Modeling the performance of early fault-tolerant quantum algorithms

Qiyao Liang [1,2,*] Yiqing Zhou [1,3] Archismita Dalal [4] and Peter Johnson [1]

[1]*Zapata AI, Boston, Massachusetts 02110, USA*

[2]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[3]*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*

[4]*Zapata Computing Canada Inc., Toronto, Ontario M5V 2Y1, Canada*

Progress in fault-tolerant quantum computation (FTQC) has driven the pursuit of practical applications with *early fault-tolerant quantum computers* (EFTQC). These devices, limited in their qubit counts and fault-tolerance capabilities, require algorithms that can accommodate some degrees of error, which are known as EFTQC algorithms. To predict the onset of early quantum advantage, a comprehensive methodology is needed to develop and analyze EFTQC algorithms, drawing insights from both the methodologies of noisy intermediate-scale quantum and traditional FTQC. To address this need, we propose such a methodology for modeling algorithm performance on EFTQC devices under varying degrees of error. As a case study, we apply our methodology to analyze the performance of randomized Fourier estimation (RFE) [Kshirsagar, Katabarwa, and Johnson, arXiv:2209.11322], an EFTQC algorithm for phase estimation. We investigate the runtime performance and the fault-tolerant overhead of RFE in comparison to the traditional quantum phase estimation algorithm. Our analysis reveals that RFE achieves significant savings in physical qubit counts while having a much higher runtime upper bound. We anticipate even greater physical qubit savings when considering more realistic assumptions about the performance of EFTQC devices. By providing insights into the performance trade-offs and resource requirements of EFTQC algorithms, our work contributes to the development of practical and efficient quantum computing solutions on the path to quantum advantage.

I. INTRODUCTION

Despite significant experimental and theoretical progress [1,2], noisy intermediate-scale quantum (NISQ) devices have yet to exhibit the capacity to solve practical real-world problems with valuable outcomes. A promising avenue towards achieving practical quantum advantage lies in the development of architectures that can support large-scale fault-tolerant quantum computations (FTQC) [3]. By incorporating robust fault-tolerance capabilities, we can suppress errors in our computations to an arbitrary extent. However, this comes at the cost of resources that far exceed the capabilities of present-day devices by several orders of magnitude. Projections by researchers indicate that millions to billions of physical qubits would be required to outperform classical computers in tasks such as factoring and ground-state energy estimation [4–6].

There exists a substantial discrepancy between the capabilities of today's quantum devices and the projected resource requirements for practical large-scale fault-tolerant architectures. This discrepancy motivates the question: *How will the intermediate generation of devices, positioned between NISQ and FTQC, deliver practical advantage?* Such devices have recently been referred to as *early fault-tolerant quantum computers* (EFTQC) [7–9]. Notably, EFTQC devices would possess a limited number of physical qubits, thus imposing constraints on the distance of the error-correcting codes they can support. These deviate from the conventional assumptions of fault-tolerant quantum computing, where such resources are presumed to be. infinitely scalable.

A recent thrust in the field of quantum computing has been the development of EFTQC algorithms tailored to address the above limitations of EFTQC devices [7,8,10–22]. So far, two key considerations are central to the pursuit of practical value using EFTQC algorithms. The first is developing quantum algorithms that reduce the number of qubits and operations per circuit, often at the expense of increased circuit runs and consequently extending the runtime [7,11,13,14,18,19]. The second is designing quantum algorithms such that they are robust against gate and measurement errors [10–12,21,22]. These recent advancements showcase the potential applications of EFTQC devices, further motivating our previously posed question. An essential next step is to develop methodologies for assessing the performance of these algorithms, enabling a deeper understanding of how intermediate devices between NISQ and FTQC can be leveraged to attain practical value.

In this work, we conduct an analytical case study, first to our best knowledge, that links logical circuit error models to the performance of a variant of the EFTQC algorithm

---

*qiyao@mit.edu

mentioned above, the robust Fourier estimation (RFE) algorithm [10]. Specifically, we develop such a methodology to achieve the following:

(i) A proof that quantifies the performance of the RFE algorithm that interpolates between using a single oracle call (suited for the high-noise NISQ setting) and using many oracle calls per circuit (suited for the low-noise FTQC setting) (see Sec. V A)

(ii) A numerical demonstration of the suitability of this algorithm for EFTQC devices in that it can reduce by an order of magnitude the number of physical qubits required for large instances of phase estimation at the cost of an increase in runtime by several orders of magnitude (see Sec. V B).

To establish these results, we introduce a modularized methodology designed to assess the impact of generic circuit-level errors on a broad class of quantum algorithms. As a case study, we demonstrate the effectiveness of our methodology on the RFE algorithm, which shares a similar structure with other EFTQC-suited algorithms, such as those used for ground-state energy estimation [13,15,16,19] and property estimation [14]. The key feature common to these algorithms is the utilization of signals derived from Hadamard test outcomes, which makes our methodology applicable and adaptable to various cases within the EFTQC framework.

Our objective is to gain a comprehensive understanding of how errors impact the RFE algorithm. By doing so, we aim to leverage this understanding and extend it to algorithms designed for tasks beyond phase estimation. While previous studies have investigated the resilience of algorithms for quantum phase estimation [23], we choose to analyze the RFE algorithm for two key reasons. First, while sharing a similar structure to many of the above-mentioned EFTQC algorithms, the RFE algorithm is analytically tractable, making it an ideal candidate for our study. Second, its ability to smoothly interpolate between the high- to low-noise settings makes it well suited for the EFTQC regime.

The paper is organized as follows. We begin in Sec. II by introducing the RFE algorithm and our modifications to the algorithm for the purpose of this study. In Sec. III, we outline the framework of our methodology, which we apply to analyze the runtime performance of the RFE algorithm. Specifically in Sec. IV, we develop a chain of noise models from physical errors to algorithmic errors and study the impact of these errors on the performance of RFE in Sec. V A. Based on these results, we provide a fault-tolerant resource estimation, comparing RFE to the traditional quantum phase estimation algorithm in Sec. V B. Finally, in Sec. VI, we conclude by highlighting our key results and provide an outlook of this work.

