Improved Error Scaling for Trotter Simulations through Extrapolation

Gumaro Rendon[1], Jacob Watkins[2], and Nathan Wiebe[3,4]

[1]Zapata Computing Inc., Boston, MA 02110, USA
[2]Department of Physics, Michigan State University, Lansing, MI 48912, USA
[3]Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada
[4]Pacific Northwest National Laboratory, Richland, WA 99352, USA

In recent years, Trotter formulas have emerged as a leading approach for simulating quantum dynamics on quantum computers, owing to their ability to exploit locality and commutator structure of the Hamiltonian. However, a major problem facing Trotter formulas is their inability to achieve poly-logarithmic scaling with the error tolerance. We address this problem by providing a well-conditioned extrapolation scheme that takes data from Trotter-Suzuki simulations obtained for specifically chosen Trotter step sizes and estimates the value that would be seen in the limit where the Trotter step size goes to zero. We show this leads, even for the first order Trotter formula, to $\tilde{O}(1/\epsilon)$ scaling for phase estimation and $\tilde{O}(t^2/\epsilon)$ scaling for estimating time-evolved expectation values for simulation time t and error tolerance ϵ. This is better scaling with the error tolerance than the best known un-extrapolated Trotter formulas. Additionally, we provide a new approach for phase estimation that is unbiased and also provide a new approach for estimating the Trotter error on a quantum computer through extrapolation which yields a new way to independently assess the errors in a Trotter simulation.

1 Introduction

Quantum simulation has become, arguably, the most promising application of quantum computing in the near-term [1, 2], with the potential to provide exponential speedups for a host of problems ranging from the electronic structure [2, 3, 4, 5] to simulation

Gumaro Rendon: gumaro.rendon@zapatacomputing.com
Jacob Watkins: watkins@frib.msu.edu
Nathan Wiebe: nawiebe@cs.toronto.edu

of scattering dynamics within quantum field theories [6, 7, 8]. The central challenge digital Hamiltonian simulation is, given a fixed Hamiltonian, simulation time and error tolerance, provide a minimal-length sequence of quantum gates that approximates the unitary dynamics within that error tolerance.

Major strides have been made in the last several years towards this goal. Specifically, several methods, including linear combinations of unitaries [9, 10, 11, 12, 13], qubitization [14, 15] and Trotter-Suzuki [16, 17, 18, 19, 20, 21] formulas, have emerged as the best methods known for simulating dynamics. Unlike the other aforementioned methods, Trotter-Suzuki expansions yield a complexity that scales with the commutators of the Hamiltonian terms, which can lead to substantial performance improvements for simulations of local Hamiltonian. In contrast, Trotter methods scale super-polynomially worse with the error tolerance than other existing methods [22, 11]. This makes such methods asymptotically inferior to other strategies in cases where high accuracy is required.

The aim of this work is to show that this gap can be bridged using polynomial interpolation. Specifically, we show that polylog$(1/\epsilon)$ scaling with the error tolerance ϵ can be attained by observables returned by quantum simulations through the use of interpolation. Interestingly, this result also holds in cases where only the lowest order Trotter formula is used and provides the first poly-logarithmic scaling method with the error that uses a constant number of ancillary qubits. We study this method in two cases. The first case uses phase estimation to extract eigenvalues of the Hamiltonian; whereas the second method uses dynamical simulation and amplitude estimation to learn expectation values of an observables in error that scales polylogarithmically. We further validate these methods numerically for a newly proposed unbiased phase estimation method that we call Gaussian phase estimation.

The layout of the paper is as follows. We review some of the critical background needed to understand this paper and also state our main results in Section 2. Specifically, we review Trotter-Suzuki formulas, polynomial extrapolation and present the well conditioned extrapolation formulas that we use to enable our approach. In Section 3 we state our main results where we apply this technique to extrapolate the eigenvalues yielded by phase estimation experiments in the limit where the duration of the Trotter step approaches zero. We further discuss applications of these ideas to the calculation of the Frobenius norm which in turn gives us a way to estimate the Trotter-Suzuki error for a dynamical simulation. In Section 4 we present our main results involving extrapolation of expectation values of time-evolved observables. This case is conceptually distinct from the case of phase estimation because here the evolutions considered are generically long and consist of several Trotter formulas whereas the phase estimation application only uses short evolutions. In Section 5 we discuss the newly proposed Gaussian phase estimation. Finally, in Section 6 we validate our claims numerically for small instances of transverse Ising models. Finally, we conclude in Section 7 and discuss future avenues of research.

2 Trotter formulas and polynomial extrapolation

We assume that we have a Hamiltonian that can be expressed as the sum of m terms

$$H = \sum_{j=1}^{m} H_j \tag{2.1}$$

such that exponential $e^{-iH_j t}$ of each term is easy to compute ("fast forwardable") and m grows polynomially in the number of subsystems (typically qubits). Examples of such H include k-local and k-sparse Hamiltonians. Under these assumptions, one can efficiently approximate the time evolution operator $U(t) = e^{-iHt}$ to some desired precision ϵ with a quantum circuit. Perhaps the best-known of these simulation techniques are product formulas (a.k.a. Trotter formulas). Canonical examples include the first order formula

$$S_1(t) := \prod_{j=1}^{m} e^{-iH_j t}, \tag{2.2}$$

the second order symmetric formula

$$S_2(t) := e^{-iH_1 t/2} \dots e^{-iH_m t/2} e^{-iH_m t/2} \dots e^{-iH_1 t/2}, \tag{2.3}$$

as well as the order $2k$ symmetric Suzuki-Trotter formula, defined recursively as

$$S_{2k}(t) := S_{2k-2}(u_k t) S_{2k-2}\left((1-4u_k)t\right) S_{2k-2}(u_k t). \tag{2.4}$$

Here $u_k := 1/\left(4 - 4^{1/(2k-1)}\right)$. Though many kinds of product formulas have been developed, the symmetric formulas of equation (2.4) will be our primary tool for this work, due to its symmetry under (naive) time reversal

$$S_{2k}(-t) = S_{2k}(t)^\dagger \tag{2.5}$$

and the fact that the order of the formula $2k$ can be taken arbitrarily large.

The error in the Trotter-Suzuki approximation satisfies

$$e^{-iHt} - S_{2k}(t) = \sum_{q=k}^{\infty} C_{2q} t^{2q+1}, \tag{2.6}$$

for bounded operators C_{2q} such that [18]

$$\| \sum_{q=k}^{\infty} C_{2q} t^{2q+1} \| \leq \alpha_{\text{comm}} t^{2k+1}, \tag{2.7}$$

where

$$\alpha_{\text{comm}} \leq \sum_{q_1,\dots,q_{2k+1}} \| [H_{q1}, \dots, [H_{q_{2k}}, H_{q_{2k+1}}] \dots] \| \leq (4m5^{k-1} \max_j \|H_j\|)^{2k+1}. \tag{2.8}$$

This shows that the error in Trotter formulas only depends on the commutators between the terms in the Hamiltonian, which can provide an advantage relative to methods such as qubitization or linear-combinations of unitaries in the event that the Hamiltonian in question is local [19].

Product formulas approximate $U(t)$ only in a neighborhood around $t = 0$. To obtain accurate approximations for more generic $t > 0$, we subdivide the time interval $[0, t]$ using r subintervals such that each interval is sufficiently small that the Trotter approximation is valid. For the simple case of a uniform mesh of r subintervals (Trotter steps), this becomes

$$S_{2k}(t/r)^r = U(t) + O(t^{2k+1}/r^{2k}) \tag{2.9}$$

where big O is understood as taking r large and t small. Clearly, we have that for bounded Hamiltonians $\lim_{r\to\infty} S_{2k}(t/r)^r = U(t)$. For the purposes of our work, we will find it convenient to reparametrize with $s := 1/r$, $s \in [-1, 1] \setminus \{0\}$. Our approximate formulas become

$$\tilde{U}_s := S_{2k}(st)^{1/s}. \tag{2.10}$$

With this parametrization, time evolutions become more exact as $s \to 0$, and more costly. The discontinuity at $s = 0$ is removable and may be filled by the limit, which tends to $U(t)$. Another benefit provided by the symmetric property is that s and $-s$ give the same formula.

$$S_{2k}(-st)^{-1/s} = \left(S_{2k}(-st)^\dagger\right)^{1/s} = \left(\left(S_{2k}(st)^\dagger\right)^\dagger\right)^{1/s} = S_{2k}(st)^{1/s} \tag{2.11}$$

Hence, we will usually assume $s > 0$ without loss of generality.

One can implement $\tilde{U}_s$ for any $s \in (0, 1]$ using fractional queries.[23] Specifically, by splitting $1/s$ into integer and fractional parts

$$1/s = r + f, \tag{2.12}$$

with $r \in \mathbb{Z}_+$ and $f \in (0, 1)$, we may simulate first r steps and then the fractional part.

$$\tilde{U}_s = S_{2k}(st)^r S_{2k}(st)^f \tag{2.13}$$

The fractional part can be computed using, say, the Quantum Singular Value Transformation. However, there is value in performing only integer Trotter steps, for conceptual simplicity as well as reduced computational overhead. We will seek to use integer $1/s$ whenever possible, particularly when it does not come at a significant computational expense. This is an important point since the choice of values of s (nodes) for interpolation can have a tremendous effect on the performance.

An important technical aspect of our work involves calculating arbitrary-order derivatives of $\tilde{U}_s$ with respect to s. Due to the annoying factor of $1/s$ in (2.10), it will be

useful to express $\tilde{U}_s$ in terms of an effective Hamiltonian H_{eff}.

$$\begin{aligned} S_{2k}(st)^{1/s} &= e^{-iH_{\text{eff}}t} \\ H_{\text{eff}} &:= \frac{i}{st}\log S_{2k}(st) \end{aligned} \tag{2.14}$$

Note that H_{eff} depends on both s and t, though explicit statement of these dependencies will typically be suppressed. For the purposes of bounding the interpolation error, a bound on the norm of H_{eff} will be useful to have. This is supplied by the following lemma.

Lemma 1. *Let $H = \sum_{j=1}^{m} H_j$ be a Hamiltonian on a finite-dimensional Hilbert space, and let $S_{2k}(st)^{1/s}$ be a symmetric, $2k$-order Suzuki-Trotter formula for e^{-iHt}, with $t > 0$ and $s \in (0,1)$, with effective Hamiltonian H_{eff} given by equation (2.14). Provided that s is chosen such that $2k(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|st \leq \pi/20$, the following bound on the derivatives of H_{eff} with respect to s holds.*

$$\|\partial_s^n H_{\text{eff}}(st)\| \leq 2t^{-1}n^n(2ke^2(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|t)^{n+1}.$$

The proof of this lemma is technical, so it is relegated to Appendix A.

2.1 Polynomial Extrapolation

The principal aim of our paper is to investigate the impact that polynomial extrapolation/interpolation of data computed using order $2k$ Suzuki-Trotter formulas. This approach is however not new. It has long been used in numerical experiments as well as in error mitigation to estimate an error-free result using a sequence of approximate estimations. There are many quantites that we could be interested in extrapolating to the zero-error limit. Examples include the energies $E_i(s)$ of the effective Hamiltonian,

$$H_{\text{eff}}\left|E_i(s)\right\rangle = E_i(s)\left|E_i(s)\right\rangle \tag{2.15}$$

as well as expectation values of time-evolved observables.

$$\begin{aligned} \langle O_s(t)\rangle &= \text{Tr}\left(\rho O_s(t)\right) \\ O_s(t) &:= S_{2k}^{\dagger}(st)^{1/s} O S_{2k}(st)^{1/s} \end{aligned} \tag{2.16}$$

While the interpolation is classical and independent of the method in which the data is generated, we will assume a quantum simulation was used when considering the computational cost. We assume all quantum operations are executed perfectly, including the exponentials $\exp(-iH_j\tau)$ for simulation. Thus, the primary sources of error are the interpolation error and error in the calculation of the data points (e.g. the Hamiltonian energies or expectation values at various points s_i). Error in the data points arise naturally and depend on the method of computing them, and will be discussed in the context of the particular application. We emphasize that these errors are intrinsic

to the algorithms used, and independent of any imperfections of the computing device (which we neglect).

We will now outline the mathematical framework for the rest of our paper. Let $f \in C^n([-a,a])$ be a n-differentiable function of a single variable $s \in [-a,a] = I_a$ and suppose we have computed f for n distinct points $s_1, s_2 \dots s_n \in I_a$ such that $s_i \neq 0$. That is, we have a set of pairs $(s_i, f(s_i))$ for $i \in [1,n]$. Let $P_{n-1}f$ be the (unique) $(n-1)$-degree polynomial interpolating f at the n points s_i. For any $s \in [-a,a]$, a standard result in polynomial interpolation tells us the error is given by

$$E_{n-1}(s) := f(s) - P_{n-1}f(s) = \frac{f^{(n)}(\xi)}{n!}\omega_n(s) \tag{2.17}$$

for some $\xi \in I_s$, where I_s is the smallest interval containing s and the interpolation points $\{s_i\}_{i=1}^n$. The nth degree nodal polynomial $\omega_n(s)$ is defined as the monic polynomial with zeros at the interpolation points.

$$\omega_n(s) := \prod_{i=0}^{n-1}(s - s_i) \tag{2.18}$$

As described above, our goal is to estimate $f(0)$ via $P_{n-1}f(0)$. Since we are interested in $s = 0$, ω_n becomes a (signed) product of the interpolation points. We can bound the interpolation error $E_n(0)$ in a way that is independent of the precise value of ξ (which is unknown and difficult to find) by maximizing over $\xi \in I_s$.

$$|E_{n-1}(0)| \leq \max_{s\in[-s_{\max},s_{\max}]} \frac{\left|f^{(n)}(s)\right|}{n!} \prod_{i=1}^{n} |s_i| \tag{2.19}$$

Here, $s_{\max} := \max_i |s_i|$. Much of the technical work in this paper involves finding suitable bounds on the size of the derivatives $f^{(n)}$.

The process for finding the coefficients for the polynomial expansion can be found in many ways. A straightforward approach is can be found through linear algebra in the following manner when we wish to find $P_{n-1}(s) = \sum_{j=1}^n a_j s^j \approx f(s)$ on a set of points $s_1, \dots, s_M$. Let $\mathbf{V} \in \mathbb{R}^{M\times n}$ be a generalized Vandermonde matrix then the coefficients a_j can be found by solving

$$\mathbf{V}\vec{a} = \vec{f}(s) \Longleftrightarrow \begin{bmatrix} 1 & s_1 & s_1^2 & \cdots & s_1^{n-1} \\ 1 & s_2 & s_2^2 & \cdots & s_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s_M & s_M^2 & \cdots & s_M^{n-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(s_1) \\ f(s_2) \\ \vdots \\ f(s_M) \end{bmatrix} \tag{2.20}$$

If a solution exists, then it is given by the Moore-Penrose pseudoinverse $\vec{a} = \mathbf{V}^+\vec{f} = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\vec{f}$. An inverse may not necessarily exist, even in the square case, if the rows are linearly dependent then the interpolation will fail. Further, even if the matrix is invertible then the practical benefits of polynomial extrapolation may be lost if the

condition number $\kappa(\mathbf{V}) := \|\mathbf{V}\|\|\mathbf{V}^{-1}\|$ is large. A large condition number implies that the fit is not robust with respect to small displacements in the data $f(s_i) \to f(s_i) + \delta$. This is a critical problem in quantum settings as the complexity of learning the dataset f within error ϵ typically scales as $\Theta(1/\epsilon)$.

The problem with using nodes on the real axis in conjunction with the Vandermonde matrix Equation (2.20) is how the condition number scales with the size of the matrix. It has been shown that the condition number for the classical Vandermonde matrix scales exponentially with its size on any node configuration that lies on the real axis [24]. This means that the error propagation on the interpolant would also grow exponentially with the number of terms, M.

We ensure a well conditioned fit here by using a set Chebyshev nodes for the s_i. This is analogous to the choice taken in the well-conditioned multi-product formula approach to quantum simulation [14]. After choosing the Chebyshev nodes, a specialized version of the bound (2.19) can be given that allows a tighter bound on the error for our purposes. This bound is given below.

Lemma 2. *Let s_i, $i = 1, 2, \ldots, n$ be the collection of Chebyshev interpolation points on the interval $I_a = [-a, a]$.*

$$s_i = a \cos\left(\frac{2i-1}{2n}\pi\right) \tag{2.21}$$

Let $f \in C^n([-a, a])$, and let $P_{n-1}f$ be the unique $n-1$-degree polynomial interpolation of f through the s_i. The error function $E_{n-1} := f - P_{n-1}f$ is bounded at $s = 0$ by

$$|E_{n-1}(0)| \leq \max_{s \in I_a} \left|f^{(n)}(s)\right| \left(\frac{a}{2n}\right)^n \tag{2.22}$$

Proof. For n odd, $s = 0$ is one of the interpolation points, so the error is zero and the bound holds. Therefore we only consider n even (which will be the case of practical interest).

Using the generic bound (2.19) with the Chebyshev nodes,

$$|E_{n-1}(0)| \leq \max_{\xi \in I_a} \left|f^{(n)}(\xi)\right| \frac{1}{n!} a^n \prod_{i=1}^{n} \left|\cos\left(\frac{2i-1}{2n}\pi\right)\right|. \tag{2.23}$$

To obtain the lemma, we just need to appropriately bound the product of cosines (standard Chebyshev nodes). Since n is even, $n = 2m$ for some $m \in \mathbb{Z}_+$. Moreover, we have a reflectional symmetry about m, in the sense that

$$\left|\cos\left(\frac{2i-1}{2n}\pi\right)\right| = \left|\cos\left(\frac{2(n-i+1)-1}{2n}\pi\right)\right|. \tag{2.24}$$

Hence, we only need to take the product over $i = 1, \ldots, m$ and square it.

$$\prod_{i=1}^{n} \left|\cos\left(\frac{2i-1}{2n}\pi\right)\right| = \left(\prod_{i=1}^{m} \cos\left(\frac{2i-1}{4m}\pi\right)\right)^2 \tag{2.25}$$

To proceed further, let's reindex the remaining product by $i \to m - i + 1$. This gives

$$\begin{aligned}\prod_{i=1}^{m} \cos\left(\frac{2i-1}{4m}\pi\right) &= \prod_{i=1}^{m} \cos\left(\frac{\pi}{2} - \frac{2i-1}{4m}\pi\right) \\ &= \prod_{i=1}^{m} \sin\left(\frac{2i-1}{4m}\right) \\ &\leq \prod_{i=1}^{m} \frac{2i-1}{4m}\end{aligned} \tag{2.26}$$

where we used the fact that $\sin(x) \leq x$ for all $x \geq 0$. Factoring out the denominator from the product, the remaining terms become a double factorial.

$$\prod_{i=1}^{m} \frac{2i-1}{4m} = \frac{(2m-1)!!}{(4m)^m} \tag{2.27}$$

The double factorial can be bounded as follows.

$$(2m-1)!!^2 \leq (2m-1)!!(2m)!! = 2m! \tag{2.28}$$

so that $(2m-1)!! \leq \sqrt{(2m)!}$. Returning to the original product of equation (2.25), and reintroducing $n = 2m$, the resulting bound is

$$\prod_{i=1}^{n} \left|\cos\left(\frac{2i-1}{2n}\pi\right)\right| \leq \left(\frac{\sqrt{n!}}{(2n)^{n/2}}\right)^2 = \frac{n!}{(2n)^n} \tag{2.29}$$

Reinserting this result into the last line of equation (2.23) gives the bound stated in the lemma. □

Although the Chebyshev nodes allow for well-conditioning, they tend to anti-cluster away from $s = 0$, so that the resulting error bound is not as strong as for, say, a sequence of points $s_i = 1/i$. Nevertheless, we will find it possible to obtain sufficiently strong bounds which yield good performance for the interpolation algorithm.

2.2 Stability analysis and low condition-number interpolation

In theory, the discussion of the previous section suggests polynomial interpolation allows us to achieve exponentially small errors in the extrapolation of a C^∞ function based on a linear number of example points provided. However, the measurement errors that we have in this dataset can be amplified by this process to such an extent that the benefits from extrapolation would be outweighted by the miniscule measurement errors needed to achieve an exacting error tolerance. We can overcome these issues by replacing the monomial expansion of (2.20) by an expansion with respect to a set of orthogonal polynomials:

$$f(s) = \sum_{j=0}^{\infty} a_{\text{opt},j} p_j(s). \tag{2.30}$$

We formalize this below.

Lemma 3. *Let $p_j(s)$ be the jth degree polynomial proportional to the jth Chebyshev polynomial $T_j(s)$ as follows.*

$$p_j(s) = \begin{cases} \sqrt{\frac{1}{M}} T_0(s), & j = 0 \\ \sqrt{\frac{2}{M}} T_j(s), & j = 1, 2, \dots \end{cases} \tag{2.31}$$

Let $s_j = \cos((2j-1)\pi/2M)$ be the roots of T_M. The matrix $\mathbf{V}_{\text{opt}}$ defined by

$$\mathbf{V}_{\text{opt}} := \begin{pmatrix} p_0(s_1) & p_1(s_1) & p_2(s_1) & \dots & p_{M-1}(s_1) \\ p_0(s_2) & p_1(s_2) & p_2(s_2) & \dots & p_{M-1}(s_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_0(s_M) & p_1(s_M) & p_2(s_M) & \dots & p_{M-1}(s_M) \end{pmatrix}$$

is an orthogonal matrix, and therefore has condition number $\kappa(\mathbf{V}_{\text{opt}}) = 1$.

