




Problem-specific classical optimization of Hamiltonian simulation

Refik Mansuroglu ^{*}, Felix Fischer , and Michael J. Hartmann 

Department of Physics, Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Staudtstraße 7, 91058 Erlangen



(Received 13 June 2023; accepted 25 September 2023; published 11 October 2023)

Nonequilibrium time evolution of large quantum systems is a strong candidate for quantum advantage. Variational quantum algorithms have been put forward for this task, but their quantum optimization routines suffer from trainability and sampling problems. Here, we present a classical preprocessing routine for variational Hamiltonian simulation that circumvents the need for a quantum optimization by expanding rigorous error bounds in a perturbative regime for suitable time steps. The resulting cost function is efficiently computable on a classical computer. We show that there always exists potential for optimization with respect to a Trotter sequence of the same order and that the cost value has the same scaling as Trotter in simulation time and system size. Unlike previous work on classical preprocessing, the method is applicable to any Hamiltonian system independent of locality and interaction lengths. Via numerical experiments for spin-lattice models, we find that our approach significantly improves digital quantum simulation capabilities with respect to Trotter sequences for the same resources. For short times, we find accuracy improvements of more than three orders of magnitude for our method as compared to Trotter sequences of the same gate number. Moreover, for a given gate number and accuracy target, we find that the preoptimization we introduce enables simulation times that are consistently more than ten times longer for a target accuracy of 0.1%.

DOI: [10.1103/PhysRevResearch.5.043035](https://doi.org/10.1103/PhysRevResearch.5.043035)

I. INTRODUCTION

Quantum simulation is believed to be one of the first applications in quantum computing that may show practical advantages over its classical counterparts [1,2]. To compile a digital quantum algorithm for simulating a time evolution $U(t)$ that will be typically generated by a p -local Hamiltonian,

$$U(t) = e^{-itH} \quad \text{with} \quad H = \sum_j c_j H_j, \quad (1)$$

a Trotter-Suzuki decomposition [3], that approximates $U(t)$ as a product of local exponentials, can be employed. Although so-called beyond-Trotter methods, such as more general product formulas [4–6], quantum signal processing techniques [7], or Krylov subspace inspired algorithms [8] have been investigated, recent studies on Trotter error scaling [9,10] have proven Trotter sequences to still be competitive.

Nonetheless, in specific applications, Trotter sequences fail to deliver the optimal solution for a product formula with fixed gate count [11]. Among other optimization strategies, variational quantum algorithms have been put forward to find optimized (system-specific) gate sequences for Hamiltonian simulation [12–17].

Although variational quantum algorithms have recently piqued significant interest as possible applications on near-term quantum hardware [18], they hide intrinsic difficulties

that make their implementation on near-term devices very challenging. Vanishing gradients (Barren plateaus) [19–21] sample overhead for efficient measurements [22,23] and effects of noise on cost function evaluations [24] are only some of the hurdles to overcome. Thus, key to a successful near-term quantum algorithm is developing strategies to use quantum hardware in the crucial solution step, but otherwise as little as possible.

A drastic, however, promising, approach to solve this is not to do quantum optimization at all, but rather push the whole optimization loop into classical preprocessing, which requires a cost function that is efficiently computable by a classical machine. This idea has been demonstrated to be applicable in specific cases, such as translational invariant models [25] in which parameters found on small systems can be reused in large systems as well as systems with efficient tensor network representations [26–29] to calculate error measures classically.

Here, we present a cost function that is classically efficiently computable and use it for classical optimization of quantum circuits that implement time evolution in preprocessing. We show that this cost function yields a faithful upper bound on the two-norm error of the ansatz unitary if one restricts the single time step to be small, an assumption that is common to all product formulas [4–6,9]. Our cost function is agnostic of the considered initial state and relies on a perturbative expansion of a rigorous upper bound that can be transformed into a polynomial in the parameters. The number of terms that need to be calculated scales with the number of noncommuting terms H_j in the Hamiltonian. While this number is generally dependent on the locality and interaction distance in H , it is bounded by a power of the qubit number for models with only two-body interactions. There are thus

^{*}refik.mansuroglu@fau.de

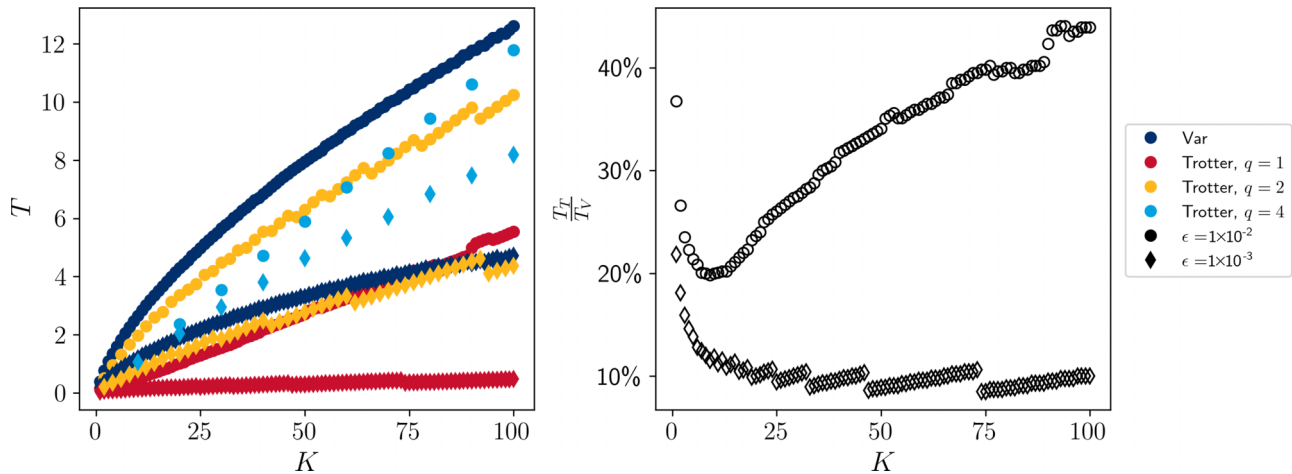


FIG. 1. Maximal simulation time T of variational and Trotter sequences of order $q \in \{1, 2, 4\}$ (left) and simulation time ratio T_T/T_V of first order sequences (right) for a fixed gate count that is determined by K and a threshold accuracy $\epsilon \in \{10^{-2}, 10^{-3}\}$ evaluated on a nearest-neighbor XY model on a 3×3 quadratic lattice and random interaction strength $J_{\mu,\nu}^{(y)}$ centered around 0.5, $J_{\mu\nu}^{(z)}$ centered around 1 and $h_\mu = 0.25$. The small jumps in T_T/T_V are due to the finite step size in the search for optimal time steps t to meet the error threshold ϵ . See Sec. IV for details of the numerical experiments to produce the data shown here.

no intrinsic limitations on the Hamiltonian for the use of our method. We further give a perturbative and classically computable error estimate that allows us to predict error scalings in system size and simulation time if the optimized gate sequence is repeated K times.

We illustrate the improvements enabled by our approach via exact numerical simulations for an XY model with random nearest-neighboring interactions. Our results (cf. Fig. 1) show that the maximally reachable simulation times T for a given target accuracy ϵ and a fixed gate budget $G = \mathcal{O}(K)$ are more than ten times longer than for a Trotter sequence of the same order for cases with $K \gtrsim 20$ and $\epsilon = 10^{-3}$. Moreover, the variational sequence even reaches slightly larger simulation times than second-order Trotter and is competitive with fourth-order Trotter, whereas these higher order formulas provide less data points for fixed gate budget.

The paper is structured as follows. We begin by introducing the second-order perturbative distance of a variational sequence from the correct time evolution operator in Sec. II and show that it bounds the algorithm error up to third-order corrections in the parameters. We further show that the perturbative distance always has a nonvanishing gradient at the point in the parameter space that corresponds to the Trotter sequence yielding a general guarantee for the existence of an optimal solution away from Trotter. In Sec. III, we compute the scaling of the error in the extrapolation of found optimal gate sequence in system size and simulation time. Finally, we demonstrate our findings and the advantages they enable in numerical experiments on the XY model in Sec. IV before we close with a summary and outlook.

II. PERTURBATIVE COST FUNCTION

We are interested in a variational ansatz for quantum simulation $U(t) = \exp(-itH)$. It has been demonstrated

before [11,25] that Trotter sequences can be optimized in a model-specific way using the ansatz

$$U_{\text{var}} = \prod_{r=1}^R \prod_{j=1}^M e^{-i\theta_{r,j} H_j}. \tag{2}$$

The inner product factorizes all M Hamiltonian terms indexed with j and the outer product adds the freedom to use in total R layers in which the parameters can be different. While the choice of ansatz is not necessarily restricted to Eq. (2), it has the advantage of containing the Trotter solution. If we use the parameters $\theta_{r,j} = \frac{c_j t}{R}$, Eq. (2) reduces to a Trotter decomposition with Trotter number R .

Finding the optimal solution for an operator approximation is hard, since error measures for the time evolution operator, such as

$$\epsilon_{\text{var}} = \|U(t) - U_{\text{var}}\|, \tag{3}$$

where $\|\cdot\|$ is a yet unspecified, unitary invariant norm, requiring exponential resources, in general. In this paper, we present a classically efficient optimization strategy that yields good results in the relevant regime of small time steps $\frac{t}{R}$, where power expansions can be applied. To this end, we introduce an approximation to an upper bound of the error measure ϵ_{var} for the time evolution operator in Eq. (3), which is provided by a quantity that we call perturbative distance.

Definition 1. Let $H = \sum_j^M c_j H_j$ be a time-independent Hamiltonian. Define the perturbative distance

$$C(\theta) = \left\| i \sum_j \xi_j(\theta) H_j - \sum_{j>j'} \chi_{j,j'}(\theta) [H_j, H_{j'}] \right\| \tag{4}$$

with the coefficients

$$\xi_j(\theta) = \sum_r \theta_{r,j} - tc_j, \quad (5)$$

$$\chi_{j,j'}(\theta) = \frac{1}{2} \left[\sum_r (\theta_{r,j} \theta_{r,j'}) + \sum_{r>r'} (\theta_{r,j} \theta_{r',j'} - \theta_{r,j'} \theta_{r',j}) \right]. \quad (6)$$

The perturbative distance $C(\theta)$ represents an approximation to an upper bound of the error measure ϵ_{var} for the time evolution operator. We make this statement precise in the following proposition.