II. RANDOMIZED FOURIER ESTIMATION

In this section, we introduce a variant of the RFE algorithm proposed in Ref. [10]. The RFE algorithm is used to solve the task of phase estimation, in which the goal is to estimate the phase angle θ defined by $U|\psi\rangle = e^{i\theta}|\psi\rangle$, assuming the ability to prepare the eigenstate $|\psi\rangle$ and to implement controlled-unitaries, c-U. The traditional approach to this problem is the quantum phase estimation (QPE) algorithm [24], which achieves the optimal performance asymptotically. However,
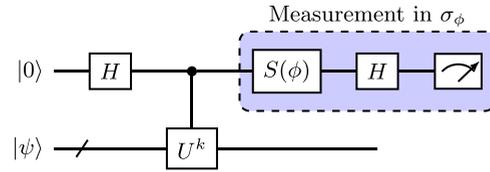
FIG. 1. Diagram of the circuits used in the RFE algorithm. The parameter k is uniformly randomly chosen among $\{0, \dots, K-1\}$ for each iteration. A key feature of the algorithm is that the parameter K controls the maximal circuit depth and is set to accommodate different degrees of error in the c-U operation: High error implies small K (low depth) and many repetitions, while low error warrants the use of large K (high depth) and fewer repetitions. The boxed-up elements in blue can be collectively interpreted as a measurement with respect to the observable $\sigma_\phi = \cos(\phi)\sigma_x - \sin(\phi)\sigma_y$, where σ_x and σ_y are the conventional Pauli operators and $S(\phi) = \begin{bmatrix} 1 & 0 \\ 0 & \exp(i\phi) \end{bmatrix}$.

the realization of the QPE algorithm requires multiple ancillary qubits and many high-fidelity quantum operations, both of which may be prohibitively costly given hardware constraints in the early fault-tolerant regime.

Given these constraints, alternative schemes for phase estimation have been proposed for near-term to early fault-tolerant devices [13,15,25,26]. In our study, we focus on the RFE algorithm [10], the performance of which can be analytically studied. Specifically, we will analyze a variant of the original RFE algorithm with slight modifications to simplify our analysis, as introduced later in this section. As the modifications made do not fundamentally change the mechanism of the algorithm, we will henceforth refer to the modified version of the algorithm as "RFE" throughout the paper.

The basic intuition behind this algorithm is summarized as follows. Each measurement outcome $z = \pm 1$ is generated from a sampled parameterized Hadamard test circuit in Fig. 1 with $k \in \{0, \dots, K-1\}$, where K sets the maximal circuit depth, and $\phi \in [0, 2\pi]$. The expected outcome of the measurement $\langle \sigma_\phi \rangle = \cos(\phi)\langle \sigma_x \rangle - \sin(\phi)\langle \sigma_y \rangle = \cos(k\theta + \phi)$ is an oscillatory function of k and θ and becomes the desired signal $e^{ik\theta}$ when averaged over a uniform distribution of ϕ between 0 and 2π, i.e., $\frac{1}{2\pi} \int_0^{2\pi} 2\cos(k\theta + \phi)e^{-i\phi} d\phi = e^{ik\theta}$.

To estimate the θ encoded in the frequency of the signal, we can construct from each outcome z an unbiased estimate $\hat{f}_j$ of the expected discrete Fourier transform f_j of signal $g(k)$: $= e^{ik\theta}$. In the Fourier domain $j \in \{0, \dots, J-1\}$, the estimate is expressed as

$$\hat{f}_j(k, \phi, z) = 2z e^{-i2\pi kj/J} e^{-i\phi}. \tag{1}$$

Given enough samples, we expect the magnitude of the averaged $\hat{f}_j$ to peak at a frequency close to θ. In the noiseless case, this peak will occur within the Fourier resolution $2\pi/J$ from the true θ, where we later set the parameter J such that the Fourier resolution matches the desired accuracy of the algorithm ϵ.

This peak frequency is then used as an estimate of θ, denoted $\hat{\theta}$. Figure 2(a) shows the real and imaginary components of the signal $g(k)$ as a function of k. The magnitude of the Fourier transformed signal $|f_j|$ is then plotted in Fig. 2(b) as a
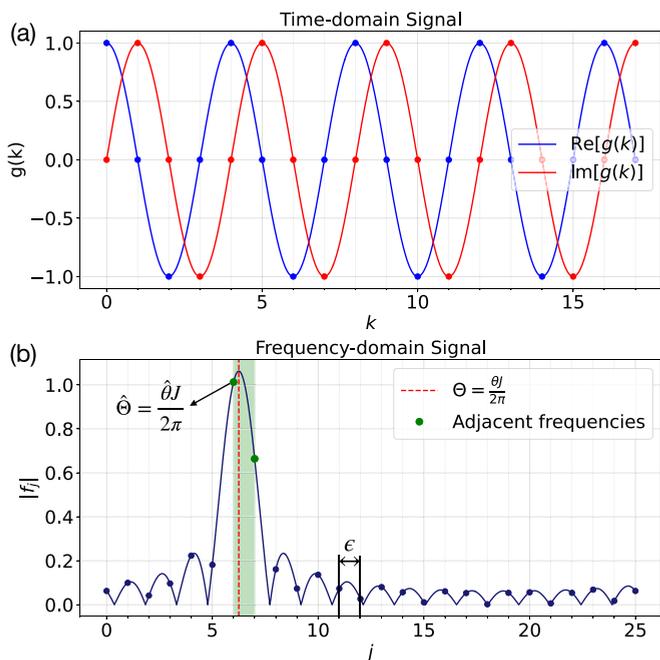
FIG. 2. (a) Real and imaginary components of the expected (noiseless) signal $\mathrm{Re}[g(k)] = \cos(k\theta)$ and $\mathrm{Im}[g(k)] = \sin(k\theta)$ as a function of the circuit depth k. (b) The magnitude of the expected Fourier transformed signal $|f_j|$. The green shaded region shows the acceptable range of values for $\hat{\theta}$, and the red dashed vertical line shows the true peak at index $\Theta = \frac{\theta J}{2\pi}$. The solid line represents the analytical functional forms where k and j are treated as continuous variables for easier visualization. The discrete values of $g(k)$ and $|f_j|$ are marked with dots. The two green dots in (b) correspond to $j = \lfloor J\theta/2\pi \rfloor$ and $\lceil J\theta/2\pi \rceil$, which we refer to as "adjacent frequencies."

function of j, where the peak occurs at an index near $\theta J/2\pi$ corresponding to the true frequency θ.