Proof. The proof of our claim follows from showing that the $\mathbf{V}_{\text{opt}}$ matrix is an orthogonal matrix and thus has unit condition number. First we will use the identity for Chebyshev polynomials that

$$T_j(\tau) = \cos(j \arccos(\tau)), \quad j = 0, 1, 2 \dots \tag{2.32}$$

This implies that

$$p_i(s_j) = \begin{cases} \sqrt{\frac{1}{M}}, & i = 0 \\ \sqrt{\frac{2}{M}} \cos(i(2j-1)\pi/2M), & i = 1, 2, \dots \end{cases} \tag{2.33}$$

First let us show that the columns of the operator $\mathbf{V}_{\text{opt}}$ are unit vectors.

$$\sum_j [\mathbf{V}_{\text{opt}}]_{j0}^2 = \sum_j 1/M = 1. \tag{2.34}$$

Next for $i \neq 0$ we have

$$\sum_j [\mathbf{V}_{\text{opt}}]_{ji}^2 = \frac{2}{M} \sum_{j=1}^{M} \cos^2 \left(\frac{i(2j-1)\pi}{2M} \right) = \sum_j \frac{2}{M} \left(\frac{1 + \cos\left(\frac{i(2j-1)\pi}{M}\right)}{2} \right) = 1. \tag{2.35}$$

Finally we need to show orthonormality of the columns. For the case where $i = 0$, it is immediately clear that the inner product of this vector with the remainder is zero. It is also true for $i \neq 0$ as can be seen by the following argument

$$\sum_j [\mathbf{V}_{\text{opt}}]_{ji} [\mathbf{V}_{\text{opt}}]_{jk} = \frac{2}{M} \sum_{j=1}^{M} \cos \left(\frac{i(2j-1)\pi}{2M} \right) \cos \left(\frac{k(2j-1)\pi}{2M} \right) = 0. \tag{2.36}$$

Thus the columns form an orthonormal basis. The exact same arguments show that the rows form an orthonormal basis. This implies that the matrix is orthogonal.

Orthogonal matrices have unit eigenvalues values, which means that $\|\mathbf{V}_{\text{opt}}\| = 1$ and $\left\|\mathbf{V}_{\text{opt}}^{-1}\right\| = 1$. This in turn means that $\kappa(\mathbf{V}_{\text{opt}}) = 1$. □

Lemma 4. *Assume we are given the vector $\tilde{f}(s_j)$ such that for all j $|\tilde{f}(s_j) - f(s_j)| \leq \epsilon/\sqrt{2M}$ with probability at least $1 - \delta/M$. Let $p \in \mathbb{R}^M$ be a vector such that $p_1(s) = T_0(s)/\sqrt{M}$ and $p_j(s) = T_j(s)\sqrt{2/M}$ for all $j > 1$. We then have that we can construct an estimate of $f(s)$, $p^T \mathbf{V}_{\text{opt}}^{-1} \tilde{f}$ such that*

$$\|p^T \mathbf{V}_{\text{opt}}^{-1} f - p^T \mathbf{V}_{\text{opt}}^{-1} \tilde{f}\| \leq \epsilon$$

with probability at least $1 - \delta$.

Proof. Now that we have these pieces in place we can move forward and analyze the sensitivity of the extrapolation of the value of the the eigenvalue. First, we notice that

$$f(s) \approx \sum_j a_{\text{opt},j} p_j(s) = \sum_{i,j} p_j(s) \left(\mathbf{V}_{\text{opt}}^{-1}\right)_{ji} f(s_j) := p^T \mathbf{V}_{\text{opt}}^{-1} f \tag{2.37}$$

We also have that

$$|T_j(s)| \leq 1. \tag{2.38}$$

Next with these in hand we will assume that $\tilde{f}$ is the vector of measurements of the energy. We then have that

$$\begin{aligned}
&\|p^T \mathbf{V}_{\text{opt}}^{-1} f - p^T \mathbf{V}_{\text{opt}}^{-1} \tilde{f}\| \leq \|p^T\| \|\mathbf{V}_{\text{opt}}^{-1}\| \|f - \tilde{f}\| \\
&\leq \sqrt{1/M + \sum \frac{2}{M}} \|f - \tilde{f}\| = \sqrt{2 - 1/M}\sqrt{M} \max_i |f(s_i) - \tilde{f}(s_i)| \\
&\leq \sqrt{2M} \max_i |f(s_i) - \tilde{f}(s_i)|.
\end{aligned} \tag{2.39}$$

Thus if each eigenvalue is measured within error $\epsilon/\sqrt{2M}$ and with failure probability at most $\delta' = \delta/M$, then by the union bound the probability of failure is at most $M\delta' = \delta$ and with error at most $\sqrt{2M} \max_i |f(s_i) - \tilde{f}(s_i)| \leq \epsilon$. □

With this semi-deterministic approach one might think that we can improve in the error/uncertainty propagation. Thus, we propose the following lemma:

Lemma 5. *Now, assume we promote $f(s_j)$ to normally distributed random variables with central values $\mu_j = f(s_j)$ and variances σ_j. This means that the corresponding promoted interpolant, also normally distributed, has a variance*

$$\sigma_P^2(s) \leq \max_j \left(\sigma_j^2\right). \tag{2.40}$$

Proof.

$$\begin{aligned}
\sigma_P^2(s) = \text{Var}\left(P_M(s)\right) &= \text{Cov}\left(P_M(s), P_M(s)\right) \\
&= \text{Cov}\left(p_i(s) \left(\mathbf{V}_{\text{opt}}^{-1}\right)_{ij} f(s_j), p_l(s) \left(\mathbf{V}_{\text{opt}}^{-1}\right)_{lk} f(s_k)\right) \\
&= p^T(s) \mathbf{V}_{\text{opt}}^{-1} \mathbf{C}_f \left(\mathbf{V}_{\text{opt}}^{-1}\right)^T p(s),
\end{aligned} \tag{2.41}$$

where $\mathbf{C}_f = \mathrm{diag}\,(\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2)$. Thus,

$$\sigma_P^2(s) \leq \|p(s)\|^2 \|\mathbf{C}_f\| \|\mathbf{V}_{\mathrm{opt}}\|^2 \leq 2 \max_j \left(\sigma_j^2\right) \tag{2.42}$$

□

Thus, finding an efficient way of sampling the observables, having the samples be normally distributed, might be convenient.

Unsatisfied with the bounds on Lemma 4, and with hopes of being able to use only a single ancillary qubit, we propose a tighter bound.

Lemma 6. *Assume we are given the vector $\tilde{f}(s_j)$ such that for all j $|\tilde{f}(s_j) - f(s_j)| \leq \epsilon/(\frac{2}{\pi}\log(M+1)+1)$ with probability at least $1-\delta/M$. Let $p \in \mathbb{R}^M$ be a vector such that $p_1(s) = T_0(s)/\sqrt{M}$ and $p_j(s) = T_j(s)\sqrt{2/M}$ for all $j > 1$. We then have that we can construct an estimate of $f(s)$, $p^T \mathbf{V}_{\mathrm{opt}}^{-1} \tilde{f}$ such that*

$$\|p^T \mathbf{V}_{\mathrm{opt}}^{-1} f - p^T \mathbf{V}_{\mathrm{opt}}^{-1} \tilde{f}\| \leq \epsilon$$

with probability at least $1-\delta$.

Proof. Now that we have these pieces in place we can move forward and analyze the sensitivity of the extrapolation of the value of the the eigenvalue. First, we notice that

$$f(s) \approx \sum_j a_{\mathrm{opt},j} p_j(s) = \sum_{i,j} p_j(s) \left(\mathbf{V}_{\mathrm{opt}}^{-1}\right)_{ji} f(s_j) := p^T \mathbf{V}_{\mathrm{opt}}^{-1} f \tag{2.43}$$

The interpolant error due to imperfect estimation can be written like

$$\begin{aligned} &\left|p^T \mathbf{V}_{\mathrm{opt}}^{-1} f - p^T \mathbf{V}_{\mathrm{opt}}^{-1} \tilde{f}\right| \\ &= \left|c^T \left(f - \tilde{f}\right)\right|, \end{aligned} \tag{2.44}$$

where $c = \left(\mathbf{V}_{\mathrm{opt}}^{-1}\right)^T p = \mathbf{V}_{\mathrm{opt}} p$. We now make use of Hölder's inequality

$$|f^T g| \leq \|f\|_p \|g\|_q$$

where $1/p + 1/q = 1$. Choosing $p = 1$ and $q = \infty$, we have

$$\left|c^T \left(f - \tilde{f}\right)\right| \leq \|c\|_1 \|f - \tilde{f}\|_\infty. \tag{2.45}$$

By Lemma 21, we know that the 1-norm of c is $L_M \leq \frac{2}{\pi}\log(M+1)+1$ and $\|f-\tilde{f}\|_\infty$ is simply $\max_i |f(s_i) - \tilde{f}(s_i)| \leq \epsilon/(\frac{2}{\pi}\log(M+1)+1)$.

Thus if each observable is measured within error $\epsilon/(\frac{2}{\pi}\log(M+1)+1)$ and with failure probability at most $\delta' = \delta/M$, then by the union bound the probability of failure is at most $M\delta' = \delta$ and error at most ϵ. □

3 Trotter errors and phase estimation

Our first example of extrapolation involves extrapolating Trotter-Suzuki errors to zero for phase estimation. The idea in essence is to perform logarithmically many phase estimation experiments evaluated at the Chebyshev nodes considered in Section 2 we then bound the error from our previous polynomial interpolation results. The following theorem bounds the performance of such an algorithm.

Theorem 7. *Let $H : \mathbb{R} \mapsto \mathbb{C}^{2^n \times 2^n}$ be a map such that for all t, $H(t) = \sum_{j=1}^{m} a_j(t) H_j$ where each H_j is a Hermitian matrix and $a_j(t)$ is real valued and in C^K for some integer $K > 0$. Further for any t let $|\ell(t)\rangle$ be an eigenstate of $H(t)$ such that $H(t)|\ell(t)\rangle = E_\ell(t)|\ell(t)\rangle$ and assume that there exists a constant $\gamma_{\text{eff}} > 0$ such that for all $\ell' \neq \ell$ and $t > 0$ $|E_\ell(t) - E_{\ell'}(t)| \geq \gamma_{\text{eff}} > 0$. Finally, assume that the evolution time t used for all such experiments satisfy $2k(5/3)^{k-1} m \max_{l\in[1,m]} \|H_l\| t \leq \pi/20$. It is then possible to use an n-point polynomial interpolation formula to estimate $E_\ell(0)$ within error ϵ with probability of failure at most $1/3$ using a number of operator exponentials that is bounded above by*

$$\widetilde{O}\left(\frac{m^2 (25/3)^k \max \|H_i\| (1+\Gamma)}{\epsilon}\right),$$

where

$$\Gamma := \max_{p=1,\ldots,n} \left(\frac{p}{n}\right) \left(\frac{2ke^2 (5/3)^{k-1} m \max_{l\in[1,m]} \|H_l\|}{\gamma_{\text{eff}}}\right)^{1/p}$$

and $n \in \widetilde{O}\left(\log\left(\frac{mk(5/3)^k \max_i \|H_i\|(1+\Gamma)}{\epsilon}\right)\right)$.

The proof of this theorem is technical and so we direct the interested reader to Appendix A for a complete proof. As a brief sketch, the proof proceeds by using perturbation theory to evaluate the derivatives of the eigenvalues and eigenvectors of the effective Hamiltonian and multiple applications of the triangle inequality and similar linearizing approximations that are valid under our assumptions on t.

The above analysis is generic for any parameterized Hamiltonian $H(s)$. However, to give a better intuition for how it could be used let us focus our attention on a phase estimation protocol. There are multiple phase estimation procedures that can be employed, but the basic idea behind phase estimation is that one provides a quantum state $|\psi\rangle = \sum_j a_j |j(s)\rangle$ and then performs a series of evolutions of the form $e^{-iH(s)s}$ to this state to yield an estimate of one of the eigenvalues, $\exp(-iE_k(s)s)$, which is randomly sampled with probability $|a_j|^2$. We will usually demand that the variance of this estimate is bounded above by δ^2 and that the expected error is at most ϵ for such a procedure, but it is also common for the accuracy guarantees to be given in terms of a probability of failure.

3.1 Estimation of Trotter Error Using Extrapolation

One important issue that our method enables is that of estimating the error in Trotter-Suzuki formulas. This is a substantial issue because existing bounds are typically not

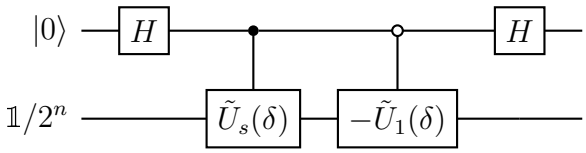

Figure 1: Quantum algorithm for computing a normalized Frobenius norm where the probability of measuring the first qubit to be 0 is $\|S_{2k}(\delta) - \tilde{U}_s(\delta)\|_F^2/(4D)$ for any fixed s and δ where $\|\cdot\|_F$ is the Frobenius norm.

tight and leading order expansions for the error are prohibitively expensive even for short evolutions [18]. Here we provide a way to address this via a method for computing a Frobenius distance between two unitaries,

$$\|S_{2k}(\delta) - e^{-iH\delta}\|_F = \lim_{s\to 0} \sqrt{\mathrm{Tr}\left((\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))(\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))^\dagger\right)} \tag{3.1}$$

Our aim is to estimate this distance by extrapolating the distance to $s = 0$. A major challenge though that arises is that we only naturally get a normalized Frobenius distance that corresponds to the average square singular value of $\tilde{U}_s(\delta) - \tilde{U}_1^\dagger(\delta)$.

Corollary 8. *If the assumptions of Theorem 7 hold then a quantum algorithm exists that can compute for the k^{th}-order Trotter Suzuki formula the quantity $\|S_{2k}(\delta) - e^{-iH\delta}\|_F^2$ for $H = \sum_{j=1}^m H_j$ within ϵ error and failure probability ϵ using a number of operator exponentials of the H_j that is in*

$$N_{\text{exp}} \in \tilde{O}\left(\frac{m^2(25/3)^k \max\|H_i\|}{\epsilon}\right)$$

Proof. Consider the circuit of Figure 1. We first demonstrate that the probability of measuring the first qubit to be zero is $\|S_{2k}(\delta) - \tilde{U}_s(\delta)\|_F^2/(4D)$. This fact follows from the analysis of the LCU lemma; however, that analysis is typically performed for pure states matrices and here we need the generalized version of it wherein the input state is a density matrix. Let $\rho = \mathbb{1}/2^n$ be the input state to the algorithm. Then the gate operations performed maps

$$\begin{aligned}
|0\rangle\langle 0| \otimes \rho &\mapsto_H \sum_{ij} \frac{|i\rangle\langle j|}{2} \otimes \rho \\
&\mapsto_{\text{SELECT}} \sum_{ij} \frac{|i\rangle\langle j|}{2} \otimes (\tilde{U}_s(\delta)^i(-\tilde{U}_1(\delta))^{1-i})\rho(\tilde{U}_s(\delta)^j(-\tilde{U}_1(\delta))^{1-j})^\dagger \\
&\mapsto_H \sum_{ijk\ell} (-1)^{ik+j\ell}\frac{|k\rangle\langle \ell|}{4} \otimes (\tilde{U}_s(\delta)^i(-\tilde{U}_1(\delta))^{1-i})\rho(\tilde{U}_s(\delta)^j(-\tilde{U}_1(\delta))^{1-j})^\dagger := \sigma
\end{aligned} \tag{3.2}$$

We then have that the probability of measuring the first qubit to be zero is

$$\mathrm{Tr}((|0\rangle\langle 0| \otimes \mathbb{1})\sigma) = \frac{1}{4}\mathrm{Tr}\left((\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))\rho(\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))^\dagger\right) \tag{3.3}$$

Taking $\rho = \mathbb{1}/2^n$ then yields

$$\frac{1}{4}\mathrm{Tr}\left((\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))\rho(\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))^\dagger\right) = \frac{\mathrm{Tr}\left((\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))(\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta))^\dagger\right)}{2^{n+2}}$$
$$= \frac{\|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2}{2^{n+2}} \tag{3.4}$$

Thus the probability of measuring 0 on the first qubit gives a normalized Frobenius distance between the two operators. The cost of doing this is $O(1)$ queries to the underlying Trotter-Suzuki formulas, each of which boils down to $O(m5^k)$ operator exponentials.

Using Amplitude estimation on the target, we can construct an operator W such that the eigenvalues of W within the subspace supporting the initial state are of the form

$$\lambda(W) = e^{\pm i \sin^{-1}\left(\sqrt{\frac{\|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2}{2^{n+2}}}\right)}. \tag{3.5}$$

We then can invoke Theorem 7 to show that the number of exponentials needed to learn the extrapolated phase under the assumptions of the Theorem are with probability greater than 2/3 within error ϵ' obeys

$$N_{\exp} \in \tilde{O}\left(\frac{m^2(25/3)^k \max \|H_i\| \log^{5/2}(\max_i \|H_i\|/\epsilon')}{\epsilon'}\right) \tag{3.6}$$

The remaining question is how small ϵ' must be to guarantee that the error is at most ϵ in our estimate of the Frobenius distance squared. Let $\hat{\phi}$ denote the estimate of the phase that is returned by our protocol which has error at most ϵ'. Our estimate of the distance squared, $\hat{D}^2$ is then found by

$$\hat{D}^2 = 2^{n+2}\sin^2(\hat{\phi}) \tag{3.7}$$

Thus we have that

$$|\hat{D}^2 - \|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2| \leq 2^{n+2} \max_\theta |\partial_\theta \sin^2(\theta)|\epsilon' \leq 2^{n+3}\epsilon'. \tag{3.8}$$

Thus we can estimate the mean square singular value of the difference between the two operators within error ϵ

$$|\hat{D}^2/2^n - \|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2/2^n| \leq \epsilon' 8 \Rightarrow \epsilon' = \epsilon/8. \tag{3.9}$$

Thus the cost for learning the mean square singular value of the difference between the two unities is

$$N_{\exp} \in \tilde{O}\left(\frac{m^2(25/3)^k \max \|H_i\| \log^{5/2}(\max_i \|H_i\|/\epsilon')}{\epsilon}\right) \tag{3.10}$$

□

Corollary 9. *Assume that* $\mathrm{Tr}(|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)|^4)/2^n - (\mathrm{Tr}(|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)|^2/2^n))^2 \leq \xi$. *If we let* E_j^2 *describe an eigenvalue of the square of the difference satisfies*

$$\Pr(|E_j^2 - \|\tilde{U}_s^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2| \leq k\sqrt{\xi}) \leq \frac{1}{k^2}$$

Further, if we have no such promise about the variance then the following weaker bound can be shown

$$\Pr(|E_j^2 - \|\tilde{U}_0^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2| \leq k\|\tilde{U}_0^\dagger(\delta) - \tilde{U}_1^\dagger(\delta)\|_F^2) \leq 1/k$$

Proof. The result trivially follows from the Chebyshev and Markov inequalities. □

This shows that we can use our extrapolation procedure to estimate largest eigenvalue of the error operator. In particular, let us let the probability of an eigenvalue being greater than the estimate be $O(1/2^n)$. This implies that the above Chebyshev inequality bound implies that it suffices to take $k \in O(\sqrt{2^n})$. Thus with high probability all of the eigenvalues for the square of the error operator will be at most $\hat{D}^2/2^n + O(\sqrt{\xi 2^n})$. Thus if $\xi \in o(\hat{D}^4/2^{3n})$ then the estimate yielded by this procedure will also estimate the spectral norm.

4 Time-evolved expectation values

Computing expectation values is a fundamental task in quantum theory, as it encodes all the information one can observe from a quantum system. Here we explore the possibility of using polynomial interpolation to compute the expectation values of a system evolved for time t. Given a quantum state ρ and observable O, the expectation value is given by $\mathrm{Tr}\,(\rho O)$. We imagine evolving our system in time according to a product formula with $1/s$ time steps (in a generalized sense, note that s is real valued). This evolution could be performed in either the Schordinger or Heisenberg picture, and we will focus on the Heisenberg picture for sake of definiteness. The expectation values we wish to compute are of the form

$$f(s) := \frac{\mathrm{Tr}\,(\rho O_s(t))}{\|O\|} \tag{4.1}$$

where

$$O_s(t) := S_{2k}^\dagger(st)^{1/s} O S_{2k}(st)^{1/s} \tag{4.2}$$

is the Trotterized Heisenberg evolution. We've normalized the expectation values by $\|O\|$ since the relative error is a more useful and natural metric. The interpolation algorithm we propose can be summarized as follows.

1. Given H, t, and ϵ, choose the appropriate interpolation interval $[-a, a]$ and number of interpolation points n. The cost of this step is negligible. The error analysis we will perform subsequently will inform the choices of a, n.

2. For each $s_1, \ldots s_{n/2}$, compute r_i, the integer part of $1/s_i$. This step also has negligible cost.

3. Compute estimates f_i of the expectation values of $\langle O_{1/r_i}(t)\rangle$ for each r_i with $i = 1, \ldots, n/2$, to an accuracy depending on ϵ and n. We will assume this step is done with Iterative Quantum Amplitude Estimation (IQAE), a state-of-the-art quantum algorithm with Heisenberg-limited efficiency. The cost of this step is defined as the number of exponentials of H_j performed on a quantum circuit, where $H = \sum_j H_j$. Note that by symmetry, we need not compute f_i for $i > n/2$. We have $f_i = f_{n-i+1}$ for all $i \in [1, n]$.

4. Perform the polynomial fit $P_{n-1}f$ through the points (s_i, f_i). Note that $P_{n-1}f$ will automatically be even. This fit is well-conditioned, and we neglect the cost of this step.

5. Evaluate the $P_{n-1}f$ at $s = 0$. This is our final estimate of $\langle O(t)\rangle$, guaranteed to have accuracy ϵ.

Note that we have chosen to incur an error by using the integer part r_i of $1/s_i$, in order to have an integer number of Trotter steps. However, the fractional part could be included as desired, using a fractional query algorithm through the machinery of Quantum Singular Value Transformations. We choose to simply ignore the fractional part for simplicity, both conceptually and to avoid large overheads for implementation on near-term hardware. In subsequent analysis, we'll find this fractional part becomes increasingly negligible as a gets smaller, in a way that does not affect the resulting asymptotics greatly.

Given an even set of Chebyshev nodes $\{s_1, \ldots, s_n\}$, and making use of Lemma 2, the interpolation error E_{n-1} assuming perfect data points is given by

$$|E_{n-1}(0)| \leq \frac{|\mathrm{Tr}\left(\rho\, \partial_s^n O_s(t)\right)|}{\|O\|\, n!} \prod_{i=1}^{n} |s_i| \leq \max_{s\in I_a} \frac{\|\partial_s^n O_s(t)\|}{\|O\|} \left(\frac{a}{2n}\right)^n \tag{4.3}$$

where $I_a = [-a, a]$. With a suitable bound on $\partial_s O(t)$, we can provide an upper bound on the interpolation error at $s = 0$. This bound is provided by the following lemma.

