

Error Mitigation and Circuit Division for Early Fault-Tolerant Quantum Phase Estimation

Alicja Dutkiewicz^{1,*†}, Stefano Polla^{1,2,*‡}, Maximilian Scheurer³, Christian Gogolin^{1,3},
William J. Huggins² and Thomas E. O'Brien^{1,2,§}

¹*Lorentz Institute for Theoretical physics, Leiden University, NL-2300 RA Leiden, Netherlands*

²*Google Quantum AI*

³*Covestro Deutschland AG, 51373 Leverkusen, Germany*



(Received 23 May 2025; accepted 11 September 2025; published 27 October 2025)

As fully fault-tolerant quantum computers capable of solving useful problems remain a distant goal, we anticipate an era of *early fault tolerance* during which limited error correction is available. We propose a framework for designing early fault-tolerant algorithms by trading between error-correction overhead and residual logical noise, and apply it to quantum phase estimation (QPE). We develop a quantum-Fourier-transform- (QFT) based QPE technique that is robust to global depolarizing noise and outperforms the previous state of the art at low and moderate noise rates. We further introduce the explicitly unbiased maximum-likelihood estimator (EUMLE), a data-processing technique that mitigates *arbitrary* errors in QFT-based QPE schemes. EUMLE provides consistent asymptotically normal error-mitigated estimates, addressing the open problem of extending error mitigation beyond expectation-value estimation. Applying this scheme to the ground-state problem of the two-dimensional Hubbard model and various molecular Hamiltonians, we find that we can roughly halve the number of physical qubits with a ~ 10 times wall-clock time overhead, but that further reduction causes a steep run-time increase. This work provides an end-to-end analysis of early fault-tolerance cost reductions and space-time trade-offs, and identifies which areas can be improved in the future.

DOI: [10.1103/mlmy-yskj](https://doi.org/10.1103/mlmy-yskj)

I. INTRODUCTION

Despite significant experimental progress in recent years [1–4], quantum computers still require many orders of magnitude more qubits to achieve advantage for practically relevant computational problems [5–7]. While quantum error correction (QEC) promises to enable fault-tolerant quantum computing with a cost that scales polylogarithmically in the size of the computation, the overhead required for the most practically realizable schemes is large. For example, the combination of the two-dimensional surface code with magic state distillation [8–11] is capable of universal quantum computation, but may require hundreds or thousands of physical qubits per logical qubit to ensure the

success of the computation. This overhead has prompted a push for work on *early fault tolerance* [12–28], i.e., the search for computational frameworks and trade-offs that enable us to take advantage of the machinery of QEC without incurring its full cost.

Initial explorations of early fault tolerance have focused along a few lines. A first line of research has aimed to reduce costs for small and medium-sized problems, by making tailored algorithm design choices. For example, the use of Trotter-based simulation instead of methods with better asymptotic scaling has been shown to offer lower costs for smaller systems [12,27,29,30]. Another approach has focused on adapting quantum algorithms to achieve a low depth per circuit. This is motivated by the observation that a given error rate p determines the maximum number of allowable operations $G \approx -\log(1-p)^{-1} \approx p^{-1}$ before the exponential decay of circuit fidelity sets in. Several works have accomplished this by trading off reduced circuit depth for an increased number of circuit repetitions [14,17–23,31–34]. A third approach has been to use error-mitigation techniques—originally designed to handle physical noise in near-term quantum devices—in order to reduce the noise-induced bias without reducing circuit

*These authors contributed equally to this work.

†Contact author: dutkiewicz@lorentz.leidenuniv.nl

‡Contact author: polla@lorentz.leidenuniv.nl

§Contact author: teobrien@google.com

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

depth, albeit at the expense of many circuit repetitions [27,35,36].

Quantum error mitigation (EM) [37] has seen increased interest in recent years, due to the development of experimental platforms which are increasingly powerful, yet remain far from enabling fault-tolerant computing. Mitigating errors caused by physical noise avoids the large constant factor requirements of fault tolerance, but results in an exponential overhead with increasing circuit depth and error rate [38–43]. A wide range of EM techniques have been conceived [37,44–51] and implemented in practice [52–57], demonstrating up to a thousandfold reductions in noise bias [56]. While these methods cannot replace scalable error correction, they have been suggested as a last-mile solution in fault-tolerant quantum algorithms, relaxing the demands on a QEC protocol by enabling a computation to tolerate $\mathcal{O}(1)$ errors per circuit repetition [13,26,35–37]. Some approaches specifically address scenarios in which the error rates vary between qubits or operations [24,27,35,36], with a particular emphasis on the non-Clifford operations that are expected to be the most costly to implement fault tolerantly. However, most EM protocols have focused on expectation-value estimation, as opposed to more general quantum computation; extending beyond this has been labeled an outstanding challenge in Ref. [37].

Quantum phase estimation [58], a workhorse of quantum computing, aims to compute an eigenvalue of a unitary matrix given its circuit implementation and an initial state with sufficiently high overlap with the target eigenstate. Traditional phase-estimation techniques [59] use a control register which, after accumulating phase information via controlled time evolution, undergoes the quantum Fourier transform (QFT) to yield a state with high density on binary representations of the desired phase. These *QFT-based QPE* algorithms achieve the Heisenberg limit, the information-theoretic optimum for phase estimation, including constant factors [60]. It is possible to remove the QFT, either by direct dequantization [58,61] or by analyzing the outputs of independent circuits that use a single control qubit via signal-processing techniques [19–21,32,62–65]. (Indeed, under some constraints on U , phase estimation without any control qubits is possible [50,66–68].) These *single-control QPE* techniques are of particular interest in early fault tolerance [13–15,18,20,21,25,27,32,69], as these protocols can tolerate a constant amount of error per circuit [62,63], avoiding the need for a high overall success probability across all circuit runs. However, it remains to be understood how this reduction in circuit depth and increase in error tolerance propagates through QEC to a trade-off between the physical qubit count and the wall-clock time (of the full estimation protocol). Additionally, besides Refs. [70,71], there has been little exploration of whether QFT-based QPE techniques can be divided into smaller circuits akin to

single-control methods. To the best of our knowledge, no prior work has applied circuit-division to make QFT-based QPE robust to noise.

A. Outline

In this work, we develop a framework that jointly optimizes QEC, mitigation, and algorithm design for early fault-tolerant quantum computation. Specifically, we study the trade-off between running fewer larger circuits at lower error rates (requiring more physical qubits per logical qubit) versus subdividing computations into many smaller circuits that can be executed at reduced code distances. Our work bridges two research directions that have so far been explored independently: the optimization of algorithms for early fault tolerance and the integration of quantum error mitigation within fault-tolerant protocols. Within this framework, we propose a new noise-robust QPE algorithm, based on subdividing the QFT-based QPE circuit of Ref. [60] used in state-of-the-art fault-tolerant benchmarks [6,7]. In the presence of global depolarizing noise, we prove that the algorithm converges to any target precision and quantify its cost as a function of noise strength. We compare our method to the noise-optimized robust-phase-estimation (RPE) scheme of Ref. [72], finding that our approach requires between 30% and 130% of the resources needed for RPE, depending on the target precision and noise strength. Then, in order to mitigate biases from more general noise models, we develop a regularized explicitly unbiased maximum-likelihood estimator (EUMLE), which extends probabilistic error-cancellation techniques beyond expectation-value estimation. This estimator, when combined with our QPE algorithm, guarantees convergence under arbitrary noise channels. For natural noise models based on local Pauli errors, we upper bound the overhead of our method relative to the global-depolarizing-noise case by a factor of 7. Finally, we present concrete and comprehensive estimates of the physical qubit count and wall-clock time for computing ground-state energies of the two-dimensional Hubbard model [6] and of various molecular Hamiltonians compressed with tensor hypercontraction [7]. Our estimates assume surface-code computing [9] with CCZ factories [11] and quantify how the previously discussed method allows to trade off an increased computation time for a reduced qubit count. We find that the available gains are limited by the error-mitigation overhead, along with the large costs associated with magic state distillation; a factor of 2 reduction in qubit count comes with an increase in the total wall time of 2 times to 100 times.

The remainder of the paper is organized as follows. The next two sections present the main narrative, summarizing our results (Sec. II) and their implications for early fault-tolerant quantum phase estimation (Sec. III). Sections IV–IV F contain the technical developments, detailed

derivations, and extended results that underpin our conclusions. This structure is intended to provide a concise overview for the general reader while giving specialists direct access to the complete methodology and analysis.

II. SUMMARY OF RESULTS

Our overall goal is to extend recent resource estimates for practical quantum phase estimation [6,7] to the case of early fault-tolerant quantum computers. We define early fault tolerance as *the operational regime in which we can use QEC to correct all but $\Omega(1)$ errors and we must design algorithms to be robust to the remainder (via, e.g., quantum error mitigation)*. In this setup, we retain the ability to spend more physical resources to reduce the logical error rate γ through QEC. In contrast to standard fault-tolerant quantum algorithm design, we consider γ as an additional free parameter to be optimized jointly with other parameters of the algorithm, subject to any constraint imposed by the device (e.g., the number of available physical qubits). The comparison of the standard fault-tolerant algorithm design process with our early fault-tolerant framework is illustrated in Fig. 1. We choose this *top-down* approach to early fault tolerance (which is similar to the approach taken in Refs. [13]) instead of a *bottom-up* approach taken by some other works [26,28], as this requires relatively few assumptions on the physical device used.

A. Circuit division and error mitigation

In order to adapt our quantum algorithm to a range of noise rates γ , we explore *circuit-division* protocols that

replace a single coherent execution of a quantum circuit consisting of G gates by the execution of M shorter circuits, each consisting of fewer than G/R gates (with $M > R$). The samples from these shorter circuits are then combined by a classical algorithm to recover the same result as the original quantum algorithm. Circuit division allows us to trade between executing a few large circuits at lower error rates γ (thus requiring more physical qubits per logical qubit) versus dividing these into many shorter circuits (which can tolerate higher γ to achieve the same success probability). Reducing the number of physical qubits required would enable a small early fault-tolerant quantum computer to perform calculations that would otherwise be out of reach, at the cost of increasing the total run-time of the calculation by a factor of $\tilde{O}(M/R)$. Note that this flavor of circuit division is different than the divide-and-conquer approaches which have been proposed for, e.g., optimization [73–75] and variational algorithms [76,77]. These are not known to be noise robust and the overhead in most cases is exponential in the number of cuts. For a circuit-division protocol to be practical, its overhead M/R needs to be reasonable. Furthermore, the classical postprocessing should be robust to faults in multiple subcircuits, as a union bound implies that the probability of failure of at least one subcircuit is larger than the probability of the original circuit failing.

Error mitigation can be further employed to boost the robustness of circuit-division protocols to the remaining subcircuit faults. The overhead of EM typically depends on the subcircuit fidelity F [37], which can be controlled either by limiting the subcircuit depth or by running

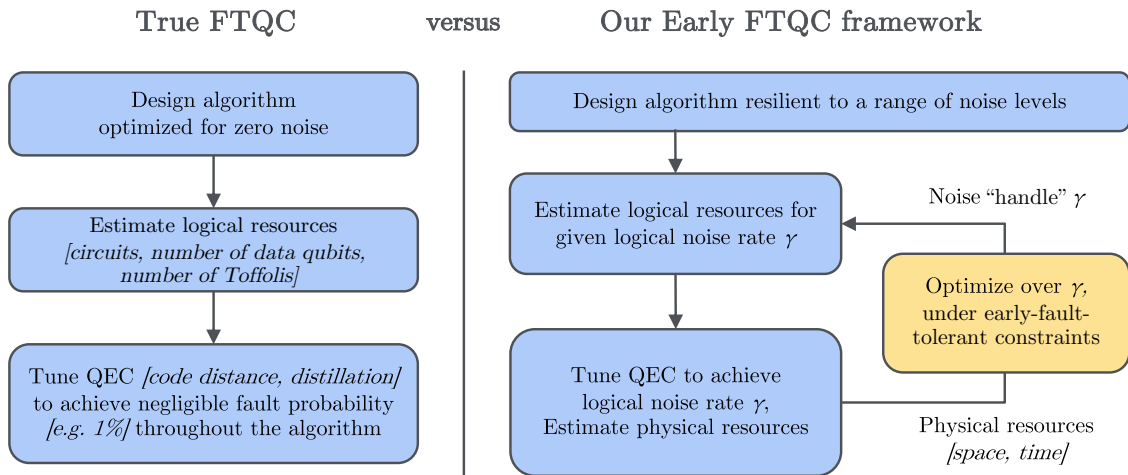


FIG. 1. An illustration of our framework for designing algorithms for early fault tolerance and how it differs from traditional fault-tolerant algorithm design and resource estimation. The standard approach (left) involves optimizing an algorithm in the absence of noise and then choosing quantum error-correction parameters such that the failure probability is small enough. We propose an alternative approach better suited to early fault tolerance. By using noise-resilient algorithms as a building block, we can allow for non-negligible levels of noise throughout the execution of a circuit, jointly optimizing the parameters of the algorithm, the quantum error-correcting scheme, and the error-mitigation technique used to address the residual error.

more error correction. We provide a general back-of-the-envelope study of the trade-off between error mitigation and error correction in Sec. IV A; we find that combining the two can lead to a significant boost of the maximum logical circuit size that can be run on a given device, in a narrow regime of physical qubit count and error rate.

Single-control QPE variants are readily adaptable to circuit division. In the presence of global depolarizing noise, they are known to be robust to faults in the subcircuits with a quadratic overhead $M \sim R^2$ [13,23,63]. As these methods rely on measuring expectation values of Hadamard-test-like circuits, they are naturally compatible with existing EM protocols. This has motivated a lot of interest in recent years [17,20–22,65,67,78]; however, little work has gone into testing the robustness of more traditional versions of QPE (that utilize a multi-qubit control register and the quantum Fourier transform).

B. QFT-based QPE algorithm robust to depolarizing noise

Our first contribution is to provide a circuit-division variant of QFT-based QPE, *maximum-likelihood sine-state QPE* (MSQPE), that is robust to global depolarizing noise. We give the algorithm in Sec. IV C 2 and prove the following result.

Theorem 1 (Theorem 6, informal). Given an initial eigenstate preparation, the MSQPE algorithm in the presence of global depolarizing noise with rate γ (per call of the unitary) converges to error ϵ in a total number of uses of the unitary given by

- (a) $\pi\epsilon^{-1}$, if $\epsilon \gg \gamma$
- (b) $C\gamma\epsilon^{-2}$ with $C \approx 20$, if $\epsilon \ll \gamma$

and interpolates between these limits when $\epsilon \sim \gamma$.

This algorithm relies on two technical developments: the extension of QFT-based QPE methods to allow for an arbitrary number of calls to the unitary (as opposed to powers of 2 only) and a noise-robust classical postprocessing of multiple QPE subroutines via maximum-likelihood estimation.

Theorem 1 demonstrates that QFT-based QPE methods can be made robust to noise, thus allowing for fair comparison with single-control QPE variants. We optimize the hyperparameters of the (QFT-based) MSQPE algorithm (details in Sec. IV C 2) and compare it with the (single-control) RPE algorithm [62], with optimized parameters from Ref. [72]. In Fig. 2, we plot the ratio between the costs of the two algorithms to achieve the same target error ϵ , where the cost $\mathcal{T}_{\text{tot}}$ is measured as the total number of unitaries across all circuits used. We observe that the two methods perform similarly; MSQPE performs better when the noise rate γ is low compared to the target error ϵ , by a

factor that increases for smaller γ . On the other hand, when the target error is small $\epsilon \ll \gamma$, the cost of RPE is about 30% smaller. In Sec. IV C 3, we expand on this comparison and note that both estimators can likely be improved; the MSQPE estimator by optimizing the sine-state in the presence of noise and the RPE estimator by numerically optimizing the RPE hyperparameters (the optimization in Ref. [72] aimed to achieve analytical bounds). However, across the range of errors and noise rates considered, the performance of both estimators falls within a factor of 10 of the information-theoretic bounds, which places a limit on any further gains.

C. Error-mitigation overhead for single-control QPE

While the global depolarizing channel is a useful analytical tool, it fails to accurately describe the residual logical noise in fault-tolerant quantum computers. Previous work has shown that other natural noise models (in particular, local depolarizing noise) can result in a biased signal in single-control QPE methods [63] and thus a biased phase estimate. Consequently, more complex EM strategies will be required to address the general noise models expected to characterize residual logical error. EM introduces a circuit-repetition overhead that scales inverse polynomially with the circuit fidelity F .

Our second contribution is to study how overhead associated with error mitigation (defined in terms of circuit

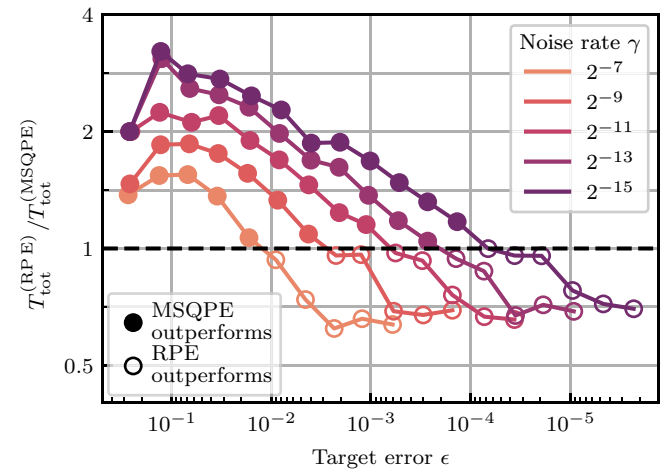


FIG. 2. The ratio of the total execution times $\mathcal{T}_{\text{tot}}$ of the optimized robust-phase-estimation (RPE) algorithm of Ref. [72] and the maximum-likelihood sine-state quantum phase-estimation (MSQPE) algorithm developed in this work across a range of target errors on the phase estimate ϵ and noise rates γ translating to an error rate per unitary of $p_{\text{err}} = (1 - e^{-\gamma})$. The dashed line at 1 denotes equivalent performance between the two methods; points above 1 give a region in which the MSQPE estimator performs better, while points below 1 show where RPE outperforms MSQPE.

repetitions) translates to an overhead in the cost of error-mitigated QPE (in terms of the number of calls to a noisy implementation of the unitary operator). We find that this overhead can be suppressed logarithmically due to circuit division.

Theorem 2 (Theorem 7, informal). Given an error-mitigation scheme for expectation values with a sample overhead scaling with a power of the circuit fidelity $F^{-\nu}$, one can obtain an unbiased error-mitigation scheme for single-control phase estimation with an overhead proportional to $\nu \gamma / \epsilon$, where ϵ is the precision of the estimate and $\gamma \leq 1/\nu$ is the noise rate per call of the unitary.

This result highlights a significant advantage, as it narrows the cost gap between EM methods with limited scope but low overhead (e.g., postselection) and more general yet costlier approaches (e.g., probabilistic error cancellation). In Table I, we summarize the circuit-repetition overheads C_{em} known for various methods and we show how these translate to unitary-call overheads for phase estimation $C_{\text{em}}^{\text{QPE}}$.

D. Error mitigation beyond expectation values

Our third contribution is to develop a general EM protocol for maximum-likelihood estimation. This is motivated by our intention to apply MSQPE in the presence of arbitrary noise. However, our protocol is more general; it can be applied to any estimation problem that would require samples from a noiseless quantum circuit, while only samples from a noisy device with a known error channel are available. Maximum-likelihood is a more general estimator than sample average, typically used for expectation-value estimation. To the best of our knowledge, prior EM schemes have exclusively targeted expectation-value estimation, leaving the extension of EM to other estimation tasks as an open problem identified in the literature [37].

Our protocol, detailed in Sec. IV E 1, is based on rewriting the distribution $P(x)$ that describes the outcomes x of a noiseless circuit as a nonconvex combination of distributions $Q_j(x)$ that can be sampled on the noisy device: $P(x) = \sum_j \alpha_j Q_j(x)$. This can be done by using the channel-decomposition techniques from probabilistic error cancellation (PEC) [44,45]. Given a parametrized model $P(x|\phi)$ for the noiseless distribution, the expected value of the log-likelihood function $\tilde{\mathcal{L}}(\phi)$ can then be expanded in terms of Q_j :

$$\tilde{\mathcal{L}}(\phi) = \sum_j \alpha_j \mathbb{E}[\log P(x|\phi) | x \sim Q_j(x)]. \quad (1)$$

This function can be estimated by taking samples from $Q_j(x)$ on the noisy quantum device, by importance sampling of j . We obtain the EUMLE by maximizing this likelihood function and we prove its convergence in the following theorem.

Theorem 3 (Theorem 8, informal). Given a parametrized distribution model $P(x|\phi)$ that describes the outcomes x of a noiseless quantum circuit and assuming access to samples from a noisy quantum computer with a known error channel, we can construct an EUMLE. This estimator converges to the true value of ϕ with a variance that scales inversely with the total number of samples.

E. QFT-based QPE with arbitrary noise

We then apply the EUMLE to MSQPE, obtaining a QFT-based QPE algorithm robust to arbitrary noise channels. To ensure that the estimator has a bounded sampling cost, we introduce a further regularization which addresses the issue of potential zeros of the likelihood function (see Sec. IV E 2). Under local depolarizing noise at rate γ , the unitary-call cost of performing phase estimation to precision ϵ is found to be $\mathcal{T}_{\text{tot}} \leq 137\gamma/\epsilon^2$. This corresponds to a QPE error-mitigation overhead of approximately $44\gamma/\epsilon$. Comparing to the cost of MSQPE with global depolarizing noise in Theorem 1, we observe that the additional

TABLE I. The error-mitigation overheads for expectation-value estimation (C_{em}) and for quantum phase estimation ($C_{\text{em}}^{\text{QPE}}$) for different error-mitigation techniques. a and b are constants depending on the QPE algorithm, Λ is a constant depending on the noise model and error-mitigation approach, γ is the noise rate, and ϵ is the target precision of phase estimation.

Postselection	$C_{\text{em}} = \Lambda F^{-1}$	$C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \frac{a\Lambda\epsilon}{\pi} \max(\frac{b\gamma}{\epsilon}, 1)$	Symmetry verification, verified phase estimation ^a
Rescaling	$C_{\text{em}} = \Lambda F^{-2}$	$C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \frac{a\Lambda\epsilon}{\pi} \max(2\frac{b\gamma}{\epsilon}, 1)$	Echo verification, ^a direct rescaling under GDN
Explicit unbiasing	$C_{\text{em}} = \Lambda F^{-4}$	$C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \frac{a\Lambda\epsilon}{\pi} \max(4\frac{b\gamma}{\epsilon}, 1)$	Probabilistic error cancellation, virtual distillation

^aThe overhead of echo verification (a.k.a. verified phase estimation) depends on the assumptions on the circuit: if the cost of state preparation is negligible, its overhead is postselection bounded; if the dominant cost comes from state preparation, it behaves like a rescaling technique due to the required doubling of the circuit depth. See Sec. IV D for a detailed discussion.

overhead of mitigating a more general noise channel is approximately a factor of 7 (see Sec. IV E 4).

F. Cost estimates

Finally, we turn to compiling MSQPE for a set of example problems to the surface code, for a range of residual logical noise rates γ . Our goal is to quantify the trade-off between the number of physical qubits and the total computation time required to estimate ground-state energies to a given precision. The details of this resource estimates are given in Sec. IV F and we provide open-source code to reproduce and extend our results [79].

To perform this resource analysis, we need to select a specific QEC architecture, but we expect that our framework can be used to extend any such architecture into the early fault-tolerant regime. We consider a quantum computer using a two-dimensional rotated surface code [8] that performs fault-tolerant Clifford gates using lattice surgery [80] and uses CCZ magic state injection and distillation to implement Toffoli gates [11]. Such an architecture is popular because it can tolerate relatively high physical error rates and because it can be implemented using a two-dimensional lattice of locally connected qubits [81,82].

The target Hamiltonians that we consider are (1) a set of $L \times L$ -sites Fermi-Hubbard models and (2) electronic

structure Hamiltonians for a selection of small to classically nontrivial molecules. In both cases, we estimate the ground-state energy to precision ΔE by applying MSQPE to a qubitized walk operator $\mathcal{W}$ [83]. For the Fermi-Hubbard Hamiltonian, we choose model parameters ($t = 1, u = 4$) and target precision ($\Delta E = 10^{-2}$) following Ref. [6], reproducing a regime that is challenging for classical methods [84]. $\mathcal{W}$ is implemented following the methods of Ref. [6]. Note that qubitization of the Hubbard model does not require data uploading by quantum read-only memories (QROMs) and the only approximation error comes from the rotations used to prepare a two-qubit state encoding the Hamiltonian coefficients. The molecular models that we consider span active space sizes between six and 26 spatial orbitals. All molecular Hamiltonians have been compressed with tensor hypercontraction (THC) [7] after applying symmetry shifts [85], to further reduce the qubitization 1-norm λ . We aim for the chemically desirable $\Delta E = 10^{-3} E_h$ accuracy of the final results, choosing the THC hyperparameters accordingly. See Sec. IV F 2 for computational details on the construction of the molecular Hamiltonians.