The algorithm that we introduce here differs from that of Ref. [10] in two regards. First, rather than taking samples corresponding to the real and imaginary parts of $g(k)$ separately by setting $\phi = 0$ or $\pi/2$, ϕ is chosen uniformly randomly such that we can construct an unbiased estimator for $g(k)$ with a single shot. This simplifies the algorithm analysis without any changes in its performance.

Second, the more substantial change is introduced to accommodate a more realistic noise model as developed in Sec. IV. This noise model results in an exponential attenuation (as a function of k) of the outcome probabilities, converging towards a uniform two-outcome distribution at large k, similarly to that of an unbiased coin toss. By appropriately setting the maximal value of k, i.e., K, we can minimize the impact of this attenuation on our estimated $\hat{\theta}$ from the signal.

In Ref. [10], the parameter K was used to set both the maximum value of k and the Fourier basis resolution $2\pi/K$. This can be an issue in the case when the attenuation is strong and high accuracy is required: Measurement outcomes for large k are uninformative because they are drawn from a nearly uniform distribution. To address this issue, we allow these two values to differ; K still labels the maximum value of k, while a new parameter J is used to set the Fourier resolution. Then, the high accuracy and high noise case is accommodated by setting J large and K small.

We now elaborate on why our algorithm works in the noiseless limit, which will provide intuition for its performance in the noisy case. One can calculate the probability of measuring the outcome z in the Hadamard circuit of Fig. 1 as

$$P(z|k, \phi; \theta) = \tfrac{1}{2}[1 + z \cdot \cos(k\theta + \phi)], \qquad (2)$$

which is an oscillatory function of k with frequency θ. The two classically sampled variables are drawn uniformly: $P(k) = 1/K$ and $P(\phi) = 1/2\pi$, where the former distribution is discrete while the latter is continuous. Using these distributions, the expected value of our constructed estimator (1) is calculated to be

$$\mathbb{E}[\hat{f}_j(k, \phi, z)] = \sum_{k=0}^{K-1} \int d\phi \sum_{z=\pm 1} P(k)P(\phi)P(z|k, \phi; \theta)\hat{f}_j(k, \phi, z) = \frac{1}{K}\frac{1 - a^K}{1 - a}, \qquad (3)$$

where $a := e^{i\theta - i(2\pi j/J)}$. For $j \in \mathbb{R}^+$, the expectation (3) is maximized at $j = \theta J/2\pi$, whereas for $j \in \mathbb{Z}^+$, i.e., in a discrete Fourier transform setting, and assuming $K \leqslant J$, the expectation achieves its maximum magnitude at $\lfloor \theta J/2\pi \rfloor$ or $\lceil \theta J/2\pi \rceil$ (or precisely at $\theta J/2\pi$ if $\theta J/2\pi \in \{0, \dots, J - 1\}$). Consequently, by setting $J = 2\pi/\epsilon$ we can then guarantee that the maximum of the expected discrete Fourier transform occurs at a frequency that is less than ϵ away from the true θ. The maximal circuit depth K will be chosen according to the target accuracy ϵ and a parameter λ [introduced in Eq. (12)] that characterizes the error strength in the c-U operation. Qualitatively, K is chosen to monotonically increase with $1/\epsilon$ and $1/\lambda$, which is described in further details in Sec. V A.

Our algorithm estimates the value of θ by using an average over multiple Fourier signal estimates (1). With M estimates $\hat{f}_j^{(n)}$ generated independently, their average will concentrate about the expected value $f_j := \mathbb{E}[\hat{f}_j(k, \phi, z)]$ (3), and a corresponding estimate of θ is given by

$$\hat{\theta} = \frac{2\pi}{J}\arg\max_j \left[\left|\frac{1}{M}\sum_{n=1}^{M}\hat{f}_j^{(n)}\right|\right]. \qquad (4)$$

This concentration will be addressed quantitatively in Sec. V A. The algorithm succeeds, i.e., yields an estimate such that $|\hat{\theta} - \theta| \leqslant \epsilon$, if one of the "adjacent" frequencies ($\lfloor \theta J/2\pi \rfloor$ or $\lceil \theta J/2\pi \rceil$) achieves the largest magnitude among all Fourier estimates $\hat{f}_j = \frac{1}{M}\sum_{n=1}^{M}\hat{f}_j^{(n)}$. In expectation, one of these "adjacent" frequencies [shown as the two green dots in Fig. 2(b)] will achieve the largest magnitude, with a finite gap between it and the magnitudes of the nonadjacent frequencies
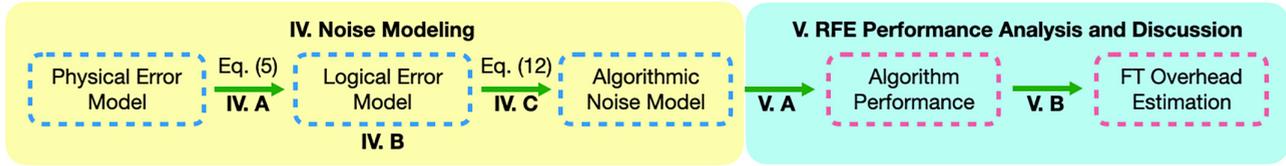
FIG. 3. Framework of our methodology.

(points that fall outside of the green shaded region). Hence, with sufficiently many samples M, the probability of failure of the algorithm can be made less than any finite failure probability δ.

III. FRAMEWORK

In this section we outline our proposed methodology for connecting the logical error model of an arbitrary quantum circuit to the success probability of a quantum algorithm, with the special case analysis of the RFE algorithm introduced in the last section as an example.

Our methodology, depicted in the flowchat of Fig. 3, can be summarized as follows. We begin by modeling the effect of error on an algorithmic level from physical to logical-level error models proposed in Sec. IV. Specifically, in Sec. IV A, we establish a fault-tolerant overhead model that relates physical error rate p_{phys} to logical error rate p_{logical} for a surface code of distance d. We then propose a generic N-qubit logical Pauli error channel in Sec. IV B and statistically quantify its impact on an N-qubit quantum circuit of depth D. To achieve this, we assume that a random N-qubit Pauli error occurs after each layer of unitaries in our circuit such that the resulting state is a mixture of random states drawn from a unitary 2-design. By computing the expected value and variance of measurement outcomes based on this probability distribution, we gain insights into their statistical properties, which allow us to develop an algorithmic noise model in Sec. IV C.