Proposition 1. Let $H = \sum_j^M c_j H_j$ be a time-independent Hamiltonian and $C(\theta)$ the perturbative distance defined in Eq. (4). Then the error ϵ_{var} defined in Eq. (3) is bounded up to third-order corrections by

$$\epsilon_{\text{var}} \leq C(\theta) + \mathcal{O}(\theta^3). \quad (7)$$

Proof. We can write the difference operator in Eq. (3) in the form

$$e^{-itH} - e^Z, \quad (8)$$

where Z is determined by merging U_{var} into one exponential via the Baker-Campbell-Hausdorff (BCH) lemma. We show in Lemma 2 (see Appendix A 2 for a proof) that up to third order

$$Z = i \sum_{r,j} \theta_{r,j} H_j - \frac{1}{2} \sum_{(r,j)>(r',j')} \theta_{r,j} \theta_{r',j'} [H_j, H_{j'}] + \mathcal{O}(\theta^3). \quad (9)$$

We assumed a specific ordering of the multi-index (r, j) for the sake of being precise. The statement can be very well applied to arbitrary orderings of the exponentials in Eq. (2). We finally make use of a well-known estimate (see Lemma 1 in Appendix A 1 for a proof)

$$\|e^{iA} - e^{iB}\| \leq \|A - B\|, \quad (10)$$

that leaves us with the proposition. ■

To make $C(\theta)$ an approximate upper bound to the error ϵ_{var} , we need to assume $\frac{t}{R} \|H\| \ll 1$ and bound the parameters θ to be at the same order of magnitude. This assumption yields a time-dependent bound on R , which is not surprising. In general, Lieb-Robinson bounds on the entanglement that is generated by time evolution gives a relation between t and R [30]. If one manages to minimize $C(\theta)$ within the perturbative regime, the leading term in the upper bound will be of order $\mathcal{O}(\theta^3)$, yielding a new error estimate that we will analyze in Sec. III.

Note that first-order Trotter sequences can be represented by Eq. (2), setting the parameters $\theta_{r,j} = \frac{tc_j}{R}$. The Trotter solution has the advantage that it makes the linear contribution in θ , respectively $\frac{t}{R}$, vanish. It is, however, not the only solution that achieves this. In fact, we can reduce the number of parameters by M if we constrain the linear terms by $\xi_j = 0 \forall j$. Let us thus fix the last layer ($r = R$) by setting

$$\theta_{R,j} = tc_j - \sum_{r \neq R} \theta_{r,j}, \quad (11)$$

so $(R - 1)M$ parameters remain. Note that Eq. (11) is a choice that turns out to be helpful as long as the operators that

contribute in linear order are all orthogonal to the operators in the second-order terms, which we assume in the following. Otherwise first and second order terms could be used to cancel each other.

We are still free to choose a specific norm. While in finite dimensions, all norms are equivalent, choosing a norm to describe the error ϵ , implies a notion of how the algorithm performs for specific initial states. If we were to choose the operator norm $\|\cdot\|_\infty$, for instance, ϵ can be connected to an infidelity for the worst case initial state, while a two-norm $\|\cdot\|_2$ yields an average case error. Later, $C(\theta)$ shall be used as a cost function in variational optimization, so it should be continuously differentiable in the parameters $\theta_{r,j}$. While this rules out the operator norm, all Schatten p -norms can be applied. If we consider the two-norm, we see that the computation of $C(\theta)$ is classically efficient. In this case,

$$C(\theta)^2 = -2^{-n} \sum_{j>j'} \sum_{k>k'} \chi_{j,j'} \chi_{k,k'} \text{Tr}([H_j, H_{j'}][H_k, H_{k'}]), \quad (12)$$

where we included a normalization factor 2^{-n} that prevents $C(\theta)$ from growing in the Hilbert space dimension. Typically, the commutators $[H_j, H_{j'}]$ follow a simple algebraic structure (for instance, if all H_j are Pauli strings). The factors $\text{Tr}([H_j, H_{j'}][H_k, H_{k'}])$ can thus efficiently be computed for the model of interest, typically even by hand, leaving $C(\theta)$ to be a fourth order polynomial in θ .

The Trotter solution, while becoming exact in the limit $R \rightarrow \infty$, can always be beaten for finite $R \geq 3$, as we will show in the following.

Proposition 2. Let $H = \sum_j c_j H_j$ a Hamiltonian and U_{var} as defined in Eq. (2) with $R \geq 3$. The Trotter solution $\theta_{r,j}^T = \frac{tc_j}{R}$ describes a (local) minimum of $C(\theta)$ if and only if the Trotter formula is exact, i.e., $C(\theta^T) = 0 \forall t$.

Proof (Sketch). For the forward direction \Rightarrow , we explicitly calculate the gradient at the Trotter solution $\nabla C(\theta^T)$. Setting this to zero, gives a relation

$$\sum_{j>l} \sum_{k>k'} \mathcal{A}_{j,k,k',l} = \frac{2r - R}{2(r - 1) - R} \sum_{j<l} \sum_{k>k'} \mathcal{A}_{j,k,k',l} \quad \forall r, l, \quad (13)$$

where $\mathcal{A}_{j,k,k',l}$ is a tensor independent of r and R . For $R \geq 3$, we show that this forces the left-hand side of Eq. (13) to vanish, making the Trotter sequence exact. For details, see Appendix A 3. The reverse direction \Leftarrow is trivial. ■

III. ERROR SCALING

Since the proposed classical optimization scheme can only make valid statements within the perturbative regime $\theta \ll \frac{1}{\|H\|}$, longer simulation times either require large R or repetitions of a small time step. A similar extrapolation can be done in system size. This section studies the scaling of the perturbative distance along these extrapolations. In the case that the second-order perturbative cost function is optimized to values that become negligible to higher orders, the error scaling is determined by a third-order term $E(\theta)$ for which

$$\epsilon_{\text{var}} \leq C(\theta) + E(\theta) + \mathcal{O}(\theta^4) \quad (14)$$

holds. First, we consider the extrapolation in system size in the presence of translation symmetries.

Proposition 3. Let $H = \sum_{a=1}^A \sum_{j=1}^n c_a H_a^{(j)}$ be a p -local, translation invariant Hamiltonian, where index a labels different interaction terms and the dependence on the qubit number, indicated by superscript (j) , only denotes on which qubit $H_a^{(j)}$ has support. Let further U_{var} be as defined in Eq. (2).

If the parameters are chosen to be translation invariant, i.e., $\theta_{r,j} = \theta_{r,a}$, then the cost value for a system of n qubits on a D -dimensional geometry is bounded by the cost value $C_{\text{unit}}(\theta)$ of a system of $(2p-1)^D$ qubits and the error is bounded by $E_{\text{unit}}(\theta)$ of a system of $(3p-2)^D$ qubits

$$C(\theta) \leq n C_{\text{unit}}(\theta) \quad \text{and} \quad E(\theta) \leq n E_{\text{unit}}(\theta). \quad (15)$$

Further, if all the commutators $[H_j, H_{j'}]$ are orthogonal with respect to the Hilbert-Schmidt product, cost and error read

$$C(\theta) = \sqrt{n} C_{\text{unit}}(\theta) \quad \text{and} \quad E(\theta) = \sqrt{n} E_{\text{unit}}(\theta). \quad (16)$$

See Appendix A 4 for a proof. Remarkably, the calculation of $C(\theta)$ on translation-invariant systems is reduced to a problem on $(2p-1)^D$ qubits only. If the optimal parameters are used for a larger system, the cost function will scale as $\mathcal{O}(n)$. In fact, a system that is small enough can be exactly diagonalized on a classical computer and an exact measure of error, such as $\|U(0, t) - U_{\text{var}}\|$ can be calculated. A successful extrapolation of optimal parameters for translation invariant systems to larger system sizes has been found in Ref. [25] and also shows a linear scaling of the squared cost function in system size n , even beyond the perturbative regime. While we restrict the discussion here to translation symmetries, the behavior of other scaling transformations such as clusterings or pooling-inspired transformations can be analogously checked.

Optimizing on longer simulation times still within the perturbative regime comes with an increase in R and therefore with the number of parameters making a successful optimization hard. Instead, long simulation times can be reached by repeating a pretrained single time step fixing the number of parameters to be RM . The scaling of the cost function under this extrapolation in simulation time is studied in the following.

Proposition 4. Let $U_{\text{var}}(\theta)$ be an ansatz for variational Hamiltonian simulation as defined in Eq. (2) with the cost value $C(\theta)$. The ansatz

$$U_{\text{var}}^K(\theta^K) = \left(\prod_{r=1}^R \prod_{j=1}^M e^{-i\theta_{r,j} H_j} \right)^K. \quad (17)$$

simulates time $T = Kt$ using KR layers and K copies of the RM parameters. The corresponding cost value reads

$$C(\theta^K) = KC(\theta). \quad (18)$$

Proof idea. The extrapolated ansatz is equivalent to the ansatz before by mapping $R \rightarrow KR$, $t \rightarrow Kt$ and identifying $\theta_{r,j} = \theta_{r \% R, j}$, where $\%$ denotes the modulo operation. Plugging this into the definition of χ from Eq. (6) yields a linear factor K . See Appendix A 5 for details. ■

In total, the cost function scales as $\mathcal{O}(K)$ when extrapolating in time. Note that the leading order Trotter error scales the

same under the above mapping

$$\|U(0, t) - U_{\text{Trotter}}\| = \mathcal{O}\left(\frac{t^2}{R}\right) \rightarrow \mathcal{O}\left(K \frac{t^2}{R}\right). \quad (19)$$

The error $E(\theta)$ of the repeated sequence does not admit a common K -dependent factor that can be pulled out, but rather introduces extra terms, Λ , that make the error scaling worse than for Trotter sequences (see Appendix A 5 for details). The extra terms will, in general, contribute to an $\mathcal{O}(K^2)$ scaling:

$$E(\theta^K) \leq KE(\theta) + \frac{K(K-1)}{2} \Lambda. \quad (20)$$

Although this result means that usually the improvement with respect to Trotter will eventually melt away at long times, our numerical experiments show that there are significant improvements for the timescales that are relevant to reach quantum advantage in digital quantum simulation.

