{v} and \mathbf{w} are vectors given by

$$\begin{aligned} \mathbf{u} &= \left(\left(\frac{1}{p_1} - 1\right) h_1^2, \dots, \left(\frac{1}{p_L} - 1\right) h_L^2 \right), \\ \mathbf{v} &= \left(\left(\frac{1}{p_1^2} - 1\right) h_1^3, \dots, \left(\frac{1}{p_L^2} - 1\right) h_L^3 \right), \\ \mathbf{w} &= \left(\left(\frac{3}{p_1} - 1\right) h_1^2, \dots, \left(\frac{3}{p_L} - 1\right) h_L^2 \right). \end{aligned}$$

A tighter bound with full details on the higher order terms in ϵ is supplied in Theorem 5 of Appendix A.

To bound the diamond distance between $\mathbb{E}(\hat{\mathcal{E}}_s^r)$ and $e^{t\mathcal{L}}$, we bound the diamond distance between $\mathbb{E}(\hat{\mathcal{E}}_s)$ and $e^{s\mathcal{L}}$. Using the triangle inequality on a telescoping sum, the unit diamond norm of all channels, and the independence of each random unitary $\hat{\mathcal{E}}_s$, we get the bound

$$\|\mathbb{E}(\hat{\mathcal{E}}_s)^r - e^{t\mathcal{L}}\|_\diamond \leq r \|\mathbb{E}(\hat{\mathcal{E}}_s) - e^{s\mathcal{L}}\|_\diamond. \quad (7)$$

To obtain an upper bound on $\|\mathbb{E}(\hat{\mathcal{E}}_s) - e^{s\mathcal{L}}\|_\diamond$, we perform a series expansion of the operators $\hat{\mathcal{E}}_s$ and $e^{s\mathcal{L}}$ with respect to the parameter s to get $\hat{\mathcal{E}}_s = \sum_{j \geq 0} \hat{\mathcal{A}}_j s^j$ and $e^{s\mathcal{L}} = \sum_{j \geq 0} \mathcal{B}_j s^j$. Using this notation, we can see that $\hat{\mathcal{A}}_0$ and \mathcal{B}_0 are both trivially the identity operator $\mathbb{1}$, and $\mathbb{E}(\hat{\mathcal{A}}_1)$ and \mathcal{B}_1 are both equal to the Liouvillian \mathcal{L} . To obtain the $\mathcal{O}(t^2\mu/G)$ and $\mathcal{O}(t^3\mu^2/G^2)$ terms in ϵ , we evaluate upper bounds on $\|\mathbb{E}(\hat{\mathcal{A}}_2) - \mathcal{B}_2\|_\diamond$ and $\|\mathbb{E}(\hat{\mathcal{A}}_3) - \mathcal{B}_3\|_\diamond$ respectively. To do this, we rewrite $\hat{\mathcal{A}}_2$ as sums over products of $\hat{\mathcal{L}}_j^{k_j}$, where each sum comprises of terms of the form $\hat{\mathcal{L}}_j^2$, $\hat{\mathcal{L}}_j \hat{\mathcal{L}}_k$, where j and k are distinct indices. Having j and k distinct allows us to find that $\mathbb{E}(\hat{\mathcal{L}}_j^2) = \mathcal{L}_j^2/p_j$ and $\mathbb{E}(\hat{\mathcal{L}}_j \hat{\mathcal{L}}_k) = \mathcal{L}_j \mathcal{L}_k$. Using a similar strategy for rewriting $\hat{\mathcal{A}}_3$, we can evaluate its expectation explicitly. Writing \mathcal{B}_2 and \mathcal{B}_3 in a similar form then allows us to compute the leading order terms in ϵ . We supply the full details of this argument in Appendix A.

We upper bound the difference between tails of $\hat{\mathcal{E}}_s$ and $e^{s\mathcal{L}}$, which are $\mathcal{O}(t^4\mu^3/G^3)$ terms, by essentially using the fundamental theorem of calculus to bound the tail of a power series from its derivatives. From this, we evaluate upper bounds on the diamond norm of $\sum_{j \geq 4} s^j (\mathbb{E}(\hat{\mathcal{A}}_j) - \mathcal{B}_j)$, and call this our tail bound. To apply the fundamental theorem of calculus, we first take the fourth derivatives of $\hat{\mathcal{E}}_{s\theta}$ and $e^{s\theta\mathcal{L}}$ with respect to θ , evaluate upper bounds on the norm of their difference over the unit interval for θ . Second, we integrate this upper bound over an appropriate region, which gives a rescaling factor of $1/4!$. Also, by obtaining polynomials in the diamond norms of \mathcal{L}_j and subsequently using the inequality $\|\mathcal{L}_j\|_\diamond \leq 2h_j$, along with the triangle inequality on the diamond norm of the difference between the ideal channel and the approximate channel, we can obtain a closed form expression for the tail bounds which we show explicitly in Theorem 5 of Appendix A.

It is important to point out that the upper bound on the simulation error in Theorem 1 depends very much on the choice of the probabilities p_1, \dots, p_L . Each p_j signifies the probability that the Hamiltonian term H_j contributes to the Trotterisation at each iteration. The smaller the value of $\mu = p_1 + \dots + p_L$, the sparser our Hamiltonian simulation is. Intuitively, different choices on the values of the probabilities p_j in Theorem 1 affect the overall simulation error of $e^{t\mathcal{L}}$. When all probabilities are equal to one, SPARSTO becomes identical to R10TROT. The simulation error, can thereby be obtained as the following corollary of Theorem 1.

Corollary 2. *When $p_1 = \dots = p_L = 1$, the simulation error is at most*

$$\epsilon = \frac{8t^3L^2}{3G^2} \left(\lambda \sum_{j=1}^L h_j^2 + \frac{2\lambda^3}{3} \right) + \mathcal{O}\left(\frac{t^4L^3}{G^3}\right).$$

While the bound that we have in Corollary 2 is tighter than [15, Theorem 1], a careful analysis of the third order terms in [15, Theorem 1] yields the same expression as that given in Corollary 2.

One might also observe that when all the probabilities in Theorem 1 are set to $p_j = 1$, we have $\|\mathbf{u}\|_1 = \|\mathbf{v}\|_1 = 0$ and $\|\mathbf{w}\|_1$ is minimized, which implies that ϵ/r which is roughly the simulation error per time segment s , is in fact minimized. This leads one to wonder what advantage might be gained by setting the probabilities to be otherwise. The solution to this conundrum lies in the penalty we pay in making such a choice. In this scenario, each $\hat{\mathcal{E}}_s$ comprises of $\mu = L$ gates, and the overall error ϵ for simulating $e^{t\mathcal{L}}$ need not be optimized since $r s^j \sim t^j (\mu/G)^{j-1}$, which appears as coefficients in Theorem 1, is in fact maximized when $\mu = L$. The resultant algorithm simulates $e^{t\mathcal{L}}$, with an expected gate count of G when $s = \mu t/G$ for all but the last repeat and $r = \lceil G/\mu \rceil$.