In Sec. V, we investigate the performance of RFE under the algorithmic noise model proposed in Sec. IV C, namely, the exponential decay noise. In Sec. V A, we give an upper bound on the algorithm runtime performance and analyze its scaling with respect to the desired degree of accuracy ϵ in the presence of various strength of decay λ. Finally, in Sec. V B, we provide a resource estimation of the RFE algorithm as compared to the standard QPE algorithm based on our original fault-tolerant overhead model proposed in Sec. IV A.

IV. NOISE MODELING

A. Fault-tolerant overhead model

We begin by establishing the connection between physical and logical error rates. Currently, physical error rates range from 10^{-3} to 10^{-4}, which are too large for reliable implementations of the RFE algorithm. To overcome this, we analyze the performance of our algorithm using lower logical error rates achievable through fault-tolerant computational protocols. Implementation errors at the circuit (logical) level arise from approximations in the operations (e.g., gate synthesis [27]) and uncorrected errors in the fault-tolerant protocols [28] (including magic state distillation and quantum error

correction). The failure probabilities in both of these cases can be systematically reduced by paying a cost in the number of physical qubits and the number of physical operations. Also, compiling abstract quantum operations to a logical architecture layout using techniques such as lattice surgery incur additional qubit costs [29–31]. These costs increase the memory and time resources of the quantum computation. In our analysis we will assume (1) that the overhead due to approximations in the operations has already been accounted for in the logical level gate counts, (2) that the logical qubit and gate overheads of magic state distillation are also accounted for in the logical level counts, and (3) that a single code distance d can sufficiently model the relationship between error reduction and quantum error correction overhead costs (noting that magic state factories and data qubits might use different code distances in practice). Reasonable deviations from these approximations would lead to slight changes in the quantitative results but would not alter the qualitative conclusions regarding the resource reduction afforded by the RFE algorithm.

In order to quantify the reduction in error rates from the physical to logical level, we adopt the model proposed in Ref. [6]. This reduction in error rate comes at the cost of an increase in time and number of physical qubits; equivalently, these resources can be thought of as convertible into error rate reduction. The quality of this conversion is governed by the ability of the particular architecture to maintain low physical gate error rates at scale. A model for this conversion as a function of resource overhead d is expressed as

$$p_{\text{logical}} = Ae^{-Bd}, \tag{5}$$

where the parameters A and B depend on the physical error rates [6]. The typical values for these parameters in the case of high (moderate) physical error rates are $A = 0.5(0.4)$ and $B = 1.6(1.1)$ [6]. Alternative values of these parameters for other approaches can be found in Table V in Ref. [32]. The overhead parameter d can be thought of as the code distance in the context of a surface code [33]. In this model, the physical qubit overhead is approximately $2d^2$. In our subsequent analysis, we approximate the logical error model as a composition of single-qubit depolarizing errors (13), where p_{logical} (5) represents the depolarizing rate. Based on these relationships, we can estimate the optimal performance of the RFE algorithm for given architecture parameters A and B, as we will elaborate in Sec. V A.

B. Logical gate error model

We consider a generic N-qubit Pauli error channel in order to study the impact of logical errors on the RFE algorithm. A generic N-qubit Pauli error channel acting on an N-qubit
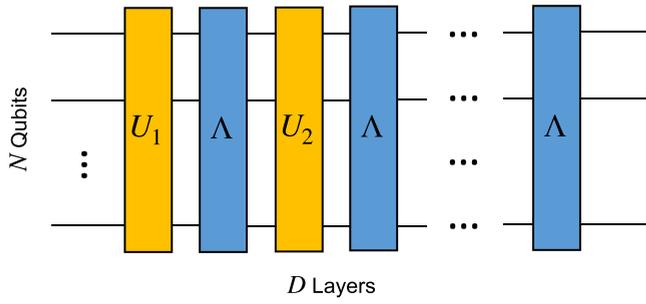
FIG. 4. Abstract circuit diagram of unitaries $U_1, \ldots, U_D$ interlaced with applications of a generic N-qubit Pauli error channel Λ in an N-qubit D-layer random quantum circuit.

density matrix ρ can be expressed by the following Kraus decomposition:

$$\Lambda(\rho) = \sum_{j=0}^{4^N-1} p_j A_j \rho A_j^\dagger, \tag{6}$$

where $A_j \in \{I, X, Y, Z\}^{\otimes N}$ and p_j is the probability of error A_j occurring. We note that here we have assumed the most generic Pauli error channel for our logical errors to follow the convention of most error-correction literature, which can later be modified into different desirable error channels based on the set of $\{p_j\}$ of our choosing. Given a quantum circuit that implements the ideal unitary $U = U_D \cdots U_2 U_1$ of depth D, we assume that the Pauli error Λ interleaves the layers of unitary $\mathcal{U}_1$ through $\mathcal{U}_D$, where $\mathcal{U}_i(\rho) := U_i \rho U_i^\dagger$ is the superoperator describing the action of the unitary U_i on the density operator ρ. The overall noisy circuit is illustrated in Fig. 4. The outcome state ρ_f on applying this noisy circuit to the initial state ρ_i is given by

$$\rho_f = \Lambda \circ \mathcal{U}_D \circ \cdots \circ \Lambda \circ \mathcal{U}_1(\rho_i)$$

$$= \sum_{j_D=0}^{4^N-1} \cdots \sum_{j_1=0}^{4^N-1} p_{j_D} \cdots p_{j_1} (A_{j_D} U_D \cdots A_{j_1} U_1) \rho_i$$

$$\times (U_1^\dagger A_{j_1}^\dagger \cdots U_D^\dagger A_{j_D}^\dagger). \tag{7}$$