Lemma 10 (Error extrapolation for time-evolved observables)**.** *Under the conditions of Lemma 1 ($ca \leq \pi/20$) where $c := 2k(5/3)^{k-1} m \max_{l\in[1,m]} \|H_l\|\, t$ the following bounds holds on the Trotterized evolution $O_s(t)$ with step parameter $s \in (0, a]$:*

1. *for $2k(5/3)^{k-1} m \max_{l\in[1,m]} \|H_l\|\, t > n$ ("long time simulation") we have have that*

$$\frac{\|\partial_s^n O_s(t)\|}{\|O\|} < (ec)^{2n}$$

which gives a relative interpolation error

$$\frac{|E_{n-1}(0)|}{\|O\|} < \left(\frac{e^2 c^2 a}{2n}\right)^n.$$

2. *For* $2k(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|\,t \leq n$ *("short time simulation") we have a derivative bound*

$$\frac{\|\partial_s^n O_s(t)\|}{\|O\|} \leq \frac{3}{2\pi}\sqrt{n-1}\left(\frac{e^2c}{2}\right)^n n! e^{4c/\sqrt{\pi}}$$

giving an interpolation error of the normalized expectation value as

$$|E_{n-1}(0)| \leq \frac{4}{3}n\left(\frac{eca}{4}\right)^n e^{4c/\sqrt{\pi}}.$$

The proof of this lemma is an exercise in repeated use of triangle inequality and the combinatorics of large derivatives, and is left to Appendix C. Note that once the derivative bound holds, the interpolation error bound follows immediately from Lemma 2.

Our main motivation for these bounds is deriving asymptotic expressions for the cost of a polynomial interpolation algorithm which utilizes estimates for Trotter-evolved expectation values at various s_i. The following theorem characterizes the performance of our algorithm.

Theorem 11. *Let* $O(t) = U^\dagger(t)OU(t)$ *be the time-evolved observable under Hamiltonian* $H = \sum_{l=1}^m H_l$*, so that* $U(t) = e^{-iHt}$*. Let* $c := 2k(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|\,t \geq 1$*, and let* $\rho \in \mathcal{D}(\mathcal{H})$ *be a quantum state. The number of exponentials* N_{exp} *required to estimate* $\text{Tr}\,(\rho O(t))$ *to precision* ϵ *using our interpolation algorithm scales as*

$$N_{\text{exp}} \in \tilde{O}\left(\frac{mc^2}{\epsilon}\right).$$

Here, $\tilde{O}$ *is big-O with terms suppresed which are logarithmically smaller. In contrast, for the case of short time evolutions the cost is*

$$N_{\text{exp}} \in \tilde{O}\left(\frac{m}{\epsilon}\right).$$

We give a sketch of the proof for the theorem above. By treating N_{exp} as our cost, we are assuming the polynomial fitting of steps 2 and 3 are relatively cheap. Thus, the entire cost is encoded in the collection of data f_i, which is obtained using a set of Suzuki-Trotter evolutions on a quantum circuit. Each estimate needs to be made essentially within ϵ of the exact Trotter value. Thankfully, the well-conditioning of the Chebyshev nodes means we do not need to be much more accurate than this to get a good interpolation. We can therefore estimate each $\langle O_i\rangle$ using $O(1/\epsilon)$ measurements of a Hadamard-test circuit, as part of an Amplitude Estimation which encodes the expectation value.

Data points taken closer to $s = 0$ will be more costly, going as $1/s$. For the particular case of a Chebyshev interpolation, we can show that the number of exponentials scales with n and a as

$$N_{\text{exp}} \in O\left(\frac{n\log n}{a}\right). \tag{4.4}$$

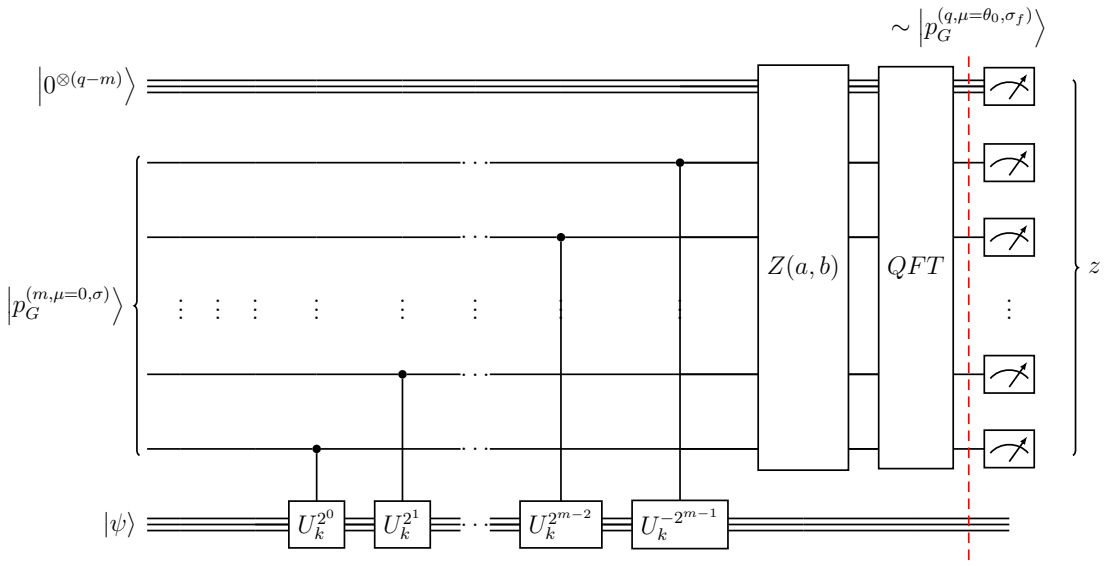

Figure 2: Circuit to implement a Gaussian m-qubit phase estimation algorithm with $(m-q)$-qubits for spectral interpolation. The U_k operator is the evolution operator to an effective Hamiltonian $H_{\text{eff},k}$, which corresponds to a Trotter step size τ_k.

Our error bounds from Lemma 10 will tell us how large n and a must be to achieve the precision ϵ. They will also depend on the order of the Trotter formula used, the length of time t, and the Hamiltonian "size" captured by $\max_{l\in[1,m]} \|H_l\|$. Obtaining this functional relationship allows us to express the scaling (4.4) directly in terms of parameters of the problem. We carry out this procedure in the formal proof provided in Appendix C.

5 Gaussian phase estimation algorithm and extrapolation

In this section, we propose an alternative, also optimally-conditioned, extrapolation approach for eigenenergy estimation in which we first prepare the effective eigenstate through the procedure in [25], except using one single qubit the semi-classical QFT, then perform Gaussian phase estimation on said eigenstate using the respective effective Hamiltonian. The result is summarized in the following theorem.

We will first demonstrate how to do the phase estimation algorithm [26], but instead of using a rectangular cutoff window, we will use a gaussian tapering window in the time domain. This will provide us with a mechanism to sample an amplitude or phase as a unbiased distribution as assumed by the proposed methods (Lemma 5) in previous sections.

The first step in the protocol involves preparing a renormalized sample of a Gaussian

distribution in the ancillary register

$$\left|p_G^{(m,\mu=0,\sigma)}\right\rangle = \frac{1}{\sqrt{\mathcal{N}(\sigma,T,m)}} \sum_{x=-(2^m/2-1)}^{2^m/2-1} \sqrt{p_G(x\,T;0,\sigma)}\,|x\rangle \tag{5.1}$$

Here, $p_G(w;\mu;\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(w-\mu)^2}{2\sigma^2}$, and T plays the role of a sampling rate.

An exact sample of the exponential distribution can be prepared through the methods proposed in Ref. [27]. However, within a target error, we can prepare a coarser Gaussian distribution and then perform an interpolation through QFT, zero-padding, and again QFT^{-1}. This is an established method for upsampling/interpolation of discrete signals in classical discrete signal processing, and was introduced for quantum distribution preparation in Ref [28].

In Figure 2 we illustrate the circuit used for this method. The operator $Z(a,b)$ is serving as the zero-padding used so often in DSP for interpolation in the conjugate space in which the padding is being done. The errors introduced by truncation and finite sample rate in the time domain are estimated in Appendix D, and can be summarized with the following

Theorem 12. *Let q, m be positive integers such that $q \geq m$ and the spacing in Fourier domain F obeys $F = \frac{1}{2^q T}$ and the sampling rate T obeys $\sigma/T = \sqrt{2^m}$. We then have that if and $\sigma_f = \frac{1}{4\sigma\pi}$ then the spectral norm of the error of the prepared wave function in Fourier domain is is*

$$\left\| \sum_k \left(\frac{1}{\mathcal{N}(\sigma_f,F,q)^{1/2}} X\left(\frac{k}{2^q T}\right)\right)|k\rangle - \right.$$
$$\left. \mathrm{QFT} \sum_x \frac{1}{\sqrt{\mathcal{N}(\sigma,T,m)}} \sqrt{p_G(x\,T,\sigma,0)}\,|x\rangle \right\| = O\left(\frac{(2^m)^{1/2}}{e^{\Omega(2^m)}}\right). \tag{5.2}$$

Here,

$$X = \mathcal{F}\{\sqrt{p_G}\}(f) = \int \sqrt{p_G(x\,T,\sigma,0)} e^{-2\pi i x f}\, dt$$

and the discrete normalization for the Gaussian in in the real-space domain is

$$\mathcal{N}(\sigma,T,m) = \sum_{n=-(2^m/2-1)}^{2^m/2-1} p_G(n\,T,\sigma,0).$$

Then using these errors in the approximate state constructed in Fourier domain, we can perform a sequence of unitaries controlled on the Gaussian distribution. This approach is similar in spirit to the Kaiser-window approach taken in [29] however a Gaussian distribution is used instead here.

We can then bound the cost for extrapolating energies using Gaussian phase estimation via these results. In the following, we use a slightly different formalism that relies on Bernstein ellipses in the complex plane rather than a Taylor series analysis of the function in real-space. The cost of the algorithm is summarized below.

Theorem 13. *Given a Hamiltonian $H = \sum_{j=1}^{m} H_j$ such that $\sum_{j=1}^{m} \|H_j\| \leq 1$ and the lower bound on its spectral gap is Δ, there exists a state preparation algorithm using controlled queries of $U_k = (S_p(ts))^{d_k^{(M)}}$, with $t = \Theta\left(\Delta^{1/p}\right)$ to avoid any discontinuities/level-crossings, and $d_k^{(M)} = \operatorname{sgn} s_k^{(M)} \left\lceil s_1^{(M)} / \left| s_k^{(M)} \right| \right\rceil$, where $s_k^{(M)} = \cos((2k-1)\pi/2M)$, with a total cost of*

$$C_{\text{prep}} = O\left(\frac{\log \frac{1}{\epsilon} \log\log \frac{1}{\epsilon}}{\Delta^{1+1/p}}\right),$$

requiring a single ancillary qubit. Here, $\epsilon = |E_m(0) - P_{M-1}E_m(0)|$ is the bias in the estimate of the energy, and $P_{M-1}E_m$ is the (unique) $(M-1)$-degree polynomial interpolating E_m at the M points $s_i^{(M)}$. Moreover, the cost of estimating all the observables using Gaussian phase estimation is

$$C_{\text{est}} = O\left(\frac{\log \frac{1}{\epsilon}}{\sigma_P}\right),$$

where σ_P is the standard deviation on the interpolated observable at $s = 0$ and requires $O\left(\log \frac{1}{\sigma_P}\right)$ ancillary qubits. Alternatively, using a single-ancillary-qubit approach,

$$C_{\text{est,1-qubit}} = O\left(\frac{\log \frac{1}{\epsilon} \log\log \frac{1}{\epsilon}}{\epsilon'}\right),$$

where ϵ' is defined through $\max_s |E_m(s) - P_{M-1}\tilde{E}_m(s)| \leq \epsilon'$ and $P_{M-1}E_m$ is the (unique) $(M-1)$-degree polynomial interpolating $\tilde{E}_m$ at the M points $s_i^{(M)}$. At the same time, $\tilde{E}_m(s_i^{(M)})$ are the estimated energies through an appropriate 1-qubit phase estimation with the following bound on their errors

$$\max_i |E_m(s_i^M) - \tilde{E}_m(s_i^M)| \leq \epsilon' / (\frac{2}{\pi} \log(M+1) + 1).$$

The bound for C_{est} can also be cast in terms of a confidence interval, $\epsilon' = w\sigma_P$, around the mean, which introduces an error rate that decreases super-exponentially with w.

This approach performs similarly to the previous methods except that with this approach we have specific claims that can be made about the bias and variance of the estimates returned by our extrapolation.

6 Numerical Experiments

Previously, we showed that extrapolation methods can be used to estimate eigenvalues using phase estimation with low-order Trotter formulas that nonetheless yield precision scaling that is asymptotically better than what would ordinarily be expected from even

very high-order Trotter formulas. Here we provide numerical evidence backing up our analytic claims that shows that improved scaling is not only possible, but also that high-order Trotter formulas need not always provide better error scaling when extrapolation is considered. Specifically, demonstrate this improved scaling for phase estimation using extrapolation for the first-order, second-order and fourth-order Trotter-Suzuki formulas. The model that serves as a testing ground is the transverse Ising model:

$$\tilde{H} = -J\left(\sum_{\langle i,j\rangle} Z_i Z_j + g\sum_j X_j\right). \tag{6.1}$$

This model is a minimal example with just two non-commuting terms and thus a product formula simulation will yield non-zero Trotter error. Our aim will be to estimate the groundstate energy of the effective Hamiltonian $\tilde{H}$ in the limit where the duration of the Trotter step approaches zero and see numerical evidence that the scaling is asymptotically improved relative to low-order Trotter formulas as suggested by Theorem 13.

We will connect $\tilde{H}$ to H shortly. In order to avoid aliasing with our Fourier spectral methods, we must satisfy the bound

$$\left\|H_{\text{eff},\ell}\left(\tau_k^{(M)}\right)\right\| d_k^{(M)}\Delta_{t,M}\tau_k^{(M)} + \Delta_{\text{pad}} \leq 1. \tag{6.2}$$

The padding is there to suppress the probability leakage wrapping around the boundary. Now, we define an ancillary operator with

$$H'_{\text{eff},\ell}\left(\tau_k^{(M)}\right) = H_{\text{eff},\ell}\left(\tau_k^{(M)}\right) d_k^{(M)}\Delta t/2, \tag{6.3}$$

such that

$$\left\|H'_{\text{eff},k}(\tau)\right\| + \Delta_{\text{pad}}/2 \leq 1/2. \tag{6.4}$$

This way, the spectrum of $H'_{\text{eff},k}(\tau)$ lies within the domain $[-1/2+\Delta_{\text{pad}}/2, 1/2-\Delta_{\text{pad}}/2]$ and now is in line with the conventions of Fourier spectral methods in [25]. With this, we now define

$$\tilde{H} = \Delta_{t,0} H'_{\text{eff},\ell}(\tau = 0). \tag{6.5}$$

Here $\Delta_{t,0}$ is sufficiently small as to fulfill Equation (6.4) and as to the method presented in this section, we must also make sure that there are no level crossings through out the sampling interval.

We perform the extrapolation for second- and fourth-order formulas at different values of M. We display the second-order extrapolations in Figure 3. For each $H'_{\text{eff},\ell}\left(\tau_k^{(M)}\right)$, we implement the ground state preparation protocol as outlined in [25] to a negligible state error, ϵ_{state}. This efficient since the algorithm scales $O(\log 1/\epsilon_{\text{state}})$. We then perform the Gaussian QPE on these states. Moreover, we have used zero-padding as described in the Section 5 in order to interpolate the the spectrum and be able to ignore digitization errors. The spectral interpolation through zero-padding is efficient

since the cost scales like $O(q^2)$ (or $\tilde{O}(q)$ if approximate QFT is used instead of exact QFT), where q is the number of total qubits after padding. The blue bands in these plots show the 1σ confidence intervals which correspond to $\sim 68\%$ confidence levels under assumptions of Gaussianity.

We calculate the systematic error numerically coming from the truncation of the polynomial expansion and plotted its value against M in Figure 4 and compare against the exact error along with the upper bounds from combining Theorem 19, Theorem 17 and Lemma 15. We see from this data that interpolation indeed improves the quality of the estimate of the energy as anticipated, even when relatively low-precision estimates are used at $\tau = 1$. Note that because of the symmetry of the Trotter error as $\tau \to -\tau$, only positive values of τ need to be actually evaluated here.

We also performed ground state energy estimation using second- and fourth-order formulas without extrapolation in order to compare the costs when plotted against the error. From the plots in Figure 5, we note that beyond $M = 2$ the extrapolation methods already outperform using second-order product formulas alone. Specifically, we see clear indications from this data that the bias in the interpolated error (for large $1/\epsilon$) scales logarithmically for these Gaussian phase estimation experiments. This agrees with our expectations from Theorems **??** and 13 wherein the systematic errors from interpolating phase estimation experiments are expected to scale as polylog$(1/\epsilon)$. In contrast, the bias from a fixed experiment using QPE is expected to be $O(1/\epsilon^{3/2})$ and $O(1/\epsilon^{5/4})$ respectively. It is worth noting that this does not violate the Heisenberg limit as this only provides an approximately unbiased estimate of the value, but does not measure the uncertainty directly. The exponential reduction in the uncertainty with the number of interpolation points considered is independently shown in Figure 4. Note that for this data we do not see an improvement from transitioning to higher-order formulas, which is consistent with our prior expectations. These results show that extrapolation tends to outperform transitioning to a higher-order product formula does not necessarily lead to an improved scaling for the problem of phase estimation. Apart from cases where the output of the quantum state is the expectation value of an observable rather than a quantum state that extrapolation will typically outperform approaches that try to achieve high accuracy by going to higher-order formulas. The most likely cases where higher-order formulas may be helpful is in cases where dynamics need to be simulated such as in estimating eigenvalues of a Floquet operator for a time-dependent Hamiltonian evolution wherein many Trotter steps could be needed to implement the time-evolution operator within sufficient accuracy; however, at present the case for high-order formulas for the time-independent case is less obvious from these results.

7 Conclusion

We have presented here two alternative approaches for Trotter extrapolation: one in which we extrapolate the Hamiltonian evolution operator and on the other we extrapo-

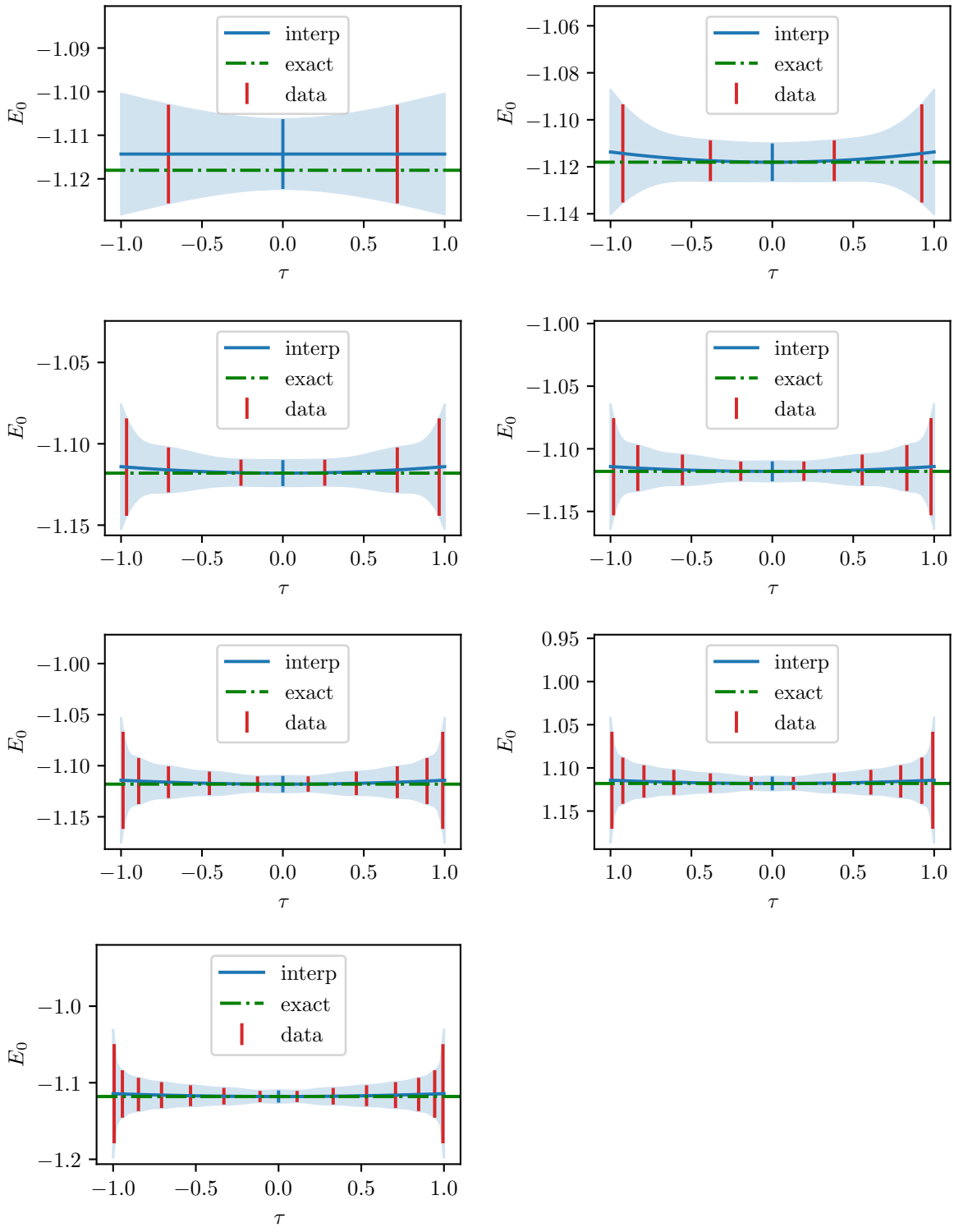

Figure 3: Here we display interpolation results using the probabilistic approach for $M = 2, 4, \ldots, 14$ using second order Trotter formula.

