

Exponentially Reduced Circuit Depths Using Trotter Error Mitigation

James D. Watson^{1,2} and Jacob Watkins^{3,4,*}

¹*Joint Center for Quantum Information & Computer Science, National Institute of Standards and Technology and University of Maryland, USA*

²*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, USA*

³*Facility for Rare Isotope Beams, Michigan State University, USA*

⁴*Global Technology Applied Research, JPMorgan Chase, USA*

 (Received 24 October 2024; revised 21 May 2025; accepted 2 July 2025; published 12 August 2025)

Product formulas are a popular class of digital quantum simulation algorithms due to their conceptual simplicity, low overhead, and performance, which often exceeds theoretical expectations. Recently, Richardson extrapolation and polynomial interpolation have been proposed to mitigate the Trotter error incurred by the use of these formulas. This work provides a rigorous, general analysis of these techniques for computing time-evolved observables, simplifying the interpolation algorithm in the process, and shows that extrapolation generically improves the performance of product formulas for this task. We demonstrate that, to achieve error ϵ in a simulation of time T using a p th-order product formula with extrapolation, circuit depths of $O(T^{1+1/p} \text{polylog}(1/\epsilon))$ are sufficient—an exponential improvement in the precision over product formulas alone. Furthermore, we prove that these algorithms achieve commutator scaling, and improve the T complexity for the interpolation algorithm. By relaxing the requirement of performing exact Chebyshev interpolation, our simplified algorithm eliminates the need for fractional implementations of Trotter steps, reducing computational overhead. Finally, we show these techniques can be combined with the classical shadows method to estimate many time-evolved local observables. Taken together, our findings provide the strongest evidence yet for the utility of Trotter error-mitigation techniques in algorithmic applications.

DOI: [10.1103/kw39-yxq5](https://doi.org/10.1103/kw39-yxq5)

I. INTRODUCTION

Quantum simulation—the task of computing dynamical properties of a quantum system—has been an early inspiration and impetus for quantum computing, and is among the most promising candidates for near-term quantum advantage. Scientific domains such as quantum chemistry, nuclear physics, materials science, and high-energy physics stand to benefit from robust, programmable quantum devices, which can implement a quantum Hamiltonian of interest [1–5]. Beyond the study of quantum phenomena, Hamiltonian simulation forms a key component to more generic routines, such as linear system solvers [6,7], and even the investigation of nonquantum phenomena [8,9].

Hamiltonian simulation is, by now, a relatively mature subfield of quantum computing. Significant attention has been given to all steps of the simulation procedure: qubit mappings [10–12], state preparation [13], time evolution [14–16], and measurement [17,18]. Time evolution has received particular attention as it is often the most expensive step in a full routine. Roughly speaking, there currently exist four families of time-evolution algorithms: product formulas (also known as Trotterization) [16,19], linear combination of unitaries (LCU) [20–23], quantum walks [24], and qubitization [15]. Despite the inferior asymptotic scaling of product formulas, they have many desirable properties, including conceptual and practical simplicity, lack of auxiliary qubits, natural incorporation of Lieb-Robinson bounds, commutator scaling, and the tendency to conserve desirable properties and symmetries of the Hamiltonian [19,25–29]. Many of these desirable properties are believed not to hold for post-Trotter methods in general [30]. Remarkably, the empirical performance of product formulas is often comparable to the LCU and qubitization methods in numerical studies [31,32], and far better than leading error bounds predict [33]. These studies

*Contact author: jacob.a.watkins@jpmchase.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

indicate a gap in our theoretical understanding of expected Trotter error.

While the unexpectedly high empirical performance of product formulas is fortunate, the current constraints on quantum hardware still make simulating interesting Hamiltonians out of reach, motivating the search for improved Trotter-based approaches that do not significantly increase quantum resources. A wide range of product formula algorithms have been suggested to optimize performance according to properties of the Hamiltonian under consideration [34–41] and by optimizing the circuit itself [42]. Additional progress has been made by recognizing that time evolution is not a full algorithm, but ultimately a subroutine embedded within a measurement protocol. With a suitable choice of measurements, the resulting classical data can be processed to improve results without going beyond product formulas. This amounts to algorithmic error mitigation, which is conceptually similar to the hardware error mitigation being developed and employed on current devices.

Several Trotter mitigation approaches have been proposed, including Richardson extrapolation [43,44], polynomial interpolation [45], and parametric matrix models [46]. Most of these results lack rigorous performance guarantees, and existing guarantees fail to capture expected properties such as “commutator scaling,” whereby the algorithmic cost is dependent on the Hamiltonian only through nested commutators of its terms (see Sec. III A for a detailed discussion of prior work). This lack of theoretical grounding may be surprising given that the Richardson extrapolation and polynomial interpolation techniques are remarkably simple and well-studied numerical techniques that are employed in various settings.

This work makes several advances in the study of Trotter error mitigation for quantum simulation. First, we provide a fully specified Richardson extrapolation algorithm with rigorous performance guarantees, utilizing well-conditioned schemes developed for multiproduct formulas [23]. Second, we improve upon the polynomial interpolation algorithm of Ref. [45] by drastically improving the error bounds and eliminating the need for fractional evolutions of Trotter formulas, which required additional quantum overhead and quantum signal processing. Finally, we conduct a rigorous and unified performance analysis of these algorithms for time-evolved expectation values. While these algorithms achieve the expected gains in precision, we also prove that they exhibit commutator scaling and $O(T^{1+1/p})$ scaling with the simulation time for a p th-order formula, both of which are not achieved by any previous literature. For the first time, our analysis considers both “coherent” and “incoherent” measurement protocols for acquiring the expectation values for various Trotter step sizes, and quantifies both the maximum circuit depth and total cost. On the way to these results, we provide a concrete analysis of the Trotter errors

associated with Heisenberg evolution of observables; in particular, we derive an explicit error series with guaranteed neighborhoods of convergence. These results are likely to be of interest beyond the applications considered presently.

We now briefly highlight some of these advances, which are discussed at greater length in Sec. III. For the short-depth method, we show that circuit depths of

$$O((\lambda T)^{1+1/p} \text{polylog}(1/\epsilon))$$

are sufficient, where λ is a factor depending on commutators of terms in the Hamiltonian, and where p is the order of the product formula. For the method optimizing total resources, we show that the overall number of Trotter steps needed scales as

$$O\left(\frac{(\lambda T)^{1+1/p}}{\epsilon} \text{polylog}(1/\epsilon)\right).$$

Essential to achieving these results is the fact that our extrapolation approaches are well conditioned, such that small errors in the data do not rapidly accumulate. We show that two methods of performing the extrapolation—Richardson extrapolation and polynomial interpolation—obtain the same asymptotic performance up to logarithmic factors. Crucially, the algorithms require no additional control gates or ancillary qubits compared to nonextrapolated product formulas, unlike the original polynomial algorithm. The short depths and reduced gate counts of extrapolated product formulas techniques should be beneficial in both the NISQ and fault-tolerant regimes. Moreover, we show the techniques can be naturally combined with classical shadows to estimate M local observable expectation values at once with only $\log(M)$ overhead. This shadow extrapolation essentially allows us to reconstruct a classical shadow representation of the final state of the idealized time evolution.

The rest of this paper is outlined as follows. In Sec. II, we provide the background on product formulas, variation of parameters, extrapolation techniques, and measurement protocols. The main results are stated in Sec. III. Section IV contains the technical error analysis of time-evolved observables under product formulas, which we then use to inform the complexity analyses of Richardson extrapolation (Sec. V) and polynomial interpolation (Sec. VI). Section VII explores additional applications for our findings and illustrations of the improvements for concrete models. Section VIII presents an algorithm combining classical shadows with extrapolation in order to estimate many time-evolved observables. Finally, we provide some discussion and concluding remarks in Sec. IX.

II. SUMMARY OF METHODS

In terms of the algorithms considered in this work, the primary elements are product formulas, amplitude measurement, and two extrapolation techniques: Richardson and polynomial. For the theoretical analysis, we rely heavily on variation of parameters from the theory of first-order ordinary differential equations. This section briefly reviews all of these components to help the reader understand the main results and subsequent proofs. Those interested only in statements of results are welcome to skip to Sec. III.

A. Product formulas

Consider a (time-independent) Hamiltonian H expressed as a sum of Γ terms

$$H = \sum_{\gamma=1}^{\Gamma} H_{\gamma} \quad (1)$$

for Hermitian H_{γ} . Product formulas are splittings of the exponential e^{-iHt} , the time-evolution operator, along the various terms H_{γ} . For example, the simplest product formula, namely first-order Trotter, is defined by

$$\mathcal{P}_1(t) := \prod_{\gamma=1}^{\Gamma} e^{-iH_{\gamma}t}, \quad (2)$$

where, by convention, we take the product going right to left. The utility of product formulas arises from the fact that there often exists decompositions of H into terms H_{γ} such that each exponential $e^{-iH_{\gamma}t}$ may be computed efficiently. For example, H_{γ} may be k local, or H may be sparse and thus decomposable into one-sparse terms.

Product formulas are meant to approximate the exact time-evolution operator for short times t . Larger times can be approximated to arbitrary precision by breaking the simulation time $T \in \mathbb{R}$ into sufficiently many steps. A product formula $\mathcal{P}$ is said to be *order* p if

$$\mathcal{P}(t) - e^{-iHt} = O(t^{p+1}) \quad (3)$$

for small t . Thus, for time $T \in \mathbb{R}$ and $r \in \mathbb{Z}_+$,

$$e^{-iHT} - \mathcal{P}(T/r)^r = O(T^{p+1}/r^p). \quad (4)$$

In this work, we only consider formulas of order at least 1. The order is roughly a proxy for accuracy, although large constant factors typically negate the advantage of high-order formulas in typical instances. There exist product formulas of arbitrarily large order, the most well-known family of which is the Trotter-Suzuki formulas S_{2k} of order $p = 2k$. These are defined recursively as follows.

For $k = 1$,

$$S_2(t) := \prod_{\gamma=\Gamma}^1 e^{-iH_{\gamma}t/2} \prod_{\gamma=1}^{\Gamma} e^{-iH_{\gamma}t/2}$$

and for $k \in \mathbb{Z}_+$ greater than 1,

$$S_{2k}(t) := [S_{2(k-1)}(u_k t)]^2 S_{2(k-1)}((1 - 4u_k)t) [S_{2(k-1)}(u_k t)]^2,$$

with a value of $u_k \in \mathbb{R}$ that is presently unimportant. One useful property of S_{2k} is that it is symmetric, meaning $S_{2k}(-t) = S_{2k}^{-1}(t)$.

We will find it useful to define $s := 1/r$ and treat s as a continuous variable. It can be shown that $\mathcal{P}(sT)^{1/s}$ is analytic in a neighborhood of $s = 0$ (see Ref. [45] or Lemma 1 below). In such a neighborhood, the product formula can be written as an evolution under an effective Hamiltonian

$$\begin{aligned} \mathcal{P}^{1/s}(sT) &= e^{-iTH_{\text{eff}}(sT)} \\ H_{\text{eff}}(sT) &= H + \sum_{j \geq p} E_{j+1} s^j T^j, \\ &= H + E(sT) \end{aligned} \quad (5)$$

where $E(t) := H_{\text{eff}}(t) - H$ and E_{j+1} are a set of coefficients with explicit form given by the Baker-Campbell-Hausdorff (BCH) formula. When $\mathcal{P}$ is symmetric, the error series is even in s , meaning $E_{j+1} = 0$ for odd j .

B. Variation of parameters

Our primary technical tool in this work will be the variation of parameters equation for linear operators. Given two matrices A, B , the formula reads

$$e^{(A+B)t} = e^{At} + \int_0^t e^{A(t-\tau)} B e^{(A+B)\tau} d\tau. \quad (6)$$

This can be understood as a special case of the formula from the theory ordinary first-order differential equations, treating B as a perturbation [20].

We apply this formula by focusing on the relevant linear operators acting on the space of observables (sometimes dubbed “superoperators”). The exact and Trotterized evolutions of an observable O under H are $O(T) = e^{iHT} O e^{-iHT}$ and $\tilde{O}(T, s) = e^{iT(H+E(sT))} O e^{-iT(H+E(sT))}$, respectively, with E defined in Eq. (5). We see that these are solutions to the first-order differential equations

$$\partial_T O(T) = i \text{ad}_H O(T) \quad \partial_T \tilde{O}(T, s) = i \text{ad}_{H+E(sT)} \tilde{O}(T, s), \quad (7)$$

where $\text{ad}_H(\cdot) := [H, \cdot]$. Taking $A = i \text{ad}_H$ and $B = i \text{ad}_{E(sT)}$, Eq. (6) reads

$$\begin{aligned}\tilde{O}(T, s) &= e^{i T \text{ad}_{H_{\text{eff}}}}(O) \\ &= e^{i T \text{ad}_H}(O) + i \int_0^T d\tau e^{i(T-\tau) \text{ad}_H} \text{ad}_{E(sT)}(\tilde{O}(\tau, s)).\end{aligned}$$

We can recursively insert the variation of parameters equation into the right-hand side. Iterating this $K - 1$ times, for a value of K we will choose appropriately, we get a series expansion for the time-evolved observable in terms of the inverse Trotter step s .

$$\tilde{O}(T, s) = O(T) + \sum_{j \geq p} s^j \tilde{E}_{j+1, K}(T)(O) + \tilde{F}_K(s, T)(O). \quad (8)$$

Here $\tilde{E}$ and $\tilde{F}$ are (super)operator coefficients. The F_K term can be thought of a type of remainder, made smaller with larger K , and the coefficients $\tilde{E}_{j+1, K}$ consist of nested commutators of the H_γ . These claims are made rigorous in Sec. IV A. The value of this expansion is that, by increasing K , we can make the remainder term irrelevant compared to the dominant errors captured by the power series in $\tilde{E}_{j+1, K}$. Thus, by characterizing the $\tilde{E}_{j+1, K}$, it becomes much easier to understand the effects of extrapolation on the resulting estimate.

C. Extrapolation methods

Equation (8) defines a series expansion in terms of the inverse Trotter step size s for the Trotterized time evolution of an observable. This, in turn, gives us a series expansion for the scalar-valued function

$$f(s) := \langle \tilde{O}(T, s) \rangle. \quad (9)$$

Ideally, we wish to estimate $f(0)$ to obtain a time-evolved expectation value without Trotter error. By taking measurements at particular values of s , we can extrapolate to the $s = 0$ ideal. Here we review the two methods to perform this extrapolation considered in this work: Richardson extrapolation and polynomial interpolation.

1. Richardson extrapolation

Given a function f with a series expansion, Richardson extrapolation is a method to get iteratively improved estimates of a particular value [47, 48]. Suppose we have a function with the expansion

$$f(s) = \sum_{k=1}^{\infty} c_k s^k$$

for some coefficients c_k , and we wish to approximate $f(0)$ when the only information about f we have access to values of $f(s)$ for $s \in \{s_1, \dots, s_m\}$. Then we can iteratively

construct an approximation as follows. Suppose we choose some $s_0 = s > 0$ and $s_1 = x/k_1$ for integer $k_1 \neq 0$. We see that if we evaluate the function at a point $f(s/k_1)$, we have

$$f(s/k_1) = f(0) + c_1 \frac{s}{k_1} + c_2 \left(\frac{s}{k_1} \right)^2 + O(s^3).$$

Since $f(s/k_1)$ and $f(s)/k_1$ match at first order, subtracting the two equations yields

$$f(s)/k_1 - f(s/k_1) = (1/k_1 - 1)f(0) + \tilde{c}_2 s^2 + O(s^3)$$

or, after dividing by the factor multiplying $f(0)$,

$$\frac{f(s)/k_1 - f(s/k_1)}{1/k_1 - 1} \equiv F^{(1)}(s) = f(0) + \tilde{c}_2 s^2 + O(s^3). \quad (10)$$

Thus, we have constructed an estimator $F^{(1)}(s)$ of $f(0)$ accurate to order $O(s^2)$. By comparison, $f(s)$ is only order $O(s)$. The precise form of $\tilde{c}_2$ is irrelevant for the present illustration. The procedure can be iterated again to eliminate the $\tilde{c}_2$ terms, and so on as desired.

In general, a linear combination of m evaluations of f , with suitably chosen coefficients b_j , removes m terms in the series expansion. An m -term Richardson extrapolation $F^{(m)}(s)$ has a small s behavior that satisfies

$$\begin{aligned}|F^{(m)}(s) - f(0)| &\leq \sum_{j=1}^m |b_j| \left| \sum_{k=m+1}^{\infty} c_k s^k \right| \\ &= O(s^{m+1}).\end{aligned}$$

Thus, by picking sufficiently small s and sufficiently large m , we achieve a more accurate estimate of $f(0)$. We can then apply this to the series expansion from Eq. (8).

When applying a well-condition Richardson extrapolation for time-evolved expectation values, we will find the error scaling as

$$|F^{(m)}(s) - \langle O(T) \rangle| = O(s^{2m} T^{2m(1+1/p)})$$

for a symmetric order p product formula. By choosing the sampling points $\{s_i\}_{i=1}^m$ such that s_1 is the largest, and all other are related by the Chebyshev nodes starting from s_1 , one can also show robustness of the estimator $F^{(m)}(s)$ to noisy data [23]. We give a more thorough analysis of Richardson extrapolation applied to Trotter-evolved expectation values in Sec. V.

2. Polynomial interpolation

Here we review the proposal of Rendon *et al.* for using polynomial interpolation as a Trotter extrapolation method [45]. Given a real-valued function $f(s)$ in a domain $D \subseteq$

$\mathbb{R}$, and given a set of values $f(s_1), \dots, f(s_m)$, there is a unique $(m-1)$ degree polynomial $P_{m-1}f$, which matches the value of f at each s_i . A concrete representation may be given terms of the Lagrange basis polynomials

$$P_{m-1}f(s) = \sum_{i=1}^m f(s_i) \mathcal{L}_i(s),$$

where

$$\mathcal{L}_i(s) := \prod_{\substack{1 \leq n \leq m \\ n \neq i}} \frac{(s - s_n)}{(s_i - s_n)}.$$

More importantly, $P_{m-1}f$ may be computed by a number of standard techniques such as barycentric interpolation, which is efficient and stable to floating-point errors [49]. By choosing the sample points s_i to be at the *Chebyshev nodes* of an interval of interest, the interpolation will be robust to errors in the computed values $f(s_i)$ [50]. These features make interpolation a plausible tool for Trotter extrapolation.

To apply polynomial interpolation to Trotter-evolved expectation values $f(s)$ of Eq. (9), we consider a neighborhood $[-\ell, \ell]$ of the origin, where $\ell \in \mathbb{R}_+$ is to be chosen based on problem parameters. This function can be computed for both positive and negative s using product formula evolutions on a quantum computer. Performing a Chebyshev interpolation of $m \in 2\mathbb{Z}_+$ sample points, the final estimate is given by $P_{m-1}f(0)$. Figure 1 depicts this setup. One can show [45] that the interpolation error at the point $s = 0$ is bounded as

$$|f(0) - P_{m-1}f(0)| \leq \max_{\xi \in [-\ell, \ell]} |f^{(m)}(\xi)| \left(\frac{\ell}{2m} \right)^m. \quad (11)$$

Utilizing the error expansion Eq. (8), we can bound the derivative term as

$$\partial_s^m \langle \tilde{O}(T, s) \rangle = O(\ell^m T^{m(1+1/p)}).$$

Hence, by choosing the interval with $\ell = O(T^{-(1+1/p)})$, we get that the approximation error from Eq. (11) decreases exponentially with m . See Sec. VI for the complete analysis.