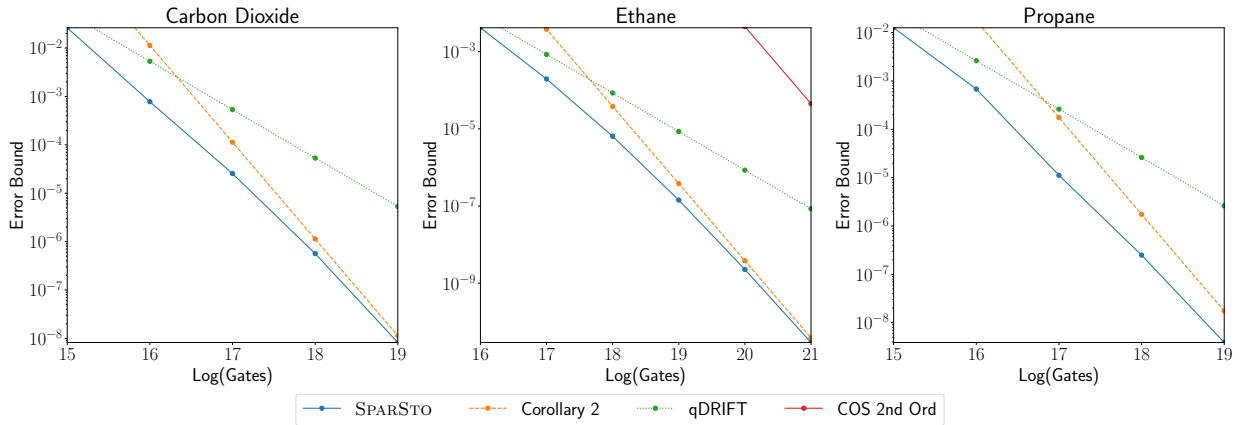


Figure 1: Error Bounds: Rigorous upper bounds on the simulation errors of various molecules in the STO-3G basis set with $L \geq 100000$ are compared with rigorous bounds for Trotterisation and qDRIFT. Here $t = 6000$. In an intermediate regime for expected the number of gates, SPARSTO requires fewer gates than both R1OTROTT and qDRIFT for a fixed simulation error. For propane and carbon dioxide, the second order Trotter error bounds (COS 2nd Order [15, Theorem 2]) are too large to be seen on the plots.

One can imagine SPARSTO to be analogous to another qDRIFT where $\mu = 1$ so that the expected number of gates per time segment s is equal to one. The tradeoff in this scenario is that $\|\mathbf{u}\|_1$ and $\|\mathbf{v}\|_1$ are potentially very large because the probabilities become very small. The key advantage of using Theorem 1 allows us to understand how ϵ interpolates between having all the probabilities to be either 1 or 0. In what follows, we consider one family of probability distributions that we use together with Theorem 1. For this example, we set $p_j = 1$ whenever h_j is above a set threshold. Otherwise, $p_j < 1$. We denote the active set A as the set of indices j for which $p_j = 1$, and the inactive set \bar{A} to be the set of indices for which $p_j < 1$. We choose the values of p_j according the following ansatz.

Definition 3 (Linear ansatz). *For every $j \in A$, we set $p_j = 1$. For every $j \in \bar{A}$, we set $p_j = ch_j$. We correspondingly have $\mu = |A| + c \sum_{j \in \bar{A}} h_j$.*

Clearly, c has to be sufficiently small so that we indeed have $p_j < 1$ for all indices j in the inactive set. By minimizing ϵ with respect to all possible values of $|A|$ and μ using our linear ansatz, we can determine which probabilities p_j to use. These probabilities can be inputted into SPARSTO, which we describe in the pseudocode Algorithm 1.

We numerically study the performance of SPARSTO using models of molecules drawn from the OpenFermion library [30], including carbon dioxide, ethane, and propane in the STO-3G basis set, and depict these results in Fig. 1. We evaluate the error bound for SPARSTO given by Theorem 5 in Appendix A. We compare the performance of SPARSTO with the Trotter bounds from [15] (Theorem 2 in their paper, setting $k = 1$), and by setting all probabilities $p_j = 1$ we also plot Corollary 2. Only the second

Algorithm 1 SPARSTO ($t, G, p_1, \dots, p_L, h_1, \dots, h_L$)

```

1:  $\mu \leftarrow p_1 + \dots + p_L$ 
2: for all rep = 1 to  $r$  do
3:   dir  $\leftarrow$  fwd or bwd with probability 1/2
4:   if rep  $\leq \lfloor G/\mu \rfloor$  then
5:      $s \leftarrow \mu t/G$ 
6:   else
7:      $s \leftarrow t - \lfloor G/\mu \rfloor \mu t/G$ 
8:   if dir = fwd then
9:     for all  $j = 1$  to  $L$  do
10:      Choose  $x$  uniformly at random from  $[0, 1]$ .
11:      if  $x \geq p_j$  then implement  $\exp(s(h_j/p_j)\mathcal{P}_j)$ 
12:   else ▷ dir = bwd
13:     for all  $j = L$  to 1 do
14:      Choose  $x$  uniformly at random from  $[0, 1]$ .
15:      if  $x \geq p_j$  then implement  $\exp(s(h_j/p_j)\mathcal{P}_j)$ 

```

order bounds from Childs *et al.* [15, Theorem 2] are visible, in the upper-right of the second plot.

We perform a limited brute force numerical optimisation over all feasible values of μ and $|A|$ for our ansatzes; we examine $|A|/L$ over the interval $[0, 1]$ with a step size of 0.1, and consider the same values for $\mu' = (\mu - |A|)/(L - |A|)$ along with 1×10^{-5} , 1×10^{-4} , and 1×10^{-3} ; we consider all pairwise combinations of these settings. Intuitively we expect that as the gate budget G increases, we ought to interpolate between the qDRIFT regime [12] and the R1OTROTT regime, and the size of the active set $|A|$ ought to go from 0 to L . We observe from our numerical study that this indeed is the case. In general, the optimal active set size increases with G , and the optimal value for μ' was usually small, and never more than 0.3.

The linear ansatz outperforms the uniform ansatz. This is expected, as the uniform ansatz is naïve and the linear ansatz can be obtained as the optimal solution of the convex program which minimizes the leading order term in the total error for constant μ (see Appendix B). In each of the molecules, the number of Hamiltonian terms is over a hundred thousand, which

is very large. When $t = 6000$ and for a range of desired error values, there is a considerable advantage in using SPARSTO over both R1OTROTT and qDRIFT. We observe similar results across other values of t and smaller molecules.

4 Discussion

While vanilla Trotterisation can simulate any Hamiltonian with sufficiently many gates, the number of these gates can become very large. This leads to the need to reduce the gate count of quantum simulation while keeping the size of simulation error fixed. Here we present a new approach to chemistry simulation on a quantum machine, using the stochastic sparsification of a target Hamiltonian to derive a hybrid approach between canonical Trotterisation and qDRIFT. Our analysis provides an upper error bound for the scheme, and optimisation over the probabilities used in sparsification allows for reductions in the simulation error over parameter regimes of interest.

It would be instructive to consider how the ideas in our hybrid approach might extend to other variants of quantum simulation schemes, such as that of the so-called “quantum signal processing” (QSP) [26] techniques, linear combinations of unitaries [7], the use of quantum walks [9, 13], qubitisation [9, 27] and post-processing techniques [18]. There has also been recent interest in the quantum simulation of time dependent Hamiltonians [8, 28], and applications of quantum simulation in phase estimation [12, 22], which may also prove amenable to stochastic sparsification. Given that random techniques can prove advantageous when applied to hybrid quantum-classical algorithms for numerical optimisation [41], our techniques might also offer some speedups in this area. Furthermore, there might exist certain families of Hamiltonians where the advantage of using our techniques over deterministic Trotterisation can be understood analytically, and we leave this as a subject for future work.

Acknowledgements.- This work was supported by the EPSRC (grant no. EP/M024261/1), and has also received research funding from Huawei. We like to thank Yuan Su for a careful reading and comments on an earlier version of this manuscript.