IV. NUMERICAL EXPERIMENT

To explore the accuracy improvement that can be enabled with our approach in specific examples, we study the performance of the derived cost function by optimizing a variational sequence with the second-order perturbative distance defined in Eq. (4) and comparing it to the exact value for the difference of time evolutions $\epsilon_{\text{var}} = \|U(t) - U_{\text{var}}\|$ and $\epsilon_{\text{Trotter}} = \|U(t) - U_{\text{Trotter}}\|$ as well as to the leading order error term $E(\theta)$ (cf. Eq. (14)) that is efficiently computable classically. For this comparison, we study the ratios

$$R_\epsilon = \frac{\epsilon_{\text{Trotter}}}{\epsilon_{\text{var}}} \quad R_E = \frac{\epsilon_{\text{Trotter}}}{E(\theta)} \quad (21)$$

of perturbative (R_E) and exact (R_ϵ) error measures comparing Trotter and variational sequences. In Eq. (21), $C(\theta^T)$ denotes the perturbative distance for choosing the variational parameters equal to the Trotter choice $\theta_{r,j} = \frac{t c_j}{R}$.

From the bound in Eq. (14), we can also infer an approximate lower bound for the exact error ratio

$$R_\epsilon \geq \frac{\epsilon_{\text{Trotter}}}{C(\theta) + E(\theta)} + \mathcal{O}(\theta^4) \approx R_E, \quad (22)$$

where we assumed that $C(\theta) \ll E(\theta)$ after a successful optimization of the parameters θ .

As an example, we consider an XY model with a fully connected interaction graph that is described by the Hamiltonian (up to unitary equivalence)

$$H_{XY} = - \sum_{\mu > \nu} (J_{\mu\nu}^{(y)} Y_\mu Y_\nu + J_{\mu\nu}^{(z)} Z_\mu Z_\nu) + \sum_{\mu} h_\mu X_\mu. \quad (23)$$

The sum over μ, ν counts every pair of qubits and the symmetric matrices $J_{\mu\nu}^{(y,z)}$ encode the interaction strengths between qubits μ and ν . In our numerical simulations, we normalize the Hamiltonian to $\|H\| = \sqrt{n}$ to fix the timescale.

To calculate $C(\theta)$ for the XY model, we define a mapping between the single index $j \in \{1, \dots, M\}$ used in the above derivations and the Hamiltonians in Eq. (23) labeled by the qubit indices $\mu \in \{1, \dots, n\}$ and the interaction type. The non-vanishing commutator terms that appear in $C(\theta)$ are of the form $[Z_\mu Z_\nu, X_\mu]$, $[Y_\mu Y_\nu, X_\mu]$ or $[Z_\mu Z_\nu, Y_\nu Y_\sigma]$. This allows the

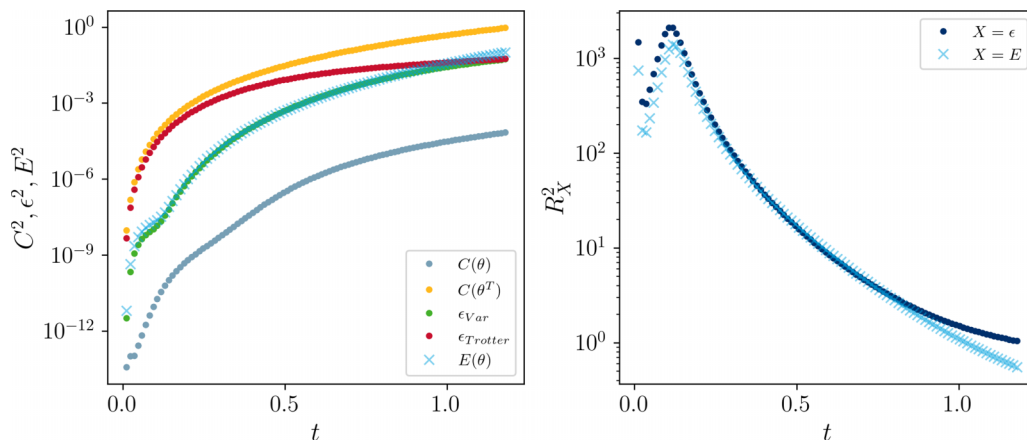


FIG. 2. Perturbative (C, E) and exact (ϵ) error measures (left) and exact error ratio R_ϵ and the ratio of error estimates R_E (right) as defined in Eq. (21) evaluated on a nearest-neighbor XY model on a 3×3 quadratic lattice and random interaction strength $J_{\mu,\nu}^{(y)}$ centered around 0.5, $J_{\mu,\nu}^{(z)}$ centered around 1 and $h_\mu = 0.25$.

reduction of $C(\theta)$ to $\mathcal{O}(n^3)$ nonvanishing terms in the fully connected case and to $\mathcal{O}(n)$ nonvanishing terms in a nearest-neighbor XY model. See Appendix C for an explicit form of the perturbative distance $C(\theta)$. Note that we are still free to turn off specific interactions from the fully connected graph to study lattice geometries of any dimension. We will now turn to a two-dimensional geometry for which there is no analytical solution using free fermion mapping, such as in one dimension.

In Fig. 2, we plot the cost values C, E and ϵ , as well as the ratios R_ϵ and R_E for different times t . The optimizer manages to find small values of $C(\theta)$ for all times. For small enough times $t < \frac{R}{\|H\|} = 1$, the perturbative distance also incorporates a faithful indicator for the exact norm in a sense that the optimal parameters also yield a significant decrease in the exact error ϵ_{var} of the variational sequence. Also, the error estimate E and the ratio R_E yield very good approximations for the true error ratio R_ϵ as predicted in Eq. (22). The numerics hence indicate that the approximate upper bound derived in Eq. (14) is tight.

In Fig. 3, the optimal parameters found before for a fixed time step $\sqrt{2nt} \in \{0.05, 0.5, 1.5\}$ are being repeated up to $K = 10$ times to simulate longer times. As predicted in Proposition 4, the square of the perturbative distance grows with a factor K^2 . For total simulation times $\sqrt{2nKt} < 1.5$, this scaling is also reflected in the exact cost ϵ . Above this critical time, higher order contributions become non-negligible, which explains why the scaling ϵ deviates from $\mathcal{O}(K^2)$. Remarkably, the error ratio between Trotter and variational sequences stays constant also far beyond this critical time until the improvement of the variational sequence eventually starts melting away at $t > 1$.

Let us take this extrapolation to a practical case in which we have a gate budget $G = \mathcal{O}(KRn)$ that will scale with the number of qubits n and the number of circuit repetitions K , and a maximal target accuracy ϵ . A natural question from a practitioner’s point of view is to ask, What is the maximally reachable simulation time $T = Kt$ with these fixed resources. We study this question in Fig. 1 where we plot T_V for variational and Trotter sequences of order $q \in \{1, 2, 4\}$ picking

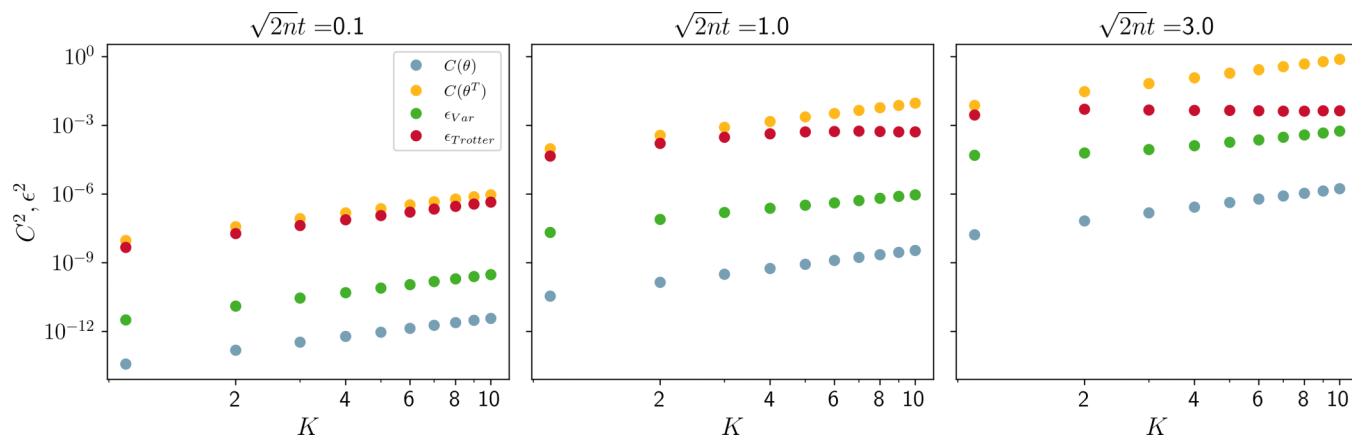


FIG. 3. Perturbative cost values C compared with exact norm cost function ϵ for a repeated sequence of variational and Trotter solution evaluated on a nearest-neighbor XY model on a 3×3 quadratic lattice and random interaction strength $J_{\mu,\nu}^{(y)}$ centered around 0.5, $J_{\mu,\nu}^{(z)}$ centered around 1 and $h_\mu = 0.25$. A single time step always corresponds to $R = 3$ layers and the single step time increases from left to right $\sqrt{2nt} \in \{0.05, 0.5, 1.5\}$. From the log-log plot, one can read off $C(\theta^K)^2 = \mathcal{O}(K^2)$ as predicted by Proposition 4.

the maximal time step t that corresponds to an error ϵ just below the threshold. While the variational sequence reaches larger simulation times than first- and second-order Trotter sequences, its maximal simulation time is comparable to the fourth-order Trotter formula in the $\epsilon = 10^{-2}$ case and in the $\epsilon = 10^{-3}$ case for circuits with $K < 25$. The simulation time of the variational sequence is then surpassed by that of a $q = 4$ Trotter formula in the case $\epsilon = 10^{-3}$ at depths $K > 25$. However, since higher order Trotter formulas require more gates for a single time step, the resolution of the time evolution for a given gate number also becomes more sparse. For NISQ applications, which are typically $K < 20$, there might be only a single- or no fourth-order data point in the implementable range, depending on the system size. We also plot the ratio of simulation times T_T/T_V for first-order sequences to show that our method can reach simulation times that are more than ten times longer than for a Trotter sequence of the same order for $\epsilon = 10^{-3}$ and $K \gtrsim 20$.