D. Taking the measurements

So far, we have been concerned with the extrapolation schemes and their accuracies. However, to implement either Richardson extrapolation or polynomial interpolation, we need to be able to take measurements of the time-evolved observable at different value of s , and these measurements come with intrinsic errors (regardless of hardware effects). For successful overall error reduction, we will require our measurements to be within some error tolerance of the exact *Trotter*-evolved expectation value, and from robustness guarantees, this allows us to extrapolate to within a small error of the *ideal* evolution. In this work, we consider two schemes for performing the expectation value measurements.

Method 1 (Incoherent): Simply time evolve the initial state ρ_0 under $\mathcal{P}^{1/s_j}(s_j T)$, then measure the expectation value by repeated measurement of O . The precision will necessarily be shot noise limited, leading to a $O(1/\epsilon^2)$ cost for precision ϵ . However, circuit depths will only be as long as needed to perform a single measurement, and often these can be estimated straightforwardly with low depth circuits (e.g., using classical shadows using randomized single-qubit measurements).

Method 2 (Coherent): Use iterative quantum amplitude amplification to achieve Heisenberg-limited scaling

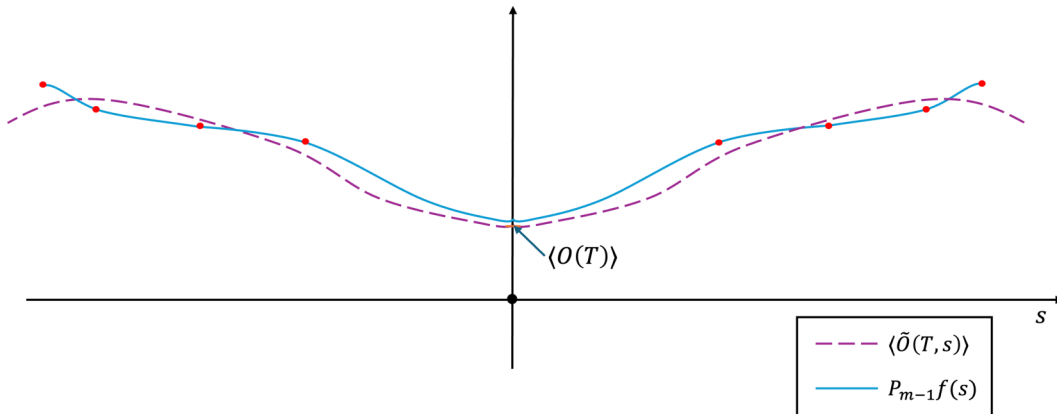


FIG. 1. Schematic of the polynomial interpolation procedure. The dotted purple line is the true value of the Trotter-evolved expectation value, $\langle \tilde{O}(T, s) \rangle$, and the blue line is the interpolating polynomial $P_{m-1}f(s)$ for $m = 8$. The red points are estimates of the time-evolved expectation value obtained via product-formula evolution with measurement, which are then used to construct the polynomial. The final estimator of the expectation value is given by $P_{m-1}f(0)$.

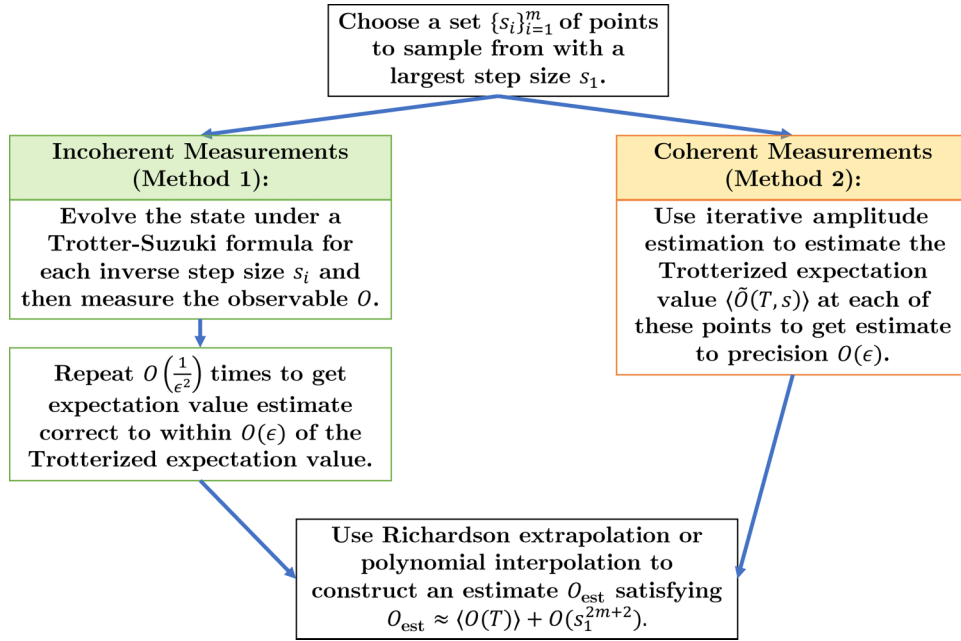


FIG. 2. Flowchart indicating the incoherent (Method 1) and coherent (Method 2) schemes for performing extrapolation of the time-evolved expectation values using product formulas, where the specified time and final error are T and ϵ , respectively. For each time-step size, we need to make a measurement of the observable’s expectation to precision ϵ . While the incoherent scheme is simpler and requires shorter circuit depths, the coherent scheme achieves a quadratic speedup from using quantum amplitude estimation. Regardless of the measurement protocol, the algorithm concludes with the same classical calculation of the acquired data.

$\tilde{O}(1/\epsilon)$ in the measurement precision. This is essentially a quadratic improvement over in error scaling Method 1 but involves longer circuit depths. Iterative protocols for amplitude estimation require few auxiliary qubits while still achieving competitive results compared to Fourier-based methods. For our analysis, we consider specifically the iterative quantum amplitude estimation protocol of Grinko *et al.* [51].

A flowchart illustrating the entire extrapolation process is given in Fig. 2. Method 1 optimizes for shorter circuit depths at the expense of greater overall resources by using incoherent measurements for observables. As such, this method may be more promising for NISQ devices. Method 2 uses coherent measurements to improve the overall scaling at the expense of longer circuits, and hence may be more useful for fault-tolerant devices.

III. RESULTS

Our results are stated for two separate measurement routines for performing the expectation value calculations, as discussed in Sec. II D. Figure 2 gives a schematic workflow for the extrapolation protocols using either measurement type. We now informally state the main results of our work; see later sections for more rigorous formulations.

Theorem 1 (Trotter extrapolation resource counts (informal)). Let O be an observable and $H = \sum_{\gamma}^{\Gamma} H_{\gamma}$ be

a time-independent Hamiltonian. Let O_{est} be the estimate for a time-evolved expectation value on an arbitrary initial state, $\langle O(T) \rangle$, produced by varying the Trotter step size of a p th-order product formula, taking m measurement samples, and then extrapolating to the zero step-size limit using either Richardson extrapolation or polynomial interpolation. Then, for a simulation time T , a relative error ϵ can be achieved, with high probability, such that

$$|\langle O(T) \rangle - O_{\text{est}}| \leq \epsilon \|O\|$$

using $O(\log(1/\epsilon))$ extrapolation points. Furthermore, the maximum circuit depth and total gate count for each of the estimation subroutines described above scales as shown in the following table.

Method	Cost	
	Max circuit depth	Total gate cost
Incoherent	$\tilde{O}\left((\lambda T)^{1+1/p} \text{polylog}\left(\frac{1}{\epsilon}\right)\right)$	$\tilde{O}\left(\frac{(\lambda T)^{1+1/p}}{\epsilon^2} \text{polylog}\left(\frac{1}{\epsilon}\right)\right)$
Coherent	$\tilde{O}\left(\frac{(\lambda T)^{1+1/p}}{\epsilon} \text{polylog}\left(\frac{1}{\epsilon}\right)\right)$	$\tilde{O}\left(\frac{(\lambda T)^{1+1/p}}{\epsilon} \text{polylog}\left(\frac{1}{\epsilon}\right)\right)$

Here $\tilde{O}$ hides $\log \log$ factors, and $\lambda \leq 4 \sum_{\gamma} \|H_{\gamma}\|$ is a nested commutator with full expression given in Lemma 7.

TABLE I. Summary of asymptotic scalings for simulation protocols for time-evolved expectation values (without extrapolation) where incoherent and coherent measurement procedures are used. Results are obtained using the current best scalings for product formulas. This should be compared to the results in Theorem 1. The quantity $\alpha_{\text{comm}}^{(p+1)} \leq 2^p (\sum_{\gamma} \|H_{\gamma}\|)^{p+1}$ is a sum of $p + 1$ nested commutators defined in Eq. (15). μ is a different norm of nested commutators defined in Ref. [20]. $\Lambda = \sum_{\gamma} \|H_{\gamma}\|$.

Method	Alternative methods without extrapolation	
	Max circuit depth	Cost
Trotter (incoherent) [19]	$\tilde{O}\left(\frac{(\alpha_{\text{comm}}^{(p+1)})^{1/p} T^{1+1/p}}{\epsilon^{1/p}}\right)$	$\tilde{O}\left(\frac{(\alpha_{\text{comm}}^{(p+1)})^{1/p} T^{1+1/p}}{\epsilon^{2+1/p}}\right)$
Trotter (coherent) [19]	$\tilde{O}\left(\frac{(\alpha_{\text{comm}}^{(p+1)})^{1/p} T^{1+1/p}}{\epsilon^{1+1/p}}\right)$	$\tilde{O}\left(\frac{(\alpha_{\text{comm}}^{(p+1)})^{1/p} T^{1+1/p}}{\epsilon^{1+1/p}}\right)$
Multiproduct formula (incoherent) [20]	$\tilde{O}(\mu T \log(1/\epsilon))$	$\tilde{O}\left(\frac{\mu T}{\epsilon^2} \log(1/\epsilon)\right)$
Multiproduct formula (coherent) [20]	$\tilde{O}\left(\frac{\mu T}{\epsilon} \log(1/\epsilon)\right)$	$\tilde{O}\left(\frac{\mu T}{\epsilon} \log(1/\epsilon)\right)$
Qubitization (incoherent) [15]	$\tilde{O}(\Lambda T + \log(1/\epsilon))$	$\tilde{O}\left(\frac{1}{\epsilon^2} (\Lambda T + \log(1/\epsilon))\right)$
Qubitization (coherent) [15]	$\tilde{O}\left(\frac{1}{\epsilon} (\Lambda T + \log(1/\epsilon))\right)$	$\tilde{O}\left(\frac{1}{\epsilon} (\Lambda T + \log(1/\epsilon))\right)$

These results demonstrate an improvement in circuit depth over the direct application of product formulas by an exponential in the error scaling, while maintaining the same time scaling (see Table I for a direct comparison for the results in Theorem 1). As we discuss in Sec. III A below, our analysis improves on prior results for both Richardson extrapolation and polynomial interpolation. The extrapolated techniques are shown to inherit all of the desirable properties of standard product-formulas methods, such as locality, commutator scaling, respecting the system's symmetries, etc.

In Appendix A, we show how the prefactors in Theorem 1 can be improved by taking into account symmetries of the system. While our commutator λ is not obviously comparable to the simpler $\alpha_{\text{comm}}^{(p+1)}$ commutator from Trotter error bounds, we show (Corollary 2) that in certain regimes our Trotter errors can indeed be bounded with $\alpha_{\text{comm}}^{(p+1)}$ essentially replacing λ^{p+1} . We then use this in Sec. VII to show concrete improvements in resource analysis by porting over well-studied properties of $\alpha_{\text{comm}}^{(p+1)}$ from prior literature [19]. Though it appears difficult to show, the λ error bound likely exhibits significant improvements over norm-bound estimates in general, even outside the parameter regimes of Corollary 2.

We note that the results have either $1/\epsilon$ or $1/\epsilon^2$ scaling for the coherent or incoherent measurement procedures, respectively; this is a consequence of the measurement protocols, not the time evolution. Indeed, this scaling would be present even for multiproduct formulas or qubitization-based simulation combined with incoherent or coherent

measurement protocols, and lower bounds can be shown from results in metrology [52].

Additionally, Method 1 (incoherent measurements) is fully compatible with classical shadow techniques for estimating many observables. To do this, in Sec. VIII we show how to extrapolate in a way that essentially constructs a classical shadow representation of the idealized final state of the evolution. This allows us to estimate M different local observables with probability $\geq 1 - \delta$, then we get an additional resource cost $O(\log(M/\delta))$.

Finally we note that the bounds in Theorem 1 are in terms of the number of Trotter steps. The total number of elementary exponential operations comes with a multiplicative factor $\Upsilon\Gamma$, which may contain additional information about scaling with system size. A true elementary circuit depth and gate cost would require more specification of a computational model, such as a k local or sparse matrix model.

A. Comparison to previous extrapolation results

1. Richardson extrapolation

The use of Richardson extrapolation to reduce Trotter error was originally proposed in Ref. [43]. Although the authors note that Richardson extrapolation should reduce the error with repeated samples, the work does not provide rigorous resource estimates in terms of the basic simulation parameters. Subsequent work in Ref. [44] employs Richardson extrapolation on actual quantum hardware, observing improvements over bare simulation results.

However, there appear to be errors in the formal treatment of the algorithmic error, which we discuss in Appendix D. As a result, a rigorous asymptotic analysis of the scaling of Richardson extrapolation for quantum simulation does not appear to be present in the literature. We hope the present work provides a more complete theoretical picture of the Richardson approach.

2. Polynomial interpolation

The polynomial interpolation approach was introduced and rigorously analyzed in Ref. [45]. There it was shown that, to measure an observable at time T , it suffices to use a total number of Trotter exponentials (in the coherent measurement case) scaling as $\tilde{O}((\chi T)^2/\epsilon \log(\chi T/\epsilon))$, where $\chi = \Gamma \max_{\gamma} \|H_{\gamma}\|$. This work improves the time scaling to $T^{1+1/p}$, which matches regular p th-order Trotter. Moreover, the norm bound χ is replaced with a commutator expression that is upper bounded by χ (up to small constants), and is often much smaller for local Hamiltonians. An important technical step to this commutator scaling is showing the effective Hamiltonian is analytic within a radius, which depends inversely on a commutator, not a norm bound on H . This may be of independent interest.

Besides improving the analysis, we relax the requirement for exact Chebyshev interpolation, avoiding the need for fractional queries of the form $\mathcal{P}^{\delta}$ for $\delta \in (0, 1)$. This leads to additional factors of $\log(1/\epsilon)$ in the complexity, but we feel the reduction in overhead and conceptual difficulty makes up for this. Our simplified approach can now more genuinely be summarized as consisting only of Trotter and classical postprocessing.

3. Other approaches using classical postprocessing

Other approaches at “dequantizing” Hamiltonian simulation techniques exist, but are not directly related to the present work. For example, Ref. [53] introduces a randomized version of the multiproduct formula presented in Ref. [23], achieving near-optimal scaling. However, the method requires additional auxiliary qubits and control operations. Further work on dequantizing the multiproduct formula can be seen in Ref. [54], in which they improve the scaling of the multiproduct formula and show it satisfies commutator scaling with a quadratic improvement in the error associated with the size of the time steps relative to standard p^{th} -order product formulas. Other approaches include constructing nonunitary channels, which can be computed using classical postprocessing using randomized compiling methods [38,55].

IV. ERROR ANALYSIS

In this section, we derive our primary technical results and develop our most important tool: an explicit series expansion, with respect to Trotter step, of a time-evolved

expectation value of an observable O evolved under a product formula. The results contained here will be directly applied to Richardson extrapolation and polynomial interpolation in the subsequent sections. Our approach is inspired by recent work of Aftab *et al.* [20], although we eventually deviate to write simpler bounds.

We begin with some essential notation and definitions. A staged product formula $\mathcal{P}$ is defined as

$$\mathcal{P}(t) := \prod_{v=1}^{\Upsilon} \prod_{\gamma=1}^{\Gamma} e^{-ita_{(v,\gamma)} H_{\pi_v(\gamma)}}, \quad (12)$$

where $\Upsilon \in \mathbb{Z}_+$ is the number of *stages*, $a_{(v,\gamma)} \in \mathbb{R}$ are coefficients and $\pi_v \in S_{\Gamma}$ are permutations [19]. For the case of Trotter-Suzuki formulas, $\Upsilon = 2 \times 5^{k-1}$ and $|a_{(v,\gamma)}| \leq 2k/3^k$ [56, Appendix A]. For $p \in \mathbb{Z}_+$, a p th-order product formula satisfies

$$\mathcal{P}(t) - e^{-iHt} = O(t^{p+1}) \quad (13)$$

in the small t limit. For ease of notation, let $[X_1 X_2 \dots X_n]$ refer to a right-nested n commutator

$$[X_1 X_2 \dots X_n] := [X_1, [X_2, [\dots [X_{n-1}, X_n] \dots]]] \quad (14)$$

and define

$$\alpha_{\text{comm}}^{(j)} := \sum_{\gamma_1 \gamma_2 \dots \gamma_j = 1}^{\Gamma} \| [H_{\gamma_1} H_{\gamma_2} \dots H_{\gamma_j}] \|, \quad (15)$$

It will be useful, for our purposes, to employ a BCH formula based on right-nested commutators [57]. We first define

$$\begin{aligned} \phi_j(Y_1, Y_2, \dots, Y_j) \\ := \frac{1}{j^2} \sum_{\sigma \in S_j} (-1)^{d_{\sigma}} \binom{j-1}{d_{\sigma}} [Y_{\sigma(1)} \dots Y_{\sigma(j)}], \end{aligned} \quad (16)$$

where S_j is the set of permutations over j elements and d_{σ} is the number of *descents* in σ . An index i is a descent of σ if $\sigma(i) > \sigma(i+1)$. In these terms, the BCH formula reads

$$\prod_{i=1}^n e^{X_i} = e^Z, \quad (17)$$

where

$$Z = \sum_{i=1}^n X_i + \sum_{j=2}^{\infty} \frac{1}{j!} \sum_{\mathcal{J}} \binom{j}{j_1 \dots j_n} \phi_j(X_1^{\times j_1}, \dots, X_n^{\times j_n}). \quad (18)$$

Here

$$\mathcal{J} := \left\{ (j_1, \dots, j_n) \in \mathbb{N}^n \mid \sum_{j=1}^n j_i = j \right\} \quad (19)$$

sums appropriately over the multinomial coefficient, and $X^{\times i}$ refers to i copies of X in the argument. Equation (18) is formal and may not converge, though a sufficient criterion for convergence is absolute convergence in the operator norm.

A. Error expansion of the time-evolved observable

We begin with a lemma concerning an error series in the effective Hamiltonian $H_{\text{eff}}(t)$ generating $\mathcal{P}$, which approximates H for small t .