References

- [1] A. Aspuru-Guzik. Simulated quantum computation of molecular energies. *Science*, 309(5741): 1704–1707, September 2005. DOI: [10.1126/science.1113479](https://doi.org/10.1126/science.1113479).
- [2] Ryan Babbush, Jarrod McClean, Dave Wecker, Alán Aspuru-Guzik, and Nathan Wiebe. Chemical basis of Trotter-Suzuki errors in quantum chemistry simulation. *Phys. Rev. A*, 91:022311, Feb 2015. DOI: [10.1103/PhysRevA.91.022311](https://doi.org/10.1103/PhysRevA.91.022311).
- [3] Ryan Babbush, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Alexandru Paler, Austin Fowler, and Hartmut Neven. Encoding electronic spectra in quantum circuits with linear T complexity. *Phys. Rev. X*, 8:041015, Oct 2018. DOI: [10.1103/PhysRevX.8.041015](https://doi.org/10.1103/PhysRevX.8.041015).
- [4] 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, Mar 2018. DOI: [10.1103/PhysRevX.8.011044](https://doi.org/10.1103/PhysRevX.8.011044).
- [5] H. Beinert. Iron-sulfur clusters: Nature’s modular, multipurpose structures. *Science*, 277(5326):653–659, August 1997. DOI: [10.1126/science.277.5326.653](https://doi.org/10.1126/science.277.5326.653).
- [6] Dominic W Berry. A random approach to quantum simulation. *Physics*, 12:91, 2019. DOI: [10.1103/physics.12.91](https://doi.org/10.1103/physics.12.91).
- [7] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Exponential improvement in precision for simulating sparse Hamiltonians. *Forum of Mathematics, Sigma*, 5, 2017. DOI: [10.1017/fms.2017.2](https://doi.org/10.1017/fms.2017.2).
- [8] Dominic W Berry, Andrew M Childs, Yuan Su, Xin Wang, and Nathan Wiebe. Time-dependent Hamiltonian simulation with L^1 -norm scaling. *arXiv preprint arXiv:1906.07115*, 2019.
- [9] Dominic W. Berry, Craig Gidney, Mario Motta, Jarrod R. McClean, and Ryan Babbush. Qubitization of Arbitrary Basis Quantum Chemistry Leveraging Sparsity and Low Rank Factorization. *Quantum*, 3:208, December 2019. ISSN 2521-327X. DOI: [10.22331/q-2019-12-02-208](https://doi.org/10.22331/q-2019-12-02-208).
- [10] Sergey Bravyi and Jeongwan Haah. Quantum Self-Correction in the 3D Cubic Code Model. *Phys. Rev. Lett.*, 111(20):200501, November 2013. DOI: [10.1103/PhysRevLett.111.200501](https://doi.org/10.1103/PhysRevLett.111.200501).
- [11] Earl Campbell. Shorter gate sequences for quantum computing by mixing unitaries. *Phys. Rev. A*, 95:042306, Apr 2017. DOI: [10.1103/PhysRevA.95.042306](https://doi.org/10.1103/PhysRevA.95.042306).
- [12] Earl Campbell. Random compiler for fast Hamiltonian simulation. *Phys. Rev. Lett.*, 123:070503, Aug 2019. DOI: [10.1103/PhysRevLett.123.070503](https://doi.org/10.1103/PhysRevLett.123.070503).
- [13] Andrew M. Childs and Dominic W. Berry. Black-box Hamiltonian simulation and unitary implementation. *Quantum Information and Computation*, 12(1-2), 2012. DOI: [10.26421/qic12.1-2](https://doi.org/10.26421/qic12.1-2).
- [14] Andrew M. Childs, Dmitri Maslov, Yunseong Nam, Neil J. Ross, and Yuan Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, 115(38):9456–9461, 2018. ISSN 0027-8424. DOI: [10.1073/pnas.1801723115](https://doi.org/10.1073/pnas.1801723115).
- [15] Andrew M. Childs, Aaron Ostrander, and Yuan

- Su. Faster quantum simulation by randomization. *Quantum*, 3:182, September 2019. DOI: 10.22331/q-2019-09-02-182.
- [16] Matthew B. Hastings. Turning gate synthesis errors into incoherent errors. *Quantum Info. Comput.*, 17(5-6):488–494, March 2017. ISSN 1533-7146. DOI: 10.26421/QIC17.5-6.
- [17] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X*, 8:031022, Jul 2018. DOI: 10.1103/PhysRevX.8.031022.
- [18] William J. Huggins, Jarrod McClean, Nicholas Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *arXiv:1907.13117*, 2019.
- [19] Alexei Yu Kitaev, Alexander Shen, Mikhail N Vyalyi, and Mikhail N Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Soc., 2002. DOI: 10.1090/gsm/047.
- [20] Ian D. Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush. Quantum simulation of electronic structure with linear depth and connectivity. *Phys. Rev. Lett.*, 120:110501, Mar 2018. DOI: 10.1103/PhysRevLett.120.110501.
- [21] Ian D. Kivlichan, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Wei Sun, Zhang Jiang, Nicholas Rubin, Austin Fowler, Alán Aspuru-Guzik, Hartmut Neven, and Ryan Babbush. Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via Trotterization. *arXiv:1902.10673*, 2019.
- [22] Ian D. Kivlichan, Christopher E. Granade, and Nathan Wiebe. Phase estimation with randomized Hamiltonians. *arXiv:1907.10070*, 2019.
- [23] Zhaokai Li, Xiaomei Liu, Hefeng Wang, Sahel Ashhab, Jiangyu Cui, Hongwei Chen, Xinhua Peng, and Jiangfeng Du. Quantum simulation of resonant transitions for solving the eigenproblem of an effective water Hamiltonian. *Phys. Rev. Lett.*, 122:090504, Mar 2019. DOI: 10.1103/PhysRevLett.122.090504.
- [24] G Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, 1976. ISSN 0010-3616. DOI: 10.1007/BF01608499.
- [25] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, August 1996. DOI: 10.1126/science.273.5278.1073.
- [26] Guang Hao Low and Isaac L. Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Physical Review Letters*, 118(1), January 2017. DOI: 10.1103/physrevlett.118.010501.
- [27] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by qubitization. *Quantum*, 3:163, July 2019. DOI: 10.22331/q-2019-07-12-163.
- [28] Guang Hao Low and Nathan Wiebe. Hamiltonian simulation in the interaction picture. *arXiv:1805.00675*, 2018.
- [29] Sam McArdle, Suguru Endo, Alan Aspuru-Guzik, Simon Benjamin, and Xiao Yuan. Quantum computational chemistry. *arXiv preprint arXiv:1808.10402*, 2018.
- [30] Jarrod R McClean, Ian D Kivlichan, Kevin J Sung, Damian S Steiger, Yudong Cao, Chengyu Dai, E Schuyler Fried, Craig Gidney, Brendan Gimby, Pranav Gokhale, et al. OpenFermion: the electronic structure package for quantum computers. *arXiv preprint arXiv:1710.07629*, 2017.
- [31] Jarrod R. McClean, Fabian M. Faulstich, Qinyi Zhu, Bryan O’Gorman, Yiheng Qiu, Steven R. White, Ryan Babbush, and Lin Lin. Discontinuous Galerkin discretization for quantum simulation of chemistry. *arXiv:1909.00028*, 2019.
- [32] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006. DOI: 10.1007/b98874.
- [33] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. Scalable quantum simulation of molecular energies. *Phys. Rev. X*, 6:031007, Jul 2016. DOI: 10.1103/PhysRevX.6.031007.
- [34] David Poulin, M. B. Hastings, D. Wecker, N. Wiebe, Andrew C. Doherty, and M. Troyer. The Trotter step size required for accurate quantum simulation of quantum chemistry. *Quantum Information & Computation*, 15(5-6):0361–0384, 2015. DOI: 10.26421/qic15.5-6.
- [35] 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(29):7555–7560, July 2017. DOI: 10.1073/pnas.1619152114.
- [36] Kanav Setia and James D. Whitfield. Bravyi-Kitaev superfast simulation of electronic structure on a quantum computer. *The Journal of Chemical Physics*, 148(16):164104, April 2018. DOI: 10.1063/1.5019371.
- [37] Rolando D. Somma. A Trotter-Suzuki ap-