We have also computed the values of the ratios R_C and R_ϵ for different times t for a transverse field Ising model in Appendix B. Also for this model, our approach leads to a dramatic improvement. This model, however, has a particularly simple commutator algebra that is routed in its low connectivity. Therefore, many of the two-qubit gates in a Trotter sequence commute with each other and, as a consequence, a first-order Trotter sequence of the Ising model is unitarily equivalent to a second-order Trotter sequence. Improving upon this sequence would require a third-order perturbative distance analogous to our discussion above.

V. CONCLUSION

In this paper, we presented a classically efficient way to estimate error measures for variational time evolution in the regime $\frac{t}{R}, \theta \ll \frac{1}{\|H\|}$ and proposed an optimization routine for variational Hamiltonian simulation that is executed completely in classical preprocessing. The method is applicable to arbitrary p -local Hamiltonians and only requires computing commutators of individual terms in the Hamiltonian.

We show that—analogue to a Trotter decomposition—the linear contribution to the error in the parameters θ can be forced to exactly vanish. The remaining parameters can then be found to minimize the second-order error terms for which the existence of a better solution than Trotter is always guaranteed within the perturbative regime. A generalization to higher orders is straightforward and remains only a matter of diligence. The Trotter-inspired error scalings in system size and extrapolations in simulation time are conserved also when deviating from the Trotter solution. Our findings are backed up with numerical experiments on the XY model with random interaction strengths.

For future work, it would be interesting to study other physically motivated models like nonlocal fermionic encodings. Also, an extension to time-dependent Hamiltonians would be interesting, although the extrapolation to larger times is not straightforward for this case. Another possibility to simplify the computation of the perturbative distance is to change the exact error measure that is to be approximated. This can be done by fixing the initial state of the quantum simulation or to consider alternative error measures, such as the conservation

of statistical moments of the Hamiltonian, as recently proposed in Ref. [17].

ACKNOWLEDGMENTS

This work received support from the German Federal Ministry of Education and Research via the funding program Quantum Technologies—from basic research to the market under Contract No. 13N16067 EQUAHUMO. It is also part of the Munich Quantum Valley, which is supported by the Bavarian state government, with funds from the Hightech Agenda Bayern Plus.

APPENDIX A: DERIVATION OF THE COST FUNCTION

1. An upper bound for the error

To derive an upper bound for the error of a variational gate sequence for time evolution, we start by reviewing a general upper bound for differences of unitaries shown in Ref. [31].

Lemma 1. Let A, B be Hermitian operators and $\|\cdot\|$ any norm that admits unitary invariance, then

$$\|e^{iA} - e^{iB}\| \leq \|A - B\|. \quad (\text{A1})$$

Proof. We begin by writing the difference operator $e^{iA} - e^{iB}$ in integral form

$$\begin{aligned} e^{iA} - e^{iB} &= \int_0^1 dx \frac{d}{dx} [e^{i(B+x(A-B))}] \\ &= \int_0^1 dx \int_0^1 dy [e^{iy(B+x(A-B))} (A - B) \\ &\quad \times e^{i(1-y)(B+x(A-B))}]. \end{aligned} \quad (\text{A2})$$

In the first step, we used the fundamental theorem of calculus and, in the second step, we used the identity for the derivative of the exponential map

$$\frac{d}{dx} e^{F(x)} = \int_0^1 dy e^{yF(x)} F'(x) e^{(1-y)F(x)} \quad (\text{A3})$$

for any operator valued function $F(x)$ and its derivative $F'(x)$. If we consider the norm of the operator in Eq. (A2), we can use the triangle inequality and its unitary invariance to get to the result

$$\begin{aligned} \|e^{iA} - e^{iB}\| &\leq \int_0^1 dx \int_0^1 dy \|e^{iy(B+x(A-B))} (A - B) \\ &\quad \times e^{i(1-y)(B+x(A-B))}\| \\ &= \int_0^1 dx \int_0^1 dy \|A - B\| = \|A - B\|. \end{aligned} \quad (\text{A4})$$

Let $H = \sum_{j=1}^M c_j H_j$ be Hermitian. We are interested in a time evolution operator built from gates of the form $e^{-i\theta H_j}$. For this choice, the variational gate sequence will have the form of Eq. (2). We can use the BCH formula to find an anti-Hermitian operator Z that combines all exponentials in Eq. (2) into one. Using Lemma 1, we then get a general bound for the

difference operator

$$\|U(0, t) - U_{\text{var}}\| = \|e^{-itH} - e^Z\| \leq \|itH + Z\|. \quad (\text{A5})$$

Z contains infinitely many terms in general. To transform Eq. (A5) into something that is classically computable, we will truncate Z in a perturbative regime.

2. Truncation of BCH formula

In the limit of small time steps $\frac{t}{R} \rightarrow 0$, a Trotter formula becomes exact. We want to keep $\frac{t}{R}$ small but finite in the following. We further assume the parameters $\theta_{r,j}$ to be of the same order as $\frac{t}{R}$. Consider the following supplementary lemma.

Lemma 2. Let $A_j, j \in \mathfrak{J}$ for some index set \mathfrak{J} be a list of bounded operators and denote the maximal operator norm $\max_j \|A_j\|_\infty =: a$. It holds that

$$\begin{aligned} \prod_{j \in \mathfrak{J}}^{\leftarrow} e^{A_j} &= e^B, \quad (\text{A6}) \\ B &= \sum_{j \in \mathfrak{J}} A_j + \frac{1}{2} \sum_{j>k} [A_j, A_k] + \frac{1}{12} \sum_{j \neq k} [A_j, [A_j, A_k]] \\ &\quad + \frac{1}{6} \sum_{j>k>l} ([A_j, [A_k, A_l]] + [A_l, [A_k, A_j]]) + \mathcal{O}(a^4), \quad (\text{A7}) \end{aligned}$$

where we assumed an implicit ordering on index set \mathfrak{J} , and the arrow above the product implies that the exponential corresponding to the smallest index is being multiplied from the right.

Proof. We proceed by induction over the size of index set \mathfrak{J} . For $|\mathfrak{J}| = 1$, the statement becomes trivial and for $|\mathfrak{J}| = 2$ equivalent to the BCH formula. We hence assume that the statement holds for an index set \mathfrak{J} of size $I - 1$. Let us rename Z from above to Z_{I-1} to emphasize the size of the index set. If we add an I th index, we get

$$e^{A_I} \prod_{j \in \mathfrak{J}}^{\leftarrow} e^{A_j} = e^{A_I} e^{Z_{I-1}} = e^{Z_I}, \quad (\text{A8})$$

where we get Z_I from the BCH lemma

$$Z_I = A_I + Z_{I-1} + \frac{1}{2}[A_I, Z_{I-1}] + \frac{1}{12}[A_I - Z_{I-1}, [A_I, Z_{I-1}]] + \mathcal{O}(a^4). \quad (\text{A9})$$

Let us recollect all orders in a . For first order, we get $A_I + \sum_{i=1}^{I-1} A_i = \sum_{i=1}^I A_i$. For the second-order term, we gather two terms from Eq. (A9):

$$\frac{1}{2} \sum_{i>j}^{I-1} [A_i, A_j] + \frac{1}{2} \left[A_I, \sum_{i=1}^{I-1} A_i \right] + \mathcal{O}(a^3) = \frac{1}{2} \sum_{i>j}^I [A_i, A_j] + \mathcal{O}(a^3). \quad (\text{A10})$$

Finally, the third-order term has three contributions from Eq. (A9),

$$\begin{aligned} &\frac{1}{12} \sum_{j \neq k}^{I-1} [A_j, [A_j, A_k]] + \frac{1}{6} \sum_{j>k>l}^{I-1} ([A_j, [A_k, A_l]] + [A_l, [A_k, A_j]]) + \frac{1}{4} \sum_{j>k}^{I-1} [A_l, [A_j, A_k]] \\ &\quad + \frac{1}{12} \sum_j^{I-1} [A_I, [A_I, A_j]] + \frac{1}{12} \sum_{j,k}^{I-1} [A_j, [A_k, A_I]] \\ &= \frac{1}{12} \sum_{j \neq k}^{I-1} [A_j, [A_j, A_k]] + \frac{1}{6} \sum_{j>k>l}^{I-1} ([A_j, [A_k, A_l]] + [A_l, [A_k, A_j]]) + \frac{1}{4} \sum_{j>k}^{I-1} [A_I, [A_j, A_k]] \\ &\quad + \frac{1}{12} \sum_j^{I-1} ([A_I, [A_I, A_j]] + [A_j, [A_j, A_I]]) + \frac{1}{12} \sum_{j>k}^{I-1} (-[A_I, [A_j, A_k]] + [A_k, [A_j, A_I]]) + \frac{1}{12} \sum_{j<k}^{I-1} [A_j, [A_k, A_I]] \\ &= \frac{1}{12} \sum_{j \neq k}^I [A_j, [A_j, A_k]] + \frac{1}{6} \sum_{j>k>l} ([A_j, [A_k, A_l]] + [A_l, [A_k, A_j]]), \quad (\text{A11}) \end{aligned}$$

where we first split the sum $\sum_{j,k} = \sum_{j=k} + \sum_{j>k} + \sum_{j<k}$ and used the Jacobi identity of the commutator in the first step. In the second, we renamed indices to absorb the fourth sum into the second terms and the rest into the first term accordingly. Taking all three orders together, we conclude with the statement. ■