Lemma 1 (Effective Hamiltonian error series). Let $\mathcal{P}$ be a staged product formula of order p , and let H_{eff} be the effective Hamiltonian of $\mathcal{P}$ defined by the relation

$$\mathcal{P}(t) = e^{-itH_{\text{eff}}(t)}$$

for $t \in \mathbb{R}$. Let $\ell \in \mathbb{R}_+$ be such that, for all $t \in (-\ell, \ell)$ the sequence $\alpha_{\text{comm}}^{(j)} (a_{\max} \Upsilon |t|)^j$ is bounded in j , with $a_{\max} := \max_{v,\gamma} |a_{(v,\gamma)}|$. Then the effective Hamiltonian admits a power-series expansion

$$H_{\text{eff}}(t) = H + \sum_{j=1}^{\infty} E_{j+1} t^j$$

for $t \in (-\ell, \ell)$, where

$$E_j := \frac{(-i)^{j-1}}{j!} \sum_{\mathcal{J}} \binom{j}{j_1 \dots j_n} \left(\prod_{i=1}^n a_i^{j_i} \right) \times \phi_j \left(H_{\gamma_1}^{\times j_1}, \dots, H_{\gamma_n}^{\times j_n} \right).$$

Here $n = \Upsilon \Gamma$ is the number of terms in $\mathcal{P}$ and $i \in [n]$ indexes the terms in order. Moreover, E_j satisfies the bound

$$\|E_j\| \leq \frac{(a_{\max} \Upsilon)^j}{j^2} \alpha_{\text{comm}}^{(j)}.$$

Remark. Note that a radius of convergence ℓ certainly exists by a norm bound on $\alpha_{\text{comm}}^{(j)}$; however our radius of convergence is commutator dependent and will generally be larger.

Proof. Consider $\mathcal{P}$ written as

$$\mathcal{P}(t) = \prod_{i=1}^n e^{-ita_i H_i}, \quad (20)$$

where $H_i \equiv H_{\gamma_i}$. By definition, $\prod_{i=1}^n e^{-ita_i H_i} = e^{-itH_{\text{eff}}}$ and may therefore express H_{eff} as a formal BCH expansion.

Using Eq. (18) with $Z = -itH_{\text{eff}}(t)$ and $X_i = -ita_i H_i$,

$$\begin{aligned} -itH_{\text{eff}}(t) &= \sum_{i=1}^n -ita_i H_i + \sum_{j=2}^{\infty} \frac{1}{j!} \sum_{\mathcal{J}} \binom{j}{j_1 \dots j_n} \\ &\quad \times \phi_j \left((-ita_1 H_1)^{\times j_1}, \dots, (-ita_n H_n)^{\times j_n} \right). \end{aligned} \quad (21)$$

Since $\mathcal{P}$ is at least first order, we have $\sum_i a_i H_i = H$. Using the multilinearity of ϕ_j ,

$$\begin{aligned} H_{\text{eff}}(t) &= H + \sum_{j=1}^{\infty} \frac{(-it)^j}{(j+1)!} \sum_{\mathcal{J}} \binom{j+1}{j_1 \dots j_n} \left(\prod_{i=1}^n a_i^{j_i} \right) \\ &\quad \times \phi_{j+1} \left(H_1^{\times j_1}, \dots, H_n^{\times j_n} \right) \\ &= H + \sum_{j=1}^{\infty} t^j E_{j+1}, \end{aligned} \quad (22)$$

where we have defined the Hermitian error operators

$$\begin{aligned} E_j &:= \frac{(-i)^{j-1}}{j!} \sum_{\mathcal{J}} \binom{j}{j_1 \dots j_n} \left(\prod_{i=1}^n a_i^{j_i} \right) \\ &\quad \times \phi_j \left(H_1^{\times j_1}, \dots, H_n^{\times j_n} \right). \end{aligned} \quad (23)$$

Applying the triangle inequality and using the definition of ϕ_j in Eq. (16)

$$\begin{aligned} \|E_j\| &\leq \frac{1}{j!} \sum_{\mathcal{J}} \binom{j}{j_1 \dots j_n} \left(\prod_{i=1}^n |a_i|^{j_i} \right) \frac{1}{j^2} \\ &\quad \times \sum_{\sigma \in S_j} \binom{j-1}{d_{\sigma}}^{-1} \| [H_{\sigma(i_1)} \dots H_{\sigma(i_j)}] \|, \end{aligned} \quad (24)$$

where by observing the argument of ϕ_j has exactly j terms, we write

$$H_{i_1} \dots H_{i_j} = H_1^{\times j_1} \dots H_n^{\times j_n} \quad (25)$$

with $i_k \in \{1, \dots, n\}$. We wish to reindex this sum to be over $i_1, \dots, i_j$, with i_k varying freely, and accomplish this by rehashing arguments used surrounding [20, Eq. (44)]. For a given $[H_{i_1} \dots H_{i_j}]$, there exists a unique $(j_1, \dots, j_n) \in \mathcal{J}$ specifying the terms in that commutator, as the j_k indices give the number of each H_k present in the commutator, and

this is a function of a full $(i_1, \dots, i_j)$ specification. Thus, we may write

$$\|E_j\| \leq \frac{1}{j!j^2} \sum_{i_1 \dots i_j=1}^n \| [H_{i_1} \dots H_{i_j}] \| \binom{j}{\vec{j}} \left(\prod_{k=1}^j |a_{i_k}| \right) \times \sum_{\sigma \in S(i_1, \dots, i_j)} \binom{j-1}{d_\sigma}^{-1}, \quad (26)$$

where $\vec{j} = (j_1, \dots, j_n)$ is the vector of counts determined by the i_k , and $S(i_1, \dots, i_j) \subseteq S_j$ refers to the permutations, which leave the sequence $(i_1, \dots, i_j)$ invariant. There are exactly $j_1!j_2! \dots j_n!$ of these. Thus,

$$\sum_{\sigma \in S(i_1, \dots, i_j)} \binom{j-1}{d_\sigma}^{-1} \leq \sum_{\sigma \in S(i_1, \dots, i_j)} 1 = j_1!j_2! \dots j_n! \quad (27)$$

and therefore

$$\|E_j\| \leq \frac{1}{j^2} \sum_{i_1 \dots i_j=1}^n \| [H_{i_1} \dots H_{i_j}] \| \prod_{k=1}^j |a_{i_k}| \leq \frac{d_{\max}}{j^2} \sum_{i_1 \dots i_j=1}^n \| [H_{i_1} \dots H_{i_j}] \|, \quad (28)$$

where $a_{\max} := \max_i |a_i|$. We now remember that each H_{i_n} runs Υ times over each term H_{i_j} in H . Thus, each sequence $H_{i_1}H_{i_2} \dots H_{i_j}$ is represented Υ^j times in the multi-index $(i_1, \dots, i_j)$. Hence,

$$\sum_{i_1 \dots i_j=1}^n \| [H_{i_1} \dots H_{i_j}] \| = \Upsilon^j \sum_{i_1, \dots, i_j=1}^{\Gamma} \| [H_{i_1} \dots H_{i_j}] \| = \Upsilon^j \alpha_{\text{comm}}^{(j)}. \quad (29)$$

This gives our upper bound on E_j from the lemma statement.

To summarize, we have that $tH_{\text{eff}}(t) = tH + \sum_{j=1}^{\infty} E_{j+1}t^{j+1}$ provided that the series converges. A sufficient condition is absolute convergence, namely the convergence of

$$\sum_{j=2}^{\infty} |t|^j \|E_j\|. \quad (30)$$

We note that $\sum_j C/j^2$ is a convergent series for all $C \in \mathbb{R}_+$, and thus, by the squeeze theorem, it suffices that there exists an $J \in \mathbb{Z}_+$ such that for all $j \geq J$, $|t|^j \|E_j\| \leq C/j^2$. Using our bound on $\|E_j\|$, this is satisfied provided

$$(a_{\max} \Upsilon |t|)^j \alpha_{\text{comm}}^{(j)} \leq C \quad (31)$$

for such j . This is equivalent to the condition provided in the lemma. ■

Observe that less stringent conditions for convergence could be derived, e.g., by bounding the series with a $1/j^{1+\epsilon}$ decay for any $\epsilon > 0$ instead of $1/j^2$, and indeed the condition under which such series converge is well studied [58,59].

A *symmetric* product formula is one for which $\mathcal{P}(-t) = \mathcal{P}^{-1}(t)$. For p th-order formulas, the lowest E_j are zero, and for symmetric formulas, the error series for H_{eff} is even, as is captured in the following lemma.

Lemma 2. Let $\mathcal{P}(t)$ be a p th-order staged product formula, and suppose the BCH convergence condition of Lemma 1 holds. Then the error operators E_{j+1} from Lemma 1 are zero for all $j < p$. Moreover, for symmetric $\mathcal{P}$, $E_{j+1} = 0$ for all odd j .

Proof. We will show that $H_{\text{eff}}(t) = H + O(t^p)$, which directly implies the first claim of the lemma. The Hamiltonians $H_{\text{eff}}(t)$ and H may be defined through the logarithm

$$H_{\text{eff}}(t) = -\frac{1}{it} \log \mathcal{P}(t), \quad H = -\frac{1}{it} \log U(t) \quad (32)$$

and, in a neighborhood of $t = 0$, the logarithm may be expanded in a power series.

$$\begin{aligned} H_{\text{eff}}(t) - H &= -\frac{1}{it} \sum_{j=0}^{\infty} \frac{(-1)^j}{j+1} [(\mathcal{P}(t) - \mathbb{1})^{j+1} - (U(t) - \mathbb{1})^{j+1}] \\ &= \frac{1}{it} \sum_{j=1}^{\infty} \frac{(-1)^j}{j} \left[\sum_{k=0}^j (-1)^{j-k} (\mathcal{P}^k(t) - U^k(t)) \right]. \end{aligned} \quad (33)$$

Since $\mathcal{P}(t) - U(t) = O(t^{p+1})$ we have that $\mathcal{P}(t) - U(t) = t^{p+1}E(t)$ for some analytic, operator-valued function E . In fact, this implies

$$\mathcal{P}^k(t) - U^k(t) = t^{p+1}E^{(k)}(t) \quad (34)$$

for some analytic $E^{(k)}$. Factoring out the t^{p+1} from the series, we find that

$$H_{\text{eff}}(t) - H = t^p \tilde{E}(t) \quad (35)$$

where, again, $\tilde{E}$ is some analytic operator-valued function. This shows that $H_{\text{eff}}(t) - H = O(t^p)$ and thus, provided the BCH series exists, all $E_j = 0$ for $j < p$.

For symmetric product formulas, the condition $\mathcal{P}(-t) = \mathcal{P}^\dagger(t)$ ($\mathcal{P}$ is unitary) implies for the effective Hamiltonian that

$$(-itH_{\text{eff}}(t))^\dagger = -i(-t)H_{\text{eff}}(-t), \quad (36)$$

which further implies $H_{\text{eff}}(t)$ is an even function. Thus, the BCH error series is even provided it exists. ■

As an aside, it is possible that the order conditions from Ref. [19, Theorem 9], could be adapted for our purposes. However, they characterize their exponential error using a time-ordered exponential, which is suitable for their objectives but less so for ours.

We have thus characterized the error terms in H_{eff} adequately for our purposes. Moving towards our primary interest, dynamical evolution of observables, we now wish to construct an error series for an observable evolved under a product formula. The following lemma provides what we need, and mimics the ideas of Ref. [20]. Our analysis is based on the variation of parameters applied to the Hamiltonian evolution of observables.

Lemma 3. Let $\mathcal{P}$ be a staged p th-order product formula and let O be an observable. For any $s \in \mathbb{R}$, let

$$\tilde{O}(T, s) := \mathcal{P}^{1/s^\dagger}(sT) O \mathcal{P}^{1/s}(sT)$$

be the product-formula evolution of O for duration $T \in \mathbb{R}_+$ and Trotter step size sT , with $s = 0$ defined via the limit. Let $\ell \in \mathbb{R}_+$ be such that, for all $s \in (-\ell, \ell)$ the sequence $\alpha_{\text{comm}}^{(j)}(a_{\text{max}} \Upsilon |s| T)^j$ is bounded in j , with $a_{\text{max}} := \max_{v, \gamma} |a_{(v, \gamma)}|$. Let $\sigma = 2$ if $\mathcal{P}$ is symmetric, 1 otherwise. Then for any $K \in \mathbb{Z}_+$, the approximation error in $\tilde{O}(T, s)$ compared with the exact evolution $O(T) := e^{iHT} O e^{-iHT}$ may be expressed as

$$\tilde{O}(T, s) = O(T) + \sum_{j \in \sigma \mathbb{Z}_+, j \geq p} s^j \tilde{E}_{j+1}(T)(O),$$

where $\tilde{E}_{j+1}(T)$ are superoperators whose induced spectral norm $\|\cdot\|$ is bounded as

$$\|\tilde{E}_{j+1}(T)\| \leq (a_{\text{max}} \Upsilon T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{(a_{\text{max}} \Upsilon T)^l}{l!} \times \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l 2 \frac{\alpha_{\text{comm}}^{(j_\kappa+1)}}{(j_\kappa+1)^2} \right).$$

Proof. By definition, we have that

$$\mathcal{P}^{1/s}(sT) = e^{-iTH_{\text{eff}}(t)}, \quad (37)$$

where $t := sT$ is the Trotter step size. By Lemma 1, H_{eff} admits a convergent power-series expansions. We write $H_{\text{eff}}(t) = H + E(t)$, where

$$E(t) = \sum_{j \in \sigma \mathbb{Z}_+, j \geq p} E_{j+1} t^j \quad (38)$$

and zero terms are excluded by Lemma 2. In what follows, the independent parameters are t and T , and any t dependence is left implicit. For observable O , and Hermitian H and E , the evolution equations for $\tilde{O}(T) := e^{i(H+E)T} O e^{-i(H+E)T}$ and $O(T) := e^{iHT} O e^{-iHT}$ are given by

$$\partial_T \tilde{O}(T) = i[H + E, \tilde{O}(T)], \quad \partial_T O(T) = i[H, O(T)]. \quad (39)$$

The solution may be written in a superoperator formalism

$$\tilde{O}(T, s) = e^{iT \text{ad}_{H+E}} O, \quad O(T) = e^{iT \text{ad}_H} O, \quad (40)$$

where $\text{ad}_X(\cdot) := [X, \cdot]$ is (anti) Hermitian with respect to the Hilbert-Schmidt inner product when X is (anti) Hermitian with respect to the standard inner product. In this context, Eq. (6) gives

$$\tilde{O}(T, s) = O(T) + \int_0^T d\tau_1 e^{i(T-\tau_1) \text{ad}_H} i \text{ad}_E(\tilde{O}(\tau_1)). \quad (41)$$

This can be viewed as a Heisenberg evolution in the interaction frame of H . Because ad_E is a bounded operator for $s \in (-\ell, \ell)$, $\tilde{O}$ admits a Dyson series expansion that can be obtained by iterating this formula. To illustrate, we carry out this procedure explicitly. Iterating once,

$$\tilde{O}(T, s) = O(T) + \int_0^T d\tau_1 e^{i(T-\tau_1) \text{ad}_H} i \text{ad}_E(O(\tau_1)) + \int_0^T d\tau_1 \int_0^{\tau_1} d\tau_2 e^{i(T-\tau_1) \text{ad}_H} i \text{ad}_E e^{i(\tau_1-\tau_2) \text{ad}_H} i \text{ad}_E(\tilde{O}(\tau_2)). \quad (42)$$

Iterating $K - 1$ times gives

$$\tilde{O}(T, s) = O(T) + \sum_{l=1}^{K-1} \int_0^T d\tau_1 \int_0^{\tau_1} d\tau_2 \dots \int_0^{\tau_{l-1}} d\tau_l e^{i(T-\tau_1) \text{ad}_H} i \text{ad}_E e^{i(\tau_1-\tau_2) \text{ad}_H} i \text{ad}_E \dots e^{i(\tau_{l-1}-\tau_l) \text{ad}_H} i \text{ad}_E(O(\tau_l)) \quad (43)$$

$$+ \int_0^T d\tau_1 \int_0^{\tau_1} d\tau_2 \dots \int_0^{\tau_{K-1}} d\tau_K e^{i(T-\tau_1) \text{ad}_H} i \text{ad}_E e^{i(\tau_1-\tau_2) \text{ad}_H} i \text{ad}_E \dots e^{i(\tau_{K-1}-\tau_K) \text{ad}_H} i \text{ad}_E(\tilde{O}(\tau_K)). \quad (44)$$

By convergence of the Dyson series, the remainder term Eq. (44) tends to zero for large K . We thus focus on the partial sums Eq. (43). Expanding the definition of E ,

$$\begin{aligned} & \sum_{l=1}^{K-1} \int_0^T d\tau_1 \int_0^{\tau_1} d\tau_2 \dots \int_0^{\tau_{l-1}} d\tau_l e^{i(T-\tau_1) \text{ad}_H} i \text{ad}_E e^{i(\tau_1-\tau_2) \text{ad}_H} i \text{ad}_E \dots e^{i(\tau_{l-1}-\tau_l) \text{ad}_H} i \text{ad}_E (O(\tau_l)) \\ &= \sum_{l=1}^{K-1} \int_0^T d\tau_1 \int_0^{\tau_1} d\tau_2 \dots \int_0^{\tau_{l-1}} d\tau_l \left(\prod_{\kappa=l}^1 \left(\sum_{j_\kappa \in \sigma \mathbb{Z}_+ \geq p} e^{i(\tau_{\kappa-1}-\tau_\kappa) \text{ad}_H} i \text{ad}_{E_{j_\kappa+1}} t^{j_\kappa} \right) \right) (O(\tau_l)) \end{aligned} \quad (45)$$

$$= \sum_{l=1}^{K-1} \int_0^T d\tau_1 \int_0^{\tau_1} d\tau_2 \dots \int_0^{\tau_{l-1}} d\tau_l \left(\sum_{j \in \sigma \mathbb{Z}_+ \geq pl} t^j \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=l}^1 e^{i(\tau_{\kappa-1}-\tau_\kappa) \text{ad}_H} i \text{ad}_{E_{j_\kappa+1}} \right) \right) (O(\tau_l)), \quad (46)$$

with $\tau_0 \equiv T$. We now reinsert $t = sT$, and make a change of variables $s_i = \tau_i/T$. This gives

$$\sum_{l=1}^{K-1} T^l \int_0^1 ds_1 \int_0^{s_1} ds_2 \dots \int_0^{s_{l-1}} ds_l \left(\sum_{j \in \sigma \mathbb{Z}_+ \geq pl} (sT)^j \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=l}^1 e^{i(s_{\kappa-1}-s_\kappa)T \text{ad}_H} i \text{ad}_{E_{j_\kappa+1}} \right) \right) (O(Ts_l)). \quad (47)$$

Next, we regroup the sum according to the degree of s , which yields

$$\begin{aligned} & \sum_{j \in \sigma \mathbb{Z}_+ \geq p} (sT)^j \sum_{l=1}^{\min\{K-1, \lfloor j/p \rfloor\}} T^l \int_0^1 ds_1 \int_0^{s_1} ds_2 \dots \int_0^{s_{l-1}} ds_l \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=l}^1 e^{iT(s_{\kappa-1}-s_\kappa) \text{ad}_H} i \text{ad}_{E_{j_\kappa+1}} \right) (O(Ts_l)) \\ &= \sum_{j \in \sigma \mathbb{Z}_+ \geq p} s^j \tilde{E}_{j+1,K}(T)(O). \end{aligned} \quad (48)$$

Here, we have defined

$$\tilde{E}_{j+1,K}(T) := \sum_{l=1}^{\min\{K-1, \lfloor j/p \rfloor\}} T^{j+l} \int_0^1 ds_1 \int_0^{s_1} ds_2 \dots \int_0^{s_{l-1}} ds_l \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=l}^1 e^{i(s_{\kappa-1}-s_\kappa)T \text{ad}_H} i \text{ad}_{E_{j_\kappa+1}} \right) e^{is_l T \text{ad}_H}. \quad (49)$$

Altogether, we have for any $K \in \mathbb{Z}_+$,

$$\tilde{O}(s, T) = O(T) = \sum_{j \in \sigma \mathbb{Z}_+ \geq p} s^j \tilde{E}_{j+1}(T)(O) + R_K(O), \quad (50)$$

where $R_K \rightarrow 0$ as $K \rightarrow 0$. Further, observe that each $\tilde{E}_{j+1,K}$ becomes constant in K for $K \geq \lfloor j/p \rfloor$. We write this operator as simply $\tilde{E}_{j+1}$, and by convergence of the Dyson series $\tilde{O}(s, T) = O(T) + \sum s^j \tilde{E}_{j+1}(T)(O)$.