- proximation for lie groups with applications to Hamiltonian simulation. *Journal of Mathematical Physics*, 57(6):062202, June 2016. DOI: [10.1063/1.4952761](https://doi.org/10.1063/1.4952761).
- [38] Masuo Suzuki. Generalized Trotter’s formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. *Comm. Math. Phys.*, 51(2):183–190, 1976. DOI: [10.1007/bf01609348](https://doi.org/10.1007/bf01609348).
- [39] Masuo Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations. *Physics Letters A*, 146(6):319–323, June 1990. DOI: [10.1016/0375-9601\(90\)90962-n](https://doi.org/10.1016/0375-9601(90)90962-n).
- [40] Masuo Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *Journal of Mathematical Physics*, 32(2):400–407, February 1991. DOI: [10.1063/1.529425](https://doi.org/10.1063/1.529425).
- [41] Ryan Sweke, Frederik Wilde, Johannes Meyer, Maria Schuld, Paul K Fährmann, Barthélémy Meynard-Piganeau, and Jens Eisert. Stochastic gradient descent for hybrid quantum-classical optimization. *arXiv preprint arXiv:1910.01155*, 2019.
- [42] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. Gate-count estimates for performing quantum chemistry on small quantum computers. *Phys. Rev. A*, 90:022305, Aug 2014. DOI: [10.1103/PhysRevA.90.022305](https://doi.org/10.1103/PhysRevA.90.022305).
- [43] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. Gate-count estimates for performing quantum chemistry on small quantum computers. *Phys. Rev. A*, 90:022305, Aug 2014. DOI: [10.1103/PhysRevA.90.022305](https://doi.org/10.1103/PhysRevA.90.022305).
- [44] James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics*, 109(5):735–750, March 2011. DOI: [10.1080/00268976.2011.552441](https://doi.org/10.1080/00268976.2011.552441).

A Upper bounds on the simulation error

In this section, we show that Theorem 1 is a corollary of Theorem 5, which we state in Section A.2. Before we can state Theorem 1, we define relevant notation in Section A.1. After that, we evaluate the leading order terms and tail terms of Theorem 5 in Section A.3 and Section A.4 respectively.

A.1 Sum over distinct indices

Given real vectors $\mathbf{a} = (a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n)$ and $\mathbf{c} = (c_1, \dots, c_n)$, we define the sums over distinct indices to be

$$\begin{aligned}\mathcal{S}(\mathbf{a}) &= \sum_{j=1}^n a_j \leq \|\mathbf{a}\|_1, \\ \mathcal{S}(\mathbf{a}, \mathbf{b}) &= \sum_{\substack{1 \leq j, k \leq n \\ j, k \text{ distinct}}} a_j b_k \leq \|\mathbf{a}\|_1 \|\mathbf{b}\|_1, \\ \mathcal{S}(\mathbf{a}, \mathbf{b}, \mathbf{c}) &= \sum_{\substack{1 \leq j, k, l \leq n \\ j, k, l \text{ distinct}}} a_j b_k c_l \leq \|\mathbf{a}\|_1 \|\mathbf{b}\|_1 \|\mathbf{c}\|_1.\end{aligned}\tag{8}$$

To perform fast computation of the above sums, we can use the following lemma which vectorises summations with distinct indices.

Lemma 4. *Let n be a positive integer. Let $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$ be real column vectors. Then*

$$\mathcal{S}(\mathbf{a}, \mathbf{b}) = A_1 B_1 - C_1 \tag{9}$$

$$\mathcal{S}(\mathbf{a}, \mathbf{b}, \mathbf{b}) = A_1 (B_1^2 - B_2) - 2C_1 B_1 + 2C_2, \tag{10}$$

$$\mathcal{S}(\mathbf{a}, \mathbf{a}, \mathbf{a}) = A_1^3 - 3A_2 A_1 + 2A_3, \tag{11}$$

where

$$A_j = \sum_{u=1}^n a_u^j, \tag{12}$$

$$B_j = \sum_{u=1}^n b_u^j, \tag{13}$$

$$C_j = \sum_{u=1}^n a_u b_u^j. \tag{14}$$

Lemma 4 can be proved iteratively by careful consideration of summation indices.

Proof of Lemma 4. The result (9) is straightforward to show. To show (10), note that we can use (9) to write

$$\begin{aligned}& \sum_{\substack{1 \leq u, v, w \leq n \\ u, v, w \text{ distinct}}} a_u b_v b_w \\ &= \sum_{u=1}^n a_u \sum_{\substack{1 \leq v, w \leq n \\ v, w \text{ distinct}}} b_v b_w - \sum_{\substack{1 \leq u, w \leq n \\ u, w \text{ distinct}}} a_u b_u b_w - \sum_{\substack{1 \leq u, v \leq n \\ u, v \text{ distinct}}} a_u b_u b_v \\ &= A_1 (B_1^2 - B_2) - 2 \sum_{u=1}^n a_u b_u B_1 + 2 \sum_{u=1}^n a_u b_u^2.\end{aligned}$$

We can specialize this to sum of distinct combinations of $a_u a_v a_w$ to get

$$\sum_{\substack{1 \leq u, v, w \leq n \\ u, v, w \text{ distinct}}} a_u a_v a_w = A_1 (A_1^2 - A_2) - 2A_2 A_1 + 2A_3,$$

which yields (11). □

A.2 Complete simulation error bound

The complete upper bound that we prove here is given by the following.

Theorem 5. *Using SPARSTO with vector of probabilities $\mathbf{p} = (p_1, \dots, p_L)$, vector of Hamiltonian coefficients $\mathbf{h} = (h_1, \dots, h_L)$, $L \geq 3$, and expected number of gates G where $G/(p_1 + \dots + p_L)$ is an integer, the error of simulating $e^{t\mathcal{L}}$ is at most ϵ where $\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_{3,1} + \epsilon_{3,2}$ and*

$$\begin{aligned}\epsilon_1 &= \frac{2t^2\mu}{G}\mathcal{S}(\mathbf{u}), \\ \epsilon_2 &= \frac{4t^3\mu^2}{3G^2}(\mathcal{S}(\mathbf{v}) + \mathcal{S}(\mathbf{w}, \mathbf{h})) + \frac{16t^3\mu^2}{9G^2}\mathcal{S}(\mathbf{h}, \mathbf{h}, \mathbf{h}), \\ \epsilon_{3,1} &= \frac{2t^4\mu^3\lambda^4}{3G^3}, \\ \epsilon_{3,2} &= \frac{2t^4\mu^3}{3G^3}(p_1 \dots p_L)\mathcal{S}(\mathbf{q})^4,\end{aligned}$$

with $\mu = \sum_{j=1}^L p_j$. Moreover, $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and \mathbf{q} are vectors given by

$$\begin{aligned}\mathbf{u} &= \left(\left(\frac{1}{p_1} - 1 \right) h_1^2, \dots, \left(\frac{1}{p_L} - 1 \right) h_L^2 \right), \\ \mathbf{v} &= \left(\left(\frac{1}{p_1^2} - 1 \right) h_1^3, \dots, \left(\frac{1}{p_L^2} - 1 \right) h_L^3 \right), \\ \mathbf{w} &= \left(\left(\frac{3}{p_1} - 1 \right) h_1^2, \dots, \left(\frac{3}{p_L} - 1 \right) h_L^2 \right), \\ \mathbf{q} &= \left(\frac{h_1}{p_1}, \dots, \frac{h_L}{p_L} \right).\end{aligned}$$

We use \mathcal{S} as defined in Section A.1. We will see that by considering explicitly the commutation structure of the matrices P_j , we can obtain a tighter bound on ϵ_2 in Theorem 5 by substituting $\mathcal{S}(\mathbf{h}, \mathbf{h}, \mathbf{h})$ for D_5 in (45).