We can now apply Lemma 1 to the BCH Hamiltonian Z of a variational decomposition of H . While we will need the third-order term of Eq. (A7) when discussing error scalings, for now we only make use of the first- and second-order term

$$\begin{aligned}
 -itH - Z &= -it \sum_{j=1}^M c_j H_j + i \sum_{r,j} \theta_{r,j} H_j - \frac{1}{2} \sum_{(r,j) > (r',j')} \theta_{r,j} \theta_{r',j'} [H_j, H_{j'}] + \mathcal{O}(\theta^3) \\
 &= i \sum_j \left(\sum_r \theta_{r,j} - t c_j \right) H_j - \frac{1}{2} \sum_{(r,j) > (r',j')} \theta_{r,j} \theta_{r',j'} [H_j, H_{j'}] + \mathcal{O}(\theta^3), \tag{A12}
 \end{aligned}$$

where we implied an order on the index set $\mathcal{J} = \{(r, j)\}_{r,j}$. The specific ordering is not important for our discussion. Without loss of generality, we will choose the order $(r, j) < (r', j')$, if and only if $r < r'$ or $r = r'$ and $j < j'$. If one fixes the last layer of parameters as in Eq. (11), the linear contribution in Eq. (A12) will vanish. For the sake of clarity, we will still write $\theta_{R,j}$ in the following even though it is not a free parameter. Finally, we can write the second-order term of the upper bound

$$\begin{aligned}
 \|itH + Z\| &= \left\| \frac{1}{2} \sum_{(r,j) > (r',j')} \theta_{r,j} \theta_{r',j'} [H_j, H_{j'}] \right\| \\
 &= \left\| \frac{1}{2} \sum_{j > j'} \left(\sum_r \theta_{r,j} \theta_{r,j'} \right) [H_j, H_{j'}] + \frac{1}{2} \sum_{j,j'} \left(\sum_{r > r'} \theta_{r,j} \theta_{r',j'} \right) [H_j, H_{j'}] \right\| \\
 &= \left\| \frac{1}{2} \sum_{j > j'} \left[\sum_r (\theta_{r,j} \theta_{r,j'}) + \sum_{r > r'} (\theta_{r,j} \theta_{r',j'} - \theta_{r',j} \theta_{r,j}) \right] [H_j, H_{j'}] \right\|. \tag{A13}
 \end{aligned}$$

We dropped vanishing commutators and regathered similar terms in the last step of Eq. (A13). For the sake of clarity, we will define an abbreviation for the coefficient of the commutator terms:

$$\chi_{j,j'} := \frac{1}{2} \sum_r \theta_{r,j} \theta_{r,j'} + \frac{1}{2} \sum_{r > r'} (\theta_{r,j} \theta_{r',j'} - \theta_{r',j} \theta_{r,j}). \tag{A14}$$

After plugging in Eq. (11), that is necessary to make the linear contribution in Eq. (A12) vanish, we get

$$\chi_{j,j'} = \frac{1}{2} t^2 c_j c_{j'} - t c_{j'} \sum_r^{R-1} \theta_{r,j} + \sum_r^{R-1} \theta_{r,j} \theta_{r,j'} + \sum_{r > r'}^{R-1} \theta_{r,j} \theta_{r',j'}. \tag{A15}$$

3. Optimization potential (proof of Proposition 2)

Proof. The backwards direction is trivial as $C(\theta) \geq 0$ per definition. As for the forward direction, assume that θ^T is a global minimum of C and hence also of C^2 . The cost value at the Trotter parameters $\theta_{r,j}^T = \frac{t c_j}{R}$ reads

$$C(\theta^T)^2 = -2^{-n} \sum_{j > j'} \sum_{k > k'} \frac{t^4 c_j c_{j'} c_k c_{k'}}{4R^2} \text{Tr}([H_j, H_{j'}][H_k, H_{k'}]), \tag{A16}$$

which is nonzero in general. For a global minimum, the gradient of C at the point θ^T must vanish. The gradient of C at the Trotter solution reads

$$\partial_{q,l} C(\theta)^2 = -2^{-n+1} \sum_{j > j'} \sum_{k > k'} (\partial_{q,l} \chi_{j,j'}) \chi_{k,k'} \text{Tr}([H_j, H_{j'}][H_k, H_{k'}]). \tag{A17}$$

Here, we introduces the shorthand $\partial_{r,j} = \frac{\partial}{\partial \theta_{r,j}}$. The factor 2 in Eq. (A17) comes from the fact that the cost function is symmetric under the exchange of the factors $\chi_{j,j'}$ and $\chi_{k,k'}$. The derivative of χ reads

$$\begin{aligned}
 2\partial_{q,l} \chi_{j,j'} &= \delta_{l,j} \theta_{q,j'} + \delta_{l,j'} \theta_{q,j} + \sum_{q > r'} (\delta_{l,j} \theta_{r',j'} - \delta_{l,j'} \theta_{r',j}) + \sum_{r > q} (\theta_{r,j} \delta_{l,j'} - \theta_{r,j'} \delta_{l,j}) \\
 &= \delta_{l,j} \left(\sum_{r \leq q} \theta_{q,j'} - \sum_{r > q} \theta_{q,j'} \right) + \delta_{l,j'} \left(\sum_{r < q} \theta_{q,j} - \sum_{r \geq q} \theta_{q,j} \right) \tag{A18}
 \end{aligned}$$

and at the Trotter solution

$$2\partial_{q,l} \chi_{j,j'} = \delta_{l,j} \frac{t c_{j'}}{R} (2q - R) + \delta_{l,j'} \frac{t c_j}{R} (2(q - 1) - R). \tag{A19}$$

Finally, the gradient of the cost function reads

$$\begin{aligned}
 (\partial_{q,l} C^2) \left(\left\{ \frac{tc_j}{R} \right\}_j \right) &= -2^{-n} \sum_{j<l} \sum_{k>k'} \frac{tc_j}{R} (2q-R) \frac{t^2 c_k c_{k'}}{2R} \text{Tr}([H_l, H_j][H_k, H_{k'}]) \\
 &\quad + 2^{-n} \sum_{j>l} \sum_{k>k'} \frac{tc_j}{R} (2(q-1)-R) \frac{t^2 c_k c_{k'}}{2R} \text{Tr}([H_l, H_j][H_k, H_{k'}]) \\
 &= 2^{-n} \left((2(q-1)-R) \sum_{j>l} - (2q-R) \sum_{j<l} \right) \frac{tc_j}{R} \sum_{k>k'} \frac{t^2 c_k c_{k'}}{2R} \text{Tr}([H_l, H_j][H_k, H_{k'}]). \tag{A20}
 \end{aligned}$$

We have $\nabla C(\theta^T) = 0$ if and only if

$$\sum_{j>l} \sum_{k>k'} c_j c_k c_{k'} \text{Tr}([H_l, H_j][H_k, H_{k'}]) = \frac{2q-R}{2(q-1)-R} \sum_{j<l} \sum_{k>k'} c_j c_k c_{k'} \text{Tr}([H_l, H_j][H_k, H_{k'}]) \quad \forall q, l \tag{A21}$$

Since $R \geq 3$, the right-hand side of Eq. (A21) will have a nontrivial dependence on q but the left-hand side will not. For instance, plug in $q = 1$ and $q = R$ with arbitrary but fixed l , then the prefactor on the right-hand side becomes $\frac{R-2}{R}$ and $\frac{R}{R-2}$. As for $R \geq 3$, these are unequal, so the left-hand side of Eq. (A21) must vanish. Since l is arbitrary, also $C(\theta) = 0 \forall t$ which means Trotterization is exact. ■

4. Translational-invariant models (proof of Proposition 3)

For the following proofs, we are interested in the leading order term $E(\theta)$ in the error from the truncation to the perturbative regime. This term will become dominant by the time we have minimized $C(\theta)$. As before, in Appendix A2, we expand $\|itH + Z\|$ now up to third order and separate C from the third-order error term E using the triangle inequality:

$$\|itH + Z\| \leq C(\theta) + E(\theta) + \mathcal{O}(\theta^4). \tag{A22}$$

Using Lemma 2, we can read

$$\begin{aligned}
 E(\theta) &= \left\| \frac{1}{12} \sum_{(r,j) \neq (r',j')} \theta_{r,j} \theta_{r,j} \theta_{r',j'} [H_j, [H_j, H_{j'}]] + \frac{1}{6} \sum_{(r,j) > (u,k) > (v,l)} \theta_{r,j} \theta_{u,k} \theta_{v,l} ([H_j, [H_k, H_l]] + [H_l, [H_k, H_j]]) \right\| \\
 &= \left\| \frac{1}{12} \sum_{j \neq j'} \left(\sum_r \theta_{r,j} \theta_{r,j} \theta_{r,j'} + \sum_{r \neq r'} \theta_{r,j} \theta_{r,j} \theta_{r',j'} \right) [H_j, [H_j, H_{j'}]] + \frac{1}{6} \left(\sum_{j>k>l} \sum_r \theta_{r,j} \theta_{r,k} \theta_{r,l} \right. \right. \\
 &\quad \left. \left. + \sum_{\substack{j>k \\ l}} \sum_{r>v} \theta_{r,j} \theta_{r,k} \theta_{v,l} + \sum_{\substack{k>l \\ j}} \sum_{r>u} \theta_{r,j} \theta_{u,k} \theta_{u,l} + \sum_{j,k,l} \sum_{r>u>v} \theta_{r,j} \theta_{u,k} \theta_{v,l} \right) ([H_j, [H_k, H_l]] + [H_l, [H_k, H_j]]) \right\|, \tag{A23}
 \end{aligned}$$

where we split the sums as before using an implicit row-major ordering on the multiindex (r, j) . For the sake of simplicity, we gather all coefficients into a tensor $\Phi_{j,k,l}$ of the twofold nested commutators and define $E(\theta) = \|\sum_{j,k,l} \Phi_{j,k,l} [H_j, [H_k, H_l]]\|$.

Proof of Proposition 3. If we impose translational symmetry on the system of interest, we can split such a symmetric, p -local (p -nearest neighbor) Hamiltonian into A different interaction types that act the same way on every qubit, i.e.,

$$H = \sum_{a=1}^A \sum_{j=1}^n c_a H_a^{(j)}. \tag{A24}$$

The index (j) labels the qubit number and a the interaction type. The Hamiltonians $H_a^{(j)}$ have nontrivial support on the p qubits starting with j . In one dimension, these are $j, j + 1, \dots, j + p - 1$. For higher dimensions, different interaction types may admit support on clusters of at most p neighboring qubits. We do not make any assumptions on lattice geometry or dimension in the following and formally denote by $\mathcal{N}_p(j)$ the set of Hamiltonian terms that correspond to a p -nearest neighbor of the j th term. The parameters are chosen to be translational invariant by just dropping the dependence on the qubit number, i.e., $\theta_{r,j} = \theta_{r,a}$. As a consequence, also $\chi_{j,j'} = \chi_{a,a'}$ and $\Phi_{j,k,l} = \Phi_{a,b,c}$.