Finally, we wish to bound the induced spectral norm of $\tilde{E}_{j+1}$. Using the triangle inequality, unitarity of $e^{i\tau \text{ad}_H}$, and evaluating the remaining integral,

$$\|\tilde{E}_{j+1}(T)\| \leq \sum_{l=1}^{\lfloor j/p \rfloor} T^{j+l} \int_0^1 \int_0^{s_1} \dots \int_0^{s_{l-1}} ds_l \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l \|\text{ad}_{E_{j_\kappa+1}}\| \right) \leq T^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{T^l}{l!} \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l \|\text{ad}_{E_{j_\kappa+1}}\| \right). \quad (51)$$

Noting that $\|\text{ad}_x\| \leq 2\|X\|$ and applying Lemma 1,

$$\begin{aligned} \|\tilde{E}_{j+1}(T)\| &\leq T^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{T^l}{l!} \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l 2\alpha_{\text{comm}}^{(j_\kappa+1)} \frac{(a_{\text{max}} \Upsilon)^{j_\kappa+1}}{(j_\kappa+1)^2} \right) \\ &= (a_{\text{max}} \Upsilon T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{(a_{\text{max}} \Upsilon T)^l}{l!} \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l 2 \frac{\alpha_{\text{comm}}^{(j_\kappa+1)}}{(j_\kappa+1)^2} \right), \end{aligned} \quad (52)$$

which is the desired bound. $\blacksquare$

The above result can be understood as providing a Taylor series expansion about $s = 0$ for the Heisenberg-evolved observable. We record this fact in a corollary.

Corollary 1. Under the assumptions and notation of Lemma 3, $\tilde{O}(T, s)$ is analytic in $s \in (-\ell, \ell)$, and the derivatives at $s = 0$ are given by

$$\partial_s^j \tilde{O}(T, 0) = j! \tilde{E}_{j+1}(T)(O)$$

for any $j \in \mathbb{Z}_+$.

Lemma 3 already provides a commutator bound on the error in $\tilde{O}(s, T)$, in terms of the expressions

$$\lambda_{j,l} := \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l 2 \frac{\alpha_{\text{comm}}^{(j_\kappa+1)}}{(j_\kappa+1)^2} \right). \quad (53)$$

Unfortunately, this bound is rather complex and challenging to reason with. For long simulation times T , a more useful bound can be derived that, in particular, is more directly comparable to the recent commutator bounds of Childs *et al.* [19].

Corollary 2. Suppose $T \in \mathbb{R}_+$ is sufficiently large such that

$$\frac{2a_{\text{max}} \Upsilon \alpha_{\text{comm}}^{(p+1)} T}{(4\Lambda)^p} \geq 1.$$

Then

$$\|\tilde{E}_{j+1}(T)\| \leq 2 \left[2\alpha_{\text{comm}}^{(p+1)1/p} (a_{\text{max}} \Upsilon T)^{1+1/p} \right]^j.$$

We provide a proof in Appendix C. This form of error leads directly to computational complexities in terms of $\alpha_{\text{comm}}^{(p+1)}$, which is useful for comparison with standard Trotter results.

More generally, we would like to ensure that $\lambda_{j,l}$ adequately captures the “size” of H , and does not grow too

large with the indices j, l . In particular, we will subsequently describe error bounds using the simpler parameter

$$\lambda := \sup_{\substack{j \in \sigma \mathbb{Z}_+ \geq \sigma m \\ 1 \leq l \leq K}} \lambda_{j,l} \quad (54)$$

but first, we must ensure λ exists.

Lemma 4. Let $\lambda_{j,l}$ be defined as in Eq. (53). Then $\lambda_{j,l} \leq 4 \sum_{\gamma=1}^{\Gamma} \|H_\gamma\|$.

Proof. Taking a triangle inequality through the expression in Lemma 6, and using a norm bound on $\alpha_{\text{comm}}^{(j)}$,

$$\lambda_{j,l} \leq 2 \left(\sum_{\gamma} \|H_\gamma\| \right) \left(\sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \prod_{\kappa=1}^l \frac{1}{(j_\kappa+1)^2} \right)^{1/(j+l)}. \quad (55)$$

Next,

$$\begin{aligned} &\left(\sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \prod_{\kappa=1}^l \frac{1}{(j_\kappa+1)^2} \right)^{1/(j+l)} \\ &\leq \left(\sum_{\substack{j_1 \dots j_l \in \mathbb{N} \\ j_1 + \dots + j_l = j}} 1 \right)^{1/(j+l)} = \binom{j+l-1}{l-1}^{1/(j+l)}. \end{aligned} \quad (56)$$

We can then use that $\binom{j+l-1}{l-1} \leq 2^{j+l-1} < 2^{j+l}$ to get that $\binom{j+l-1}{l-1}^{1/(j+l)} \leq 2$. Put together, this provides the stated bound. $\blacksquare$

It will be convenient to adopt two simpler bounds on $\tilde{E}_{j+1}$ depending on the evolution time T . From the bounds previously derived in Lemma 3 we can straightforwardly obtain a bound on the derivatives in terms of fundamental simulation parameters.

Corollary 3. For any $j \in \mathbb{Z}_+$, we have, for “short times” $a_{\max} \Upsilon \lambda T \leq 1$

$$\|\tilde{E}_{j+1}(T)\| \leq 2(a_{\max} \Upsilon \lambda T)^j,$$

and for “long times” $a_{\max} \Upsilon \lambda T > 1$,

$$\|\tilde{E}_{j+1}(T)\| \leq 2(a_{\max} \Upsilon \lambda T)^{j(1+1/p)}.$$

Proof. For convenience we reproduce the bound on $\|\tilde{E}_{j+1}(T)\|$ from Lemma 3

$$\begin{aligned} \|\tilde{E}_{j+1}(T)\| &\leq (a_{\max} \Upsilon T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{(a_{\max} \Upsilon T)^l}{l!} \\ &\quad \times \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{k=1}^l 2^{\frac{\alpha_{\text{comm}}^{(j_k+1)}}{(j_k+1)^2}} \right) \\ &\leq \sum_{l=1}^{\lfloor j/p \rfloor} \frac{(a_{\max} \Upsilon \lambda T)^{j+l}}{l!}, \end{aligned} \quad (57)$$

where λ is defined by Eq. (54). First, consider the case $a_{\max} \Upsilon \lambda T \leq 1$. We have

$$\begin{aligned} \|\tilde{E}_{j+1}(T)\| &\leq (a_{\max} \Upsilon \lambda T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{1}{l!} < (a_{\max} \Upsilon \lambda T)^j (e - 1) \\ &< 2(a_{\max} \Upsilon \lambda T)^j. \end{aligned} \quad (58)$$

This provides the first bound of the corollary. For the long-time case, we instead use the bound

$$\begin{aligned} \|\tilde{E}_{j+1}(T)\| &\leq (a_{\max} \Upsilon \lambda T)^{j+\lfloor j/p \rfloor} \sum_{l=1}^{\lfloor j/p \rfloor} \frac{1}{l!} \\ &< 2(a_{\max} \Upsilon \lambda T)^{j(1+1/p)} \end{aligned} \quad (59)$$

completing the lemma. $\blacksquare$

V. RICHARDSON EXTRAPOLATION

Having laid the technical groundwork in the previous section, we now apply these results to analyze Richardson extrapolation for time-evolved observables. Before we begin, we briefly provide a more detailed overview Richardson extrapolation to supplement Sec. II C 1. A more general discussion of the method can be found in Ref. [48].

In our context $f : [0, \ell] \rightarrow \mathbb{R}$ is a smooth function, such that

$$f(s) = f(0) + \sum_{j=1}^n c_j s^{\sigma_j} + O(s^{\sigma_{n+1}}), \quad (60)$$

where the $\sigma_j \in \mathbb{Z}_+$ form an increasing sequence. Moreover, suppose $f(s_k)$ can be computed for a monotonically decreasing sequence of inputs $s_k \in (0, \ell]$. Richardson extrapolation provides a new function $F^{(m)}(s)$ satisfying

$$F^{(m)}(s) = f(0) + \sum_{j=m}^n \tilde{c}_j s^{\sigma_j} + O(s^{\sigma_{n+1}}) \quad (61)$$

for some $1 \leq m \leq n$. Thus, the convergence rate to $f(0)$ for small s is boosted from $O(s^{\sigma_1})$ to $O(s^{\sigma_m})$. The procedure can be applied to C^m functions on $[0, c]$, and in this case the relevant expansion is the Taylor polynomial.

There are several algorithms for performing Richardson extrapolation. Regardless of how it is performed, the result is a linear combination

$$F^{(m)}(s) = \sum_{k=1}^m b_k f(s_k), \quad (62)$$

where the s_k are themselves functions of s , and $s = \max_k s_k = s_1$. Moreover, $b = (b_1, \dots, b_m)$ solves the linear system

$$Vb = \hat{e}_1, \quad (63)$$

where $\hat{e}_j$ is the j th standard basis vector and V is a generalized $m \times m$ Vandermonde matrix, with elements $V_{jk} = s_k^{\sigma_j - 1}$ and $\sigma_0 \equiv 0$. For certain values of σ_j , an exact solution to Eq. (63) is known. In particular, if $\sigma_j = \eta j$ for some $\eta \in \mathbb{Z} \setminus \{0\}$ and $j = 0, 1, \dots, m-1$, then V is the standard Vandermonde matrix in terms of $y_k \equiv s_k^\eta$. Thus, the inverse is known via the theory of Lagrange interpolation

$$b_k = (V^{-1})_{k0} = \prod_{i \neq k} \frac{s_i^\eta}{s_i^\eta - s_k^\eta}. \quad (64)$$

For more general σ_j , Eq. (63) still holds, but it is less clear if a closed-form solution exists in the mathematics literature. This has important implications for higher-order product-formula simulations that we will discuss later in this section.

A proper choice of sample points is crucial for the Richardson method to be well-conditioned, hence robust to computational imprecision. For example, certain natural choices, such as $s_k = s/k$ for $k > 0$, lead to poor conditioning. In our context, the relevant condition number is

the one norm $\|b\|_1$ of b , which can grow very large despite the summation constraint

$$\sum_k b_k = 1 \quad (65)$$

enforced by Eq. (63). This results in a sign problem, where small numerical errors in $f(s_k)$ are magnified by enormous b_k and hence large $\|b\|_1$, while the size of the answer $f(0)$ remains $O(1)$. Some choices of s_k , such as a geometric sequence $s_k = \omega^{k-1}s_1$ for $\omega \in (0, 1)$ are provably well conditioned. However, in many applications such as our own, computing s_k closer to 0 becomes intolerably expensive. To achieve low-depth Trotter evolutions, it is preferable to keep the sample points s_k as far from the origin as allowable.

Additionally, for product-formula simulation, we would also like to choose $s_k = 1/r_k$ for nonzero integer r_k , since otherwise we would have to apply fractional Trotter steps. Although this is possible using quantum signal-processing techniques, the overhead is potentially large and undesirable—particularly if using a NISQ-era device. s_1 is a parameter chosen appropriately to minimize the number of Trotter steps necessary while also ensuring sufficiently accurate simulations.

To meet the well conditioning and integer query conditions we desire, we make use of the well-conditioned extrapolation scheme of Ref. [23]. There the authors were concerned with extrapolations of the unitary evolution operator, known as a multiproduct formula, but the same scheme works for expectation values by linearity. In particular, we let

$$r_k = r_{\text{scale}} \left\lceil \frac{R}{\sin(\pi(2k-1)/8m)} \right\rceil, \quad k \in \{1, \dots, m\}. \quad (66)$$

At a high level, this choice comes from rounding a subset of the Chebyshev nodes to the nearest reciprocal integer. The well-conditioned property comes from orthogonality relations on the Chebyshev nodes; see the original work for more details. We will make the explicit choice $R = \sqrt{8m}/\pi$, chosen such that the r_k are distinct, and $r_{\text{scale}} \in \mathbb{Z}_+$ scales to ensure the r_k are large enough to accommodate longer evolutions. This scaling has no effect on b and hence the conditioning. A useful upper bound for the purpose of resource estimation is [60]

$$m \leq r_k/r_{\text{scale}} \leq 3m^2. \quad (67)$$

One can prove the following about this choice of extrapolation points.

Lemma 5 (Well-conditioned Richardson extrapolation [23]). Let $f \in C^{2m+2}([-1, 1])$ be an even, real-valued

function of s , and let P_j and R_j be the degree j Taylor polynomial and Taylor remainder, respectively, such that $f(s) = P_j(s) + R_j(s)$. Let

$$F^{(m)}(s) = \sum_{k=1}^m b_k f(s_k)$$

be the unique Richardson extrapolation of f at points $s_1, s_2, \dots, s_m$ given by

$$s_k = \frac{s}{r_k},$$

for r_k defined in Eq. (66) and b_k given in Eq. (64) for $\eta = 2$. Then

$$F^{(m)}(s) = f(0) + \sum_{k=1}^m b_k R_{2m}(s_k). \quad (68)$$

Moreover, $\|b\|_1 = O(\log m)$.

Proof. By the extrapolation properties of $F^{(m)}$, $\sum_{k=0}^{m-1} b_k P_{2m}(s_k) = f(0)$. Equation (68) follows immediately. The scaling of $\|b\|_1$ with m is proven in Theorem 1 of [23] for $r_{\text{scale}} = 1$. However, b is invariant under a change in scale. Indeed, if each r_k is multiplied by r_{scale} , then in Eq. (64) one has

$$b'_k = \prod_{i \neq k} \frac{r_{\text{scale}}^2 r_i^2}{r_{\text{scale}}^2 (r_i^2 - r_k^2)} = b_k. \quad (69)$$

Hence, for any $r_{\text{scale}} \neq 0$, $\|b\|_1 = O(\log m)$. ■

A. Application to time-evolved observables

Richardson extrapolation may be applied to the problem of computing time-evolved expectation values, where the Trotter step s is the extrapolation parameter and the function of interest is

$$f(s) = \langle O_p(T, s) \rangle := \text{tr}(\rho_0 \mathcal{P}^{1/s\dagger}(sT) O \mathcal{P}^{1/s}(sT)). \quad (70)$$

Here, p is the order of the product formula $\mathcal{P}$. Note that $f(0)$ may be defined via the limit, and corresponds to the ideal value.

In what follows, we will assume the extrapolation always starts from the linear or quadratic error terms, depending on the symmetry of the product formula. We do not start the extrapolation from the smallest nonzero power according to p . Although this may seem unnecessary for large p , we wish to eventually utilise the well-conditioned extrapolation of Lemma 5, which begins at second order. It is possible that these concerns are primarily academic, and practitioners may find it more reasonable to use an alternative scheme without the theoretical guarantees.

The following lemma characterizes the error in Richardson approach thus described, without committing yet to a specific choice of sampling points.

Lemma 6 (Richardson extrapolation error). Let O be an observable, $H = \sum_{\gamma}^{\Gamma} H_{\gamma}$ be a time-independent Hamiltonian. Let $\mathcal{P}$ be a staged p th-order product formula of symmetry class σ , where $\sigma = 2$ if $\mathcal{P}$ is symmetric, 1 otherwise. Let

$$\langle O_{p,m}(T) \rangle := \sum_{k=1}^m b_k \langle O_p(T, s_k) \rangle$$

be an m -term Richardson extrapolation with $\sigma m \geq p$, and ascending sequence of Trotter steps $r_k = 1/s_k \in \mathbb{Z}_+$. Let $T \in \mathbb{R}_+$ and suppose that the sequence $(a_{\max} \Upsilon s_1 T)^j \alpha_{\text{comm}}^{(j)}$ is bounded in j (which implies the same for all s_k). Then, the error in the extrapolation, as compared to the exact expectation value $\langle O(T) \rangle$, for any initial quantum state ρ_0 , satisfies

$$|\langle O(T) \rangle - \langle O_{p,m}(T) \rangle| \leq \|O\| \|b\|_1 \sum_{\substack{j \in \sigma \mathbb{Z}_+ \\ j \geq \sigma m}} s_1^j \|\tilde{E}_{j+1}(T)\|,$$

where $\|b\|_1 = \sum_k |b_k|$.

Proof. For any $s \in (0, s_1]$, the hypotheses of Lemma 3 are satisfied, and thus the error series

$$\langle \tilde{O}(T, s) \rangle = \langle O(T) \rangle + \sum_{j \in \sigma \mathbb{Z}_+ \geq p} s^j \langle \tilde{E}_{j+1}(T)(O) \rangle \quad (71)$$

is valid. The Richardson extrapolation procedure with m samples will remove all terms up to $O(s^{\sigma(m-1)})$ in the series, so that

$$\langle O_{p,m}(T) \rangle - \langle O(T) \rangle = \sum_{k=1}^m b_k \langle R_{\sigma(m-1)}(T, s_k)(O) \rangle, \quad (72)$$

where R_q is the Taylor remainder of degree q and satisfies

$$R_{\sigma(m-1)}(T, s) := \sum_{j \in \sigma \mathbb{Z}_+ \geq \sigma m} s^j \tilde{E}_{j+1}(T). \quad (73)$$

Applying Hölder's inequality,

$$\begin{aligned} |\langle O_{p,m}(T) \rangle - \langle O(T) \rangle| &\leq \|b\|_1 \max_{k \in [m]} \|R_{\sigma(m-1)}(T, s_k)(O)\| \\ &\leq \|b\|_1 \|O\| \max_{k \in [m]} \|R_{\sigma(m-1)}(T, s_k)\|, \end{aligned} \quad (74)$$

and we now focus on the remainder of Eq. (73). By the triangle inequality,

$$\|R_{\sigma(m-1)}(T, s)\| \leq \sum_{j \in \sigma \mathbb{Z}_+ \geq \sigma m} s^j \|\tilde{E}_{j+1}(T)\|. \quad (75)$$

To obtain the lemma statement we simply substitute $s = s_1$ and note that this produces maximum remainder bound over all s_k . ■

With the relevant error bounds in hand, we turn to the question of algorithmic cost to achieve an error within tolerance ϵ . Given Trotter steps $(r_1, \dots, r_m) \in \mathbb{Z}_+^m$ listed in ascending order, the maximum Trotter depth is r_m and the total Trotter depth is $\sum_k r_k$ for a single evolution. Multiple evolutions are generally necessary for the measurement.

We start by using our error bounds to derive a sufficient Trotter depth to achieve precision ϵ in the estimator, assuming exactly computed expectation values throughout.