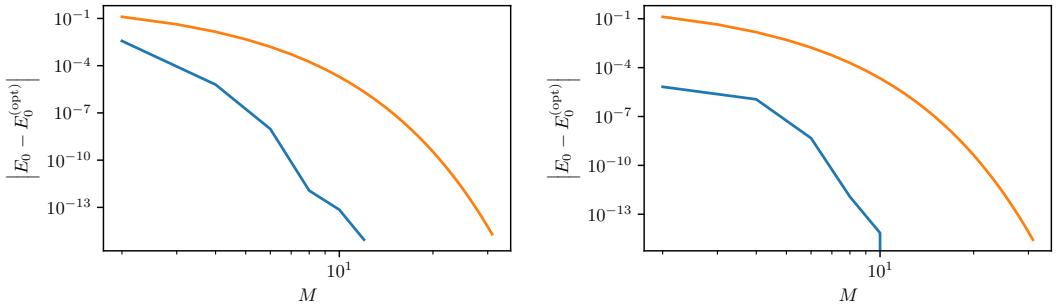

Figure 4: The systematic error coming from truncating the polynomial expansion. In blue we have the exact result for the error and in orange are the upper bounds for the error. On the left, we used a second-order formula and on the right a fourth-order one. As we can see from this log-log plot, the convergence is exponential as expected.

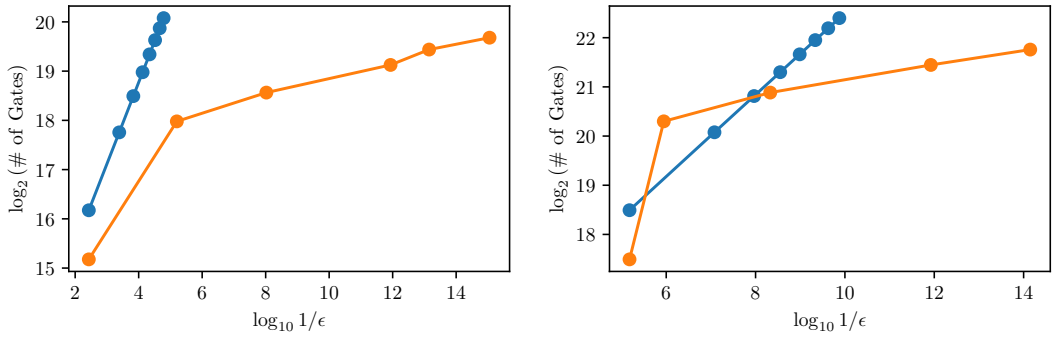

Figure 5: The gate cost from using second- (left) and fourth-order (right) formulas for ground state estimation plotted against the systematic error on the eigenvalue estimation. In blue, we have the cost of using single long-depth circuits using product formulas alone. In orange, we have the gate cost of using multiple shallower circuits and then using the extrapolation methods presented in this work.

late the ground state energy of the effective Hamiltonian. This approach is highly favorable for quantum phase estimation and leads to a polynomial speedup from $O(1/\epsilon^{3/2})$ for the second order Trotter-formula to $\tilde{O}(1/\epsilon)$ by extrapolating the results for phase estimation down to zero time step duration for the Trotter-step. We find scalings that are on the order of $\tilde{O}(t^2/\epsilon)$, which is a polynomial improvement over the first-order Trotter formula; however, the extrapolation may not provide improved results for larger values of t. This issue may arise because of looseness in the bounds used in our analysis.

We apply these methods to Trotter error estimation and show that the Frobenius distance between two unitary matrices can be estimated at cost $\tilde{O}(1/\epsilon)$ for constant dimension. This provides a constructive method to estimate the Trotter error and in some cases gives a way to address the fact that the error in the Trotter-Suzuki approximation is difficult to accurately bound using existing approaches. We further demonstrate the effectiveness of our approach numerically using our Gaussian phase estimation for transverse Ising models.

There are a number of open questions revealed by this work. The first question involves asking how transitioning to high-order formulas impacts the extrapolation error. At present, our bounds appear to suggest that for phase estimation that there's no asymptotic benefits from transitioning to higher-order formulas. In practice, this is unlikely to be true as higher order formulas can sometimes take fewer exponentials to get into the regime where the error switches from exponential to polynomial. and thus becomes ammenable to extrapolation.

An important second question involves the application of these ideas. As Trotter formulas provide the best scaling algorithms for chemistry, investigating how extrapolation can be used to improve them is a significant remaining question. Such investigations will require more detailed analysis of the commutator bounds in the expansion since these commutators are vital for proving the scalings found in these results.

More broadly, this work shows that error extrapolation can be a powerful method for estimating energies using low-order formulas. Thus this work suggests that investigation of highly efficient low-order formulas for simulation is likely to be an increasingly important topic given that our present results ammeliorate the worst aspects of these formulas. Improvements that arise from the superior constant factors that low order formulas can provide may lead to a new generation of vastly more efficient algorithms for simulating chemistry and physics using quantum computers.

8 Acknowledgements

During the inception of this work and the initial write-up which included most of Sections 1 and 2.2 and appendix D, G. R. was supported through Brookhaven National Laboratory, the Laboratory Directed Research and Development (LDRD) program No. 21-043, by Program Development Fund No. NPP PD 19-025, by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under the Contract No. DE-SC0012704 (BNL). During later additions to the manuscript, G.R. was supported

by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under contract number DE-SC0012704 (BNL). G.R. was also supported and employed by Zapata Computing, Inc. through part of the writing of the manuscript. G. R. is thankful to Christoph Lehner, Taku Izubuchi and Yuta Kikuchi for the insightful discussions.

References

[1] S. Lloyd, *Universal quantum simulators*, *Science* **273** (1996) 1073 [https://science.sciencemag.org/content/273/5278/1073.full.pdf].

[2] M. Reiher, N. Wiebe, K.M. Svore, D. Wecker and M. Troyer, *Elucidating reaction mechanisms on quantum computers*, *Proceedings of the national academy of sciences* **114** (2017) 7555.

[3] J.D. Whitfield, J. Biamonte and A. Aspuru-Guzik, *Simulation of electronic structure hamiltonians using quantum computers*, *Molecular Physics* **109** (2011) 735.

[4] J. Lee, D.W. Berry, C. Gidney, W.J. Huggins, J.R. McClean, N. Wiebe et al., *Even more efficient quantum computations of chemistry through tensor hypercontraction*, *PRX Quantum* **2** (2021) 030305.

[5] V. von Burg, G.H. Low, T. Häner, D.S. Steiger, M. Reiher, M. Roetteler et al., *Quantum computing enhanced computational catalysis*, *Physical Review Research* **3** (2021) 033055.

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

[7] A.F. Shaw, P. Lougovski, J.R. Stryker and N. Wiebe, *Quantum algorithms for simulating the lattice schwinger model*, *Quantum* **4** (2020) 306.

[8] N. Klco, M.J. Savage and J.R. Stryker, *Su (2) non-abelian gauge field theory in one dimension on digital quantum computers*, *Physical Review D* **101** (2020) 074512.

[9] A.M. Childs and N. Wiebe, *Hamiltonian simulation using linear combinations of unitary operations*, *arXiv preprint arXiv:1202.5822* (2012) .

[10] G.H. Low, V. Kliuchnikov and N. Wiebe, *Well-conditioned multiproduct hamiltonian simulation*, *arXiv preprint arXiv:1907.11679* (2019) .

[11] D.W. Berry, A.M. Childs, R. Cleve, R. Kothari and R.D. Somma, *Simulating hamiltonian dynamics with a truncated taylor series*, *Physical review letters* **114** (2015) 090502.

[12] G.H. Low and N. Wiebe, *Hamiltonian simulation in the interaction picture*, *arXiv preprint arXiv:1805.00675* (2018) .

[13] M. Kieferová, A. Scherer and D.W. Berry, *Simulating the dynamics of time-dependent hamiltonians with a truncated dyson series*, *Physical Review A* **99** (2019) 042314.

[14] G.H. Low and I.L. Chuang, *Hamiltonian Simulation by Qubitization*, *Quantum* **3** (2019) 163.

[15] R. Babbush, C. Gidney, D.W. Berry, N. Wiebe, J. McClean, A. Paler et al., *Encoding electronic spectra in quantum circuits with linear t complexity*, *Physical Review X* **8** (2018) 041015.

[16] D.W. Berry, G. Ahokas, R. Cleve and B.C. Sanders, *Efficient quantum algorithms for simulating sparse hamiltonians*, *Communications in Mathematical Physics* **270** (2006) 359–371.

[17] N. Wiebe, D.W. Berry, P. Høyer and B.C. Sanders, *Simulating quantum dynamics on a quantum computer*, *Journal of Physics A: Mathematical and Theoretical* **44** (2011) 445308.

[18] A.M. Childs, Y. Su, M.C. Tran, N. Wiebe and S. Zhu, *Theory of trotter error with commutator scaling*, *Physical Review X* **11** (2021) 011020.

[19] J. Haah, M.B. Hastings, R. Kothari and G.H. Low, *Quantum algorithm for simulating real time evolution of lattice hamiltonians*, *SIAM Journal on Computing* (2021) FOCS18.

[20] M. Hagan and N. Wiebe, *Composite quantum simulations*, *arXiv preprint arXiv:2206.06409* (2022) .

[21] G.H. Low, Y. Su, Y. Tong and M.C. Tran, *On the complexity of implementing trotter steps*, *arXiv preprint arXiv:2211.09133* (2022) .

[22] G.H. Low and I.L. Chuang, *Optimal hamiltonian simulation by quantum signal processing*, *Physical Review Letters* **118** (2017) .

[23] A. Gilyén, Y. Su, G.H. Low and N. Wiebe, *Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics*, STOC 2019, (New York, NY, USA), p. 193–204, Association for Computing Machinery, 2019, DOI.

[24] W. Gautschi, *How (un)stable are vandermonde systems? asymptotic and computational analysis*, in *Lecture Notes in Pure and Applied Mathematics*, pp. 193–210, Marcel Dekker, Inc, 1990.

[25] G. Rendon, T. Izubuchi and Y. Kikuchi, *Effects of Cosine Tapering Window on Quantum Phase Estimation*, *arXiv e-prints* (2021) arXiv:2110.09590 [2110.09590].

[26] D.S. Abrams and S. Lloyd, *Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors*, *Physical Review Letters* **83** (1999) 5162 [quant-ph/9807070].

[27] N. Klco and M.J. Savage, *Minimally entangled state preparation of localized wave functions on quantum computers*, *Physical Review A* **102** (2020) 012612 [1904.10440].

[28] J. José García-Ripoll, *Quantum-inspired algorithms for multivariate analysis: from interpolation to partial differential equations*, *arXiv e-prints* (2019) arXiv:1909.06619 [1909.06619].

[29] W. Górecki, R. Demkowicz-Dobrzański, H.M. Wiseman and D.W. Berry, *π-corrected heisenberg limit*, *Physical review letters* **124** (2020) 030501.

[30] R.A. Horn and C.R. Johnson, *Matrix analysis*, Cambridge university press (2012).

[31] L.N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition*, SIAM (2019).

[32] F.L. Bauer and C.T. Fike, *Norms and exclusion theorems*, *Numer. Math.* **2** (1960) 137–141.

[33] A.M. Childs, Y. Su, M.C. Tran, N. Wiebe and S. Zhu, *A theory of trotter error*, 2020.

[34] D. Grinko, J. Gacon, C. Zoufal and S. Woerner, *Iterative quantum amplitude estimation*, *npj Quantum Information* **7** (2021) 52 [1912.05559].

[35] B.L. Higgins, D.W. Berry, S.D. Bartlett, H.M. Wiseman and G.J. Pryde, *Entanglement-free Heisenberg-limited phase estimation*, *Nature* **450** (2007) 393 [0709.2996].

[36] R.B. Griffiths and C.-S. Niu, *Semiclassical Fourier Transform for Quantum Computation*, *Physical Review Letters* **76** (1996) 3228 [quant-ph/9511007].

[37] A.Y. Kitaev, *Quantum measurements and the abelian stabilizer problem*, *arXiv preprint quant-ph/9511026* (1995) .

A Proof of lemma 1: bounds on derivatives of effective Trotter Hamiltonian

Proof of Lemma 1. For convenience, here we recall the definition of the effective Hamiltonian.

$$H_{\text{eff}} := \frac{i}{st} \log S_{2k}(st) \tag{A.1}$$

We will understand $\log S_{2k}(st)$ through a power series expansion about the identity.

$$\log S_{2k}(st) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j+1} (S_{2k}(st) - \mathbb{1})^{j+1} \tag{A.2}$$

The convergence of this series is guaranteed precisely when the eigenvalues $e^{i\varphi}$ of $S_{2k}(st)$ satisfy

$$\left|e^{i\varphi} - 1\right| \leq 1 \implies |\varphi| \leq \pi/3 \tag{A.3}$$

for φ chosen in the branch $[-\pi, \pi]$. Since $\lim_{s\to 0} S_{2k}(st) = \mathbb{1}$, this can be guaranteed for sufficiently small s. To be more precise, let's use the fact that

$$S_{2k}(st) = e^{-i(Hst + \mathcal{E}(st))} \tag{A.4}$$

where the error operator $\mathcal{E}(ts)$ is in $O((ts)^{2k})$. Without loss of generality we may take $\|H\| \leq 1$ by increasing t appropriately, and in practice we will only consider the regime where $st < 1$, since this is required for accurate Trotter simulations in the first place. We see then that this condition is also approximately what we need for convergence of the series.

Within a radius of convergence $st < R$ (for some $R > 0$) this series can be differentiated term by term. Since $\log S_{2k}(0) = 0$, $s = 0$ is a zero of order at least one. We want to absorb the diverging $1/s$ term and better understanding the leading dependence in s. To facilitate this we write

$$H_{\text{eff}} = -\frac{1}{it} \sum_{j=0}^{\infty} \frac{(-1)^j}{j+1} s^j \Delta S_{2k}(st)^{j+1} \tag{A.5}$$

where we defined

$$\Delta S_{2k}(st) := \frac{S_{2k}(st) - \mathbb{1}}{s}. \tag{A.6}$$

Note that ΔS_{2k} is analytic in s, and is a finite difference around $s = 0$, i.e. $\lim_{s\to 0} \Delta S_{2k}(st) = -iHt$. Through the series expansion (A.5) we may bound derivatives of H_{eff} via bounds on derivatives of ΔS_{2k}. We first obtain a power series of ΔS_{2k} by Taylor expanding every term in the product formula S_{2k}. Regrouping in powers of st, the result its

$$\Delta S_{2k}(st) = \sum_{j=1}^{\infty} \frac{s^{j-1}(-it)^j}{j!} \sum_{J} \binom{j}{j_1 \dots j_{N_k}} \prod_{l=1}^{N_k} (H_l \tau_l)^{j_l} \tag{A.7}$$

where $N_k = 2m5^{k-1}$ denotes the number of exponentials in S_{2k} and τ_l denotes the lth weight for the exponential. The sum $\sum_J$ is over all values of $J = (j_1, \ldots, j_{N_k})$ such that $\sum_k j_k = j$. The derivatives with respect to s are now easy to compute. Using the fact that

$$\partial_s^n s^{j-1} = \frac{(j-1)!}{(j-1-n)!} s^{j-n-1} \tag{A.8}$$

for $p > n$ (and zero otherwise), we have

$$\begin{aligned} \partial_s^n \Delta S_{2k}(st) &= \sum_{j=n+1}^{\infty} \frac{s^{j-n-1}(-it)^j}{j!} \frac{(j-1)!}{(j-1-n)!} \sum_J \binom{j}{j_1 \ldots j_{N_k}} \prod_{l=1}^{N_k} (H_l \tau_l)^{j_l} \\ \|\partial_s^n \Delta S_{2k}(st)\| &\le \sum_{j=n+1}^{\infty} \frac{t^j}{(j-n-1)!} s^{j-n-1} (\tau_{\max} N_k \Lambda)^j \end{aligned} \tag{A.9}$$

where $\Lambda = \max_l H_l$ and $\tau_{\max} = \max_l |\tau_l|$. Factoring out powers of $n+1$ and reindexing, we are left with the following bound on derivatives of ΔS_{2k}.

$$\|\partial_s^n \Delta S_{2k}(st)\| \le (\tau_{\max} N_k \Lambda t)^{n+1} e^{s \tau_{\max} N_k \Lambda t} \tag{A.10}$$

This expression is quite elegant; it is as if we were taking $n+1$ derivatives of the exponential e^{cs} with $c = \tau_{\max} N_k \Lambda t$. The $+1$ as compared to the number of derivatives n can be understood from the fact that ΔS_{2k} is a finite difference of S_{2k} which cancels the leading identity term. Factors of c will occur frequently in what follows, so we find it convenient to adopt this symbol as shorthand.

Stepping back again, we return to bounding the derivatives of powers of $\Delta S_{2k}(st)$ as in equation (A.5).

$$\partial_s^n \Delta S_{2k}(st)^{j+1} \tag{A.11}$$

(The power is taken before the derivative.) We reduce this to the previous case by performing a multinomial expansion.

$$\partial_s^n \Delta S_{2k}(st)^{j+1} = \sum_N \binom{n}{n_0 \ldots n_j} \prod_{l=0}^{j} \partial_s^{n_l} \Delta S_{2k}(st) \tag{A.12}$$

As usual, the capital letter N denotes the set of all nonnegative indices $n_0, \ldots, n_j$ summing to n. Applying the triangle inequality and submultiplicativity, and employing the bound (A.10),

$$\begin{aligned} \left\|\partial_s^n \Delta S_{2k}(st)^{j+1}\right\| &\le \sum_N \binom{n}{n_0 \ldots n_j} \prod_{l=0}^{j} \|\partial_s^{n_l} \Delta S_{2k}(st)\| \\ &\le \sum_N \binom{n}{n_0 \ldots n_j} \prod_{l=0}^{j} c^{n_l+1} e^{cs} \\ &= e^{(j+1)cs} c^{n+j+1} \sum_N \binom{n}{n_0 \ldots n_j}, \end{aligned} \tag{A.13}$$

where we've used the sum property of the n_l where appropriate. The remaining sum over the multinomial coefficient is given by $(j+1)^n$. Hence,

$$\left\|\partial_s^n \Delta S_{2k}(st)^{j+1}\right\| \leq ((j+1)c)^n (ce^{cs})^{j+1} \tag{A.14}$$

Notice that, when $j = 0$, this is consistent with equation (A.10).

With result (A.14) in hand, we return to the power series (A.5). Differentiating term by term

$$\partial_s^n H_{\text{eff}} = -\frac{1}{it} \sum_{j=0}^{\infty} \frac{(-1)^j}{j+1} \partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) \tag{A.15}$$

and performing a binomial expansion for each term

$$\partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) = \sum_{q=0}^{n} \binom{n}{q} \left(\partial_s^q s^j \right) \left(\partial_s^{n-q} \Delta S_{2k}(st)^{j+1} \right) \tag{A.16}$$

will allow us to apply our previous results. It will be helpful to consider two cases separately: $j \leq n$ and $j > n$. These regimes are somewhat qualitatively different, since the derivatives of s^j may or may not vanish depending on the number of derivatives. Focusing on the case $j \leq n$, we have

$$\partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) = \sum_{q=0}^{j} \binom{n}{q} \frac{j!}{(j-q)!} s^{j-q} \left(\partial_s^{n-q} \Delta S_{2k}(st)^{j+1} \right). \tag{A.17}$$

Note that the sum runs only to j, not n. Taking a triangle inequality upper bound using (A.14), we may upper bound (A.17) as

$$\begin{aligned} &\sum_{q=0}^{j} \binom{n}{q} \frac{j!}{(j-q)!} s^{j-q} ((j+1)c)^{n-q} (ce^{cs})^{j+1} \\ &= (ce^{cs})^{j+1} \sum_{q=0}^{j} \binom{j}{q} \frac{n!}{(n-q)!} s^{j-q} ((j+1)c)^{n-q} \end{aligned} \tag{A.18}$$

where we have factored out terms not involving q from the sum, and manipulated the factorials for reasons which will be seen presently. Taking the upper bound $n!/(n-q)! < n^q$, and factoring out $n-j$ powers of $(j+1)c$, we may upper bound the above expression by

$$\begin{aligned} &(ce^{cs})^{j+1} ((j+1)c)^{n-j} \sum_{q=0}^{j} \binom{j}{q} n^q ((j+1)cs)^{j-q} \\ &= (ce^{cs})^{j+1} ((j+1)c)^{n-j} (n + (j+1)cs)^j \end{aligned} \tag{A.19}$$

Thus, with some minor polishing, we may express the bound on (A.16) for $j \leq n$ as

$$\left\|\partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right)\right\| \leq e^{(j+1)cs} c^{n+1} (j+1)^n \left(\frac{n}{j+1} + cs \right)^j \tag{A.20}$$

Now let's move on to the $j > n$ case. Here, there are not enough derivatives to kill off the s^j term, so the binomial sum in (A.17) will run from $q = 0$ to n.

$$\partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) = \sum_{q=0}^{n} \binom{n}{q} \frac{j!}{(j-q)!} s^{j-q} \left(\partial_s^{n-q} \Delta S_{2k}(st)^{j+1} \right) \tag{A.21}$$

Similar to before, we use the bound (A.14), to obtain

$$\begin{aligned} \left\| \partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) \right\| &\leq \sum_{q=0}^{n} \binom{n}{q} \frac{j!}{(j-q)!} s^{j-q} ((j+1)c)^{n-q} (ce^{cs})^{j+1} \\ &= (ce^{cs})^{j+1} s^{j-n} \sum_{q=0}^{n} \binom{n}{q} \frac{j!}{(j-q)!} ((j+1)cs)^{n-q}. \end{aligned} \tag{A.22}$$

Taking $j!/(j-q)! < j^q$, a simpler upper bound is given by

$$(ce^{cs})^{j+1} s^{j-n} \sum_{q=0}^{n} \binom{n}{q} j^q ((j+1)cs)^{n-q} = (ce^{cs})^{j+1} s^{j-n} (j + (j+1)cs)^n. \tag{A.23}$$

With some minor rearrangements, this gives the following upper bound for $j > n$.

$$\left\| \partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) \right\| \leq e^{(j+1)cs} c^{n+1} (j+1)^n (cs)^{j-n} \left(\frac{j}{j+1} + cs \right)^n \tag{A.24}$$