Since we introduced another index $j \rightarrow (a, j)$, we need to refine the chosen ordering. Without loss of generality, we choose

$$(r, a, j) : (1, 1, 1) \leq (1, 1, j) \leq (1, a, j) \leq (r, a, j). \tag{A25}$$

To simplify the cost function defined in Eq. (12), we decompose the sum

$$\sum_{(a,j)>(a',j')} = \sum_{\substack{j>j' \\ a=a'}} + \sum_{a>a'} \sum_{j,j'} \tag{A26}$$

We plug in the above structure into the definition of the cost function, and use the locality of the Hamiltonian to get

$$\begin{aligned} C(\theta) &= \left\| \sum_{j>j'} \chi_{a,a} [H_a^{(j)}, H_a^{(j')}] + \sum_{a>a'} \sum_{j,j'} \chi_{a,a'} [H_a^{(j)}, H_a^{(j')}] \right\| \\ &= \left\| \sum_j \left[\sum_{\substack{j' \in \mathcal{N}_p(j) \\ j'>j}} \chi_{a,a} [H_a^{(j')}, H_a^{(j)}] + \sum_{a>a'} \sum_{j' \in \mathcal{N}_p(j)} \chi_{a,a'} [H_a^{(j)}, H_a^{(j')}] \right] \right\| \\ &\leq n \left\| \sum_{\substack{j' \in \mathcal{N}_p(j) \\ j'>j}} \chi_{a,a} [H_a^{(j')}, H_a^{(j)}] + \sum_{a>a'} \sum_{j' \in \mathcal{N}_p(j)} \chi_{a,a'} [H_a^{(j)}, H_a^{(j')}] \right\| \text{ for one fixed } j \\ &=: nC_{\text{unit}}(\theta), \end{aligned} \tag{A27}$$

where we used the triangle inequality in the last step. From translational invariance, we know that the second to last term yields the same value for every j yielding a factor n . If the commutators $[H_j, H_{j'}]$ are orthogonal with respect to the Hilbert-Schmidt product, the sum \sum_j can be pulled out of the squared norm yielding n equal addends and hence $C(\theta) = \sqrt{n}C_{\text{unit}}(\theta)$.

The same way, we can reduce the computation of $E(\theta)$ to a unit cell. However, as we consider twofold nested commutators here, the unit cell is of size $(3p - 2)^D$:

$$\begin{aligned} E(\theta) &= \left\| \sum_j \left[\sum_{a,b,c} \sum_{k \in \mathcal{N}_p(j)} \sum_{l \in \mathcal{N}_p(k)} \Phi_{a,b,c} [H_a^{(j)}, [H_b^{(k)}, H_c^{(l)}]] \right] \right\| \\ &\leq nE_{\text{unit}}(\theta) \end{aligned} \tag{A28}$$

Again, if the commutators $[H_a^{(j)}, [H_b^{(k)}, H_c^{(l)}]]$ are orthogonal with respect to the Hilbert-Schmidt inner product, we get $E(\theta) = \sqrt{n}E_{\text{unit}}(\theta)$. ■

5. Long simulation times (proof of Proposition 4)

Proof. As Eq. (17) is equivalent to the ansatz before by just mapping $R \rightarrow KR, t \rightarrow Kt$ and identifying $\theta_{r,j} = \theta_{r \% R, j}$, the cost function will have the same form as before with modified χ ,

$$\begin{aligned} \chi_{j,j'}^{(KR)} &= \frac{1}{2} \sum_{r=1}^{KR} (\theta_{r,j} \theta_{r,j'}) + \frac{1}{2} \sum_{r>r'}^{KR} (\theta_{r,j} \theta_{r',j'} - \theta_{r,j'} \theta_{r',j}) \\ &= \frac{K}{2} \sum_{r=1}^R (\theta_{r,j} \theta_{r,j'}) + \frac{K}{2} \sum_{r>r'}^R (\theta_{r,j} \theta_{r',j'} - \theta_{r,j'} \theta_{r',j}) = K\chi_{j,j'}^{(R)}, \end{aligned} \tag{A29}$$

where we denoted the number of layers as a superscript $\chi^{(R)}$ to better differentiate the two coefficients. The factor K comes from the identification of $\theta_{r,j} = \theta_{r \% R, j}$. While the first factor is trivial, the second comes from splitting the antisymmetric sum into

$$\sum_{r>r'}^{KR} = \frac{K(K+1)}{2} \sum_{r>r'}^R + \frac{(K-1)K}{2} \sum_{r<r'}^R = \frac{K(K+1)}{2} \sum_{r>r'}^R - \frac{(K-1)K}{2} \sum_{r>r'}^R = K \sum_{r>r'}^R. \tag{A30}$$

In the same way, we can plug in the duplicated parameters θ^K into $E(\theta)$. We get the following identities: ■

$$\sum_r^{KR} = K \sum_r^R \quad \text{and} \quad \sum_{r \neq r'}^{KR} = K^2 \sum_{r \neq r'}^R + K(K-1) \sum_{r=r'}^R, \tag{A31}$$

which yields a global factor K^2 for the first sum $\sum_{j \neq j'}$ in Eq. (A23). For the second sum, unfortunately we do not have a common factor to pull out. We have

$$\begin{aligned} \sum_{r>u}^{KR} &= \frac{K(K+1)}{2} \sum_{r>u}^R + \frac{K(K-1)}{2} \sum_{r \leq u}^R = \sum_{r>u}^R + \frac{K(K-1)}{2} \sum_{r,u}^R \\ \sum_{r>u>v}^{KR} &= K \sum_{r>u>v}^R + \frac{K(K-1)(K+1)}{6} \sum_r^R \sum_{u>v}^R + \frac{K(K-1)(K-2)}{6} \sum_r^R \sum_{u \leq v}^R \\ &= K \sum_{r>u>v}^R + \frac{K(K-1)}{2} \sum_r^R \sum_{u>v}^R. \end{aligned} \tag{A32}$$

In the last step, we used that the sum over $\sum_{r,u,v}$ vanishes because of antisymmetry of the commutator. While some of the above contributions multiply the error before duplication by a factor, others are nonzero extra terms. We thus get an upper bound

$$E(\theta^K) \leq KE(\theta) + \frac{K(K-1)}{2} \Lambda, \tag{A33}$$

involving the rest term

$$\Lambda = \left\| \left(\sum_l \sum_{j>k} - \sum_l \sum_{j>k} - 2 \sum_l \sum_{j \leq k} \right) \theta_{rj} \theta_{uk} \theta_{vl} ([H_j, [H_k, H_l]] + [H_k, [H_j, H_l]]) + 3 \sum_{j,k,l} \sum_{r>u>v} \theta_{rj} \theta_{uk} \theta_{vl} ([H_k, [H_j, H_l]]) \right\|. \tag{A34}$$

APPENDIX B: PERTURBATIVE DISTANCE OF A TRANSVERSE FIELD ISING MODEL

We apply the presented method to a general transverse field Ising model (TFIM) with full connectivity, which is described by the Hamiltonian

$$H_{\text{TFIM}} = - \sum_{\mu, \nu} J_{\mu\nu} Z_{\mu} Z_{\nu} + \sum_{\mu} h_{\mu} X_{\mu}. \tag{B1}$$

The sum over μ, ν counts every pair of qubits and the symmetric matrix $J_{\mu\nu}$ encodes the interaction strengths between qubits μ and ν . To fix a timescale, we normalize the Hamiltonian to $\|H\| \approx \sqrt{n}$. In total, the Hamiltonian admits $\frac{n(n-1)}{2}$ ZZ terms and n X terms. To use the notation introduced before, one needs to map the interactions between two qubits to a single index j :

$$\begin{aligned} H_j &= \begin{cases} Z_{\mu(j)} Z_{\nu(j)} & \text{for } j \leq \frac{n(n-1)}{2} \\ X_{\mu(j)} & \text{for } \frac{n(n-1)}{2} < j \leq \frac{n(n+1)}{2} \end{cases} \\ c_j &= \begin{cases} J_{\mu(j)\nu(j)} & \text{for } j \leq \frac{n(n-1)}{2} \\ h_{\mu(j)} & \text{for } \frac{n(n-1)}{2} < j \leq \frac{n(n+1)}{2}. \end{cases} \end{aligned} \tag{B2}$$

The exact ordering of the indices is not important, in general. To give an explicit example, consider row major ordering $j = (\mu - 1)(n - \frac{n}{2}) + \nu$, for $\mu \leq \nu$. We plug H_j into Eq. (12) and observe that the only nonzero commutator terms are of the form $Z_{\mu} Y_{\nu}$. To also have a nonvanishing trace (Pauli-string squares to $\mathbb{1}$), both commutators need to match, i.e.,

$$\begin{aligned} \text{Tr}([H_j, H_{j'}][H_k, H_{k'}]) &= -2^{n+2} \delta_{j,k} \delta_{j',k'} (\delta_{\mu(j),\mu(j')} + \delta_{\mu(j),\nu(j')}) \\ &\quad \Theta_{j - \frac{n(n-1)}{2}} \Theta_{\frac{n(n-1)}{2} + 1 - j'}, \end{aligned} \tag{B3}$$

where we used the implicit one-to-one mapping between indices $j \in \{1, \dots, \frac{n(n+1)}{2}\}$ and $\mu, \nu \in \{1, \dots, n\}$, as well as the

Kronecker delta and the discretized Heavyside step function:

$$\delta_{j,k} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{else} \end{cases} \quad \Theta_x = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{else} \end{cases}. \tag{B4}$$

The Heavyside factors ensure that only ZZ and X terms are composed in the commutator. The Kronecker deltas in the bracket account for the two cases where X can hit the first or second qubit of ZZ. The factors $\delta_{j,k} \delta_{j',k'}$ ensure that the Pauli strings match yielding a nonvanishing trace. In total, we get

$$\begin{aligned} C(\theta)^2 &= 4 \sum_{j>j'} \sum_{k>k'} \chi_{j,j'} \chi_{k,k'} \delta_{j,k} \delta_{j',k'} (\delta_{\mu(j),\mu(j')} + \delta_{\mu(j),\nu(j')}) \\ &\quad \times \Theta_{j - \frac{n(n-1)}{2}} \Theta_{\frac{n(n-1)}{2} + 1 - j'} \\ &= 4 \sum_{j=M_z+1}^{M_z+n} \sum_{j'=1}^{M_z} \chi_{j,j'}^2 (\delta_{\mu(j),\mu(j')} + \delta_{\mu(j),\nu(j')}) \\ &= 4 \sum_{\mu<\nu}^n (\chi_{j(\mu),j(\mu,\nu)}^2 + \chi_{j(\nu),j(\mu,\nu)}^2), \end{aligned} \tag{B5}$$

where we shorthanded the number of ZZ interactions as $M_z = \frac{n(n-1)}{2}$. Note that every addend in Eq. (B5) is positive, so minimizing $C(\theta)$ reduces to the simultaneous minimization of each addend. However, the equations $\chi_{j,j'}^2 = 0$ do not decouple in general, so there is no exact formula for the roots of Eq. (B5).