Lemma 7 (Sufficient Trotter depth). Suppose s_1 satisfies the radius of convergence conditions of Lemma 6, and furthermore is chosen such that $s_1(a_{\max} \Upsilon \lambda T)^{1+1/p} < 1/2$. To achieve an error ϵ such that

$$|\langle O(T) \rangle - \langle O_{p,m}(T) \rangle| \leq \epsilon \|O\|,$$

it suffices to choose a minimum number of Trotter steps $r_1 = 1/s_1$ such that

$$r_1 \geq (a_{\max} \Upsilon \lambda T) \left(\frac{4\|b\|_1}{\epsilon} \right)^{\frac{1}{\sigma m}}$$

for $a_{\max} \Upsilon \lambda T \leq 1$ (“short times”) and

$$r_1 \geq (a_{\max} \Upsilon \lambda T)^{1+1/p} \left(\frac{4\|b\|_1}{\epsilon} \right)^{\frac{1}{\sigma m}}$$

for $a_{\max} \Upsilon \lambda T > 1$ (“long times”).

Proof. From Lemma 6,

$$|\langle O(T) \rangle - \langle O_{p,m}(T) \rangle| \leq \|O\| \|b\|_1 \sum_{j \in \sigma \mathbb{Z}_+ \geq \sigma m} s_1^j \|\tilde{E}_{j+1}(T)\|. \quad (76)$$

We then simply apply the two bounds of Corollary 3. For short times,

$$\begin{aligned} |\langle O(T) \rangle - \langle O_{p,m}(T) \rangle| &\leq 2\|O\| \|b\|_1 (s_1 a_{\max} \Upsilon \lambda T)^{\sigma m} \sum_{j \in \sigma \mathbb{N}} (s_1 a_{\max} \Upsilon \lambda T)^j, \end{aligned} \quad (77)$$

where we factored out the leading power. The remaining series is bounded by 2 by our assumptions on s_1 . To

achieve an relative error (i.e., normalized by $\|O\|$) of ϵ , it suffices then to choose $r_1 = 1/s_1$ satisfying

$$r_1 \geq a_{\max} \Upsilon \lambda T \left(\frac{4\|b\|_1}{\epsilon} \right)^{\frac{1}{\sigma m}}, \quad (78)$$

as claimed.

For $a_{\max} \Upsilon \lambda T > 1$, we proceed in an entirely parallel manner as the short time case: factoring out a leading σm power and noting that what remains in the series is bounded by 2. This directly leads to the condition stated in the lemma for long times, and concludes the proof. ■

So far we have derived a set of bounds for the minimum number of Trotter steps required to reach a given error. We now examine the asymptotic scaling of the parameters.

Corollary 4 (Asymptotic Richardson costs). Consider an m -term Richardson extrapolation of a time-evolved expectation value in the setting of the previous lemma, where $\mathcal{P}$ is symmetric. Choose the Trotter step size according to Eq. (66), with r_{scale} large enough such that r_1 satisfies the “long-time” condition of Lemma 7 for a choice of m scaling as $O(\log(1/\epsilon))$. Then, the maximum number of Trotter steps that needs to be implemented scales as

$$\max_k r_k = O\left((a_{\max} \Upsilon \lambda T)^{(1+1/p)} \log(1/\epsilon)\right).$$

Proof. By satisfying the conditions of Lemma 7, the extrapolation scheme achieves an ϵ relative to $\|O\|$, in the sense stated in that lemma. Since $\|b\|_1 = O(\log m)$, $\|b\|_1^{1/2m} = O(1)$. Choose $m = p \lceil \log(1/\epsilon) \rceil$. Then,

$$\left(\frac{4}{\epsilon} \right)^{\frac{1}{2p \lceil \log(1/\epsilon) \rceil}} = O(1). \quad (79)$$

Thus, from Lemma 7,

$$r_1 = O\left((a_{\max} \Upsilon \lambda T)^{(1+1/p)}\right) \quad (80)$$

is the minimum number of Trotter steps required to achieve error ϵ . To obtain an upper bound on the maximum number of steps, we utilize the bounds Eq. (67) to obtain

$$r_k \leq 3mr_1 = O\left((a_{\max} \Upsilon \lambda T)^{(1+1/p)} m\right) \quad (81)$$

given our choice of m , this yields the scaling stated in the corollary. ■

We remark that, in the above proof, the choice to make m a multiple of p is mainly for simplicity of proof. For our purposes, we treat p as fixed and not scaling with the simulation parameters.

B. Resource estimates for Richardson extrapolation

So far, we have determined error bounds for the Richardson procedure and provided partial results on the resources required. In this section, we will derive full resource costs. We examine two metrics: the maximum circuit depth $D_{\max}$ of Trotter steps needed, and the total number C_{Trot} of Trotter steps required. The former is arguably the most relevant metric for NISQ-era devices where only short-depth circuits are possible to implement, whereas C_{Trot} is more relevant for fault-tolerant devices. We note that C_{Trot} is proportional to the total number of elementary operations required for the protocol.

There are two primary sources of error intrinsic to our protocol: the extrapolation error (associated with the Richardson extrapolation procedure) and the error in the function values themselves. We suppose the estimates for the function f at points $\{s_k\}_{k=1}^m$ are given by $\tilde{f}(s_k)$, such that the final estimate we have is $\tilde{F}^{(m)}(s) = \sum_{k=1}^m b_k \tilde{f}(s_k)$. The error in our final prediction is $\epsilon = |f_B(0) - \tilde{F}^{(m)}(s)|$, which can be broken down as

$$\begin{aligned} & |f_B(0) - \tilde{F}^{(m)}(s)| \\ & \leq |f_B(0) - F^{(m)}(s)| + |F^{(m)}(s) - \tilde{F}^{(m)}(s)| \\ & \leq |f_B(0) - F^{(m)}(s)| + \left| \sum_{k=0}^{m-1} b_k f(s_k) - \sum_{k=0}^{m-1} b_k \tilde{f}(s_k) \right| \\ & \leq \|O\|(\epsilon_{\text{ext}} + \|b\|_1 \epsilon_{\text{data}}), \end{aligned} \quad (82)$$

where ϵ_{ext} is the (relative) interpolation error and $|f(s_k) - \tilde{f}(s_k)| \leq \|O\| \epsilon_{\text{data}}$ is the maximum (relative) error associated with each individual measurement point. To satisfy a total relative error tolerance ϵ , it thus suffices to ensure that

$$\epsilon_{\text{ext}} \leq \frac{\epsilon}{2}, \quad \epsilon_{\text{data}} \leq \frac{\epsilon}{2\|b\|_1}. \quad (83)$$

The resources required to satisfy the first of these inequalities is essentially the content of Corollary 4 (the factor of 1/2 is asymptotically irrelevant). To derive the theoretical results of this section and Sec. VI, we will assume that data errors only result from the measurement protocols on errorless quantum computers. The effects of hardware imperfections can be straightforwardly incorporated, and this is discussed in Sec. VII A.

1. Incoherent measurements

Within the incoherent scheme, let us consider how many measurements are needed to ensure $\epsilon_{\text{data}} \leq \epsilon/(2\|b\|_1)$. From Hoeffding’s inequality, we see that to achieve an estimate $\langle O \rangle'$ satisfying

$$|\text{tr}[\mathcal{P}^{\dagger 1/s}(sT) O \mathcal{P}^{1/s}(sT) \rho_0] - \langle O \rangle'| \leq \epsilon_{\text{data}} \|O\| \quad (84)$$

with probability $\geq (1 - \delta')$, it suffices to choose a number of samples N satisfying

$$N \geq \frac{1}{2\epsilon_{\text{data}}^2} \log \left(\frac{2}{\delta'} \right). \quad (85)$$

By the union bound, it suffices to choose $\delta' = \delta/m$ to have an overall success probability of $1 - \delta$ for all measurements. Taking $m = O(\log(1/\epsilon))$ from Corollary 4, we have that N scales as

$$\begin{aligned} N &= O \left(\frac{\|b\|_1^2}{\epsilon^2} \log \left(\frac{m}{\delta} \right) \right) = O \left(\frac{(\log m)^3}{\epsilon^2} \right) \\ &= O \left(\frac{(\log \log 1/\epsilon)^3}{\epsilon^2} \right). \end{aligned} \quad (86)$$

Each measurement requires one product-formula evolution. The maximum Trotter step size is given in Corollary 4, and this directly gives the maximum Trotter depth. Meanwhile, the total number of Trotter steps required (which is proportional to the total resources) scales as

$$\begin{aligned} C_{\text{Trot}} &\leq N \sum_{k=1}^m r_k \\ &\leq Nr_1 \sum_{k=1}^m (r_k/r_1) \\ &\leq Nr_1 \frac{1}{m} \sum_{k=1}^m \left\lceil \frac{R}{\sin(\pi(2k+1)/8m)} \right\rceil, \end{aligned} \quad (87)$$

where we have used that $1/r_1 \leq 1/(r_{\text{scale}}m)$ from Eq. (67). Next, we have

$$\begin{aligned} &\leq Nr_1 \frac{1}{m} \sum_{k=1}^m \left(\frac{R}{\sin \left(\frac{\pi(2k+1)}{8m} \right)} + 1 \right) \\ &\leq Nr_1 \left(\sum_{k=1}^m \frac{8R}{\pi(2k+1)} + 1 \right) \\ &\leq Nr_1 \left(R \frac{4}{\pi} (2 + \log m) + 1 \right) \\ &= O(N(a_{\text{max}} \Upsilon \lambda T)^{1+1/p} m \log(m)). \end{aligned} \quad (88)$$

Using the scaling from Eq. (86), and $m = O(\log(1/\epsilon))$, we arrive at our main result concerning the use of Richardson extrapolation with incoherent measurements.

Theorem 2 (Resource costs for incoherent measurements). Let $\langle O_{p,m}(T) \rangle$ be the m -term Richardson extrapolation estimate for $\langle O(T) \rangle$, taken by varying the step size of a p th-order staged product formula, with samples taken

at the rescaled Chebyshev nodes as specified in Eq. (66). The resource costs for computing this estimate such that

$$|\langle O(T) \rangle - \langle O_{p,m}(T) \rangle| \leq \epsilon \|O\|,$$

when using incoherent measurements uses a number of sample points $m = O(1/\epsilon)$. Moreover, the maximum Trotter depth and total Trotter steps scales as

$$\begin{aligned} D_{\text{max}} &= O((a_{\text{max}} \Upsilon \lambda T)^{1+1/p} \log(1/\epsilon)), \\ C_{\text{Trot}} &= O \left(\frac{(a_{\text{max}} \Upsilon \lambda T)^{1+1/p}}{\epsilon^2} \log(1/\epsilon) (\log \log(1/\epsilon))^4 \right). \end{aligned}$$

Here $\lambda \leq 4 \sum_{\gamma} \|H_{\gamma}\|$ is defined in Lemma 7 and there is a failure probability of $\delta = 0.01$.

2. Coherent measurements

Although incoherent measurements are conceptually simple and easy to implement, they give $O(1/\epsilon^2)$ scaling rather than the optimal Heisenberg scaling of $O(1/\epsilon)$. In this section, we consider the iterative quantum amplitude estimation (IQAE) scheme developed by Grinko *et al.* for this problem. Although the method deviates by $O(\log \log 1/\epsilon)$ from the ideal Heisenberg scaling, it achieves better constant factors among rigorously analyzed methods with true Heisenberg scaling that do not require standard quantum phase estimation (QPE), a relatively intensive routine with large qubit overhead [51]. Thus, the method represents a practical yet rigorous algorithm suitable for our purposes. We follow a similar analysis to Ref. [45].

We assume the problem of measuring the expectation value can be written as the problem of estimating an amplitude, as is often done in practice via, say, the Hadamard test. Assuming $\|O\| \leq 1$ (if otherwise, we can rescale it), then performing a Hadamard test gives an amplitude $(1 + \langle O_p(T, s) \rangle)/2$. This amplitude can now be estimated using IQAE. In particular, for each sample point s_k , we need a number of calls to a Grover oracle N_G scaling as Ref. [45, Section E]

$$N_G \leq \frac{100}{\epsilon_{\text{data}}} \log \left(\frac{2m}{\delta} \log \left(\frac{\pi}{\epsilon_{\text{data}}} \right) \right), \quad (89)$$

where δ/m is the probability of failure per data point, which by the union bound ensures an overall success rate of $1 - \delta$. For a particular value of Trotter step s , we need a total Trotter depth scaling as N_G/s . The maximum Trotter step size needed to compute $F^{(m)}(s)$ is $\max_k r_k$, which is characterized asymptotically in Corollary 4. Thus, the

maximum Trotter depth is

$$\begin{aligned}
D_{\max} &\leq N_G \max_k r_k \\
&= O\left((a_{\max} \Upsilon \lambda T)^{1+1/p} \log(1/\epsilon) \frac{\|b\|_1}{\epsilon}\right. \\
&\quad \times \log\left(\frac{\log(1/\epsilon)}{\delta} \log\left(\frac{\|b\|_1}{\epsilon}\right)\right)\Bigg) \\
&= O\left(\frac{(a_{\max} \Upsilon \lambda T)^{1+1/p}}{\epsilon} \log(1/\epsilon) (\log \log(1/\epsilon))^2\right).
\end{aligned}$$

Meanwhile, the total number of Trotter steps, C_{Trot} , is bounded as

$$C_{\text{Trot}} \leq N_G \sum_{k=1}^m r_k. \quad (90)$$

Using the same reasoning as in Eq. (88), with N replaced by N_G , we obtain

$$\begin{aligned}
C_{\text{Trot}} &= O(N_G (a_{\max} \Upsilon \lambda T)^{1+1/p} m \log m) \\
&= O\left(\frac{(a_{\max} \Upsilon \lambda T)^{1+1/p}}{\epsilon} \log(1/\epsilon) (\log \log(1/\epsilon))^3\right).
\end{aligned} \quad (91)$$

We summarize the results of the above analysis in our main result for Richardson extrapolation using coherent measurements.

Theorem 3 (Resource costs for coherent measurements). Let $\langle O_{p,m}(T) \rangle$ be the m -term Richardson extrapolation estimate for $\langle O(T) \rangle$, taken by varying the step size of a p th-order symmetric staged product formula, with samples taken at the rescaled Chebyshev nodes as specified in Eq. (66). The resource costs for computing this estimate such that

$$|\langle O(T) \rangle - \langle O_{p,m}(T) \rangle| \leq \epsilon \|O\|,$$

when using coherent measurements is given by

$$\begin{aligned}
D_{\max} &= O\left(\frac{(a_{\max} \Upsilon \lambda T)^{1+1/p}}{\epsilon} \log(1/\epsilon) (\log \log(1/\epsilon))^2\right), \\
C_{\text{Trot}} &= O\left(\frac{(a_{\max} \Upsilon \lambda T)^{1+1/p}}{\epsilon} \log(1/\epsilon) (\log \log(1/\epsilon))^3\right),
\end{aligned}$$

where $\lambda \leq \sum_{\gamma} \|H_{\gamma}\|$ is defined in Lemma 7 and there is a failure probability of $\delta = 0.01$. Moreover, $m = O(\log(1/\epsilon))$.

VI. POLYNOMIAL INTERPOLATION

We now consider the polynomial interpolation algorithm of Rendon *et al.* for mitigating Trotter errors. Let $f(s) \in C^m[-\ell, \ell]$ be a real-valued function, and suppose we have the value of f at points $s_1, s_2, \dots, s_m$. Let $P_{m-1}f$ be the unique degree $m-1$ polynomial interpolating f at the s_k . It is possible to show [61] that the error of the approximating polynomial in the interval $[-\ell, \ell]$ is bounded as

$$|f(s) - P_{m-1}f(s)| \leq \max_{\xi \in [-\ell, \ell]} \frac{|f^{(m)}(\xi)|}{m!} |\omega_m(s)|,$$

where

$$\omega_m(s) := \prod_{k=1}^m (s - s_k)$$

is the monic nodal polynomial. If we choose our samples to be taken at the Chebyshev nodes on $[-\ell, \ell]$, given by

$$s_i = \ell \cos\left(\frac{2i-1}{2m}\pi\right), \quad i \in \{1, 2, \dots, m\}, \quad (92)$$

then the interpolation satisfies a number of nice properties, such as robustness to errors in the interpolation values. As in the Richardson extrapolation case, we are interesting in bounding the error at $s = 0$. To do so, we can use the following lemma.

Lemma 8 (Lemma 2, [45]). Let $s_1, s_2, \dots, s_m$ be the collection of Chebyshev interpolation points on the interval $[-\ell, \ell]$. Then the error of the interpolating polynomial $P_{m-1}f$ with respect to f at $s = 0$ is bounded as

$$|f(0) - P_{m-1}f(0)| \leq \max_{s \in [-\ell, \ell]} |f^{(m)}(s)| \left(\frac{\ell}{2m}\right)^m.$$

A. Application to time-evolved observables

From Lemma 8, we see that the key property in bounding the error to polynomial interpolation is to bound the derivative $f^{(m)}$ on $[-\ell, \ell]$. We have already developed the requisite tools in Sec. IV.

Theorem 4 (Polynomial interpolation error). Consider a Chebyshev interpolation $P_{m-1}f(s)$ of the time-evolved expectation value $f(s)$ on the interval $[-\ell, \ell]$, for long simulation time $a_{\max} \Upsilon \lambda T > 1$, with

$$\ell = \frac{1}{2} (a_{\max} \Upsilon \lambda T)^{-(1+1/p)}.$$

The approximation error at $s = 0$ may be bounded as

$$\frac{|f(0) - P_{m-1}f(0)|}{\|O\|} \leq c e^{-\gamma m},$$

where $\gamma > 1.5$ and $c < 11$.

Proof. From Corollary 1, the Taylor series for $\tilde{O}(T, s)$ is given by

$$\tilde{O}(T, s) = O(T) + \sum_{j \in \sigma \mathbb{Z}, j \geq p} s^j \tilde{E}_{j+1}(T)(O).$$

The m th derivative of $\tilde{O}$ is then given by

$$\begin{aligned} \partial_s^m \tilde{O}(T, s) &= \sum_{\substack{j \in \sigma \mathbb{Z} \\ j \geq \max\{p, m\}}} \frac{j!}{(j-m)!} s^{j-m} \tilde{E}_{j+1}(T)(O) \\ &= m! \sum_{\substack{j \in \sigma \mathbb{Z} \\ j \geq \max\{p, m\}}} \binom{j}{m} \tilde{E}_{j+1}(T)(O) s^{j-m}. \end{aligned} \quad (93)$$

Applying the triangle inequality and noting that $|s| \leq \ell$,

$$\begin{aligned} \max_{s \in [-\ell, \ell]} \|\partial_s^m \tilde{O}(T, s)\| &\leq m! \|O\| \sum_{\substack{j \in \sigma \mathbb{Z} \\ j \geq \max\{p, m\}}} \binom{j}{m} \ell^{j-m} \|\tilde{E}_{j+1}(T)\|. \end{aligned} \quad (94)$$

Using Corollary 3 in the long-time regime,

$$\begin{aligned} \max_{s \in [-\ell, \ell]} \|\partial_s^m \tilde{O}(T, s)\| &\leq 2 \frac{m!}{\ell^m} \|O\| \sum_{\substack{j \in \sigma \mathbb{Z} \\ j \geq \max\{p, m\}}} \binom{j}{m} (\ell(a_{\max} \Upsilon \lambda T)^{1+1/p})^j. \end{aligned} \quad (95)$$

We now choose $\ell = \frac{1}{2}(a_{\max} \Upsilon \lambda T)^{-(1+1/p)}$. Doing so, the infinite series is given by

$$\sum_{\substack{j \in \sigma \mathbb{Z} \\ j \geq \max\{p, m\}}} \binom{j}{m} \frac{1}{2^j} \quad (96)$$

and may be upper bounded by

$$\sum_{j=m}^{\infty} \binom{j}{m} \frac{1}{2^j} = 2. \quad (97)$$

Thus, the derivative is upper bounded as

$$\max_{s \in [-\ell, \ell]} \|\partial_s^m \tilde{O}(T, s)\| \leq 4 \|O\| \frac{m!}{\ell^m}. \quad (98)$$

Applying Lemma 8 gives

$$|f(0) - P_{m-1}f(0)| \leq 4 \|O\| \frac{m!}{(2m)^m}. \quad (99)$$

Using a simplified Stirling-type upper bound $m! < \sqrt{2\pi m} (m/e)^m e^{1/(12m)}$, one obtains

$$|f(0) - P_{m-1}f(0)| \leq 4\sqrt{2\pi} e^{1/12} \sqrt{m} (2e)^{-m} \|O\| < c e^{-\gamma m} \|O\|, \quad (100)$$

where $\gamma \equiv \ln 2 + 1 - 1/(2e) \approx 1.509$ and $c \equiv 4\sqrt{2\pi} e^{1/12} \approx 10.9$. This immediately leads to the statement of the lemma. ■

It is interesting to note that, while the factor of $1/2$ in ℓ was chosen for simplicity, other choices will lead to different values of c, γ . In particular, a factor approaching 1 should make the exponential decay more shallow and increase the constant factor. However, it is interesting that the form of the error dependence remains exponentially decaying.