We estimate the Toffoli costs of phase estimation on the qubitization walk operators $\mathcal{W}$ using QUALTRAN [86]. We consider using one or multiple CCZ factories [11] to produce magic states for implementing Toffoli gates and we optimize the data-qubit code distances d and the

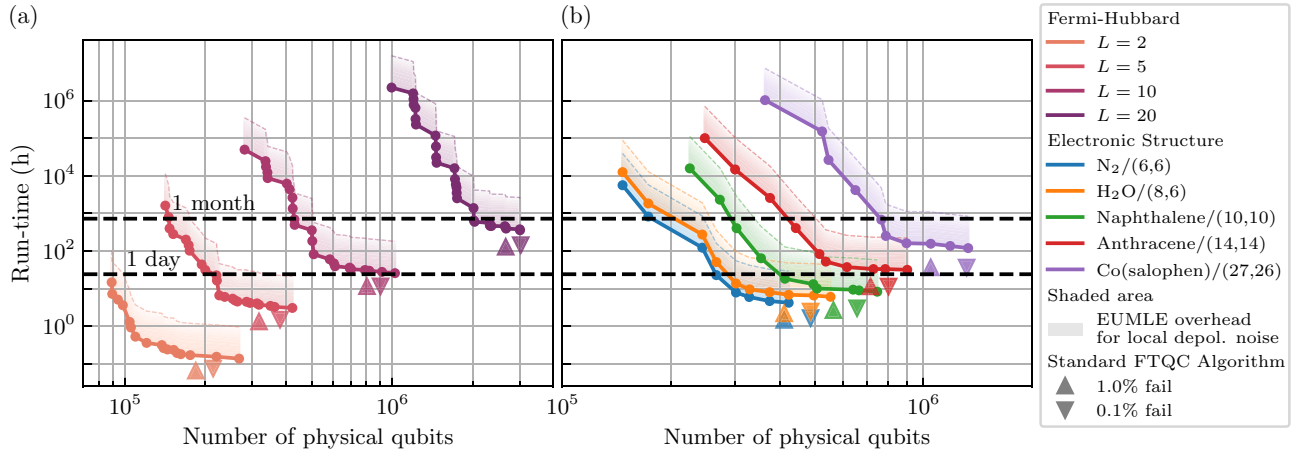


FIG. 3. The physical costs for MSQPE applied to (a) the qubitized Fermi-Hubbard model Hamiltonian and (b) the active-space molecular electronic structure Hamiltonians of some chemical systems. For the Fermi-Hubbard Hamiltonian, we choose the hopping parameter $t = 1$ to set the energy units and interaction strength as $u/t = 4$. For the electronic structure Hamiltonians, the active spaces sizes (number of electrons and spatial orbitals) are notated in the legend. The target precision on the energy is $\Delta E = 10^{-2}$ for the Hubbard model and $\Delta E = 10^{-3} E_h$ for the chemical systems. The physical resources are estimated assuming computation in the surface code using CCZ resource states produced by a single factory; the size of the surface-code and CCZ factory parameters are chosen to minimize the computation volume while keeping the error rate below a chosen γ , assuming a physical error rate of 10^{-3} , a surface-code clock cycle of $1 \mu s$, and a 5% routing overhead. The residual error is assumed to follow a global-depolarizing-noise model (solid lines). We also report the sample overhead bound of 7 times required to mitigate local depolarizing noise of the same strength within the EUMLE framework (shaded areas). Changing γ allows us to trade between the number of physical qubits (used as a resource for implementing better error correction) and the total run-time. The values of γ used to generate this plot are in the range $[10^{-9}, 10^{-2}]$ for Fermi-Hubbard and $[10^{-8}, 10^{-2}]$ for electronic structure. For comparison, we report the physical costs of the standard fault-tolerant single-circuit implementation of sine-state QPE, accepting a failure probability of 1% or 0.1% (triangles).

two-stage distillation code distances d_0 and d_1 to minimize the computational volume of $c\mathcal{W}$ while keeping the error rate (i.e., the logarithm of the logical error probability per call of $c\mathcal{W}$) bounded by γ . We then use the methods detailed earlier in the paper to optimize a robust estimator from oracle calls to the noisy $\mathcal{W}$, with a final target standard deviation of $\Delta\phi = \Delta E/\lambda$ (where λ is the qubitization 1-norm). (Note that this is different than requiring precision $\Delta\phi$ with a fixed success probability, which is a common choice in the literature. We expand on this comparison in Sec. IV F 3.)

In Fig. 3, we plot the resulting estimates of the wall-clock time and the number of physical qubits required to execute our MSQPE algorithm (assuming global depolarizing noise), as well as a bound on the EUMLE cost for a more realistic logical noise model (local depolarizing). Across the range of experiments considered, we observe that circuit splitting allows for a factor of ~ 2 decrease in the qubit count at the cost of a factor of ~ 10 increase in the wall-clock time. All the curves demonstrate an “elbow-like” behavior; reducing the number of physical qubits below a specific point incurs a significant penalty in total run-time. The effect becomes more pronounced for larger systems. This aligns with our back-of-the-envelope characterization of circuit division (Sec. IV A 1), where we observe that there only is a thin region of error rate and number of available qubits in which significant circuit-depth enhancements are possible by combining error correction with error mitigation. As anticipated, it is apparent that the simulation of chemistry is more resource intensive than the simulation of the Hubbard model at the same state-space size. Our analysis makes this quantitative: compared to the $L = 5$ Hubbard model with 25 spatial modes, the 26 spatial orbital Co(salophen) molecule requires roughly 4 times more qubits and around 1.5 orders of magnitude longer wall-clock time. This is due to the higher complexity of molecular Hamiltonians and the costs associated with the QROMs required to upload the Hamiltonian coefficients for realizing the qubitization oracle.

III. DISCUSSION AND CONCLUSIONS

Besides various technical contributions to phase estimation and error mitigation, the key focus of this work is to develop a framework that adapts fully fault-tolerant algorithms to the early fault-tolerant regime and quantifies the savings achieved by trading error rates for physical qubit counts. Such a framework requires a full-stack treatment of quantum algorithm compilation, including a robust treatment of the output of a noisy quantum computer. Improvements in logical gates, QEC codes, or Hamiltonian simulation algorithms can be immediately tested within our code, which relies on QUALTRAN [86] for computing logical costs, in a plug-and-play fashion. The importance of this adaptability is highlighted by recent advances in

T-gate cultivation [87], which will significantly reduce the resource estimates presented here and necessitate going beyond simple T-gate counting. Additionally, our approach allows for exploring alternative QEC architectures, such as those with long-range connectivity and using nonsurface codes [2,3]. As illustrated in Fig. 3, with current methods, qubit counts can be reduced by a factor of 2–3, but with significant overheads of a factor of 10–100 in run-time. This result, while somewhat sobering for early fault-tolerance research, stems from the fact that the gain in physical qubit count from increased logical circuit depth [Eq. (9)] is only logarithmic.

We would like this paper to be understood as an invitation to try to beat the provided early fault-tolerant estimates, and we can already suggest various improvements that could be achieved within our framework. In this work, we have focused on developing noise-robust postprocessing for data sampled using the sine-state QPE circuit, which is optimal only in the noiseless case. Optimizing this circuit while taking noise into account could further improve MSQPE and bring its performance closer to the information-theoretic limit [88], potentially saving a factor of ≈ 2 (see Sec. IV C 3). Additionally, applying the unary-iteration-based technique from Ref. [7] to the iterated qubitized walk operator could yield another factor of 2 reduction in logical costs. Furthermore, in Sec. IV E 5, we outline a filtering procedure for the EUMLE, which could reduce the error-mitigation overhead toward the bound from postselection, potentially achieving a factor of 4 reduction in total cost (analogous to Theorem 2). Additional savings could also be achieved in the compilation of our algorithms by allowing higher algorithmic error. For instance, reducing the precision of the rotation angles could lower the physical resource requirements, though balancing this against energy-estimation accuracy remains nontrivial.

We emphasize that the EUMLE introduced in this work has potential applications beyond QPE, in more general estimation problems. On the other hand, the EUMLE needs accurate characterization of the noise model in order to construct a quasiprobability decomposition. Investigating the robustness of EUMLE against inaccurate noise characterization is a clear direction for future work. Similarly, further research is needed to characterize residual logical noise in specific fault-tolerant computation models. This characterization could be carried out analytically, by analyzing error-correcting codes, decoders, and distillation protocols, or experimentally, by extending noise-characterization techniques such as gate set tomography [89–91] to fault-tolerant computation.

We have aimed to be as complete as possible in the resource estimate of MSQPE. However, we have neglected a significant complicating factor for ground-state energy estimation: the preparation of a good initial state. Efficient protocols for preparing high-fidelity initial states are

a topic of active research in the field of quantum simulation for chemistry [92–95] and the cost of preparing a state $|\psi\rangle$ with overlap $a = |\langle\phi|\psi\rangle|$ with the ground state $|\phi\rangle$ can vary widely depending on the problem. If only an approximate ground state is available (i.e., $a \neq 1$) the energy can still be estimated, provided that the first excited state of the Hamiltonian is well separated from the ground state, by an energy gap Δ . This can be achieved in two ways: (1) energy-filtering and amplitude amplification, which allow us to distill the exact ground state $|\phi\rangle$ [96,97]; or (2) classical signal-processing techniques that isolate the ground-state phase information [20,22,32,63,65]. While the first method integrates immediately with our algorithm, it requires long state-preparation circuits, with a cost scaling as $\mathcal{O}(\Delta^{-1}a^{-1})$. This cost needs to be added for each circuit sample, which penalizes circuit division. The second class of methods require only one preparation of $|\psi\rangle$ for each circuit sample, but introduce a sampling overhead scaling as $\mathcal{O}(a^{-2})$ and require QPE circuits of at least depth $\mathcal{O}(\Delta^{-1})$ [98]. Drawing inspiration from these methods, we propose a filtered EUMLE algorithm in Sec. IV E 5, which we anticipate will achieve similar performance. However, extending the classical postprocessing of the EUMLE to the multiple-eigenstate scenario is nontrivial and bounding the resulting overhead is left to future work. All the options described above establish a lower bound on the minimum circuit depth required for QPE of $\Omega(\Delta^{-1})$, which translates in our framework to a maximum tolerable error rate γ (and thus a minimum required number of qubits). If the error rate exceeds this threshold, the cost of estimating the ground-state energy becomes exponential in ϵ [98]. Therefore, we suggest that early fault-tolerant-algorithm research should prioritize optimizing initial state preparation and identifying applications where the state-preparation costs are reasonably low.

IV. TECHNICAL DETAILS AND EXTENDED RESULTS

A. Designing algorithms for early fault tolerance

Most analysis of fault-tolerant quantum algorithms makes the assumption that QEC enables error-free computation. Under reasonable physical assumptions, the space and time overheads of QEC are at most polylogarithmic in the size of the computation. In the limit where devices are large and error rates are low, this asymptotic behavior dominates and it is affordable to suppress errors to arbitrarily small levels. However, before we reach this regime, we expect an era of “early fault tolerance,” where we have access to quantum computers that are capable of benefiting from quantum error correction, but are not so large and performant that we can abstract noise away entirely.

Some works on early fault tolerance consider other limitations besides size and error rate that might impose limits

on QEC [13]. For example, QEC is designed to suppress uncorrelated errors, but early experimental evidence with superconducting qubits has demonstrated that large correlated errors can arise from cosmic ray impact events [1,99]. Experimental progress has shown that this effect can be suppressed [100,101], although unknown mechanisms still appear to set limitations on the performance of QEC in state-of-the-art experiments [4]. In this work, we generally make the assumption that scalable QEC is possible and that any confounding factors will be addressed by scientific and engineering progress. However, it is possible that the early fault-tolerant era will be prolonged by unforeseen difficulties, in which case the techniques we explore may prove especially useful.

There is a growing body of work focused on understanding and maximizing the power of early fault-tolerant quantum computers. One frequently taken approach is to assume that some combination of techniques will allow us to execute much larger circuits than we can on the noisy devices of today, but that practical concerns will still impose limits on the number of qubits and the circuit size. Works in this direction focus on applications, usually picking a simple computational task and exploring different design choices in order to minimize the resources required to perform the task in the absence of error. The focus on early fault tolerance in these works appears mainly in the choice of the computational task, which might be a toy example designed to demonstrate algorithmic primitives [30] or a scientific application just beyond the reach of classical computation [12,29]. When the constant factors are taken into account, and when applicable (finite) target precisions are fixed, it is common to find that the simpler methods with worse asymptotic scaling (such as Trotterized time evolution) are superior to their asymptotically optimal counterparts (such as quantum signal processing) [12,29].

Other works have focused on developing new algorithmic tools for the setting in which we have large but finite upper bounds on the maximum circuit size [14,17–22,32–34,78]. These algorithms focus on reducing the number of gates, the circuit depth, or the number of ancillary qubits by classically combining the results from multiple separate circuit executions. Underlying many of these approaches is a trade-off between using deeper circuits to estimate some quantity at the Heisenberg limit and using more repetitions of a shallow circuit to estimate the same quantity at the shot-noise limit. For example, Refs. [17,22,78] all allow for ground-state energy using a total run-time that scales with the precision ϵ as $\epsilon^{-\alpha}$ for an $\alpha \in [1, 2]$ that decreases as the maximum circuit depth is increased. Other works make a related trade-off, choosing between running many copies of a shorter circuit and postselecting on observing some rare event, or obtaining a quadratic speedup with amplitude amplification and the deeper circuits that it requires [21,34].

1. Combining quantum error mitigation and quantum error correction

Early fault-tolerant quantum computers will not be able to arbitrarily suppress error rates with a negligible resource overhead, which makes it desirable to consider computations that can tolerate a nonzero error rate. This can be achieved either by developing algorithms that have some natural robustness to error [15,25,102] or by augmenting error correction with quantum EM [26,35,36,103]. Years of work in the noisy intermediate-scale quantum (NISQ) era has produced a large body of work on algorithms robust to error and we do not attempt to review that literature here. For similar reasons, we limit our discussion of error mitigation to those works that have explicitly explored the combination of quantum error correction and error mitigation. Broadly, error mitigation allows us to estimate noise-free expectation values at the expense of increasing the overall computation time exponentially in the number of expected errors [37]. For example, given a known noise model, PEC lets us express the noise-free expectation value of an observable as a quasiprobability distribution over quantities sampled from modified versions of the original circuit [44]. In this case, the exponential overhead manifests as an increased variance or, equivalently, as a larger number of samples required to estimate an expectation value to a fixed precision.

The combination of quantum error mitigation and quantum error correction can take a number of different forms. Motivated by the fact that fault-tolerant non-Clifford gates are particularly costly to implement, some works have focused on combining error-corrected Clifford gates with noisy non-Clifford gates [27,35,36]. By using error mitigation to handle the noise in the non-Clifford gates, such schemes avoid the difficulties and overheads of, e.g., magic state distillation. Other works have explored the use of error mitigation for both Clifford and non-Clifford gates, potentially allowing the use of lower code distances and fewer qubits. In some cases, researchers have explicitly considered the optimization of quantum error-correction parameters for use in conjunction with error mitigation [13,26,36].

To develop some intuition, we illustrate a toy example of how QEC and EM can be combined in Fig. 4. In Fig. 4(a), we estimate the depth of the largest circuit that one could perform on 100 logical qubits for a variety of physical error rates and numbers of (physical) qubits. In this model, we demand an “effective sampling rate” of one sample per minute. In other words, the combination of error correction and error mitigation should emulate the effect of an idealized error-free quantum computer that executes the circuit and performs a measurement of a desired observable once per minute. We approximate the effective time per sample using the expression $N_{\text{parallel}}^{-1} D t_{\text{gate}} \gamma^2$, where N_{parallel} is the number of copies of the computation that we can perform in parallel given our physical resources,

D is the circuit depth, t_{gate} is the time per gate, and γ^2 is the overhead required to mitigate the logical errors using PEC. We optimize over the choice of whether or not to use error correction and the code distance d , finding the choice that allows for the largest D such that the effective time per sample is less than one minute.

For simplicity, in this example we assume a circuit composed entirely of controlled-NOT (CNOT) gates executed on a superconducting quantum processor similar to the one described in Ref. [1]. We model the error-suppression factor of the surface code using the expression

$$\Lambda = \frac{0.01}{p_{\text{phys}}}, \quad (2)$$

where 0.01 is an estimate of the surface-code threshold and p_{phys} is a single parameter that describes the error rate of the physical gate and measurement operations [9]. This means that we can estimate the error per surface-code cycle as $0.1 \Lambda^{-(d+1)/2}$, where the constant 0.1 is an estimate of the true prefactor. We can perform CNOT gates using lattice surgery using two qubits operating for $3d$ surface-code cycles, so we can therefore estimate the logical error probability for a gate as

$$p_{\text{logical}} = \begin{cases} p_{\text{phys}}, & \text{if } d = 1, \\ 1 - \left(1 - 0.1 \frac{0.01}{p_{\text{phys}}}^{-(d+1)/2}\right)^{6d}, & \text{if } d > 1, \end{cases} \quad (3)$$

using the convention that $d = 1$ indicates that we are not using QEC at all. Using a physical gate time of 20 ns and a surface-code cycle time of 970 ns [1], we therefore have

$$t_{\text{gate}} = \begin{cases} 20 \times 10^{-9}, & \text{if } d = 1, \\ 3 \times d \times 970 \times 10^{-9}, & \text{if } d > 1. \end{cases} \quad (4)$$

A logical qubit in the surface code requires $2d^2 - 1$ physical qubits, so we can execute

$$N_{\text{parallel}} = \left\lfloor \frac{N_{\text{phys}}}{100(2d^2 - 1)} \right\rfloor \quad (5)$$

copies of a 100-logical-qubit computation in parallel. Treating p_{logical} as the probability of applying a two-qubit depolarizing channel, we can follow Ref. [44], finding that the overhead incurred by applying PEC is given by

$$\gamma^2 = \left(\frac{1 + 7/8 p_{\text{logical}}}{1 - p_{\text{logical}}} \right)^{2G}, \quad (6)$$

where $G = 49.5D$ is the number of noisy gates in the circuit. (While the specifics will change with the error model

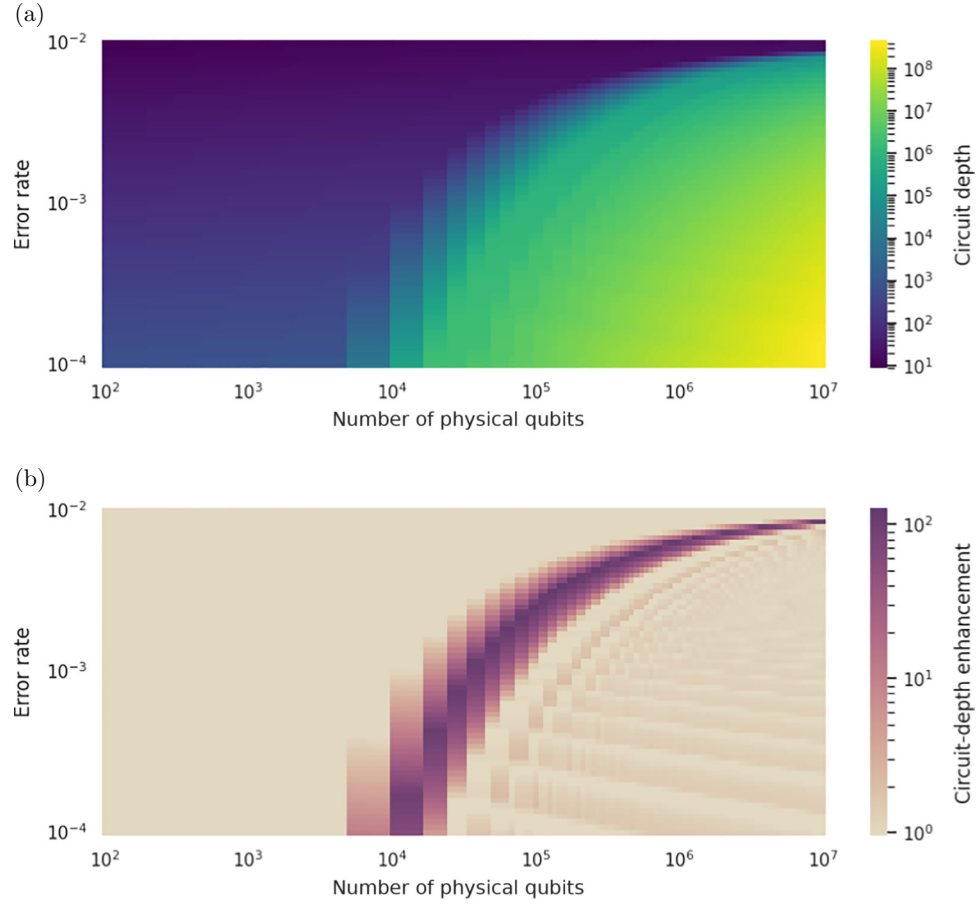


FIG. 4. (a) The maximum circuit depth for a circuit with 100 logical qubits and an effective sampling rate of one sample per minute over a variety of physical error rates and qubit numbers. We assume a circuit composed entirely of controlled-NOT (CNOT) gates, a raw two-qubit gate time of 20 ns, and a surface-code cycle time of 970 ns. Errors are mitigated using probabilistic error cancellation (PEC) or by a combination of surface-code quantum error correction (QEC) and PEC. We maximize the achievable depth by optimizing over the choice of whether or not to use error correction, and the code distance. (b) The increase in circuit depth allowed by this combination of QEC and EM, when compared with a protocol that either uses EM or QEC (with QEC parameters chosen such that the error probability throughout the circuit is less than 1%; 100 logical qubits). The existence of the dark band reveals a regime in which this increase is substantial.

and circuit, the goal of this section is illustrate the consequences of combining error mitigation and correction in a general setting.)

In this simplified model, we can attempt to quantify the benefit of combining quantum error mitigation and quantum error correction. To do so, we perform a second calculation to determine the maximum depth achievable when either (i) using physical gates and error mitigation alone or (ii) using QEC alone. For (i), we calculate the maximum depth as above, restricting $d = 1$. For (ii), we calculate p_{logical} as above and choose the largest D such that the overall probability of an error in the circuit is less than 1%, i.e.,

$$D_{\text{QEC max}} = \max \left\{ D : 1 - (1 - p_{\text{logical}})^{49.5D} \leq .01 \right\}. \quad (7)$$

In Fig. 4(b), we plot the ratio of the circuit depth achievable with the combination of EM and QEC compared to the circuit depth achievable using either technique alone. We see that there is a regime on the boundary between NISQ and full fault tolerance in which the combination of QEC and EM allows for a substantial enhancement in the circuit depth.

2. Circuit division in early fault tolerance

In the previous section we explored the ability of quantum error mitigation to boost the maximum depth allowed in a (logical) quantum circuit when being executed on a fixed number of physical qubits and a fixed error rate. However, this still bounds the circuit depths available to quantum algorithmists on a fixed quantum device to some D_{max} . In principle one can consider running multiple quantum circuits j of depth $D_j < D_{\text{max}}$, given access

to a classical postprocessing method that is robust against the failure of some of these circuits. Perhaps surprisingly, such circuits exist for many problems. The main contribution of our paper is to explore this “circuit division” in the context of on such problem class—quantum phase estimation (QPE)—to combine this with optimized error-mitigation techniques and compile the resulting algorithms to a realistic physical implementation.

Before introducing the relevant background on QPE and explaining our results in detail, let us roughly estimate the potential benefit of circuit division in an early fault-tolerant setting.

Definition 1. A circuit division of a quantum algorithm $\mathcal{A}$ that requires a circuit of G gates corresponds to a set of M “subcircuits” of length $\leq G/R$ and a classical postprocessing routine which can be applied to the subcircuits to solve the problem targeted by $\mathcal{A}$.

We target phase-estimation routines, for which (under some problem constraints [98]) this division is allowed at a cost $M \sim R^2$. For a circuit with G gates, we can approximate the overall probability for the circuit to execute without error as $1 - p_{\text{circuit}} = 1 - (1 - p_{\text{logical}})^G \approx e^{-Gp_{\text{logical}}}$, where p_{logical} is the error rate per gate. Each logical qubit in the surface code requires $2d^2 + 1$ physical qubits (not accounting for overheads from routing or magic state distillation), where the code distance d is an odd number that we can choose freely. For a given physical error rate p_{phys} , we can approximate the error-suppression factor of the surface code Λ as in Eq. (2), which implies an error per logical qubit per cycle proportional to $\Lambda^{-(d+1)/2}$. Without specifying a particular choice of gate and compilation into surface-code operations, we can still make the approximation that $p_{\text{logical}} \propto d\Lambda^{-(d+1)/2}$ (since most gates require a number of cycles proportional to d and the error rates involved are small). The constant of proportionality is determined by the details of the error model and the particular choice of gates, but we can set it to 1 to obtain a qualitative understanding, yielding

$$1 - p_{\text{circuit}} \approx \exp(-Gd\Lambda^{-(d+1)/2}). \quad (8)$$

This exponential decay puts a cutoff on the circuit sizes that we can accept before the failure probability grows too large.