1. Two-dimensional square lattice

Physical models typically yield local interactions. Let us reduce the above discussion to a two-dimensional TFIM that only includes nearest-neighbor interactions. This is straightforwardly done by imposing

$$J_{\mu\nu} = 0 \quad \text{and} \quad \theta_{r,j(\mu,\nu)} = 0 \quad \text{if } \mu \approx \nu, \tag{B6}$$

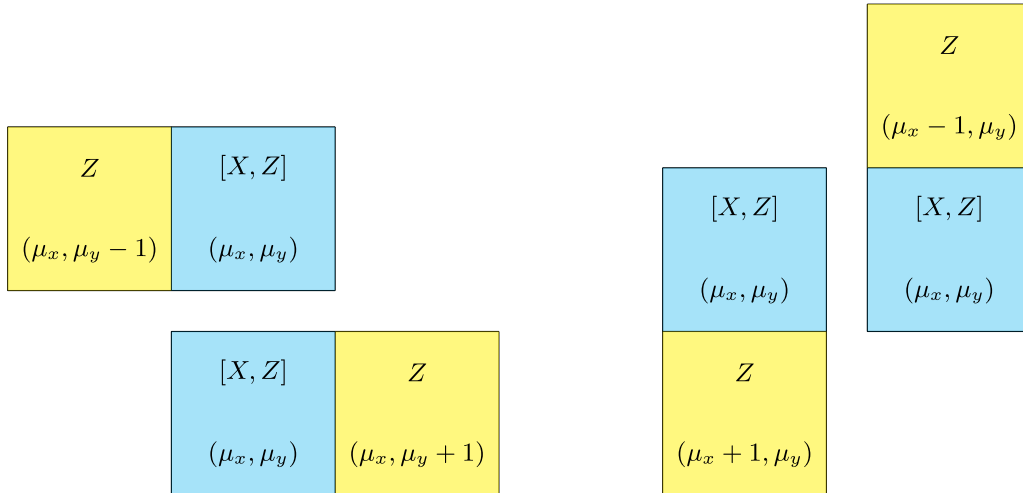


FIG. 4. Visualization of the perturbative distance for a two-dimensional nearest-neighbor Ising model. Instead of defining cumbersome notation as in Eq. (B7), it is often simpler to think about the commutator structure in this pictorial manner.

where $\mu \approx \nu$ denotes that μ and ν are not neighbors. On a two-dimensional square lattice, it is handy to further refine the qubit indices into two $\mu = (\mu_x, \mu_y)$, where $\mu_x \in \{1, \dots, n_x\}$ and $\mu_y \in \{1, \dots, n_y\}$ count the lattice sites in each dimension, respectively. $\theta_{r,j} = 0$ also implies $\chi_{j,j'} = 0 \forall j'$ which further reduces the number of terms in Eq. (B5):

$$\begin{aligned}
 C(\theta) = 4 \sum_{\mu=(1,1)}^{(n_x, n_y)} & \left(\chi_{j(\mu), j(\mu, (\mu_x, \mu_y + 1))}^2 \right. \\
 & + \chi_{j(\mu, (\mu_x, \mu_y + 1)), j(\mu, (\mu_x, \mu_y + 1))}^2 + \chi_{j(\mu), j(\mu, (\mu_x + 1, \mu_y))}^2 \\
 & \left. + \chi_{j((\mu_x + 1, \mu_y)), j(\mu, (\mu_x + 1, \mu_y))}^2 \right). \tag{B7}
 \end{aligned}$$

Although the cost function becomes tedious to write, it only contains a linear number of terms and is thus efficiently computable. The four contributions in Eq. (B5) account for the different combinations of the appearing ZY string and are depicted in Fig. 4.

2. Numerical results

In Fig. 5, we study the performance of the derived cost function by optimizing a variational sequence with the second-order cost function derived in Eq. (B7) and comparing it to the exact value of the difference of time evolutions ϵ for different times t . The optimizer manages to find small values of $C(\theta)$ for all times. For small enough times $t < \frac{R}{\|H\|} \approx 1$, the perturbative cost function also incorporates a faithful indicator for the exact norm in a sense that the optimal parameters also yield a significant decrease in the exact error of the variational sequence.

APPENDIX C: PERTURBATIVE DISTANCE OF AN XY MODEL IN A TRANSVERSE FIELD

In total, the XY Hamiltonian from Eq. (23) admits $\frac{n(n-1)}{2}$ YY terms, $\frac{n(n-1)}{2}$ ZZ terms, and nX terms. To use the notation introduced before, one needs to map the interactions between

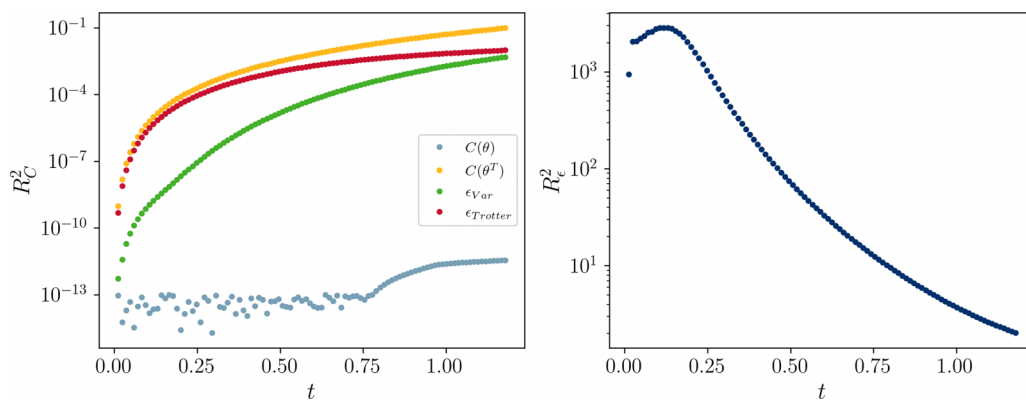


FIG. 5. Perturbative (C) and exact (ϵ) error measures (left) and exact error ratio R_ϵ and the ratio of error estimates R_E (right) as defined in Eq. (21) evaluated on a nearest-neighbor Ising model on a 3×3 quadratic lattice and random interaction strength $J_{\mu,\nu}$ centered around 1 and $h_\mu = 0.25$.

two qubits to a single index j :

$$H_j = \begin{cases} X_{\mu(j)} & \text{for } j \leq n \\ Y_{\mu(j)}Y_{\nu(j)} & \text{for } n < j \leq \frac{n(n+1)}{2} \\ Z_{\mu(j)}Z_{\nu(j)} & \text{for } \frac{n(n+1)}{2} < j \leq n^2, \end{cases} \quad c_j = \begin{cases} h_{\mu(j)} & \text{for } j \leq n \\ J_{\mu(j)\nu(j)}^{(y)} & \text{for } n < j \leq \frac{n(n+1)}{2} \\ J_{\mu(j)\nu(j)}^{(z)} & \text{for } \frac{n(n+1)}{2} < j \leq n^2. \end{cases} \quad (\text{C1})$$

The exact ordering of the indices is not important, in general. To give an explicit example, consider row major ordering $j = (\mu - 1)(n - \frac{\mu}{2}) + \nu$ for $\mu \leq \nu$. We plug H_j into Eq. (12) and observe that the only nonzero commutator terms are of the form $Z_\mu Y_\nu$ and $Y_\mu X_\nu Z_\sigma$. To also have a nonvanishing trace (Pauli-string squares to $\mathbb{1}$), both commutators always need to match. While this is always given if $j = k$ and $j' = k'$, now we can also have the case $[Y_\mu Y_\nu, X_\mu] = -[Z_\mu Z_\nu, X_\nu] = -2iZ_\mu Y_\nu$,

$$C(\theta)^2 = 4 \sum_{\mu < \nu}^n (\chi_{j_y(\mu, \nu), j(\mu)}^2 + \chi_{j_y(\mu, \nu), j(\nu)}^2 + \chi_{j_z(\mu, \nu), j(\mu)}^2 + \chi_{j_z(\mu, \nu), j(\nu)}^2 - \chi_{j_z(\mu, \nu), j(\mu)} \chi_{j_y(\mu, \nu), j(\nu)} - \chi_{j_z(\mu, \nu), j(\nu)} \chi_{j_y(\mu, \nu), j(\mu)}) \\ + 4 \sum_{\mu < \nu < \sigma}^n (\chi_{j_z(\mu, \nu), j_y(\nu, \sigma)}^2 + \chi_{j_z(\mu, \sigma), j_y(\nu, \sigma)}^2 + \chi_{j_z(\nu, \sigma), j_y(\mu, \nu)}^2 + \chi_{j_z(\nu, \sigma), j_y(\mu, \sigma)}^2 + \chi_{j_z(\mu, \nu), j_y(\mu, \sigma)}^2 + \chi_{j_z(\mu, \sigma), j_y(\mu, \nu)}^2), \quad (\text{C2})$$

where we mapped back the lattice indices μ, ν to the single index j by reverting the mapping for all cases in Eq. (C1). The terms in the first line of Eq. (C2) come from commutator terms of the form YZ and the terms in the second line from terms of the form XYZ.