With the bounds (A.20) and (A.24), we can return to bounding $\partial_s^n H_{\text{eff}}$. Still separating the two cases $j \leq n$ and $j > n$, we can write

$$\begin{aligned} \|\partial_s^n H_{\text{eff}}\| \, t &\leq \sum_{j=0}^{n} \frac{1}{j+1} \left\| \partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) \right\| + \sum_{j=n+1}^{\infty} \frac{1}{j+1} \left\| \partial_s^n \left(s^j \Delta S_{2k}(st)^{j+1} \right) \right\| \\ &= B_l + B_h \end{aligned} \tag{A.25}$$

where B_l and B_h refer to bounds on the "low" and "high" parts of the series. Employing the bounds from equations (A.20) and (A.24), we have

$$\begin{aligned} B_l &\leq \sum_{j=0}^{n} \frac{1}{j+1} e^{(j+1)cs} c^{n+1} (j+1)^n \left(\frac{n}{j+1} + cs \right)^j \\ &= c^{n+1} \sum_{j=0}^{n} e^{(j+1)cs} (j+1)^{n-1} \left(cs + \frac{n}{j+1} \right)^j \end{aligned} \tag{A.26}$$

and

$$\begin{aligned} B_h &\leq \sum_{j=n+1}^{\infty} \frac{1}{j+1} e^{(j+1)cs} c^{n+1} (j+1)^n (cs)^{j-n} \left(\frac{j}{j+1} + cs \right)^n \\ &\leq c^{n+1} \sum_{j=n+1}^{\infty} e^{(j+1)cs} (j+1)^{n-1} (cs)^{j-n} (1+cs)^n \\ &= c^{n+1} (1+cs)^n \sum_{j=n+1}^{\infty} e^{(j+1)cs} (j+1)^{n-1} (cs)^{j-n}. \end{aligned} \tag{A.27}$$

Let's begin by bounding B_l. We will at this point make the assumption that s is sufficiently small such that $cs < 1$. This will necessarily factor into the cost later. This simplification yields

$$\begin{aligned} B_l &\leq c^{n+1} \sum_{j=0}^{n} e^{j+1}(j+1)^{n-1} \left(1 + \frac{n}{j+1}\right)^j \\ &\leq c^{n+1} \sum_{j=0}^{n} e^{j+1}(j+1)^{n-1} e^n \\ &\leq \frac{1}{e}(e^2 c)^{n+1} \sum_{j=1}^{n+1} j^{n-1} \end{aligned} \tag{A.28}$$

The remaining sum can be bounded by $(n+1)^n$, hence

$$B_l \leq \frac{(e^2(n+1)c)^{n+1}}{e(n+1)} \leq (e^2 c)^{n+1} n^n, \tag{A.29}$$

where the definition that $0^0 = 1$ handles the edge case. Let's turn our attention to B_h. We will start by reindexing so that the series begins at $j = 0$ in (A.27).

$$B_h \leq c^{n+1}(1+cs)^n \sum_{j=0}^{\infty} e^{(j+n+2)cs}(j+n+2)^{n-1}(cs)^{j+1} \tag{A.30}$$

$$= (ce^{cs})^{n+1}(1+cs)^n \sum_{j=0}^{\infty} (cse^{cs})^{j+1}(j+n+2)^{n-1} \tag{A.31}$$

The series converges if and only if

$$cse^{cs} < 1. \tag{A.32}$$

This function is monotonically increasing with cs and further the optimal value is attained when $cs = \text{LambertW}(1) \approx 0.567 > 0.5$. Moreover, we have the bound

$$(j+n+2)^{n-1} = (n+2)^{n-1} \left(1 + \frac{j}{n+2}\right)^{n-1} \leq (n+2)^{n-1} e^j. \tag{A.33}$$

Thus, we have

$$\begin{aligned} B_h &\leq (cse^{cs})^{n+1}(1+cs)^n (n+2)^{n-1} \sum_{j=0}^{\infty} (ecse^{cs})^j \\ &= (cse^{cs})^{n+1}(1+cs)^n (n+2)^{n-1} \frac{1}{1 - ecse^{cs}}. \end{aligned} \tag{A.34}$$

To be concrete, let's take $ecse^{cs} < 1/2$, which is implied by $cs < \pi/20$

$$B_h \leq 2e^{\pi/20}(cs)^{n+1}(3e^{\pi/20}/2)^n (n+2)^{n-1} \leq 4(cs)^{n+1}(9/5)^n (n+2)^{n-1}. \tag{A.35}$$

Since $(n+2)^{n-1} \leq e^2 n^n/2$ (using $0^0 := 1$ for the edge case $n = 0$), we have

$$B_h \leq 2e^2 (cs)^{n+1} (9/5)^n n^n. \tag{A.36}$$

Altogether, using $s \leq 1$

$$\begin{aligned} \|\partial_s^n H_{\text{eff}}\| \, t &\leq n^n (e^2 c)^{n+1} \left(1 + 2(9/5e^2)^n\right) \\ &\leq 2n^n (e^2 c)^{n+1} \end{aligned} \tag{A.37}$$

The final result then follows from substituting for c and noting that the duration of each time step is at most $2k/3^{k-1}$ using the results of [17]. □

B Proof of Theorem 7

In order to use the result of Lemma 1 in the proof of Theorem 7, we need to re-express the derivatives of the eigenvalues in terms of the effective Hamiltonian above. The following lemma provides such a bound on the derivatives under the assumption that the Hamiltonian is gapped.

Lemma 14. *Assume that $H_{\text{eff}} : \mathbb{R} \mapsto \mathbb{C}^{N \times N}$ yields a Hermitian matrix for each time such that $H_{\text{eff}}(t)$ has a minimum eigenvalue gap at of $\gamma_{\text{eff}}(t) := \min_{k \neq j}(|E_k(t) - E_j(t)|)$ where $E_j(t)$ and $E_k(t)$ are eigenvalues of $H_{\text{eff}}(t)$. Further assume that $H_{\text{eff}}(t)$ is in C^p for some positive integer p. We then define*

$$\Lambda_\gamma := \max_p \left(\sup_t \left(\|\partial_t^p H_{\text{eff}}(t)\|^{1/(p+1)}, \left(\frac{\|\partial_t^p H_{\text{eff}}(t)\|}{\gamma_{\text{eff}}(t)} \right)^{1/p} \right) \right),$$

we then have that for all eigenvalues $E_k(t)$

$$|\partial_t^p E_k(t)| \leq 8^{p-1}(p-1)! \Lambda_\gamma^{p+1}.$$

Proof. The essence of this proof is to examine the derivatives of the eigenvalues of an effective Hamiltonian by expressing them in terms of derivatives of the effective Hamiltonian, which we bound above. Specifically, let $H_{\text{eff}}(t) |k(t)\rangle = E_k(t) |k(t)\rangle$ and assume that the eigenvectors' phases are chosen such that $\left\langle k(t) \middle| \dot{k}(t) \right\rangle = 0$. Standard results from matrix analysis [30] show that

$$\begin{aligned} \partial_t E_k(t) &= \langle k(t)| \left(\partial_t H_{\text{eff}}(t)\right) |k(t)\rangle \\ \partial_t |E_k(t)\rangle &= \sum_{j \neq k} |j(t)\rangle \frac{\langle j(t)| \partial_t H_{\text{eff}}(t) |k(t)\rangle}{E_k(t) - E_j(t)} \\ \partial_t (E_k(t) - E_j(t)) &= \frac{\langle k(t)| \partial_t H_{\text{eff}}(t) |k(t)\rangle - \langle j(t)| \partial_t H_{\text{eff}}(t) |j(t)\rangle}{(E_k(t) - E_j(t))^2} \end{aligned} \tag{B.1}$$

Further, note that

$$\|\partial_t |E_k(t)\rangle\|^2 \leq \sum_{j\neq k} \left| \frac{\langle j(t)| \partial_t H_{\text{eff}}(t) |k(t)\rangle}{E_k(t) - E_j(t)} \right|^2 \leq \frac{\|\dot{H}_{\text{eff}}(t)\|^2}{\gamma_{\text{eff}}^2(t)}, \tag{B.2}$$

where $\gamma_{\text{eff}}(t)$ is the minimum eigenvalue gap of the effective Hamiltonian at time t. The next step in our proof involves using the above properties to evaluate the second- and higher-order derivatives of $E_k(t)$.

Next note that from recursively evaluating the derivatives of the eigenvalues are sums of products of the form

$$\partial_t^q E_k(t) = \sum_{\ell=1}^{q} \sum_{x,y} c_{\vec{x},\vec{y},\ell} \langle x_1(t)| \partial_t^{y_\mu} H(t) |x_2(t)\rangle \prod_{\mu=2}^{\ell} \frac{\langle x_{2\mu-1}(t)| \partial_t^{y_\mu} H(t) |x_{2\mu}(t)\rangle}{(E_{x_{2\mu}}(t) - E_{x_{2\mu-1}}(t))}. \tag{B.3}$$

Note that there is a pattern to this series. If there are m $H^{(p)}$ (for $p \geq 0$) present in the expansion there must be $m-1$ powers of the gaps present in the terms. Before going into more detail let us define the "degree" of a term to be the number of differentiable factors (eigenvectors, Hamiltonians or inverse gaps) present in a product. For example $\langle x(t)| \dot{H}_{\text{eff}}(t) |x(t)\rangle$ is degree 3.

We have that if A is a term of degree $\deg(A)$. Then it follows from (B.1) that

$$\deg(\partial_t A) \leq \deg(A) + 4, \tag{B.4}$$

since from the product rule the derivative of the product of the factors is the sum of the distributed sum of the derivatives of the factors and that if $|E_k\rangle$ or $(E_j - E_k)^{-1}$ are differentiated then the degree increases by at most 4. While the derivative of the effective Hamiltonian does not increase the degree, the increase in the degree is still at most 4 in this case thus in the worst case scenario the degree is increased by 4. As the degree $\deg(\partial_t E_k(t)) = 3 \leq 4$, it is straight forward to see that

$$\deg(\partial_t^q E_k(t)) = 4q \tag{B.5}$$

Next, let us assume that B is a sum of $N_{\text{terms}}(B)$ products of the above factors. We then have from (B.1) that

$$N_{\text{terms}}(\partial_t B) \leq 2N_{\text{terms}}(B)\deg(B). \tag{B.6}$$

Next we will show that the number of terms present in $\partial_t^q E_k(t)$ obeys

$$N_{\text{terms}}(\partial_t^q E_k(t)) \leq 8^{q-1}(q-1)!. \tag{B.7}$$

We prove this by induction. The number of terms present when $q = 1$ is 1. This demonstrates the base case. Assume that for some value q' that the induction hypothesis holds. Then from (B.5) and (B.6) that

$$N_{\text{terms}}(\partial_t^{q'+1} E_k(t)) \leq 2(8^{q'-1}(q'-1)!)(4q') = 8^{q'} q'! \tag{B.8}$$

which proves the relation by induction.

We then see that if any factor is differentiated ν times then its norm is bounded above by $\Lambda_\gamma^{\nu+1}$ where

$$\Lambda_\gamma := \max_p \left(\max \left(\|\partial_t^p H_{\text{eff}}(t)\|^{1/(p+1)}, \left(\frac{\|\partial_t^p H_{\text{eff}}(t)\|}{\gamma_{\text{eff}}(t)} \right)^{1/p} \right) \right). \tag{B.9}$$

We therefore have that since the maximum value of the derivative is the number of terms multiplied by the maximum norm of the product of all the factors that

$$|\partial_t^q E_k(t)| \leq 8^{q-1}(q-1)!\Lambda_\gamma^{q+1}. \tag{B.10}$$

□

Proof of Theorem 7. We proceed by proving the theorem in stages first, from the Baker Campbell Hausdorff formula $S_{2k}(t) = e^{-iHt+O(t^{2k+1})}$ and thus $\log(S_{2k}(t))/t \in O(1)$ for all $t > 0$. Thus an eigenvalue $E_{\text{eff}}(t)$ exists for $H_{\text{eff}}(t)$ for all $t > 0$ which demonstrates the first claim.

Next we show from Lemma 14 that provided that $H_{\text{eff}}(t)$ is in C^p for positive integer p then $\partial_t^p E_k(t)$ exists and is bounded above by $8^{p-1}(p-1)!\Lambda^{p+1}$ since

$$\partial_t H_{\text{eff}}(t) = \frac{1}{t}\partial_s H_{\text{eff}}(st) \tag{B.11}$$

The value of Λ_γ can then be bounded above using Lemma 1 (under the assumption that $2k(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|t \leq \pi/20$ for all t) by

$$\begin{aligned}
\Lambda_\gamma &\leq \left(2t^{-n-1}n^n(2ke^2(5/3)^{k-1}m \max_{l\in[1,m]} \|H_l\|t)^{n+1} \right)^{1/(n+1)} \\
&\quad + \max_p \left(\frac{2t^{-p-1}p^p(2ke^2(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|t)^{p+1}}{\gamma_{\text{eff}}} \right)^{1/p} \\
&\leq 4nke^2(5/3)^{k-1}m \max_{l\in[1,m]} \|H_l\|) \left(1 + \max_p \left(\frac{p}{n} \right) \left(\frac{2ke^2(5/3)^{k-1}m\max_{l\in[1,m]}\|H_l\|}{\gamma_{\text{eff}}} \right)^{1/p} \right) \\
&:= 4nke^2(5/3)^{k-1}m \max_{l\in[1,m]} \|H_l\|)(1+\Gamma)
\end{aligned} \tag{B.12}$$

This implies that

$$|\partial_t^q E_k(t)| \leq 8^{q-1}(q-1)!\Lambda^{q+1} \leq \frac{q!q^q}{64} \left(32ke^2(5/3)^{k-1}m \max_{l\in[1,m]} \|H_l\|)(1+\Gamma) \right)^{q+1}. \tag{B.13}$$

For $t \in [-a, a]$ we have from Lemma 2 that for an $n-1$ point interpolation formula

$$|E_k(0) - P(0)| \leq \frac{n!}{64} \left(32ke^2(5/3)^{k-1}m \max_{l\in[1,m]} \|H_l\|)(1+\Gamma) \right)^{n+1} \left(\frac{a}{2} \right)^n \tag{B.14}$$

Given that we wish to have the interpolation error at most ϵ_{int} we can solve for the value of a such that $t \in [-a, a]$ for the interpolation that is needed in order to ensure that the error bound. Specifically,

$$a = \frac{2(64\epsilon_{\text{int}})^{1/n}}{(n!)^{1/n}} \left(\frac{1}{32ke^2(5/3)^{k-1} m \max_{l\in[1,m]} \|H_l\|)(1+\Gamma)} \right)^{1+1/n} \tag{B.15}$$

The number of exponentials needed in the degree n polynomial interpolation is, using the fact that the error at $-t$ is equal to the error at t is from the fact that the j^{th} chebyshev node location scales like a/j and the fact that the cost of phase estimation within error ϵ_{PE} with probability of failure at most δ_{PE} scales as $O(\log(1/\delta_{\text{PE}})/\epsilon_{\text{PE}}$

$$\begin{aligned} O\left(\sum_{j=1}^{n/2} \frac{m5^{k-1}}{ja} \right) &= O\left(\frac{n\log(n) m5^{k-1} \log(1/\delta_E)}{a\epsilon_{PE}} \right) \\ &= O\left(\frac{n^2 \log(n) \log(1/\delta_{PE})(n!)^{1/n} m^{2+1/n} (5/3)^{2k+k/n} (\max \|H_i\|(1+\Gamma)k)^{1+1/n}}{\epsilon_{\text{int}}^{1/n} \epsilon_{\text{PE}}} \right) \\ &= \widetilde{O}\left(\frac{n^2 m^{2+1/n} \log(1/\delta_{PE}) 5^k (5/3)^{k+k/n} (\max \|H_i\|(1+\Gamma)k)^{1+1/n}}{\epsilon_{\text{int}}^{1/n} \epsilon_{\text{PE}}} \right) \end{aligned} \tag{B.16}$$

There are two competing tendencies in the cost. The number of operator exponentials increases polynomially with n, but the scaling with the error tolerance improves exponentially with n. Setting these two equal to each other to estimate the optimal scaling yields

$$n^2 = \left(\frac{mk(5/3)^k \max_i \|H_i\|(1+\Gamma)}{\epsilon_{\text{int}}} \right)^{1/n} \tag{B.17}$$

Under the assumption that $\epsilon_{\text{int}} \leq 5/3$ we have that the solution is of the form

$$n = \frac{\ln\left(\frac{mk(5/3)^k \max_i \|H_i\|(1+\Gamma)}{\epsilon_{\text{int}}} \right)}{2\text{LambertW}\left(\frac{\ln\left(\frac{mk(5/3)^k \max_i \|H_i\|(1+\Gamma)}{\epsilon_{\text{int}}} \right)}{2} \right)} \in O\left(\log\left(\frac{mk(5/3)^k \max_i \|H_i\|(1+\Gamma)}{\epsilon_{\text{int}}} \right) \right) \tag{B.18}$$

Finally, we have from Lemma 4 that the error in the extrapolated energy as a function of the number of points used for the measurement for an n-point formula we would like to be at most ϵ which is implied by

$$\sqrt{n} \max_i |E_k(s_i) - \widetilde{E}_k(s_i)| \leq \sqrt{n}\epsilon_{\text{PE}} \leq \epsilon \tag{B.19}$$

This implies that the total number of exponentials needed to learn each point within error ϵ_{PE} within probability of failure $\delta_{\text{PE}} = 1/3n$ which by the union bound guarantees

that the total probability of failure is at most 1/3

$$N_{\text{exp}} \in \widetilde{O}\left(\frac{m^2(25/3)^k \max \|H_i\|(1+\Gamma)\log^{7/2}(\max_i \|H_i\|/\epsilon)}{\epsilon}\right). \tag{B.20}$$

□

C Cost asymptotics for expectation values

Here we provide proofs for Lemma 10 and Theorem 11 as stated in Section 4 of the main paper.

Proof of Lemma 10. For scalar functions $f(s)$, derivatives of $\exp(f(s))$ can be expressed through the complete Bell polynomials via the Faa di Bruno formula.

$$\partial_s^n e^{f(s)} = Y_n(f'(s), f''(s), \ldots, f^{(n)}(s))e^{f(s)} \tag{C.1}$$

For operator exponentials such as $\exp(-iH_{\text{eff}}t)$, derivatives can be expressed via repeated application of Duhamel's formula. Yet these expressions are always upper bounded by the commuting (scalar) case, so that

$$\left\|\partial_s^n e^{-iH_{\text{eff}}t}\right\| \leq Y_n\left(t\left\|\partial_s H_{\text{eff}}\right\|, t\left\|\partial_s^2 H_{\text{eff}}\right\|, \ldots, t\left\|\partial_s^n H_{\text{eff}}\right\|\right). \tag{C.2}$$

Note that the exponential disappeared in the bound since it has norm one. Applying Lemma 1 and invoking the fact that Y_n is monotonic in each argument, this is upper bounded by

$$Y_n\left((2j^j c^{j+1})_{j=1}^n\right) \tag{C.3}$$

where $c := 2ke^2(5/2)^{k-1}m \max_{l\in[1,m]} \|H_l\| t$. An explicit formula for this is given by

$$Y_n\left((2j^j c^{j+1})_{j=1}^n\right) = \sum_D \frac{n!}{d_1! \ldots d_n!} \prod_{j=1}^n \left(\frac{2j^j c^{j+1}}{j!}\right)^{d_j} \tag{C.4}$$

where D is a sum over indices $\{d_j\}_{j=1}^n$ such that

$$\sum_{j=1}^n d_j j = n. \tag{C.5}$$

Using a Stirling-type bound

$$\frac{1}{j!} \leq \left(\frac{e}{j}\right)^j \frac{1}{\sqrt{2\pi}} \tag{C.6}$$

allows us to write

$$\begin{aligned} Y_n\left((2j^j c^{j+1})_{j=1}^n\right) &\leq \sum_D \frac{n!}{d_1!\dots d_n!}\prod_{j=1}^n\left(\sqrt{\frac{2}{\pi}}e^j c^{j+1}\right)^{d_j} \\ &= (ec)^n \sum_D \frac{n!}{d_1!\dots d_n!}\prod_{j=1}^n\left(\sqrt{\frac{2}{\pi}}c\right)^{d_j} \\ &= (ec)^n Y_n(\sqrt{2/\pi}c, \sqrt{2/\pi}c, \dots, \sqrt{2/\pi}c) \\ &= (ec)^n B_n(\sqrt{2/\pi}c). \end{aligned} \tag{C.7}$$

In the second line we brought out n factors of ec using the condition on the indices D, and we identified Y_n evaluated the same at every argument to be the single-variable Bell (or Touchard) polynomial. We can bound the size of $B_n(\sqrt{2/\pi}c)$ by

$$B_n(\sqrt{2/\pi}c) \leq \left(\frac{n}{\log(1+\sqrt{\frac{\pi}{2}}n/c)}\right)^n \tag{C.8}$$

so that

$$\left\|\partial_s^n e^{-iH_{\text{eff}}t}\right\| \leq \left(\frac{ecn}{\log(1+\sqrt{\frac{\pi}{2}}n/c)}\right)^n \leq \left(\frac{ecn}{2}\right)^n \left(1+\sqrt{\frac{8}{\pi}\frac{c}{n}}\right)^n \tag{C.9}$$

where we've used the bound $1/\log(1+x) \leq 1/2 + 1/x$.

With this bound on the Suzuki-Trotter formula derivatives, we now turn to bounding $\partial_s^n O_s(t)$. Applying the binomial theorem and triangle inequality to (4.2),

$$\begin{aligned} \|\partial_s^n O_s(t)\| &\leq \|O\| \sum_{p=0}^n \binom{n}{p} \left\|\partial_s^p e^{itH_{\text{eff}}}\right\| \left\|\partial_s^{n-p} e^{-itH_{\text{eff}}}\right\| \\ &\leq \|O\| \left(\frac{ec}{2}\right)^n \sum_{p=0}^n \binom{n}{p} p^p (n-p)^{n-p} \left(1+\sqrt{\frac{8}{\pi}\frac{c}{p}}\right)^p \left(1+\sqrt{\frac{8}{\pi}\frac{c}{n-p}}\right)^{n-p}. \end{aligned} \tag{C.10}$$

At this point, it will be fruitful to consider two regimes. Recall that c encodes information about the time length of the simulation.