After obtaining the noisy output quantum state (7), we now focus on the various statistical properties of the final measurement with respect to an N-qubit Pauli observable. Here we introduce the notion of a "trajectory state," defined as $|\psi_{\mathbf{j}}\rangle \equiv |\psi_{j_1 \ldots j_D}\rangle := A_{j_D} U_D \ldots A_{j_1} U_1 |0^N\rangle$, where $\mathbf{j}$ is the index tuple that includes all of $j_1, \ldots, j_D$. Each of these trajectory states $|\psi_{\mathbf{j}}\rangle$ corresponds to one combination of errors occurring on ρ_i among the 4ND possibilities with probability $p_{\mathbf{j}} = p_{j_1} \ldots p_{j_D}$. Expressing the expectation of an observable P with respect to the state ρ_f in Eq. (7) in terms of these trajectory states, we get

$$\langle P \rangle = \text{Tr}[\Lambda \circ \mathcal{U}_D \circ \cdots \circ \Lambda \circ \mathcal{U}_1(\rho_i)P], \tag{8}$$

$$= p_{\mathbf{0}} \text{Tr}[|\psi_{\mathbf{0}}\rangle\langle\psi_{\mathbf{0}}|P] + \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}} \text{Tr}[|\psi_{\mathbf{j}}\rangle\langle\psi_{\mathbf{j}}|P], \tag{9}$$

where $|\psi_{\mathbf{0}}\rangle = U_D \ldots U_1 |0^N\rangle$ is the ideal state on which no error has occurred. To provide a tractable analysis of $\langle P \rangle$, we propose the *unitary 2-design model*. Though different in their applications, other works that consider 2-designs include [34–36]. Under this approximation, we replace each of the noisy trajectory states $|\psi_{\mathbf{j}}\rangle$ with a randomly sampled state from a spherical 2-design [37] (i.e., any distribution over pure states whose first and second moments match those of the Haar distribution of unitaries over N-qubits applied to a reference state). This assumption establishes the statistics of $\langle P \rangle$, allowing us to compute its mean and variance as

$$\mathbb{E}[\langle P \rangle] = p_{\mathbf{0}}\langle\psi_{\mathbf{0}}|P|\psi_{\mathbf{0}}\rangle \text{ and} \tag{10}$$

$$\text{Var}[\langle P \rangle] = \frac{1}{2^N+1} \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2, \tag{11}$$

respectively. By our assumption, each of these noisy trajectories shares the same mean. Consequently, as an increasing number of trajectories are averaged over to form the second term in Eq. (10), their collective average will converge towards this mean. A detailed calculation of the above quantities is provided in Appendix A. We also point out that the unitary 2-design model proposed here to establish Eq. (10) and (11) assumes that the noisy trajectory states to be drawn uniformly randomly from all possible states, which is a strong assumption for errors happening in a worst-case scenario. An improved version of the model would likely benefit from system-specific knowledge of noises in the early fault-tolerant devices and their corresponding distributions of noisy trajectory states, which warrants further investigation.

Last, we briefly mention the potential impact of state preparation errors on our analysis. Small amounts of error in the initial state will not significantly affect the results of our paper. However, larger amounts of error can have a more pronounced effect. For example, if the initial state is not perfectly prepared in the ground state, then the peak in Fig. 2 will be suppressed in height and additional peaks may appear, each with a height proportional to the overlap between the imperfect initial state $|\tilde{\psi}\rangle$ and other eigenstates $|\psi\rangle$ of U, i.e., $|\langle\tilde{\psi}|\psi\rangle|^2 < 1$. To avoid overcomplicating the analysis, we choose to omit the consideration of state preparation errors in our noise modeling. We further note that a recent study [20] investigates the performance of a similar-spirited EFTQC algorithm for phase estimation as RFE under the effect of imperfect state preparation. Combining our work with the analysis of Ref. [20] would provide additional insights into the interplay between different error sources in the early fault-tolerant regime.

C. Logical gate error model to algorithmic noise model

Based on the statistics of $\langle P \rangle$ from Equations (10) and (11), we propose an algorithmic noise model that captures the effect of noise as a combination of exponential decay and random fluctuations on our algorithm. The exponential decay term stems from the $p_{\mathbf{0}}$ term in the mean of $\langle P \rangle$, whereas the random fluctuation is related to the variance of $\langle P \rangle$. For the RFE algorithm, the impact of error is to alter the probability

(2) of the measurement outcome z as

$$\Pr(z|k, \phi; \theta) = \frac{1}{2}[1 + ze^{-\lambda k}\cos(k\theta - \phi) + \eta_{k,\phi}], \quad (12)$$

where we introduce $\eta_{k,\phi}$ to represent a noise bias, i.e., random fluctuations, and λ to parametrize an exponential attenuation of outcome probability with respect to circuit depth k. Letting $\eta_{k,\phi}$ and λ vary arbitrarily, this "algorithmic" noise model is completely general and the algorithm would clearly not succeed in all settings. The key to establishing a reasonable algorithm performance is to limit, at least statistically, the magnitude of $\eta_{k,\phi}$ under different strengths of λ, as we elaborate later in this subsection.

We now analyze how RFE works in a noisy setting under our proposed algorithmic noise model (12). Given, for example, a single-qubit depolarizing error channel

$$\mathcal{D}(\rho) = (1 - r)\rho + \frac{r}{3}(X\rho X + Y\rho Y + Z\rho Z), \quad (13)$$

the N-qubit composite error channel is given by

$$\Lambda(\rho) = \mathcal{D}^{\otimes N}(\rho), \quad (14)$$

a special case of the N-qubit Pauli error channel from Eq. (6). On applying the operation $c\text{-}U$, where U is assumed to have D layers, k repetitions in the Hadamard test circuit, the probability of the signal remaining noiseless at the end is

$$p_{\text{total}} = p_{\mathbf{0}}^k = (1 - r)^{NDk}, \quad (15)$$

which decreases as a function of $c\text{-}U$ depth number k. From Eq. (10), we learn that the expected value of the quantum expectation value $\langle P \rangle$ decays with the total probability p_{total} after k iterates. This corresponds to a decay in the signal, i.e.,

$$e^{-\lambda} = (1 - r)^{ND}, \quad (16)$$

which makes the exponential decay parameter λ that we introduced in Eq. (12) to be $\lambda = -\ln[(1 - r)^{ND}]$. Similarly, we can substitute r from Eq. (13) into the variance expression $\text{Var}[\langle P \rangle]$ from Eq. (11)

$$\text{Var}[\langle P \rangle] = \frac{1}{2^N + 1}\{[(1 - r)^2 + 3(r/3)^2]^{NDk} - (1 - r)^{2NDk}\}. \quad (17)$$

We note that $\text{Var}[\langle P \rangle]$ here establishes the expected deviation β_j on our signal f_j in our algorithmic error model to be discussed next.

