Article:

Statistical Phase Estimation and Error Mitigation on a Superconducting Quantum Processor

Nick S. Blunt[1,*,‡] Laura Caune,[1,†,‡] Róbert Izsák,[1] Earl T. Campbell,[1,2] and Nicole Holzmann[1,3]

[1]*Riverlane, Cambridge CB2 3BZ, United Kingdom*

[2]*Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, United Kingdom*

[3]*Astex Pharmaceuticals, 436 Cambridge Science Park, Cambridge, CB4 0QA, United Kingdom*

Quantum phase estimation (QPE) is a key quantum algorithm, which has been widely studied as a method to perform chemistry and solid-state calculations on future fault-tolerant quantum computers. Recently, several authors have proposed statistical alternatives to QPE that have benefits on early fault-tolerant devices, including shorter circuits and better suitability for error-mitigation techniques. However, experimental investigations of the algorithm on real quantum processors are lacking. Here, we implement statistical phase estimation on Rigetti's superconducting processors. Specifically, we use a modification of the Lin and Tong [PRX Quantum 3, 010318 (2022)] algorithm with the improved Fourier approximation of Wan *et al.* [Phys. Rev. Lett. 129, 030503 (2022)] and apply a variational-compilation technique to reduce the circuit depth. We then incorporate error-mitigation strategies including zero-noise extrapolation and readout-error mitigation with bit-flip averaging. We propose a new method to estimate energies from the statistical phase estimation data, which is found to improve the accuracy in the final energy estimates by 1–2 orders of magnitude with respect to prior theoretical bounds, reducing the cost of performing accurate phase-estimation calculations. We apply these methods to chemistry problems for active spaces up to four electrons in four orbitals, including the application of a quantum embedding method, and use them to correctly estimate energies within chemical precision. Our work demonstrates that statistical phase estimation has a natural resilience to noise, particularly after mitigating coherent errors, and can achieve far higher accuracy than suggested by previous analysis, demonstrating its potential as a valuable quantum algorithm for early fault-tolerant devices.

I. INTRODUCTION

Quantum phase estimation (QPE) [1,2] is one of the most widely studied quantum algorithms, due to its potential for exponential speed-ups in a wide range of problems. While this potential is promising, the quantum circuits involved in useful applications of QPE have a high depth. Because of this, QPE is often described as a fault-tolerant quantum algorithm, which will require large-scale quantum error correction (QEC) for nontrivial applications. Many studies have been performed in recent years to assess the resources required to apply QPE to active spaces at the limit of current classical algorithms, often estimating the need for tens or hundreds of millions of qubits using current QPE and QEC schemes [3–6].

In the past few years, statistical modifications to the QPE algorithm have been proposed [7–14] that use lower-depth circuits and far fewer auxiliary qubits than techniques based on qubitization [6,15] and so are better suited to near-term devices. Importantly, statistical QPE is also more readily combined with error-mitigation techniques, many of which are primarily designed for estimating expectation values. "Textbook" and many other QPE approaches measure a discrete output, namely, the bits of the energy estimate, and are therefore not compatible with most error-mitigation techniques. In contrast, the circuits involved in statistical phase estimation methods are typically Hadamard tests, the output of which is an expectation value, as shown in Fig. 1. Multiple such circuits are performed, with the resulting expectation values used to construct an appropriate function, from which the desired eigenvalues may be estimated.

---

*nick.blunt@riverlane.com

†laura.caune@riverlane.com
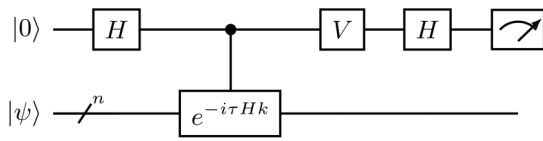
‡These authors contributed equally to this work.

FIG. 1. The Hadamard test circuits that are considered in this work. Setting $V = \mathbb{1}$ or $V = S^{\dagger}$ allows estimation of Re[$\langle\psi| e^{-i\tau Hk}|\psi\rangle$] and Im[$\langle\psi| e^{-i\tau Hk}|\psi\rangle$], respectively.

Error mitigation has been a major theme in early practical applications of quantum computing, including applications to chemistry and condensed-matter physics problems. Current quantum processors are often referred to as noisy intermediate-scale quantum (NISQ) devices due to their high error rates and low qubit counts. In the absence of QEC, which requires higher qubit counts and lower error rates than currently available, error mitigation has been widely investigated. This includes techniques for readout error mitigation [16,17], as well as numerous methods for mitigating gate errors, such as zero-noise extrapolation (ZNE) [18–20], probabilistic error cancellation (PEC) [18,21,22], Clifford data regression (CDR) [23,24], noise tailoring techniques such as randomized compiling (RC) [25,26], and symmetry constraints or postselection [27–29]. See also Ref. [30] for a recent review of error-mitigation techniques and Ref. [31] for a recent study testing ZNE and PEC on multiple quantum computing platforms.

In addition to NISQ studies, there is good reason to believe that error mitigation will continue to be important in the early fault-tolerant regime, where partial error correction is possible [32–35]. For example, a recent study has investigated the resources required to perform even the simplest ground-state calculation on an H_2 molecule using the surface code [36]; even here, it has been found that thousands of physical qubits are required. The largest overhead comes from the need for magic state distillation factories and decomposing rotation gates into T gates. This cost could be significantly reduced using error mitigation. For example, Piveteau *et al.* [32] have discussed approaches to perform noisy T gates and apply error mitigation. It should be emphasized that error-mitigation techniques in general have a cost that scales exponentially with the circuit depth [37]. However, in the near-term regime of NISQ and early fault tolerance, they offer valuable methods to improve the performance of statistical QPE.

Examples of prior QPE calculations on quantum hardware include an early study of Kitaev's iterative QPE algorithm using superconducting qubits [38]; an application of textbook QPE on a neutral-atom quantum computer [39]; and, very recently, a study of Bayesian QPE on a trapped-ion quantum computer [40]. All of these studies consider a minimal H_2 model, using a Hamiltonian of only one or two qubits, demonstrating the challenge of performing QPE in practice.

In this paper, we present the first combination of statistical phase estimation and error mitigation on a current quantum device. The methodology presented allows us to perform ground-state energy calculations to accuracy better than 0.1 mHa for the systems considered, which is comparable to, or better than, the accuracy that could be previously achieved with the variational quantum eigensolver (VQE) and current hardware [41–43]. Statistical phase estimation also avoids problematic barren plateaus that hamper the scalability of VQE, although the speed-up of all QPE algorithms does depend on the quality of the initial trial states [44]. For error mitigation, we focus on ZNE, symmetrized readout-error mitigation [17], and RC methods. Although our results are performed on current NISQ devices, the error-mitigation techniques presented can also be applied in the early fault-tolerant regime. For example, we apply RC to mitigate coherent errors for controlled-Z (CZ) gates in this work but, as described in Ref. [25], it can equally be applied to T gates in an error-corrected scheme.

We focus on a statistical QPE approach based on the cumulative distribution function (CDF) of the spectral measure of the Hamiltonian, which has been introduced by Lin and Tong [8], and use the improved Fourier approximation derived by Wan *et al.* [9] (but do not investigate the randomized compilation approach for Hamiltonian simulation introduced in the same paper). We also implement and test the quantum eigenvalue estimation algorithm (QEEA) by Somma [7]. In both approaches, we apply importance sampling as described in Ref. [8]. We apply these methods to several molecules, including two examples motivated by pharmaceutical applications, using a chemical-embedding approach to construct relevant active spaces [45]. These results are achieved by using a variational-circuit-compilation strategy to allow the required operations to be performed with a low circuit depth, suitable for current quantum processors. We also study a Trotterized example, which avoids the exponential preprocessing step of the variational-compilation method, using two-qubit gate depths up to 100.

We show that mitigating coherent errors [25,26] is important in statistical phase estimation and that this methodology can be effectively combined with importance sampling, allowing energies to be extracted with confidence, even in the presence of significant QPU errors. We further introduce a simple approach to estimate energies from the QPU data with significantly improved accuracy, compared to the theoretical bounds derived in Ref. [9]. The largest example that we study is a model of a pharmaceutically relevant molecule, where the ground-state energy is again obtained with an error of less than 0.1 mHa.

The structure of the paper is as follows. In Sec. II A, we cover the theory of statistical phase estimation, particularly the methods of Refs. [8,9], including importance sampling. In Sec. II B, we discuss a variational-compilation method to reduce the circuit depth. Section II C introduces the

error-mitigation techniques to be applied. Results are then presented from Rigetti's Aspen devices, first applying the statistical phase estimation method to an example with two electrons in two orbitals, followed by a study of larger active spaces in combination with error-mitigation strategies, and a final example using a Trotterized expansion of the time-evolution operator.

II. THEORY

A. Statistical phase estimation

We consider n-qubit Hamiltonians of the form

$$H = \sum_{l=1}^{L} c_l P_l, \tag{1}$$

where P_l are n-qubit Pauli operators, and denote the eigenvalues and eigenvectors of H by $\{\lambda_i; |\Psi_i\rangle\}$. We will be concerned with estimating $\{\lambda_i\}$ using statistical phase-estimation methods. For these techniques, it is necessary to bound the Hamiltonian in a known range and we therefore work with a scaled Hamiltonian τH, where $\tau > 0$.

In this paper, we focus on circuits of the form shown in Fig. 1. This is a Hadamard test circuit, where setting $V = \mathbb{1}$ or $V = S^\dagger = |0\rangle\langle0| - i|1\rangle\langle1|$ allows measurement in the X or Y bases, respectively. For $V = \mathbb{1}$, defining a random variable X equal to $+1$ for $|0\rangle$ measurements and -1 for $|1\rangle$ measurements, it can be shown that

$$E[X] = \text{Re}[\langle\psi|e^{-i\tau Hk}|\psi\rangle]. \tag{2}$$

where throughout $E[\cdot]$ denotes an expectation value. Similarly, for $V = S^\dagger$ measurements, defining a random variable Y equal to $+1$ for $|0\rangle$ measurements and -1 for $|1\rangle$ measurements gives

$$E[Y] = \text{Im}[\langle\psi|e^{-i\tau Hk}|\psi\rangle]. \tag{3}$$

Performing the circuits of Fig. 1 therefore allows estimation of

$$g_k = \langle\psi|e^{-i\tau Hk}|\psi\rangle \tag{4}$$

up to statistical errors, which are controlled by averaging over multiple repetitions of the circuit, or "shots." This g_k is the main quantity of interest that we seek to estimate by quantum computation.

In general, $|\psi\rangle$ will not be an exact eigenstate of H. We denote the expansion of $|\psi\rangle$ in the eigenbasis of H by

$$|\psi\rangle = \sum_i v_i|\Psi_i\rangle, \tag{5}$$

which gives

$$g_k = \sum_i p_i e^{-i\tau\lambda_i k}, \tag{6}$$

where we define $p_i = |v_i|^2$. Therefore, g_k will in general consist of a sum of oscillating signals, the frequencies of which are determined by the energies of H and the amplitudes of which are determined by the components of the corresponding eigenstates in $|\psi\rangle$. The goal of statistical phase estimation methods is to extract (some of) the phases λ_i from the noisy estimates of g_k and hence estimate the energies of H.

Multiple such methods have been introduced in recent years, each suggesting different techniques to construct eigenvalue estimates from the g_k estimates. Roughly speaking, these methods involve identifying some function $f(H)$ that allows eigenvalues to be identified. The function $f(H)$ is then expanded in a Fourier series, which can be constructed, after truncation, using the g_k estimates. Truncation is needed to put a finite limit on the unitary time evolution τk. Note that once g_k has been constructed, the task of estimating the desired λ_i is a purely classical task and is related to similar problems in signal processing.

We begin by reviewing the approach of Wan et al. [9]. This approach is based on a similar method by Lin and Tong in Ref. [8] but uses an alternative Fourier approximation that allows the authors to prove a better asymptotic complexity. Note that Ref. [9] also introduces a randomized compiling approach to implement $e^{-i\tau Hk}$ with reduced circuit depth, but this approach is not tested in this paper.

1. CDF-based statistical phase estimation

In the approach of Refs. [8,9], we wish to calculate a cumulative distribution function (CDF) associated with the Hamiltonian and state $|\psi\rangle$,

$$C(x) = \sum_{i:\tau\lambda_i\leq x} p_i. \tag{7}$$

If $C(x)$ could be constructed, then it would allow us to identify the eigenvalues of H through its discontinuities. In practice, the CDF will be constructed as a sum of terms e^{ikx} for integer values of k and so it is necessary to define $C(x)$ to be 2π periodic. We therefore instead define

$$C(x) = \int_{-\pi/2}^{\pi/2} p(y)\Theta(x - y)dy, \tag{8}$$

where $\Theta(x)$ is a 2π-periodic Heaviside step function and $p(y)$ is the probability distribution of τH associated with $|\psi\rangle$,

$$p(y) = \sum_i p_i\delta(y - \tau\lambda_i). \tag{9}$$

It is straightforward to check that this definition gives the desired $C(x) = \sum_{i:\lambda_i\leq x} p_i$ for $|x| \leq \pi/2$. For $|x| > \pi/2$, this is not true and we should therefore choose τ such that

$\|\tau H\| \leq \pi/2$, where $\|\ldots\|$ denotes the operator norm. The discontinuities of $C(x)$ in the range $|x| \leq \pi/2$ can then be used to identify eigenvalues of τH.

One can proceed by considering a Fourier-series expansion of the step function. Reference [9] defines

$$F(x) = \sum_{|k| \leq N} F_k e^{ikx}, \tag{10}$$

where $F(x)$ is an approximation to $\Theta(x)$, which is derived by truncating the Chebyshev polynomial expansion for an erf($\cdot$) function. The authors carefully constructed this approximation to satisfy various bounds on its error and on the scaling of $\sum_{|k| < N} |F_k|$ with N. The derived Fourier coefficients are

$$F_0 = 1/2,$$

$$F_{2j+1} = -i\sqrt{\frac{\beta}{2\pi}} e^{-\beta} \frac{I_j(\beta) + I_{j+1}(\beta)}{2j+1}, \quad 0 \leq j \leq d-1$$

$$F_{2d+1} = -i\sqrt{\frac{\beta}{2\pi}} e^{-\beta} \frac{I_d(\beta)}{2d+1}, \tag{11}$$

where d is related to N above by $N = 2d + 1$. In addition, we have $F_{-k} = -F_k$ for all $k \neq 0$. $I_n(\beta)$ is the nth modified Bessel function of the first kind. The Fourier coefficient F_k is nonzero for odd k only; even k do not contribute.