We now have the ability to estimate the number of Trotter steps necessary to achieve a certain accuracy in the interpolation, assuming the values $f(s_k)$ are computed exactly.

Lemma 9. In the setting of Theorem 4, to achieve a relative error ϵ such that

$$|O(T) - P_{m-1}f(0)| \leq \epsilon \|O\|,$$

it suffices to choose $m = O(\log(1/\epsilon))$ and $\ell = \frac{1}{2}(a_{\max} \Upsilon \lambda T)^{-(1+1/p)}$. This interpolation protocol requires a maximum number of Trotter steps

$$\max_k r_k = O((a_{\max} \Upsilon \lambda T)^{(1+1/p)} \log(1/\epsilon)).$$

Proof. From Theorem 4, to achieve relative error ϵ , it suffices to choose m such that

$$c e^{-\gamma m} \leq \epsilon, \quad (101)$$

i.e., $m = O(\log(1/\epsilon))$. To find the maximum number of Trotter steps needed for a particular sampling point, we see from Eq. (92) that the smallest s_k is given by $k = m/2$ and is of size $O(m^{-1})$. Hence,

$$\max_k r_k = O(\ell m) = O((a_{\max} \Upsilon \lambda T)^{(1+1/p)} \log(1/\epsilon)). \quad (102)$$

■

B. Resource estimates

As for the Richardson extrapolation case, for polynomial interpolation we will have to deal with both the algorithmic extrapolation error ϵ_{ext} and the measurement error ϵ_{data} . We assume that each $f(s_i)$ is measured with

error $\leq \epsilon_{\text{data}}$, i.e., we measure a set of data points $\{f_i\}_i$ such that $|f(s_i) - \tilde{f}_i| \leq \epsilon_{\text{data}}$. Let $\tilde{P}_m f$ be the polynomial obtained from fitting to the data $\{\tilde{f}_i\}_i$, then it can be shown [50] that

$$\max_{s \in [-\ell, \ell]} |P_m f(s) - \tilde{P}_m f(s)| \leq L_m \epsilon_{\text{data}},$$

where L_m is the Lebesgue constant, which for Chebyshev interpolation is bounded by $2/\pi \log(m+1) + 1$.

Thus, in terms of the total error, ϵ , we can make the partition

$$\epsilon_{\text{ext}} = \frac{\epsilon}{2}, \quad \epsilon_{\text{data}} = \frac{\epsilon}{2L_m}.$$

To find the total resource costs for the polynomial interpolation methods, observe that the analysis is identical to the analysis performed in Secs. VB 1 and VB 2.

1. Stability to imperfect Chebyshev nodes

As in the Richardson extrapolation case, when performing polynomial interpolation, we wish to ensure our samples are taken from inverse integer values, which are at or close to the ideal Chebyshev nodes. However, in Appendix B, we see that by choosing a slightly larger value of ℓ , we get robustness to this sampling error but with a new Lebesgue constant L'_m , which satisfies $L'_m \leq 2L_m$. Thus, rather than sampling from the Chebyshev nodes exactly, we can instead sample from the closest inverse integer. The robustness condition is satisfied by choosing

$$\begin{aligned} 1/\ell &= O((a_{\text{max}} \Upsilon \lambda T)^{1+1/p} m^2 \log(m)) \\ &= O((a_{\text{max}} \Upsilon \lambda T)^{1+1/p} \log^2(1/\epsilon) \log \log(1/\epsilon)). \end{aligned} \quad (103)$$

Thus the maximum number of Trotter steps scales as

$$O((a_{\text{max}} \Upsilon \lambda T)^{1+1/p} \log^3(1/\epsilon) \log \log(1/\epsilon)).$$

Considering the total number C_{Trot} of Trotter steps only changes the overall scaling by $\log \log$ factors in $1/\epsilon$, which we neglect. We summarize these findings in the following theorem.

Theorem 5 (Resource costs for polynomial interpolation). Let $\tilde{P}_m f$ be the m -degree polynomial fit taken at the inverse integers closest to the Chebyshev nodes in the interval $[-\ell, \ell]$. The resource requirements to achieve error

$$|\langle O(T) \rangle - \tilde{P}_m f(0)| \leq \epsilon \|O\|$$

for incoherent and coherent measurement scales as the following table.

Scaling	Max depth, D_{max}	Total resources, C_{Trot}
Coherent	$\tilde{O}\left(\frac{(a_{\text{max}} \Upsilon \lambda T)^{1+1/p}}{\epsilon} \log^3(1/\epsilon)\right)$	$\tilde{O}\left(\frac{(a_{\text{max}} \Upsilon \lambda T)^{1+1/p}}{\epsilon} \log^3(1/\epsilon)\right)$
Incoherent	$\tilde{O}\left((a_{\text{max}} \Upsilon \lambda T)^{1+1/p} \log^3(1/\epsilon)\right)$	$\tilde{O}\left(\frac{(a_{\text{max}} \Upsilon \lambda T)^{1+1/p}}{\epsilon^2} \log^3(1/\epsilon)\right)$

Here, $\lambda \leq 4 \sum_{\gamma} \|H_{\gamma}\|$ is defined in Lemma 7, $\tilde{O}$ hides $\log \log$ factors, and there is a failure probability of $\delta = 0.01$. Moreover, $m = O(\log(1/\epsilon))$.

VII. APPLICATIONS

A. Error resilience

In the algorithmic analysis above, we made the standard theoretical assumption of faultless quantum operations. At the present time, errors due to hardware imperfections are an unavoidable and significant difficulty in quantum computing. We argue that the reduced circuit depths of our extrapolation algorithms lead to improved robustness to hardware errors compared with standard product formulas, even without implementing hardware error mitigation.

Noise effects can be accounted for simply by returning to Eq. (82) and noting that ϵ_{data} may consist of both measurement and intrinsic hardware error: $\epsilon_{\text{data}} = \epsilon_{\text{meas}} + \epsilon_{\text{hw}}$. We model the effects of hardware noise with a single parameter η that gives a systematic error “per unit length” of the circuit. Thus, a product formula in r_j steps will accumulate error going as ηr_j . Of course, this linear approximation should only be expected for timescales for which $\eta \lambda T$ is small. However, considering the expectation value $f(\eta)$ as a function of the noise, our linear approximation is an upper bound on the error when f is convex or concave in η and converges as $\eta \rightarrow \infty$. This occurs in an exponential noise model, for example. Hence, our model captures the essential features of the noise, or at least upper bounds them.

For concreteness, let us consider the Richardson protocol. Under the above assumptions, the error in the final estimate due to hardware can be at worst

$$\eta \sum_j |b_j| r_j \leq \eta \|b\|_1 r_{\text{max}}. \quad (104)$$

Recall that for long simulation times, $r_j = O((\lambda T)^{1+1/p})$ are required even in the absence of hardware noise. If ϵ is the desired precision of the estimator, then we find

$$\eta \leq \tilde{O}\left(\frac{\epsilon}{(\lambda T)^{1+1/p}}\right) \quad (105)$$

suffices to ensure an overall relative error $O(\epsilon)$ in the expectation value. This noise tolerance compares favorably with Trotter alone, for which r_{max} goes as $O(\epsilon^{-1/p})$ leading to an extra factor of $\epsilon^{1/p}$ in Eq. (105). While the

above analysis is quite general, it is perhaps most relevant to systematic errors (e.g., decoherence), since unbiased random errors could be mitigated by increasing the number of measurements.

B. Improvements to polynomial algorithm

In Sec. III A we have highlighted the improvements this work provides to the original polynomial interpolation algorithm of Ref. [45]: commutator scaling, improved T dependence, and an elimination of fractional queries in the algorithm. Here we illustrate how these features lead to improved resource estimates in regimes of interest.

1. Estimating local observables

In their important work on Trotter error, Childs *et al.* showed that commutator-dependent Trotter error bounds can be significantly tighter than norm bounds for local Hamiltonians and other physically relevant systems [19], corresponding to improvements in algorithmic complexity. As an example, for geometrically local Hamiltonians on a lattice of n qubits, one has that $\alpha_{\text{comm}}^{(p+1)} = O(n)$. In contrast, a simple norm bound on H satisfies $\sum_{\gamma} \|H_{\gamma}\| = O(n)$, leading to error scaling as $O(n^{p+1})$ for constant T, ϵ .

Our error generally depends on a different commutator than their work; however, Corollary 2 leads to an error dependence on $\alpha_{\text{comm}}^{(p+1)}$ provided T is sufficiently large. For example, for the symmetric second-order Trotter formula ($p = 2$), one requires $T = \Omega(n)$ to satisfy the conditions of the corollary. For a constant dimension d of the lattice, this is a natural timescale to consider. In this regime, one may replace λ in our main results with $\sqrt{\alpha_{\text{comm}}^{(3)}} = O(\sqrt{n})$. For $T = \Theta(n)$ this leads to an overall complexity of $O(n^2)$,

taking $\epsilon \in O(1)$. In contrast, the prior bound in Ref. [45] of $O(T^2 \Gamma^2 \max_{\gamma} \|H_{\gamma}\|^2)$, which leads to a $O(d^2 n^4)$ scaling. Looking at n alone, then, our bounds lead to quadratic improvement over the original bounds of Rendon *et al.*

2. Nuclear physics

As another example, we consider the bounds for a model of nuclear interactions known as pionless effective field theory, which was studied in the work [5]. Using the bounds from Corollary 2 and combining with the commutator bounds developed in Ref. [5], we can compare this to the bounds in Ref. [45, Section 4]. By considering the minimum number of Trotter steps needed when $p = 1$ and $p = 2$ product formulas with extrapolation, we can compare this to the resource estimates from Ref. [45] and plot the ratio of these bounds in Fig. 3. (Taking the ratio removes the need to specify the desired error and probability of success.) This demonstrates that even the new bounds in this work drastically outperform previous bounds, and hence give improved performance guarantees.

We note that because the polynomial interpolation algorithm in this paper is similar to that in Ref. [45], in practice we expect empirically similar performance once an initial sampling point is chose. However, our work shows that we can drastically reduce the resource costs by orders of magnitude to achieve certain error guarantees when choosing the sampling points, thereby reducing resources of users of the algorithm. Put more simply, the work in this paper gives users significantly stronger error guarantees.

C. Numerics: Heisenberg chain

As a final example, we consider Richardson and polynomial extrapolation of local observables on a random

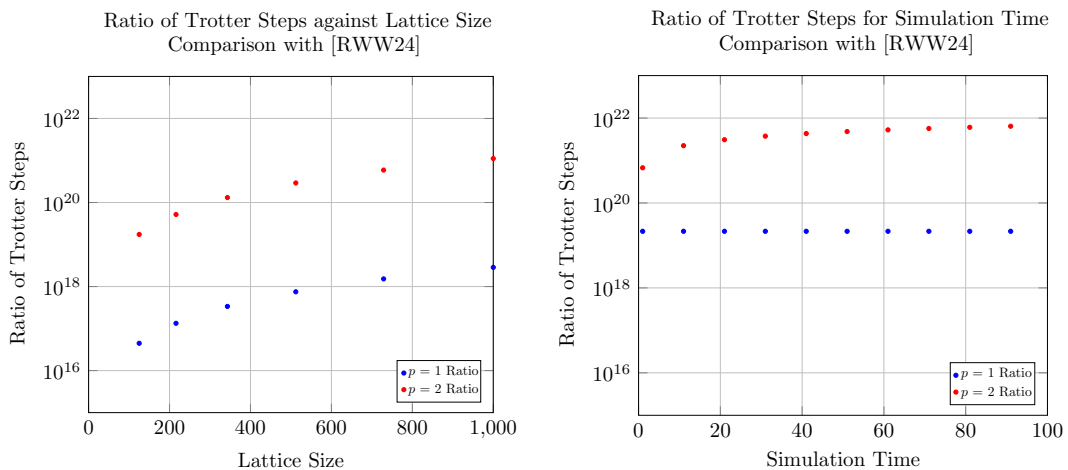


FIG. 3. Left: ratios of the number of Trotter steps $p = 1$ and $p = 2$ bounds in this work vs Ref. [45] needed to simulate the nuclear EFT model. Cost estimates are for a time period of the crossing time of a single nucleon with 10 MeV of energy. Different lattice sizes are considered. Right: the same as left, but where the simulation time is varied and the lattice size is held constant on a $10 \times 10 \times 10$ lattice.

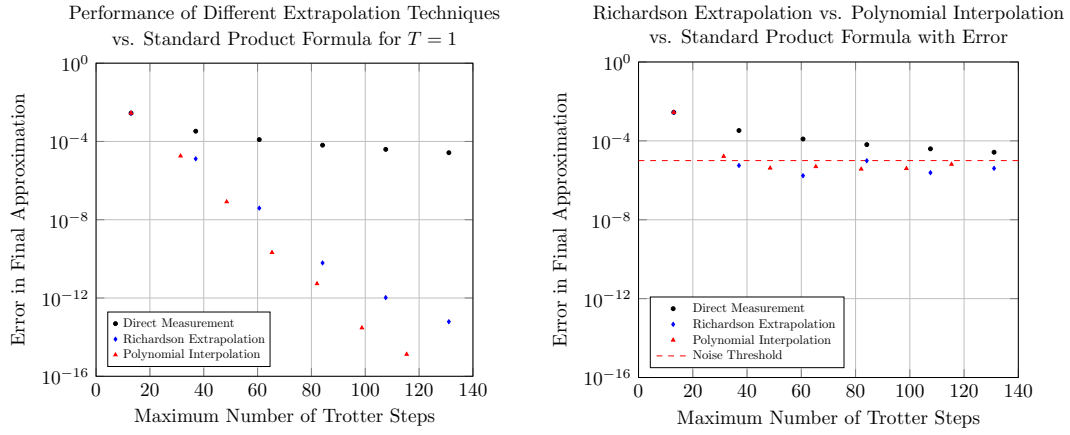


FIG. 4. Left: error comparison between the observable measured on the time-evolved state using Trotterization (black circles) vs the extrapolated error (blue diamonds and red triangles) for different maximum numbers of Trotter steps per run on a system of six qubits. The initial number of steps corresponds to approximately $(\Lambda T)^{3/2}$. We see the plot levels out at the bottom due to floating point precision. Right: the same as left, but where the extrapolation procedures are performed with a measurement error of 10^{-6} for each measurement. We see that the extrapolation procedures are limited by the measurement error.

Heisenberg chain. The Hamiltonian is given by

$$H = \sum_{i=1}^{L-1} (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1}) + \sum_{i=1}^L h_i Z_i,$$

where $\{h_i\}_{i=1}^L$ are sampled independently and uniformly over $[-1, 1]$. This Hamiltonian is well studied in the context of Trotter simulation [32] as it is a simple system with nontrivial Trotter error, and captures the essence of interesting condensed-matter phenomena such as localization. For all simulations shown, we consider the common $p = 2$ symmetric Trotter formula on $L = 6$ qubits. For the initial state, we choose a randomly chosen classical bit string: $|x\rangle$, $x \in \{0, 1\}^n$ and we choose a randomly chosen sum of three Pauli strings to act as the observable. In order to generate the figures in the following section, we have used a minimum number of Trotter steps $r = 1/s_{\min} = \lceil (\Lambda T)^{3/2} \rceil$. Here $\Lambda = \sum_{\gamma} \|H_{\gamma}\|$, and the decomposition H_{γ} is the naïve one along the Pauli terms.

We consider the error associated with choosing a fixed maximum circuit depth that we can utilize (e.g., if we are limited by noise in the physical circuit), and then use this to compare results from time simulation. That is, suppose we want to predict $\langle O(T) \rangle$ but are restricted to some maximum circuit depth (i.e., Trotter steps). How does the unextrapolated error compare to the extrapolated error?

Figure 4 (left) shows that if we consider a fixed time and as we increase the number of steps from a minimum value of $r = \lceil (\Lambda T)^{3/2} \rceil$, using Trotter extrapolation drastically reduces the error compared with just measuring the state directly. However, as demonstrated by Fig. 4 (right), the benefits of this are limited by the error associated by the measurements at each point. As one might expect, neither extrapolation techniques can improve results beyond the

limit to which we measure the observable we are extrapolating. We can also consider the case where we run the simulation for different amounts of time and see how the error scales with the number of nodes. From Fig. 5 we see that the extrapolation holds for much longer simulation times. We point the reader to Ref. [62, Chapter 3] for similar numerics [63].

VIII. CLASSICAL SHADOWS

In this section, we consider the estimation of many time-evolved expectation values using the classical shadow technique. To use classical shadows alongside extrapolation, simply run a classical shadow protocol on each of

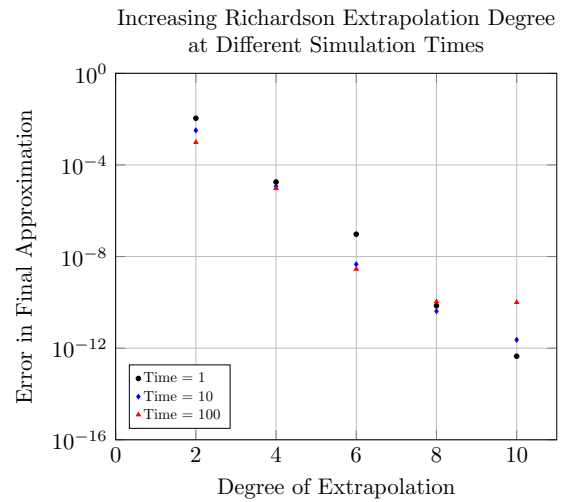


FIG. 5. A comparison in the performance of Richardson extrapolation as a function of the Richardson extrapolation degree for different simulation times.