Under the effects of random fluctuations and exponential decay, we now arrive at a noisy expression of the expected $\hat{f}_j(k, \phi, z)$, similarly to the calculation in Eq. (3)

$$\mathbb{E}[\hat{f}_j(k, \phi, z)] = \beta_j + \frac{1}{K}\left[\frac{1 - e^{K(i\theta - i2\pi j/J - \lambda)}}{1 - e^{i\theta - i2\pi j/J - \lambda}}\right], \quad (18)$$

where $\beta_j = \mathbb{E}_k[\eta_{j,k}]$ is the expected deviation of our signal related to the quantity $\text{Var}[\langle P \rangle]$. We delay the detailed discussion of the variance to Appendix A 2. There we numerically show that the standard deviation of $\langle P \rangle$, $\sigma_{\langle P \rangle} := \sqrt{\text{Var}[\langle P \rangle]}$, is negligible compared to the signal magnitude ~ 1 for the parameter settings of our interest in the EFTQC regime. This means that the variance term drops out of our algorithmic error model (12), leaving exponential decay on the ideal signal as the sole source of error to be considered for the remaining
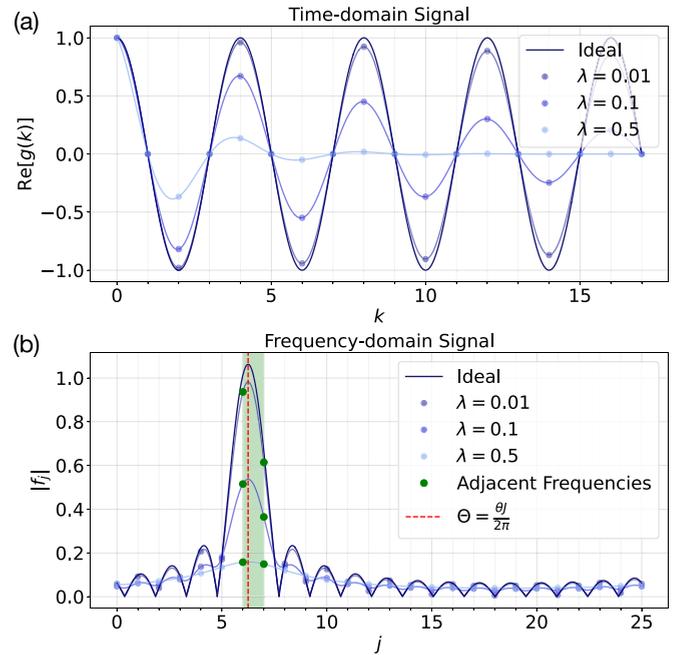
FIG. 5. Expected $\text{Re}[g(k)]$ (a) and $|f_j|$ (b) under the exponential decay noise with decay parameter $\lambda = 0.01, 0.1, 0.5$.

analysis. This result is indicative that our approach for modeling the noisy trajectory states with a unitary 2-design model is indeed consistent with a global depolarizing noise, the effect of which has been thoroughly studied in many EFTQC algorithms [12,21,26]. Hence, we set the β_j term to 0 for the rest of the discussion.

The effect of exponential decay on the expected signal from Eq. (18) is shown as a function of decay strength λ in Fig. 5. We note that as λ increases, the signal $g(k)$ decays exponentially as a function of circuit depth k, which results in a flatter spectrum in the expected Fourier signal amplitude $|f_j|$. This is problematic as the RFE algorithm relies on distinguishing the largest peak in the signal amplitude as our predicted phase index. A flattened spectrum means that the amplitude contrasts between neighboring peaks will decrease and, as a result, more measurements are needed to overcome shot noise in order to better discern the highest-amplitude point in the spectrum.

V. RFE PERFORMANCE ANALYSIS AND DISCUSSION

A. Link to algorithm performance analysis

The goal of this section is to establish the connection between the algorithmic noise model proposed in the previous section and the algorithm performance guarantee. As a reminder, the algorithm is said to succeed when the estimate $\hat{\theta}$ is within ϵ of θ, see Fig. 2(b). Our analysis determines an upper bound on the number of samples that are needed to ensure success with probability greater than $1 - \delta$.

The estimated $\hat{\theta}$ is calculated based on the discrete frequency point $2\pi j/J$ corresponding to the highest amplitude of $|\hat{f}_j|$ [as in Eq. (4)]. As a reminder, the consideration of a successful estimate of θ depends on whether θ is one of the values of $2\pi j/J$ or not. In the case when θ falls onto one of the

discrete frequency points, there are a total of three values of j leading to successful estimates, i.e., $j = J\theta/2\pi, J\theta/2\pi \pm 1$. In the case when θ falls between two points in the discrete frequency spectrum, as shown in Fig. 5(b), there are a total of two neighboring points that constitute successful guesses, namely $\lfloor J\theta/2\pi \rfloor$ and $\lceil J\theta/2\pi \rceil$.

Let j_* indicate the index of the "best estimate," that is, the integer closest to $J\theta/2\pi$. Any index j that is more than 1 away from $J\theta/2\pi$ will not correspond to an ϵ-accurate estimate according to our criteria for success. We refer to such estimates as "bad estimates." Hence, to guarantee that the algorithm succeeds, it suffices that $|\hat{f}_j|^2 < |\hat{f}_{j*}|^2$ for all j with $|j - J\theta/2\pi| > 1$. In other words, this condition ensures that no bad estimate has the largest $|f_j|$ so that the algorithm picks one of the ϵ-accurate estimates. Note that this is not a necessary condition; the algorithm could succeed even if $|\hat{f}_{j*}|^2$ were smaller than that of one of the bad estimates, as long as the other (or one among the two others) ϵ-accurate estimate had a magnitude larger than the rest.

The above algorithm success condition can be violated in the presence of exponential decay noise, the effect of which is shown in Fig. 5(b). The noise reduces contrast between neighboring peaks, which can invalidate our success condition if not enough measurement samples are collected to overcome the uncertainty due to shot noise. One way to counteract this deleterious effect is to perform extra measurements. Thus the performance of our algorithm can be quantified by establishing an upper bound on the number of measurements needed to guarantee a certain success probability of the algorithm, for various noise strengths and the desired accuracies. We point interested readers to Appendix B for a proof of the algorithm performance bound.