The Fourier coefficients, and hence the approximate Heaviside function $F(x)$, depend on a parameter $\beta > 0$, which represents the sharpness of the erf($\cdot$) function. For untruncated k, the approximation becomes more accurate as β increases. However, the $\{F_k\}$ decay more slowly with increasing β and so for a truncated summation, there is a trade-off in the choice of N and β.

An approximate periodic CDF can then be expressed as

$$\tilde{C}(x) = \int_{-\pi/2}^{\pi/2} p(y) F(x-y) dy \tag{12}$$

$$= \sum_{|k| \leq N} F_k e^{ikx} \langle \psi | e^{-i\tau Hk} | \psi \rangle \tag{13}$$

$$= \sum_{|k| \leq N} F_k e^{ikx} g_k. \tag{14}$$

Noting that $g_{-k} = g_k^*$ and that $F_k = -i|F_k|$ for $k > 0$ and $F_k = i|F_k|$ for $k < 0$, this can be simplified to the following final expression:

$$\tilde{C}(x) = \frac{1}{2} + 2\sum_{k=1}^{N} |F_k| \left[\text{Re}[g_k] \sin(kx) + \text{Im}[g_k] \cos(kx) \right]. \tag{15}$$

In practice, one only needs to estimate g_k for $k \geq 1$ and only for odd values of k.

2. Importance sampling

The summation to be estimated in the CDF-based QPE approach takes the form given in Eq. (15), where each term is weighted by a Fourier coefficient, $|F_k|$. These Fourier coefficients decay rapidly, such that contributions at large k may be several orders of magnitude smaller than those at low k.

For this reason, Ref. [8] suggests performing importance sampling of this summation. Here, terms are randomly sampled with probabilities proportional to $|F_k|$,

$$P_k = \frac{|F_k|}{\mathcal{S}}, \tag{16}$$

where $\mathcal{S} = \sum_{k=1}^{N} |F_k|$. We obtain a set of N_S values $\{k_1, \ldots, k_{N_S}\}$, where each k_i is sampled with probability P_{k_i}. The importance-sampled CDF (which we denote by $\tilde{H}(x)$) can then be constructed as

$$\tilde{H}(x) = \frac{1}{2} + \frac{2\mathcal{S}}{N_S} \sum_{i=1}^{N_S} \left[\text{Re}[g_{k_i}] \sin(k_i x) + \text{Im}[g_{k_i}] \cos(k_i x) \right], \tag{17}$$

which is an unbiased estimator for $\tilde{C}(x)$. The estimates of $\text{Re}[g_{k_i}]$ and $\text{Im}[g_{k_i}]$ for each k_i are then each obtained by performing Hadamard tests as in Fig. 1, with the resulting real and imaginary components denoted by r_i and s_i, respectively. In our experiments, each r_i and s_i estimate will be averaged over multiple shots in practice, since on current cloud-based platforms it is inefficient to perform a circuit for a single shot, due to the overhead in submitting a circuit. The CDF can then be constructed by

$$\tilde{G}(x) = \frac{1}{2} + \frac{2\mathcal{S}}{N_S} \sum_{i=1}^{N_S} [r_i \sin(k_i x) + s_i \cos(k_i x)], \tag{18}$$

which again is an unbiased estimator for $\tilde{C}(x)$. Note that new estimates of r_i and s_i are taken for each sample k_i (rather than only obtaining single estimates for each *unique* k). Also note that we use the same set of samples $\{k_1, \ldots, k_{N_S}\}$, $\{r_1, \ldots, r_{N_S}\}$ and $\{s_1, \ldots, s_{N_S}\}$ for every value of x when constructing $\tilde{G}(x)$ (rather than performing a fresh sample for each x).

We emphasize that there are two separate levels of sampling here. We refer to $|\tilde{C}(x) - \tilde{H}(x)|$ as "importance-sampling error," whereas $|\tilde{C}(x) - \tilde{G}(x)|$ contains importance-sampling error and also "shot noise."

To aid with discussion later, it will be helpful to write Eq. (18) in an alternative form. Let n_k denote the number of times that k is sampled during importance sampling. We
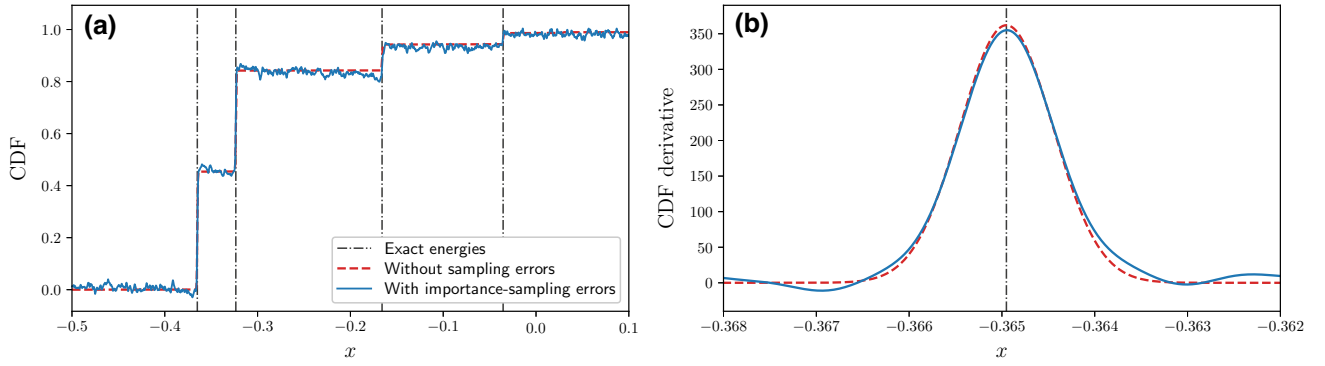
FIG. 2. (a) The CDF for H_4 STO-3G, taking the Hartree-Fock wave function as the initial state, calculated with and without sampling errors. In this example, the results are simulated. The parameters are $\beta = 10^6$ and $d = 5000$. The "Without sampling errors" CDF is obtained via Eq. (15), using exact values of g_k. The CDF "with importance-sampling errors" is obtained from Eq. (17) with $N_S = 5000$ but again using exact g_k values. The dashed vertical lines are exact energies. (b) The derivative of the CDF, enlarged in the region around the ground-state energy. The maximum of the CDF derivative provides an accurate energy estimate.

can then write

$$\tilde{G}(x) = \frac{1}{2} + \frac{2\mathcal{S}}{N_S} \sum_{k=1}^{N} n_k \left[\tilde{r}_k \sin(kx) + \tilde{s}_k \cos(kx) \right], \quad (19)$$

where

$$\tilde{r}_k = \frac{1}{n_k} \sum_{i:k_i=k} r_i, \quad \tilde{s}_k = \frac{1}{n_k} \sum_{i:k_i=k} s_i \quad (20)$$

are estimates of the real and imaginary parts of g_k, averaged over all repeated samples of k. In order to mitigate coherent errors when running on a QPU, we will perform a separate Pauli twirl for each sample of k. Therefore, $\tilde{r}_k$ ($\tilde{s}_k$) will denote the estimate of Re$[g_k]$ (Im$[g_k]$) averaged over n_k Pauli twirls of the appropriate Hadamard test (with each single-Pauli-twirl estimate, r_i or s_i also averaged over multiple shots in practice).

As an example, Fig. 2(a) presents simulated results for H_4 in a STO-3G basis, with a square geometry of side length 1.28 Å. The state $|\psi\rangle$ is taken to be the Hartree-Fock state. The CDF estimates used here are $\tilde{C}(x)$ (red) and $\tilde{H}(x)$ (blue) with $N_S = 5000$ (thus shot noise is not present in these examples). Because this problem is multireference, the CDF has multiple "jumps," each corresponding to an energy eigenvalue of τH. While importance sampling introduces noise into the CDF estimate, the first few eigenvalues can still be clearly identified.

Figure 3 plots the values $|F_k|$ against k for the CDF-QPE method and compares them to those from Somma's QEEA, which is discussed in Appendix B. In the QEEA, we take the half bin width as $\epsilon = 3 \times 10^{-3}$. In the CDF-QPE method, we set $\beta = 10^5$, which has been chosen to target an equivalent accuracy of around 3×10^{-3} in estimates of λ_i. The Fourier coefficients F_k decay rapidly in both methods, although the decay is much more rapid in

the CDF-QPE method. This rapid decay is the reason for the large efficiency gain in using importance sampling.

3. Estimating energies in the CDF-QPE method

Equation (15) provides a formula to construct the approximate CDF, which in the limit of large β and N can be used to obtain the exact energies $\tau\lambda_i$ of τH through its jump discontinuities. In practice, this function is only constructed to a finite precision and a method is needed to estimate λ_i from the approximate $\tilde{C}(x)$.

Reference [9] proves that $F(x)$ can be constructed with a guaranteed level of accuracy, provided that sufficient β and d are chosen. In particular, it is proven that for any $\epsilon > 0$
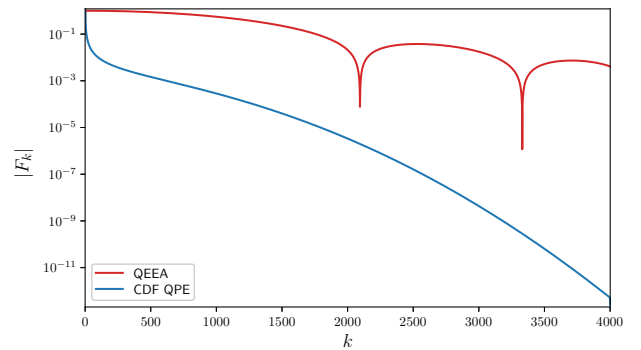
FIG. 3. A comparison of the Fourier coefficients from the two statistical phase estimation methods considered (rescaled so that $|F_1| = 1$), for similar target accuracies. In the QEEA [7], we choose a bin size of 3×10^{-3}. In the CDF-QPE method, we set $\beta = 10^5$, which gives at least a similar level of accuracy with high probability. The Fourier coefficients decay rapidly in both methods, which allows efficient importance sampling.

and $\delta \in (0, \pi/2)$, the condition

$$|\Theta(x) - F(x)| \le \epsilon \quad \forall \, x \in [-\pi + \delta, -\delta] \cup [\delta, \pi - \delta] \tag{21}$$

is satisfied, provided that

$$\beta = \max \left\{ \frac{1}{4 \sin^2 \delta} W \left(\frac{3}{\pi \epsilon^2} \right), 1 \right\} \tag{22}$$

and with a sufficient d, which can be chosen as $d = \mathcal{O}(\delta^{-1} \log(\epsilon^{-1}))$, and $W(\cdot)$ denotes the principal branch of the Lambert W function. Throughout this paper, we use Eq. (22) to choose β for a given target accuracy of δ, after loosely setting $\epsilon = 0.1$. Note that it is always possible to construct $F(x)$ via Eq. (10) to check the accuracy of $F(x)$ for a given β and d; we use this approach to choose d after first choosing β from Eq. (22). Given the accuracy guarantees above, the authors of Ref. [9] estimate $\tau \lambda_i$ with a procedure similar to a binary search, suggested in Ref. [8]. This approach allows a careful proof of the scaling of the algorithm for a given target accuracy δ.

In this paper, we take a different practical approach to estimating $\tau \lambda_i$ from $\tilde{C}(x)$, which we find can give more accurate estimates than the target accuracy δ by 1–2 orders of magnitude or more. This is performed by maximizing the derivative of $\tilde{C}(x)$ in the region of each jump. From Eq. (15), the derivative can be calculated (up to an unimportant constant) as

$$\tilde{C}'(x) = \sum_{k=1}^{N} |F_k| \, k \, [\mathrm{Re}[g_k] \cos(kx) - \mathrm{Im}[g_k] \sin(kx)]. \tag{23}$$

This can be viewed as an objective function and the locations of its local maxima (in the regions of jumps in $\tilde{C}(x)$) can be used to estimate each $\tau \lambda_i$. When performing importance sampling and averaging over shots, we instead take

$$\tilde{G}'(x) = \sum_{k=1}^{N} n_k \, k \, [\tilde{r}_k \cos(kx) - \tilde{s}_k \sin(kx)] \tag{24}$$

as an objective function, which provides an unbiased estimator of $\tilde{C}'(x)$, up to an unimportant overall constant. Here, n_k, $\tilde{r}_k$, and $\tilde{s}_k$ are as defined in Sec. II A 2. Note that although $C(x)$ is discontinuous in the exact case, for any finite d and β, the function $\tilde{C}'(x)$ will be well defined and so this does not lead to practical issues.

To motivate why the above provides accurate estimates of $\tau \lambda_i$, recall that $\tilde{C}(x)$ is defined as a convolution between the approximate Heaviside function, $F(x)$, and the probability density function, $p(y)$, as in Eq. (8). $F(x)$ is constructed to meet the accuracy condition in Eq. (21),

restricting the jump to a region of width approximately 2δ. However, from the Fourier definition of $F(x)$, it can be seen that the maximum of the derivative of $F(x)$ lies at exactly $x = 0$, even for small β and d. To see this, consider the derivative of $F(x)$,

$$F'(x) = \sum_{|k| \le N} S_k e^{ikx}, \tag{25}$$

where we have denoted $S_k = ikF_k$. The coefficients S_k obey $S_k = S_{-k} \ge 0$ for all k. Therefore, $F'(0) = 2 \sum_{k=1}^{N} S_k$. Since $S_k \ge 0$, this is the maximum value of $F'(x)$, as claimed. Now consider the simple case when $p_0 = 1$, so that the initial state $|\psi\rangle$ is an exact eigenstate of τH with energy $\tau \lambda_0$. In this case, $p(y) = \delta(y - \tau \lambda_0)$. Since $\tilde{C}(x)$ is a convolution between $F(x)$ and $p(y)$, the maximum of its derivative will then lie at *exactly* $\tau \lambda_0$, even for small β. In the more general case where multiple p_i values are nonzero, the CDF derivative will be a sum of such contributions that will overlap and this argument no longer holds exactly. However, if the gap between eigenvalues $\tau \lambda_i$ is much greater than δ, then it is expected to remain a significantly better approximation than the bound provided by Eq. (21).

These arguments will be affected by the presence of noise, including shot-noise and importance-sampling errors. Also note that the additional factors of k in $\tilde{G}'(x)$ will increase noise from high-k contributions, potentially making the derivative more susceptible to errors. We will show in our results, however, that this approach is often robust in practice.