1. Two-dimensional square lattice

If we only admit nearest-neighbor interaction and set all other $J_{\mu\nu}^{(y,z)} = 0$, we get

$$C(\theta)^2 = 4 \sum_{\mu=(1,1)}^{(n_x, n_y)} (\chi_{j_y(\mu, (\mu_x, \mu_y+1)), j(\mu)}^2 + \chi_{j_y(\mu, (\mu_x, \mu_y+1)), j((\mu_x, \mu_y+1))}^2 + \chi_{j_y(\mu, (\mu_x+1, \mu_y)), j(\mu)}^2 + \chi_{j_y(\mu, (\mu_x+1, \mu_y)), j((\mu_x+1, \mu_y))}^2 \\ + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j(\mu)}^2 + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j((\mu_x, \mu_y+1))}^2 + \chi_{j_z(\mu, (\mu_x+1, \mu_y)), j(\mu)}^2 + \chi_{j_z(\mu, (\mu_x+1, \mu_y)), j((\mu_x+1, \mu_y))}^2 \\ - \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j(\mu)} \chi_{j_y(\mu, (\mu_x, \mu_y+1)), j((\mu_x, \mu_y+1))} - \chi_{j_z(\mu, (\mu_x+1, \mu_y)), j(\mu)} \chi_{j_y(\mu, (\mu_x+1, \mu_y)), j((\mu_x+1, \mu_y))} \\ - \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j((\mu_x, \mu_y+1))} \chi_{j_y(\mu, (\mu_x, \mu_y+1)), j(\mu)} - \chi_{j_z(\mu, (\mu_x+1, \mu_y)), j((\mu_x+1, \mu_y))} \chi_{j_y(\mu, (\mu_x+1, \mu_y)), j(\mu)} \\ + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j_y((\mu_x, \mu_y+1), (\mu_x, \mu_y+2))}^2 + \chi_{j_z(\mu, (\mu_x+1, \mu_y)), j_y((\mu_x+1, \mu_y), (\mu_x+2, \mu_y))}^2 \\ + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j_y(\mu, (\mu_x+1, \mu_y))}^2 + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j_y((\mu_x, \mu_y+1), (\mu_x+1, \mu_y+1))}^2 \\ + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j_y((\mu_x-1, \mu_y), \mu)}^2 + \chi_{j_z(\mu, (\mu_x, \mu_y+1)), j_y((\mu_x-1, \mu_y+1), (\mu_x, \mu_y+1))}^2 \\ + \chi_{j_z((\mu_x, \mu_y+1), (\mu_x, \mu_y+2)), j_y(\mu, (\mu_x, \mu_y+1))}^2 + \chi_{j_z((\mu_x+1, \mu_y), (\mu_x+2, \mu_y)), j_y(\mu, (\mu_x+1, \mu_y))}^2 \\ + \chi_{j_z(\mu, (\mu_x+1, \mu_y)), j_y(\mu, (\mu_x, \mu_y+1))}^2 + \chi_{j_z((\mu_x, \mu_y+1), (\mu_x+1, \mu_y+1)), j_y(\mu, (\mu_x, \mu_y+1))}^2 \\ + \chi_{j_z((\mu_x-1, \mu_y), \mu), j_y(\mu, (\mu_x, \mu_y+1))}^2 + \chi_{j_z((\mu_x-1, \mu_y+1), (\mu_x, \mu_y+1)), j_y(\mu, (\mu_x, \mu_y+1))}^2).$$

Although being tedious to write, Eq. (C3) admits a linear number of addends that follow simple systematics. The first two lines gather the 2D-analogs of terms 1–4 in Eq. (C2) and lines 3 and 4 gather the former terms 5 and 6. The remaining terms are gathered in lines 5–10 and resemble every combination of nontrivial three-body commutators.

2. More parameters

We explore the behavior under changing depth R within a single time step, which also alters the number of parameters to be optimized. In Fig. 6, we show a numerical experiment analog to Fig. 2, but using $R = 4$ and $R = 6$. As with increasing R ,

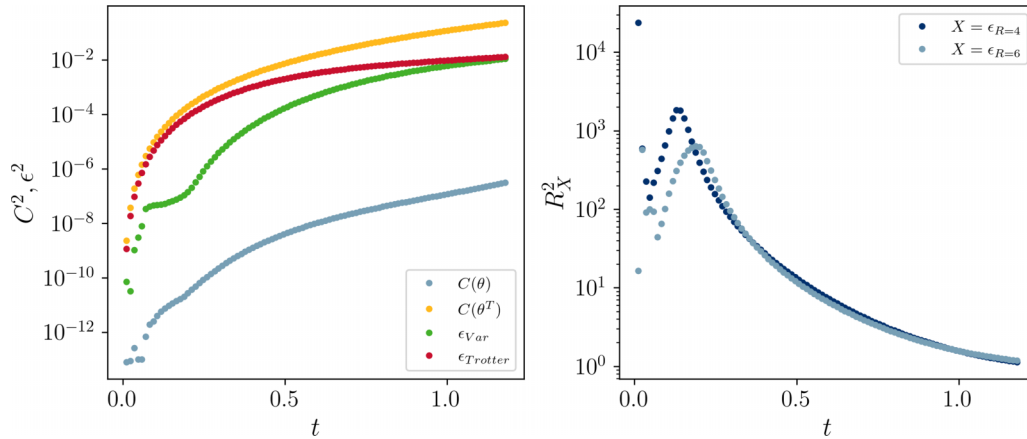


FIG. 6. Perturbative (C) and exact (ϵ) error measures (left) and exact error ratio R_ϵ (right) for $R = 4$ and $R = 6$ evaluated on a nearest-neighbor XY model on a 3×3 quadratic lattice and random interaction strength $J_{\mu,\nu}^{(y)}$ centered around 0.5, $J_{\mu,\nu}^{(z)}$ centered around 1 and $h_\mu = 0.25$. On the left panel, results for $R = 6$ are plotted.

$\frac{t}{R}$ decreases and the Trotter errors that we compare with get smaller. The optimizer still manages to find solutions that yield a comparable improvement ratio as seen in Fig. 2. For very small times, the improvement ratios become much smaller, since the cost function C , as well as the error ϵ , approach numerical accuracy.

-
- [1] A. Smith, M. S. Kim, F. Pollmann, and J. Knolle, Simulating quantum many-body dynamics on a current digital quantum computer, *npj Quantum Inf.* **5**, 106 (2019).
- [2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [3] M. Suzuki, Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics, *J. Math. Phys.* **26**, 601 (1985).
- [4] H. Yoshida, Construction of higher order symplectic integrators, *Phys. Lett. A* **150**, 262 (1990).
- [5] T. Barthel and Y. Zhang, Optimized Lie–Trotter–Suzuki decompositions for two and three non-commuting terms, *Ann. Phys.* **418**, 168165 (2020).
- [6] J. Ostmeier, Optimised Trotter decompositions for classical and quantum computing, *J. Phys. A: Math. Theor.* **56**, 285303 (2023).
- [7] G. H. Low and I. L. Chuang, Optimal Hamiltonian simulation by quantum signal processing, *Phys. Rev. Lett.* **118**, 010501 (2017).
- [8] K. Bharti and T. Haug, Quantum-assisted simulator, *Phys. Rev. A* **104**, 042418 (2021).
- [9] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter error with commutator scaling, *Phys. Rev. X* **11**, 011020 (2021).
- [10] A. Schubert and C. B. Mendl, Trotter error with commutator scaling for the Fermi-Hubbard model, *arXiv:2306.10603*.
- [11] B. D. M. Jones, G. O. O’Brien, D. R. White, E. T. Campbell, and J. A. Clark, Optimising Trotter–Suzuki decompositions for quantum simulation using evolutionary strategies, *arXiv:1904.01336* [cs.NE].
- [12] S. Barison, F. Vicentini, and G. Carleo, An efficient quantum algorithm for the time evolution of parameterized circuits, *Quantum* **5**, 512 (2021).
- [13] M. Benedetti, M. Fiorentini, and M. Lubasch, Hardware-efficient variational quantum algorithms for time evolution, *Phys. Rev. Res.* **3**, 033083 (2021).
- [14] C. Cirstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, *npj Quantum Inf.* **6**, 82 (2020).
- [15] J. Gibbs, K. Gili, Z. Holmes, B. Commeau, A. Arrasmith, L. Cincio, P. J. Coles, and A. Sornborger, Long-time simulations for fixed input states on quantum hardware, *npj Quantum Inf.* **8**, 135 (2022).
- [16] M. S. J. Tepaske, D. Hahn, and D. J. Luitz, Optimal compression of quantum many-body time evolution operators into brickwall circuits, *SciPost Phys.* **14**, 073 (2023).
- [17] H. Zhao, M. Bukov, M. Heyl, and R. Moessner, Making Trotterization adaptive and energy-self-correcting for NISQ devices and beyond, *PRX Quantum* **4**, 030319 (2023).
- [18] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [19] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [20] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).
- [21] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, *Nat. Commun.* **12**, 6961 (2021).

- [22] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [23] Z. Cai, Resource estimation for quantum variational simulations of the Hubbard model, *Phys. Rev. Appl.* **14**, 014059 (2020).
- [24] D. S. França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, *Nat. Phys.* **17**, 1221 (2021).
- [25] R. Mansuroglu, T. Eckstein, L. Nützel, S. A. Wilkinson, and M. J. Hartmann, Variational Hamiltonian simulation for translational invariant systems via classical pre-processing, *Quantum Sci. Technol.* **8**, 025006 (2023).
- [26] M. Streif and M. Leib, Training the quantum approximate optimization algorithm without access to a quantum processing unit, *Quantum Sci. Technol.* **5**, 034008 (2020).
- [27] C. M. Keever and M. Lubasch, Classically optimized Hamiltonian simulation, *Phys. Rev. Res.* **5**, 023146 (2023).
- [28] E. Cervero Martín, K. Plekhanov, and M. Lubasch, Barren plateaus in quantum tensor network optimization, *Quantum* **7**, 974 (2023).
- [29] A. Kotil, R. Banerjee, Q. Huang, and C. B. Mendl, Riemannian quantum circuit optimization for Hamiltonian simulation, [arXiv:2212.07556](https://arxiv.org/abs/2212.07556).
- [30] J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, Quantum algorithm for simulating real time evolution of lattice Hamiltonians, *SIAM J. Comput. FOCS18* (2021).
- [31] S. Chakraborty, A. Gilyén, and S. Jeffery, The power of block-encoded matrix powers: Improved regression techniques via faster Hamiltonian simulation, in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, Leibniz International Proceedings in Informatics (LIPIcs), edited by C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi (Schloss Dagstuhl–Leibniz Center for Informatics, Dagstuhl, Germany, 2019), Vol. 132, pp. 33:1–33:14.