A protocol that replaces a single execution of a circuit with G gates by R^2 executions of circuits with G/R gates increases the overall time required to obtain a solution by a factor of R . If we demand a constant error rate per circuit and neglect logarithmic terms, then Eq. (8) suggests that we should take $G \propto \Lambda^{(d+1)/2}$. To obtain a concrete estimate, let us consider $\Lambda = 10$ and $G = 10^6$, which implies that we need a code distance of $d = 11$. If we let d' be the code distance required for a circuit with G/R gates, then

we have $d - d' \approx 2 \log_{\Lambda} R$. The number of physical qubits required per logical qubit is approximately proportional to d^2 , so taking $N'_{\text{phys}} \propto (d')^2$ and $N_{\text{phys}} \propto d^2$, we can approximate the savings in terms of the number of physical qubits as

$$\frac{N'_{\text{phys}}}{N_{\text{phys}}} \approx 1 - \frac{4 \log_{\Lambda} R}{d} + \frac{4 \log_{\Lambda}^2 R}{d^2}. \quad (9)$$

If these approximations hold, then we could achieve a factor of 2 reduction in the number of physical qubits with $R \approx 40$, but it would appear that a factor of 10 reduction would require R to be unfeasibly large. While the model of errors that we have considered above is oversimplified, we shall show that a more detailed analysis yields a qualitatively similar conclusion.

B. Background on quantum phase estimation

The term “quantum phase estimation” (QPE) refers to a family of computational problems related to estimating eigenphases ϕ_j of a unitary operator U , $U|\phi_j\rangle = e^{i\phi_j}|\phi_j\rangle$, to some target precision ϵ . In this work, we target the estimation of a single specific eigenvalue ϕ ; e.g., if $U = e^{iHt}$ for a Hamiltonian H , one might target the lowest eigenvalue of H . (Note that much recent work on QPE considers the estimation of multiple eigenvalues; see, e.g., Refs. [20,63–65,97].) Following common practices in metrology [6], we define “precision” to mean a bound on the Holevo error of the error in our estimator $\tilde{\phi}$:

$$\epsilon = \sqrt{\mathbb{E}[4 \sin^2(\frac{\tilde{\phi} - \phi}{2})]} \quad (10)$$

(as opposed to the computer science literature [59], where one typically requires $P(|\phi - \tilde{\phi}| \leq \epsilon) \geq (1 - p)$ for fixed p). This metrological definition is slightly stricter [62] as it requires bounding the tail of the distribution of the estimator $\tilde{\phi}_j$.

The problem of determining ground states is known to be hard even with a quantum computer [i.e., quantum Merlin Arthur (QMA) hard] for even 2-local Hamiltonians [104]), but it reduces to a problem solvable by a quantum computer [i.e., one in bounded-error quantum polynomial time (BQP)] when one assumes access to an initial state $|\psi\rangle$ which has large overlap $\langle\psi|\phi\rangle$ with the target eigenstate [105,106]. In this work, we further require that our target eigenvalue ϕ is separated from other eigenvalues ϕ_j (at least those for which $|\langle\phi_j|\psi\rangle| > 0$) by a gap $\min_j |\phi - \phi_j| =: \Delta \gg \epsilon$. This is necessary by information-theoretic constraints [98] which set Δ^{-1} as a lower bound on the required evolution by controlled- U in order to resolve an eigenvalue (without an overhead that grows exponentially in the number of nearby eigenvalues). This in turn places an upper bound on the amount by which

we can divide a circuit, which will later coincide with a minimum fidelity requirement on a noisy implementation of U . Under this assumption, one can project the starting state $|\psi\rangle$ onto the eigenstate $|\phi\rangle$ at a cost proportional to $\Delta^{-1}|\langle\psi|\phi\rangle|^{-1}$ [96,97]. In this work we absorb this cost onto the cost of initial state preparation.

Definition 2 (QPE for single eigenstates). Let U be a unitary operator with eigendecomposition $U|\phi_j\rangle = e^{i\phi_j}|\phi_j\rangle$. Assume access to a preparation unitary $V_{|\phi\rangle}$ for an eigenstate $|\phi\rangle$ of U and access to a (possibly noisy) implementation $\tilde{U}$ of controlled- U (U_c), and fix some $\epsilon_t > 0$. The QPE problem is to produce an estimator $\tilde{\phi}$ of ϕ with Holevo error $\epsilon \leq \epsilon_t$. We say that the (oracular) cost $\mathcal{T}_{\text{tot}}$ of solving the QPE problem is the total number of applications of $\tilde{U}$ required to implement the estimator. $\mathcal{T}$ is the maximum number of applications of $\tilde{U}$ in a single quantum circuit.

In this section, we review the well-established algorithms to solve the noiseless case of this problem (in preparation for adding noise in the remainder of the text).

Definition 3 (Noiseless QPE for single eigenstates). The problem in Definition 2 in the absence of noise in the quantum circuit, i.e., $\tilde{U} = U_c^\dagger(\cdot)U_c$.

Existing quantum algorithms for phase estimation can be divided into two classes: *quantum-Fourier-transform- (QFT) based* methods (also called *parallel* or *entanglement-based*) and *single-control based* methods (also called *sequential*, *iterative*, or *single-control*). In the first class, the phase estimate is obtained from a single run of a circuit with multiple control qubits. Improving the precision requires increasing the dimension of the control register and the circuit depth. In the second class, the phase is extracted by classically processing expectation values of multiple simpler circuits. Better precision can be achieved either by increasing the depth of the circuits or the number of samples used to estimate the expectation values. We describe both classes in more details in the following two sections.

1. QFT-based QPE

In this approach we obtain the first n bits of the binary expansion of the phase from a single measurement of an additional n -qubit register. The circuit consists of three stages: first, a probe state is prepared; then, the register is used to control the U ; and finally we use QFT. Using n control qubits yields precision of $\epsilon = \mathcal{O}(2^{-n})$ [59].

Possibly the most well-known variant of phase estimation is that presented in Nielsen and Chuang [59], based on previous work from Ref. [107]; this is sometimes known in the field as “textbook” QPE. This algorithm uses

two registers, a $\log_2(K)$ -qubit control register and a system register. The first step of the algorithm is to prepare a “probe state” on the control register (in the textbook algorithm, this is done by the Hadamard transform, preparing $1/\sqrt{k} \sum_k |k\rangle$). The second step is to apply a repeated time-evolution oracle U^k on the system register, with k being controlled by the control register state. Finally, an inverse quantum Fourier transform is applied before measuring the control register in a computational basis. The output of this measurement is a binary expansion of the phase estimate.

It has been shown that, in the absence of noise, the estimator constructed following this algorithm is optimal in terms of accuracy when there is no prior information on the phase to estimate [60]. The probe state that optimizes the Holevo variance of the estimator, however, is not the uniform superposition $1/\sqrt{K} \sum_{k=0}^{K-1} |k\rangle$ but rather the state

$$|s_K\rangle = \sqrt{\frac{2}{K+1}} \sum_{j=0}^{K-1} \sin\left(\frac{j+1}{K+1}\pi\right) |j\rangle, \quad (11)$$

which can be prepared with cost $\mathcal{O}(\log K)$ [6] (see Fig. 5). The Holevo error of this estimator is $\tan(\pi/(K+1)) \approx \pi/K$ [108]. This algorithm has been the most utilized in fault-tolerant application compilation research due to this optimality [6,7,109].

Algorithm 1. [Single circuit sine-state phase-estimation algorithm] Input: target precision ϵ_t , oracle access to initial state $|\phi\rangle$ and cU .

- (1) Let $n = \lceil \log_2(\pi \arctan(\epsilon)^{-1} - 2) \rceil$, $K = 2^n$.
- (2) Run SinQPE circuit (Fig. 5) with control dimension K to obtain x .
- (3) Output $\tilde{\phi} = 2\pi(x/K)$.

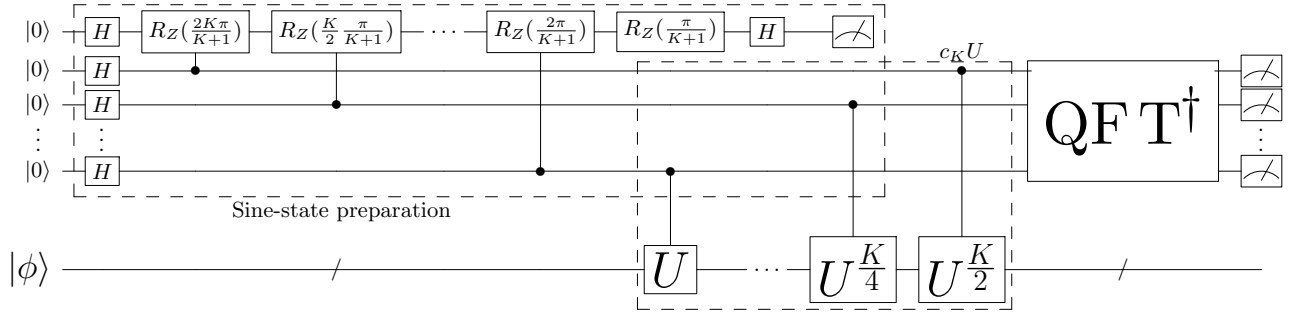
[Note that it is also possible to choose $K = \lceil \arctan(\pi/\epsilon_t - 2) \rceil \neq 2^n$. This is discussed later in Fig. 7, where we explicitly construct the sine-state QPE circuit for general control dimension $K \in \mathbb{N}$ with a log-log overhead.]

We distinguish here between the “target precision” ϵ_t (which is the number given to the algorithm that defines the input parameter Δ) and the actual precision ϵ of the algorithm. As we will see in Sec. IV C, these parameters are no longer the same in the presence of noise.

Theorem 4. [108] Algorithm 1 with $\epsilon_t = \epsilon$ solves Problem 3 with cost $\mathcal{T}_{\text{tot}} = \mathcal{T} = K - 1 \approx \pi/\epsilon$, the lowest possible $\mathcal{T}_{\text{tot}}$.

2. Single-control QPE

The $\lceil \log(\epsilon) \rceil$ -qubit control register is not essential for a quantum speedup in quantum phase estimation. This has



This circuit produces a random variable $x \in 0, 1, \dots, K-1$ with probability distribution

$$P_{K-1}^{(\text{SinQPE})}(x|\phi) = \frac{\sin^2 \frac{\pi}{K+1}}{K(K+1)} \frac{1 + \cos[(K+1)(\phi - 2\pi \frac{x}{K})]}{(\cos(\phi - 2\pi \frac{x}{K}) - \cos \frac{\pi}{K+1})^2}.$$

FIG. 5. A sine-state QPE circuit with control dimension K . First, the state $|s_K\rangle$ is prepared on the control register of $\lceil \log_2 K \rceil$ qubits (conditional on the ancillary qubit measurement yielding 1) and the system register is prepared in state $|\phi\rangle$. Then, controlled unitary $c_K U = \sum_{k=0}^{K-1} |k\rangle \langle k| \otimes U^k$ is applied. Finally, $\text{QFT}^\dagger$ is applied on the control register and the register is measured in the computational basis to obtain a bit string x . The oracular cost is $K-1$.

been known since its conception; Kitaev's original presentation of quantum phase estimation [58] used only a single control qubit, a semiclassical quantum Fourier transform using only one control qubit was presented in Ref. [110] around the same time, and though Shor's original algorithm [111] used a control register, this was simplified shortly after [112–115]. These original algorithms relied on performing the QFT sequentially; reading bits of ϕ one by one via the single control qubit while partially projecting the system register into the corresponding eigenstate $|\phi\rangle$. This is unnecessary when $|\phi\rangle$ is already prepared in an eigenstate, in which case the measurement does not affect the system register. Discarding and repreparing the system register separates the algorithm into classical postprocessing of a series of single-qubit Hadamard tests, which can be advantageous in the presence of noise. However, as an optimal probe state is no longer prepared, these single-control QPE algorithms do not obtain the Heisenberg limit [116]. (This is not an issue for NP algorithms such as factoring, as one can classically confirm the result and repeat until success.) This problem has been solved up to a $\log^*$ factor by the maximum-likelihood algorithm of Ref. [117] and later up to a constant factor by the RPE algorithm of Ref. [62] and the Bayesian phase-estimation algorithm of Ref. [118]. The RPE algorithm, which repeats lower-order Hadamard tests multiple times to increase the confidence in more significant bits of the phase ϕ , has been further optimized in Refs. [72,119]. The latter paper has provided an analytical optimization of the hyperparameters of the RPE algorithm that we will use in this work, proving a separation of around a factor of 25 from the strict Heisenberg limit $\mathcal{T}_{\text{tot}} = \pi/\epsilon$. Indeed, this limit cannot be achieved with

only a single control qubit; it has been shown in Ref. [61] that this requires two control qubits (and can be achieved with such).

The replacement of the control register by a single qubit and the circuit-depth reduction from going to a single qubit have generated much interest in single-control QPE methods for NISQ or early fault-tolerant research. This has been made stronger by the demonstration that one can remove control qubits in some cases [66], which leads to a natural error-mitigation strategy via verification of the initial state [50]. However, the restriction to start with an eigenstate is impractical for quantum simulation applications, unless one allows access to state-preparation methods such as those in Refs. [96,97] (which require additional control registers). In Ref. [63], one of the authors has demonstrated (a) that single-control methods do not require an initial state and (b) that one has a freedom of choice to trade between shorter QPE circuits and more repetitions (with a quadratic overhead as one is reduced to the sampling-noise limit). This freedom of choice is ultimately bounded by the gap between eigenenergies, which has been shown earlier in Ref. [98]; the problem of phase estimation of a continuous (or near-continuous) spectrum has been formalized as the quantum eigenvalue estimation problem in Ref. [64]. Further work based on Bayesian [120] and integral transform [121] among other methods has appeared; however, the Heisenberg limit was not achieved with single-control methods without eigenstate access until Ref. [20] and separately in Ref. [65]. This has been optimized further in Refs. [17,22,32].

For future use, we now state the RPE algorithm as taken from Ref. [72].

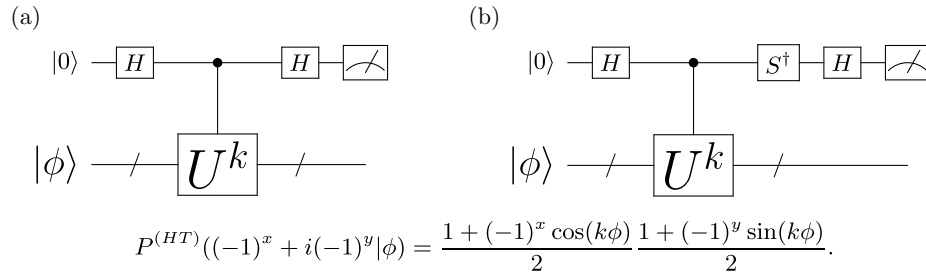


FIG. 6. Hadamard test circuits with exponent k . First, a $|+\rangle$ state is prepared on the control qubit and state $|\phi\rangle$ is prepared on the system register. Then, cU is applied k times. Finally, the control qubit is measured in the X basis (a) to obtain a single bit x . The procedure is repeated with for measurement in the Y basis (b), to obtain another bit y . The output is defined as $Z = (-1)^x + i(-1)^y$. The oracular cost is $2k$. This circuit produces a random variable $Z \in \{\pm 1 \pm i\}$ with probability distribution.

Algorithm 2. [Robust phase estimation (RPE)] Input: number of orders J , vector of numbers of samples $\vec{M} \in \mathbb{N}^J$.

For $j = 0, 1, 2, 3, \dots, J - 1$:

- (1) Fix $k = 2^j$.
- (2) Run M_j repetitions of the circuits of Fig. 6 with exponent k to obtain a set of bits $\{Z_i\}$.
- (3) Calculate the average $\bar{Z}$ of measurement results $\{Z_i\}$.
- (4) Compute $\tilde{\theta}^{(j)} = \text{Arg}[\bar{Z}] \in [0, 2\pi)$.
- (5) If $j = 0$, set $\tilde{\phi}^{(0)} = \tilde{\theta}^{(0)}$. Else, set $\tilde{\phi}^{(j)}$ to be the unique value in the interval $[\tilde{\phi}^{(j-1)} - \pi/k, \tilde{\phi}^{(j-1)} + \pi/k)$ (with periodic boundaries) such that

$$k \tilde{\phi}^{(j)} = \tilde{\theta}^{(j)} \pmod{2\pi}. \quad (12)$$

Return $\tilde{\phi} = \tilde{\phi}^{(J-1)}$.

In the above algorithm we do not fix the values of the maximum order J , nor the number of repeat measurements at each order M_j . This is because these numbers will change in the presence of global depolarizing noise in the next section. However, in the noiseless setting, near-optimal choices of M_j and J are known due to Ref. [62].

Algorithm 3. [RPE noiseless hyperparameters] Input: target precision Δ , metaparameters α, β . Run Algorithm 2 with $J = \lceil \log_2(1/\Delta) \rceil$, and $M_j = \alpha(J - j - 1) + \beta$.

It has been proven in Ref. [62] that for a range of α and β , RPE achieves the Heisenberg limit up to a constant multiplicative factor; $\mathcal{T}_{\text{tot}} \leq c\epsilon^{-1}$. Near-optimal values for these constant factors have been found and bounds tightened in Ref. [72], yielding the following result.

Theorem 5. [72] Fix the target precision $\epsilon_t > 0$. With input parameters $\Delta = 0.409\epsilon_t$, $\alpha = 4.0835$ and $\beta = 11$, RPE solves Definition 3 for $\epsilon_t = \epsilon$ with oracular cost $\mathcal{T}_{\text{tot}} \leq 24.26\pi\epsilon^{-1}$ and maximal depth $\mathcal{T}_{\text{max}} \leq 2\Delta^{-1} \leq 5\epsilon^{-1}$.

Note again the distinction between the input parameter ϵ_t and the noise bound ϵ that has been made in Algorithm 1. The above result is an analytical bound that is not perfectly tight; numerically, we find that with the chosen parameters, RPE achieves error $\epsilon \approx 5\pi/\mathcal{T}_{\text{tot}}$.

3. Information-theoretic bounds

To benchmark the algorithms for Problem 2, it is useful to know what lower bounds exist. Every phase-estimation algorithm has some distribution $P^{(\text{alg})}(\tilde{\phi}|\phi)$ of outputs $\tilde{\phi}$ depending on the true phase ϕ . We are interested in the average Holevo error

$$\begin{aligned} \epsilon^{(\text{alg})}(\mathcal{T}_{\text{tot}}) &= \sqrt{\int_0^{2\phi} \frac{d\phi}{2\pi} \int_0^{2\pi} d\tilde{\phi} P^{(\text{alg})}(\tilde{\phi}|\phi) 4 \sin^2\left(\frac{\tilde{\phi} - \phi}{2}\right)}. \end{aligned} \quad (13)$$

We want to optimize this and find the minimum error that can be achieved at a fixed cost $\mathcal{T}_{\text{tot}}$, but this is not practical as we do not have a simple form of $P^{(\text{alg})}$ for an arbitrary algorithm. Instead, one can fix a choice of quantum circuits (e.g., repetitions of SinQPE with different control dimensions K or repetitions of single Hadamard tests on U^k for different k) and bound the performance of any classical algorithm on the classical output distributions $\vec{x}$ of these circuits. The Cramer-Rao theorem [122,123], bounds the Holevo from below by the inverse of the Fisher information,

$$\epsilon^2 \geq \frac{1}{\mathcal{I}(\phi)}, \quad \mathcal{I}(\phi) = \int dx P(\vec{x}|\phi) \left[\frac{d \log P(\vec{x}|\phi)}{d\phi} \right]^2. \quad (14)$$

For the Hadamard test, it turns out that $\mathcal{I}(\phi)$ is independent of ϕ [65]; however, this is not true for SinQPE. To compensate for this, it is common to take the Bayesian Cramer-Rao bound [122,124], where the Fisher information is replaced

by the average over all possible phases,

$$\epsilon^2 \geq \frac{1}{\bar{\mathcal{I}}}, \quad \bar{\mathcal{I}} = \int_0^{2\pi} \frac{d\phi}{2\pi} \mathcal{I}(\phi). \quad (15)$$

We will invoke this definition and the Cramer-Rao bound for fixed probability distributions later in this work, in order to bound the error in a QPE algorithm from above. As the Fisher information is additive, we can use invoke these results on a single circuit at a time and add their results. We will find that for all our considered algorithms, the lowest cost is achieved by running an increasing number M of circuits with depth $\mathcal{T}_*$ that maximizes the specific Fisher information $\bar{\mathcal{I}}(\mathcal{T})/\mathcal{T}$ [65,88],

$$\epsilon^2 \geq \frac{1}{\bar{\mathcal{I}}(\mathcal{T})M} = \frac{\mathcal{T}}{\bar{\mathcal{I}}(\mathcal{T})\mathcal{T}_{\text{tot}}} \geq \frac{\mathcal{T}_*}{\bar{\mathcal{I}}(\mathcal{T}_*)\mathcal{T}_{\text{tot}}}. \quad (16)$$

In this work we fix our quantum circuits and optimize over the parameters of the probability distributions, but in some cases it is possible to optimize the Fisher information over all choices of measurements on all states generated using a fixed oracle. This yields the channel Fisher information, which characterizes the black box in which the parameter ϕ is encoded. For the noiseless version of QPE, Problem 3, the channel Fisher information can be taken to yield the ultimate bound on the error: the *Heisenberg limit* $\epsilon \geq \pi/\mathcal{T}_{\text{tot}}$ [125]. This is saturated by the probability distribution $P_T^{\text{(SinQPE)}}$ [Eq. (12)] [60]. For a noisy oracle subject to global depolarizing noise with rate γ (see Sec. IV C), the channel Fisher information can be bounded from above, yielding $\epsilon \geq \sqrt{\gamma/\mathcal{T}_{\text{tot}}}$ [65].

C. Quantum phase estimation with global depolarizing noise

Hardware noise in quantum devices can take many forms and some knowledge of the noise model is often required in order to perform error mitigation. In this section we focus on mitigating a simple well-characterized noise model in a QPE experiment; global depolarizing noise with a fixed strength γ . This will allow us to more rigorously analyze and compare the performance of different estimators; we will consider the more general form in Sec. IV E.

Global depolarizing noise (GDN) [49,59,63,126,127] is a stochastic noise model where, with some probability p , a random global unitary is applied to the system, replacing the state of the system with the maximally mixed state. This maps our controlled unitary U_c to the quantum channel $\tilde{\mathcal{U}}_\gamma^{(\text{GDN})}$, defined by

$$\tilde{\mathcal{U}}_\gamma^{(\text{GDN})}[\rho] = e^{-\gamma} U_c \rho U_c^\dagger + (1 - e^{-\gamma}) \rho_{\text{MC}}. \quad (17)$$

This definition is exact, as the global-depolarizing-noise process commutes with any subsequent unitary operations.

This allows us to calculate the effect of repeated GDN channels interleaved by arbitrary operations as

$$\begin{aligned} (\tilde{\mathcal{U}}_\gamma^{(\text{GDN})})^k[\rho] &= e^{-\gamma k} (U_c)^k \rho (U_c^\dagger)^k + (1 - e^{-\gamma k}) \rho_{\text{MC}} \\ &= \tilde{\mathcal{U}}_{k\gamma}^{(\text{GDN})}[\rho]. \end{aligned} \quad (18)$$

We see that this yields an exponential decay in circuit fidelity regardless of the unitary; with probability $1 - e^{-\gamma k}$, the system yields the maximally mixed state and all useful information about the target problem (e.g., in our case information about the phase ϕ) is lost. This makes global depolarizing noise a particularly simple noise model to analyze, though it comes with an exponential decay in fidelity that cannot be mitigated or even error corrected without an overhead that is exponential in k [37]. This is in contrast to, e.g., *local* independent identically distributed (IID) depolarizing noise on individual qubits. These noises do not commute with unitary operations and so a simple form such as Eq. (18) cannot be found, but this noise model allows error correction when error rates are below the fault-tolerant threshold [9].

In this section we optimize and compare two different variants of QPE robust to global depolarizing noise. One of these will be based on the optimized RPE algorithm of Ref. [72], while for the other we develop a robust version of sine-state QPE, using maximum-likelihood estimation. The main outcome of this analysis is a black box that takes in an error ϵ and fixed noise rate γ and tells the number of calls to the oracle $\mathcal{T}_{\text{tot}}$ that would be required to solve the following.

Definition 4 (QPE in the presence of depolarizing noise). The problem in Definition 2 in the presence of global depolarizing noise with a known fixed error rate γ per application of the unitary oracle, i.e., $\tilde{\mathcal{U}} = \tilde{\mathcal{U}}_\gamma^{(\text{GDN})}$.

This will be later used in Sec. IV F to optimize surface-code parameters according to the framework defined in Fig. 1.