Figure 2(b) plots the CDF derivative for the H$_4$ example described above, enlarged in the region of the ground-state energy. It can be seen that the maximum is an accurate estimate of $\tau \lambda_0$, even after applying importance sampling. By numerically maximizing this function, the estimate of $\tau \lambda_0$ is in error by only 2.5×10^{-7} Ha without importance sampling [i.e., using Eq. (15)], which increases to 9.2×10^{-6} Ha with importance sampling [using Eq. (17)]. This can be compared to the width of the jump region, which is approximately 10^{-3} Ha.

We note that a comparable approach has recently been suggested by Wang *et al.* [13]. In this, the method of Lin and Tong is used to find an approximate region where the ground-state energy is located. A more accurate estimate is then obtained by finding the maximum of $(f_\sigma * p)(x)$ in this region, where $f_\sigma(x)$ is a Gaussian filter kernel. We expect that our approach is comparable from a practical point of view, although it avoids working with a separate Gaussian kernel. Instead, we work with the derivative of $F(x)$ in place of $f_\sigma(x)$.

B. Variational-circuit compilation

Applying statistical phase estimation requires estimating $g_k = \langle \psi | e^{-i\tau H k} | \psi \rangle$ using the Hadamard test of Fig. 1.

This requires implementation of the controlled-$e^{-i\tau Hk}$ unitary for a range of k values. One commonly considered approach to approximately implement this unitary is through Trotter product formulas. On current NISQ devices, the circuit depth required to achieve this is far too high for nontrivial problems, particularly for *ab initio* Hamiltonians, where there are many terms in H, and for the high values of k required for good precision in QPE.

Instead, in this work we primarily use a variational-compilation technique that allows the action of each controlled-$e^{-i\tau Hk}$ operation to be compiled to a constant circuit depth, up to a negligible error. This technique has recently been used in other studies applying closely related circuits on superconducting processors [46,47]. More generally, the task of variationally optimizing the action of time-evolution operators on a given input state has been considered in Ref. [48] and generalized to approximate the whole unitary using tensor-network methods in Ref. [49]. The results in these papers suggest that variationally optimized circuits can be made significantly shorter and more accurate than Trotter product circuits. A related approach has also been applied to run quantum signal processing on a trapped-ion device, using parametrized quantum circuits that are again optimized variationally [50].

We consider a circuit ansatz as shown in Fig. 4, consisting of alternating one- and two-qubit layers. The one-qubit layers consist of U_3 gates, each of which allows an arbitrary one-qubit rotation. The two-qubit layers are constructed using CZ gates, entangling alternating pairs of qubits in each layer with a "brickwork" pattern. The U_3 gates are parametrized by Euler angles (θ, ϕ, and λ) and implemented in native gates for Rigetti's processors as

$$U_3(\theta, \phi, \lambda) = R_Z(\phi) R_X(-\pi/2) R_Z(\theta) R_X(\pi/2) R_Z(\lambda), \tag{26}$$

where rotation gates are defined by $R_Z(\theta) = e^{-i\theta Z/2}$ and $R_X(\theta) = e^{-i\theta X/2}$. The parameters in the U_3 gates can be variationally optimized such that the circuit ansatz closely matches the action of the desired unitary on a given input state. Specifically, following Refs. [47,48], we define the loss function

$$L(\mathbf{p}) = \| U|\Psi\rangle - \tilde{U}(\mathbf{p})|\Psi\rangle \|, \tag{27}$$

where the L_2 norm is used. Here, U is the target unitary, $\tilde{U}(\mathbf{p})$ is that of the ansatz circuit with parameters $\mathbf{p}$, and $|\Psi\rangle = |+\rangle \otimes |\psi\rangle$ is the input state to U in the circuit. We apply this variational-compilation procedure to the controlled-$e^{-i\tau Hk}$ operation. Other components in the circuit are constructed directly as native gates.

We perform the minimization of $L(\mathbf{p})$ classically. Constructing and optimizing this loss function requires constructing the action of U on $|\Psi\rangle$. As such, the current methodology is not scalable to systems beyond classical computation but is nonetheless valuable for near-term NISQ studies. Alternative loss functions based on the reduced density matrices can be used instead [46]; this compilation strategy could then be applied to subcircuits on fewer qubits than the total circuit, allowing the approach to scale to large numbers of total qubits. Alternatively, approaches based on tensor networks could be used [49]. We do not consider these alternative approaches here and instead work with Eq. (27). While the use of constant-depth circuits simplifies some aspects of the error mitigation, there are many aspects of the error-mitigation task that remain challenging and important to treat carefully, as we shall see. To investigate the additional challenges introduced by using Trotterization, we also study a Trotterized example on a QPU at the end of Sec. IV. Additionally, a Trotterized example is studied in Appendix C in the presence of both unitary and depolarizing errors.

The circuit optimization is implemented using the JAX library [51], which enables automatic differentiation of PYTHON functions. We use the BFGS algorithm implemented in the JAX library to perform the minimization of $L(\mathbf{p})$. We find BFGS to be far more robust than optimizations using stochastic gradient-descent methods for this task.

C. Error mitigation

1. Zero-noise extrapolation

In this paper, we apply ZNE [18–20,52] to mitigate errors in the expectation values of the Hadamard tests. ZNE is one of the most commonly studied error-mitigation methods in the literature. The core idea of ZNE is to execute the target circuit at varying error rates, denoted by λ, and extrapolate the results to obtain an estimate at a reduced error rate. Expectation values are estimated for
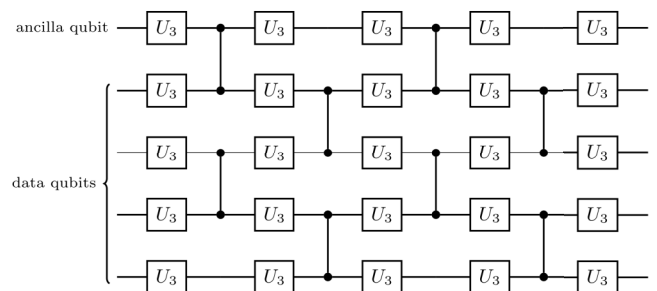

FIG. 4. The circuit ansatz used to compile controlled-$e^{-i\tau Hk}$ operations. The one-qubit layers consist of U_3 gates. Each U_3 gate is specified by three parameters that are optimized to approximately match the action of the desired unitary. Each U_3 gate is applied as five native gates on the quantum processor (see Eq. (26)). The two-qubit layers are formed from CZ gates.

the original circuit, defined by $\lambda = 1$, in addition to circuits at increased error rates $\lambda > 1$. A function is fitted to these expectation values and used to extrapolate to error rate $\lambda = 0$, which gives the error-mitigated estimate.

There are various possible methods to increase the error rate λ. Examples in the literature include parameter noise scaling and pulse stretching [20,52]. In our implementation of ZNE, we increase λ using identity insertion (sometimes referred to as "unitary folding") [53]. Identity insertion $n > 0$ times replaces a unitary operation U according to

$$U \rightarrow U(U^{\dagger}U)^n. \tag{28}$$

The error rate λ is then defined as

$$\lambda = 1 + 2n. \tag{29}$$

According to this definition, $\lambda = 1$ corresponds to $n = 0$, meaning that no identity insertion is performed.

The circuits in this study are performed in layers of one- or two-qubit gates that are executed in parallel (see Sec. II B and Fig. 4). In our ZNE implementation, we fold full two-qubit gate layers, which typically have higher error rates than one-qubit gates on superconducting devices. Folding layers of gates rather than individual gates helps to ensure a consistent error profile with folding, e.g., ensuring that the crosstalk will be consistent in each layer.

In our implementation of ZNE, we execute circuits at error rates $\lambda = 1$, 3, and 5. To obtain circuits at error rates $\lambda = 3$ and $\lambda = 5$, we apply identity insertion as in Eq. (28), with $n = 1$ and $n = 2$, respectively. A possible drawback of using such large error rates as 3 or 5 is that for low gate fidelities, the final error can be too large to perform a reliable extrapolation. However, in the examples studied in this paper, this is not found to be a significant problem. Furthermore, the use of only odd-integer λ values means that every two-qubit gate layer is folded, avoiding complications around having to pick a subset of layers to fold.

2. Mitigating coherent errors

As discussed in Sec. II C 1, we fold the two-qubit layers, which consist of CZ gates in this work. Such CZ gates typically suffer from significant coherent errors on current superconducting devices. To mitigate these coherent errors, we apply a form of RC [25,26]. This is achieved by applying random Pauli gates, uniformly sampled from $\{I, Z, X, Y\}$, to each qubit before a CZ layer. Because CZ gates are Clifford operators, it is always possible to then apply corresponding cancelling Paulis after the CZ layer [54,55]. This essentially has the effect of Pauli twirling the CZ gate layer [56,57]. This twirling process is usually averaged over multiple instances of the same circuit, with a new set of random Paulis applied in each. One can additionally average each twirl of the circuit over multiple

shots, as we do in this work (for details, see Sec. III B). After inserting the Pauli layers, each circuit contains subsequent layers of Pauli and U_3 gates, which can be merged into a single U_3 gate layer. This procedure is known to convert an arbitrary error channel into a Pauli error channel, thus eliminating coherent errors.

Our implementation of randomized compiling differs from previous descriptions due to the use of importance sampling. Remember that our goal is to estimate the CDF, $\tilde{C}(x)$, defined in Eq. (15). As discussed in Sec. II A 2, we importance sample this summation, as contributions at high k will typically be orders of magnitude smaller than at small k. We incorporate the twirling procedure into the importance sampling of $\tilde{C}(x)$. If n_k denotes the number of times that k is sampled during importance sampling, then the estimates for $\text{Re}[g_k]$ and $\text{Im}[g_k]$ are each averaged over n_k independent twirls of the corresponding circuits. These estimates are denoted by $\tilde{r}_k$ and $\tilde{s}_k$, as defined in Eq. (20). Therefore, estimates for $k = 1$ will typically be averaged over a large number of twirls, while many circuits for large values of k will be performed for just a single twirl (i.e., without averaging). This will lead to poorer results at large k (both larger coherent errors and larger statistical errors), which will also impact on the performance of ZNE. However, because these terms are weighted by n_k, their contribution will be small and thus it is to be expected that the corresponding errors will not significantly impact $\tilde{C}(x)$. The converse is also true; using this approach, errors at low k will be much smaller, which is beneficial for the final estimate of $\tilde{C}(x)$ due to their high weight in the summation.

The mitigation of coherent errors is known to improve the performance of ZNE, allowing a more reliable fitting of the expectation values with λ. This has been demonstrated in, e.g., Ref. [58], which has provided theoretical arguments to justify this finding, and will be further demonstrated in our results. Important improvements have also been demonstrated in the VQE algorithm [59]. Specifically, we find an exponential fit to be accurate in most cases. It should be pointed out, however, that the theoretical arguments for well-behaved exponential ZNE extrapolations in Ref. [58] only hold for depolarizing channels and, indeed, the authors show that this result does not hold in general for Pauli channels with nonequal Pauli weights. Despite this, we will see that ZNE performs well after randomized compiling. We mention that the noiseless output extrapolation (NOX) method [60] has been proposed to overcome the above potential shortcomings, although we do not consider this method here.

Given the above, our strategy for ZNE is to attempt an exponential fit for $\text{Re}[g_k]$ and $\text{Im}[g_k]$ at every value of k sampled. However, in some cases (particularly at large k where n_k is small), this fit may be unstable or of poor quality. Since we know that all $\text{Re}[g_k]$ and $\text{Im}[g_k]$ values must lie in the range $[-1, +1]$, we loosely check that the

extrapolated estimate is less 1.2 in magnitude. If this is not the case, then we instead switch to a quadratic fit for that data point. This strategy avoids extremely large g_k estimates due to unstable exponential fits.

3. Readout-error mitigation

We apply readout-error mitigation to our results using the symmetrized approach described in Ref. [17]. Other readout-mitigation strategies include methods based on assuming readout noise to be local or based on continuous-time Markov processes [16,61].

Suppose that we are mitigating the readout of n qubits. Define a calibration matrix A as

$$A_{ij} = P(\text{measure } |i\rangle \mid \text{prepared } |j\rangle), \qquad (30)$$

where $|i\rangle$ and $|j\rangle$ are computational basis states. Also define a vector C such that the ith element is equal to the number of times the n-qubit state is measured to state $|i\rangle$. Then let C_{ideal} be such a vector C under perfect readout. Applying matrix A to C_{ideal} gives an estimate of the results under noisy readout,

$$E[C_{\text{noisy}}] = A C_{\text{ideal}}, \qquad (31)$$

where $E[\cdot]$ denotes an expectation value. Therefore, an estimate of C_{ideal} can be obtained by inverting A,

$$C_{\text{ideal}} \approx A^{-1} C_{\text{noisy}}. \qquad (32)$$

To estimate the matrix A, we repeatedly prepare and measure each of the 2^n computational basis states. We use the measurement outcomes to estimate the probabilities as in Eq. (30) and thus the matrix A.

In practice, the readout error when measuring state $|1\rangle$ is often higher than the readout error when measuring state $|0\rangle$, so that the calibration matrix A will not be symmetric. Moreover, readout errors often drift quite rapidly, so that a given estimate of A may not be accurate throughout an experiment. We can nonetheless symmetrize the calibration matrix by bit-flip averaging [17], in which an X gate is applied directly before measurement for half of the shots performed (for every circuit involved in estimating each g_k). This symmetrizes any errors that remain after applying readout mitigation, resulting in a better-behaved final CDF estimate. Nonsymmetric readout errors will generally result in additive errors in the g_k estimates, which can show up as a spurious signal at $\lambda_i = 0$ in the Fourier transform. Symmetrizing readout errors in this way can therefore significantly improve the quality of the results. This simple approach of applying X before measurement in 50% of shots is found to be very effective in practice.

III. METHODS

A. Software

To generate phase-estimation circuits, including variational-circuit compilation, we use software developed by Riverlane. These circuits are then converted to pyQuil format and compiled to executables that are run on Rigetti's Aspen quantum processors, using their Quantum Cloud Services (QCS) platform [62] and associated software [63]. We also use pyQuil extensively to perform prior testing on quantum virtual machines (QVMs).

We use the ORCA program package [64] to perform embedding calculations and PySCF [65,66] to perform Hartree-Fock to generate the fermionic Hamiltonian for other systems. The Gaussian 16 program package [67] is used to perform geometry optimization for one structure detailed in Appendix A. The OpenFermion library [68] is used to perform mapping from fermionic to qubit operators using the Bravyi-Kitaev mapping [69], discussed further below. Variational-circuit compilation is performed using the JAX library [51].