Shadow Extrapolation

Input: A set of local Hamiltonian terms $\{h_j\}_j$, and a list of time-step sizes T_{list} , and an overall simulation time. A set of local observables $\{O_i\}_{i=1}^M$

Oracles: $\text{SHADOW}_p(|\psi\rangle, \{O_i\}_{i=1}^M, N_{\text{shadow}}, T, t, \{h_j\}_j)$ which runs a classical shadow routine using N_{shadow} samples, on a state $|\psi\rangle$ which has been time-evolved under the p^{th} -order Trotter operator for time T and time-step size t . It outputs a list of estimates for each of the $\langle\psi| e^{iHT} O_i e^{-iHT} |\psi\rangle$ which is calculated using N_{shadow} copies of the evolved-state.

Output: A set of estimates $\tilde{O}_i$ for each local observable expectation values.

1. $N_{\text{shadow}} \leftarrow \left\lceil \frac{128 \max_j \|O_j\|^2}{\epsilon^2} \log \left(\frac{2M|T_{\text{list}}|}{\delta} \right) \right\rceil$
2. for t_k in T_{list} :
 - (a) $\{\tilde{O}_j(t_k)\}_{j=1}^M \leftarrow \text{SHADOW}_p(|\psi\rangle, \{O_i\}_{i=1}^M, N_{\text{shadow}}, T, t_k, \{h_j\}_j)$
3. for O_j in $\{O_j\}_{j=1}^M$:
 - (a) for t_k in T_{list} :
 - i. $b_k \leftarrow \prod_{\substack{t_j \in T_{\text{list}} \\ t_j \neq t_k}} \frac{1}{1 - (t_k/t_j)}$
 - (b) $\tilde{O}_{j,\text{est}} \leftarrow \sum_{k=1}^{|T_{\text{list}}|} b_k \tilde{O}_j(t_k)$
4. Return $\{\tilde{O}_{j,\text{est}}\}_{j=1}^M$

FIG. 6. Pseudocode for shadow extrapolation, using Richardson extrapolation for concreteness. For the implementation of the SHADOW_p oracle we refer readers to Ref. [64].

the m states $\{\rho_k\}_{k=1}^m$ resulting from a Trotter evolution with inverse step size s_k . Then compute the expectation values of interest using these shadows and perform the extrapolation. We call this procedure *shadow extrapolation* for brevity. Pseudocode for shadow extrapolation, using Richardson extrapolation for concreteness, is given in Fig. 6. Polynomial interpolation can also be used. We note that this procedure is essentially allowing us to reconstruct a classical shadow representation of the final state of the un-Trotterized time evolution.

Classical shadows requires repeated computational basis measurements on the time-evolved states ρ_k , which is cheap but limits us to shot-noise accuracy. From Ref. [64], to estimate the expectation value of M local observables $\{O_i\}_{i=1}^M$ to accuracy $|\text{tr}[\rho O_i] - \tilde{O}_i| \leq \epsilon_{\text{data}} \|O_i\|$ with probability $\geq (1 - \delta)$, it suffices to take a number of samples of ρ_k , which is bounded by

$$\leq \frac{128}{\epsilon_{\text{data}}^2} \max_i \|O_i\|^2 \log \left(\frac{2Mm}{\delta} \right),$$

where m is the order of the Richardson extrapolator. Each of these samples requires a single Trotter evolution whose cost is determined by k (the number of Trotter steps). The following theorem is a simple corollary of the above results on Richardson and polynomial extrapolation. We note that shadow extrapolation is limited to $O(1)$ -local observables as opposed to the extrapolation procedures given earlier, which apply to general observables.

Theorem 6. Let $\{O_i(T)\}_{i=1}^M$ be a set of M $O(1)$ -local time-evolved observables $O(T) = e^{iHT} O e^{-iHT}$. To get an estimate $|\text{tr}(\rho O_i(T)) - \tilde{O}_i| \leq \epsilon_{\text{data}} \|O_i\|$ with probability $\geq (1 - \delta)$, the maximum Trotter depth scales as

$$D_{\text{max}} = O((\lambda T)^{1+1/p} \text{polylog}(1/\epsilon)),$$

where we need to run experiments at no more than $m = O(\log(1/\epsilon))$ sampling points. Meanwhile, the total number of Trotter steps for the entire protocol scales as

$$C_{\text{Trot}} = \tilde{O} \left(\frac{1}{\epsilon^2} (\lambda T)^{1+1/p} \text{polylog} \left(\frac{1}{\epsilon} \right) \log \left(\frac{M}{\delta} \right) \right),$$

where $\tilde{O}$ suppresses doubly logarithmic multiplicative factors.

IX. DISCUSSION AND CONCLUSIONS

In this work, we have shown that by performing product formula evolutions with different time-step sizes, we can use classical extrapolation techniques to predict the expectation values of observables with gate complexity scaling as $O(T^{1+1/p}/\epsilon \text{polylog}(1/\epsilon))$, where we are ignoring $\log \log$ factors. We emphasize that the $O(1/\epsilon)$ scaling arises from the measurement error, and that the effective Trotter error scales as $O(\text{polylog}(1/\epsilon))$.

In the NISQ regime, circuit depth is arguably the most important cost metric. For this, we consider extrapolation using shot-limited measurements of the observable

or classical shadows. Here the fact coherent measurements are not needed drastically reduces the constant overheads needed and we see only circuit depths scaling as $O(T^{1+1/p} \text{polylog}(1/\epsilon))$ are required. We expect the resource scaling with error to be optimal up to log factors. The Heisenberg measurement limit is a fundamental measurement limit, which requires at least $\Omega(1/\epsilon)$ scaling [52]. Thus we should expect our total resource cost to scale as $\Omega(1/\epsilon)$ in general.

1. Beyond standard product formulas

There are many proposals for modifying product formulas, or otherwise combining them with simulations algorithms [40,65]. In follow-up work it has been shown that the Richardson extrapolation approach can be fruitfully applied to the qDRIFT algorithm to exponentially improve error scaling [66]. It remains open whether these modified algorithms can be combined with the Richardson or polynomial extrapolation approaches considered here to improve their performance.

2. State-dependent bounds

The bounds given in this work are also sensitive to the “physics” of the system. That is, if the system has symmetries, or is restricted to a particular subspace, then we can improve the error estimates given here by replacing the spectral norm with a symmetry-respecting norm. We demonstrate this in Appendix A, where we show that if the evolution is restricted to a particular subspace, the various norms or commutators characterizing the error can be replaced with quantities projected onto the relevant subspace.

3. More tractable expressions for commutator scaling

Although the results in this paper demonstrate commutator scaling in the error of Richardson extrapolation, actually computing these expressions is likely to be extremely computationally expensive. Developing an expression, which can be computed efficiently, is an enormously important task, as it allows us to upper bound the amount of quantum resources needed to reach a guaranteed precision. We hope that future work will find simplified expressions for the commutator error.

4. Lower bounds

The work here gives an improvement on the Trotter error for measured observables. In the general setting, lower bounds have been proven for the performance of Trotterization methods [67], but it remains to be seen if similar bounds on the query complexity can be proven if postprocessing is allowed.

ACKNOWLEDGMENTS

The authors would like to acknowledge very useful discussions with Gyula Lakos on the convergence of the BCH and Magnus expansions, as well as Dong An, Andrew Childs, and Alessandro Roggero on the topic of Trotter error bounds. The authors would also like to thank Abhishek Rajput for useful discussions. The authors also recognize concurrent work by Gumaro Rendon [68] and thank the author for useful discussions and for sharing manuscripts. J.D.W. acknowledges support from the United States Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Accelerated Research in Quantum Computing program, and also NSF QLCI Grant No. OMA-2120757. J.W. acknowledges support from the National Science Foundation under Grants No. DGE-1848739 & No. PHY-2310620, as well as the Department of Energy under Grant No. DE-SC0023658. We thank ECT* and the ExtreMe Matter Institute EMMI at GSI, Darmstadt, for support in the framework of an ECT*/EMMI Workshop during which this work has been initiated.

This paper was prepared for informational purposes by the Global Technology Applied Research Center of JPMorgan Chase & Co. This paper is not a product of the Research Department of JPMorgan Chase & Co., or its affiliates. Neither JPMorgan Chase & Co. nor any of its affiliates makes any explicit or implied representation or warranty and none of them accept any liability in connection with this paper, including, without limitation, with respect to the completeness, accuracy, or reliability of the information contained herein and the potential legal, compliance, tax, or accounting effects thereof. This document is not intended as investment research or investment advice, or as a recommendation offer, or solicitation for the purchase or sale of any security, financial instrument, financial product or service, or to be used in any way for evaluating the merits of participating in any transaction.

No potential conflict of interest was reported by the author(s).

DATA AVAILABILITY

Data for this study is publicly available at Ref. [69].

APPENDIX A: IMPROVED BOUNDS WITH PHYSICAL KNOWLEDGE

Here we consider the case where the Hamiltonian may have some physical symmetry. In this case, we may be able to exploit this symmetry to improve our convergence bounds. Let $H = \sum_j h_j$, where $\{h_j\}_j$ are the terms used in the Trotter decomposition. Then assume that there exists some symmetry subspace denoted $\mathcal{S}$, such that $\Pi_{\mathcal{S}}$ is a projector onto the subspace $\mathcal{S}$. We assume that the both

the overall Hamiltonian and the individual terms in the Trotter decomposition preserve the symmetry. That is, the following holds:

$$[H, \Pi_S] = 0 \quad [h_j, \Pi_S] = 0.$$

Consider a state $|\psi_S\rangle \in \mathcal{S}$. Then we see that

$$\begin{aligned} e^{-ih_j t} |\psi_S\rangle &= e^{-ih_j t} \Pi_S |\psi_S\rangle \\ &= \Pi_S e^{-i\Pi_S h_j \Pi_S t} \Pi_S |\psi_S\rangle, \end{aligned}$$

where we have used that $|\psi_S\rangle = \Pi_S |\psi_S\rangle$ and $\Pi_S = \Pi_S \Pi_S$. Applying this to product formulas,

$$\begin{aligned} \mathcal{P}(t) |\psi_S\rangle &= \prod_j e^{-ia_j h_j t_j} |\psi_S\rangle \\ &= \prod_j e^{-ia_j \Pi_S h_j \Pi_S t_j} |\psi_S\rangle. \end{aligned}$$

Thus, because the simulation is restricted to a particular subspace, we can consider the Hamiltonian restricted to that subspace. In this case we no longer need to consider the full operator norm and can consider a “symmetry-projected” norm. We can define “symmetry-respecting” norms, which characterize the rate of convergence.

$$\Lambda_S := \sum_j \|\Pi_S h_j \Pi_S\|.$$

For the case of commutators, from Lemma 6,

$$\lambda_{j,l} := \left(\sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \prod_{\kappa=1}^l 2^{\frac{\alpha_{\text{comm}}^{(j_\kappa+1)}}{(j_\kappa+1)^2}} \right)^{1/(j+l)}$$

and

$$\alpha_{\text{comm}, \mathcal{S}}^{(j)} = \sum_{\gamma_1 \gamma_2 \dots \gamma_j = 1}^{\Gamma} \|\Pi_S [H_{\gamma_1} H_{\gamma_2} \dots H_{\gamma_j}] \Pi_S\|.$$

These symmetry-respecting quantities can then be used in place of $\Lambda, \alpha_{\text{comm}}^{(j)}$ in Lemma 6 and elsewhere, thus giving us improved convergence results. We believe that similar results should hold when the restriction to the subspace $\mathcal{S}$ is not strictly preserved, i.e., there is some leakage to other subspaces. For example, if one restricts to low-energy states as per Ref. [25].

APPENDIX B: POLYNOMIAL INTERPOLATION: ERROR FROM IMPERFECT CHEBYSHEV NODES

As described earlier, if the Chebyshev nodes do not coincide with the inverse integers, we may not be able to

sample exactly from the Chebyshev nodes as, if we restrict ourselves to integer applications of the Trotter evolution, we require $s = 1/r$, for $r \in \mathbb{Z}$. For any $\xi \in \mathbb{Z}$, the spacing between inverse integers is

$$\begin{aligned} \left(\frac{1}{\xi} - \frac{1}{\xi+1} \right) &= \frac{1}{\xi} \sum_{k=1}^{\infty} \frac{(-1)^k}{\xi^k} \\ &\leq \frac{1}{\xi} \sum_{k=1}^{\infty} \frac{1}{\xi^k} \\ &\leq \frac{2}{\xi^2}, \quad \text{for sufficiently large } \xi, \quad (\text{B1}) \end{aligned}$$

where one can use the standard bound on a geometric sum to get the last line. We then use the following theorem.

Theorem B1 (Perturbed Chebyshev nodes, Sec. 3 of Ref. [70]). Let $x \in [c, d]$, and let L_m Lebesgue constant of the $m+1$ Chebyshev points in $[c, d]$. The Lebesgue constant of the points after they have been perturbed by $\leq \epsilon_N$ away from the Chebyshev node, denoted L'_m , is bounded by

$$L'_m \leq \frac{L_m}{1 - \alpha},$$

where ϵ_m satisfies

$$\epsilon_m = \frac{\alpha(d-c)}{m^2 L_m}.$$

We can apply this to the case where we are not sampling from the exact Chebyshev nodes. We now consider the interval $[-\ell, \ell]$ on which we wish to learn the function $f_B(s)$. Applying Theorem B1 with $\alpha = 1/2$ gives

$$L'_m \leq 2L_m,$$

provided for

$$\epsilon_m = \frac{\ell}{m^2 L_m}.$$

If we consider the interval $[-\ell, \ell]$, then the maximum distance between inverse point is at the boundaries. Hence from Eq. (B1), the maximum distance between inverse integers in $[-\ell, \ell]$ is

$$\begin{aligned} \epsilon_m &\leq \max_{\xi \in [-1/\ell, 1/\ell]} \frac{1}{\xi^2} \\ &\leq \ell^2. \end{aligned}$$

Thus, for a given m , we need to ensure that $\ell \leq (m^2 L_m)^{-1}$. We note that in general ℓ scales in terms of $O(T^{1+1/q})$,

while $m = O(\log(1/\epsilon))$ and $L_m = O(\log \log(1/\epsilon))$. To satisfy both bounds we can choose

$$\ell \leq \frac{1}{(a_{\max} \Upsilon \lambda T)^{1+1/q} m^2 L_m},$$

and thus the minimum number of Trotter steps scales as

$$O(m/\ell) = O\left((a_{\max} \Upsilon \lambda T)^{1+1/q} \log^3\left(\frac{1}{\epsilon}\right) \log \log\left(\frac{1}{\epsilon}\right)\right).$$

APPENDIX C: SIMPLER BOUNDS FOR COMMUTATOR ERROR

In this section we reformulate our bounds in such a way to make them easier to compute with knowledge of low-nested commutators.

Proof of Corollary 2. We first note the following “nesting” property of $\alpha_{\text{comm}}^{(j)}$: for any positive integers k, j such that $2 \leq k \leq j$

$$\alpha_{\text{comm}}^{(j)} \leq (2\Lambda)^{j-k} \alpha_{\text{comm}}^{(k)}. \quad (\text{C1})$$

In particular, in the norm bound on $\tilde{E}_{j+1}$ from Lemma 3, $j_\kappa \geq p$ for each $\kappa \in [l]$. Thus,

$$\begin{aligned} \|\tilde{E}_{j+1}\| &\leq (a_{\max} \Upsilon T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{(a_{\max} \Upsilon T)^l}{l!} \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l \frac{(2\Lambda)^{j_\kappa - p} \alpha_{\text{comm}}^{(p+1)}}{(j_\kappa + 1)^2} \right) \\ &= (a_{\max} \Upsilon T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{(a_{\max} \Upsilon T)^l}{l!} (2\Lambda)^{j-pl} (\alpha_{\text{comm}}^{(p+1)})^l \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \prod_{\kappa=1}^l \frac{2}{(j_\kappa + 1)^2} \\ &= (2a_{\max} \Upsilon \Lambda T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{1}{l!} \left(\frac{a_{\max} \Upsilon \alpha_{\text{comm}}^{(p+1)} T}{(2\Lambda)^p} \right)^l \sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \prod_{\kappa=1}^l \frac{2}{(j_\kappa + 1)^2}. \end{aligned}$$

The inner sum can be bounded as

$$\begin{aligned} &\sum_{\substack{j_1 \dots j_l \in \sigma \mathbb{Z}_+ \geq p \\ j_1 + \dots + j_l = j}} \left(\prod_{\kappa=1}^l \frac{2}{(j_\kappa + 1)^2} \right) \\ &\leq \frac{2^l}{(p+1)^{2l}} \sum_{\substack{j_1 \dots j_l \in \mathbb{N} \\ j_1 + \dots + j_l = j - pl}} 1 \\ &= \frac{2^l}{(p+1)^{2l}} \binom{j+l-1}{l-1} \leq 2^{j-pl+l}, \quad (\text{C2}) \end{aligned}$$

leading to

$$\|\tilde{E}_{j+1}\| \leq (4a_{\max} \Upsilon \Lambda T)^j \sum_{l=1}^{\lfloor j/p \rfloor} \frac{1}{l!} \left(\frac{2a_{\max} \Upsilon \alpha_{\text{comm}}^{(p+1)} T}{(4\Lambda)^p} \right)^l. \quad (\text{C3})$$

By our assumptions on T , the parenthetical term in Eq. (C3) is greater than one, and the sum can be bounded

by the largest power. Thus,

$$\begin{aligned} \|\tilde{E}_{j+1}(T)\| &\leq (e-1)(4a_{\max} \Upsilon \Lambda T)^j \left(\frac{2a_{\max} \Upsilon \alpha_{\text{comm}}^{(p+1)} T}{(4\Lambda)^p} \right)^{\lfloor j/p \rfloor} \\ &\leq 2 \left[42^{1/p} (a_{\max} \Upsilon T)^{1+1/p} \frac{1}{4} \alpha_{\text{comm}}^{(p+1)1/p} \right]^j \\ &\leq 2 \left[2\alpha_{\text{comm}}^{(p+1)1/p} (a_{\max} \Upsilon T)^{1+1/p} \right]^j, \quad (\text{C4}) \end{aligned}$$

which is the corollary statement. ■

APPENDIX D: PREVIOUS LITERATURE ON RICHARDSON EXTRAPOLATION

Here we discuss the previous work on Richardson extrapolation by Vazquez *et al.* [44], which appears to have incorrect derivations for the error in Richardson extrapolation of observables.