1. Single-control QPE in the presence of noise

If we directly apply RPE (Algorithm 3) in the presence of depolarizing noise, the (Holevo) error ϵ_H in the estimate will plateau as the target precision ϵ_t is decreased, in spite of the growing cost. This is because the algorithm requires running ever-deeper circuits, the output of which converges exponentially quickly to uniform noise. In Ref. [119], it has been shown that the RPE algorithm is robust to the probabilities $p_j(Z)$ being corrupted by an arbitrary noise term of strength no greater than a model-dependent number ~ 0.13 . For global depolarizing noise, this holds for all distributions p_j till $j = j_{\text{max}} = \lfloor \log_2[\ln(2/\sqrt{8})/\gamma] \rfloor$, which estimates ϕ with high confidence to error $\epsilon \leq 2\pi/2^j \sim \mathcal{O}(\gamma)$. The algorithm will work with reduced

confidence slightly beyond this point, but will quickly start to fail. To adapt the algorithm to the presence of noise, we can instead stop at some maximum depth around $\mathcal{T}_{\max} \sim 2^{j_{\max}}$ and repeat these last circuits many times to increase precision at the sampling noise limit (as has been suggested in Refs. [65,72,88]). To adapt RPE to GDN, we identify two distinct regimes: first, when $\epsilon \gg \gamma$, in which case the number of samples is rescaled to ensure that the final error is within the required precision and, second, when $\epsilon \lesssim \gamma$, in which case the maximum depth is fixed to some $\mathcal{T}_*$ and most samples are taken at this maximum depth.

As discussed in Sec. IV B 3, the optimal maximal depth $\mathcal{T}_*$ can be quantified by specific Fisher information [65,88]. In the case of the Hadamard test with global depolarizing noise with strength γ , the Fisher information $\bar{\mathcal{I}}_\gamma(\mathcal{T}) = e^{-2\gamma\mathcal{T}}\mathcal{T}^2$ can easily be calculated by substituting the probability distribution from Eq. (13) into Eq. (15). The depth that maximizes $\bar{\mathcal{I}}_\gamma(\mathcal{T})/\mathcal{T}$ is $\mathcal{T}_* = 1/2\gamma$ [65]. Asymptotically, the best performance that we can get is given by Eq. (16)

$$\mathcal{T}_{\text{tot}} \geq \frac{2\mathcal{T}_*}{\bar{\mathcal{I}}_\gamma(\mathcal{T}_*)\epsilon^2} = \frac{4e\gamma}{\epsilon^2}. \quad (19)$$

In practice, this bound will not be saturated, as we need to allocate a number of shots for circuits with lower depths in order to determine the most significant bits of the phase with sufficiently high confidence. In Ref. [72], the authors optimize the number of measurements M_j of Algorithm 2 to account for global depolarizing noise. This results in a well-defined algorithm, the cost of which scales asymptotically as $\mathcal{T}_{\text{tot}} \sim \mathcal{O}(\epsilon^{-2}\gamma)$ (we will later fit the prefactor numerically).

Algorithm 4. [RPE with global depolarizing noise [72]]
Fix a target precision ϵ_t and define the constants

$$J = \lfloor \log_2(1/\epsilon_t) \rfloor, \quad \beta = 11 \quad \text{if } \epsilon_t > \gamma \quad (20)$$

$$J = \lfloor \log_2(1/\gamma) \rfloor, \quad \beta = 11 \frac{\gamma^2}{\epsilon_t^2} \quad \text{if } \epsilon_t < \gamma \quad (21)$$

Run Algorithm 2 with J orders and numbers of samples $M_j = e^{2\gamma k_j} (\alpha(J-j) + C\gamma(2^J - 2^j) + \beta)$, with $\alpha = 4.0835$ and $C = 1.3612$.

Several alternative single-control QPE algorithms to RPE have been proposed for the early fault-tolerant regime [17,20,22] and studied in the presence of GDN [23,25]. In Ref. [18] the costs of surface-code implementations of several algorithms have been analyzed, comparing the performance of various estimators under GDN. The study has found that when the initial state is an eigenstate, as in our case, the total run-time is minimized by either RPE or the quantum complex exponential least-squares algorithm

of Ref. [22]. For simplicity, we choose RPE with the parameters of Ref. [72] as the benchmark against which to compare.

2. Multi-circuit QFT-based QPE

Similar to the single-control case, in the presence of noise a sine-state QPE circuit will lose information as it grows arbitrarily deep. This effect is even more drastic than for RPE, because a single measurement will be used; if we try to apply Algorithm 1 using a noisy oracle, the actual error ϵ will become larger for decreasing target error ϵ_t (instead of plateauing). This can be seen from Eq. (17); the final measurement will return an estimate with the same error as the noiseless algorithm with probability $e^{-\gamma\mathcal{T}}$ and a fully random estimate with error $\mathcal{O}(1)$ with probability $1 - e^{-\gamma\mathcal{T}}$, yielding a Holevo variance

$$\epsilon_H^2(\mathcal{T}, \gamma) = e^{-\gamma\mathcal{T}} \tan^2\left(\frac{\pi}{\mathcal{T} + 2}\right) + 2(1 - e^{-\gamma\mathcal{T}}). \quad (22)$$

The above has a minimum $\min_{\mathcal{T}} \epsilon_H(\mathcal{T}, \gamma) = \sqrt{3} \sqrt[3]{\pi\gamma} + \mathcal{O}(\gamma) \approx 2.54\gamma^{1/3} > 0$ at $\mathcal{T} = \sqrt[3]{\pi^2/\gamma} + \mathcal{O}(\gamma^{1/3})$ for any $\gamma > 0$. Indeed, achieving arbitrary precision using a single QPE circuit (as in Fig. 8) is impossible in the presence of global depolarizing noise. Moreover, naively averaging the outcomes over many circuit runs would result in $\mathcal{T}_{\text{tot}} \propto \gamma^{1/3}\epsilon^{-2}$, as compared to the scaling $\mathcal{T}_{\text{tot}} \propto \gamma\epsilon^{-2}$ in the previous section for RPE. We would hope to recover this scaling.

As in the previous section, to adapt Algorithm 1 to the presence of noise, we want to limit the maximum depth of a circuit and instead run multiple circuits. This requires both defining the best choice of circuits to run and determining a method to combine the results of these circuits to optimize the final precision. We can write down the probability distribution of sine-state QPE in the presence of global depolarizing noise with strength γ as

$$P_{\gamma,\mathcal{T}}^{(\text{SinQPE})}(x|\phi) = e^{-\mathcal{T}\gamma} P_{\mathcal{T}}^{(\text{SinQPE})}(x|\phi) + (1 - e^{-\mathcal{T}\gamma}) \frac{1}{\mathcal{T} + 1}, \quad (23)$$

where $P_{\mathcal{T}}^{(\text{SinQPE})}$ is the distribution of the noiseless circuit [Eq. (12)]. This probability distribution is characterized by average Fisher information

$$\begin{aligned} \bar{\mathcal{I}}_\gamma(\mathcal{T}) &= \int_0^{2\phi} \frac{d\phi}{2\pi} \sum_{x=0}^{\mathcal{T}} P_{\gamma,\mathcal{T}}^{(\text{SinQPE})}(x|\phi) \\ &\times \left[\frac{d \log P_{\gamma,\mathcal{T}}^{(\text{SinQPE})}(x|\phi)}{d\phi} \right]^2, \end{aligned} \quad (24)$$

calculated by substituting the probability distribution $P_{\gamma, T}^{(\text{SinQPE})}$ into Eq. (15). An asymptotically optimal estimator is then to perform maximum-likelihood estimation on the joint noisy distributions $P_{\gamma, T}^{(\text{SinQPE})}(x|\phi)$ taken from different runs. (This is slightly different, but related to the maximum-likelihood estimation performed on single-control QPE data by Refs. [22,117].) The maximum-likelihood routine can be performed classically with classical cost $\mathcal{O}(\epsilon^{-2})$; due to the constant overheads of quantum computation, we assume that this cost is negligible compared to the total cost of the quantum algorithm $\mathcal{T}_{\text{tot}} \sim \mathcal{O}(\epsilon^{-2})$. We numerically find that for the sine-state circuit with GDN, $\arg \max_T \tilde{\mathcal{I}}_\gamma(T)/T \approx \gamma^{-1}$. It remains to choose what to do for larger target errors.

a. Algorithm details. There are three regimes characterized by the target precision ϵ relative to the noise strength γ . When $\epsilon \gg \gamma$, the effects of noise are negligible and the optimal strategy is the same as for the noiseless case—run a single circuit of sufficient depth, such that the variance of the output is $\epsilon_H^2(\mathcal{T}, \gamma) \leq \epsilon^2$. As the depth increases, the output becomes more and more influenced by noise up to a maximal depth $\mathcal{T}_{\text{max}} \sim \mathcal{O}(\gamma^{-1})$, beyond which running larger depths is never efficient. Instead, asymptotically small errors can be achieved by repeating this single circuit multiple times. In this regime the best possible precision can be achieved by processing the circuit samples through maximum-likelihood estimation. Maximum-likelihood estimation is only optimal in the limit of a large number of samples; as a consequence, in the intermediate regime ($\gamma \approx \epsilon$), we will need to run a large enough number of samples for circuits of depth smaller than for the asymptotic regime, effectively interpolating circuit depth and number of shots between the two edge cases.

To optimize our measurement strategy for the intermediate regime, we first choose a maximal depth $\mathcal{T}_1$ for any single experiment. With a small number of samples, the error is characterized by the variance of the circuit output probability distribution. We need to find $\mathcal{T}_1$ such that when the total resources are $\mathcal{T}_1$, the expected error is smaller when running a circuit with depth $\mathcal{T}_1$ once than if the circuit of depth $\mathcal{T}_1/2$ is repeated twice. In the first case, the expected error is simply $\epsilon_H(\mathcal{T}_1, \gamma)$, defined in Eq. (22). In the second case, we get two independent samples, each coming from the target distribution with probability $p = e^{-\gamma \mathcal{T}_1/2}$. There are three possibilities to consider for this second case: if both samples come from the target distribution, the output estimate $\tilde{\phi}$ is their average and the error is $\approx \frac{1}{\sqrt{2}}(\pi/\mathcal{T}_1/2)$; if one of the samples comes from the target distribution, the error is either $\pi/\mathcal{T}_1/2$ or 2; if both samples come from the noise distribution, the output is random and the error is 2. Summing over these possibilities for the second case, we find the threshold depth $\mathcal{T}_1$ by

solving

$$p^2 \frac{\pi^2}{\mathcal{T}_1^2} + (1-p^2) \times 2 \leq p^2 \times \frac{4\pi^2}{2\mathcal{T}_1^2} + 2p(1-p) \times \left(\frac{1}{2} \times \frac{4\pi^2}{\mathcal{T}_1^2} + \frac{1}{2} \times 2 \right) + (1-p)^2 \times 2. \quad (25)$$

One can check that this condition is equivalent to

$$\frac{\pi^2}{\mathcal{T}_1^2} \leq 2 \frac{1-p}{1+3p/4} = 8 \frac{1-e^{-\gamma \mathcal{T}_1/2}}{4-3e^{-\gamma \mathcal{T}_1/2}} = 2\gamma \mathcal{T}_1 + \mathcal{O}(\gamma^2 \mathcal{T}_1^2). \quad (26)$$

To first order in $\gamma \mathcal{T}_1$, we get that $\mathcal{T}_1 = \sqrt[3]{\pi^2/2\gamma} \approx 1.70\gamma^{-1/3}$ (which validates the above expansion for small γ). The corresponding error is $\epsilon_H(\mathcal{T}_1, \gamma) \approx 1.84\gamma^{1/3}$.

We have not yet chosen the depth $\mathcal{T}_{\text{max}}$ for the small-error regime. Again, it is best to choose the depth $\mathcal{T}_{\text{max}} = \mathcal{T}_2$ that maximizes the specific Fisher information $\mathcal{T}_2 = \arg \max_T \tilde{\mathcal{I}}_\gamma(T)/T$. Integrating $\tilde{\mathcal{I}}_\gamma(T)$ in Eq. (24) numerically, we verify that the specific Fisher information is maximized at depth $\mathcal{T}_2 = \gamma^{-1}$, and $\tilde{\mathcal{I}}_\gamma(\mathcal{T}_2) \approx 0.04\gamma^{-2}$. In the limit of many samples M (in practice, we assume that the estimator has converged for $M > M_2 = 100$), the variance of the MLE is $(\tilde{\mathcal{I}}M)^{-1}$, so to achieve $\epsilon \ll \gamma$ the number of samples required is $M = 1/\tilde{\mathcal{I}}_\gamma(\mathcal{T}_2)\epsilon^2$ leading to the total cost $\mathcal{T}_{\text{tot}} = \mathcal{T}_2 M = \mathcal{T}_2/\tilde{\mathcal{I}}_\gamma(\mathcal{T}_2)\epsilon^2$.

It remains to define the algorithm in the intermediate regime to interpolate between depths $\mathcal{T}_1$ and $\mathcal{T}_2$ and numbers of samples 1 and M_2 . This is not easy, because while we know the expected asymptotic performance of the MLE, it is hard to predict what happens when there are few samples. Instead, we construct a model for the error $\epsilon(\mathcal{T}, M, \gamma)$ and then choose $\mathcal{T}$ and M so that the total cost $\mathcal{T}_{\text{tot}} = \mathcal{T}M$ is minimal, while $\epsilon(\mathcal{T}, M, \gamma) \leq \epsilon_t$. We write the estimator error as a sum of two cases:

$$\epsilon^2 \approx p_F \times 2 + (1-p_F) \frac{1}{\tilde{\mathcal{I}}M}. \quad (27)$$

That is, with probability p_F , the estimator fails, the outcome is completely random, and the average square error is $\int_0^{2\pi} \frac{d\tilde{\phi}}{2\pi} 4 \sin^2(\frac{\tilde{\phi}-\phi}{2})^2 = 2$, but with probability $1-p_F$, the estimator succeeds, and the average square error satisfies the Cramer-Rao bound $1/\tilde{\mathcal{I}}M$. The failure probability p_F depends on the number of samples M and the circuit fidelity $F = e^{-\gamma T}$. We conjecture that the probability of failure decreases exponentially with the number of samples $p_F = e^{-\alpha M}$ (motivated by the asymptotic normality of the maximum-likelihood estimator). Furthermore, we choose $\alpha = \gamma T/2$, so that $p_F(M=2) = e^{-\gamma T}$ —we expect this behavior because, when we have only two samples, there is no way to tell which of the two is more likely to be produced by the signal or by the

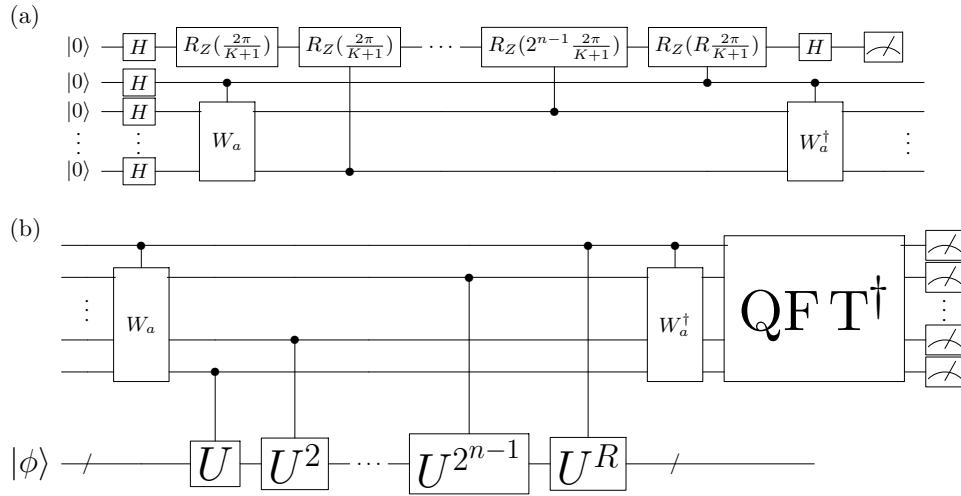


FIG. 7. The implementation of the sine-state QPE circuit for $K \in \mathbb{N}$. Let $n = \lfloor \log_2(K) \rfloor$ and $R = K - 2^n$. If $R = 0$ (where K is a power of 2), then the circuit can be implemented from Fig. 5. Otherwise, the $c_K U$ operation can be realized using a control register of $n + 1$ qubits in the following way, as shown in panel (b). First, the most significant qubit controls an addition gate $W_a = \sum_{j=0}^{2^{n-1}-a-1} |j+a\rangle \langle j|$ for $a = 2^n - R$ in the remaining register (W_a can be constructed with up to n Toffoli gates and as many auxiliary qubits [128,129]) and then R applications of U_c . Then, the remaining register controls a $c_{2^n} U$. Finally, the addition of a is uncomputed. We can apply the same trick to the controlled rotations in the preparation of the sine-state to prepare $|s_K\rangle$ as shown in panel (a). QFT of order K can be realized with cost $\tilde{O}(\log(K))$ [130,131]. The oracular cost is $K - 1$, as in the case of $K = 2^n$ (Fig. 5).

noise; in that case MLE is forced to randomly pick one, the estimator for two samples reduces to that with one sample, and the probability of success is equal to the fidelity $e^{-\gamma T}$. Finally, we choose the threshold precision ϵ_2 , below which the outcome of the optimization is negligibly close to the asymptotic expression. For $\epsilon_t \ll \gamma$, the optimization yields the asymptotically optimal depth $T \sim \gamma^{-1}$ and the number of samples $M \sim \bar{\mathcal{I}}_\gamma(\mathcal{T}_2)^{-1} \epsilon_t^{-2} \sim 20\gamma^2 \epsilon_t^2$. The failure probability $p_F \sim e^{-M/2}$ can be neglected whenever the other term $\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)^{-1} M^{-1}$ dominates and $Me^{-M/2} \ll \bar{\mathcal{I}}_\gamma(\mathcal{T}_2)^{-1} \sim 20\gamma^2$. This is satisfied for $M \geq M_2 = \max(100, \lceil 5 \log \gamma^{-1} \rceil)$ and we set $\epsilon_2 = 1/\sqrt{\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)M_2}$. We note that this threshold is introduced solely to simplify the presentation and avoid unnecessary calculations in the numerical simulations. In principle, one could perform the optimization for $\epsilon_t < \epsilon_2$: the algorithm would still prescribe the optimal depth $T \sim \gamma^{-1}$ and a marginally larger number of samples M to account for the exponentially small failure probability $\sim e^{-M/2}$.

As we want to optimize the performance of the algorithm for relatively small oracle depths, it is worth noting that we are not restricted to powers of 2. In Fig. 7, we show how to generalize the SinQPE circuit of Fig. 5 to a general control dimension $K \in \mathbb{N}$.

Algorithm 5. [Multi-circuit sine-state QPE with GDN]
Input: Target precision ϵ_t , noise strength γ .

- (1) Let the threshold parameters be $T_1 = \lfloor \sqrt[3]{2\pi^2/3\gamma} \rfloor$, $\epsilon_1 = \epsilon_H(\gamma, T_1)$ (see Eq. (22)), and $T_2 = \lfloor \gamma^{-1} \rfloor$

$$M_2 = \max(100, \lceil 5 \log \gamma^{-1} \rceil), \quad \epsilon_2 = 1/\sqrt{\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)M_2} \quad (\text{see Eq. (24)}).$$

- (2) Choose the number of shots M and circuit depth T :

- (a) If $\epsilon_t \geq \epsilon_1$: $T = \lceil \frac{\pi}{\arctan(\epsilon_t)} - 2 \rceil$, $M = 1$
- (b) If $\epsilon_t \leq \epsilon_2$: $T = T_2$, $M = 1/\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)\epsilon_t^2$.
- (c) If $\epsilon_1 < \epsilon_t < \epsilon_2$: $T, M = \arg \min_{T \in \mathbb{N}, M \in \mathbb{N}} (T \times M)$ under the constraint

$$2e^{-\gamma TM/2} + (1 - e^{-\gamma TM/2}) \frac{1}{\bar{\mathcal{I}}_\gamma(T)M} \leq \epsilon_t^2. \quad (28)$$

- (3) Run M shots of the SinQPE circuit [(Fig. 7)] with control dimension $K = T + 1$ to obtain M samples $\{x_j\}$.
- (4) Return $\arg \min_{\tilde{\phi}} L(\{x_j\}, \tilde{\phi})$.

Note that for $\gamma > 10^{-10}$ it is sufficient to fix $M_2 = 100$. As before, we again define this algorithm in terms of a parameter ϵ_t that allows us to fix the circuit parameters, but does not necessarily form a bound on the final Holevo error. However, we find numerically that the true Holevo error bound ϵ lies very close to this target.

b. Performance guarantees. Here we state and prove the formal version of Theorem 1.

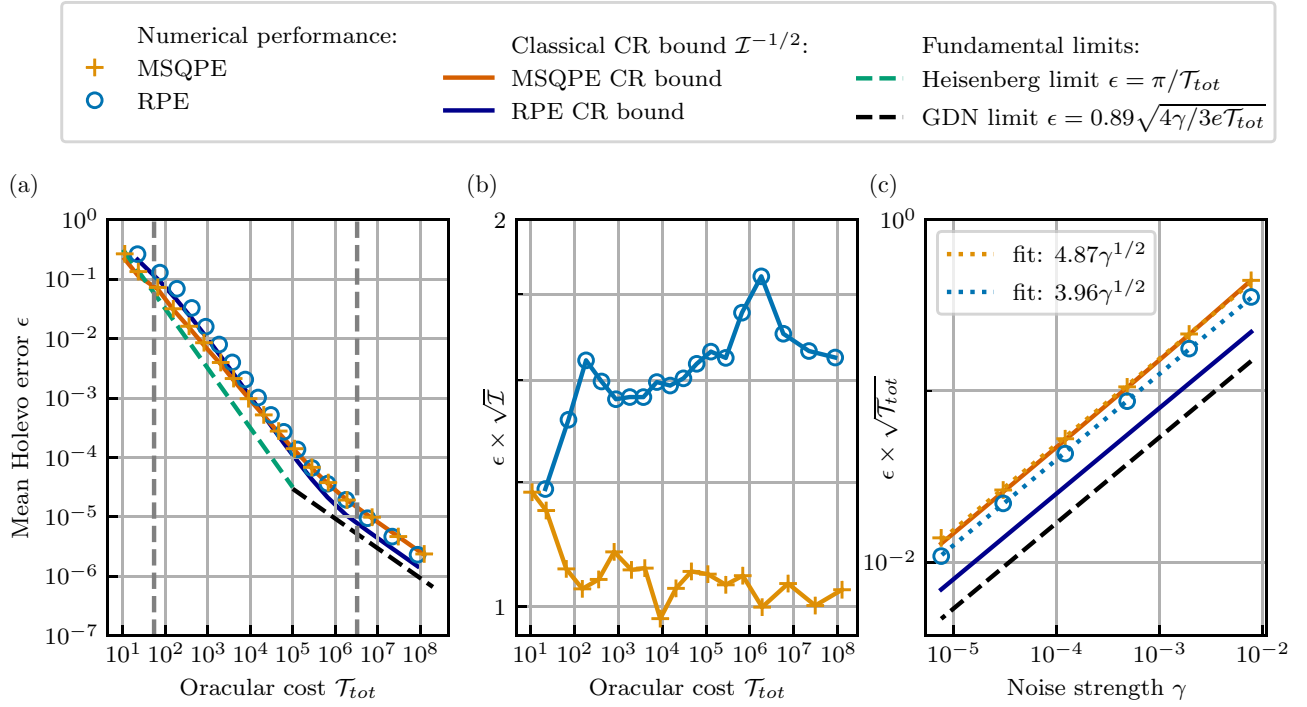


FIG. 8. A comparison of RPE (Algorithm 4) and MSQPE (Algorithm 5) in the presence of global depolarizing noise. (a) The numerical mean Holevo error ϵ as a function of the total cost $\mathcal{T}_{\text{tot}}$ for a fixed noise rate $\gamma = 2^{-15}$. The vertical dashed lines represent $\mathcal{T}_{\text{tot}}(\epsilon_1)$ and $\mathcal{T}_{\text{tot}}(\epsilon_2)$ separating different regimes in Algorithm 5. The solid lines correspond to the Cramer-Rao (CR) bound $\epsilon \geq \mathcal{I}^{-1/2}$, where $\mathcal{I}$ is the total Fisher information of all the quantum circuits used. (b) The numerical error ϵ relative to the minimal error allowed by the Cramer-Rao bound $\mathcal{I}^{-1/2}$. (c) The performance in the limit as $\epsilon \rightarrow 0$ as a function on the noise strength γ . The blue circles and orange pluses are obtained by fitting $\epsilon \propto \mathcal{T}_{\text{tot}}^{1/2}$ to the numerical errors ϵ in the small error regime for different noise rates γ and the solid lines by fitting $\mathcal{I}^{-1/2} \propto \mathcal{T}_{\text{tot}}^{1/2}$. The dotted lines represent the least-squares fit to these points. The error bars in all cases are insignificant.