B. Implementation details

Here, we discuss some specifics related to implementation of circuits for Rigetti's software stack and QPUs.

As discussed in Sec. II C 2, we perform a separate Pauli twirl of the circuit for each value k_i obtained during importance sampling. For example, if $k = 1$ is sampled $n_1 = 50$ times during the importance-sampling step, then the estimates for $\tilde{r}_1$ and $\tilde{s}_1$ will each be averaged over 50 independent twirls of the corresponding $k = 1$ circuits.

Ideally, we would like to perform as many Pauli twirls as possible and so would seek to perform an independent twirl for each shot. In practice, each separate twirl must be submitted to the QPU as a separate circuit. Because circuit loading takes significantly longer than circuit running, it is inefficient to perform only one shot per twirl. Instead, we perform 100 shots of each, with 50 each for the original and bit-flipped versions of the circuit in order to symmetrize readout errors. Therefore, when constructing the estimates $\tilde{r}_k$ and $\tilde{s}_k$ as in Eq. (20), it should be understood that the estimates r_i and s_i are each averaged over 100 shots in this manner, before averaging over n_k independent twirls to construct $\tilde{r}_k$ and $\tilde{s}_k$. These are the final estimates plotted for $\text{Re}[g_k]$ and $\text{Im}[g_k]$ in subsequent sections and also used in constructing the final CDF estimates via Eq. (19).

To improve the effectiveness of ZNE, it is important for each gate layer to be as consistent as possible, thus ensuring a similar noise profile for each. To help achieve this, we add FENCE statements around all two-qubit layers in the Quil circuit, which ensures identical pulse timings. Additionally, we decompose all one-qubit gates with the structure $R_Z(\phi) R_X(-\pi/2) R_Z(\theta) R_X(\pi/2) R_Z(\lambda)$. This
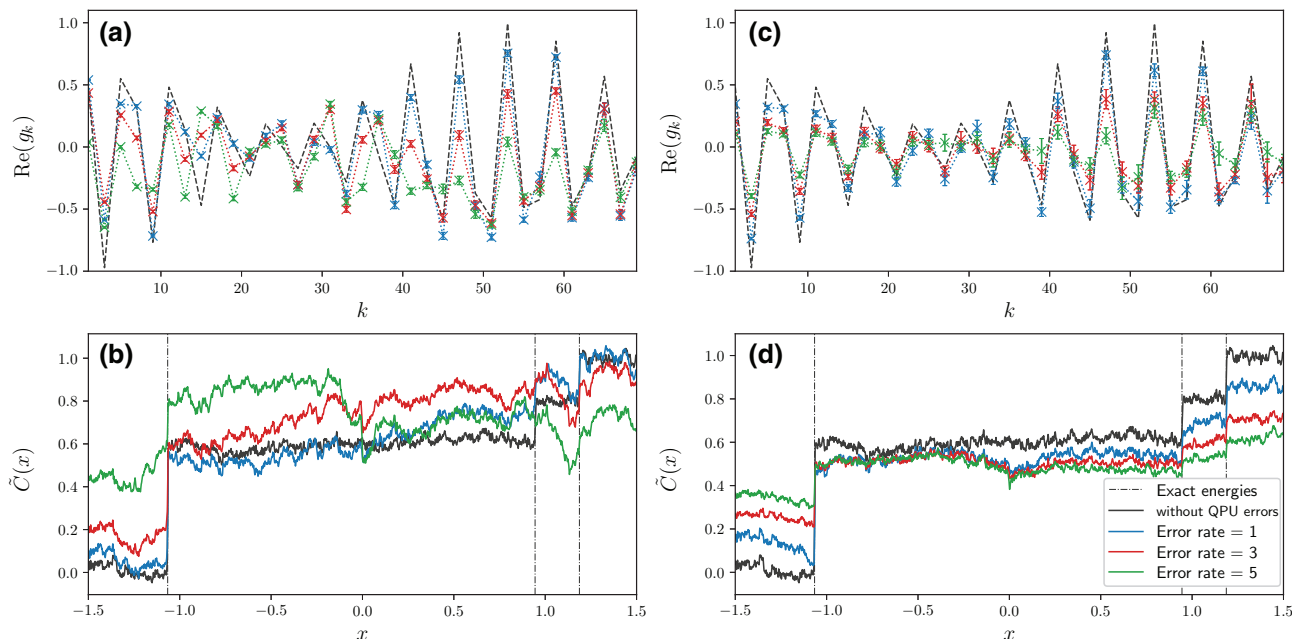
FIG. 5.   Results from the CDF-QPE method performed on the Aspen-11 QPU at three different error rates (ZNE-extrapolated results are not presented here), for methanethiol ($2e$, $2o$) with a stretched SH bond. (a),(b) Results performed without mitigation of coherent errors. (c),(d) Results with mitigation of coherent errors by RC. There are three significant eigenstates present, the exact eigenvalues of which are marked by dashed lines in the CDF. The mitigation of coherent errors leads to significantly better-behaved estimates of g_k with an increasing error rate, which leads to improved estimates of $\tilde{C}(x)$. The locations of jumps in the CDF match the exact eigenvalues closely, even at high error rates.

includes trivial Pauli gates such as I and Z. Our aim here is to again ensure that gate layers are as consistent as possible. It should be noted that R_Z gates are performed virtually on Rigetti's quantum processors [70].

C. Chemical systems

We study five separate systems in active spaces from two to four spatial orbitals. Here, the active space refers to a particular set of orbitals and a number of electrons used to occupy those orbitals. An active space with n electrons in m spatial orbitals ($2m$ spin orbitals) is denoted (ne, mo). Two of our example systems are motivated by pharmaceutical applications and use a recently developed embedding method to target a chemically relevant region of the molecule with a small active space [45]. The first of these systems is methanethiol, using a ($2e$, $2o$) active space for a minimal model of hydrogen abstraction. The second is a structure that we refer to as "clusterTS," taking a ($4e$, $4o$) active space. Both of these systems are described in Appendix A.

We also study H_3^+ and H_3^- in the STO-3G basis as example three-orbital systems. The geometry is an equilateral triangle in both cases, with a bond distance of 0.9 Å in H_3^+ and 1.75 Å in H_3^-. Lastly, we study H_2 to investigate a minimal example of Trotterization. Here, a STO-3G basis is used once again, with a stretched bond length of 2.0 Å.

The qubit Hamiltonian is generated using the Bravyi-Kitaev qubit mapping [69] for all systems. Specifically, we use the approach of Ref. [71], which allows two qubits to be tapered due to spin- and particle-number symmetries, and is implemented in OpenFermion [68]. Thus, for an active space of M spatial orbitals, the corresponding qubit Hamiltonian requires $2M - 2$ qubits. For the Trotterized H_2 example, we taper a further qubit, which is possible due to reflection symmetry. This allows the Hamiltonian for H_2 STO-3G to be represented by a single qubit.

The Hartree-Fock wave function is taken as the initial wave function, $|\psi\rangle$, for all systems. The PySCF input files used to generate fermionic integrals for H_2, H_3^+, H_3^-, and H_4, including the molecular geometries, are included in the additional data [72].

IV. RESULTS

A. Methanethiol ($2e$, $2o$)

As a first example, we consider application to methanethiol in a ($2e$, $2o$) active space, using orbitals centered on the sulfur-hydrogen (SH) bond, as described in Appendix A. To model the dissociation limit, we take the SH distance to be 4 Å. This is a minimal model but it results in a multireference problem. We focus first on demonstrating how the CDF is constructed from the QPU
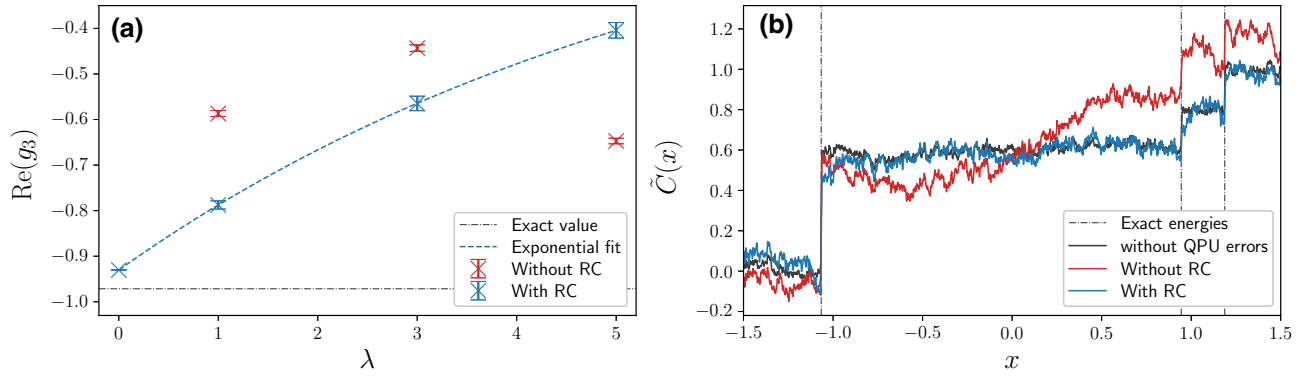
FIG. 6. The results for methanethiol (2e, 2o) with a stretched SH bond, using g_k estimates obtained from Aspen-11. (a) Estimates of Re[g_k] for $k = 3$ as an example, performed at error rates $\lambda = 1$, 3, and 5. An exponential fit is accurate after performing randomized compiling, leading to an improved ZNE estimate at $\lambda = 0$. Such a fit cannot be reliably performed for data obtained without mitigation of coherent errors. (b) CDFs constructed using ZNE-extrapolated g_k estimates.

calculations and, in particular, on the effect of incorporating RC into the importance-sampling procedure. In Sec. IV B, we then apply this approach to more challenging Hamiltonians with four and six qubits.

The qubit Hamiltonian for this system can be constructed using two qubits, so that three qubits are required for each Hadamard test. We variationally compile each controlled-$e^{-i\tau Hk}$ operation to a brickwork circuit ansatz. Here, the variational compilation can be performed with negligible errors using only three layers of CZ gates; the value of $L(\mathbf{p})$ is typically smaller than 10^{-6} after optimization. The total number of CZ gates is three, nine, and 15 for error rate $\lambda = 1$, 3, and 5 circuits, respectively.

The calculations are performed on Aspen-11 using qubits 11, 26, and 27, with qubit 11 taken as the ancilla. CDF Fourier parameters are taken as $\beta = 10^5$ and $d = 2 \times 10^3$. From Eq. (22) together with $\epsilon = 0.1$, this corresponds to an accuracy of $\delta \sim 0.003$. $N_S = 2 \times 10^3$ samples are taken for the importance sampling and 100 shots are performed per sample, thus the total number of shots is 2×10^5. Due to the very rapid decay of F_k, most samples are performed at small k. For example, 530 samples are performed at $k = 1$ and 153 samples are performed at $k = 3$, whereas only 28 samples are taken in total for all values $k > 1000$.

In all of the results, we compare with the CDF labeled "Without QPU errors," which is calculated numerically with importance sampling using exactly the same samples $\{k_i\}$ but in the absence of QPU errors or shot noise; this corresponds to $\tilde{H}(x)$ in Eq. (17). ZNE aims to correct gate errors by improving estimates of g_k but cannot correct importance-sampling noise, and so this is the fair comparison to make.

Figure 5 presents results for the real components of g_k (up to $k = 79$) and the corresponding CDF estimates, for error rates $\lambda = 1$, 3, and 5, performed both with and without randomized compiling. Without RC, the CDF at

$\lambda = 1$ can be used to correctly estimate the energies of τH through its jumps but this becomes challenging for the excited states at higher error rates and the general quality of the CDF is poor. This is improved significantly by applying RC. The energies are clearly identifiable at all error rates and the shape of the CDF is largely correct. By inspecting the values of Re[g_k], it can be seen that the expectation values decay with increasing error rate in a more systematic manner when RC is applied than without. Figure 6(a) emphasizes this behavior by showing an example ZNE extrapolation for the real component of g_k at $k = 3$. An exponential fit is seen to be accurate with RC applied, which leads to an improved estimate. This exponential decay is not observed when RC is not incorporated, making ZNE less effective.

Figure 6(b) shows the CDFs obtained from ZNE-extrapolated estimates of g_k. The same general behavior is observed; with RC applied, the ZNE-corrected CDF has roughly the correct amplitude compared with the exact result. In contrast, the ZNE-corrected CDF obtained without RC is of relatively poor quality and has large fluctuations (even beyond the expected importance-sampling noise) away from the jump regions. These results demonstrate the importance of mitigating coherent errors in statistical phase estimation experiments.

B. Four- and six-qubit Hamiltonians

We next apply these methods to larger systems, first to H_3^-, which requires five-qubit circuits, and then to the clusterTS system defined in Appendix A, which requires seven-qubit circuits. Given the improvements observed by applying RC in Sec. IV A, it is applied for all results in this section.