Before we proceed to prove Theorem 5, we prove that Theorem 1 is a straightforward consequence of Theorem 5. Note that

Proof of Theorem 1. By overcounting (8), it is easy to see that

$$\mathcal{S}(\mathbf{w}, \mathbf{h}) \leq \|\mathbf{w}\|_1 \|\mathbf{h}\|_1 = \lambda \|\mathbf{w}\|_1,$$

and

$$\mathcal{S}(\mathbf{h}, \mathbf{h}, \mathbf{h}) \leq \|\mathbf{h}\|_1 \|\mathbf{h}\|_1 \|\mathbf{h}\|_1 = \lambda^3.$$

Moreover, since \mathbf{u} and \mathbf{v} are non-negative vectors, we have $\mathcal{S}(\mathbf{u}) = \|\mathbf{u}\|_1$ and $\mathcal{S}(\mathbf{v}) = \|\mathbf{v}\|_1$. Furthermore, we have $\epsilon_{3,j} = \mathcal{O}(t^4\mu^3/G^3)$. This completes the proof. \square

The proof of Theorem 5 then arises from the evaluation of (1) the leading order terms ϵ_1 and ϵ_2 , and (2) the higher order terms $\epsilon_{3,1}$ and $\epsilon_{3,2}$. This will proceed in the next two subsections. We emphasize that in what follows, because of the telescoping argument we mentioned in the main text, it suffices to only analyze $\|\hat{\mathcal{E}}_s - e^{s\mathcal{L}}\|_\diamond$, and the overall simulation error will just be r times of this diamond norm.

A.3 The leading order terms ϵ_1 and ϵ_2 in Theorem 5

Here, we show that the leading order terms in the simulation error are as given by ϵ_1 and ϵ_2 . First recall that we have the Taylor series expansions $\hat{\mathcal{E}}_s = \sum_{j \geq 0} \hat{\mathcal{A}}_j s^j$ and $e^{s\mathcal{L}} = \sum_{j \geq 0} \mathcal{B}_j s^j$.

Note that when $L \geq 3$, we have

$$\begin{aligned}
\hat{\mathcal{T}}_{s,\rightarrow} &= \prod_{j=1}^L \left(\mathbf{1} + s\hat{\mathcal{L}}_j + \frac{s^2}{2}\hat{\mathcal{L}}_j^2 + \frac{s^3}{6}\hat{\mathcal{L}}_j^3 + \dots \right) \\
&= \mathbf{1} + s \sum_{j=1}^L \hat{\mathcal{L}}_j + \frac{s^2}{2} \sum_{j=1}^L \hat{\mathcal{L}}_j^2 + \frac{2s^2}{2} \sum_{1 \leq j < k \leq L} \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k \\
&\quad + \frac{s^3}{6} \sum_{j=1}^L \hat{\mathcal{L}}_j^3 + s \frac{s^2}{2} \sum_{1 \leq j < k \leq L} \left(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k + \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2 \right) \\
&\quad + s^3 \sum_{1 \leq j < k < l \leq L} \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k \hat{\mathcal{L}}_l + \dots
\end{aligned} \tag{15}$$

Similarly,

$$\begin{aligned}
\hat{\mathcal{T}}_{s,\leftarrow} &= \mathbf{1} + s \sum_{j=1}^L \hat{\mathcal{L}}_j + \frac{s^2}{2} \sum_{j=1}^L \hat{\mathcal{L}}_j^2 + \frac{2s^2}{2} \sum_{1 \leq k < j \leq L} \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k \\
&\quad + \frac{s^3}{6} \sum_{j=1}^L \hat{\mathcal{L}}_j^3 + s \frac{s^2}{2} \sum_{1 \leq j < k \leq L} \left(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k + \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2 \right) \\
&\quad + s^3 \sum_{1 \leq l < k < j \leq L} \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k \hat{\mathcal{L}}_l + \dots
\end{aligned} \tag{16}$$

Since $\hat{\mathcal{E}}_s = \frac{1}{2} (\hat{\mathcal{T}}_{s,\rightarrow} + \hat{\mathcal{T}}_{s,\leftarrow})$, (15) and (16) imply that

$$2! \hat{\mathcal{A}}_2 = \sum_{j=1}^L \hat{\mathcal{L}}_j^2 + \sum_{j \neq k} \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k, \tag{17}$$

$$3! \hat{\mathcal{A}}_3 = \sum_{j=1}^L \hat{\mathcal{L}}_j^3 + \frac{6}{4} \sum_{j \neq k} \left(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k + \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2 \right) + \frac{6}{2} \sum_{\substack{1 \leq j < k < l \leq L \\ 1 \leq l < k < j \leq L}} \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k \hat{\mathcal{L}}_l. \tag{18}$$

Moreover, we know that

$$\mathcal{L}^2 = \sum_{j=1}^L \mathcal{L}_j^2 + \sum_{j \neq k} \mathcal{L}_j \mathcal{L}_k, \tag{19}$$

$$\begin{aligned}
\mathcal{L}^3 &= \sum_{j=1}^L \mathcal{L}_j^3 + \sum_{j \neq k} \left(\mathcal{L}_j^2 \mathcal{L}_k + \mathcal{L}_j \mathcal{L}_k^2 + \mathcal{L}_j \mathcal{L}_k \mathcal{L}_j \right) \\
&\quad + \sum_{\substack{1 \leq j < k < l \leq L \\ 1 \leq l < k < j \leq L}} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l + \sum_{\substack{1 \leq k < j < l \leq L \\ 1 \leq k < l < j \leq L}} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l + \sum_{\substack{1 \leq j < l < k \leq L \\ 1 \leq l < j < k \leq L}} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l.
\end{aligned} \tag{20}$$

Clearly $\mathcal{L} - \mathbb{E}(\hat{\mathcal{L}}) = 0$. Next note that

$$\begin{aligned}
\mathcal{L}^2 - \mathbb{E}(2! \hat{\mathcal{A}}_2) &= \sum_{j=1}^L \left(\mathcal{L}_j^2 - \mathbb{E}(\hat{\mathcal{L}}_j^2) \right) + \sum_{j \neq k} \left(\mathcal{L}_j \mathcal{L}_k - \mathbb{E}(\hat{\mathcal{L}}_j \hat{\mathcal{L}}_k) \right) \\
&= \sum_{j=1}^L \left(\mathcal{L}_j^2 - \mathbb{E}(\hat{\mathcal{L}}_j^2) \right).
\end{aligned} \tag{21}$$

Now $\mathbb{E}(\hat{\mathcal{L}}_j^2) = p_j \frac{\mathcal{L}_j^2}{p_j^2}$, which implies that for $0 < p_j \leq 1$, we have

$$\mathcal{L}^2 - \mathbb{E}(2! \hat{\mathcal{A}}_2) = \sum_{j=1}^L \left(1 - \frac{1}{p_j} \right) \mathcal{L}_j^2. \tag{22}$$

Since $p_j \leq 1$, we have

$$\|\mathcal{L}^2 - \mathbb{E}(2!\hat{\mathcal{A}}_2)\|_\diamond \leq \sum_{j=1}^L \left(\frac{1}{p_j} - 1\right) (4h_j^2). \quad (23)$$

Since $\mathcal{B}_2 = \mathcal{L}^2/2!$, we get the upper bound

$$\begin{aligned} \|s^2\mathbb{E}(\hat{\mathcal{A}}_2) - s^2\mathcal{B}_2\|_\diamond &= \frac{s^2}{2!} 4 \sum_{j=1}^L \left(\frac{1}{p_j} - 1\right) h_j^2 \\ &= 2s^2 \sum_{j=1}^L \left(\frac{1}{p_j} - 1\right) h_j^2, \end{aligned} \quad (24)$$

where $s = t\mu/G$. Multiplying this by $r = G/\mu$ gives us ϵ_1 .