Theorem 6. Given an initial eigenstate preparation, the total number of uses $\mathcal{T}_{\text{tot}}$ of the unitary required for Algorithm 5 to produce a phase estimate with Holevo error $\epsilon \rightarrow 0$ in the presence of global depolarizing noise with strength $\gamma \rightarrow 0$ depends on the order in which these limits are taken. Specifically:

- (1) $\lim_{\epsilon_t \rightarrow 0} \lim_{\gamma \rightarrow 0} \mathcal{T}_{\text{tot}}\epsilon = \pi$ (i.e., $\gamma \ll \epsilon$)
- (2) $\lim_{\gamma \rightarrow 0} \lim_{\epsilon_t \rightarrow 0} \mathcal{T}_{\text{tot}}\epsilon^2/\gamma = C$ (i.e., $\epsilon \ll \gamma$), where $C \approx 20$ is a constant of proportionality

Proof. We start by proving statement (1), by noting that for $\gamma \rightarrow 0$ we recover the noiseless case and Algorithm 5 behaves identically to Algorithm 1, as described in Theorem 4. Specifically, as $\gamma \rightarrow 0$ for a fixed ϵ_t , Algorithm 5 prescribes a single circuit with depth $\mathcal{T} = \lceil \pi/\arctan(\epsilon_t) - 2 \rceil$ and the total cost is $\mathcal{T}_{\text{tot}} = \mathcal{T}$. The error ϵ is given by the Holevo variance of the noisy distribution, given in Eq. (22):

$$\epsilon = \sqrt{e^{-\gamma\mathcal{T}_{\text{tot}}} \tan^2\left(\frac{\pi}{\mathcal{T}_{\text{tot}} + 2}\right) + 2(1 - e^{-\gamma\mathcal{T}_{\text{tot}}})}. \quad (29)$$

Taking the limit as $\gamma \rightarrow 0$, this becomes the same as in the noiseless case, $\epsilon \xrightarrow{\gamma \rightarrow 0} \tan(\pi/(\mathcal{T}_{\text{tot}} + 2))$. Thus the total cost satisfies

$$\lim_{\epsilon_t \rightarrow 0} \mathcal{T}_{\text{tot}}\epsilon = \lim_{\mathcal{T}_{\text{tot}} \rightarrow \infty} \mathcal{T}_{\text{tot}} \times \tan\left(\frac{\pi}{\mathcal{T}_{\text{tot}} + 2}\right) = \pi. \quad (30)$$

We turn to proving statement (2). When $\epsilon_t \rightarrow 0$ for a fixed noise rate γ , Algorithm 5 uses a large number of samples $M = 1/\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)\epsilon_t^2$ of circuit 5 with depth $\mathcal{T}_2 = \lfloor \gamma^{-1} \rfloor$, so the total cost is $\mathcal{T}_{\text{tot}} = M\mathcal{T}_2$. By the limiting properties of the maximum-likelihood estimator (which saturates the Cramér-Rao bound for large number of samples M), the algorithm will converge to the target precision $\epsilon = \epsilon_t$, with the total cost satisfying

$$\lim_{\epsilon_t \rightarrow 0} \mathcal{T}_{\text{tot}}\epsilon^2 = \frac{\mathcal{T}_2}{\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)}. \quad (31)$$

It remains to calculate the limit $\lim_{\gamma \rightarrow 0} \mathcal{T}_2/\gamma\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)$. For brevity of notation, let $f_T(\Delta\phi)$ be such that the probability

distribution in Eq. (12) can be written as

$$P_{0,T}(x|\phi) = f_T\left(\phi - \frac{2\pi x}{T+1}\right), \quad (32)$$

$$f_T(\Delta\phi) = \frac{\sin^2 \frac{\pi}{T+2}}{(T+1)(T+2)} \frac{1 + \cos[(T+2)\Delta\phi]}{\left[\cos(\Delta\phi) - \cos\left(\frac{\pi}{T+2}\right)\right]^2}. \quad (33)$$

The fidelity of the circuit depth T under GDN with strength γ is $e^{-\gamma T}$ and the associated probability distribution is $P_{\gamma,T}(x|\phi) = e^{-\gamma T} P_{0,T}(x|\phi) + (1 - e^{-\gamma T})/(T+1)$. The Fisher information is by definition [Eq. (15)]

$$\begin{aligned} \bar{\mathcal{I}}_\gamma(T) &= \int_0^{2\pi} \frac{d\phi}{2\pi} \\ &\times \sum_{x=0}^T \frac{e^{-2\gamma T} f_T'^2(\phi - 2\pi \frac{x}{T+1})}{e^{-\gamma T} f_T(\phi - 2\pi \frac{x}{T+1}) + (1 - e^{-\gamma T})/(T+1)} \end{aligned} \quad (34)$$

$$= \sum_{x=0}^T \int_{-\pi-2\pi \frac{x}{T+1}}^{\pi-2\pi \frac{x}{T+1}} \frac{ds}{2\pi} \frac{e^{-2\gamma T} f_T'^2(s)}{e^{-\gamma T} f_T(s) + (1 - e^{-\gamma T})/(T+1)} \quad (35)$$

$$= (T+1) \int_{-\pi}^{\pi} \frac{ds}{2\pi} \frac{e^{-2\gamma T} f_T'^2(s)}{e^{-\gamma T} f_T(s) + (1 - e^{-\gamma T})/(T+1)} \quad (36)$$

$$= (T+1) e^{-\gamma T} \int_{-\pi}^{\pi} \frac{ds}{2\pi} \frac{f_T'^2(s)}{f_T(s) + (e^{\gamma T} - 1)/(T+1)}, \quad (37)$$

where we have used the periodicity of f_T to remove the sum and dependence on ϕ . We can substitute $s \mapsto (T+2)s$ to rewrite this as

$$\begin{aligned} \bar{\mathcal{I}}_\gamma(T) &= e^{-\gamma T} \times (T+2)(T+1) \\ &\times \int_{-(T+2)\pi}^{(T+2)\pi} \frac{1}{2\pi} \frac{(T+2)^{-2} f_T'^2(\frac{s}{T+2}) ds}{f_T(\frac{s}{T+2}) + (e^{\gamma T} - 1)/(T+1)}. \end{aligned} \quad (38)$$

Using L'Hôpital's rule, we can show that in the limit $T \rightarrow \infty$, the function f_T converges pointwise to

$$f(s) = \lim_{T \rightarrow \infty} f_T\left(\frac{s}{T+2}\right) = (2\pi)^2 \frac{1 + \cos s}{(\pi^2 - s^2)^2}, \quad (39)$$

and the derivative converges as

$$\frac{1}{T+2} f_T' \left(\frac{s}{T+2} \right) \xrightarrow{T \rightarrow \infty} f'(s). \quad (40)$$

Therefore, in the limit as $\gamma \rightarrow 0$, where the fidelity is $e^{\gamma T_2} \rightarrow e$ the integrand converges pointwise to

$$j(s) = \frac{1}{2\pi} \frac{f'^2(s)}{f(s)} \quad (41)$$

$$\begin{aligned} &= \frac{1}{(s^2 - \pi^2)^2} \left[(1 - \cos s) + \frac{16s^2}{(s^2 - \pi^2)^2} (1 + \cos s) \right. \\ &\quad \left. + \frac{8s}{s^2 - \pi^2} \sin s \right], \end{aligned} \quad (42)$$

which we can bound by bounding all the trigonometric functions as $|\sin s|, |\cos s| \leq 1$ as

$$|j(s)| \leq \frac{1}{(s^2 - \pi^2)^2} \left| 2 + 2 \frac{16s^2}{(s^2 - \pi^2)^2} + \frac{8s}{s^2 - \pi^2} \right|. \quad (43)$$

For large s , we can bound this as $|j(s)| = \mathcal{O}(|s|^{-4})$, so the integral converges and

$$\frac{\gamma}{T_2} \bar{\mathcal{I}}_\gamma(T_2) \xrightarrow{T \rightarrow \infty} e^{-1} \int_{-\infty}^{+\infty} j(s) ds \equiv C^{-1}. \quad (44)$$

We numerically find that the value of the integral is $e/C \approx 0.13$ and $C \approx 20.8$. ■

3. Comparison between single-control and QFT-based methods

We compare numerically the MSQPE algorithm against RPE across a range of target precisions ϵ_t and error rates γ . For RPE, we choose the optimized parameters from Ref. [72], fixing the target precision $\epsilon_t = 2^{-j}$ to give optimal performance for the algorithm. (As currently defined in Algorithm 4, to target $2^{-(j-1)} > \epsilon_t > 2^{-j}$ has the same cost as estimating at a precision 2^{-j} , though this could be improved upon—e.g., by extending Algorithm 2 in a similar way as in Algorithm 7.) To calculate the Holevo error $\epsilon^{(\text{RPE})}$ achieved by RPE, we run 1000 simulations of Algorithm 4 targeting a different random phase on $[0, 2\pi)$. This yields the orange pluses in Fig. 8. With the Holevo errors $\epsilon^{(\text{RPE})}$ from RPE determined, we use this as a target error $\epsilon_t = \epsilon^{(\text{RPE})}$ for MSQPE. We run 1000 simulations, resulting in the blue circles in Fig. 8. We observe that the mean Holevo error $\epsilon^{(\text{MSQPE})}$ for MSQPE is very close to the target precision ϵ_t , which validates our choice of error model used to set the parameters of the algorithm [Eq. (27)]. For comparison, we plot the Cramer-Rao bounds from the simulated quantum device data that are fed into either algorithm (orange for RPE and blue for SinQPE). We additionally state the standard Heisenberg limit and the global-depolarizing-noise limit calculated in Ref. [88] (see Sec. IV B). To construct Fig. 2, we repeat the above calculation for a different noise rate $\gamma = 2^{-m}$ and plot the relative total cost $\mathcal{T}_{\text{tot}}$ of RPE versus MSQPE at a fixed $\epsilon^{(\text{RPE})} \approx \epsilon^{(\text{MSQPE})}$.

We observe that MSQPE achieves the target precision as expected; the orange and blue points in Fig. 8(a) have nearly identical γ values, but differ in the cost $\mathcal{T}_{\text{tot}}$. The numerical error for the MSQPE is very close to the classical Cramer-Rao bound for the data taken, while the error of RPE is about 50% above its Cramer-Rao bound [Fig. 8(b)]. However, the Cramer-Rao bounds for the two algorithms are different (as they take different data), so this does not directly translate into a performance advantage of MSQPE at all regimes in Fig. 2. In the regime of large target error ϵ_t , the noise is not significant and both algorithms perform as in the noiseless case, achieving near-Heisenberg scaling $\epsilon \propto 1/\mathcal{T}_{\text{tot}}$ [Fig. 8(a)]. However, MSQPE has a better scaling constant; it satisfies the Heisenberg limit $\epsilon \approx \pi/\mathcal{T}_{\text{tot}}$, while RPE has a small gap $\epsilon \approx 5\pi/\mathcal{T}_{\text{tot}}$. This matches known results for the noiseless case [60,61]. The advantage of MSQPE continues in the intermediate regime and is more significant for smaller noise strengths γ (as can be seen in Fig. 2). In the limit of small errors, both algorithms asymptotically show $\epsilon \propto \sqrt{\gamma}/\mathcal{T}_{\text{tot}}$ as expected. In this regime, RPE outperforms MSQPE by a factor of ≈ 1.3 . Note that MSQPE continues to closely follow the Cramer-Rao bound in this case, which implies that the performance loss is driven by the sine-state being suboptimal in the presence of noise, rather than the classical estimation routine performing badly. For MSQPE, the scaling factor is ≈ 4.9 (which is close to the analytical result in Theorem 6, as $\sqrt{C} \approx 4.6$), and for RPE ≈ 4.0 —slightly worse than $\sqrt{4e} \approx 3.3$ bound from the specific Fisher information in Eq. (19). Both are significantly (by a factor of 2) above the fundamental limit $0.89\sqrt{4\gamma/3e} \approx 1.7$ [88].

D. Error-mitigation overhead for QPE

The depolarizing-noise model is a useful toy model to understand the idealized performance of a quantum algorithm (especially in terms of the variance cost), but it fails to capture the bias present in a typical noise model for a quantum device [37]. Over recent years, many error-mitigation techniques have emerged, promising a range of fidelity overheads and unbiasing ability. In this section we consider the effect that realistic noise- and error-mitigation techniques would have on the results in Sec. IV C. Our primary result is to incorporate a fidelity overhead of F^α directly into our $\mathcal{T}$ optimization for QPE. We find that we can trade an increased α (typically $\alpha = 1, 2$, or 4) against $\mathcal{T}$ to yield a multiplicative overall cost $\mathcal{T}_{\text{tot}} \rightarrow \alpha \mathcal{T}_{\text{tot}}$. This suggests that the overhead for explicitly unbiased methods such as PEC [44,45] may be affordable, but it also suggests that improving error mitigation will yield little further gains for QPE.

Error-mitigation theory typically focuses on expectation-value estimation tasks: measuring to precision ϵ an expectation value $\langle \psi | O | \psi \rangle$ on a state $|\psi\rangle$ produced by a given noiseless quantum circuit [37]. This can theoretically be

obtained running the ideal circuit M_1 times. Only having access to a noisy device, an error-mitigation scheme attempts to reproduce the same estimate by running a larger number M_F of shots of noisy quantum circuits. The error-mitigation overhead $C_{\text{em}}(F) = M_F/M_1$ depends on the fidelity $F = \langle \psi | \rho | \psi \rangle$ to which the original circuit can be run on the noisy device (where $|\psi\rangle$ is the state that would be produced by a noiseless run of the logical circuit and ρ is the mixed state actually produced by the noisy run on a device).

Definition 5 (Error mitigation for expectation values). Consider a quantum circuit that prepares a state $|\psi\rangle$, a noisy device that can run this circuit preparing a mixed state ρ with fidelity $F = \langle \psi | \rho | \psi \rangle$, and an observable O that can be sampled on the prepared state. An *error-mitigation scheme* is an algorithm that produces an estimate $\tilde{O}$ of $\langle \psi | O | \psi \rangle$ by running $\tilde{M}$ quantum circuits on the noisy device and processing the outcome samples. (The circuits run can be different from the original logical circuit and the algorithm can depend on additional knowledge of the device-noise model.) An error-mitigation scheme is *unbiased* if it produces an estimate that converges to the true value $\tilde{O} \rightarrow \langle \psi | O | \psi \rangle$ in the large number of samples $\tilde{M} \rightarrow \infty$; otherwise, it converges to $\tilde{O} \rightarrow \langle \psi | O | \psi \rangle + b$, where we call b the *bias*. For an unbiased scheme, the *error-mitigation overhead* is defined as $C_{\text{em}}(F) = \lim_{\tilde{M} \rightarrow \infty} \tilde{M} \text{Var}[\tilde{O}] / \text{Var}[O]$, where $\text{Var}[\tilde{O}]$ is the expected variance of the estimate and $\text{Var}[O] = \langle \psi | O^2 | \psi \rangle - \langle \psi | O | \psi \rangle^2$.

Error-mitigation techniques can be classified based on the overhead scaling as a function of F . If errors can be detected deterministically and associated runs discarded, error mitigation can be achieved with overhead $C_{\text{em}} \sim F^{-1}$, which we term the *postselection bound*. This is achieved by symmetry verification [47,48,132], and echo verification when state preparation is cheap compared to measurement [50]. Postselection schemes are typically problem dependent and can only mitigate some errors, leaving a large residual bias b .

A more general set of error-mitigation schemes relies on measuring a noisy signal and processing it classically to remove the bias, which is estimated analytically under specific assumptions on the problem and noise model. This is particularly effective for global depolarizing noise [133] or when the effect of the noise can be simulated effectively [134]. Due to error propagation in the required rescaling, the overhead for these techniques is $C_{\text{em}} \sim F^{-2}$; we call this the *rescaling bound*. Echo verification matches this bound when the state preparation is expensive compared to measurement, as the volume of the noisy circuits that need to be run is double that of the original circuit [49,50].

Finally, a last set of error-mitigation techniques is prescribed to measure explicitly the bias on the signal caused

by circuit errors and subtract it from the biased signal. This is the case of PEC, which can completely remove the bias from an observable expectation value by rewriting the noiseless signal as a nonconvex combination of signals from noisy “basis” circuits with coefficients α_j , assuming good characterization of the noise itself. In the case of local stochastic noise, the overhead of PEC is $C_{\text{em}}(\sum_j |\alpha_j|)^2 \propto F^{-4}$ [37,44,45]. This can be interpreted as a cost F^{-2} due to the necessity of explicitly measuring the bias and F^{-2} due to the necessity of rescaling of the resulting signal after subtracting the bias.

When implementing quantum phase estimation in the presence of noise, we have an additional handle on the depth to which the circuits to be run; choosing this as a function of the noise rate impacts the scaling of the error-mitigation overheads. We consider a stochastic error model and we define the noise rate γ as the logarithm of the error probability when implementing the noisy channel

$$\mathcal{U}(\rho) = e^{-\gamma} cU\rho cU^\dagger + (1 - e^{-\gamma})\mathcal{N}(\rho), \quad (45)$$

which approximates the controlled unitary oracle cU , where $\mathcal{N}$ is the action of the channel conditioned on at least one noise event happening (a.k.a. the pure-noise channel). In the noiseless case $\gamma = 0$, we can achieve an accuracy $\leq \epsilon$ on the estimated phase ϕ by running the optimal Sin-QPE (Algorithm 1) at oracle depth $T_{\text{tot}} = \lceil \pi/\epsilon \rceil$. In the presence of noise, we can choose to run multiple circuits with smaller depth $T < T_{\text{tot}}$, resulting in fidelities bounded by $F > e^{-\gamma T}$. Under a given scheme, we define $T_{\text{tot}}(\gamma, \epsilon)$ as the cost of estimating ϕ to accuracy ϵ . By analogy to the definition of C_{em} , we define the error-mitigation overhead for this scheme as $C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) = T_{\text{tot}}(\gamma, \epsilon)/T_{\text{tot}}(0, \epsilon)$.

Theorem 7 (Error-mitigation overhead for QPE). Assume an unbiased error-mitigation scheme for expectation values with overhead $C_{\text{em}}(F) = \Lambda F^{-\nu}$ and the ability to perform single-control QPE circuits with T calls to the noisy implementation $\mathcal{U}$ of U in Eq. (45) for a fidelity $F = e^{-\gamma T}$ (with error rate $\gamma < 1/\nu$). Then, we can realize an error-mitigated single-control QPE protocol that estimates ϕ to an accuracy ϵ with an overhead $C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \propto \nu\gamma/\epsilon$ in the number of calls to $\mathcal{U}$, compared to the optimal noiseless QPE protocol.

Proof. In the absence of noise, we know from Theorem 4 that the optimal QPE algorithm (sine-state QPE - Algorithm 1) estimates ϕ with a precision ϵ in time $T_{\text{tot}}^{\text{opt}} = \pi/\epsilon$. From Theorem 5 we know that, using RPE, we can estimate ϕ to precision $\tilde{\epsilon} \leq \epsilon$ using noiseless circuits of depth bounded by $T = b/\tilde{\epsilon}$ for a cost $T_{\text{tot}}^{\text{RPE}}(\tilde{\epsilon}) = a/\tilde{\epsilon}$ (for algorithm-specific constants $b = 5$ and a analytically bounded by $a < 24.26\pi$ and numerically found to be $b \approx 5\pi$). We can recover an estimate to precision ϵ by repeating $\tilde{\epsilon}^2/\epsilon^2$ times this RPE algorithm to precision $\tilde{\epsilon}$

and averaging the results; the maximum depth for the RPE circuits remains unchanged and the total time becomes $T_{\text{tot}}^{\text{SC}}(T, \epsilon) = a\tilde{\epsilon}/\epsilon^2 = ab/T\epsilon^2$. As the single-control QPE circuits measure expectation values, we can apply error mitigation to recover the noiseless signal. The fidelity of these circuits is bounded by $F \geq e^{-\gamma T}$, so the error-mitigation overhead is bounded by $C_{\text{em}}(T, \gamma) \leq \Lambda e^{\nu\gamma T}$. The QPE error-mitigation overhead is then calculated as

$$C_{\text{em}}^{\text{QPE}}(T, \gamma, \epsilon) = \frac{T_{\text{tot}}^{\text{SC}}(T, \epsilon) C_{\text{em}}(T, \gamma)}{T_{\text{tot}}^{\text{opt}}(\epsilon)} \leq \frac{ab\Lambda}{T\pi\epsilon} e^{\nu\gamma T}$$

$$\xrightarrow[\text{over } T]{\text{minimize}} C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \frac{a\Lambda e}{\pi} \frac{b\nu\gamma}{\epsilon}, \quad (46)$$

with the bound minimized by the choice of $T = 1/\gamma\nu$. As we are limited to an integer number of repetitions, this holds for $b/T = \tilde{\epsilon} \leq \epsilon$, i.e., for $\gamma > \epsilon/b\nu$. For smaller error rates, we simply have to run RPE once while, on every circuit, paying the error-mitigation overhead, bounded by $C_{\text{em}}(e^{-\gamma b/\epsilon}) = \Lambda e^{\nu\gamma b/\epsilon}$ as the maximum depth is b/ϵ . The total cost is bounded by $C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon)a/\epsilon$ and the QPE overhead is bounded by

$$C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \Lambda e^{\nu\gamma b/\epsilon} \frac{a}{\pi} \leq \frac{a\Lambda e}{\pi}. \quad (47)$$

Summarizing the results for all $\gamma < 1/\nu$, we have

$$C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \frac{a\Lambda e}{\pi} \max\left(\frac{b\nu\gamma}{\epsilon}, 1\right). \quad (48)$$

The residual overhead even for $\gamma \rightarrow 0$ is due to (1) the fact that we are using single-control QPE circuits (factor b/π), (2) the assumption that the error-mitigation scheme is applied anyway (factor Λ) and (3) the fact that we are using a loose bound in Eq. (47) (factor e). ■

The above analysis considers the increased variance from quantum error mitigation (and the associated increase in the cost of measurement), but not the effect of any residual bias b . For all orders before the $(J - 1)$ th, Algorithm 4 is robust to a bias $b \leq 0.13$ [119], in that we can promise with high confidence to estimate the first $J - 2$ bits of the phase ϕ . In the absence of error mitigation, the bias in an expectation value $\bar{Z}$ is worst-case bounded by $b \leq 2(1 - F)$, which for the $(J - 2)$ th circuit executed in Theorem 7 yields in turn $b \leq 2(1 - e^{-1/2\nu})$. This corresponds for $\nu = 1, 2, 4$ to $b \leq 0.79, 0.44, 0.24$ respectively, which implies that working error mitigation (with sufficiently small residual bias) is required to achieve the depths that we target for RPE. At the final order, robustness cannot be guaranteed and a bias b in the estimation of $\bar{Z}$ at the $(J - 1)$ th order propagates through to the final estimate

of ϕ ,

$$\begin{aligned} b_\phi &= |\langle \tilde{\phi}^{(J-1)} \rangle - \phi| = \frac{\pi}{2^{J-1}} (\text{Arg}[\bar{Z} + b] - \text{Arg}[\bar{Z}]) \\ &\leq \frac{\pi b}{2^{J-1}}. \end{aligned} \quad (49)$$

E. Explicit unbiasing of multi-control QPE

In sine-state QPE, the estimation task cannot be reduced to sample averaging, as the estimate of the phase is not an expectation value. In this section, we extend explicit unbiasing techniques to a more complex data-processing scheme: maximum-likelihood estimation. Extending error-mitigation techniques beyond expectation-value estimation is identified as a relevant open problem in the literature [37].

1. The explicitly unbiased maximum-likelihood estimator

Similar to what is done in PEC for expectation values, we begin by decomposing the unitary channel that implements the ideal SinQPE circuit $\mathcal{U}[\rho] = U\rho U^\dagger$ as a nonconvex linear combination of “basis operations” $\{\mathcal{B}_j\}$, i.e., channels that can be run on the noisy device:

$$\mathcal{U} = \sum_j \alpha_j \mathcal{B}_j. \quad (50)$$

This decomposition can be learned via either a sparse Pauli-Lindblad decomposition [54] or from the output of QEC stabilizer measurements [135]. For stochastic Pauli channels (which can represent a model for logical noise in surface code even when the underlying physical noise is coherent [136]), the decomposition procedure is known. In particular, for a circuit of volume A and local depolarizing noise with constant rate per volume element r , the decomposition has 1-norm $\|\alpha\|_1 = e^{2rA} = F^{-2}$, where $F = e^{-rA}$ is the circuit fidelity. (The cost of standard PEC for mitigating expectation values is then $C_{\text{em}} = \|\alpha\|_1^2 = F^{-4}$.)

In multi-circuit sine-state QPE, we are interested in estimating the phase ϕ given a set of M measurements $\{x_i\}_{i=1,\dots,M}$. In the ideal noiseless case, these are IID samples obtained from the SSQPE circuit (see Fig. 5) with distribution $x \sim P(x) = \text{Tr} \Pi_x \mathcal{U}[\rho_0]$, where $\{\Pi_x\}$ is the projector of the control register onto the bit string x and ρ_0 is the input state. We can decompose this distribution using Eq. (50) as

$$P(x) = \sum_j \alpha_j Q_j(x), \quad Q_j(x) = \text{Tr} \Pi_x \mathcal{B}_j[\rho_0]. \quad (51)$$

Our goal is to define a consistent estimator of ϕ based on samples taken from the latter distributions, from which we can sample using the noisy device.