Figure 7 shows the results for H_3^-. Here, controlled-$e^{-i\tau Hk}$ unitaries are compiled with seven CZ layers for the $\lambda = 1$ circuits, with two CZ gates per layer. The CDF
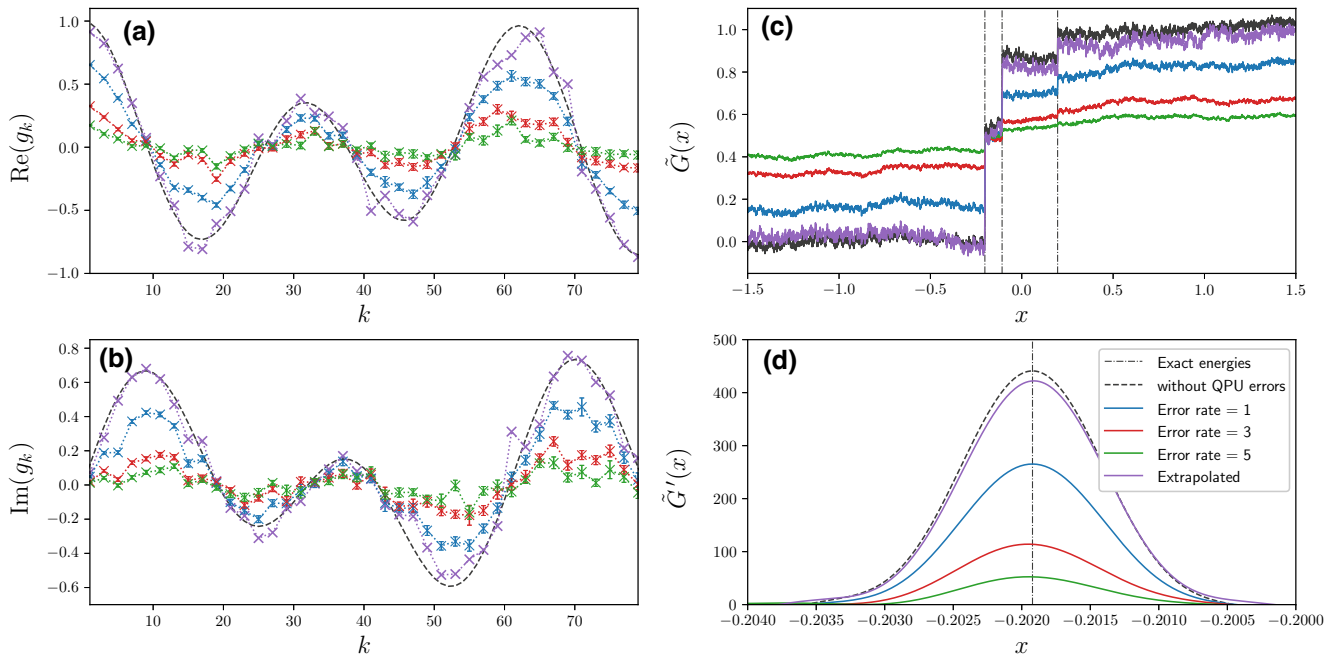
FIG. 7. Results from the CDF-QPE method for H_3^- in a STO-3G basis, performed on Aspen-M-3. Here, the ground-state wave function is multireference, leading to three jump regions in the CDF, each indicating an energy eigenvalue. (a),(b) Estimates of the real and imaginary parts of g_k for error rates 1, 3, and 5, and the subsequent ZNE-extrapolated estimates. Note that we set $d = 5000$ but only present results up to $k = 79$ for clarity. (c) The CDF itself and (d) its derivative, enlarged in the region of the ground-state energy. The derivative of the CDF can be used to obtain an extremely accurate estimate of each energy. Extrapolation improves the amplitude of the CDF, although the location of the jump is not affected.

Fourier parameters are $\beta = 10^6$ and $d = 5 \times 10^3$. $N_S = 4 \times 10^3$ samples are taken for importance sampling. The circuits are performed on Aspen-M-3 using qubits 30 and 34–37, with 34 taken as the ancilla. The corresponding CZ fidelities, as estimated by randomized benchmarking, are between 97.5% and 99.3%.

It can be seen that ZNE does a good job at correcting g_k estimates, particularly at low k, where more samples are taken. This system is multireference, with three eigenstates having a significant overlap with the initial Hartree–Fock state, leading to three jump regions visible in the CDF. The ground-state energy is clearly identifiable at all error rates, although there is no clear signal from excited states at error rate $\lambda = 5$. Also shown is the CDF derivative, $\tilde{G}'(x)$, enlarged in the region of the ground-state energy; the maximum of this objective function provides an extremely accurate estimate of the true energy at all error rates.

Figure 8 presents equivalent results for the clusterTS system, which has four orbitals in the active space. Here, we take CDF Fourier parameters $\beta = 10^6$ and $d = 5 \times 10^3$, and $N_S = 4.8 \times 10^3$ for importance sampling. We are able to obtain a good representation of each circuit using nine CZ layers for each controlled-$e^{-i\tau Hk}$ operation. Each layer contains three CZ gates. Thus, for $\lambda = 1$, each circuit contains 27 CZ gates. For the highest error rate, $\lambda = 5$, each circuit contains 135 CZ gates. The final CDF is constructed

by averaging over a large number of circuits; when considering both X and Y measurement bases, each ZNE error rate, the large number samples N_S (each corresponding to a separate Pauli twirl), and bit-flip averaging to symmetrize readout errors, 57 600 separate circuits are performed to construct the CDFs, with 50 shots of each. The circuits are again performed on Aspen-M-3, using qubits 30–32 and 34–37, with qubit 34 again taken as the ancilla. The corresponding CZ fidelities, as estimated by randomized benchmarking, are between 96.7% and 99.2%.

Figures 8(a) and 8(b) show the real and imaginary components of g_k at each error rate and the ZNE-extrapolated results. It is again observed that g_k estimates decay sensibly with λ within statistical errors and that the extrapolated estimates are a significant improvement for most k values. The final CDF estimates are shown in Fig. 8(c). Here, the ground-state wave function is single reference, leading to a single jump in the CDF, which is distinguishable at each error rate. The maximum of the CDF derivative in Fig. 8(d) again allows an accurate estimate of λ_0 to be obtained.

As one method of quantifying the improvement made to the CDF estimates by applying ZNE, we consider the distance metric

$$W = \int_{-\alpha}^{\alpha} |\tilde{C}_{\text{approx}}(x) - \tilde{C}_{\text{exact}}(x)| \, dx, \qquad (33)$$
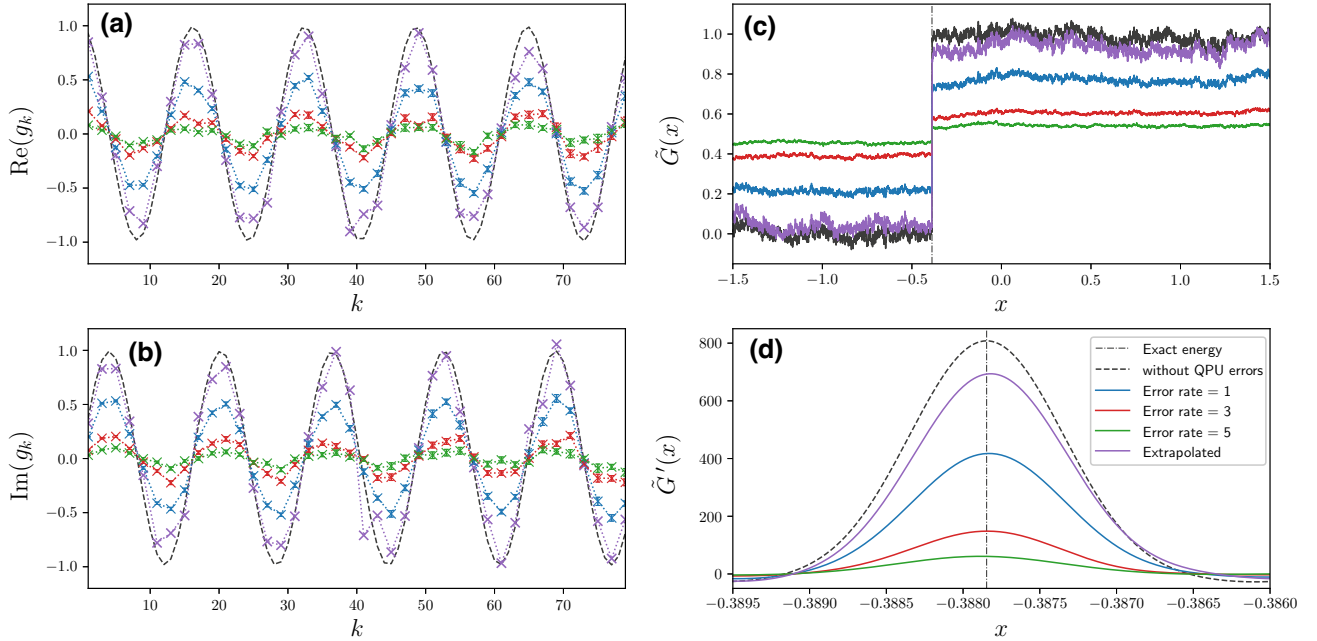
FIG. 8.   The results from the CDF-QPE method for the six-qubit "clusterTS" system, performed on Aspen-M-3. Here, the ground-state wave function is single reference, leading to a single jump region in the CDF. (a),(b) Estimates of the real and imaginary parts of g_k for error rates 1, 3, and 5, and the subsequent ZNE-extrapolated estimates. (c) The CDF itself and (d) the CDF derivative, enlarged in the region of the ground-state energy.

where $[-\alpha, \alpha]$ is the range in which the CDF is calculated, with $\alpha = 1.5$ here, and we calculate the integral numerically. An exact estimate of the CDF corresponds to $W = 0$. Table I gives the percentage reduction in W after performing ZNE ($\lambda = 0$) compared to the unmitigated estimates ($\lambda = 1$). The CDF is improved by around 77–80% for the larger systems studied.

This improvement demonstrates the potential of ZNE to mitigate errors in expectation values, and agrees with previous ZNE studies. However, we find that the benefit of performing ZNE is somewhat limited in statistical phase estimation. Ultimately, the energies of the Hamiltonian are estimated through the jumps in the CDF; thus,

two important metrics are the final energy estimates and the ability to distinguish these jumps from the sampling noise. Here, we find no improvement upon performing ZNE. Table I gives errors in ground-state energy estimates for each system, calculated as $\Delta\lambda_0 = \lambda_0^{estimate} - \lambda_0^{exact}$. No systematic improvement is observed by performing ZNE and in many cases the error is slightly increased (although statistical errors may account for this). Similarly, while ZNE increases the amplitude of the CDF, the importance-sampling noise is also inevitably amplified. This latter result may be expected; remember that ZNE only aims to address QPU errors, and not statistical sampling errors. Moreover, ZNE comes with a significant sampling overhead in order to estimate each g_k with sufficient precision at high λ, as required for a reliable extrapolation.

The lack of improvement by performing ZNE seems to be associated with the natural resilience of statistical phase estimation to noise, particularly after mitigating coherent errors and symmetrizing readout errors. Indeed, even in the presence of significant QPU errors, the ground-state energy errors are found to be extremely small, typically smaller than 0.1 mHa. Again, note that for all of the examples studied here, β is chosen according to Eq. (22) for a target accuracy of $\delta \sim 1$ mHa. Thus the final accuracy is often found to significantly exceed this target, even in the presence of noise. Therefore, while applying ZNE is found to give little practical benefit, we find that mitigating coherent errors by RC is very beneficial and can lead to an algorithm with natural noise tolerance. Since

TABLE I.   Metrics assessing the effect of ZNE. All results use RC and readout-error mitigation. The application of ZNE leads to a significant improvement in the distance metric W for all systems studied. The error in the ground-state energy estimate, $\Delta\lambda_0$, is extremely small both with and without ZNE applied; however, there is no improvement made within statistical errors by applying ZNE and the estimate is even worsened in some cases.

System	Improvement in W after ZNE (%)	$\Delta\lambda_0$ before ZNE (mHa)	$\Delta\lambda_0$ after ZNE (mHa)
Methanethiol	63.1	0.022	0.009
H_3^+	79.8	0.050	0.071
H_3^-	77.6	−0.002	0.019
ClusterTS	77.3	0.089	0.149

statistical phase estimation requires averaging over a large number of circuits, the required twirls can be incorporated at minimal cost compared to the bare method, unlike ZNE, which requires additional circuits to be performed, the signal of which decays exponentially with increasing λ. Furthermore, incorporating the twirls directly into the importance-sampling procedure is found to be practically effective.

C. Trotterization

The previous sections have investigated constructing the CDF for circuits the depth of which is independent of k. This requires mitigation of various errors that reduce the performance of the algorithm. However, for a scalable approach, we require circuits the length of which grows at least linearly with k. As a final example, we investigate a minimal model of H_2 using Trotterization and investigate the performance of the same error-mitigation techniques.

We consider H_2 in a STO-3G basis set, with a stretched internuclear distance of 2.0 Å. Using the Bravyi-Kitaev transformation with qubit tapering, this Hamiltonian can be represented by a single data qubit [71]. In particular, qubits are tapered due to particle number, spin, and reflection symmetries. This allows the qubit Hamiltonian to be written as

$$H = c_1 Z + c_2 X, \qquad (34)$$

where $c_1 = 0.121256$ and $c_2 = 0.259138$ and we have discarded a constant shift of -0.662537. The two eigenstates of H correspond to the lowest bonding and antibonding states of H_2. We choose $\tau = 1.5/(c_1 + c_2) = 3.943$ and work with the scaled Hamiltonian τH. We then perform first-order Trotterization with a single Trotter step, i.e., taking $e^{-i\tau Hk} \sim (e^{-i\tau c_2 X} e^{-i\tau c_1 Z})^k$. The exact energies of τH before and after Trotterization are ± 1.128189 Ha and ± 1.089119 Ha, respectively. Thus there is a Trotter error of 39 mHa, or 10 mHa after rescaling by τ^{-1}. We are not concerned with Trotter error here but, rather, with the performance of the statistical phase estimation method and error mitigation and thus we only compare to the Trotterized energies from now on. Note that a similar H_2 Hamiltonian has been used in a study of textbook QPE on a neutral-atom quantum computer, performed to three bits of precision [39].

The circuit for a single Trotter step is given in Appendix C and has a CZ depth of 4. Therefore, circuits to estimate g_k have a CZ depth of $4k$. The results are performed on Aspen-M-2 using qubits 121 and 122. The estimated CZ-gate fidelity from randomized benchmarking is $99.22 \pm 0.1662\%$. The ancilla is taken as qubit 121, with an estimated readout fidelity of 98.4%.