To evaluate ϵ_2 , we proceed to write

$$\mathcal{L}^3 - \mathbb{E}(3!\hat{\mathcal{A}}_3) = D_1 + D_2 + D_3 + D_4 + D_5, \quad (25)$$

where

$$D_1 = \sum_{j=1}^L \left(\mathcal{L}_j^3 - \mathbb{E}(\hat{\mathcal{L}}_j^3)\right), \quad (26)$$

$$D_2 = \sum_{j \neq k} \left(\mathcal{L}_j^2 \mathcal{L}_k - \mathbb{E}(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k)\right), \quad (27)$$

$$D_3 = \sum_{j \neq k} \left(\mathcal{L}_j \mathcal{L}_k^2 - \mathbb{E}(\hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2)\right), \quad (28)$$

$$D_4 = \sum_{j \neq k} \left(\mathcal{L}_j \mathcal{L}_k \mathcal{L}_j - \frac{1}{2} \mathbb{E}(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k + \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2)\right), \quad (29)$$

$$D_5 = \sum_{\substack{1 \leq k < j < l \leq L \\ 1 \leq k < l < j \leq L}} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l + \sum_{\substack{1 \leq j < l < k \leq L \\ 1 \leq l < j < k \leq L}} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l - 2 \sum_{\substack{1 \leq j < k < l \leq L \\ 1 \leq l < k < j \leq L}} \mathbb{E}(\hat{\mathcal{L}}_j \hat{\mathcal{L}}_k \hat{\mathcal{L}}_l). \quad (30)$$

We now proceed to simplify D_j for $j = 1, \dots, 5$. Note that $\mathbb{E}(\hat{\mathcal{L}}_j^3) = p_j \frac{\mathcal{L}_j^3}{p_j^3}$. This implies that

$$D_1 = \sum_{j=1}^L \left(1 - \frac{1}{p_j^2}\right) \mathcal{L}_j^3. \quad (31)$$

Next, multiplicativity of the expectation for independent random variables implies that

$$D_2 = \sum_{j \neq k} \left(1 - \frac{1}{p_j}\right) \mathcal{L}_j^2 \mathcal{L}_k, \quad (32)$$

$$D_3 = \sum_{j \neq k} \left(1 - \frac{1}{p_k}\right) \mathcal{L}_j \mathcal{L}_k^2. \quad (33)$$

Now we can write

$$D_4 = \frac{1}{2} \sum_{j \neq k} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_j + \frac{1}{2} \sum_{j \neq k} \mathcal{L}_k \mathcal{L}_j \mathcal{L}_k - \sum_{j \neq k} \left(\frac{1}{2} \mathbb{E}(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k + \hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2)\right). \quad (34)$$

Clearly, we have $\mathbb{E}(\hat{\mathcal{L}}_j^2 \hat{\mathcal{L}}_k) = \mathcal{L}_j^2 \mathcal{L}_k / p_j$ and $\mathbb{E}(\hat{\mathcal{L}}_j \hat{\mathcal{L}}_k^2) = \mathcal{L}_j \mathcal{L}_k^2 / p_k$. Next by swapping the roles of j and k in the summation, we get

$$\sum_{j \neq k} \mathcal{L}_j \mathcal{L}_k^2 / p_k = \sum_{j \neq k} \mathcal{L}_k \mathcal{L}_j^2 / p_j.$$

By pairing the first term with the third term and the second term with the fourth term in (34), this implies that

$$D_4 = \frac{1}{2} \sum_{j \neq k} \mathcal{L}_j (\mathcal{L}_k \mathcal{L}_j - \mathcal{L}_j \mathcal{L}_k / p_j) + \frac{1}{2} \sum_{j \neq k} (\mathcal{L}_j \mathcal{L}_k - \mathcal{L}_k \mathcal{L}_j / p_k) \mathcal{L}_k, \quad (35)$$

where we swap the roles of j and k in the second sum. From the above, we can see that

$$\|D_1\|_\diamond \leq \sum_{j=1}^L \left(\frac{1}{p_j^2} - 1 \right) 8h_j^3, \quad (36)$$

$$\|D_2\|_\diamond \leq \sum_{j \neq k} \left(\frac{1}{p_j} - 1 \right) 8h_j^2 h_k, \quad (37)$$

$$\|D_3\|_\diamond \leq \sum_{j \neq k} \left(\frac{1}{p_k} - 1 \right) 8h_j h_k^2, \quad (38)$$

$$\|D_4\|_\diamond \leq \sum_{j \neq k} \left(1 + \frac{1}{p_j} \right) 8h_j^2 h_k. \quad (39)$$

From this, we can obtain the first two terms in ϵ_2 . To see this, note that

$$\frac{1}{3!} \|D_1\|_\diamond \leq \frac{4}{3} \sum_{j=1}^L \left(\frac{1}{p_j^2} - 1 \right) h_j^3, \quad (40)$$

and

$$\frac{1}{3!} \sum_{i=2}^4 \|D_i\|_\diamond \leq \frac{4}{3} \sum_{j \neq k} \left(3 \frac{1}{p_j} - 1 \right) h_j^2 h_k. \quad (41)$$

Multiplying the right sides of (40) and (41) by r gives the first two terms in ϵ_2 .

We proceed to simplify D_5 . Note from (30) that

$$\begin{aligned} D_5 = & \sum_{1 \leq k < j < l \leq L} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l + \sum_{1 \leq k < l < j \leq L} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l + \sum_{1 \leq j < l < k \leq L} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l \\ & + \sum_{1 \leq l < j < k \leq L} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l - 2 \sum_{1 \leq j < k < l \leq L} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l - 2 \sum_{1 \leq l < k < j \leq L} \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l. \end{aligned} \quad (42)$$

Now by ordering all the indices in the same way we get

$$\begin{aligned} D_5 = & \sum_{1 \leq j < k < l \leq L} \mathcal{L}_k \mathcal{L}_j \mathcal{L}_l + \sum_{1 \leq j < k < l \leq L} \mathcal{L}_l \mathcal{L}_j \mathcal{L}_k + \sum_{1 \leq j < k < l \leq L} \mathcal{L}_j \mathcal{L}_l \mathcal{L}_k \\ & + \sum_{1 \leq j < k < l \leq L} \mathcal{L}_k \mathcal{L}_l \mathcal{L}_j - 2 \sum_{1 \leq j < k < l \leq L} (\mathcal{L}_j \mathcal{L}_k \mathcal{L}_l + \mathcal{L}_l \mathcal{L}_k \mathcal{L}_j). \end{aligned} \quad (43)$$

By pairing the first term with the fifth term, and the third term with the fifth term in (43), we get $\mathcal{L}_k \mathcal{L}_j \mathcal{L}_l - \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l = [\mathcal{L}_k, \mathcal{L}_j] \mathcal{L}_l$ and $\mathcal{L}_j \mathcal{L}_l \mathcal{L}_k - \mathcal{L}_j \mathcal{L}_k \mathcal{L}_l = \mathcal{L}_j [\mathcal{L}_l, \mathcal{L}_k]$. By pairing the second term with the sixth term, and the fourth term with the sixth term in (43), we get $\mathcal{L}_l \mathcal{L}_j \mathcal{L}_k - \mathcal{L}_l \mathcal{L}_k \mathcal{L}_j = \mathcal{L}_l [\mathcal{L}_j, \mathcal{L}_k]$ and $\mathcal{L}_k \mathcal{L}_l \mathcal{L}_j - \mathcal{L}_l \mathcal{L}_k \mathcal{L}_j = [\mathcal{L}_k, \mathcal{L}_l] \mathcal{L}_j$. We can thus rewrite (43) as

$$D_5 = \sum_{1 \leq j < k < l \leq L} ([\mathcal{L}_k, \mathcal{L}_j] \mathcal{L}_l + \mathcal{L}_j [\mathcal{L}_l, \mathcal{L}_k] + \mathcal{L}_l [\mathcal{L}_j, \mathcal{L}_k] + [\mathcal{L}_k, \mathcal{L}_l] \mathcal{L}_j) \quad (44)$$

Collecting the terms in the above summation in terms of commutators again, we get

$$D_5 = \sum_{1 \leq j < k < l \leq L} ([\mathcal{L}_l, [\mathcal{L}_j, \mathcal{L}_k]] + [[\mathcal{L}_k, \mathcal{L}_l], \mathcal{L}_j]). \quad (45)$$

A trivial upper bound on the diamond norm of this is

$$\|D_5\|_\diamond \leq 8 \frac{8}{6} \mathcal{S}(\mathbf{h}, \mathbf{h}, \mathbf{h}), \quad (46)$$

where the first factor of 8 arises from going from the diamond norm of the Liouvillean \mathcal{L}_j to the operator norm of H_j , and the numerator 8 in the fraction arises from the total number of summations over non-decreasing indices, and 6 arises from the number of ways to permute the indices j, k and l . From the bounds we have on the diamond norms of D_1, D_2, D_3, D_4 and D_5 , we obtain

$$\frac{1}{3!} \left\| \mathcal{L}^3 - \mathbb{E} \left(3! \hat{\mathcal{A}}_3 \right) \right\|_{\diamond} \leq \frac{1}{6} \sum_{j=1}^5 \|D_j\|_{\diamond} = \epsilon_2/r. \quad (47)$$

Multiplying this by r gives us the error in ϵ that is $\mathcal{O}(t^3 \mu^2/G^2)$. We have thus completed bounding the leading order errors in Theorem 5.