In the case where $c > n$, we have

$$\begin{aligned} \frac{\|\partial_s^n O_s\|}{\|O\|} &\leq \left(\frac{ec}{2}\right)^n \sum_{p=0}^n \binom{n}{p} c^n \left(1+\sqrt{8/\pi}\right)^n \\ &= c^{2n}\left(\frac{e}{2}(1+\sqrt{8/\pi})\right)^n 2^n \\ &= \left(\sqrt{e(1+\sqrt{8/\pi})c}\right)^{2n} < (ec)^{2n}. \end{aligned} \tag{C.11}$$

This implies a relative error in the polynomial fit bounded by

$$\frac{|E_{n-1}(0)|}{\|O\|} \leq \left(\frac{e^2c^2a}{2n}\right)^n \tag{C.12}$$

In the regime where $c \leq n$ (otherwise known as the short-time regime), the approximation

$$\left(1+\sqrt{\frac{8}{\pi}}\frac{c}{p}\right)^p < e^c\sqrt{8/\pi} \tag{C.13}$$

is not so crude. Thus,

$$\frac{\|\partial_s^n O_s\|}{\|O\|} \leq \left(\frac{ec}{2}\right)^n n! \sum_{p=0}^{n} \frac{p^p}{p!}\frac{(n-p)^{n-p}}{(n-p)!} e^{4c/\sqrt{\pi}}. \tag{C.14}$$

Regrouping and employing a Stirling bound where appropriate,

$$\begin{aligned}\frac{\|\partial_s^n O_s\|}{\|O\|} &\leq e^{4c/\sqrt{\pi}}\left(\frac{ec}{2}\right)^n n!\left(2+\frac{e^n}{2\pi}\sum_{p=1}^{n-1}\frac{1}{\sqrt{p(n-p)}}\right)\\ &\leq e^{4c/\sqrt{\pi}}\left(\frac{e^2c}{2}\right)^n n!\left(\frac{2}{e^n}+\frac{\sqrt{n-1}}{2\pi}\right)\\ &\leq \frac{3}{2\pi}\sqrt{n-1}\left(\frac{e^2c}{2}\right)^n n! e^{4c/\sqrt{\pi}}.\end{aligned} \tag{C.15}$$

This gives a corresponding relative interpolation error of

$$\frac{E_{n-1}(0)}{\|O\|} \leq \frac{4}{3}n\left(\frac{eca}{4}\right)^n e^{4c/\sqrt{\pi}}. \tag{C.16}$$

□

Proof of Theorem 11. Let $f(s) = \langle O_s(t)\rangle / \|O\|$ consist of the normalized expectation values. Our interpolation algorithm produces an estimate $\bar{f}$ of $f(0)$ which we require to be accurate within ϵ.

$$\left|f(0) - \bar{f}\right| \leq \epsilon \tag{C.17}$$

Let's consider the various types of errors accrued by the algorithm.

1. A polynomial fit $P_n f(0)$ over Chebyshev nodes has error characterized by 2.
2. Our nodes are not exactly the Chebyshev nodes. They are approximate Chebyshev nodes so that the reciprocal is an integer
3. Our Trotter expectation values are only approximate, being estimated by Amplitude estimation

There is the interpolation error from the polynomial $P_{n-1}f$ fitting f assuming perfect interpolation points $(s_i, f(s_i))$. But $f(s_i)$ can only be estimated; let's call f_i this estimate. The error in f_i comes from two sources (in our analysis): the statistical error inherent in amplitude estimation and the systematic error of using integer Trotter steps (i.e. r_i as the integer part of $1/s_i$). We can unpack all of these sources of error via the triangle inequality.

$$\begin{aligned}\left|f(0) - \bar{f}\right| &\leq |f(0) - P_{n-1}f(0)| + \left|P_{n-1}f(0) - \tilde{P}_{n-1}f(0)\right| \\ &\leq \epsilon_{\text{int}} + L_n\epsilon_{\text{data}}\end{aligned} \tag{C.18}$$

Here L_n is the norm of the linear projection which is the nth-degree polynomial fit, essentially a condition number, and ϵ_{data} is an upper bound on the error in the data. $\tilde{P}_{n-1}f$ is the fit to the imperfect data and $P_{n-1}f$ the fit to the perfect data $(s_i, f(s_i))$. For generic interpolation nodes, L_n can grow rapidly, however for the set of Chebyshev nodes we obtain an optimal value.

$$L_n \leq \frac{2}{\pi}\log(n+1) + 1 \tag{C.19}$$

The error $|f(s_i) - f_i|$ of the ith data point can be further divided into two pieces: one from the use of integer Trotter steps (without fractional queries) and one due to the usual finite statistics.

$$|f(s_i) - f_i| \leq \left|f(s_i) - \frac{\text{Tr}\left(S_{2k}(s_it)^{r_i}OS_{2k}(s_it)^{r_i}\right)}{\|O\|}\right| + \left|\frac{\text{Tr}\left(S_{2k}(s_it)^{r_i}OS_{2k}(s_it)^{r_i}\right)}{\|O\|} - f_i\right| \tag{C.20}$$

For the integer query part we have

$$\left|f(s_i) - \frac{\text{Tr}\left(S_{2k}(s_it)^{r_i}OS_{2k}(s_it)^{r_i}\right)}{\|O\|}\right| \leq \frac{\left\|[O, S_{2k}(s_it)^{\delta_i}]\right\|}{\|O\|} \tag{C.21}$$

where $\delta_i \in (0,1)$ is the fractional part of $1/s_i$.

$$\left|\tilde{f}_i - f(s_i)\right| \leq \epsilon_{\text{data}} \tag{C.22}$$

and suppose the interpolation error is within ϵ_{int}. We can bound the error by splitting the error source (data vs interpolation) through the triangle inequality.

$$\begin{aligned}\left|\tilde{P}_{n-1}f(0) - f(0)\right| &\leq \left|\tilde{P}_{n-1}f(0) - P_{n-1}f(0)\right| + |P_{n-1}f(0) - f(0)| \\ &\leq \left\|\tilde{P}_{n-1}f - P_{n-1}f\right\|_\infty + \epsilon_{\text{int}} \\ &\leq L_n\epsilon_{\text{data}} + \epsilon_{\text{int}}\end{aligned} \tag{C.23}$$

Here L_n is the norm of the linear operator corresponding to the polynomial fit, essentially a condition number. This number can grow rapidly in n for generic interpolation nodes, however for the set of Chebyshev nodes we obtain an optimal value.

$$L_n \leq \frac{2}{\pi} \log(n+1) + 1 \tag{C.24}$$

Since we want the total error to be within a threshold ϵ, we can require

$$\epsilon_{\text{data}} \leq \frac{\epsilon}{2L_n}, \quad \epsilon_{\text{int}} \leq \frac{\epsilon}{2}. \tag{C.25}$$

Given these error bounds, we can now turn to the cost of acquiring the data points. Let's assume the expectation value can be encoded as an amplitude estimation problem. This occurs, for example, when O can be block encoded in a unitary $\mathcal{U}$, with scaling parameter γ such that $\|O\|/\gamma \leq 1$. Performing a Hadamard test gives the amplitude

$$a = \frac{1 + \langle O_{s_i}(t) \rangle / \gamma}{2}. \tag{C.26}$$

If we estimate this amplitude to within accuracy $\epsilon_{\text{data}} \|O\| / 2\gamma$, we can estimate $f(s_i)$ within ϵ_{data}. Using Iterative Quantum Amplitude Estimation, we can obtain this estimate using a Grover iterate G constructed from a two Hadamard test oracles. The number of Grover oracles N_G required is given by

$$N_G \leq \frac{200\gamma L_n}{\|O\| \epsilon} \log\left(\frac{2}{\alpha} \log_2\left(\frac{\gamma L_n \pi}{\|O\| \epsilon}\right)\right) \tag{C.27}$$

where α is the failure probability. Each G requires two Hadamard tests, and each Hadamard oracle calls a Suzuki-Trotter evolution once. We take as our metric for the cost of a Suzuki-Trotter simulation $N_k / |s_i|$, where $1/|s_i|$ generalizes the notion of "number of Trotter steps". Altogether, the "number of exponentials" for a single data point is given by

$$N_G \times 2 \times \frac{N_k}{|s_i|}. \tag{C.28}$$

For our analysis, we will assume the most costly part of this entire procedure is the time evolution. Then it is reasonable to define the cost of the interpolation algorithm solely in terms of the number of Hamiltonian exponentials $e^{-iH_j t}$ required to collect the data. Therefore, the cost C_{data} of generating all $n/2$ data points (we only need half due to symmetry) is defined to be

$$C_{\text{data}} := 2N_G N_k \sum_{i=1}^{n/2} \frac{1}{s_i}. \tag{C.29}$$

For our application, we have chosen the Chebyshev nodes $s_i = a\sin(\pi(2i-1)/2n)$ due to their well-conditioning via L_n. For this case, we have the following upper bound.

$$N_k \sum_{i=1}^{n/2} \frac{1}{s_i} = \frac{N_k}{a} \sum_{i=1}^{n/2} \frac{1}{\sin(\pi \frac{2i-1}{2n})} \leq \frac{2N_k}{a} \sum_{i=1}^{n/2} \frac{2n}{(2i-1)\pi} = \frac{2nN_k}{\pi a} \sum_{i=1}^{n/2} \frac{1}{i-1/2} \tag{C.30}$$

Here, we've used the fact that $\sin(x) > x/2$ for $x \in [0, \pi/2]$. The remaining sum can be upper bounded as follows.

$$\sum_{i=1}^{n/2} \frac{1}{i-1/2} \leq 2 + \int_1^{n/2} \frac{1}{x-1/2} dx = 2 + \log(n-1) \tag{C.31}$$

Using this bound in equation (C.29), along with the IQAE bound (C.27),

$$\begin{aligned} C_{\text{data}} &\leq \frac{800\gamma N_k L_n n}{\pi \|O\| \epsilon a} (\log(n-1)+2) \log\left(\frac{2}{\alpha} \log_2\left(\frac{\gamma L_n \pi}{\|O\| \epsilon}\right)\right) \\ &\in \tilde{O}\left(\frac{n \log^2 n}{a\epsilon}\right) \end{aligned} \tag{C.32}$$

where $\tilde{O}$ suppresses factors on the order of log log or slower, and we treat other parameters as fixed for asymptotic purposes. The number of nodes n and the half-width of the interpolation interval a will be determined by ϵ_{int}, the interpolation error assuming perfect data. To apply our error bounds from the previous subsection, we work in a regime where a satisfies the conditions of Lemma 1. We require

$$\left(\frac{e^2 c^2 a}{2n}\right)^n < \frac{\epsilon}{2}. \tag{C.33}$$

since we need to make the error arbitrarily small, we require $a/n \in \Omega(c^2) = \Omega((\Lambda t)^2)$. In fact, we will be able to reach this. Since we require $ac \leq \pi/20$, let's set $a = \pi/(20c)$. Rearranging (C.33), and substituting in our expression for a,

$$\begin{aligned} \left(\frac{\pi n}{10e^2 c}\right)^n &> \frac{2}{\epsilon} \\ n &> \frac{\log 2/\epsilon}{\text{LambertW}(\pi \log(2/\epsilon)/(10e^2 c))} \end{aligned} \tag{C.34}$$

This gives

$$n \in \Omega(c + \log 1/\epsilon) \tag{C.35}$$

Altogether,

$$C_{\text{data}} \in \tilde{O}\left(\Lambda t(\Lambda t + \log 1/\epsilon)/\epsilon\right) \tag{C.36}$$

For the short time case, we require

$$\frac{4}{3}n\left(\frac{eca}{4}\right)^n e^{4c/\sqrt{\pi}} < \epsilon/2 \tag{C.37}$$

We can satisfy this, using $c < n$, by

$$\frac{4}{3}n\left(\frac{e^{1+4/\sqrt{\pi}}na}{4}\right)^n < \epsilon/2. \tag{C.38}$$

Setting a such that $na = \pi/42$ we satisfy our required condition on a and also have

$$\frac{4}{3}n\frac{1}{2^n} < \epsilon/2 \tag{C.39}$$

So that

$$n > -\frac{\text{LambertW}_{-1}(-3\epsilon/8\log 2)}{\log 2} \in O(\log(1/\epsilon)) \tag{C.40}$$

Thus

$$C_{\text{data}} \in \tilde{O}\left(\frac{\log 1/\epsilon}{\epsilon}\right) \tag{C.41}$$

Note here that we are not considering asymptotic time scaling Λt, since this is a short time regime. □

D Proof of Sampling Error

We would like to estimate the error between the DFT and corresponding samples of the Fourier transform. We split this analysis in three parts: calculating time-domain truncation errors, estimating frequency-domain truncation errors (non-zero sampling rate), and finally calculate the error coming from renormalization.

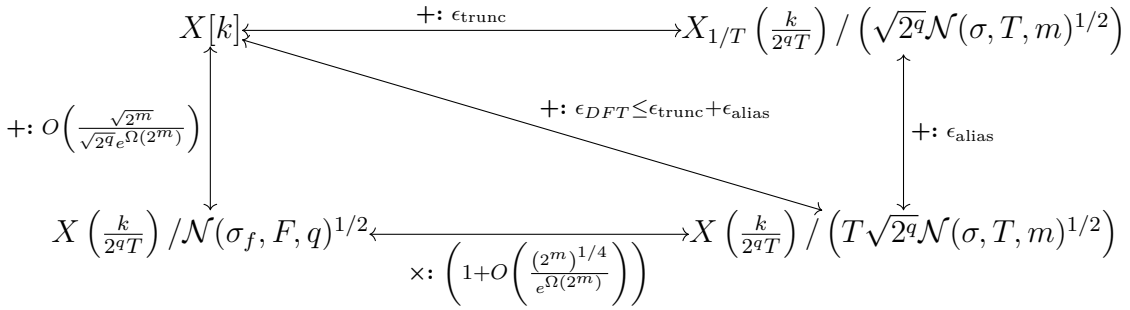

Figure 6: Schematic representation of the error estimation between the DFT of $x[n]$, $X[k]$, and the re-normalized samples of the Fourier transform $X\left(\frac{k}{2^qT}\right)$. At the edges of the diagram we have noted the error between the expressions at the nodes, where a $+$ denotes additive and $\times$ multiplicative.

Here are some useful quantities which we will use for estimating the error on the final register result of Figure 2 with respect to the desired result.

$$p_G(t,\sigma,\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \tag{D.1}$$

$$x(t) = \sqrt{p_G(t,\sigma,\mu)} \tag{D.2}$$

$$X_{1/T}\left(\frac{k}{2^q T}\right) = \sum_{n=-\infty}^{\infty} x(nT)\cdot e^{-i2\pi\frac{k}{2^q}n} \qquad k = -2^q/2,\ldots,2^q/2-1 \tag{D.3}$$

$$X = \mathcal{F}\{x\}(f) = \int x(t)e^{-2\pi i t f}dt \tag{D.4}$$

$$x[n] = \begin{cases} \frac{1}{\mathcal{N}(\sigma,T,m)^{1/2}} x(nT) & n = -2^m/2-1,\ldots,2^m/2-1 \\ 0 & n = -2^m/2 \end{cases} \tag{D.5}$$

$$\mathcal{N}(\sigma,T,m) = \sum_{n=-(2^m/2-1)}^{2^m/2-1} |x(nT)|^2 \tag{D.6}$$

$$X[k] = \frac{1}{\sqrt{2^q}} \sum_{n=-(2^m/2-1)}^{2^m/2-1} x[n]\cdot e^{-i2\pi\frac{k}{2^q}n} \tag{D.7}$$

D.1 Truncation Error

The error between the DTFT

$$\frac{1}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}} X_{1/T}\left(\frac{k}{2^q T}\right) \tag{D.8}$$

and the DFT can be estimated to be:

$$\begin{aligned} \epsilon_{\text{trunc}} &= \frac{1}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}} \left(\sum_{x=2^m/2}^{\infty} \sqrt{p_G(xT;\sigma,0)} \right. \\ &\qquad \left. + \sum_{x=-\infty}^{-2^m/2} \sqrt{p_G(xT;\sigma,0)} \right) \\ &= \frac{\tilde{\epsilon}_{\text{trunc}}}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}} \end{aligned} \tag{D.9}$$

We can estimate an upper bound on the left term by using a right-Riemann sum approximation of the error function:

$$\sum_{n=2^m/2}^{\infty} \sqrt{p_G(nT;\sigma,\mu)} \leq \sqrt[4]{\frac{\pi}{2}} \frac{\sqrt{\sigma}}{T} \operatorname{erfc}\left(\frac{T(2^m/2-1)}{2\sigma}\right)$$

We can do something similar to the right term, but using a left-Riemann sum

$$\sum_{n=-\infty}^{-2^m/2} \sqrt{p_G(nT;\sigma,0)} \leq \sqrt[4]{\frac{\pi}{2}} \frac{\sqrt{\sigma}}{T} \operatorname{erfc}\left(\frac{T(2^m/2-1)}{2\sigma}\right)$$

Therefore,

$$\begin{aligned}\tilde{\epsilon}_{\text{trunc}} &\leq 2^{3/4}\pi^{1/4}\frac{\sqrt{\sigma}}{T}\operatorname{erfc}\left(\frac{T(2^{m-1}-1)}{2\sigma}\right)\\ &\leq 2^{3/4}\pi^{1/4}\frac{\sqrt{\sigma}}{T}e^{-\left(\frac{T(2^{m-1}-1)}{2\sigma}\right)^2}.\end{aligned}$$

Finally, we can express the truncation error using (D.8) with

$$\epsilon_{\text{trunc}} \leq \frac{2^{3/4}\pi^{1/4}\frac{\sqrt{\sigma}}{T}e^{-\left(\frac{T(2^{m-1}-1)}{2\sigma}\right)^2}}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}} = O\left(\sqrt{\frac{\sigma}{T}}\sqrt{\frac{1}{2^q}}e^{-\Omega(2^m/(\sigma/T))}\right). \tag{D.10}$$

D.2 Aliasing error

Now that we have estimated the error on the DTFT by time-domain truncation, we would also like to estimate the aliasing error, that is, the difference between the DTFT and the Fourier transform over a period $1/T$. This is coming from having a finite sampling rate of a function that is not bounded in the frequency domain. In order to estimate the aliasing error we must look at the Fourier transform of a Gaussian distribution:

$$\begin{aligned}\int \sqrt{p_G(t;\sigma,0)}e^{-2\pi itf}dx &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{1/2} 2\sigma\sqrt{\pi}\cdot e^{-(2\sigma\pi f)^2}\\ &= \pi^{1/4}2^{3/4}\sqrt{\sigma}\cdot e^{-(2\sigma\pi f)^2}\\ &= \left(4\sigma\pi/\sqrt{2\pi}\right)^{1/2}\cdot e^{-\frac{f^2}{4/(4\sigma\pi)^2}}\\ &= \sqrt{p_G(f;\sigma_f,0)}\\ \sigma_f &= \frac{1}{4\sigma\pi}.\end{aligned} \tag{D.11}$$

Now, the DTFT can be expressed in terms of a Fourier transform in the following way

$$X_{1/T}(f) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X(f-k/T). \tag{D.12}$$

Therefore, the aliasing error is

$$\tilde{\epsilon}_{\text{alias}} = \frac{1}{T}\sum_{k=-\infty}^{-1} X(f-k/T) + \frac{1}{T}\sum_{k=1}^{\infty} X(f-k/T). \tag{D.13}$$

The aliasing error has a critical point at $f = 0$ and its second derivative is strictly positive throughout $f = [-f_s/2, f_s/2]$ and thus we know that the error is largest at the boundaries of the DTFT. That is, at $f = -f_s/2, f_s/2$. The error at these two points is expected to be the same so we just choose $f = f_s/2$ in order to bound the error from above. That means

$$\begin{aligned}
\tilde{\epsilon}_{\text{alias}} &\leq \frac{1}{T}\sum_{k=-\infty}^{-1} X(f_s/2 - kf_s) + \frac{1}{T}\sum_{k=1}^{\infty} X(f_s/2 - kf_s) \\
&\leq \frac{1}{T}\sum_{k=-\infty}^{\infty} X(f_s/2 - kf_s) \\
&= \pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot\left(\theta_2(e^{-(2\sigma\pi f_s)^2})\right) \\
&= \pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot\left(2e^{-\pi^2(\sigma/T)^2}\sum_{0}^{\infty}(e^{-4\pi^2(\sigma/T)^2})^{n(n+1)}\right)
\end{aligned} \tag{D.14}$$

For $e^{-4\pi^2(\sigma/T)^2} \leq 1$

$$\begin{aligned}
\tilde{\epsilon}_{\text{alias}} &\leq \pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot\left(2e^{-\pi^2(\sigma/T)^2}\sum_{0}^{\infty}(e^{-4\pi^2(\sigma/T)^2})^{n}\right) \\
&= \pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot\left(2e^{-\pi^2(\sigma/T)^2}\frac{1}{1-e^{-4\pi^2(\sigma/T)^2}}\right)
\end{aligned} \tag{D.15}$$

For $e^{-4\pi^2(\sigma/T)^2} \leq 1/2$

$$\begin{aligned}
\tilde{\epsilon}_{\text{alias}} &\leq \pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot\left(2e^{-\pi^2(\sigma/T)^2}\frac{1}{1-e^{-4\pi^2(\sigma/T)^2}}\right) \\
&\leq 4\pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot e^{-\pi^2(\sigma/T)^2}
\end{aligned} \tag{D.16}$$

The sampling rate and width on this new and imperfect Gaussian are

$$\begin{aligned}
F &= \frac{1}{T2^q} \\
\sigma_f &= \frac{1}{4\sigma\pi}.
\end{aligned} \tag{D.17}$$

Therefore, the error between

$$\frac{1}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}}X_{1/T}\left(\frac{k}{2^qT}\right) \tag{D.18}$$

and

$$\frac{1}{T\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}}X\left(\frac{k}{2^qT}\right)$$

$$= \frac{F^{1/2}}{T^{1/2}\mathcal{N}(\sigma,T,m)^{1/2}} X\left(\frac{k}{2^q T}\right) \tag{D.19}$$

can be estimated using (D.16) to be

$$\epsilon_{\text{alias}} = \frac{\tilde{\epsilon}_{\text{alias}}}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}} \leq \left(\frac{4\pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot e^{-\pi^2(\sigma/T)^2}}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}}\right). \tag{D.20}$$

Thus, with a choice of

$$\frac{\sigma}{T} \sim \sqrt{2^m}, \tag{D.21}$$

both sources of error can be bounded with

$$\begin{aligned}\epsilon_{\text{trunc}} &= O\left(\frac{(2^m)^{1/4}}{(2^q)^{1/2}e^{\Omega(2^m)}}\right)\\ \epsilon_{\text{alias}} &= O\left(\frac{(2^m)^{1/4}}{(2^q)^{1/2}e^{\Omega(2^m)}}\right).\end{aligned} \tag{D.22}$$