In the case of multiple samples $\{x_i\}$ taken from noiseless circuits, an optimal estimator $\hat{\phi}$ of ϕ is defined by maximizing the log-likelihood of the samples over the parametric model of the outcome distribution $P(x|\phi)$ given in Eq. (12):

$$\hat{\phi} = \arg \max_{\phi} \mathcal{L}(\phi), \quad \mathcal{L}(\phi) = \frac{1}{M} \sum_{i=1}^M \log P(x_i|\phi). \quad (52)$$

We have added a constant factor M^{-1} compared to the usual definition of log-likelihood, making clear that the $\mathcal{L}(\phi)$ is evaluated as a sample average of the logarithm of the model probability. In the large-sample limit,

$$\lim_{M \rightarrow \infty} \mathcal{L}(\phi) = \bar{\mathcal{L}}(\phi) = \mathbb{E} [\log P(x|\phi) | x \sim P(x)], \quad (53)$$

which we can decompose through Eq. (51) as

$$\bar{\mathcal{L}}(\phi) = \sum_j \alpha_j \mathbb{E} [\log P(x|\phi) | x \sim Q_j(x)] \quad (54)$$

$$= \|\alpha\|_1 \mathbb{E} \left[\text{sgn}(\alpha_j) \mathbb{E} [\log P(x|\phi) | x \sim Q_j(x)] \middle| j \sim \frac{|\alpha_j|}{\|\alpha\|_1} \right] \quad (55)$$

$$= \|\alpha\|_1 \mathbb{E} [\text{sgn}(\alpha_j) \log P(x|\phi) | j, x \sim \tilde{Q}(j, x)], \quad (56)$$

where we have defined the joint distribution $\tilde{Q}(j, x) = |\alpha_j|/\|\alpha\|_1 Q_j(x)$. We can sample data points $\{(j_i, y_i)\}_{i=1,\dots,\tilde{M}}$ by importance sampling $j_i \sim |\alpha_{j_i}|/\|\alpha\|_1$ and then sampling $y_i \sim Q_{j_i}(y_i)$ on the noisy quantum device. We can then compute an estimate $\tilde{\mathcal{L}}$ of $\bar{\mathcal{L}}$ as a sample average

$$\tilde{\mathcal{L}}(\phi) = \frac{\|\alpha\|_1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} \text{sgn}(\alpha_{j_i}) \log P(y_i|\phi). \quad (57)$$

Maximizing this gives an estimator $\tilde{\phi}$ of ϕ , which we define below. We call this estimator the *explicitly unbiased maximum-likelihood estimator* (EUMLE).

Definition 6. [Explicitly unbiased maximum-likelihood estimator] Given (1) a parametrized model $P(x|\phi)$ for the distribution of x obtained from a noiseless quantum circuit and (2) a nonconvex decomposition of $P(x|\phi^*) = \sum_j \alpha_j Q_j(x)$ in terms of distributions $Q_j(x)$ that can be sampled using a noisy quantum device, the EUMLE of the phase ϕ^* is defined as $\tilde{\phi} = \arg \max_{\phi} \tilde{\mathcal{L}}(\phi)$, where $\tilde{\mathcal{L}}(\phi)$ in Eq. (57) is constructed by sampling $\tilde{M}$ points $\{(j_i, x_i) | j_i \sim |\alpha_{j_i}|/\|\alpha\|_1, x_i \sim Q_{j_i}(x_i)\}_{i=1,\dots,\tilde{M}}$.

The EUMLE is a consistent estimator, i.e., in the large-sample limit as $\tilde{M} \rightarrow \infty$, it converges to the true value of ϕ . The objective function of the optimization $\tilde{\mathcal{L}}(\phi) \rightarrow$

$\tilde{\mathcal{L}}(\phi)$ asymptotically matches the log-likelihood of the noiseless data in Eq. (53). Here we prove the consistency of the EUMLE by showing that the maximum of the likelihood for $\tilde{M} \rightarrow \infty$ remains the same. We also prove the stronger property of asymptotic normality and recover an expression for the asymptotic variance of the estimator.

Theorem 8 (Asymptotic normality of the EUMLE). The EUMLE $\tilde{\phi}$ of the phase ϕ is asymptotically normal with mean ϕ and variance $\sigma^2 \propto \tilde{M}^{-1}$.

Proof. For this proof we follow a strategy similar to the one used to prove asymptotic normality of the maximum-likelihood estimator in Theorem 10.1.12 of Ref. [124]. First, we Taylor expand $\tilde{\mathcal{L}}'(\phi)$ (the primes are derivatives with respect to ϕ) around the true value $\phi^* = \arg \min_{\phi} \tilde{\mathcal{L}}(\phi)$:

$$\tilde{\mathcal{L}}'(\phi) = \tilde{\mathcal{L}}'(\phi^*) + (\phi - \phi^*)\tilde{\mathcal{L}}''(\phi^*) + \mathcal{O}((\phi - \phi^*)^2). \quad (58)$$

We ignore the second order term, as it is subdominant for small errors under reasonable regularity conditions, and we use the stationary point condition $\tilde{\mathcal{L}}'(\tilde{\phi}) = 0$ to get

$$\tilde{\phi} - \phi^* = \frac{-\tilde{\mathcal{L}}'(\phi^*)}{\tilde{\mathcal{L}}''(\phi^*)}. \quad (59)$$

The denominator is

$$\tilde{\mathcal{L}}''(\phi^*) = \frac{|\alpha|_1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} \text{sgn}(\alpha_{j_i}) [\partial_{\phi}^2 \log P(y_i|\phi)]_{\phi^*} \quad (60)$$

$$\xrightarrow[\tilde{M} \rightarrow \infty]{\text{p}} |\alpha|_1 \mathbb{E} [\text{sgn}(\alpha_j) [\partial_{\phi}^2 \log P(y|\phi)]_{\phi^*} | j, y \sim \tilde{Q}(j, y)] \quad (61)$$

$$= \sum_j \alpha_j \int dQ_j(y) [\partial_{\phi}^2 \log P(y|\phi)]_{\phi^*} \quad (62)$$

$$= \mathbb{E} [[\partial_{\phi}^2 \log P(y|\phi)]_{\phi^*} | y \sim P(y)] \\ = \mathbb{E} \left[- \underbrace{[\partial_{\phi} \log P(y|\phi)]_{\phi^*}^2}_{[s(y)]^2} | y \sim P(y) \right] =: -\mathcal{I}(\phi^*), \quad (63)$$

where the convergence in probability $\xrightarrow{\text{p}}$ is justified by the law of large numbers and $\mathcal{I}(\theta^*)$ is the Fisher information of an ideal sample x from the noiseless distribution $P(x)$. For brevity of further notation, we define the score of

the probability model $s(x) := [\partial_{\phi} \log P(x|\phi)]_{\phi=\phi^*}$. By the central limit theorem, the numerator

$$-\tilde{\mathcal{L}}'(\phi^*) = -\frac{|\alpha|_1}{\tilde{M}} \sum_{i=1}^{M'} \text{sgn}(\alpha_{j_i}) s(y_i) \quad (64)$$

converges in distribution to a normal distribution with mean $-\mathbb{E}[s(x)|x \in P(x)] = 0$ and variance $(|\alpha|_1^2/\tilde{M}) \mathbb{E} [s(y)^2 | y \sim \tilde{Q}(y)]$, with $\tilde{Q}(y) = \sum_j \tilde{Q}(j, y)$ notating the marginal distribution. Thus, the estimator error $\tilde{\phi} - \phi$ is asymptotically normal with mean 0 and variance

$$\sigma^2 = \frac{|\alpha|_1^2}{\tilde{M}} \frac{\mathbb{E} [s^2(y) | y \sim \tilde{Q}(y)]}{\mathbb{E} [s^2(y) | y \sim P(y)]^2}. \quad (65)$$

■

2. Regularizing the EUMLE

Maximizing $\tilde{\mathcal{L}}$ as is leads to unstable behavior, as our probability distribution $P(x|\phi)$ [Eq. (12)] can be arbitrarily close to 0. This is especially problematic in our case, as we are sampling from $\tilde{Q}(x)$ instead of $P(x|\phi)$ and so we do not expect singular behavior of $\tilde{\mathcal{L}}$ to be a rare event. To alleviate this problem we regularize our model $P(x|\phi)$ by adding a small constant; $P(x|\phi) \rightarrow P_c(x|\phi) = P(x|\phi) + c$. (In principle we can add an arbitrary ϕ -independent function to P , which may be better for specific noise cases. This extension is left for future research.) This yields a new form of the log-likelihood:

$$\tilde{\mathcal{L}}_c(\phi) = \frac{|\alpha|_1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} \text{sgn}(\alpha_{j_i}) \log P_c(y_i|\phi) + c \sum_{x \in \mathcal{D}} \log P_c(x|\phi), \quad (66)$$

where $\mathcal{D}$ is the domain of the samples x (e.g., strings of $\log_2(\mathcal{T})$ bits).

For periodic distributions with continuous domain $\mathcal{D} = [0, 2\pi)$, where $P(x|\phi) = P(x - \phi \bmod 2\pi)$ holds, the second term is not dependent on ϕ and can be neglected. Such a distribution can be obtained from QPE adding a random offset to the phase being estimated [137,138]. Here, however, we introduce a general regularized EUMLE which uses this modified likelihood and prove its asymptotic normality.

Definition 7. [Regularized EUMLE] The regularized explicitly unbiased maximum-likelihood estimator (rEUMLE) of the phase ϕ^* is derived from the EUMLE (definition 6); it requires the same inputs plus an additional regularization constant c , it uses the same samples from the quantum device, and it is defined as $\tilde{\phi} = \arg \max_{\phi} \tilde{\mathcal{L}}_c$ using the regularized log-likelihood given in Eq. (66).

Theorem 9 (Asymptotic normality of the rEUMLE). The rEUMLE $\tilde{\phi}$ of the phase ϕ is asymptotically normal with mean ϕ^* and variance

$$\sigma^2 = \frac{|\alpha|_1^2 \mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)]}{\tilde{M} \mathbb{E}[s_c^2(y)|y \sim P_c(y)]^2}, \quad (67)$$

where s_c is the regularized score function $s(x) := [\partial_\phi \log P_c(x|\phi)]_{\phi=\phi^*}$.

Proof. First, we show that the estimator is consistent, i.e., in the limit of infinite samples $\tilde{M}$ the estimator converges to the true value ϕ^* . The regularized likelihood function $\tilde{\mathcal{L}}_c$ converges to the expectation value

$$\begin{aligned} \mathcal{L}_c(\phi) &\equiv \lim_{\tilde{M} \rightarrow \infty} \tilde{\mathcal{L}}_c(\phi) = |\alpha|_1 \mathbb{E}[\text{sgn}(\alpha_j) \log P_c(x|\phi) | \\ &\quad (j, x) \sim \frac{|\alpha_j|}{|\alpha|_1} Q_j(x)] + c \sum_{x \in \mathcal{D}} \log P_c(x|\phi) \end{aligned} \quad (68)$$

$$= \sum_{x \in \mathcal{D}} P(x) \log P_c(x|\phi) + c \log P_c(x|\phi) \quad (69)$$

$$= \sum_{x \in \mathcal{D}} P_c(x) \log P_c(x|\phi). \quad (70)$$

$\mathcal{L}_c(\phi)$ has a maximum at $\phi = \phi^*$ since

$$\mathcal{L}'_c(\phi^*) = \partial_\phi \left[\sum_{x \in \mathcal{D}} P_c(x|\phi) \right] = \partial_\phi [1 + c] = 0, \quad (71)$$

$$\begin{aligned} \mathcal{L}''_c(\phi^*) &= - \sum_{x \in \mathcal{D}} \frac{[\partial_\phi P_c(x|\phi)]_{\phi=\phi^*}^2}{P_c(x|\phi^*)} \\ &= - \mathbb{E}[s_c^2(x)|x \sim P_c(x)] < 0. \end{aligned} \quad (72)$$

The proof of asymptotic normality then follows the same steps as the proof of Theorem 8 using the regularized versions of the likelihood $\tilde{\mathcal{L}}_c$ and score s_c . ■

Next, we calculate the sample overhead of this error-mitigation technique. This is defined as the ratio between the number of samples $\tilde{M}$ needed to achieve σ^2 to the number of samples M that would be sufficient in the noiseless case ($\tilde{Q} = P$, $\alpha = 1$ and no regularization required $c = 0$):

$$\begin{aligned} \mathcal{C}_{\text{EUMLE}} &= \frac{\tilde{M}}{M} \\ &= |\alpha|_1^2 \frac{\mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)]}{\mathbb{E}[s_c^2(x)|x \sim P_c(x)]} \frac{\mathbb{E}[s_0^2(x)|x \sim P(x)]}{\mathbb{E}[s_c^2(x)|x \sim P_c(x)]}. \end{aligned} \quad (73)$$

Here, the term $\mathbb{E}[s_0^2(x)|x \sim P(x)]$ is the Fisher information of the noiseless distribution. The overhead consists of

three factors: (1) the rescaling constant $|\alpha|_1^2$, (2) the ratio between the average (regularized) score of the importance-sampled noisy distribution and the noiseless distribution, and (3) the overhead due to the loss of information caused by regularization. The regularization constant c ensures that term (2) remains bounded, but it also causes term (3) to increase, so it should be chosen to optimize the product of the two. Quantifying $\mathcal{C}_{\text{EUMLE}}$ requires additional assumptions about the noise model and the consequent basis circuit decomposition, and about the specific model $P(x|\phi)$.

Let us finally consider the effect of residual bias on the regularized EUMLE. This appears if the assumption of perfect noise modeling is broken (in which case PEC is no longer unbiased), leading us to sample from an incorrect set of probability distributions: $\tilde{Q}_j(x) \rightarrow Q_j(x) + b_j(x)$. In the $\tilde{M} \rightarrow \infty$ limit this sends $\tilde{\mathcal{L}}_c(\phi) \rightarrow \mathcal{L}_c(\phi) + B(\phi)$, where

$$\begin{aligned} B(\phi) &= |\alpha|_1 \mathbb{E}[\text{sgn}(\alpha_j) \log P_c(x|\phi) | (j, x) \\ &\quad \sim \frac{|\alpha_j|}{|\alpha|_1} (Q_j(x) + b_j(x))] \end{aligned} \quad (74)$$

$$- |\alpha|_1 \mathbb{E}[\text{sgn}(\alpha_j) \log P_c(x|\phi) | (j, x) \sim \frac{|\alpha_j|}{|\alpha|_1} Q_j(x)] \quad (75)$$

$$= \sum_{x \in \mathcal{D}} b(x) \log P_c(x|\phi), \quad (76)$$

and $b(x) = \sum_j \alpha_j b_j(x)$ is the difference between the distribution produced and our target $P(x)$. Adding $B(\phi)$ to $\mathcal{L}_c(\phi)$ shifts the minimum from ϕ^* , yielding a bias b_ϕ . To calculate this, we Taylor expand the derivative of $\mathcal{L}_c$ at the new point

$$\begin{aligned} &\frac{\partial}{\partial \phi} [\mathcal{L}_c(\phi) + B(\phi) - B(\phi)]_{\phi=\phi^*+b_\phi} \\ &= \frac{\partial}{\partial \phi} [\mathcal{L}_c(\phi)]_{\phi=\phi^*} + b_\phi \frac{\partial^2}{\partial \phi^2} [\mathcal{L}_c(\phi)]_{\phi=\phi^*} + \mathcal{O}(b_\phi^2). \end{aligned} \quad (77)$$

We have that $\mathcal{L}_c + B$ has a minimum at $\phi^* + b_\phi$, and $\mathcal{L}_c$ has a minimum at ϕ^* , so we can calculate to lowest order

$$b_\phi = - \frac{B'(\phi^* + b_\phi)}{\mathcal{L}''_c(\phi^*)}. \quad (78)$$

We identify the denominator $\mathcal{L}''_c(\phi^*) = \mathcal{I}(\phi^*)$ as the Fisher information. To bound the numerator, we calculate

$$B'(\phi) = \sum_{x \in \mathcal{D}} b(x) \frac{P'_c(x|\phi)}{P_c(x|\phi)}. \quad (79)$$

This gives a final bound on our bias of

$$b_\phi \leq b\kappa \quad \kappa = \frac{1}{\mathcal{I}(\phi^*)} \sum_{x \in \mathcal{D}} \frac{|P'_c(x|\phi^* + b_\phi)|}{P_c(x|\phi^* + b_\phi)}, \quad (80)$$

where $b = \max_x |b(x)|$. The additional factor κ depends heavily on the shape of the regularized probability distribution P_c . In particular, in the next section we show $|P'_c(x)/P_c(x)| \leq 0.4$ and $\mathcal{I} \approx 0.01\mathcal{T}^2$ for sine-state QPE distribution in Eq. (12) with depth $\mathcal{T}$ and $c = 0.4$, leading to

$$b_\phi \leq 40 \frac{b}{\mathcal{T}}. \quad (81)$$

This bounds the resulting propagated bias to a factor 13 times that of the equivalent propagated bias of RPE in Eq. (49) for $\mathcal{T} = 2^{J-1}$ (assuming that the Hadamard test bias b in RPE and the $b = \max_x |b(x)|$ considered here are comparable). We note that this calculation is based on loose analytical bounds and we expect our method to perform much better in practice.

3. Applying the regularized EUMLE to sine-state QPE

In this section, we explore the application of the rEUMLE to phase estimation, using the samples from the SinQPE circuit. The regularization is necessary as the SinQPE distribution $P(x|\phi)$ vanishes for certain values of ϕ . Here we bound the variance of the resulting rEUMLE.

Theorem 10. [EUMLE for SinQPE] We consider the rEUMLE (Definition 7) applied to the SinQPE probability distribution [Eq. (12)], which can be expressed using a PEC decomposition [Eq. (51)] with 1-norm $|\alpha|_1$. For any constant c , there exists a threshold depth $\mathcal{T}_{\text{thr}}$ and a constant B such that, for any SinQPE circuit with depth $\mathcal{T} > \mathcal{T}_{\text{thr}}$, the variance of the EUMLE σ^2 is bounded as

$$\sigma^2 \leq B \frac{|\alpha|_1^2}{M\mathcal{T}^2}, \quad (82)$$

where M is the number of samples and the total oracular cost of the algorithm is $\mathcal{T}_{\text{tot}} = M\mathcal{T}$.

Proof. The goal of this theorem is to bound the variance of the rEUMLE in Eq. (67) for the sine-state QPE outcome probability distribution (from Theorem 6):

$$P_{\mathcal{T}}(x|\phi) = f_{\mathcal{T}}\left(\phi - \frac{2\pi x}{\mathcal{T}+1}\right), \quad (83)$$

$$f_{\mathcal{T}}(\Delta\phi) = \frac{\sin^2 \frac{\pi}{\mathcal{T}+2}}{(\mathcal{T}+1)(\mathcal{T}+2)} \frac{1 + \cos[(\mathcal{T}+2)\Delta\phi]}{\left[\cos(\Delta\phi) - \cos\left(\frac{\pi}{\mathcal{T}+2}\right)\right]^2}. \quad (84)$$

In order to bound σ^2 , we calculate asymptotic bounds for $\mathbb{E}[s_c^2(y)|y \sim P(y)]$ (bounded from below as $\Omega(\mathcal{T})^2$) and

for $\mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)]$ (bounded from above as $\mathcal{O}(\mathcal{T}^2)$). This allows us to bound σ^2 for sufficiently large $\mathcal{T}$.

We first show that $(1/\mathcal{T}^2) \mathbb{E}[s_c^2(y)|y \sim P(y|\phi)]$ converges to a constant a in the limit as $\mathcal{T} \rightarrow \infty$. Substituting the definition of the probability, we can write

$$\begin{aligned} & \frac{1}{\mathcal{T}^2} \mathbb{E}[s_c^2(y)|y \sim P(y|\phi)] \\ &= \sum_{x=0}^{\mathcal{T}} \frac{\mathcal{T}^{-2} f'_{\mathcal{T}}\left(\phi - \frac{2\pi x}{\mathcal{T}+1}\right)^2}{\left[f_{\mathcal{T}}\left(\phi - \frac{2\pi x}{\mathcal{T}+1}\right) + c\right]^2} f_{\mathcal{T}}\left(\phi - \frac{2\pi x}{\mathcal{T}+1}\right). \end{aligned} \quad (85)$$

For $\mathcal{T} \rightarrow \infty$, we can replace the normalized sum $1/(\mathcal{T}+1) \sum_{x=0}^{\mathcal{T}}$ with an integral

$$\begin{aligned} & \frac{1}{\mathcal{T}^2} \mathbb{E}[s_c^2(y)|y \sim P(y|\phi)] \rightarrow \frac{\mathcal{T}+1}{2\pi} \\ & \times \int_0^{2\pi} ds \frac{\mathcal{T}^{-2} f'_{\mathcal{T}}(\phi - s)^2}{[f_{\mathcal{T}}(\phi - s) + c]^2} f_{\mathcal{T}}(\phi - s) \end{aligned} \quad (86)$$

$$= \frac{\mathcal{T}+1}{2\pi} \int_{-\pi}^{\pi} ds \frac{\mathcal{T}^{-2} f'_{\mathcal{T}}(s)^2}{[f_{\mathcal{T}}(s) + c]^2} f_{\mathcal{T}}(s) \quad (87)$$

$$\begin{aligned} &= \frac{\mathcal{T}+1}{2\pi(\mathcal{T}+2)} \int_{-(\mathcal{T}+2)\pi}^{(\mathcal{T}+2)\pi} ds' \frac{\mathcal{T}^{-2} f'_{\mathcal{T}}\left(\frac{s'}{\mathcal{T}+2}\right)^2}{\left[f_{\mathcal{T}}\left(\frac{s'}{\mathcal{T}+2}\right) + c\right]^2} \\ & f_{\mathcal{T}}\left(\frac{s'}{\mathcal{T}+2}\right), \end{aligned} \quad (88)$$

where, in the second step, we have used the periodicity and evenness of $f_{\mathcal{T}}$ to remove the dependence on ϕ and shift the integration bounds, and in the third step we have performed a change of variable $s' = (\mathcal{T}+2)s$. The limit of this integral as $\mathcal{T} \rightarrow \infty$ can be calculated following the same steps as in the proof of Theorem 6. Using the convergence property $\lim_{\mathcal{T} \rightarrow \infty} f(s/(\mathcal{T}+2)) = f(s)$ with $f(s)$ in Eq. (39), we can write

$$\begin{aligned} & \frac{1}{\mathcal{T}^2} \mathbb{E}[s_c^2(y)|y \sim P(y|\phi)] \\ & \xrightarrow{\mathcal{T} \rightarrow \infty} \int_{-\infty}^{\infty} \frac{ds}{2\pi} \frac{f'(s)^2}{[f(s) + c]^2} f(s) =: a. \end{aligned} \quad (89)$$

This integral converges to a finite constant a , as the integrand (independent on $\mathcal{T}$) can be bounded by $j(s)$ defined in Eq. (41), which can be bounded by $\mathcal{O}(|s|^{-4})$ [see Eq. (43)].

Second, we show that $\lim_{\mathcal{T} \rightarrow \infty} (1/\mathcal{T}^2) \mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)]$ is bounded by bounding $\lim_{\mathcal{T} \rightarrow \infty} 1/\mathcal{T} |s_c(x)|$ for any x . By definition, the score is

$$s_c(x) = [\partial_\phi \log P_c(x|\phi)]_{\phi=\phi^*} = \frac{f'_{\mathcal{T}}\left(\frac{2\pi x}{\mathcal{T}+1} - \phi\right)}{f\left(\frac{2\pi x}{\mathcal{T}+1} - \phi\right) + c} \quad (90)$$

and we can bound it as $|s_c(x)| \leq |f'_T(\frac{2\pi x}{T+1} - \phi)|/c$. As per Eq. (40), the limit function $f'(s) = \lim_{T \rightarrow \infty} \frac{1}{T} f'(s/(T+2))$ is continuous and $f'(s) \xrightarrow{s \rightarrow \pm\infty} 0$ so f' is bounded by a constant $b := \sup_s |f'(s)|$. Combining all the steps, we get the bound

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T^2} \mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)] &\leq \lim_{T \rightarrow \infty} \frac{1}{T^2} \max_x |s_c^2(x)| \\ &\leq \lim_{T \rightarrow \infty} \frac{1}{T^2} \frac{\max_s |f'_T(s)|^2}{c^2} \leq \frac{b^2}{c^2}. \end{aligned} \quad (91)$$

Combining the above limits,

$$\lim_{T \rightarrow \infty} T^2 \frac{\mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)]}{\mathbb{E}[s_c^2(y)|y \sim P_c(y)]^2} \leq \frac{b^2}{a^2 c^2}. \quad (92)$$

For any constant $B > b^2/a^2 c^2$, we can then use this limit to bound the variance σ^2 for any sufficiently large depth $T > T_B$:

$$\sigma^2 = \frac{\mathbb{E}[s_c^2(y)|y \sim \tilde{Q}(y)]}{\mathbb{E}[s_c^2(y)|y \sim P_c(y)]^2} \frac{|\alpha|_1^2}{\tilde{M} T^2} \leq B \frac{|\alpha|_1^2}{\tilde{M} T^2}. \quad (93)$$

Numerically, we find that the limits for $T \rightarrow \infty$ act as bounds in the correct direction: for any T , $\max f'_T \leq (T+2) \max f'$ and $\mathbb{E}[s_c^2(y)|y \sim P(y)] \geq a(T+2)^2$. We find the value of $b \approx 0.16$. For an example choice of $c = 0.4$, we find $a \approx 0.01$. Hence, the bound on the variance in Eq. (82) is satisfied for any depth T for $B = b^2/a^2 c^2 \approx 12.6$. ■

4. Overhead for QPE with local Pauli noise

To quantify the EUMLE overhead in quantities that we can relate to the rest of this work, we assume a local stochastic noise channel with uniform error rate that is locally invertible [37,44,45]. (This is the case for local Pauli channels with constant and qubit-independent error rate, e.g., local depolarizing noise.) Under this assumption, Eq. (50) can be written

$$\mathcal{U} = \frac{1}{F} \mathcal{B}_0 + \sum_{j>0} \alpha_j \mathcal{B}_j, \quad |\alpha|_1 = \frac{1}{F^2}, \quad (94)$$

where F is the logical circuit fidelity and $\alpha_0 = 1/F$ is the weight of the noisy implementation $\mathcal{B}_0$ of the logical circuit.