The simulation parameters are chosen as $\beta = 50$, $d = 30$, and $N_S = 500$. These are considerably lower than those used in previous sections due to the need to limit the
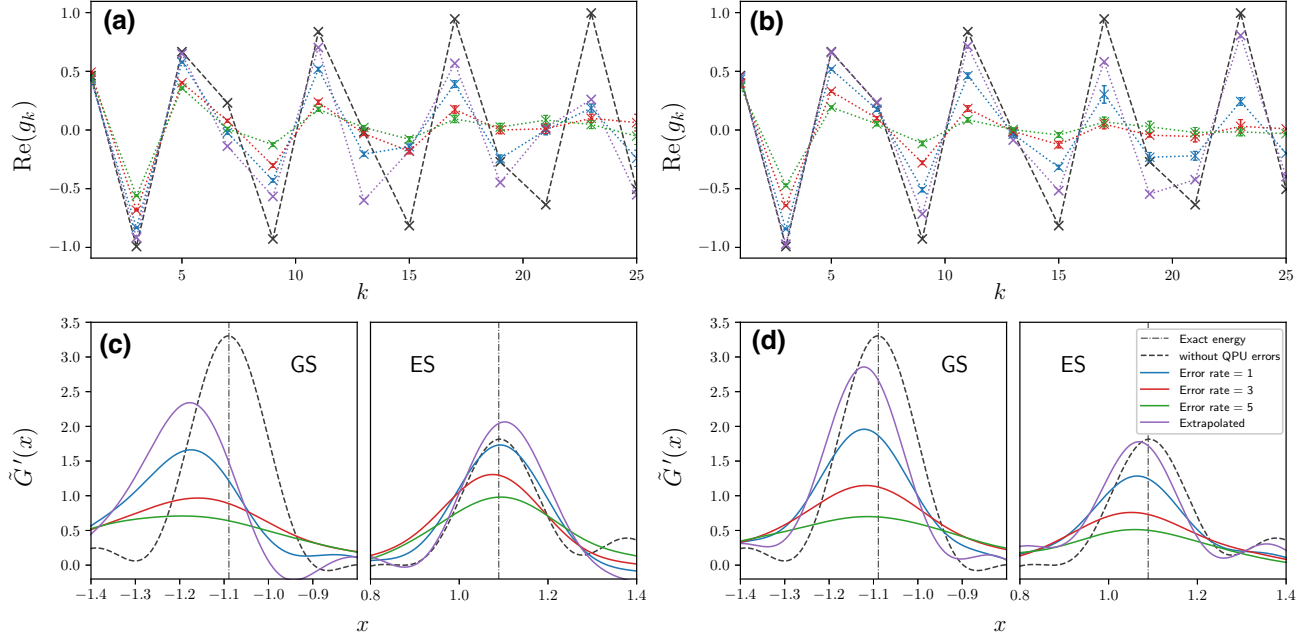

FIG. 9. The results for a stretched H_2 molecule, using first-order Trotterization to construct $e^{-i\tau Hk}$, performed on the Aspen-M-2 QPU. (a),(b) Estimates of the real part of g_k, estimated (a) without and (b) with randomized compiling. The dashed lines are added between estimates for clarity. (c),(d) The derivative of the CDF constructed from g_k, (c) without and (d) with randomized compiling. The subplots labeled "GS" and "ES" are enlarged in the region of the ground state and excited state, respectively. The results are performed at error rates 1, 3, and 5 and extrapolated. RC leads to better final energy estimates and a more accurate relative signal between the two states. ZNE improves the signal but does not improve the energy estimates.

circuit depth. Taking $\epsilon = 0.1$, this corresponds to $\delta \approx 0.13$ in Eq. (22); thus we expect a low resolution in the CDF. The highest value of k sampled is $k = 25$. The highest CZ depth is therefore 100 with $\lambda = 1$ or 500 with $\lambda = 5$.

The results are presented in Fig. 9, performed with (left) and without (right) RC applied: they display behavior similar to that observed in Sec. IV A. In particular, the decay of g_k with increasing λ is better behaved for the twirled results, leading to more accurate extrapolations for ZNE. A better energy estimate is also obtained (as determined by maximizing the CDF derivative) with RC applied than without. For the ground-state energy, the error in the ZNE-extrapolated estimate is reduced from -22 mHa to -8 mHa after applying RC. These errors remain large in both cases but this is expected due to the low precision used. As observed in Sec. IV B, we again find that ZNE improves the amplitude of the CDF toward the exact value but does not lead to improvements in the energy estimates.

We note that a related result has been observed in simulations performed in Ref. [73], where the authors consider a control-free variant of phase estimation. They show that coherent noise causes errors in the phases of the unitary and prove that these errors are removed to first order with RC, verified with simulations. Such a result can be similarly motivated in statistical phase estimation. For $g_k = \sum_i p_i e^{-i\tau\lambda_i k}$, the Fourier transform gives a sum of delta functions, shifted by the energies, λ_i. In the ideal case of a depolarizing noise model, we expect g_k to decay exponentially with k relative to the exact result, i.e., $g_k \sim e^{-\gamma|k|} \sum_i p_i e^{-i\tau\lambda_i k}$, for some decay rate γ. The energies λ_i can still be extracted exactly by taking a Fourier transform of this g_k; the corresponding poles will be broadened but the maxima of the poles, and therefore the energy estimates, are unchanged. This also motivates why ZNE should not be expected to improve energy estimates further. In Appendix C, simulated results are performed for the same system but with higher precision, applying depolarizing noise in one example and unitary errors in another. The results are found to confirm these ideas; increasing the depolarizing error rate broadens each peak in the CDF derivative but does not affect the energy estimates, beyond statistical noise. Under unitary errors, there is a significant error in the ground-state energy estimate, which is largely removed by incorporating RC into the importance-sampling procedure, at the cost of broadened signals in the CDF. Lastly, we note that Ref. [74] has also analyzed a statistical version of phase estimation and given theoretical arguments to justify the approach having tolerance to noise under certain models.

V. CONCLUSIONS

In this study, we have implemented statistical phase estimation techniques on Rigetti's quantum processors, in combination with error-mitigation and chemical-embedding methods, allowing accurate energy estimation for several small chemical problems. In addition, a variational-compilation technique has been used to reduce the circuit depth. We find this combination of techniques to be robust in practice, allowing accurate estimation of the ground-state energy with high confidence, even in the presence of significant QPU errors. The variational-compilation technique is also found to be robust and should be seen as a valuable tool for near-term NISQ studies. In the longer term, there is an interesting possibility of using this technique to optimize repeated sub-blocks within larger circuits.

We have demonstrated that the CDF-based approach of Lin and Tong [8], using the optimized Fourier approximation of Wan *et al.* [9], can be used to give significantly better energy estimates in practice than suggested by previous bounds. In particular, the derivative of the estimated CDF can be viewed as an objective function, the maximum of which gives accurate energy estimates. This estimate can be orders of magnitude more accurate than the rigorous bound derived in Ref. [9], even in the presence of both QPU and importance-sampling errors. This improved accuracy will permit the use of shorter circuits, allowing useful applications of phase estimation to be performed sooner. In the future, it would be interesting to derive a theoretical explanation of our observed improved accuracy.

We have shown that error mitigation and noise tailoring are important for improving the quality of the estimated CDF. In particular, there is a significant benefit in mitigating coherent errors. The twirling procedure in randomized compiling is incorporated into the importance sampling of the CDF, thus performing a much larger number of twirls at small k than at large k. Indeed, many circuits at large k are performed for only a single twirl in this approach. Despite this, we find the approach to be robust, leading to noise resilience in the statistical phase estimation results. We do not find further improvements in the energy estimates after applying ZNE; although ZNE does improve the signal of the CDF in each case, the energy estimates themselves and the signal-to-noise ratio are not systematically improved. Given the very high additional sampling cost associated with performing ZNE, overall we find it preferable to not use ZNE for this application. Lastly, we mention that an alternative error-mitigation approach has been suggested recently, which involves postselecting the data qubits of the Hadamard test circuits to be measured in the starting state [29]; we have not tested this approach but believe that it could be combined with the methodology developed here to improve the results further.

The results presented suggest the possibility of performing large-scale phase-estimation experiments in a manner that allows noise resilience and also that the required circuit depth for a given accuracy can often be made much lower than in traditional QPE approaches. The use of

shorter circuits and accurate results in the presence of noise will be crucial for successful applications of phase estimation on early fault-tolerant devices. Taken together with other recent results [13,14], we believe that this shows significant promise for statistical phase estimation techniques, emphasizing their potential as valuable techniques for chemistry and materials problems and motivating further development work in this direction. As we enter the regime of early fault-tolerant quantum computing, it will be important to explore whether our observed results hold. This suggests further work to repeat our experiments on a real or simulated early fault-tolerant quantum computer.

APPENDIX A: DETAILS OF CHEMICAL SYSTEMS

Here, we provide details of the two systems studied in this paper that have been motivated by pharmaceutical applications.

Methanethiol is a model system for the amino acid cysteine, a naturally occurring amino acid in proteins. Hydrogen abstraction from the thiol group of cysteine is an essential step in many enzymatic and drug-binding processes [75]. Methanethiol is the smallest neutral self-contained system that features a carbon-connected thiol group. The orbitals in the (2*e*, 2*o*) active space selected are located on the SH bond, allowing a minimal model of hydrogen abstraction.

As a larger example, we have studied a cut-out of ibrutinib in its binding pocket, a drug approved for treatment of non-Hodgkin lymphoma [76] that binds covalently to the cysteine of Bruton's tyrosine kinase (BTK). Corresponding to the hydrogen abstraction in the model system, we have considered a transition-state structure for the hydrogen transfer from thiol to an adjacent water molecule, which we refer to as "clusterTS." There are the two electrons of the thiol bond and the forming bond between the transferred hydrogen and water, which need to be included in a (4*e*, 4*o*) active space. This pharmaceutically relevant system has been used in a previous study of quantum algorithm resource estimation by some of the current authors [5] and we refer to a more detailed description of this system in that work. The geometry optimization of

methanethiol was carried out with the Gaussian 16 program package [67] at the MP2 level of theory [77] and using the aug-cc-pVQZ orbital basis set [78–80] level of theory. The SH distance has then been fixed to 4 Å. The transition-state structure for the drug-protein cut-out has been taken from one of our previous studies [81].

APPENDIX B: SOMMA'S QUANTUM EIGENVALUE ESTIMATION ALGORITHM (QEEA)

Some alternative statistical phase estimation methods primarily differ in their classical analysis. Therefore, once a set of values $\{g_k\}$ have been estimated, it is relatively easy to test alternative statistical phase estimation approaches to estimate the eigenvalues $\{\lambda_i\}$.

In addition to the CDF-QPE approach discussed in the main text, we have also implemented the quantum eigenvalue estimation algorithm (QEEA) of Somma [7] and tested it with estimates of $\{g_k\}$ from Rigetti's QPUs. Although the underlying approach to estimate $\{\lambda_i\}$ is quite different, it will be seen that the final objective function takes an identical form, primarily differing in the Fourier coefficients of the target function. We also extend the QEEA method by implementing importance sampling and briefly demonstrate its performance here.

1. Theory

In the QEEA, a range $[-\alpha, \ \alpha]$ is considered in which the Hamiltonian is known to be bounded and where $\alpha \leq \pi$. $\alpha = 1/2$ is chosen in the original presentation. The range $[-\alpha, \ \alpha]$ is divided into M bins, each with width 2ϵ for a small $\epsilon > 0$. These bins are constructed to overlap with each other, such that the center of the jth bin is given by $-\alpha + j\epsilon$. The use of bins that overlap helps to ensure appropriate normalization of results; we will return to this point shortly.

Consider a state $|\psi\rangle$, which will be chosen to have a large overlap with a state (or set of states) the energy of which we seek to estimate. For example, for a chemical system where we wish to estimate the ground-state energy, $|\psi\rangle$ might be taken as the Hartree-Fock state. For each bin, the QEEA aims to assess whether corresponding eigenstates (the eigenvalues of which lie within the bin) have a significant overlap with $|\psi\rangle$. By making ϵ sufficiently small, we can then obtain accurate estimates of the desired λ_i.

More precisely, for each bin, we define a function f_j that acts as a window function for that bin. This means that $f_j(\lambda_i)$ will be nonzero only if λ_i is within the jth bin. For an eigenstate $|\Psi_i\rangle$ of τH with eigenvalue $\tau \lambda_i$,

$$f_j(\tau H)|\Psi_i\rangle = f_j(\tau \lambda_i)|\Psi_i\rangle. \tag{B1}$$

The goal of the QEEA is then to construct the vector

$$p_j = \langle \psi | f_j(\tau H) | \psi \rangle. \tag{B2}$$

This is referred to as the *probability vector*. Only bins containing eigenstates supported by $|\psi\rangle$ will have nonzero values p_j and the magnitudes p_j will be related to the corresponding overlaps, $\langle \Psi_i | \psi \rangle$. Thus, by assessing the bins with a large p_j, one can estimate the desired eigenvalues of τH.

We can estimate p_j from a set of g_k estimates by expanding each f_j as a Fourier series,

$$f_j(\tau H) = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} F_j(k) e^{i\tau H k}. \tag{B3}$$

Inserting this expansion into Eq. (B2) and truncating such that $|k| < N$, for some $N \in \mathbb{N}$, gives

$$\tilde{p}_j = \frac{1}{\sqrt{2\pi}} \sum_{|k|<N} F_j(k) g_k^*. \tag{B4}$$

Here, $\tilde{p}_j$ denotes the approximate probability vector, with an error introduced due to truncation.

It only remains to choose the precise form of $f_j(x)$. There are many choices that could be made but because the Fourier series is truncated, it is important that the Fourier coefficients $F_j(k)$ decay rapidly. Somma chooses

$$f_j(x) = \int_{-\infty}^{\infty} h_\epsilon(y-x) 1_j(y) \, dy, \tag{B5}$$

where 1_j is the indicator function for the jth bin, and h_ϵ is a rescaled bump function. Specifically,

$$h(x) = \begin{cases} a\, e^{-1/(1-x^2)}, & \text{if } |x| < 1 \\ 0, & \text{if } |x| \geq 1 \end{cases} \tag{B6}$$

with a such that $\int_{-1}^{1} h(x) dx = 1$ and then

$$h_\epsilon(x) = \frac{2}{\epsilon} h(2x/\epsilon). \tag{B7}$$

The key properties for this choice are that the Fourier coefficients $F_j(k)$ decay superpolynomially and that $\sum_{j=1}^{M} f_j(x) = 1$ for $x \in [-\alpha, \alpha]$. This second point is important to ensure that contributions are appropriately normalized; this is why bins are chosen to overlap, preventing potential issues if an eigenvalue lies near the edge of a bin.

As stated in Ref. [7, Appendix A], the Fourier coefficients can be derived as

$$F_j(k) = H(k\epsilon/2) e^{-i\lambda_j k} \frac{\sin(k\epsilon/2)}{k}, \tag{B8}$$

$$\equiv F_k e^{-i\lambda_j k}, \tag{B9}$$

where λ_j is the center of the jth bin and $H(k)$ is the Fourier transform of $h(x)$, which we calculate numerically through a fast Fourier transform after discretization. The coefficients $F_j(k)$ only depend on the bin index j through a phase factor, which allows us to write the second line and thus define F_k independent of j. We can therefore rewrite Eq. (B4) as

$$\tilde{p}_j = \frac{1}{\sqrt{2\pi}} \sum_{|k|<N} F_k e^{-i\lambda_j k} g_k^*. \tag{B10}$$

Note that this expression has an almost identical form to that in Eq. (14) in the CDF-QPE method. The main difference is in the Fourier coefficients used. These are plotted for $\epsilon = 3 \times 10^{-3}$ in Fig. 3 and compared to those in the method of Wan *et al.*, aiming for a similar final accuracy by using Eq. (22) to select β. The Fourier coefficients decay less rapidly in the QEEA. However, as pointed out by Somma [7], there are advantages to this binning approach; in particular, if the gaps between eigenvalues are very small (e.g., in a solid-state system with bands of energy values), then solving the QEEA is much less ambitious than distinguishing individual eigenvalues to very high precision and so there may be advantages in such cases.