We now present the main result of our algorithm performance analysis as follows. For target accuracy ϵ and exponential decay parameter λ, a success probability greater than $1 - \delta$ can be ensured by using M samples satisfying

$$M \geqslant 8W(K(\epsilon, \lambda), 2\pi/\epsilon, \lambda) \log(16\pi/\delta\epsilon), \qquad (19)$$

where $W(K, J, \lambda)$ is a complicated function whose explicit form is described in Eq. (B49) of Appendix B and K is chosen as a function of ϵ and λ as described below. The algorithm performance measured by the runtime upper bound in units of c-U operations are plotted against $1/\epsilon$ for various values of λ in Fig. 6. In the low-noise regime, the runtime scales as $1/\epsilon$, which resembles that of the original QPE algorithm [24]. In the high-noise regime, the runtime scales as $1/\epsilon^2$. For a moderate amount of noise, the runtime performance of the algorithm interpolates between the low- and high-noise performance.

Due to the limited coherence time from the exponential decay, we choose the maximal circuit depth K based on ϵ and λ, given by $K(\epsilon, \lambda) = \max\{c\lfloor \frac{1}{c(a\lambda + \epsilon b/2\pi)} \rfloor, 2\}$, where $a, b, c \in \mathbb{R}^+$. We note that here we set K to be tunable based on λ and ϵ such that the runtime of the algorithm interpolates between the Heisenberg-limit scaling [i.e., $O(1/\epsilon)$] in the regime of $J\lambda \gg 1$ and the shot-noise-limit scaling [i.e., $O(1/\epsilon^2)$] in the regime of $J\lambda \ll 1$. For the specific functional form of K, we employ the floor function and set $c = 10$ to bypass values of K between 3 and 9, which were numerically found to be sub-
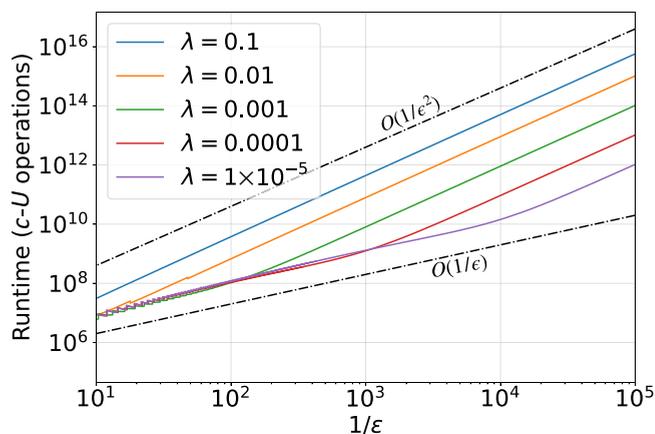
FIG. 6. This plot shows the runtime (measured in total number of calls to the c-U operation) as a function of the target accuracy inverse $1/\epsilon$ in the presence of various exponential decay errors of strength $\lambda = 0.1, 0.01, 0.001, 0.0001,$ and 0.00001. As λ decreases, the runtime transitions from a $1/\epsilon^2$ scaling to a $1/\epsilon$ scaling. The upper and lower dotted lines show the $O(1/\epsilon^2)$ and $O(1/\epsilon)$ scaling, respectively.

optimal. The parameters $a = 2$ and $b = 1.5$ were then chosen within this form to roughly minimize the runtime upper bound over a range of values of J and λ. We note that since this functional form of K is empirically derived to minimize the runtime upper bound, in practice, a more rigorous treatment for developing an optimal strategy for choosing K is necessary in the future.

Our analysis of the algorithm in Appendix B differs from that of Ref. [10] in two regards: (1) we separate the roles of K and J enabling high-accuracy estimates in the high-noise setting and (2) we take into consideration the correlation between the values of neighboring $\hat{f}_j$, enabling a reduction in sample complexity in the high-noise (low-K) setting. These two features enable us to establish a unifying expression that captures the performance of the algorithm in a wide variety of scenarios, ranging from using (effectively) a Bernoulli estimation approach (i.e., $K = 2$) to a Heisenberg-limited phase estimation approach [i.e., $K = O(1/\epsilon)$]. A signature of the switch in the algorithm's strategy from the Bernoulli to the Heisenberg approach to accommodate different scenarios of error is marked by the cusp in the runtime upper bound of RFE in Fig. 7.

Finally, we address the issue of determining the exponential decay parameter λ. Our algorithm relies on some knowledge of λ when setting the appropriate $K(\epsilon, \lambda)$. This raises the question of whether it is necessary to accurately determine λ and, if not, to what extent a discrepancy between the presumed and actual values of λ would compromise the algorithm's performance. The precise level of accuracy required to maintain the derived runtime remains an open question that we defer to future investigation. Nonetheless, intuitively, we anticipate that overestimating λ will result in using more samples than necessary, while underestimating λ will lead to using fewer samples than required, thus slightly increasing the probability of failure. We further highlight that several established benchmarking techniques, such as randomized benchmarking [34] and cross-entropy benchmarking [38], can be utilized to
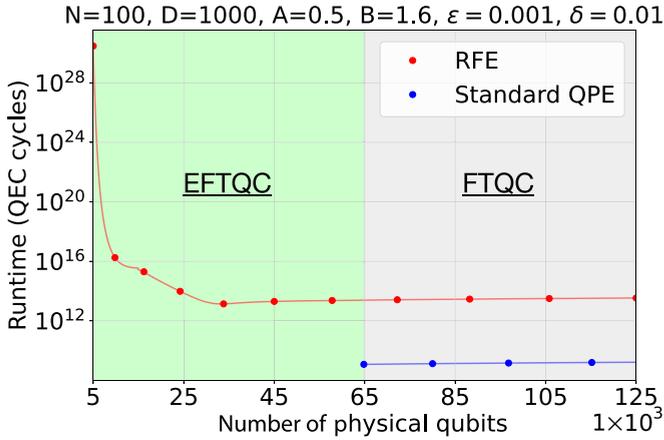
FIG. 7. Comparison of runtime upper bounds (in units of error correction cycles) between RFE (red) and the traditional QPE algorithm (blue) as a function of physical qubits available. These upper bounds are proved in Appendices B and C, respectively. The upper bound on the runtime of RFE may be higher than the standard approach to QPE, RFE can be run using an order-of-magnitude fewer physical qubits. The cusp occurring in the runtime upper bound of RFE at $\sim 10^4$ physical qubits marks the switch of the algorithm's strategy from using $K = 2$ to $K > 2$. In the regime of low physical qubits, the runtime becomes extremely high due to the sampling overhead that compensates for the high degree of error. The minimal runtime for RFE occurs around 34 000 physical qubits, where the increase in fault-tolerant runtime overhead is balanced with the decrease in number of $c\text{-}U$ queries through the transition from $O(1/\epsilon^2)$ to $O(1/\epsilon)$ scaling (see Fig. 6). Note that the runtime of standard QPE may be larger than presented; we have optimistically assumed that a single sample of the QPE circuit suffices, though in practice one might have to take several samples to ensure the failure rate is below some tolerance.

estimate essential noise parameters like the depolarizing error rate. These estimates can then be applied to determine the value of λ using Eq. (16) in our methodology.