A.4 Tail bounds in Theorem 1

Here, we explain how the tail bounds $\epsilon_{3,1}$ and $\epsilon_{3,2}$ in Theorem 5 arise.

To evaluate the higher order terms in ϵ , we consider a convergent power series in θ given by $\mathcal{F}_\theta = \sum_{k \geq 0} f_k \theta^k$. Here, f_k is independent of θ , and \mathcal{F}_θ and f_k belong to a Banach algebra. Define $[\theta^j] \mathcal{F}_\theta = f_j$ as the j th coefficient in the power series expansion of \mathcal{F}_θ . A useful technique to bound quantities in a Banach algebra relies on the fundamental theorem of calculus, and has been used for example in Ref [10] and Ref [8]. This for example can be used to obtain the well-known integral form of the remainder term of the Taylor series of the power series \mathcal{F}_s where $s > 0$.

Lemma 6. *Let $\theta_0 = 1, s > 0$ and $\mathcal{F}_s = \sum_{k \geq 0} f_k s^k$. For every positive integer t , we have*

$$\sum_{k \geq t} f_k s^k = \int_0^{\theta_0} d\theta_1 \cdots \int_0^{\theta_{t-1}} d\theta_t \frac{d^t}{d\theta_t} \mathcal{F}_{s\theta_t}.$$

Proof. The proof of this is well-known but we provide the complete details for completeness. We first note that $\frac{d^t}{d\theta_t} \mathcal{F}_{s\theta_t} = \sum_{k \geq t} f_k (s\theta_t)^{k-t} k_{\underline{t}}$, where $k_{\underline{t}} = (k) \dots (k-t+1)$ denotes the falling factorial.

Applying the fundamental theorem of calculus on monomials in θ_t , we have

$$\begin{aligned} \int_0^{\theta_{t-1}} d\theta_t \frac{d^t}{d\theta_t} \mathcal{F}_{s\theta_t} &= \int_0^{\theta_{t-1}} d\theta_t \sum_{k \geq t} f_k (s\theta_t)^{k-t} k_{\underline{t}} \\ &= \sum_{k \geq t} f_k s^{k-t} \theta_{t-1}^{k-t-1} k_{\underline{t-1}}. \end{aligned}$$

Applying this argument iteratively gives the result. \square

Now let us denote a single timeslice of the ideal channel and SPARSTO for time $s\theta$ as $U_\theta = e^{s\theta\mathcal{L}}$ and $\hat{\mathcal{E}}_{s\theta} = \frac{1}{2}(e^{s\theta\hat{\mathcal{L}}_1} \dots e^{s\theta\hat{\mathcal{L}}_L} + e^{s\theta\hat{\mathcal{L}}_L} \dots e^{s\theta\hat{\mathcal{L}}_1})$ respectively. We proceed to evaluate the fourth derivatives of a single timeslice of U_θ and \hat{V}_θ , which are respectively given by

$$\frac{d^4}{d\theta^4} U_\theta = (s\mathcal{L})^4 U_\theta, \quad (48)$$

$$\begin{aligned} \frac{d^4}{d\theta^4} \hat{\mathcal{E}}_{s\theta} &= \frac{s^4}{2} \sum_{\substack{n_1 + \dots + n_L = 4 \\ n_1, \dots, n_L \in \mathbb{N}}} \binom{4}{n_1, \dots, n_L} \hat{\mathcal{L}}_1^{n_1} e^{s\theta\hat{\mathcal{L}}_1} \dots \hat{\mathcal{L}}_L^{n_L} e^{s\theta\hat{\mathcal{L}}_L} \\ &+ \frac{s^4}{2} \sum_{\substack{n_1 + \dots + n_L = 4 \\ n_1, \dots, n_L \in \mathbb{N}}} \binom{4}{n_1, \dots, n_L} \hat{\mathcal{L}}_L^{n_L} e^{s\theta\hat{\mathcal{L}}_L} \dots \hat{\mathcal{L}}_1^{n_1} e^{s\theta\hat{\mathcal{L}}_1}, \end{aligned} \quad (49)$$

where in the second equation we used the general Leibniz rule and the $\binom{4}{n_1, \dots, n_L} = 4!/(n_1! \dots n_L!)$ denotes the multinomial coefficient. The diamond norm of the tail of U_θ is therefore at most

$$\begin{aligned} \frac{s^4}{4!} \|\mathcal{L}^4 U_\theta\|_{\diamond} &\leq \frac{s^4}{4!} \|\mathcal{L}^4\|_{\diamond} \|U_\theta\|_{\diamond} \\ &\leq \frac{(2s)^4}{4!} \lambda^4 \|U_\theta\|_{\diamond} \\ &= \frac{(2s\lambda)^4}{4!}, \end{aligned} \quad (50)$$

where the last equality arises because U_θ is a quantum channel.

Now note that by the linearity of the derivative and expectation operator,

$$\begin{aligned} \frac{d^4}{d\theta^4} \mathbb{E}(\hat{E}_{s\theta}) &= \frac{s^4}{2} \sum_{\substack{n_1+\dots+n_L=4 \\ n_1, \dots, n_L \in \mathbb{N}}} \binom{4}{n_1, \dots, n_L} \mathbb{E}(\hat{\mathcal{L}}_1^{n_1} e^{s\theta \hat{\mathcal{L}}_1} \dots \hat{\mathcal{L}}_L^{n_L} e^{s\theta \hat{\mathcal{L}}_L}) \\ &\quad + \frac{s^4}{2} \sum_{\substack{n_1+\dots+n_L=4 \\ n_1, \dots, n_L \in \mathbb{N}}} \binom{4}{n_1, \dots, n_L} \mathbb{E}(\hat{\mathcal{L}}_L^{n_L} e^{s\theta \hat{\mathcal{L}}_L} \dots \hat{\mathcal{L}}_1^{n_1} e^{s\theta \hat{\mathcal{L}}_1}). \end{aligned} \quad (51)$$

By the independence of every stochastic Trotter step, the expectation is multiplicative so that

$$\begin{aligned} \frac{d^4}{d\theta^4} \mathbb{E}(\hat{E}_{s\theta}) &= \frac{s^4}{2} \sum_{\substack{n_1+\dots+n_L=4 \\ n_1, \dots, n_L \in \mathbb{N}}} \binom{4}{n_1, \dots, n_L} \mathbb{E}(\hat{\mathcal{L}}_1^{n_1} e^{s\theta \hat{\mathcal{L}}_1}) \dots \mathbb{E}(\hat{\mathcal{L}}_L^{n_L} e^{s\theta \hat{\mathcal{L}}_L}) \\ &\quad + \frac{s^4}{2} \sum_{\substack{n_1+\dots+n_L=4 \\ n_1, \dots, n_L \in \mathbb{N}}} \binom{4}{n_1, \dots, n_L} \mathbb{E}(\hat{\mathcal{L}}_L^{n_L} e^{s\theta \hat{\mathcal{L}}_L}) \dots \mathbb{E}(\hat{\mathcal{L}}_1^{n_1} e^{s\theta \hat{\mathcal{L}}_1}). \end{aligned} \quad (52)$$