Lastly, noting that $g_{-k} = g_k^*$ and $F_{-k} = F_k$ allows us to write

$$\tilde{p}_j = \frac{\epsilon}{2\pi} + \sqrt{\frac{2}{\pi}} \sum_{k=1}^{N-1} F_k \big[\text{Re}[g_k] \cos(k\lambda_j)$$
$$- \text{Im}[g_k] \sin(k\lambda_j) \big]. \tag{B11}$$

2. Importance sampling

We have additionally implemented importance sampling in the QEEA, which was not considered in the original presentation of the method. As pointed out in Ref. [8], this can be used to reduce the complexity of the QEEA, bringing it closer to that of other statistical phase estimation approaches, some of which have proven Heisenberg-limited scaling [8,11]. We do not perform a study of scaling here. However, importance sampling allows the use of a much smaller bin width (and therefore higher precision) for a given number of circuits to perform.

The probability vectors in the QEEA can be constructed using importance sampling in an identical manner to that described in Sec. II A 2, starting from Eq. (B11). The main difference is that the sign of F_k can vary. Therefore, these
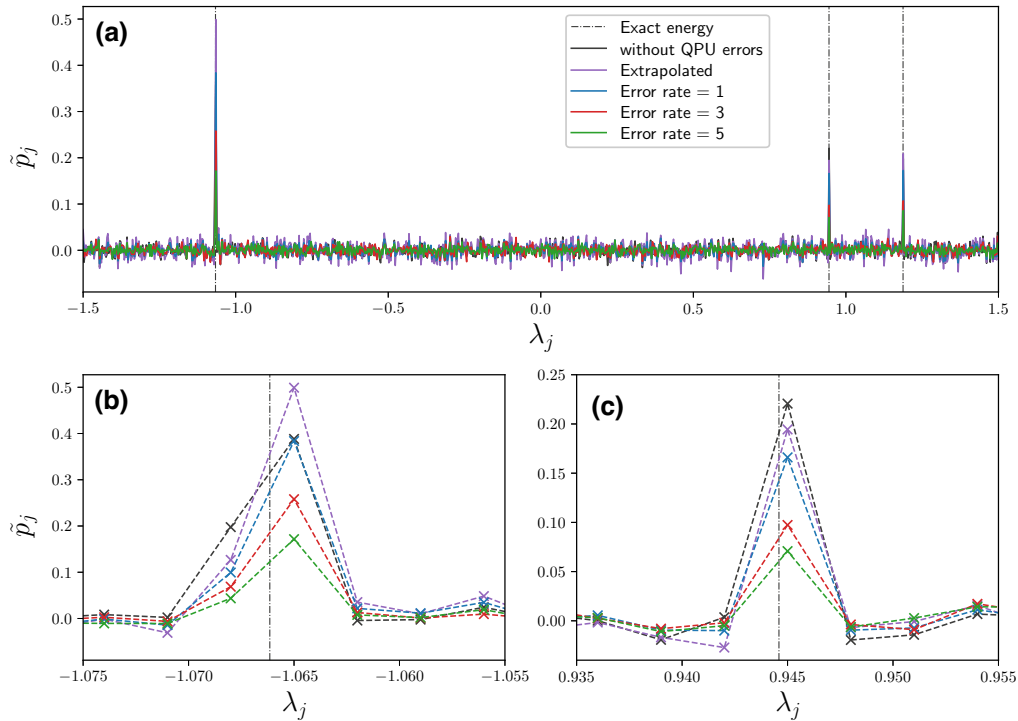
FIG. 10. Results from Aspen-11, performing the QEEA on methanethiol $(2e, 2o)$ with a stretched SH bond. (a) The full probability vector from -1.5 to $+1.5$. (b),(c) Plots enlarged in the region of the ground-state and first-excited state, respectively. The results presented are for error rates $\lambda = 1$, 3, and 5 and the ZNE-extrapolated $(\lambda = 0)$ result. The half bin width is set to $\epsilon = 3 \times 10^{-3}$ and we set $N = 4001$. For importance sampling, $N_S = 2000$ samples are taken. The correct energy is estimated for every error rate, to the precision considered.

signs must also be absorbed into the importance-sampled summation but the approach is otherwise unchanged. As described in the main text, we have also performed RC in the following results, incorporating it into the importance-sampling procedure by performing one twirl for each sample.

3. Results

We have performed the QEEA on Rigetti's Aspen-11 for the same methanethiol system studied in Sec. IV A, which requires three qubits in total. Each controlled-$e^{-i\tau H k}$ operation has been compiled to a circuit ansatz with three CZ layers, as in Sec. IV A. The half bin width has been

set to $\epsilon = 3 \times 10^{-3}$ and the corresponding Fourier coefficients have been importance sampled with $N_S = 2000$. A separate Pauli twirl has been performed for each sample. The Fourier summation has been truncated at $N = 4001$.

The results are presented in Fig. 10 and can be compared to equivalent results from the CDF-QPE method in Fig. 5. As for the CDF-QPE method, we find the QEEA to be robust and that each eigenvalue can be clearly and correctly identified from the probability vector, within the accuracy determined by roughly half the bin width.

As found in Sec. IV B, there is no improvement to the energy estimates after performing ZNE. In this case, the estimates are already correct within the resolution determined by ϵ and so there is no improvement to be made.
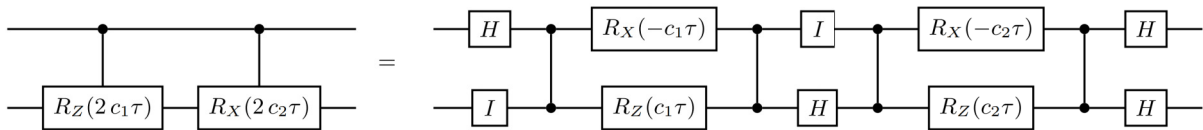

FIG. 11. The circuit for a single Trotter step for H_2 in a STO-3G basis, where the Hamiltonian has the form $H = c_1 Z + c_2 X$. The circuit on the right is reduced so that the only two-qubit gates are CZ gates, using standard identities. The one-qubit gates are each implemented in native Rigetti gates via the structure $R_Z(\phi) R_X(-\pi/2) R_Z(\theta) R_X(\pi/2) R_Z(\lambda)$, which ensures consistent gate layers.
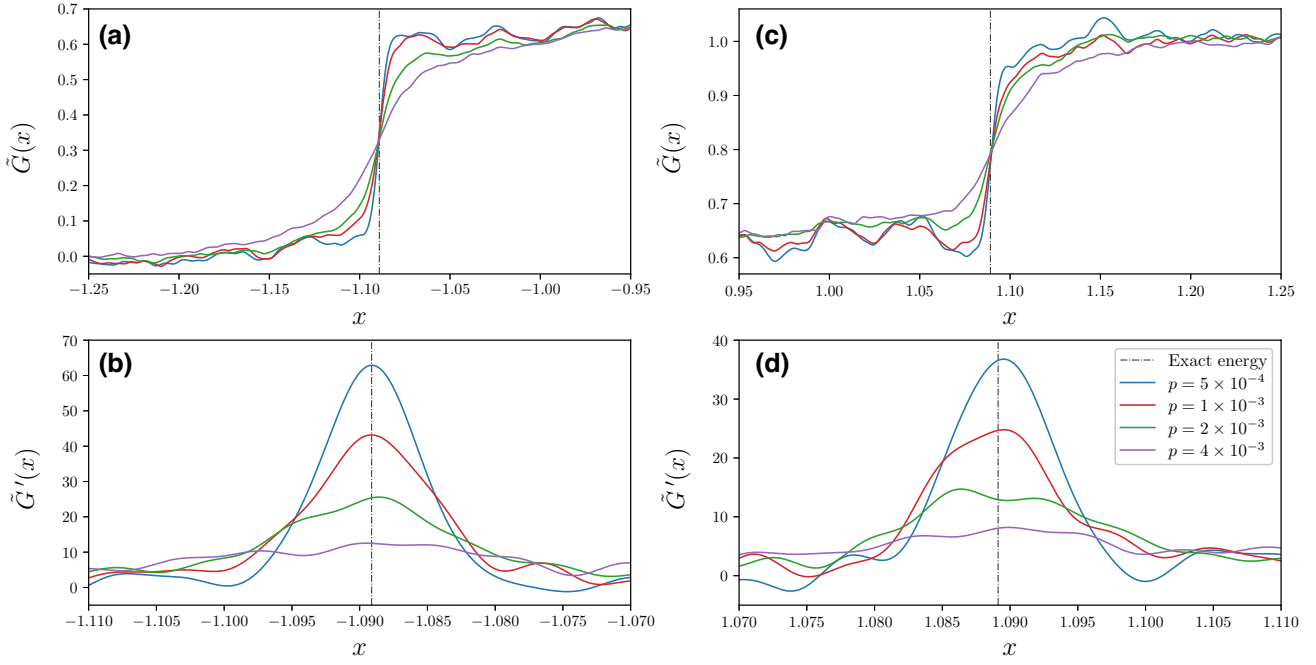
FIG. 12. CDFs and their derivatives for stretched H_2 STO-3G, performed with Trotterization using the Trotter step in Fig. 11. The simulations are performed on a pyQuil quantum virtual machine (QVM). Depolarizing noise is applied to CZ gates with four different error rates, p. The subplots are enlarged in the region of (a),(b) the ground state and (c),(d) the excited state, respectively. Increasing the error rate broadens the "jump" region of the CDF and the corresponding peak of the derivative. The peak of the CDF remains roughly correct regardless, although statistical noise at higher p can lead to errors in the final energy estimate. Note that different x-axis scales are used between subplots.

ZNE boosts the signal from the probability vector but overshoots considerably for the ground-state bin. One reason for this is that because the coefficients $|F_k|$ decay much more slowly in the QEEA, very few samples are performed for any particular k, even at small k. This makes mitigation of coherent errors less successful and also increases the statistical error bars on each g_k estimate, thus lowering the quality of each extrapolation and therefore also the ZNE estimate of the probability vector.

APPENDIX C: SIMULATED TROTTERIZATION RESULTS WITH COHERENT AND INCOHERENT ERRORS

Figure 11 presents the circuit used for each Trotter step in Sec. IV C. Since the Hamiltonian has the form $H = c_1 Z + c_2 X$, a single first-order Trotter step is taken as

$$U_{\text{Trotter}} = e^{-ic_2\tau X} \, e^{-ic_1\tau Z} \tag{C1}$$

$$= R_X(2 \, c_2 \tau) \, R_Z(2 \, c_1 \tau), \tag{C2}$$

which leads to the circuit on the left. The circuit on the right is then expressed with CZ as the only two-qubit gate, which can be obtained through standard circuit identities.

In addition to the results in the main text, we have performed simulated results using the pyQuil QVM. This

allows us to investigate higher precision and the effect of varying error rates. The same H_2 example is considered with an identical Trotter step. However, CDF-QPE parameters of $\beta = 5 \times 10^4$, $d = 511$, and $N_S = 1000$ are taken. This choice of β corresponds to $\delta \approx 0.004$ in Eq. (22), after choosing $\epsilon = 0.1$. As for the results in the main text, we perform 100 shots for each k_i value obtained during importance sampling.

First, we consider applying depolarizing noise to each CZ gate, before next considering the effect of coherent errors. The depolarizing channel is defined as

$$\Delta(\rho) = (1-p)\rho + \frac{p}{2^n}\mathbb{1}, \tag{C3}$$

where n is the number of qubits, equal to 2 when applied to a CZ gate, and p is the depolarizing error parameter. In the following results, we vary p from 5×10^{-4} to 4×10^{-3}. All other gates and measurements are applied without error. Figure 12 presents the CDF, $\tilde{G}(x)$, and its derivative, $\tilde{G}'(x)$. The subplots are enlarged in the region of the ground state (GS) and excited state (ES). As might be expected, the jumps in the CDF become broader as the error rate is increased. Despite this, the energy estimate obtained by maximizing the CDF derivative remains accurate in each case.
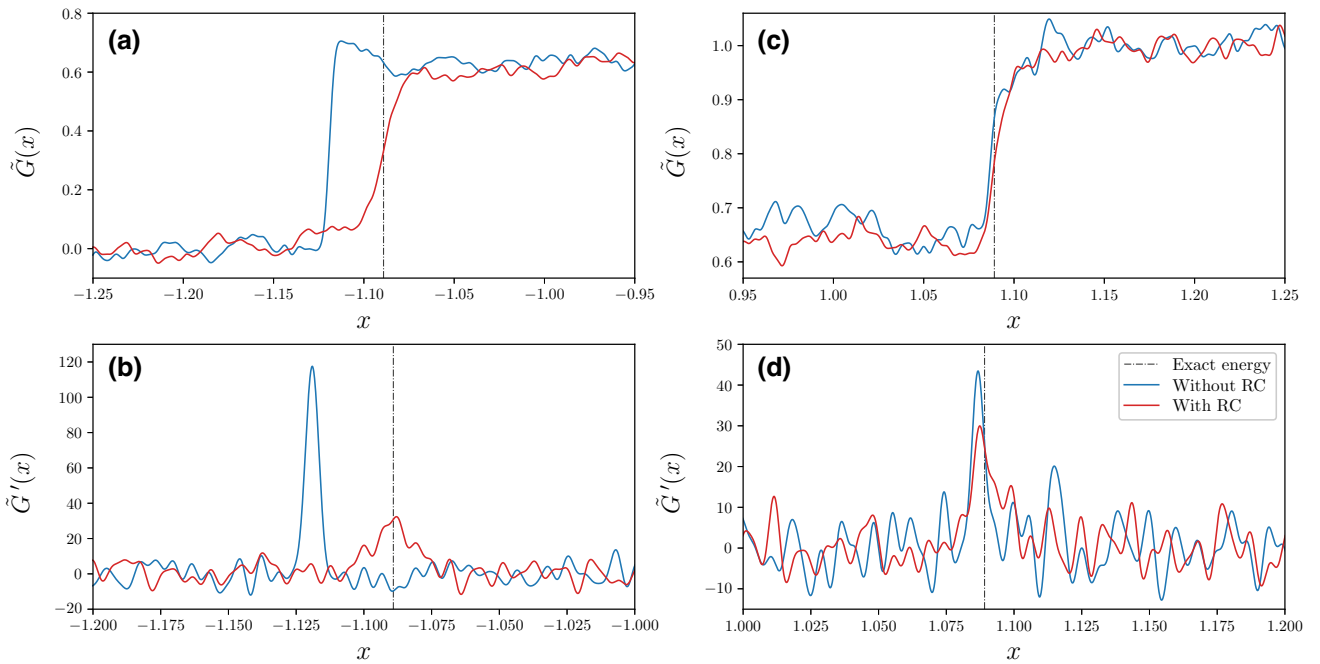
FIG. 13. CDFs and their derivatives for stretched H_2 STO-3G, performed with Trotterization using the Trotter step in Fig. 11. The simulations are performed on a pyQuil QVM. A unitary error of $e^{-i(\theta/2)Z\otimes Z}$ with $\theta = 0.1$ is applied after every CZ gate. We then consider how this affects the CDF, both before and after applying randomized compiling. (a),(b) Without RC, there is a large error in the ground-state energy. This error is effectively removed by RC, at the cost of reduced signal. Interestingly, the error in the excited-state energy is much smaller but still slightly improved by RC. (c),(d) The excited state is harder to distinguish in the CDF derivative due to statistical noise. Note that different x-axis scales are used between subplots.