This means that after the DFT, we know that we will have in the register the amplitude

$$\frac{F^{1/2}}{T^{1/2}\mathcal{N}(\sigma,T,m)^{1/2}} X\left(\frac{k}{2^q T}\right) + \epsilon_{DFT}, \tag{D.23}$$

where ϵ_{DFT} is the total error from both truncation as well as aliasing. We then bound the error by summing (D.16) and (D.8) to find

$$\begin{aligned}\epsilon_{DFT} &\leq \frac{2^{3/4}\pi^{1/4}\frac{\sqrt{\sigma}}{T}e^{-\left(\frac{T(2^{m-1}-1)}{2\sigma}\right)^2}}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}} + \left(\frac{4\pi^{1/4}2^{3/4}\sqrt{\sigma}\frac{1}{T}\cdot e^{-\pi^2(\sigma/T)^2}}{\sqrt{2^q}\mathcal{N}(\sigma,T,m)^{1/2}}\right)\\ &\leq 2\max\left(\epsilon_{\text{alias}},\epsilon_{\text{trunc}}\right) = O\left(\frac{(2^m)^{1/4}}{(2^q)^{1/2}e^{\Omega(2^m)}}\right).\end{aligned} \tag{D.24}$$

D.3 Renormalization error and ultimate additive error

We know that the signal after DFT has to be normalized, we also know that it is locally ϵ_{DFT} close to Equation (D.19), which is not necessarily normalized. Therefore, for the norm of samples of Equation (D.19) we obtain the following upper bound on normalization ratios

$$\begin{aligned}\left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\sum_k\left|X\left(\frac{k}{2^q T}\right)\right|^2\right) &= \left(\frac{F\mathcal{N}(\sigma_f,F,q)}{T\mathcal{N}(\sigma,T,m)}\right)\\ &= 1 - 2\left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2}\Re\left(\vec{\epsilon}_{DFT}\cdot\vec{X}\right) + \|\vec{\epsilon}_{DFT}\|^2\end{aligned}$$

$$
\begin{aligned}
&\leq 1 + \|\vec{\epsilon}_{DFT}\|^2 \\
&\leq 1 + \|\vec{\epsilon}_{DFT}\| \\
&\leq 1 + \epsilon_{DFT}\sqrt{2^q}, \qquad \text{(D.25)}
\end{aligned}
$$

where we have used the assumption $\|\vec{\epsilon}_{DFT}\| \leq 1$. Similarly, the lower bound is:

$$
\begin{aligned}
\left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\sum_k \left|X\left(\frac{k}{2^q T}\right)\right|^2\right) &= \left(\frac{F\mathcal{N}(\sigma_f,F,q)}{T\mathcal{N}(\sigma,T,m)}\right) \\
&\geq 1 - 2\left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} \Re\left(\|\vec{\epsilon}_{DFT}\|\|\vec{X}\|\right) \\
&\geq 1 - 2\epsilon_{DFT}\sqrt{2^q}\left(\frac{F\mathcal{N}(\sigma_f,F,q)}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} \\
&\geq 1 - 2\epsilon_{DFT}\sqrt{2^q}\left(1 + \|\vec{\epsilon}_{DFT}\|^2\right) \\
&\geq 1 - 4\epsilon_{DFT}\sqrt{2^q} \qquad \text{(D.26)}
\end{aligned}
$$

Thus, if we take the full inequality

$$
1 - 4\epsilon_{DFT}\sqrt{2^q} \leq \left(\frac{F\mathcal{N}(\sigma_f,F,q)}{T\mathcal{N}(\sigma,T,m)}\right) \leq 1 + \epsilon_{DFT}\sqrt{2^q} \qquad \text{(D.27)}
$$

and take the square root, we obtain,

$$
1 - 2\epsilon_{DFT}\sqrt{2^q} \leq \left(\frac{F\mathcal{N}(\sigma_f,F,q)}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} \leq 1 + \epsilon_{DFT}\sqrt{2^q}, \qquad \text{(D.28)}
$$

where we have used $1 + \sqrt{x} \leq 1 + x$ and $1 - \sqrt{x} \geq 1 - x/2$ for $x \leq 1$. We now divide every side by $\mathcal{N}(\sigma_f, F, q)$ to obtain

$$
\begin{aligned}
&1/\mathcal{N}(\sigma_f,F,q)^{1/2} - 2\epsilon_{DFT}\sqrt{2^q}/\mathcal{N}(\sigma_f,F,q) \leq \left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} \leq \\
&1/\mathcal{N}(\sigma_f,F,q)^{1/2} + \epsilon_{DFT}\sqrt{2^q}/\mathcal{N}(\sigma_f,F,q)^{1/2}. \qquad \text{(D.29)}
\end{aligned}
$$

Subtract $1/\mathcal{N}(\sigma_f, F, q)$, which gives

$$
\begin{aligned}
&- 2\epsilon_{DFT}\sqrt{2^q}/\mathcal{N}(\sigma_f,F,q)^{1/2} \leq \left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} - 1/\mathcal{N}(\sigma_f,F,q)^{1/2} \\
&\leq \epsilon_{DFT}\sqrt{2^q}/\mathcal{N}(\sigma_f,F,q)^{1/2}. \qquad \text{(D.30)}
\end{aligned}
$$

We multiply by $\max|X(k)|$

$$\begin{aligned}
&- 2\epsilon_{DFT}\sqrt{2^q}/\mathcal{N}(\sigma_f,F,q)^{1/2}/(\sigma_f^{1/2}2^{1/4}\pi^{1/4})\\
&\leq \left(\left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} - 1/\mathcal{N}(\sigma_f,F,q)^{1/2}\right)\max|X(k)|\\
&\leq \epsilon_{DFT}\sqrt{2^q}/\mathcal{N}(\sigma_f,F,q)^{1/2}/(\sigma_f^{1/2}2^{1/4}\pi^{1/4}).
\end{aligned} \tag{D.31}$$

We take the absolute value to have a two-sided inequality, and unwind some replacements:

$$\begin{aligned}
&\left|\left(\frac{F}{T\mathcal{N}(\sigma,T,m)}\right)^{1/2} - 1/\mathcal{N}(\sigma_f,F,q)^{1/2}\right|\max|X(k)|\\
&\leq 2(\sigma/T)^{1/2}\epsilon_{DFT}/\left(F\mathcal{N}(\sigma_f,F,q)\right)^{1/2}/((1/(2\pi^{1/2}))2^{1/4}\pi^{1/4})\\
&\leq 4\pi^{1/2}(\sigma/T)^{1/2}\epsilon_{DFT}/\left(F\mathcal{N}(\sigma_f,F,q)\right)^{1/2}/(2^{1/4}\pi^{1/4})\\
&\leq 2^{7/8}\pi^{3/4}(\sigma/T)^{1/2}\epsilon_{DFT}/\left(F\mathcal{N}(\sigma_f,F,q)\right)^{1/2}\\
&\leq 22^{7/8}\pi^{3/4}(\sigma/T)^{1/2}\left(2^{3/4}\pi^{1/4}\left(\sigma/T\right)^{1/2}/\left(2^q\cdot T\mathcal{N}(\sigma,T,m)\right)^{1/2}\right)\\
&\max\left(4e^{-\pi^2(\sigma/T)^2}, e^{-\left(\frac{2^{m-1}-1}{2\sigma/T}\right)^2}\right)/\left(F\mathcal{N}(\sigma_f,F,q)\right)^{1/2}\\
&\leq 42^{5/8}\pi(\sigma/T)\max\left(4e^{-\pi^2(\sigma/T)^2}, e^{-\left(\frac{2^{m-1}-1}{2\sigma/T}\right)^2}\right)\\
&/\left(2^q\cdot F\mathcal{N}(\sigma_f,F,q)\cdot T\mathcal{N}(\sigma,T,m)\right)^{1/2}
\end{aligned} \tag{D.32}$$

It is worth noting that, for the purpose of estimating $\left(T\mathcal{N}(\sigma,T,m)\right)^{1/2}$, we can increase 2^q arbitrarily (or decrease F arbitrarily). Through this, we see that:

$$\begin{aligned}
\lim_{F\to 0} F\mathcal{N}(\sigma_f,F,q) &= \lim_{F\to 0}\operatorname{erf}\left(\frac{(2^{q-1}-1/2)F}{\sqrt{2}\sigma_f}\right) = \operatorname{erf}\left(\frac{4\sigma\pi}{2\sqrt{2}T}\right)\\
&\geq 1 - \left(\frac{1}{e^{\left(\frac{4\pi}{2\sqrt{2}}\right)^2\left(\frac{\sigma}{T}\right)^2}}\right)
\end{aligned} \tag{D.33}$$

For our choice of $\frac{\sigma}{T}$ this would mean that

$$\left(T\mathcal{N}(\sigma,T,m)\right)^{1/2} = 1 - O\left(\frac{(2^m)^{1/4}}{e^{\Omega(2^m)}}\right), \tag{D.34}$$

and

$$T^{1/2}\mathcal{N}(\sigma,T,m)^{1/2} = F^{1/2}\mathcal{N}(\sigma_f,F,q)^{1/2}$$

$$
\begin{aligned}
&- T^{1/2}\mathcal{N}(\sigma, T, m)^{1/2} O\left(\frac{(2^m)^{1/4}}{e^{\Omega(2^m)}}\right) \\
&= F^{1/2}\mathcal{N}(\sigma_f, F, q)^{1/2} - O\left(\frac{(2^m)^{1/4}}{e^{\Omega(2^m)}}\right).
\end{aligned} \tag{D.35}
$$

Therefore, the amplitudes that we get on the ancillary register register are

$$
\begin{aligned}
=&\frac{1}{\mathcal{N}(\sigma_f, F, q)^{1/2}} X\left(\frac{k}{2^q T}\right) \\
&+ O\left(\frac{(2^m)^{1/2}}{(2^q)^{1/2} e^{\Omega(2^m)}}\right) + \epsilon_{DFT}.
\end{aligned} \tag{D.36}
$$

E Proof of Theorem 13

To begin to prove the previous theorem, we first need to introduce the following alternative lemma for the convergence rate of Chebyshev interpolation:

Lemma 15. *Given an analytic function $f(z)$ over $z \in [-1, 1]$ with an analytic continuation of the function to the complex plane such that $|f(z)| \leq f_{\max}$ for all points on the Bernstein ρ-ellipse, the Chebyshev polynomial interpolation for any $\zeta \in [-1, 1]$ satisfies*

$$
|f(\zeta) - P_n(\zeta)|_\infty \leq \frac{4 f_{\max} \rho^{-n}}{\rho - 1}.
$$

(Theorem 8.2 of [31])

Lemma 15 shows that the error in a Chebyshev interpolation shrinks exponentially with the number of points in the series provided that $\rho < 1$ is a fixed value and $f_{\max}$ is finite. However, in order to specialize the lemma to apply to an eigenvalue $E_k(z)$ of the Hamiltonian $H_{\text{eff}}(z)$ we first recall a result by Bauer and Fike [32] which relates the shift in the operator to the shift in eigenvalues:

Lemma 16. *If A is a normal matrix such that*

$$
K^{-1} A K = \operatorname{diag} \lambda_i \tag{E.1}
$$

where $\|K^{-1}\|\|K\| = 1$. Then, if λ is an eigenvalue of $A + B$, it means that

$$
|\lambda - \lambda_i| \leq \|B\|_2 \quad \text{for at least one value of } i \tag{E.2}
$$

Proof. The matrix $A + B - \lambda I$ is singular and thus the right side of

$$
K^{-1}(A + B - \lambda I)K = \operatorname{diag}(\lambda_i - \lambda) + K^{-1} B K \tag{E.3}
$$

is also singular. Now, we distinguish two cases:

- Case 1. $\lambda = \lambda_i$ for some i.
- Case 2. $\lambda \neq \lambda_i$ for any i, thus we can write

$$\operatorname{diag}(\lambda_i - \lambda) + K^{-1}BK = \operatorname{diag}(\lambda_i - \lambda)\left[I + \operatorname{diag}(\lambda_i - \lambda)^{-1} K^{-1}BK\right] \tag{E.4}$$

Taking determinants we see that the term in brackets must be singular. Now, for that to happen, we must have

$$\|\operatorname{diag}(\lambda_i - \lambda)^{-1} K^{-1}BK\| \geq 1, \tag{E.5}$$

which gives us

$$\max|\lambda_i - \lambda|^{-1}\|K^{-1}\|\|B\|\|K\| \geq 1 \tag{E.6}$$

$$|\lambda - \lambda_i| \leq \|B\|_2 \quad \text{for at least one value of } i \tag{E.7}$$

□

In order to determine the radius of analyticity, we assume a specific form for the upper bound on $\|H - H_{\text{eff}}(z)\|$, here is the resulting theorem

Theorem 17. *If there is an eigenvalue of $H_{\text{eff}}(z)$, $\lambda_m(z)$, with a lower bound Δ on its spectral gap and $\|H - H_{\text{eff}}(z)\| \leq \frac{\alpha|z|^p}{(p+1)!}e^{\beta|z|}$ we have that $\lambda_m(z)$ is analytic on the origin-centered discs with radii r, where*

$$r \leq r_{\text{cross}}^{(low)} = \frac{p}{\beta}\text{LambertW}\left(\frac{\beta}{p}\left(\frac{\Delta(p+1)!}{2\alpha}\right)^{1/p}\right) \leq r_{\text{cross}}. \tag{E.8}$$

Proof. First, by Lemma 15 and Lemma 16 we have that over the domain, D_ρ, defined by the Bernstein ρ-ellipse, we can upper bound the truncation error of a Chebyshev interpolation of the eigenvalues

$$\max_{z\in D_\rho}|E_k(z) - P(z)| \leq \frac{4\max_{z\in D_\rho}|E_k(z)|\rho^{-M}}{\rho - 1},$$

where an upper bound on $\max_z|E_k(z)|$ is given through Lemma 16, which is

$$\max_{z\in D_\rho}|E_k(z)| \leq \sum_{m''}^{m}\|H_{m''}\| + \max_{z\in D_\rho}\|H - H_{\text{eff}}(z)\|$$

Considering Lemma 16,

$$\max_{z\in m_r}|\lambda_{m-1}(0) - \lambda_j(z)| \leq \max_{z\in m_r}\|H - H_{\text{eff}}(z)\|$$

$$\max_{z\in m_r}|\lambda_m(0)-\lambda_k(z)|\leq\max_{z\in m_r}\|H-H_{\text{eff}}(z)\|$$
$$\max_{z\in m_r}|\lambda_{m+1}(0)-\lambda_l(z)|\leq\max_{z\in m_r}\|H-H_{\text{eff}}(z)\| \tag{E.9}$$

for some j, k, and l. We know that $j\neq m$, $l\neq m$ and $k=m$ if

$$\max_{z\in D_r}\|H-H_{\text{eff}}(z)\|\leq\Delta/2. \tag{E.10}$$

Or in the case that we have the upper bound

$$\|H-H_{\text{eff}}(z)\|\leq\alpha\frac{|z|^p}{(p+1)!}e^{\beta|z|}, \tag{E.11}$$

we also know that $j\neq m$, $l\neq m$ and $k=m$ if have the looser constraint

$$\Delta/2\geq\max_{z\in D_r}\left(\alpha\frac{|z|^p}{(p+1)!}e^{\beta|z|}\right)=\alpha\frac{r^p}{(p+1)!}e^{\beta r}. \tag{E.12}$$

Solving for r on the last inequality gives us

$$r\leq\frac{p}{\beta}\text{LambertW}\left(\frac{\beta}{p}\left(\frac{\Delta(p+1)!}{2\alpha}\right)^{1/p}\right). \tag{E.13}$$

□

However, since we need an analytic domain bounded by an ellipse, the following result becomes useful

Corollary 18. *Given that a function $f(z)$ is analytic within discs of radii $r\leq r_{\text{cross}}^{(\text{low})}$, then, the Bernstein ρ-ellipses that we can draw in which $f(z)$ is analytic are*

$$1\leq\rho\leq\rho_{\text{cross}}^{(\text{low})}=r_{\text{cross}}^{(\text{low})}+\sqrt{\left(r_{\text{cross}}^{(\text{low})}\right)^2-1} \tag{E.14}$$

Proof. The major axis of a ρ-ellipse is:

$$x=(\rho+\rho^{-1})/2.$$

Thus, the largest ellipse that is circumvented by $r_{\text{cross}}^{(\text{low})}$ solves the equation

$$(\rho_{\text{cross}}^{(\text{low})}+1/\rho_{\text{cross}}^{(\text{low})})/2=r_{\text{cross}}^{(\text{low})} \tag{E.15}$$

Solving for $\rho_{\text{cross}}^{(\text{low})}$, we obtain:

$$\rho_{\text{cross}}^{(\text{low})}=r_{\text{cross}}^{(\text{low})}+\sqrt{\left(r_{\text{cross}}^{(\text{low})}\right)^2-1}.$$

□

Having derived the main result of this subsection, we will continue by justifying our choice of form on the upper bound for $\|H - H_{\text{eff}}(z)\|$ used in Theorem 17.

Here, we follow closely the work done in [33] for estimating the operator error in the effective Hamiltonian. In order to make assertions over the general complex plane we will extend their results to complex time. The extension will be straight forward because

$$e^{i\tau A}$$

has no singularities except for non-removable singularities at $\text{Im}(\tau) \in \{-\infty, \infty\}$ for a Hermitian A and at $\text{Im}(\tau) = -\infty$ when it is positive semi-definite. Thus, we can replace all the real-axis integrals

$$\int_0^t f(\tau)\mathrm{d}\tau$$

for the contour integrals

$$\int_C f(\tau)\mathrm{d}\tau,$$

where C is any path from 0 to a generally complex t as long as that path does not include $\text{Im}(\tau) \in \{-\infty, \infty\}$. Our results are summarized by

Theorem 19. *The norm of operator error can be estimated*

$$\|H - H_{\mathit{eff}}(z)\| \leq \sum_{(\upsilon,m)} \alpha_{\text{comm}}\Big(a_{(\Upsilon,m)}H_{\pi_\Upsilon(m)}, \ldots, a_{(\upsilon,m+1)}H_{\pi_\upsilon(m+1)}, a_{(\upsilon,m)}H_{\pi_\upsilon(m)}\Big) \cdot \frac{|z|^p}{(p+1)!} e^{2|z|\sum_j^m \|H_j\|}$$

where $\alpha_{\text{comm}}\Big(A_s, \ldots, A_1, B\Big) := \sum_{q_1+\cdots+q_s=p} \binom{p}{q_1 \ \cdots \ q_s} \Big\|\text{ad}_{A_s}^{q_s} \cdots \text{ad}_{A_1}^{q_1}(B)\Big\|$

Proof. First, we define a general product formula the following way

$$\mathscr{S}(t) := \prod_{\upsilon=1}^{\Upsilon} \prod_{m=1}^{m} e^{ita_{(\upsilon,m)}H_{\pi_\upsilon(m)}}, \tag{E.16}$$

where the coefficients $a_{(\upsilon,m)}$ are real numbers. The parameter Υ denotes the number of *stages* of the product formula. For example, for the Suzuki formula $\mathscr{S}_{2k}(t)$, we have $\Upsilon = 2 \cdot 5^{k-1}$. The permutation π_υ controls the ordering of operator summands within stage υ of the product formula. We then consider the error in the effective Hamiltonian with the help of the definition of the path-ordered exponential:

$$\mathscr{S}(t) = \exp_{\mathcal{C}}\left(i\int_C \mathrm{d}\tau\Big(H + \mathscr{E}(\tau)\Big)\right). \tag{E.17}$$

Using this definition, we can write the operator error on the Hamiltonian the following way:

$$H_{\text{eff}}(t) - H = \left(\int_C \mathrm{d}\tau\mathscr{E}(\tau)\right)/t. \tag{E.18}$$

For convenience, authors in [33] proposed the lexicographic order on a pair of tuples (υ, m) and (υ', m'), defined as follows: one writes $(\upsilon, m) \succeq (\upsilon', m')$ if $\upsilon > \upsilon'$, or if $\upsilon = \upsilon'$ and $m \geq m'$. We have $(\upsilon, m) \succ (\upsilon', m')$ if both $(\upsilon, m) \succeq (\upsilon', m')$ and $(\upsilon, m) \neq (\upsilon', m')$ hold. Notations $(\upsilon, m) \preceq (\upsilon', m')$ and $(\upsilon, m) \prec (\upsilon', m')$ are defined in a similar way, except that we reverse the directions of all the inequalities. Imposing this convention on the evolution operator, it

$$\mathscr{S}(t) = \overleftarrow{\prod_{(\upsilon,m)}} e^{ita_{(\upsilon,m)}H_{\pi_\upsilon(m)}}, \tag{E.19}$$

where the symbol $\prod^{\leftarrow}_{(\upsilon,m)}$ means that the operations are applied from right to left in ascending lexographical order defined by the tuple (υ, m), $\prod^{\rightarrow}_{(\upsilon,m)}$ means in descending order. After differentiating the evolution operator and some algebraic manipulation one obtains that $\mathscr{E}(\tau)$ can be expressed as

$$\mathscr{E}(\tau) = \sum_{(\upsilon,m)} \overleftarrow{\prod_{(\upsilon',m')\succ(\upsilon,m)}} e^{i\tau a_{(\upsilon',m')}H_{\pi_{\upsilon'}(m')}} \left(a_{(\upsilon,m)}H_{\pi_\upsilon(m)}\right) \overrightarrow{\prod_{(\upsilon',m')\succ(\upsilon,m)}} e^{-i\tau a_{(\upsilon',m')}H_{\pi_{\upsilon'}(m')}} - H. \tag{E.20}$$