It remains to optimize T , in an equivalent manner to Theorem 7. To achieve precision ϵ using a circuit of depth T , we need to choose the number of samples $\tilde{M} = \sigma^2/\epsilon^2$.

Using Theorem 10, we can write

$$\mathcal{T}_{\text{tot}} = T \tilde{M} = T \frac{\sigma^2}{\epsilon^2} \leq \frac{B |\alpha|_1^2}{\epsilon^2 T}. \quad (95)$$

Using $|\alpha|_1 = F^{-2}$ and $F = e^{-\gamma T}$,

$$\mathcal{T}_{\text{tot}} \leq \frac{B e^{4\gamma T}}{\epsilon^2 T}. \quad (96)$$

This is optimized at $T = 1/4\gamma$, yielding

$$\mathcal{T}_{\text{tot}}^{(\text{Pauli noise})} = 4B e \gamma \epsilon^{-2} \approx 137 \gamma \epsilon^{-2}. \quad (97)$$

Comparing to Theorem 6, where $\mathcal{T}_{\text{tot}} = C \gamma \epsilon^{-2}$ with $C \approx 20$, we find the overhead from mitigating local Pauli noise to be a constant factor of 7 larger. As the bound in Theorem 10 holds for arbitrary adversarial noise, while the expression in Theorem 8 saturated in the limit as $\epsilon \rightarrow 0$, in practice we expect this gap to be significantly smaller. Furthermore, one could consider replacing our constant regularization with a noise-specific additional term, which may lead to further optimization.

5. Improving the overhead by filtering

In order to improve the error-mitigation overhead of the EUMLE, we propose to combine the explicit unbiased method with a postselection method, which we refer to as filtering. In the single-eigenstate case, we expect the noiseless samples from the QPE circuit to cluster in some region, around the true phase value. By postselecting the outcomes of all our circuit runs based on whether they lie in this clustering region, we can avoid the explosion of the final estimator variance due to the spread of the noisy samples. Note that this is already somewhat achieved by the regularization, as the regularized likelihood is constant on phase estimates sufficiently far from the true value. However, filtering gives us a means to further add prior information, such as energy estimates from classical methods, to boost our QPE performance.

Here we develop a filtering procedure that assumes prior knowledge of a clustering region of size $\Delta \gg \epsilon$ where the phase should lie. The proposed filtering procedure can be applied offline, as the circuits that need to be sampled are the same described above for the EUMLE, the only difference being in the data processing.

Given a guess ϕ_{guess} for the phase that we are attempting to estimate and a phase distance Δ within which we expect to find the true phase, we define the filtering region $[\phi_{\text{guess}} - \Delta, \phi_{\text{guess}} + \Delta]$ and the relative indicator function

$$R(x) = [1 \text{ if } |x - \phi_{\text{guess}}| < \Delta, \text{ else } 0]. \quad (98)$$

We start from a set of measurements $\{(j_i, y_i)\}_{i=1, \dots, \tilde{M}}$ sampled from the same distribution $\tilde{Q}(j, y)$ defined in the

section above, obtained by importance sampling j from the PEC decomposition Eq. (51). We then reject all samples that lie outside the filtering region, obtaining a subset of the data

$$\mathcal{S} = \left\{ (j_i, y_i) \in \{(j_i, y_i)\}_{i=1, \dots, \tilde{M}} \mid y_i \in [\phi_{\text{guess}} - \Delta, \phi_{\text{guess}} + \Delta] \right\}. \quad (99)$$

This subset represents samples from the filtered distribution $\tilde{Q}(j, y)R(y)$ (up to an irrelevant normalization factor). The corresponding filtered model for the noiseless signal is then $P_R(x|\phi) = P(x|\phi)R(x)/\mathcal{N}_R(\phi)$, where $\mathcal{N}_R$ is a ϕ -dependent normalization factor $\mathcal{N}_R(\phi) = \sum_x R(x)P(x|\phi)$. We can then construct a filtered explicitly unbiased maximum-likelihood estimator as in Definition 6, by maximizing

$$\tilde{\mathcal{L}}_R(\phi) = \frac{|\alpha|_1}{|\mathcal{S}|} \sum_{(j_i, y_i) \in \mathcal{S}} \text{sgn}(\alpha_{j_i}) \log P_R(y_i|\phi). \quad (100)$$

Regularization can be reintroduced analogously, as in Sec. IV E 2. Depending on the chosen size of the filtering window Δ , the regularization constant c can be lowered compared to the unfiltered case, or it may not be needed at all.

Let us formalize the above.

Definition 8. [Filtered EUMLE] Given the same assumptions as in Definition 6, as well as an indicator function $R(x)$ (taking values in $\{0, 1\}$), the filtered EUMLE of the phase ϕ is defined as $\tilde{\phi} \sim \arg \max_{\phi} \tilde{\mathcal{L}}_R(\phi)$, where $\tilde{\mathcal{L}}_R(\phi)$ [Eq. (100)] is constructed by sampling $\tilde{M}$ points $\{(j_i, y_i) | j_i \sim |\alpha_{j_i}|/|\alpha|_1, y_i \sim Q_{j_i}(y_i)\}_{i=1, \dots, \tilde{M}}$ and constructing a subset $\mathcal{S}$ [Eq. (99)] by rejecting all points at which $R(y_i) = 0$.

The filtered EUMLE is also a consistent estimator, as $\tilde{\mathcal{L}}_R(\phi)$ matches the log-likelihood of filtered noiseless data coming from the probability $P_R(x) = P(x)R(x)/\mathcal{N}_R$ [which matches the model $P_R(x|\phi)$ for the true value of ϕ]:

$$\lim_{\tilde{M} \rightarrow \infty} \tilde{\mathcal{L}}_R(\phi) = \mathcal{N}_R \mathbb{E} [\log P(x|\phi)R(x)/\mathcal{N}_R(\phi) | x \sim P(x)R(x)/\mathcal{N}_R]. \quad (101)$$

Theorem 11. The filtered EUMLE estimate $\tilde{\phi}$ of a phase $\phi \in [\phi_{\text{guess}} - \Delta, \phi_{\text{guess}} + \Delta]$ is asymptotically normal with mean ϕ and variance $\sigma^2 \propto \tilde{M}^{-1}$.

This result can be proven with the same technique as the result without filtering in Theorem 8.

The scaling of the filtered EUMLE variance with fidelity depends on the details of the noise and of the filtering window. Let us assume that $\sigma \ll \Delta$ and that ϕ lies in

our window. Then, in our worst-case scenario, ϕ lies at the edge of our window, so we reject half of the data from $P(x)$, while an adversarial noise model gives samples within the window. In the best-case scenario, all noise falls outside the window and ϕ lies in the center. In this case, filtering works exactly as postselection. In practice, we suggest that the result may be closer to the latter than the former, as good classical bounds can be many times smaller than the spectral range of the Hamiltonian and we have no reason to expect the outcomes due to noise to fall within the filtering window with a particularly high probability.

Filtering can also be useful when we do not have access to an exact eigenstate $|\phi\rangle$. Assume that the starting state has nonzero overlap with multiple eigenphases separated by gap Δ and that the precision of phase estimation for any single circuit $1/T \ll \Delta$. Then, a filtering procedure similar to the one we have described should enable us to separate the signal coming from each eigenstate. The development of a filtering procedure that works both for eigenvalue projection and variance reduction, as well as the details on the choice of the center and width of the filter and the calculation of the filtering-EUMLE overhead, are left to future work.

F. Surface-code resource estimate and optimization

In this section, we aim to benchmark our circuit-division QPE method and get an idea of the run-time overhead implied by a limitation on the number of available physical qubits. To do this, we estimate the physical resources for running fixed-precision circuit-division sine-state QPE on qubitized block encodings of fermionic Hamiltonians. In particular, we consider the two-dimensional Fermi-Hubbard Hamiltonian [6] and a selection of molecular Hamiltonians for small to medium active spaces, using a block encoding based on the tensor-hypercontraction factorization [7]. We assume free and perfect state preparation—including a cost for state preparation would increase the circuit-division overhead, as the state would need to be prepared for every run; imperfect state preparation would require filtering or a different analysis of data processing. We also assume that the residual noise (after error correction) follows the global depolarizing model. This allows us to apply the results from Sec. IV C 2 and choose the optimal depth $\mathcal{T}$ and the number of samples M as a function of the error rate γ as detailed in Algorithm 5. In practice, in the presence of a more realistic noise model, the techniques proposed in Sec. IV E would need to be applied, resulting in an additional overhead which depends strongly on the assumptions about the noise model.

The assumptions of perfect eigenstate preparation and of global depolarizing noise will need to be lifted for practical applications, which will result in additional overhead. We have introduced an error-mitigation technique

which allows us to extend MSQPE to more general noise models—the EUMLE—in Sec. IV E 1. The overhead associated with error mitigation will depend on the details of the noise model. The comparison between different noise models is not immediate, as the meaning of the noise strength γ can differ depending on the assumptions on noise. However, if we assume a local Pauli noise channel with the same overall error-event probability as GDN, we can bound the overhead of EUMLE-based error mitigation by a factor ≈ 7 (see Sec. IV E 4). Extending our protocol to a general initial state is a clear direction for future research, which could build on the filtered EUMLE technique introduced in Sec. IV E 5.

In the following sections, we describe our procedure for estimating the logical qubit and Toffoli gate count of the QPE circuits for the target problems. In order to translate these to a physical resource estimate, we must make some assumptions about the error correction and magic state distillation model used. We consider a rotated surface code, with Clifford gates performed via lattice surgery [9] and using the magic state factories defined in Ref. [11] to produce CCZ states (as our algorithm only requires Toffoli gates as non-Clifford resources). We consider both using a single factory, to minimize the number of physical qubits (as in Ref. [6]), and using two factories to reduce the run-time. We assume 50% of routing overhead [11] on top of the physical qubits used for data encoding and magic state distillation. We assume a physical gate error rate of $p = 0.001$ and a surface-code cycle time of 1 μ s [9]. The code distances for data qubits encoding d and for the two levels of distillation d_0 and d_1 are then optimized (by grid search) to minimize the physical computation volume while keeping the total error probability per circuit below $1 - e^{-\gamma^T}$.

1. Cost estimates for the qubitized Fermi-Hubbard Hamiltonian

We consider quantum phase estimation of the qubitized walk operator for the two-dimensional Fermi-Hubbard model. We consider a model with $L \times L$ sites, for a total of $2L^2$ spin-orbitals, and interaction strength $u/t = 4$ (with the hopping parameter t setting the energy dimensions). This choice of interaction strength is common in the resource-estimates literature, as it represents the most challenging regime to treat at scale with classical algorithms [84]. We compute the Toffoli cost of the controlled-walk operator $\mathcal{W}$ using QUALTRAN [86] and following the procedure of Ref. [6]. We assume that the rotation angles in the PREPARE step are encoded using $\beta = 10$ bits and performed with the phase-gradient technique (we neglect the cheap one-time preparation of the phase-gradient resource state). We separately compute the Toffoli cost of sin-state preparation and quantum Fourier transformation on $\lceil \log_2(\mathcal{T} + 1) \rceil$ qubits, also using QUALTRAN. We estimate

the number of logical data qubits analytically as

$$\begin{aligned} & \overbrace{2L^2}^{\text{spin orbitals}} + \overbrace{\lceil \log_2(\mathcal{T} + 1) \rceil}^{\text{QPE control qubits}} + \overbrace{\lceil 12 + 3 \log_2(2L^2) \rceil}^{\text{qubitization ancillas from Eq. 64 of [6]}} \\ & + \overbrace{\beta + 1}^{\text{controlled phase gradient}} \end{aligned} \quad (102)$$

The walk operator has the same spectrum as $\pm \arccos H/\lambda$, where $\lambda/t = 8L^2$ is the qubitization 1-norm. We fix the goal of estimating the energy to a precision $\Delta E/t = 0.01$, which can be achieved by requiring a precision on the eigenphase of $\mathcal{W}$ of $\epsilon = \Delta E/\lambda = 0.01/8L^2$ [6].

With the required precision on the phase estimation of the walk operator $\mathcal{W}$ fixed, we estimate the logical cost of implementing the circuit-division QPE algorithm for different choices of the noise rate γ (defined per application of $\mathcal{W}$) using the protocols developed previously in this work. We first use Algorithm 5 to choose the optimal $\mathcal{T}$ and $\mathcal{T}_{\text{tot}}$ based on γ and ϵ . Then, multiplying these by the Toffoli cost of $\mathcal{W}$, we obtain the single-circuit Toffoli cost and total Toffoli cost, respectively.

The resulting physical resources as a function of γ are reported in Fig. 9, for four lattice-side sizes of $L = 2, 5, 10, 20$. Increasing the error rate allowed shortens the largest circuit that is optimal to run (following Algorithm 5, $\mathcal{T} \leq \gamma^{-1}$). As we reduce $\mathcal{T}$, we must increase the number of repetitions quadratically to achieve the same fixed error rate. Thus the total run-time scales inversely with the run-time of the largest circuit. At the same time, increasing the allowed error rate allows us to choose smaller code distances for data (d) and distillation (d_0, d_1), reducing the physical qubit requirements. In Fig. 9(b), we show the time required to run the algorithm under a limitation on the total physical footprint. In, e.g., the $L = 5$ case, we see that our algorithm requires about 2 h with 3×10^5 physical qubits, or about 200 h with half that number of physical qubits. This plot also indicates the number of physical qubits required for it to be more convenient to use multiple factories.

We observe that the curves representing the computational space-time trade-off [in Fig. 9(b)] have an “elbow-like” shape. We expect that our method will lead to the most significant advantage around the elbow of the curve, as further increasing the number of available qubits affects the total run-time only marginally, while further reducing the computational time requires unreasonable space overhead. This elbow point represents the regime in which is most advantageous to combine error correction and error mitigation, similar to the enhancement regime discussed in Sec. IV A 1 and represented in Fig. 4(b). We find that, across system sizes, the elbow point corresponds to the algorithm prescribing around $M \approx 15$ –40 samples, i.e., to the middle regime in Algorithm 5, and to the regime in which MSQPE slightly outperforms RPE (see Fig. 2).

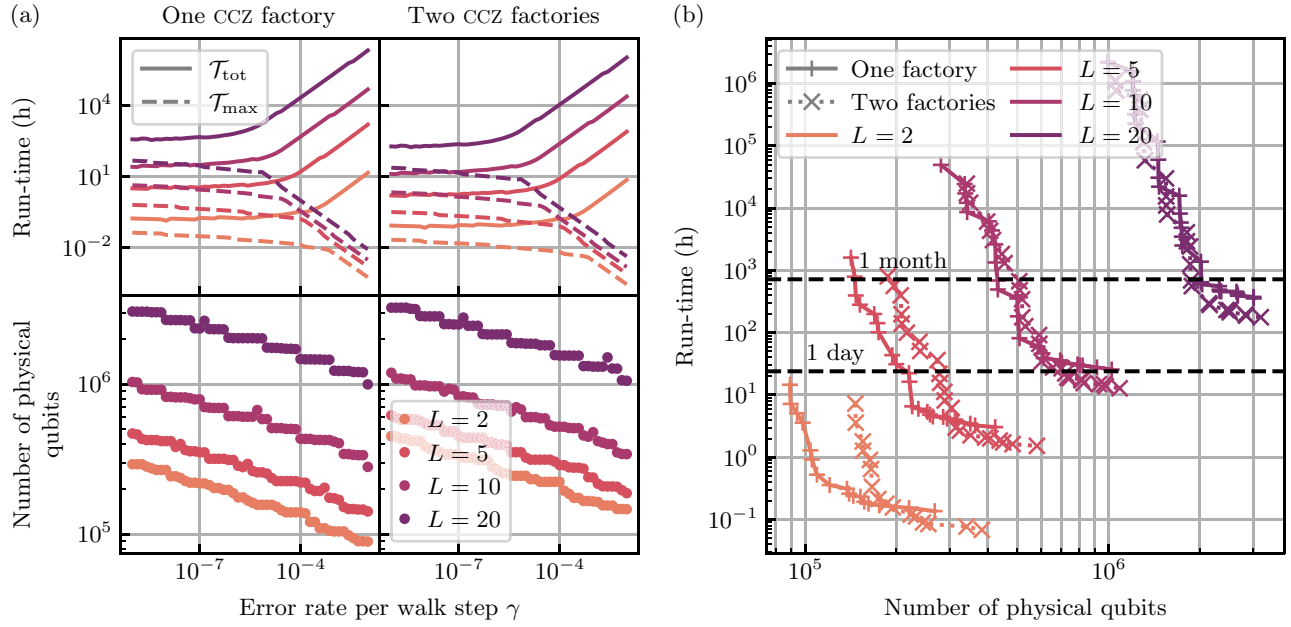


FIG. 9. The physical costs for MSQPE applied to the qubitized Hubbard model Hamiltonian, with a target precision on the phase of $\epsilon = 10^{-2}/\lambda$ (where λ is the qubitization 1-norm). (a) The run-time and qubit costs as a function of the error rate per step of the qubitized walk γ . (b) The trade-off between the number of physical qubits and the run-time. We assume that the residual error (originating from imperfect error correction and imperfect distillation) can be modeled as a global-depolarizing-noise channel of strength γ . The logical circuits are defined according to Algorithm 5 (multi-circuit sine-state QPE with global depolarizing noise), with the Toffoli cost for the controlled-walk operator estimated using QUALTRAN. We assume a model of fault-tolerant computation based on CCZ resource states and show cost estimates based on using either a single CCZ factory or two of them; this allows us to highlight the crossover point when it can be convenient to invest more physical qubits to implement multiple magic state factories. The size of the surface-code and CCZ factory parameters are chosen to minimize computation volume while keeping the error rate below γ , assuming a physical error rate of 10^{-3} and a surface-code clock cycle of $1 \mu\text{s}$.

2. Cost estimates for electronic structure calculations using THC

We compute the quantum resources for a number of molecular systems of small to medium size to illustrate how the higher complexity of molecular Hamiltonians and the more stringent precision requirements affect the required quantum resources and run times. All quantum chemistry calculations are performed with the help of PySCF [139]. For H_2O , N_2 , naphthalene, and anthracene, the active spaces have been constructed either from restricted Hartree-Fock (RHF) orbitals or using the atomic valence active space technique (AVAS) [140] as implemented in PySCF. The CAS(27, 26) of the Co(salophen) complex has been determined with the help of the Active Space Finder (ASF) [141] software developed by HQS Quantum Simulations and Covestro (for details, see Table II).

The resulting active-space Hamiltonians have first been symmetry shifted as proposed in Ref. [85]. This means finding shifts a_1 and a_2 and modifying the molecular Hamiltonian H by adding powers of the total-particle-number operator N_e according to

$$H' = H + a_1 N_e + a_2 N_e^2. \quad (103)$$

This leaves the eigenvectors and the spectrum of the Hamiltonian inside the particle-number sectors intact up to a particle-number-dependent offset that can be easily subtracted classically but reduces the norm of the coefficients of the Hamiltonian. Concretely, one picks a_1 as the median of the eigenvalues of the one-body part of the Hamiltonian and a_2 as the median of a certain diagonal of the two-body tensor [85].

The shifted Hamiltonian H' of each molecule has then been factorized with the tensor-hypercontraction (THC) procedure with L2 regularization with the penalty parameter ρ_{THC} as implemented in OpenFermion [142], leading to a λ factor denoted by λ' . Symmetry shifting leaves the coupled cluster with singles and doubles (CCSD) energy unaffected but does cause very small changes in the triples correction contained in CCSD(T). It has thus been more convenient to choose the THC rank based on the constraint that the CCSD energy error, ΔE_{CCSD} , should stay below 1 mEh. Consistent with what has been reported in Ref. [94], we have also observed roughly a factor of 2 reduction of the lambda value due to symmetry shifting.

The logical resources for obtaining QPE estimates with $\epsilon = 10^{-3}/\lambda'$ precision (see Fig. 10) have then been computed using QUALTRAN [86], from the THC tensors. The

TABLE II. The orbital selection method, the selected active spaces, and the THC factorization parameters and results for the molecular systems used in the resource estimates. The AVAS [140] and ASF [141] methods have been used for some systems.

Molecule	Active space	Selection method	THC rank	ρ_{THC}	λ'	ΔE_{CCSD} (mEh)
N ₂	CAS(6, 6)	Around HOMO-LUMO	30	10^{-4}	4.66	0.62
H ₂ O	CAS(8, 6)	AVAS (H 1s, O 2s2p)	30	10^{-4}	6.97	0.63
Naphthalene	CAS(10, 10)	AVAS (C 2p _z)	45	10^{-5}	6.45	0.67
Anthracene	CAS(14, 14)	AVAS (C 2p _z)	80	10^{-5}	12.12	0.19
Co(salophen)	CAS(27, 26)	ASF default settings	100	10^{-5}	34.58	0.44

logical circuit for the controlled-walk operator is constructed using the algorithmic primitives included in QUALTRAN and from that, both the number of logical qubits and Toffoli gates are computed (assuming that rotations in PREPARE and SELECT are implemented to 10-bit and 16-bit precision, respectively [7]). The logical cost for the controlled walk operator is converted into physical costs for Algorithm 5 as a function of the residual noise rate γ , under the same assumptions made for the Hubbard model.

3. Comparison with the fully fault-tolerant algorithm

For the sake of comparison in Fig. 3, we have also estimated the resources required for running the standard fault-tolerant single-circuit sine-state QPE algorithm. The standard fault-tolerant algorithm is defined by choosing an acceptable failure probability δ as a free parameter. We fix the logical circuit to depth $T = T_{\text{tot}} = \lambda\pi/\Delta E$ to ensure a phase estimate to the desired precision in the absence

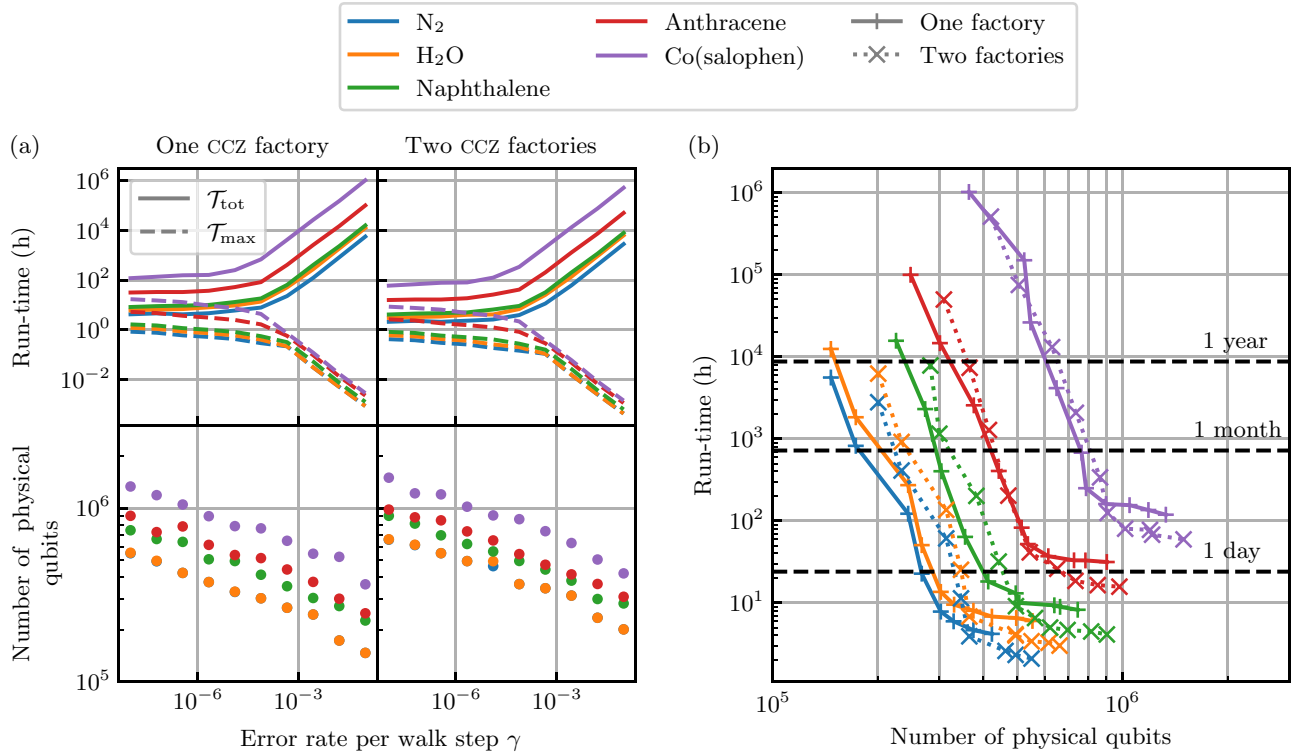


FIG. 10. The physical costs for MSQPE applied to the electronic structure Hamiltonian factorized with THC, with a target precision on the phase of $\epsilon_t = 10^{-3}/\lambda'$ (where λ' is the symmetry-shifted THC lambda value). (a) The run-time and qubit costs as a function of the error rate γ per step of the qubitized walk. (b) The relation between the number of physical qubits and the run-time. We assume that the residual error (originating from imperfect error correction and imperfect distillation) can be modeled as a global-depolarizing-noise channel of strength γ . The logical circuits are defined according to Algorithm 5 (multi-circuit sine-state QPE with global depolarizing noise), with the Toffoli cost and logical qubit requirements for the walk operator estimated using QUALTRAN (assuming that rotations in PREPARE and SELECT are specified to 10 and 16 bits of precision, respectively). We assume a model of fault-tolerant computation based on CCZ resource states and show cost estimates based on using either a single CCZ factory or two of them; this allows us to highlight the crossover point at which it can be convenient to invest more physical qubits to implement multiple magic state factories. The size of the surface-code and CCZ factory parameters are chosen to minimize the computation volume while keeping the error rate below γ , assuming a physical error rate of 10^{-3} and a surface-code clock cycle of 1 μ s.

of errors. The surface code and distillation distances are then chosen to optimize the computational volume while keeping the total error probability below δ , under the same assumptions as above. The resulting requirements are reported in Fig. 3 for $\delta = 1\%$ ($\blacktriangle$) and $\delta = 0.1\%$ ($\blacktriangledown$), together with the requirements for the circuit-division algorithm.