It is straightforward to see that depolarizing noise does not prevent us from obtaining accurate energy estimates. Under depolarizing noise, the expectation values will decay as $e^{-\gamma|k|}$, for some decay rate γ. We then expect

$$g_k \sim e^{-\gamma|k|} \sum_i p_i e^{-i\tau\lambda_i k}. \qquad (C4)$$

This decay factor does not affect the frequencies present in g_k but it should be expected that it becomes more challenging to reliably estimate each λ_i with increasing γ. To be precise, the Fourier transform of $e^{-\gamma|k|}$ is a Lorentzian function centered about 0, and the width of this function grows as γ increases. The results in Fig. 12 are as expected, given these arguments.

A separate type of error consists of coherent (or unitary) errors, which preserve the purity of the input state. Reference [73] considers a type of control-free phase estimation and demonstrates that unitary errors cause errors in the final phase estimates, which can be largely removed with RC. Here, we demonstrate a similar result with the methodology developed in this paper, with the CDF-based method of Ref. [9] and integrating RC with importance sampling. Following Ref. [73], we apply each CZ gate with a unitary error, so that $U'_{CZ} = \Lambda U_{CZ}$, with

$$\Lambda = e^{-i(\theta/2)Z\otimes Z}. \qquad (C5)$$

We choose $\theta = 0.1$, which is a very large error in practice. All other gates and measurements are applied without error.

The results are presented in Fig. 13. All simulation parameters are the same as for the results in Fig. 12, including β, d, and N_S. Application of the CZ unitary error Λ leads to a large error in the ground-state energy of -7.6 mHa (after rescaling by τ^{-1}). This is reduced to $+0.3$ mHa after applying RC. Interestingly, the error in the excited-state energy is less than 1 mHa in both cases, although there is a slight improvement with RC applied. These results demonstrate that coherent errors can cause significant errors in energy estimates from statistical phase estimation but that RC is a promising approach to help mitigate these errors in practice.

[1] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, ArXiv:quant-ph/9511026.

[2] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. R. Soc. A **454**, 339 (1998).

[3] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proc. Natl. Acad. Sci. **114**, 7555 (2017).

[4] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more

efficient quantum computations of chemistry through tensor hypercontraction, PRX Quantum **2**, 030305 (2021).

[5] N. S. Blunt, J. Camps, O. Crawford, R. Izsák, S. Leontica, A. Mirani, A. E. Moylett, S. A. Scivier, C. Sünderhauf, P. Schopf, J. M. Taylor, and N. Holzmann, Perspective on the current state-of-the-art of quantum computing for drug discovery applications, J. Chem. Theory Comput. **18**, 7001 (2022).

[6] A. V. Ivanov, C. Sünderhauf, N. Holzmann, T. Ellaby, R. N. Kerber, G. Jones, and J. Camps, Quantum computation for periodic solids in second quantization, ArXiv:2210.02403 (2022).

[7] R. D. Somma, Quantum eigenvalue estimation via time series analysis, New J. Phys. **21**, 123025 (2019).

[8] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, PRX Quantum **3**, 010318 (2022).

[9] K. Wan, M. Berta, and E. T. Campbell, Randomized quantum algorithm for statistical phase estimation, Phys. Rev. Lett. **129**, 030503 (2022).

[10] T. E. O'Brien, B. Tarasinski, and B. M. Terhal, Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments, New J. Phys. **21**, 023022 (2019).

[11] A. Dutkiewicz, B. M. Terhal, and T. E. O'Brien, Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits, Quantum **6**, 830 (2022).

[12] L. Clinton, J. Bausch, J. Klassen, and T. Cubitt, Phase estimation of local hamiltonians on NISQ hardware, New J. Phys. **25**, 033027 (2023).

[13] G. Wang, D. Stilck-França, R. Zhang, S. Zhu, and P. D. Johnson, Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision, ArXiv:2209.06811 (2022).

[14] Z. Ding and L. Lin, Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation, ArXiv:2211.11973 (2022).

[15] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear T complexity, Phys. Rev. X **8**, 041015 (2018).

[16] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, Phys. Rev. A **103**, 042605 (2021).

[17] A. W. R. Smith, K. E. Khosla, C. N. Self, and M. S. Kim, Qubit readout error mitigation with bit-flip averaging, Sci. Adv. **7**, eabi8009 (2021).

[18] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[19] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[20] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2020), p. 306.

[21] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors, ArXiv:2201.09866 (2022).

[22] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system, Nat. Commun. **11**, 587 (2020).

[23] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5**, 592 (2021).

[24] P. Czarnik, M. McKerns, A. T. Sornborger, and L. Cincio, Improving the efficiency of learning-based error mitigation, ArXiv:2204.07109 (2022).

[25] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[26] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, Phys. Rev. X **11**, 041039 (2021).

[27] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, Low-cost error mitigation by symmetry verification, Phys. Rev. A **98**, 062339 (2018).

[28] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, Phys. Rev. Lett. **122**, 180501 (2019).

[29] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error mitigation via verified phase estimation, PRX Quantum **2**, 020317 (2021).

[30] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, ArXiv:2210.00921 (2022).

[31] V. Russo, A. Mari, N. Shammah, R. LaRose, and W. J. Zeng, Testing platform-independent quantum error mitigation on noisy quantum computers, ArXiv:2210.07194 (2022).

[32] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, Error mitigation for universal gates on encoded qubits, Phys. Rev. Lett. **127**, 200505 (2021).

[33] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the NISQ to the fault-tolerant quantum computing eras, PRX Quantum **3**, 010345 (2022).

[34] M. Lostaglio and A. Ciani, Error mitigation and quantum-assisted simulation in the error corrected regime, Phys. Rev. Lett. **127**, 200506 (2021).

[35] Y. Xiong, D. Chandra, S. X. Ng, and L. Hanzo, Sampling overhead analysis of quantum error mitigation: Uncoded vs. coded systems, IEEE Access **8**, 228967 (2020).

[36] N. S. Blunt, G. P. Gehér, and A. E. Moylett, Compilation of a simple chemistry application to quantum error correction primitives, ArXiv:2307.03233 (2023).

[37] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, npj Quantum Inf. **8**, 114 (2022).

[38] P. J. J. O'Malley, *et al.*, Scalable quantum simulation of molecular energies, Phys. Rev. X **6**, 031007 (2016).

[39] T. M. Graham, *et al.*, Multi-qubit entanglement and algorithms on a neutral-atom quantum computer, Nature **604**, 457 (2022).

[40] K. Yamamoto, S. Duffield, Y. Kikuchi, and D. M. Ramo, Demonstrating Bayesian quantum phase estimation with quantum error detection, ArXiv:2306.16608 (2023).

[41] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[42] Y. Nam, *et al.*, Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, npj Quantum Inf. **6**, 33 (2020).

[43] L. Zhao, J. Goings, K. Shin, W. Kyoung, J. I. Fuks, J.-K. Kevin Rhee, Y. M. Rhee, K. Wright, J. Nguyen, J. Kim, and S. Johri, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, npj Quantum Inf. **9**, 60 (2023).

[44] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, *et al.*, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, Nat. Commun. **14**, 1952 (2023).

[45] R. Izsák, C. Riplinger, N. S. Blunt, B. de Souza, N. Holzmann, O. Crawford, J. Camps, F. Neese, and P. Schopf, Quantum computing in pharma: A multilayer embedding approach for near future applications, J. Comput. Chem. **44**, 406 (2023).

[46] S.-N. Sun, M. Motta, R. N. Tazhigulov, A. T. K. Tan, G. K.-L. Chan, and A. J. Minnich, Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution, PRX Quantum **2**, 010317 (2021).

[47] R. N. Tazhigulov, S.-N. Sun, R. Haghshenas, H. Zhai, A. T. K. Tan, N. C. Rubin, R. Babbush, A. J. Minnich, and G. K.-L. Chan, Simulating models of challenging correlated molecules and materials on the Sycamore quantum processor, PRX Quantum **3**, 040318 (2022).

[48] M. Benedetti, M. Fiorentini, and M. Lubasch, Hardware-efficient variational quantum algorithms for time evolution, Phys. Rev. Res. **3**, 033083 (2021).

[49] C. M. Keever and M. Lubasch, Classically optimized Hamiltonian simulation, ArXiv:2205.11427 (2022).

[50] Y. Kikuchi, C. M. Keever, L. Coopmans, M. Lubasch, and M. Benedetti, Realization of quantum signal processing on a noisy quantum computer, ArXiv:2303.05533 (2023).

[51] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of PYTHON+NumPy programs, http://github.com/google/jax (2018).

[52] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature **567**, 491 (2019).

[53] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, Zero-noise extrapolation for quantum-gate error mitigation with identity insertions, Phys. Rev. A **102**, 012426 (2020).

[54] E. Knill, Fault-tolerant postselected quantum computation: Threshold analysis, ArXiv:quant-ph/0404104 (2004).

[55] E. Knill, Quantum computing with realistically noisy devices, Nat. Commun. **434**, 39 (2005).

[56] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Purification of noisy entanglement and faithful teleportation via noisy channels, Phys. Rev. Lett. **76**, 722 (1996).

[57] M. R. Geller and Z. Zhou, Efficient error models for fault-tolerant architectures and the Pauli twirling approximation, Phys. Rev. A **88**, 012314 (2013).

[58] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, Scalable error mitigation for noisy quantum circuits produces competitive expectation values, ArXiv:2108.09197 (2021).

[59] T. Kurita, H. Qassim, M. Ishii, H. Oshima, S. Sato, and J. Emerson, Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver, ArXiv:2212.11198 (2022).

[60] S. Ferracin, A. Hashim, J.-L. Ville, R. Naik, A. Carignan-Dugas, H. Qassim, A. Morvan, D. I. Santiago, I. Siddiqi, and J. J. Wallman, Efficiently improving the performance of noisy quantum computers, ArXiv:2201.10672 (2022).

[61] B. Yang, R. Raymond, and S. Uno, Efficient quantum readout-error mitigation for sparse measurement outcomes of near-term quantum devices, Phys. Rev. A **106**, 012423 (2022).

[62] P. J. Karalekas, N. A. Tezak, E. C. Peterson, C. A. Ryan, M. P. da Silva, and R. S. Smith, A quantum-classical cloud platform optimized for variational hybrid algorithms, Quantum Sci. Technol. **5**, 024003 (2020).

[63] R. S. Smith, M. J. Curtis, and W. J. Zeng, A practical quantum instruction set architecture, ArXiv:1608.03355 (2016).

[64] F. Neese, Software update: The ORCA program system, version 5.0, Wiley Interdiscip. Rev. Comput. Mol. Sci. **12**, e1606 (2022).

[65] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, PySCF: The PYTHON-based simulations of chemistry framework, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[66] Q. Sun, *et al.*, Recent developments in the PySCF program package, J. Chem. Phys. **153**, 024109 (2020).

[67] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, *et al.*, Gaussian 16 Revision C.01, Gaussian Inc. Wallingford CT, 2016, https://gaussian.com/citation.

[68] J. R. McClean, *et al.*, OpenFermion: The electronic structure package for quantum computers, Quantum Sci. Technol. **5**, 034014 (2020).

[69] S. B. Bravyi and A. Y. Kitaev, Fermionic quantum computation, Ann. Phys. (NY) **298**, 210 (2002).

[70] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, Phys. Rev. A **96**, 022330 (2017).

[71] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, ArXiv:1701.08213 (2017).

[72] http://doi.org/10.5281/zenodo.10201197.

[73] Y. Gu, Y. Ma, N. Forcellini, and D. E. Liu, Noise-resilient phase estimation with randomized compiling, ArXiv:2208.04100 (2022).

[74] R. Kshirsagar, A. Katabarwa, and P. D. Johnson, On proving the robustness of algorithms for early fault-tolerant quantum computers, ArXiv:2209.11322 (2022).

[75] L. E. S. Netto, M. A. de Oliveira, G. Monteiro, A. P. D. Demasi, J. R. R. Cussiol, K. F. Discola, M. Demasi, G. M. Silva, S. V. Alves, V. G. Faria, and B. B. Horta, Reactive cysteine in proteins: Protein folding, antioxidant defense, redox signaling and more, Comp. Biochem. Physiol. Part C: Toxicol. Pharmacol. **146**, 180 (2007), Fourth Special Issue of CBP dedicated to "The Face of Latin American Comparative Biochemistry and Physiology".

[76] U.S. Food and Drug Administration, FDA expands approved use of Imbruvica for rare form of non-Hodgkin lymphoma, https://wayback.archive-it.org/7993/20170112 222810/http://www.fda.gov/NewsEvents/Newsroom/Press Announcements/ucm432123.htm.

[77] C. Møller and M. S. Plesset, Note on an approximation treatment for many-electron systems, Phys. Rev. **46**, 618 (1934).

[78] T. H. Dunning Jr, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, J. Chem. Phys. **90**, 1007 (1989).

[79] R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions, J. Chem. Phys. **96**, 6796 (1992).

[80] D. E. Woon and T. H. Dunning Jr, Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon, J. Chem. Phys. **98**, 1358 (1993).

[81] N. Holzmann and P. Schopf, New insight into the covalent binding mechanism of ibrutinib to Bruton's tyrosine kinase—a QM cluster study (to be published).