Using the triangle inequality and the submultiplicativity of the diamond norm, the diamond norm of the tail of $\hat{\mathcal{E}}_{s\theta}$ is at most

$$\sum_{\substack{n_1+\dots+n_L=4 \\ n_1, \dots, n_L \in \mathbb{N}}} \frac{s^4 \binom{4}{n_1, \dots, n_L}}{4!} \|\mathbb{E}(\hat{\mathcal{L}}_1^{n_1} e^{s\theta \hat{\mathcal{L}}_1})\|_\diamond \dots \|\mathbb{E}(\hat{\mathcal{L}}_L^{n_L} e^{s\theta \hat{\mathcal{L}}_L})\|_\diamond. \quad (53)$$

When $n_j \geq 1$, we appeal to the Taylor series expansion for the exponential function, to get

$$\begin{aligned} \mathbb{E}(\hat{\mathcal{L}}_j^{n_j} e^{s\theta \hat{\mathcal{L}}_j}) &= \mathbb{E} \left(\sum_{k \geq 0} \frac{(s\theta)^k}{k!} \hat{\mathcal{L}}_j^{k+n_j} \right) \\ &= \sum_{k \geq 0} \frac{(s\theta)^k}{k!} \mathbb{E}(\hat{\mathcal{L}}_j^{k+n_j}) \\ &= \sum_{k \geq 0} \frac{(s\theta)^k}{k!} \mathcal{L}_j^{k+n_j} / p_j^{k+n_j-1} \\ &= p_j (\mathcal{L}_j / p_j)^{n_j} e^{s\theta \mathcal{L}_j / p_j}. \end{aligned} \quad (54)$$

It is clear that (54) also holds when $n_j = 0$. Using (54), we get the upper bound

$$\|\mathbb{E}(\hat{\mathcal{L}}_j^{n_j} e^{s\theta \hat{\mathcal{L}}_j})\|_\diamond \leq p_j (\|\mathcal{L}_j\|_\diamond / p_j)^{n_j}. \quad (55)$$

Using (55) with the multinomial theorem on (53), the diamond norm of the tail of $\hat{\mathcal{E}}_{s\theta}$ is at most

$$\frac{s^4 (p_1 \dots p_L)}{4!} (\|\mathcal{L}_1\|_\diamond / p_1 + \dots + \|\mathcal{L}_L\|_\diamond / p_L)^4. \quad (56)$$

Applying the identity $\|\mathcal{L}_j\|_\diamond \leq 2h_j$, we find that the diamond norm of the tail of $\hat{\mathcal{E}}_{s\theta}$ is at most

$$\frac{(2s)^4 (p_1 \dots p_L)}{4!} (h_1 / p_1 + \dots + h_L / p_L)^4. \quad (57)$$

By setting $\theta = 1$ and multiplying the results that we obtained from the tail bounds on a single timeslice s by a factor of r , we thus find that the expected contribution to the simulation error from the tail bounds on r repeats of $e^{s\mathcal{L}}$ and $\hat{\mathcal{E}}_s$ is at most $\epsilon_{3,1}$ and $\epsilon_{3,2}$ respectively, where

$$\begin{aligned} \epsilon_{3,1} &= \frac{2^4 r s^4}{4!} \lambda^4 \\ \epsilon_{3,2} &= \frac{2^4 r s^4}{4!} (p_1 \dots p_L) (h_1 / p_1 + \dots + h_L / p_L), \end{aligned} \quad (58)$$

from which the result follows.

B Convex programming on leading order error terms

Here we minimise the leading order term in the simulation error by optimizing over the probabilities p_j . In particular, by restricting our minimisation to only the leading order term of ϵ_1 in ϵ , we find that the leading order simulation error is $rs^2 \sum_{j=1}^L h_j^2/p_j$.

The way we find the optimal probability is by taking the derivative of the corresponding Lagrangian function, and thereby determine its turning points. If the primal and dual solutions are furthermore feasible and satisfy complementary slackness, then we know from the convexity of our problem that these primal and dual solutions are optimal for the primal and dual optimisation problems respectively.

More formally, in convex optimisation theory, we know that a primal problem and its dual problem are both optimal if and only if (1) Slater's constraint qualification holds, and (2) the primal and dual variables satisfy the so-called Karush-Kuhn-Tucker (KKT) conditions [32]. Of these two conditions (1) easily holds. The notion of the active set appears in one of the optimality conditions of (2), which is known as complementary slackness.

Now we only optimize over the p_j for which j belongs to the inactive set \bar{A} . Hence we consider the optimisation problem

$$\begin{aligned} & \text{minimize} && rs^2 \sum_{j \in \bar{A}} \frac{h_j^2}{p_j} \\ & p_j > 0, j \in \bar{A} && \\ & \text{subject to} && \sum_{j \in \bar{A}} p_j = \bar{\mu}, \\ & && p_j \leq 1, \end{aligned} \tag{59}$$

where $\bar{\mu} = \mu - |A|$. Note that the objective function here is convex in p_j , and the constraint function is linear in p_j . By treating $\bar{\mu}$ as a constant, we analytically derive the optimal value of this optimisation problem from the first order KKT conditions [32]. Since Slater's condition is satisfied, the KKT condition is necessary and sufficient for optimality. The KKT conditions require (1) the turning points of the Lagrangian to be zero, (2) primal feasibility, (3) feasibility of the Lagrange dual, (4) and complementary slackness. Complementary slackness requires the Lagrange multiplier of a constraint to be zero when that constraint is not tight.

Denoting u as the Lagrange multiplier for the equality constraint and v_j as Lagrange multipliers for the inequality constraints, the Lagrangian of (59) is

$$\begin{aligned} L &= rs^2 \sum_{j \in \bar{A}} \frac{h_j^2}{p_j} + u \left(\sum_{j \in \bar{A}} p_j - \bar{\mu} \right) + \sum_{j \in \bar{A}} (p_j - 1)v_j \\ &= \sum_{j \in \bar{A}} \left(\frac{rs^2 h_j^2}{p_j} + (u + v_j)p_j - v_j \right) - u\bar{\mu}. \end{aligned} \tag{60}$$

Note that

$$\frac{\partial L}{\partial p_j} = \frac{-rs^2 h_j^2}{p_j^2} + u + v_j, \tag{61}$$

and hence the turning point of the Lagrangian L occurs when

$$p_j = \frac{\sqrt{rs}h_j}{\sqrt{u+v_j}}. \tag{62}$$

Note that we have $u \in \mathbb{R}$ and $v_j \geq 0$. From complementary slackness, we know that if the optimal $p_j < 1$, then we correspondingly have $v_j = 0$. When $p_j = 1$, the constraint corresponding to v_j is active, and $v_j > 0$. Hence it follows that whenever $p_j < 1$, we have

$$\sqrt{u}p_j = \sqrt{rs}h_j. \tag{63}$$

Conversely, when $p_j = 1$, we have

$$u + v_j = rs^2 h_j^2. \tag{64}$$

Note here that we have not verified that the problem is primal feasible, namely, that we need to check that $\sum_{j \in \bar{A}} p_j = \bar{\mu}$. This can be satisfied whenever we have

$$\bar{\mu} = s\sqrt{\frac{r}{u}} \sum_{j \in \bar{A}} h_j. \tag{65}$$

Thus our ansatz for p_j is

$$p_j = \frac{\bar{\mu} h_j}{\sum_{j \in \bar{A}} h_j}, \quad (66)$$

with the regularity condition that this formula satisfies $p_j < 1$ for $j \in \bar{A}$.