The costs of this standard approach should only be considered a reference and not a fair comparison, as the circuit-division algorithm targets the estimate of the phase with variance bounded by ϵ^2 , without allowing for any failure probability. A failure probability of δ , assuming that the error produces a random phase estimate, would result in a final estimate with a much larger variance $(1 - \delta)\epsilon^2 + \delta\lambda^2\pi^2$. While prior knowledge about the target phase can be used to filter these results with a repeat-until-success strategy, effectively reducing the expected error, the same technique can be applied to circuit-division QPE as suggested in Sec. IV E 5.

Note that in most cases, the cost of phase-estimation circuits on a qubitized walk operator can be improved by an additional factor of ≈ 2 by using the unary iteration circuit to control the reflect operator, as described in Ref. [7]. This improvement is not considered in this work, as it would require a subtle modification of the parameter-selection procedure for MSQPE. Nevertheless, we expect that this would yield an improvement by about a factor of 2 in the total run-time.

ACKNOWLEDGMENTS

We acknowledge useful discussions with Oumarou Oumarou, Kianna Wan, Ryan Babbush, and Rafał Demkowicz-Dobrzański, we thank Rolando Somma and Nobuyuki Yoshioka for useful feedback on the manuscript, and we thank Carlo Beenakker for support. We thank the Lorentz Center for hosting us during the workshop “Bridging the gap between classical & quantum simulation,” enabling the progress of this work. S.P. acknowledges support from Shell Global Solutions BV during his time in Leiden. A.D. is supported by a Google Ph.D. Fellowship.

A. D., S. P., and T. E. O. conceived the main ideas, developed the algorithms, and conceived and proved the theorems. S. P. and T. E. O. conceived the error-mitigation scheme. W. J. H. provided the performance estimate for early fault tolerance in Sec. IV A. S. P., A. D., and M. S. developed the cost-estimates code. M. S. provided the molecular models for the chemistry cost estimates. All authors contributed to interpreting the results, writing, and revising the manuscript.

DATA AVAILABILITY

The data that support the findings of this paper as well as PYTHON code for performing the logical and physical resource estimates of MSQPE and all used

molecular Hamiltonians and their factorizations are openly available [79].

-
- [1] R. Acharya *et al.*, (Google Quantum AI), Suppressing quantum errors by scaling a surface code logical qubit, *Nature* **614**, 676 (2023).
 - [2] M. P. da Silva *et al.*, Demonstration of logical qubits and repeated error correction with better-than-physical error rates, *Arxiv:2404.02280*.
 - [3] D. Bluvstein *et al.*, Logical quantum processor based on reconfigurable atom arrays, *Nature* **626**, 58 (2024).
 - [4] R. Acharya *et al.*, Quantum error correction below the surface code threshold, *Arxiv:2408.13687*.
 - [5] 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).
 - [6] 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).
 - [7] 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).
 - [8] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
 - [9] A. G. Fowler and C. Gidney, Low overhead quantum computation using lattice surgery, *Arxiv:1808.06709*.
 - [10] D. Litinski, A game of surface codes: Large-scale quantum computing with lattice surgery, *Quantum* **3**, 128 (2019).
 - [11] C. Gidney and A. G. Fowler, Efficient magic state factories with a catalyzed $|CCZ\rangle$ to $2|T\rangle$ transformation, *Quantum* **3**, 135 (2019).
 - [12] E. T. Campbell, Early fault-tolerant simulations of the Hubbard model, *Quantum Sci. Technol.* **7**, 015007 (2022).
 - [13] A. Katabarwa, K. Gratsea, A. Caesura, and P. D. Johnson, Early fault-tolerant quantum computing, *PRX Quantum* **5**, 020101 (2024).
 - [14] R. Zhang, G. Wang, and P. Johnson, Computing ground state properties with early fault-tolerant quantum computers, *Quantum* **6**, 761 (2022).
 - [15] R. Kshirsagar, A. Katabarwa, and P. D. Johnson, On proving the robustness of algorithms for early fault-tolerant quantum computers, *Quantum* **8**, 1531 (2024).
 - [16] G. Wang, S. Sim, and P. D. Johnson, State preparation boosters for early fault-tolerant quantum computation, *Quantum* **6**, 829 (2022).
 - [17] G. Wang, D. S. 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, *Quantum* **7**, 1167 (2023).
 - [18] J. S. Nelson and A. D. Baczewski, An assessment of quantum phase-estimation protocols for early fault-tolerant quantum computers, *Arxiv:2403.00077*.

- [19] K. Wan, M. Berta, and E. T. Campbell, Randomized quantum algorithm for statistical phase-estimation, *Phys. Rev. Lett.* **129**, 030503 (2022).
- [20] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, *PRX Quantum* **3**, 010318 (2022).
- [21] Y. Dong, L. Lin, and Y. Tong, Ground state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices, *PRX Quantum* **3**, 040305 (2022).
- [22] 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, *PRX Quantum* **4**, 020331 (2023).
- [23] Z. Ding, Y. Dong, Y. Tong, and L. Lin, Robust ground-state energy estimation under depolarizing noise, *ArXiv:2307.11257*.
- [24] D. Bultrini, S. Wang, P. Czarnik, M. H. Gordon, M. Cerezo, P. J. Coles, and L. Cincio, The battle of clean and dirty qubits in the era of partial error correction, *Quantum* **7**, 1060 (2023).
- [25] Q. Liang, Y. Zhou, A. Dalal, and P. Johnson, Modeling the performance of early fault-tolerant quantum algorithms, *Phys. Rev. Res.* **6**, 023118 (2024).
- [26] Y. Akahoshi, K. Maruyama, H. Oshima, S. Sato, and K. Fujii, Partially fault-tolerant quantum computing architecture with error-corrected Clifford gates and space-time efficient analog rotations, *PRX Quantum* **5**, 010337 (2024).
- [27] R. Toshio, Y. Akahoshi, J. Fujisaki, H. Oshima, S. Sato, and K. Fujii, Practical quantum advantage on partially fault-tolerant quantum computer, *ArXiv:2408.14848*.
- [28] Y. Akahoshi, R. Toshio, J. Fujisaki, H. Oshima, S. Sato, and K. Fujii, Compilation of Trotter-based time evolution for partially fault-tolerant quantum computing architecture, *ArXiv:2408.14929*.
- [29] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, *Proc. Natl. Acad. Sci.* **115**, 9456 (2018).
- [30] N. S. Blunt, G. P. Gehér, and A. E. Moylett, Compilation of a simple chemistry application to quantum error correction primitives, *Phys. Rev. Res.* **6**, 013325 (2024).
- [31] D. Wang, O. Higgott, and S. Brierley, Accelerated variational quantum eigensolver, *Phys. Rev. Lett.* **122**, 140504 (2019).
- [32] Z. Ding and L. Lin, Simultaneous estimation of multiple eigenvalues with short-depth quantum circuit on early fault-tolerant quantum computers, *Quantum* **7**, 1136 (2023).
- [33] H. Ni, H. Li, and L. Ying, On low-depth algorithms for quantum phase-estimation, *Quantum* **7**, 1165 (2023).
- [34] I. Y. Akhalwaya, S. Ubaru, K. L. Clarkson, M. S. Squillante, V. Jejjala, Y.-H. He, K. Naidoo, V. Kalantzis, and L. Horesh, Topological data analysis on noisy quantum computers, *ArXiv:2209.09371*.
- [35] 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).
- [36] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error-minimization technique: Applications from NISQ to FTQC eras, *PRX Quantum* **3**, 010345 (2022).
- [37] 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, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [38] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation, *ArXiv:quant-ph/9611028*.
- [39] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023 (Association for Computing Machinery, New York, 2023), p. 945.
- [40] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, *npj Quantum Inf.* **8**, 114 (2022).
- [41] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, *ArXiv:2210.11505*.
- [42] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, *Phys. Rev. Lett.* **131**, 210602 (2023).
- [43] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Universal cost bound of quantum error mitigation based on quantum estimation theory, *Phys. Rev. Lett.* **131**, 210601 (2023).
- [44] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [45] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, *Phys. Rev. X* **8**, 031027 (2018).
- [46] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [47] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, *Phys. Rev. Lett.* **122**, 180501 (2019).
- [48] 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).
- [49] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, *Phys. Rev. X* **11**, 041036 (2021).
- [50] 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).
- [51] A. Montanaro and S. Stanisic, Error mitigation by training with fermionic linear optics, *ArXiv:2102.02120*.
- [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] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, and L. DiCarlo, Experimental error mitigation via symmetry verification in a variational quantum eigensolver, *Phys. Rev. A* **100**, 010302(R) (2019).

- [54] E. van den Berg, Z. K. Mineev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors, *Nat. Phys.* **19**, 1116 (2023).
- [55] S. Stanisic, J. L. Bosse, F. M. Gamba, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer, *Nat. Commun.* **13**, 5743 (2022).
- [56] T. E. O'Brien *et al.*, Purification-based quantum error mitigation of pair-correlated electron simulations, *Nat. Phys.* **19**, 1787 (2023).
- [57] F. Arute *et al.*, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
- [58] A. Y. Kitaev, Quantum measurements and the Abelian Stabilizer Problem, [ArXiv:quant-ph/9511026](https://arxiv.org/abs/quant-ph/9511026).
- [59] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2001), 2nd ed., Vol. 54, p. 60.
- [60] W. van Dam, G. M. D'Ariano, A. Ekert, C. Macchiavello, and M. Mosca, Optimal quantum circuits for general phase-estimation, *Phys. Rev. Lett.* **98**, 090501 (2007).
- [61] P. Najafi, P. C. S. Costa, and D. W. Berry, Optimum phase-estimation with two control qubits, [ArXiv:2303.12503](https://arxiv.org/abs/2303.12503).
- [62] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate-set via robust phase-estimation, *Phys. Rev. A* **92**, 062315 (2015).
- [63] 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).
- [64] R. D. Somma, Quantum eigenvalue estimation via time series analysis, *New J. Phys.* **21**, 123025 (2019).
- [65] 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).
- [66] S. Lu, M. C. Bañuls, and J. I. Cirac, Algorithms for quantum simulation at finite energies, *PRX Quantum* **2**, 020321 (2021).
- [67] Y. Gu, Y. Ma, N. Forcellini, and D. E. Liu, Noise-resilient phase-estimation with randomized compiling, *Phys. Rev. Lett.* **130**, 250601 (2023).
- [68] Y. Yang, A. Christianen, M. C. Bañuls, D. S. Wild, and J. I. Cirac, Phase-sensitive quantum measurement without controlled operations, *Phys. Rev. Lett.* **132**, 220601 (2024).
- [69] O. Kiss, U. Azad, B. Reuena, A. Roggero, D. Wakeham, and J. M. Arrazola, Early fault-tolerant quantum algorithms in practice: Application to ground-state energy estimation, *Quantum* **9**, 1682 (2025).
- [70] G. Rendon and P. D. Johnson, Low-depth gaussian state energy estimation, [ArXiv:2309.16790](https://arxiv.org/abs/2309.16790).
- [71] D. Castaldo, S. Jahangiri, A. Migliore, J. M. Arrazola, and S. Corni, A differentiable quantum phase-estimation algorithm, [ArXiv:2406.14113](https://arxiv.org/abs/2406.14113).
- [72] F. Belliard and V. Giovannetti, Achieving Heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error, *Phys. Rev. A* **102**, 042613 (2020).
- [73] V. Dunjko, Y. Ge, and J. I. Cirac, Computational speedups using small quantum devices, *Phys. Rev. Lett.* **121**, 250501 (2018).
- [74] Y. Ge and V. Dunjko, A hybrid algorithm framework for small quantum computers with application to finding Hamiltonian cycles, *J. Math. Phys.* **61**, 012201 (2020).
- [75] A. M. Childs, R. Kothari, M. Kovacs-Deak, A. Sundaram, and D. Wang, Quantum divide and conquer, *ACM TQC* **6**, 17 (2025).
- [76] K. Fujii, K. Mizuta, H. Ueda, K. Mitarai, W. Mizukami, and Y. O. Nakagawa, Deep variational quantum eigensolver: A divide-and-conquer method for solving a larger problem with smaller size quantum computers, *PRX Quantum* **3**, 010346 (2022).
- [77] A. Pérez-Salinas, R. Draškić, J. Tura, and V. Dunjko, Shallow quantum circuits for deeper problems, *Phys. Rev. A: At., Mol. Opt. Phys.* **108**, 062423 (2023).
- [78] G. Wang, D. S. França, G. Rendon, and P. D. Johnson, Efficient ground-state-energy estimation and certification on early fault-tolerant quantum computers, *Phys. Rev. A* **111**, 012426 (2025).
- [79] S. Polla and A. Dutkiewicz, StefanoPolla/EarlyFT_QPE: V0.1.0, [GitHub repository](https://github.com/StefanoPolla/EarlyFT_QPE), [Zenodo release](https://zenodo.org/record/10000000), 2025.
- [80] D. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, Surface code quantum computing by lattice surgery, *New J. Phys.* **14**, 123011 (2012).
- [81] R. Raussendorf and J. Harrington, Fault-tolerant quantum computation with high threshold in two dimensions, *Phys. Rev. Lett.* **98**, 190504 (2007).
- [82] A. G. Fowler, A. M. Stephens, and P. Groszkowski, High-threshold universal quantum computation on the surface code, *Phys. Rev. A* **80**, 052312 (2009).
- [83] G. H. Low and I. L. Chuang, Hamiltonian simulation by qubitization, *Quantum* **3**, 163 (2019).
- [84] J. P. F. LeBlanc *et al.*, Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms, *Phys. Rev. X* **5**, 041041 (2015).
- [85] D. Rocca, C. L. Cortes, J. F. Gonthier, P. J. Ollitrault, R. M. Parrish, G.-L. Anselmetti, M. Degroote, N. Moll, R. Santagati, and M. Streif, Reducing the runtime of fault-tolerant quantum simulations in chemistry through symmetry-compressed double factorization, *J. Chem. Theory Comput.* **20**, 4639 (2024).
- [86] M. P. Harrigan, T. Khattar, C. Yuan, A. Peduri, N. Yosri, F. D. Malone, R. Babbush, and N. C. Rubin, Expressing and analyzing quantum algorithms with QUALTRAN, [ArXiv:2409.04643](https://arxiv.org/abs/2409.04643).
- [87] C. Gidney, N. Shutty, and C. Jones, Magic state cultivation: Growing T states as cheap as CNOT gates, [ArXiv:2409.17595](https://arxiv.org/abs/2409.17595).
- [88] J. Kołodyński and R. Demkowicz-Dobrzański, Efficient tools for quantum metrology with uncorrelated noise, *New J. Phys.* **15**, 073043 (2013).
- [89] S. T. Merkel, J. M. Gamba, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, *Phys. Rev. A—At., Mol. Opt. Phys.* **87**, 062119 (2013).
- [90] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of

- qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nat. Commun.* **8**, 14485 (2017).
- [91] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
- [92] N. M. Tubman, C. Mejuto-Zaera, J. M. Epstein, D. Hait, D. S. Levine, W. Huggins, Z. Jiang, J. R. McClean, R. Babbush, M. Head-Gordon, and K. B. Whaley, Postponing the orthogonality catastrophe: Efficient state preparation for electronic structure simulations on quantum devices, *ArXiv:1809.05523*.
- [93] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, M. Kastoryano, R. Babbush, J. Preskill, D. R. Reichman, E. T. Campbell, E. F. Valeev, L. Lin, and G. K.-L. Chan, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, *Nat. Commun.* **14**, 1952 (2023).
- [94] S. Fomichev, K. Hejazi, M. S. Zini, M. Kiser, J. F. Morales, P. A. M. Casares, A. Delgado, J. Huh, A.-C. Voigt, J. E. Mueller, and J. M. Arrazola, Initial state preparation for quantum chemistry on quantum computers, *ArXiv:2310.18410*.
- [95] D. W. Berry, Y. Tong, T. Khattar, A. White, T. I. Kim, S. Boixo, L. Lin, S. Lee, G. K.-L. Chan, R. Babbush, and N. C. Rubin, Rapid initial state preparation for the quantum simulation of strongly correlated molecules, *ArXiv:2409.11748*.
- [96] Y. Ge, J. Tura, and J. I. Cirac, Faster ground state preparation and high-precision ground energy estimation with fewer qubits, *J. Math. Phys.* **60**, 022202 (2019).
- [97] L. Lin and Y. Tong, Near-optimal ground state preparation, *Quantum* **4**, 372 (2020).
- [98] A. Moitra, in *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15 (Association for Computing Machinery, New York, 2015), p. 821.
- [99] M. McEwen *et al.*, Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits, *Nat. Phys.* **18**, 107 (2022).
- [100] X. Pan, Y. Zhou, H. Yuan, L. Nie, W. Wei, L. Zhang, J. Li, S. Liu, Z. H. Jiang, G. Catelani, L. Hu, F. Yan, and D. Yu, Engineering superconducting qubits to reduce quasiparticles and charge noise, *Nat. Commun.* **13**, 7196 (2022).
- [101] M. McEwen *et al.*, Resisting high-energy impact events through gap engineering in superconducting qubit arrays, *Phys. Rev. Lett.* **133**, 240601 (2024).
- [102] W. J. Huggins, B. A. O’Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, *Nature* **603**, 416 (2022).
- [103] M. Lostaglio and A. Ciani, Error mitigation and quantum-assisted simulation in the error corrected regime, *Phys. Rev. Lett.* **127**, 200506 (2021).
- [104] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, *SIAM J. Comput.* **35**, 1070 (2006).
- [105] P. Wocjan and S. Zhang, Several natural BQP-complete problems, *ArXiv:quant-ph/0606179*.
- [106] S. Gharibian and F. L. Gall, Dequantizing the quantum singular value transformation: Hardness and applications to quantum chemistry and the quantum PCP Conjecture, *SIAM J. Comput.* **52**, 1009 (2023).
- [107] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, *Proc. R. Soc. London. Series A: Math. Phys. Eng. Sci.* **454**, 339 (1998).
- [108] D. W. Berry and H. M. Wiseman, Optimal states and almost optimal adaptive measurements for quantum interferometry, *Phys. Rev. Lett.* **85**, 5098 (2000).
- [109] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, *Phys. Rev. Res.* **3**, 033055 (2021).
- [110] R. B. Griffiths and C.-S. Niu, Semiclassical Fourier transform for quantum computation, *Phys. Rev. Lett.* **76**, 3228 (1996).
- [111] P. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (IEEE, Santa Fe, NM, 1994), p. 124.
- [112] M. Mosca and A. Ekert, The Hidden Subgroup Problem and Eigenvalue Estimation on a Quantum Computer, *ArXiv:quant-ph/9903071*.
- [113] C. Zalka, Fast versions of Shor’s quantum factoring algorithm, *ArXiv:quant-ph/9806084*.
- [114] S. Parker and M. B. Plenio, Efficient factorization with a single pure qubit and $\log N$ mixed qubits, *Phys. Rev. Lett.* **85**, 3049 (2000).
- [115] S. Beauregard, Circuit for Shor’s algorithm using $2n + 3$ qubits, *QIC* **3**, 175 (2003).
- [116] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Entanglement-free Heisenberg-limited phase-estimation, *Nature* **450**, 393 (2007).
- [117] K. M. Svore, M. B. Hastings, and M. Freedman, Faster phase-estimation, *ArXiv:1304.0741*.
- [118] N. Wiebe and C. Granade, Efficient Bayesian phase-estimation, *Phys. Rev. Lett.* **117**, 010503 (2016).
- [119] A. E. Russo, K. M. Rudinger, B. C. A. Morrison, and A. D. Baczewski, Evaluating energy differences on a quantum computer with robust phase-estimation, *Phys. Rev. Lett.* **126**, 210501 (2021).
- [120] E. van den Berg, Efficient Bayesian phase-estimation using mixed priors, *Quantum* **5**, 469 (2021).
- [121] A. Roggero, Spectral density estimation with the Gaussian integral transform, *Phys. Rev. A* **102**, 022409 (2020).
- [122] Harald Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).
- [123] C. R. Rao, in *Breakthroughs in Statistics*, edited by S. Kotz and N. L. Johnson (Springer New York, 1992), p. 235.
- [124] G. Casella and R. L. Berger, *Statistical Inference* (Thomson Learning, Pacific Grove, CA, 2002), 2nd ed.
- [125] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, *Science (New York)* **306**, 1330 (2004).
- [126] C. King, The capacity of the quantum depolarizing channel, *ArXiv:quant-ph/0204172*.
- [127] J. Vovrosh, K. E. Khosla, S. Greenaway, C. Self, M. S. Kim, and J. Knolle, Simple mitigation of global depolarizing errors in quantum simulations, *Phys. Rev. E* **104**, 035309 (2021).

- [128] C. Gidney, Halving the cost of quantum addition, [Quantum](#) **2**, 74 (2018).
- [129] Y. R. Sanders, D. W. Berry, P. C. S. Costa, L. W. Tessler, N. Wiebe, C. Gidney, H. Neven, and R. Babush, Compilation of fault-tolerant quantum heuristics for combinatorial optimization, [PRX Quantum](#) **1**, 020312 (2020).
- [130] L. Hales and S. Hallgren, in *Proceedings 41st Annual Symposium on Foundations of Computer Science* (IEEE, Redondo Beach, CA, 2000), p. 515.
- [131] M. Mosca and C. Zalka, Exact quantum fourier transforms and discrete logarithms, [ArXiv:quant-ph/0301093](#).
- [132] The $C_{\text{em}} \sim F^{-1}$ bound for symmetry verification assumes that the number of symmetries increases with the system size using, e.g., schemes developed in Ref. [143]. Technically, when postselecting on a single symmetry, symmetry verification achieves constant C_{em} , but except in special cases this only removes around half the error, which is typically insufficient for error mitigation unless combined with a separate scheme..
- [133] X. Mi *et al.*, Information scrambling in quantum circuits, [Science](#) **374**, 1479 (2021).
- [134] S. Filippov, M. Leahy, M. A. C. Rossi, and G. García-Pérez, Scalable tensor-network error mitigation for near-term quantum computing, [ArXiv:2307.11740](#).
- [135] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Learning logical Pauli noise in quantum error correction, [Phys. Rev. Lett.](#) **130**, 200601 (2023).
- [136] S. Bravyi, M. Englbrecht, R. Koenig, and N. Peard, Correcting coherent errors with surface codes, [npj Quantum Inf.](#) **4**, 55 (2018).
- [137] N. Linden and R. de Wolf, Average-case verification of the quantum Fourier transform enables worst-case phase-estimation, [Quantum](#) **6**, 872 (2022).
- [138] J. van Apeldoorn, A. Cornelissen, A. Gilyén, and G. Nannicini, Quantum tomography using state-preparation unitaries, [ArXiv:2207.08800](#).
- [139] Q. Sun *et al.*, Recent developments in the PySCF program package, [J. Chem. Phys.](#) **153**, 024109 (2020).
- [140] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, Automated construction of molecular active spaces from atomic valence orbitals, [J. Chem. Theory Comput.](#) **13**, 4063 (2017).
- [141] HQS Quantum Simulations ActiveSpaceFinder, GitHub repository (2024)..
- [142] J. R. McClean *et al.*, OpenFermion: The electronic structure package for quantum computers, [Quantum Sci. Technol.](#) **5**, 034014 (2020).
- [143] C. Derby, J. Klassen, J. Bausch, and T. Cubitt, Compact fermion to qubit mappings, [Phys. Rev. B](#) **104**, 035118 (2021).