In Appendix A, Eq. (13) of Ref. [44], it is claimed that the iterated product formula has an expansion

$$\mathcal{P}^k(T/k) = e^{-iHT} + \sum_{n=1}^{\infty} A_n \frac{T^{n+1}}{k^n}, \quad (\text{D1})$$

where the A_n are size $O(1)$ in T and consist of nested commutators of the Hamiltonian terms. However, this expansion is incorrect, which can be seen by comparison to a counterexample [71]. Consider the Hamiltonian $H = X + Z$ and the associated Trotterization $e^{-iXT/k}e^{-iZT/k}$. Making the identification $s = 1/k$, using the standard BCH formula we get an effective Hamiltonian

$$\begin{aligned} \log(e^{-iXsT}e^{-iZsT}) &= -isT(X + Z) - (sT)^2 \frac{1}{2}[X, Z] + \frac{i(sT)^3}{12}([X, [X, Z]] + [Z, [Z, X]]) - \frac{(sT)^4}{24} \underbrace{[Z[X, [X, Z]]]}_{=0} + O(s^5T^5) \\ &= -i(X + Z)(sT) + iY(sT)^2 + i(Z + X)\frac{(sT)^3}{3} + O((sT)^5). \end{aligned} \quad (\text{D2})$$

We can then exponentiate this and consider the iterated first-order product formula to get

$$\mathcal{P}^{1/s}(sT) = (e^{-iXsT}e^{-iZsT})^{1/s} = \exp\left(-i(X + Z)T + iYsT^2 + i(Z + X)\frac{s^2T^3}{3} + O(s^4T^5)\right). \quad (\text{D3})$$

To obtain a polynomial expansion in s , we use the variation of parameters equation

$$e^{A+B} = e^A + \int_0^1 d\tau e^{(1-\tau)A} B e^{(A+B)\tau}, \quad (\text{D4})$$

which we iterate 4 times to get the following.

$$\begin{aligned} e^{A+B} &= e^A + \int_0^1 d\tau_1 e^{(1-\tau_1)A} B e^{A\tau_1} + \int_0^1 d\tau_1 \int_0^{\tau_1} d\tau_2 e^{(1-\tau_1)A} B e^{A(\tau_1-\tau_2)} B e^{A\tau_2} \\ &\quad + \int_0^1 d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 e^{(1-\tau_1)A} B e^{A(\tau_1-\tau_2)} B e^{A(\tau_2-\tau_3)} B e^{A\tau_3} \\ &\quad + \int_0^1 d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 e^{(1-\tau_1)A} B e^{A(\tau_1-\tau_2)} B e^{A(\tau_2-\tau_3)} B e^{A(\tau_3-\tau_4)} B e^{A\tau_4} + O(s^4T^5). \end{aligned} \quad (\text{D5})$$

We now make the explicit choice

$$\begin{aligned} A &= -iT(X + Z) \\ B &= iYsT^2 + i(Z + X)\frac{s^2T^3}{3} + O(s^4T^5) \end{aligned} \quad (\text{D6})$$

and neglect any terms above $O(s^3)$. Since B is $O(s)$, then the last line of Eq. (D5) is $O(s^4)$, hence we neglect it. Keeping some A terms implicit for brevity, we have

$$\begin{aligned} \mathcal{P}^{1/s}(sT) &= e^{-iT(X+Z)} + \int_0^1 d\tau_1 e^{A(1-\tau_1)} (iYsT^2 + i(Z + X)\frac{s^2T^3}{3}) e^{A\tau_1} \\ &\quad + \int_0^1 d\tau_1 \int_0^{\tau_1} d\tau_2 \left[e^{A(1-\tau_1)} \left(iYsT^2 + i(Z + X)\frac{s^2T^3}{3} \right) e^{A(\tau_1-\tau_2)} \left(iYsT^2 + i(Z + X)\frac{s^2T^3}{3} \right) e^{A\tau_2} \right] \\ &\quad + \int_0^1 d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 e^{A(1-\tau_1)} (iYsT^2) e^{A(\tau_1-\tau_2)} (iYsT^2) e^{A(\tau_2-\tau_3)} (iYsT^2) e^{A\tau_3} + O(s^4). \end{aligned} \quad (\text{D7})$$

Grouping the above according to combinations of s and T , the expression may be written as

$$\mathcal{P}^{1/s}(sT) = e^{-iHT} + sT^2 G_1(T) + s^2 T^3 G_2(T) + s^2 T^4 G_3(T) + s^3 T^5 G_4(T) + s^3 T^6 G_5(T) + O(s^4), \quad (\text{D8})$$

where $G_1, G_2, G_3, G_4, G_5 = O(1)$ are matrix functions of T only. We note in particular that there are terms with scaling $O(s^2 T^4)$, $O(s^3 T^6)$, and $O(s^2 T^5)$, which do not appear with the ratio $s^m T^{m+1}$, which would be necessary for Eq. (D1) to be valid. It can be checked that this occurs for higher-order terms as well.

The incorrect series expansion given in Eq. (D1) is then used in Appendix C of Ref. [44] and hence the errors unfortunately propagate into this later section. In particular, through Eqs. (15)–(18), which is where the error in the Richardson extrapolation is derived. Following from Ref. [44, eq. (17)] (and broadly using their notation) with $\mathcal{P}^{k_j}(T/k_j) = U + E_j$, then

$$\langle O(k_j^{-1}, T) \rangle = \langle \psi | U^\dagger O U | \psi \rangle + \langle \psi | U^\dagger O E_j | \psi \rangle + \langle \psi | E_j^\dagger O U | \psi \rangle + \langle \psi | E_j^\dagger O E_j | \psi \rangle.$$

We now explicitly consider Richardson extrapolation with $l = 3$ extrapolation points, for the Hamiltonian $H = X + Z$. Following Ref. [44, eq. (17)],

$$\epsilon = \sum_{j=1}^{l=3} a_j \langle \psi | U^\dagger O E_j | \psi \rangle \quad (\text{D9})$$

$$= \sum_{j=1}^{l=3} a_j \langle \psi | U^\dagger O \left(T^2 \frac{1}{k_j} G_1 + \frac{1}{k_j^2} T^3 G_2 + \frac{1}{k_j^2} T^4 G_3 + \frac{1}{k_j^3} T^5 G_4 + \frac{1}{k_j^3} T^6 G_5 \right) | \psi \rangle + \mathcal{O}\left(\frac{1}{k_1^4}\right). \quad (\text{D10})$$

We see that the Richardson conditions cancels off all terms up to order $O(k_1^{-2})$. This gives

$$\epsilon = \sum_{j=1}^{l=3} a_j \langle \psi | U^\dagger O \left(\frac{1}{k_j^3} T^4 G_4 + \frac{1}{k_j^3} T^6 G_5 \right) | \psi \rangle + \mathcal{O}\left(\frac{1}{k_1^4}\right).$$

Thus $\epsilon = \mathcal{O}(T^6/k_1^3)$ is the asymptotic scaling of the error. This gives an overall error scaling of the Richardson estimator as

$$\sum_{j=1}^{l=3} a_j \langle O(k_j^{-1}, T) \rangle = \langle \psi | U^\dagger O U | \psi \rangle + \mathcal{O}\left(\frac{T^6}{k_1^3}\right). \quad (\text{D11})$$

The result is that the asymptotic behaviour of the error given in Eq. (7) in the main text of Ref. [44] appears to

be incorrect. The correct version of Eq. (D1) appears in Ref. [20, Lemma 7]. We also realize that for the error in Eq. (D11) to shrink, we require that $k_1 = O(T^2)$, and hence we would require a number of Trotter steps (and hence circuit depths) scaling as $O(T^2)$. As such, the error scaling in Eq. (D11) is consistent with the bounds present in the current work, as seen in Theorem 1 where the time scaling is $O(T^2)$ for $p = 1$ product formulas.

-
- [1] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, and Nicolas P. D. Sawaya, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).
 - [2] Laura Clinton, Toby Cubitt, Brian Flynn, Filippo Maria Gambetta, Joel Klassen, Ashley Montanaro, Stephen Piddock, Raul A. Santos, and Evan Sheridan, Towards near-term quantum simulation of materials, *Nat. Commun.* **15**, 211 (2024).
 - [3] Alessandro Roggero, Andy C. Y. Li, Joseph Carlson, Rajan Gupta, and Gabriel N. Perdue, Quantum computing for neutrino-nucleus scattering, *Phys. Rev. D* **101**, 074038 (2020).
 - [4] Alexander F. Shaw, Pavel Lougovski, Jesse R. Stryker, and Nathan Wiebe, Quantum algorithms for simulating the lattice Schwinger model, *Quantum* **4**, 306 (2020).
 - [5] James D. Watson, Jacob Bringewatt, Alexander F. Shaw, Andrew M. Childs, Alexey V. Gorshkov, and Zohreh Davoudi, Quantum algorithms for simulating nuclear effective field theories, *ArXiv:2312.05344*.
 - [6] B. David Clader, Bryan C. Jacobs, and Chad R. Sprouse, Preconditioned quantum linear system algorithm, *Phys. Rev. Lett.* **110**, 250504 (2013).
 - [7] Aram W. Harrow, Avinandan Hassidim, and Seth Lloyd, Quantum algorithm for linear systems of equations, *Phys. Rev. Lett.* **103**, 150502 (2009).
 - [8] Ryan Babbush, Dominic W. Berry, Robin Kothari, Rolando D. Somma, and Nathan Wiebe, Exponential quantum speedup in simulating coupled classical oscillators, *Phys. Rev. X* **13**, 041041 (2023).
 - [9] Pedro C. S. Costa, Stephen Jordan, and Aaron Ostrander, Quantum algorithm for simulating the wave equation, *Phys. Rev. A* **99**, 012323 (2019).
 - [10] Charles Derby, Joel Klassen, Johannes Bausch, and Toby Cubitt, Compact fermion to qubit mappings, *Phys. Rev. B* **104**, 035118 (2021).
 - [11] Mark Steudtner and Stephanie Wehner, Fermion-to-qubit mappings with varying resource requirements for quantum simulation, *New J. Phys.* **20**, 063010 (2018).
 - [12] Yuchen Wang, LeeAnn M. Sager-Smith, and David A. Mazziotti, Quantum simulation of bosons with the contracted quantum eigensolver, *New J. Phys.* **25**, 103005 (2023).
 - [13] Anthony N. Ciavarella and Ivan A. Chernyshev, Preparation of the SU(3) lattice Yang-Mills vacuum with variational quantum methods, *Phys. Rev. D* **105**, 074504 (2022).
 - [14] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma, Simulating

- Hamiltonian dynamics with a truncated Taylor series, *Phys. Rev. Lett.* **114**, 090502 (2015).
- [15] Guang Hao Low and Isaac L. Chuang, Hamiltonian simulation by qubitization, *Quantum* **3**, 163 (2019).
- [16] Seth Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [17] Emanuel Knill, Gerardo Ortiz, and Rolando D. Somma, Optimal quantum measurements of expectation values of observables, *Phys. Rev. A—At. Mol. Opt. Phys.* **75**, 012328 (2007).
- [18] Rolando D. Somma, Quantum eigenvalue estimation via time series analysis, *New J. Phys.* **21**, 123025 (2019).
- [19] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu, Theory of trotter error with commutator scaling, *Phys. Rev. X* **11**, 011020 (2021).
- [20] Junaid Aftab, Dong An, and Konstantina Trivisa, Multiproduct Hamiltonian simulation with explicit commutator scaling, *ArXiv:2403.08922*.
- [21] Andrew M. Childs and Nathan Wiebe, Hamiltonian simulation using linear combinations of unitary operations, *Quantum Inf. Comput.* **12**, 901 (2012).
- [22] Jeongwan Haah, Matthew B. Hastings, Robin Kothari, and Guang Hao Low, Quantum algorithm for simulating real time evolution of lattice Hamiltonians, *SIAM J. Comput.* **52**, FOCS18 (2021).
- [23] Guang Hao Low, Vadym Kliuchnikov, and Nathan Wiebe, Well-conditioned multiproduct Hamiltonian simulation, *ArXiv:1907.11679*.
- [24] Dominic W. Berry and Andrew M. Childs, Black-box hamiltonian simulation and unitary implementation, *Quantum Inf. Comput.* **12**, 29 (2012).
- [25] Burak Şahinoğlu and Rolando D. Somma, Hamiltonian simulation in the low-energy subspace, *npj Quantum Inf.* **7**, 119 (2021).
- [26] Minh C. Tran, Su-Kuan Chu, Yuan Su, Andrew M. Childs, and Alexey V. Gorshkov, Destructive error interference in product-formula lattice simulation, *Phys. Rev. Lett.* **124**, 220502 (2020).
- [27] Minh C. Tran, Yuan Su, Daniel Carney, and Jacob M. Taylor, Faster digital quantum simulation by symmetry protection, *PRX Quantum* **2**, 010323 (2021).
- [28] Qi Zhao, You Zhou, Alexander F. Shaw, Tongyang Li, and Andrew M. Childs, Hamiltonian simulation with random inputs, *Phys. Rev. Lett.* **129**, 270502 (2022).
- [29] Qi Zhao, You Zhou, and Andrew M. Childs, Entanglement accelerates quantum simulation, *ArXiv:2406.02379*.
- [30] Alexander Zlokapa and Rolando D. Somma, Hamiltonian simulation for low-energy states with optimal time dependence, *Quantum* **8**, 1449 (2024).
- [31] 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 (2015).
- [32] Andrew M. Childs, Dmitri Maslov, Yunseong Nam, Neil J. Ross, and Yuan Su, Toward the first quantum simulation with quantum speedup, *Proc. Natl. Acad. Sci.* **115**, 9456 (2018).
- [33] Markus Heyl, Philipp Hauke, and Peter Zoller, Quantum localization bounds Trotter errors in digital quantum simulation, *Sci. Adv.* **5**, eaau8342 (2019).
- [34] Jan Lukas Bosse, Andrew M. Childs, Charles Derby, Filippo Maria Gambetta, Ashley Montanaro, and Raul A. Santos, Efficient and practical Hamiltonian simulation from time-dependent product formulas, *Nat. Commun.* **16**, 2673 (2025).
- [35] Earl Campbell, Random compiler for fast Hamiltonian simulation, *Phys. Rev. Lett.* **123**, 070503 (2019).
- [36] Huo Chen, Niladri Gomes, Siyuan Niu, and Wibe Albert de Jong, Adaptive variational simulation for open quantum systems, *Quantum* **8**, 1252 (2024).
- [37] Mauro E. S. Morales, Pedro C. S. Costa, Giacomo Pantaleoni, Daniel K. Burgarth, Yuval R. Sanders, and Dominic W. Berry, Greatly improved higher-order product formulae for quantum simulation, *ArXiv:2210.15817*.
- [38] Kouhei Nakaji, Mohsen Bagherimehrab, and Alán Aspuru-Guzik, High-order randomized compiler for Hamiltonian simulation, *PRX Quantum* **5**, 020330 (2024).
- [39] Yingkai Ouyang, David R. White, and Earl T. Campbell, Compilation by stochastic Hamiltonian sparsification, *Quantum* **4**, 235 (2020).
- [40] Kunal Sharma and Minh C. Tran, Hamiltonian simulation in the interaction picture using the magnus expansion, *ArXiv:2404.02966*.
- [41] Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C. Benjamin, Theory of variational quantum simulation, *Quantum* **3**, 191 (2019).
- [42] Conor Mc Keever and Michael Lubasch, Classically optimized Hamiltonian simulation, *Phys. Rev. Res.* **5**, 023146 (2023).
- [43] Suguru Endo, Qi Zhao, Ying Li, Simon Benjamin, and Xiao Yuan, Mitigating algorithmic errors in a Hamiltonian simulation, *Phys. Rev. A* **99**, 012334 (2019).
- [44] Almudena Carrera Vazquez, Daniel J. Egger, David Ochsner, and Stefan Woerner, Well-conditioned multiproduct formulas for hardware-friendly Hamiltonian simulation, *Quantum* **7**, 1067 (2023).
- [45] Gumaro Rendon, Jacob Watkins, and Nathan Wiebe, Improved accuracy for Trotter simulations using Chebyshev interpolation, *Quantum* **8**, 1266 (2024).
- [46] Patrick Cook, Danny Jammooa, Morten Hjorth-Jensen, Daniel D. Lee, and Dean Lee, Parametric matrix models, *ArXiv:2401.11694*.
- [47] Lewis Fry Richardson, The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam, *Philos. Trans. R. Soc. A* **210**, 307 (1911).
- [48] Avram Sidi, in *Practical Extrapolation Methods: Theory and Applications*, Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge, UK, 2003), p. 21.
- [49] Nicholas J. Higham, The numerical stability of barycentric Lagrange interpolation, *IMA J. Numer. Anal.* **24**, 547 (2004).
- [50] T. J. Rivlin, *Dover Books on Mathematics Chebyshev Polynomials* (Dover Publications, Garden City, NY, 2020).
- [51] Dmitry Grinko, Julien Gacon, Christa Zoufal, and Stefan Woerner, Iterative quantum amplitude estimation, *npj Quantum Inf.* **7**, 52 (2021).

- [52] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, Quantum metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
- [53] Paul K. Faehrmann, Mark Steudtner, Richard Kueng, Maria Kieferova, and Jens Eisert, Randomizing multi-product formulas for Hamiltonian simulation, *Quantum* **6**, 806 (2022).
- [54] Sergiy Zhuk, Niall Robertson, and Sergey Bravyi, Trotter error bounds and dynamic multi-product formulas for Hamiltonian simulation, *Phys. Rev. Res.* **6**, 033309 (2024).
- [55] W. Gong, Yaroslav Kharkov, Minh C. Tran, Przemyslaw Bienias, and Alexey V. Gorshkov, Improved digital quantum simulation by non-unitary channels, [ArXiv:2307.13028](https://arxiv.org/abs/2307.13028).
- [56] Nathan Wiebe, Dominic Berry, Peter Hoyer, and Barry C. Sanders, Higher order decompositions of ordered operator exponentials, *J. Phys. A: Math. Theor.* **43**, 065203 (2010).
- [57] Ana Arnal, Fernando Casas, and Cristina Chiralt, A note on the Baker–Campbell–Hausdorff series in terms of right-nested commutators, *Mediterr. J. Math.* **18**, 1 (2021).
- [58] Gyula Lakos, Convergence estimates for the Magnus expansion I. Banach algebras, [ArXiv:1709.01791](https://arxiv.org/abs/1709.01791).
- [59] Gyula Lakos, Convergence estimates for the Magnus expansion II. C^* -algebras, [ArXiv:1910.03328](https://arxiv.org/abs/1910.03328).
- [60] Jacob Watkins, Nathan Wiebe, Alessandro Roggero, and Dean Lee, Time-dependent Hamiltonian simulation using discrete-clock constructions, *PRX Quantum* **5**, 040316 (2024).
- [61] Alfio Quarteroni, Riccardo Sacco, and Fausto Saleri, *Numerical Mathematics* (Springer Science & Business Media, Berlin, Germany, 2010), Vol. 37.
- [62] Jacob Watkins, New quantum algorithms and analyses for Hamiltonian simulation. PhD thesis, Michigan State University, 2024, p. 193.
- [63] We note a small difference in the scaling of the error compared to Ref. [62] where the gradient changes with T . This is due to the different choice in the minimum number of Trotter steps.
- [64] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [65] Abhishek Rajput, Alessandro Roggero, and Nathan Wiebe, Hybridized methods for quantum simulation in the interaction picture, *Quantum* **6**, 780 (2022).
- [66] James D. Watson, Randomly compiled quantum simulation with exponentially reduced circuit depths, [ArXiv:2411.04240](https://arxiv.org/abs/2411.04240).
- [67] Alexander Hahn, Paul Hartung, Daniel Burgarth, Facchi Paolo, and Kazuya Yuasa, Lower bounds for the Trotter error, *Phys. Rev. A* **111**, 022417 (2025).
- [68] Gumaro Rendon, All you need is Trotter, [ArXiv:2311.01533](https://arxiv.org/abs/2311.01533).
- [69] <https://doi.org/10.5281/zenodo.16369083>.
- [70] Marco Vianello and Federico Piazzon, Stability inequalities for Lebesgue constants via Markov-like inequalities, *Dolomites Res. Notes Approx.* **11**, 1 (2018).
- [71] We add that the series in Eq. (D1) (i.e., Ref. [44, Eq. (13)]) is claimed to come from Ref. [72], however, we were unable to verify this explicit series appears in Ref. [72].
- [72] Siu A. Chin, Multi-product splitting and Runge-Kutta-Nyström integrators, *Celest. Mech. Dyn. Astron.* **106**, 391 (2010).