We can bound the norm of this term through the following Taylor expansion of the term

$$\begin{aligned}
&e^{i\tau A_s}\cdots e^{i\tau A_2}e^{i\tau A_1}Be^{-i\tau A_1}e^{-i\tau A_2}\cdots e^{-i\tau A_s} \\
&= C_0 + C_1\tau + \cdots + C_{p-1}\tau^{p-1} \\
&\quad + i^p \sum_{k=1}^{s} \sum_{\substack{q_1+\cdots+q_k=p \\ q_k\neq 0}} e^{i\tau A_s}\cdots e^{i\tau A_{k+1}} \\
&\qquad\qquad \cdot \int_C \mathrm{d}\tau_2\; e^{i\tau_2 A_k}\mathrm{ad}_{A_k}^{q_k}\cdots\mathrm{ad}_{A_1}^{q_1}(B)e^{-i\tau_2 A_k} \cdot \frac{(\tau-\tau_2)^{q_k-1}\tau^{q_1+\cdots+q_{k-1}}}{(q_k-1)!q_{k-1}!\cdots q_1!} \\
&\qquad\qquad \cdot e^{-i\tau A_{k+1}}\cdots e^{-i\tau A_s},
\end{aligned} \tag{E.21}$$

where $\mathrm{ad}_A(B) = [A, B]$. We can bound the norm of that last term in Equation (E.21) through triangle inequality for contour integrals, and for sake of simplicity, we fix contour C going straight from 0 to τ. We also assume for the moment that matrices A_j are positive definite. Thus, the upper bound becomes:

$$\begin{aligned}
&\sum_{k=1}^{s} \sum_{\substack{q_1+\cdots+q_k=p \\ q_k\neq 0}} \prod_{j=k+1}^{s} \overbrace{\|e^{i\tau A_j}\|\|e^{-i\tau A_j}\|}^{\leq e^{2|\mathrm{Im}(\tau)|\,\|A_j\|}} \frac{|\tau|^{q_1+\cdots+q_{k-1}}}{(q_k-1)!q_{k-1}!\cdots q_1!} \left\|\mathrm{ad}_{A_k}^{q_k}\cdots\mathrm{ad}_{A_1}^{q_1}(B)\right\| \\
&\qquad \int_C \mathrm{d}\tau_2|\tau-\tau_2|^{q_k-1}\overbrace{\|e^{i\tau_2 A_k}\|\|e^{-i\tau_2 A_k}\|}^{\leq e^{2|\mathrm{Im}(\tau_2)|\,\|A_k\|}}
\end{aligned} \tag{E.22}$$

$$\leq \sum_{k=1}^{s} \sum_{\substack{q_1+\cdots+q_k=p \\ q_k\neq 0}} \binom{p}{q_1\ \cdots\ q_k} \frac{|\tau|^p}{p!} \left\|\mathrm{ad}_{A_k}^{q_k}\cdots\mathrm{ad}_{A_1}^{q_1}(B)\right\| e^{2|\tau|\sum_{j=k}^{s}\|A_j\|} \tag{E.23}$$

$$\leq \sum_{q_1+\cdots+q_s=p} \binom{p}{q_1 \ \cdots \ q_s} \frac{|\tau|^p}{p!} \left\| \mathrm{ad}_{A_s}^{q_s} \cdots \mathrm{ad}_{A_1}^{q_1}(B) \right\| e^{2|\tau| \sum_{j=1}^{s} \|A_j\|} \tag{E.24}$$

$$= \alpha_{\mathrm{comm}}\Big(A_s, \ldots, A_1, B\Big) \frac{|\tau|^p}{p!} e^{2|\tau| \sum_{j=1}^{s} \|A_j\|}, \tag{E.25}$$

where

$$\alpha_{\mathrm{comm}}\Big(A_s, \ldots, A_1, B\Big) := \sum_{q_1+\cdots+q_s=p} \binom{p}{q_1 \ \cdots \ q_s} \left\| \mathrm{ad}_{A_s}^{q_s} \cdots \mathrm{ad}_{A_1}^{q_1}(B) \right\|. \tag{E.26}$$

Thus, the upper bound on $\|\mathscr{E}(z)\|$ is

$$\|\mathscr{E}(z)\| \leq \sum_{(\upsilon,m)} \alpha_{\mathrm{comm}}\Big(a_{(\Upsilon,m)} H_{\pi_\Upsilon(m)}, \ldots, a_{(\upsilon,m+1)} H_{\pi_\upsilon(m+1)}, a_{(\upsilon,m)} H_{\pi_\upsilon(m)}\Big) \cdot \frac{|z|^p}{p!} e^{2|z| \sum_{(\upsilon',m') \succ (\upsilon,m)} \left\| a_{(\upsilon',m')} H_{\pi_{\upsilon'}(m')} \right\|}, \tag{E.27}$$

where $(\upsilon, m+1) = (\upsilon+1, 1)$. With this, the upper bound in $\|H_{\mathrm{eff}}(t) - H\|$

$$\begin{aligned} \|H - H_{\mathrm{eff}}(z)\| &\leq \sum_{(\upsilon,m)} \alpha_{\mathrm{comm}}\Big(a_{(\Upsilon,m)} H_{\pi_\Upsilon(m)}, \ldots, a_{(\upsilon,m+1)} H_{\pi_\upsilon(m+1)}, a_{(\upsilon,m)} H_{\pi_\upsilon(m)}\Big) \\ &\qquad \cdot \frac{|z|^p}{(p+1)!} e^{2|z| \sum_{(\upsilon',m') \succ (\upsilon,m)} \left\| a_{(\upsilon',m')} H_{\pi_{\upsilon'}(m')} \right\|} \\ &\leq \sum_{(\upsilon,m)} \alpha_{\mathrm{comm}}\Big(a_{(\Upsilon,m)} H_{\pi_\Upsilon(m)}, \ldots, a_{(\upsilon,m+1)} H_{\pi_\upsilon(m+1)}, a_{(\upsilon,m)} H_{\pi_\upsilon(m)}\Big) \\ &\qquad \cdot \frac{|z|^p}{(p+1)!} e^{2|z| \sum_{j}^{m} \|H_j\|} \end{aligned} \tag{E.28}$$

□

While the above theorem is a useful bound for estimating the difference between the two Hamiltonians a problem arises because bounding each of the commutators individually is often impractical computationally. In such cases, $\alpha_{\mathrm{comm}}\Big(A_s, \ldots, A_1, B\Big)$ can be upper bounded as

$$\alpha_{\mathrm{comm}}\Big(A_s, \ldots, A_1, B\Big) \leq \|B\| 2^p (\sum_{j=1}^{s} \|A_j\|)^p. \tag{E.29}$$

For the Gaussian QPEA defined in the last section, we must define the U_k operator. A first attempt looks the following way

$$U_k' = \left(S_p\left(t s_k^{(M)}\right)\right)^{d_k'^{(M)}} = \left(e^{its_k^{(M)} H_{\mathrm{eff},p}}\right)^{d_k'^{(M)}}, \tag{E.30}$$

where $d_k'^{(M)}$ is chosen such that the simulation times are equivalent for each U_k', thus keeping the resolution of the corresponding QPEA the same for each k. That is,

$$d_k'^{(M)} = \frac{s_1^{(M)}}{s_k^{(M)}}. \tag{E.31}$$

This, however, will require fractional powers of (fractional number of queries of) $S_p\left(ts_k^{(M)}\right)$. Although this can be efficiently achieved through quantum signal processing and one extra ancillary qubit, within this work, we would like to keep it theoretically simpler. Here, we will set out to recover integer query numbers $d_k^{(M)}$, where

$$d_k^{(M)} = \operatorname{sgn} s_k^{(M)} \left\lceil \frac{s_1^{(M)}}{\left|s_k^{(M)}\right|} \right\rceil. \tag{E.32}$$

This has the effect of increasing the resolution and the cost of QPEA by a factor of

$$\frac{d_k^{(M)}}{d_k'^{(M)}} \leq 2. \tag{E.33}$$

Asymptotically, the cost can still be upper bounded by a quantity proportional to L_0. Finally, similarly to U_k' , U_k is defined through

$$U_k = \left(S_p\left(ts_k^{(M)}\right)\right)^{d_k^{(M)}} = \left(e^{its_k^{(M)} H_{\text{eff},p}}\right)^{d_k^{(M)}}. \tag{E.34}$$

As a last step before jumping into the proof of Theorem 13 we will first discuss some estimates for the coefficients of the terms in the extrapolation. The first such result is given in In the framework we propose here, most algorithms' cost will scale proportionally to $\|d\|_1 \leq 2\|d'\|_1$. The following Lemma provides a convenient bound on this norm which will be needed in the proof of Theorem 13.

Lemma 20. *Given the vector d' where $d_k' = 1/\cos\left(\frac{2k-1}{2M}\pi\right)$. Then, its 1-norm can be calculated to be*

$$\|d'\|_1 \leq \frac{2M}{\pi}\left(\gamma + \log(M+1)\right),$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

Proof.

$$\|d'\|_1 = \sum_{k=1}^{M} \frac{1}{\left|\cos\left(\frac{2k-1}{2M}\pi\right)\right|} = \sum_{k=1}^{M} \frac{1}{\left|\sin\left(\frac{M-2k+1}{2M}\pi\right)\right|} \tag{E.35}$$

We can rewrite the last expression the following way

$$\sum_{k=1}^{M} \frac{1}{\left|\sin\left(\frac{M-2k+1}{2M}\pi\right)\right|} \tag{E.36}$$

$$= \sum_{k=1}^{M/2} \frac{1}{\left|\sin\left(\frac{M-2k+1}{2M}\pi\right)\right|} \tag{E.37}$$

$$+ \sum_{k=M/2+1}^{M} \frac{1}{\left|\sin\left(\frac{M-2k+1}{2M}\pi\right)\right|} \tag{E.38}$$

$$= 2\sum_{k=1}^{M/2} \frac{1}{\sin\left(\frac{2k-1}{2M}\pi\right)} \tag{E.39}$$

$$\tag{E.40}$$

We can now use the following lower bound on $\sin(x)$

$$\sin x \geq x/2 \quad (0 \leq x \leq \pi/2) \tag{E.41}$$

in order to bound the last term in Equation (E.70)

$$2\sum_{k=1}^{M/2} \frac{1}{\sin\left(\frac{2k-1}{2M}\pi\right)} \tag{E.42}$$

$$\leq \frac{4M}{\pi} \sum_{k=1}^{M/2} 1/\left(2k-1\right) \tag{E.43}$$

$$= \frac{4M}{\pi}\left(H_M - \frac{1}{2}H_{M/2}\right) \tag{E.44}$$

$$\leq \frac{4M}{\pi} H_M = \frac{2M}{\pi}\left(\gamma + \psi(M+1)\right)$$

$$\leq \frac{2M}{\pi}\left(\gamma + \log(M+1)\right)$$

$$\leq M\left(1 + \frac{2M}{\pi}\log(M+1)\right). \tag{E.45}$$

Here, H_n denotes the nth harmonic number and $\psi(n)$ is the digamma function.

□

We would like to get a tighter and also better estimate on the cost for ground state phase estimation (although this result could be extended to amplitude estimation of any state observable). We will allow the variances of observables to vary across nodes and then minimize the cost function:

$$L_0 = \sum_{k=1}^{M/2} \frac{1}{\sigma_k \cos\frac{2k-1}{2M}\pi} \tag{E.46}$$

with the constraint

$$2\sum_{k=1}^{M/2} c_k^2 \sigma_k^2 - \sigma_{F_{\text{opt}}}^2 = 0. \tag{E.47}$$

Thus, we defined the cost function with Lagrange multiplier:

$$L = L_0 + \lambda \left(2 \sum_{k=1}^{M/2} c_k^2 \sigma_k^2 - \sigma_{F_{\text{opt}}}^2 \right) \tag{E.48}$$

We minimize by solving for

$$\frac{\partial L}{\partial \sigma_k} = 0 \tag{E.49}$$

We obtain

$$4\lambda c_k^2 \sigma_k^2 = \frac{1}{\sigma_k \cos \frac{2k-1}{2M}\pi} \tag{E.50}$$

To solve for the Lagrange multiplier we sum both sides over $k = 1, \ldots, M/2$ obtaining:

$$\lambda = \frac{L_0}{2\sigma_{F_{\text{opt}}}^2} \tag{E.51}$$

Replacing back into Equation (E.50) we obtain:

$$\sigma_k = \left(\frac{\sigma_{F_{\text{opt}}}^2}{2c_k^2 \cos\left(\frac{2k-1}{2M}\pi\right) L_0} \right)^{1/3} \tag{E.52}$$

Inserting back into Equation (E.80) and solving for L_0 we obtain:

$$L_0 = \frac{2^{1/2}}{\sigma_P} \left(\sum_{k=1}^{M/2} \frac{|c_k|^{2/3}}{\cos^{2/3}\left(\frac{2k-1}{2M}\pi\right)} \right)^{3/2} \tag{E.53}$$

Finally, the optimal distribution for σ_k's is

$$\sigma_k = \frac{\sigma_P}{2^{1/2}|c_k|^{2/3} \cos^{1/3}\left(\frac{2k-1}{2M}\pi\right) \left(\sum_{k=1}^{M/2} \frac{|c_k|^{2/3}}{\cos^{2/3}\left(\frac{2k-1}{2M}\pi\right)} \right)^{1/2}}. \tag{E.54}$$

We will now estimate the scaling of the cost function L_0 with respect to M. L_0 is of the form

$$L_0 = \frac{2^{1/2}}{\sigma_P} \left(\sum_{k=1}^{M/2} e_k \right)^{3/2}, \tag{E.55}$$

where

$$e_k = \frac{|c_k|^{2/3}}{\cos^{2/3}\left(\frac{2k-1}{2M}\pi\right)} = \left| \frac{\sec^2\left(\frac{\pi - 2k\pi}{2M}\right) \sin\left(\frac{(M-1)(2k+M-1)\pi}{2M}\right)}{M} \right|^{2/3}. \tag{E.56}$$

Equivalently, we can replace the summand by the monotonically increasing function

$$e(k) = \frac{2\left|\sin\left(\frac{\pi-2k\pi}{2M}\right)\right|}{M\cos\left(\frac{\pi-2\pi k}{M}\right)+M} \tag{E.57}$$

Then, we can split the sum $\sum_{k=1}^{M/2} e_k$ into

$$e(M/2) + \sum_{k=1}^{M/2-1} e(k). \tag{E.58}$$

We can take the second term and approximate it from above with the integral

$$\int_1^{M/2} e(k)\,dk. \tag{E.59}$$

From this, we obtain that

$$\int_1^{M/2} e(k)\,dk = \Theta\left(M^{2/3}\right) \tag{E.60}$$

and that the single term is also $e(M/2) = O(M^{2/3})$. Thus,

$$L_0 = O\left(\frac{M}{\sigma_P}\right). \tag{E.61}$$

With this result in hand, it is straight forward to verify that the following bound the product of $\mathbf{V}_{\text{opt}}p(s)$, which is needed to estimate the maximum extrapolation error that can be seen as we extrapolate to $s = 0$.

Lemma 21. *Let $d(s) \in \mathbb{R}^M$ be a vector defined through $d(s) = \mathbf{V}_{\text{opt}}p(s)$ where*

$$\left(\mathbf{V}_{opt}\right)_{kj} = \begin{cases} \sqrt{\frac{1}{M}}\cos\left(j\left(\frac{2k-1}{2M}\pi\right)\right), & j = 0 \\ \sqrt{\frac{2}{M}}\cos\left(j\left(\frac{2k-1}{2M}\pi\right)\right), & j = 1, 2, \ldots M-1, \end{cases} \tag{E.62}$$

and

$$p_i(s) = \begin{cases} \sqrt{\frac{1}{M}}, & i = 0 \\ \sqrt{\frac{2}{M}}\cos(i\arccos(s)), & i = 1, 2, \ldots. \end{cases} \tag{E.63}$$

Then,

$$\|d(s=0)\|_1 \leq 1 + \frac{2}{\pi}\log\left(M+1\right).$$

Proof. First, we define

$$\begin{aligned} \mathbf{d}(s) &= \left(\mathbf{V}_{\text{opt}}^{-1}\right)^{\top} \cdot \mathbf{p}(s) \\ &= \mathbf{V}_{\text{opt}} \cdot \mathbf{p}(s), \end{aligned} \tag{E.64}$$

where

$$(\mathbf{V}_{\text{opt}})_{kj} = \begin{cases} \sqrt{\frac{1}{M}} \cos\left(j\left(\frac{2k-1}{2M}\pi\right)\right), & j = 0 \\ \sqrt{\frac{2}{M}} \cos\left(j\left(\frac{2k-1}{2M}\pi\right)\right), & j = 1, 2, \ldots M-1. \end{cases} \tag{E.65}$$

Contracting indices we get

$$d_k(\tau) = \frac{1}{M} + \sqrt{\frac{2}{M}} \sum_{j=1}^{M-1} \cos\left(j\left(\frac{2k-1}{2M}\pi\right)\right) p_j(s), \tag{E.66}$$

where

$$p_i(s) = \begin{cases} \sqrt{\frac{1}{M}}, & i = 0 \\ \sqrt{\frac{2}{M}} \cos(i \arccos(s)), & i = 1, 2, \ldots, M-1. \end{cases} \tag{E.67}$$

Thus,

$$d_k(s) = \frac{i^{3M} \sin\left(\frac{\pi(M-1)(2k+M-1)}{2M}\right) \cos\left(M \cos^{-1}(s)\right)}{M \cos\left(\frac{\pi-2\pi k}{2M}\right) - Ms} \tag{E.68}$$

which, when $s = 0$ simplifies to

$$d_k(s=0) = \frac{\sin\left(\frac{\pi(M-1)(2k+M-1)}{2M}\right) \sec\left(\frac{\pi-2\pi k}{2M}\right)}{M}. \tag{E.69}$$

Now, the 1-norm of $d(s=0)$ is

$$\|d(s=0)\|_1 \leq \frac{1}{M} \sum_{k=1}^{M} \frac{1}{\cos\left(\frac{2k-1}{2M}\pi\right)} = \frac{1}{M} \|d'\|_1.$$

Finally, using Lemma 20, we obtain

$$\|d(s=0)\|_1 \leq 1 + \frac{2}{\pi} \log(M+1).$$

□

Finally, with these preliminaries in place, we can conclude with the proof of Theorem 13, which provides estimates of the cost of interpolation using the prepare first approach.

Proof of Theorem 13. In the framework we propose here, most algorithms' cost will scale from Lemma 21

$$\|d\|_1 \leq 2\|d'\|_1 = \sum \frac{2}{\left|\cos\left(\frac{2k-1}{2M}\pi\right)\right|} = O\left(M \log M\right). \tag{E.70}$$

For example, the ground state preparation methods detailed in [25] would scale proportionally to $\|d\|_1$. More precisely, the cost of state preparation scales with the minimum eigenvalue gap as

$$C_{\text{prep}} = O\left(\frac{\|d\|_1}{t \min_k \left(E_1(s_k) - E_0(s_k)\right)}\right), \tag{E.71}$$

where, from Lemma 16 and the fact that the spectral gap is bounded below by Δ by assumption in Theorem 13, we have that

$$\begin{aligned}\min_k \left(E_1(s_k) - E_0(s_k)\right) &\geq \Delta - \max_{x \in I_{-t,t}} \|H - H_{\text{eff}}(x)\| \\ &\geq \Delta - O\left(t^p\right). \end{aligned} \tag{E.72}$$

Thus, for small $t = O(\Delta^{1/p})$

$$\min_k \left(E_1(s_k) - E_0(s_k)\right) = \Omega(\Delta) \tag{E.73}$$

From Theorem 17, we also have that

$$r = \Omega\left(\left(\frac{(p+1)!}{2^{p+1}}\right)^{1/p} \frac{\Delta^{1/p}}{t}\right) \tag{E.74}$$

Recalling Lemma 15 that the interpolation error

$$\epsilon = O\left(\frac{1}{\rho^{M+1}}\right) \tag{E.75}$$

Thus, solving for M we find:

$$M = O\left(\frac{\log 1/\epsilon}{\log \rho}\right) \tag{E.76}$$

We also have that the radius of the Bernstein elipse needed satisfies $\rho = \Theta(r)$ from Corollary 18; hence,

$$M = O\left(\frac{\ln \frac{1}{\epsilon}}{\ln\left(\frac{\Delta^{1/p}}{t}\right) + \left(\ln \frac{(p+1)!}{2^{p+1}}\right)/p}\right). \tag{E.77}$$

It is clear that a constraint is

$$t = \Omega\left(\Delta^{1/p}\right), \tag{E.78}$$

since we need the radius to be a quantity greater than 1 and there is no advantage in going to a much smaller t. Finally, the cost for state preparation is

$$C_{\text{prep}} = O\left(\frac{\ln \frac{1}{\epsilon} \ln\ln \frac{1}{\epsilon}}{\Delta^{1+1/p}}\right). \tag{E.79}$$

Now, the cost of estimating the energy through GPE would, naively speaking, also scale like $O(M \log M)$. However, to get a tighter cost bound, we will allow the variances of observables to vary across nodes and then minimize the cost function:

$$L_0 = \sum_{k=1}^{M/2} \frac{1}{\sigma_k \cos \frac{2k-1}{2M}\pi} \tag{E.80}$$

with the constraint

$$2\sum_{k=1}^{M/2} c_k^2 \sigma_k^2 - \sigma_P^2 = 0. \tag{E.81}$$

Thus, we the cost function with Lagrange multiplier is:

$$L = L_0 + \lambda \left(2\sum_{k=1}^{M/2} c_k^2 \sigma_k^2 - \sigma_P^2 \right). \tag{E.82}$$

Note that the terms in the cost function L_0 go like $1/\sigma_k$ due to Fourier duality and Theorem 12 and the other factor is from the cost of a single U_k implementation. After doing the minimization we are left with (See (E.53) and surrounding equations):

$$\begin{aligned} L_0 &= \frac{2^{1/2}}{\sigma_P} \left(\sum_{k=1}^{M/2} \frac{|c_k|^{2/3}}{\cos^{2/3}\left(\frac{2k-1}{2M}\pi\right)} \right)^{3/2} \\ \sigma_k &= \frac{\sigma_P}{2^{1/2}|c_k|^{2/3} \cos^{1/3}\left(\frac{2k-1}{2M}\pi\right) \left(\sum_{k=1}^{M/2} \frac{|c_k|^{2/3}}{\cos^{2/3}\left(\frac{2k-1}{2M}\pi\right)} \right)^{1/2}}. \end{aligned} \tag{E.83}$$

The optimal L_0 follows

$$L_0 = O\left(\frac{M}{\sigma_P}\right). \tag{E.84}$$

Now, the cost for estimating the interpolant with a variance of σ_P and a bias ϵ using Gaussian phase estimation is

$$C_{\text{est}} = O\left(\frac{\ln \frac{1}{\epsilon}}{\sigma_P}\right) \tag{E.85}$$

On the other hand, using Lemma 6, if one wishes to estimate the energy using a single ancillar qubit approach, the cost of estimating the energy is

$$C_{\text{est,1-qubit}} = O\left(\frac{\log \frac{1}{\epsilon} \log^2 \log \frac{1}{\epsilon}}{\epsilon'}\right), \tag{E.86}$$

where ϵ' is the semi-deterministic error coming from the Heisenberg-limited estimation algorithms like IQAE[34] or the one in Ref. [35], or non-Heisenberg-limited alternatives like the semi-classical QPE[36, 37], the single-qubit version of the textbook QPE[26]. □