B. FT overhead estimation comparison

As an application of our established performance bound in Eq. (19), in this section we provide a cost analysis of implementing the RFE algorithm accounting for the error correction overhead. Specifically, we aim to compare the overall runtime performance (measured in units of QEC cycles) of the traditional QPE versus the RFE algorithm as a function of the number of physical qubits required, which relates to distance of the surface code d as described in Eq. (5).

In Fig. 7, we show this comparison for the case of a logical $c\text{-}U$ acting on 100 qubits with a unitary circuit depth 1000, aiming for an accuracy of 0.1% and failure probability below 1%. These order-of-magnitude parameters are chosen to roughly correspond to the logical qubit and logical gate counts involved in quantum computations that might outperform corresponding classical computations such as in Ref. [39]. One notable characteristic of the traditional QPE algorithm is that there is a minimal number of physical qubits or code distance below which the desired algorithm success probability cannot be achieved (refer to the detailed derivation in Appendix C).

Therefore, in order to implement the traditional QPE algorithm reliably, a certain distance code or number of physical qubits have to be attained. This is, however, not required by the RFE algorithm. In principle, a higher runtime cost can always be paid in exchange for fewer physical qubits or a lower code distance in implementing the RFE algorithm. This is particularly valuable in the EFTQC regime, where devices have limited number of physical qubits, as depicted schematically in the green region of Fig. 7.

The RFE algorithm offers the advantage of requiring an order-of-magnitude fewer qubits compared to the traditional QPE algorithm. However, Fig. 7 also shows that the runtime upper bound of RFE in the FTQC regime is approximately four orders of magnitude larger than that of traditional QPE. Nevertheless, we expect that the aforementioned upper bound is conservative for two main reasons. First, the analysis incorporates several analytical bounds that result in a more cautious choice of M than what is likely necessary (a well-known phenomenon in algorithm analysis). We expect that the empirical runtime would be significantly smaller than the derived upper bounds and leave an investigation of this to future work. Second, the algorithm itself can be optimized to enhance performance. Examples of potential improvements include (1) tailoring the distribution from which k is sampled based on the noise characteristics of the device and (2) employing more sophisticated fitting strategies for extracting $\hat{\theta}$ from the Fourier transformed data. In addition, we note that the runtime of the RFE algorithm in reality can be further reduced via parallelization across multiple noisy-intermediate scale devices, as opposed to FTQC algorithms which require the execution of a long circuit on a fault-tolerant machine.

VI. CCONCLUSIONS AND OUTLOOK

In summary, we have developed a methodology for systematically analyzing the performance of a class of quantum algorithms suited for early fault-tolerant quantum computers. This is motivated by the need to understand the quantum resources necessary for these algorithms to achieve quantum advantage. Our approach can be extended to encompass various error models, fault-tolerant resource overhead models, and a wide range of quantum algorithms. By offering a generalized framework, our methodology paves the way for a comprehensive understanding of resource requirements and performance trade-offs in the realm of early fault-tolerant quantum computing.

As an application of our methodology, we analyzed the performance of the recently proposed RFE algorithm [10]. Studying the circuit-level error under our proposed Haar trajectory model, we found that the noise can be best described by an exponential decay at the algorithmic level. We developed a variant of this algorithm that interpolates between low depth and high depth (based on the strength of the exponential decay) and analytically derived its runtime upper bound as a function of target accuracy and failure rate. Studying the algorithm under a continuum of noise strengths, we found that the runtime upper bound interpolates between $O(1/\epsilon^2)$ (shot-noise limited scaling) and $O(1/\epsilon)$ (Heisenberg scaling) in the high- to low-noise limits.

Based on the runtime upper bound, we then carried out a fault-tolerant overhead analysis, comparing RFE with the traditional QPE in the problem instance of 100 logical qubits. Our analysis showed that the RFE algorithm can be implemented with an order-of-magnitude fewer physical qubits than the traditional QPE algorithm, albeit with a substantial increase in runtime. This trade-off allows for an earlier onset of practical quantum advantage in the EFTQC regime, where error correction remains expensive.

There are several crucial research directions that hold promise for delivering quantum computation with practical advantage:

First, it is imperative to develop more realistic circuit-level error models. Our methodology relies on the assumption that the sampled Haar trajectory states are derived from a unitary 2-design, or at the very least, their statistics can be well-approximated by those of a 2-design. To better understand the limitations of this model, further numerical investigations are required. In future studies, it would be valuable to develop an empirically derived model that captures the expected bias and variance of various circuit-level errors. We anticipate that the structure of the specific quantum circuits and error models of interest will introduce nonuniformity in the distribution of noisy trajectories. This nonuniformity could in principle shift $\mathbb{E}[\langle P \rangle]$ and increase $\mathrm{Var}[\langle P \rangle]$, leading to a worse algorithm performance than that predicted by the current analysis. By addressing these aspects, future studies can provide more realistic noise settings and evaluate the performance of RFE under those settings.

Second, there is a need to develop fault-tolerant overhead models specifically tailored for early fault-tolerant quantum computers [9]. Recent work [40] has shown that small deviations of fault-tolerant architectures from the ideal assumptions can substantially impact the performance of quantum algorithms. We anticipate that by including such realistic models of fault-tolerant architectures into the methodology of this work, the importance of robust quantum algorithms will be increasingly apparent.

Third, as briefly mentioned in Sec. IV B, we have exclusively considered circuit-level errors while neglecting the effect of state preparation errors in our noise modeling. Extending our methodology to include state preparation errors is a plausible avenue for further exploration and would likely require an algorithm-specific analysis. We note that parallel to our work, some related recent studies (e.g., Ref. [20]) provide complementary insights by studying the impact of state preparation errors on the performance of a similar-spirited EFTQC algorithm to RFE. A potential future direction is to study the expected performance of these relatively shallow-circuit algorithms under a mixture of the realistic error models, which early fault-tolerant devices are known to suffer from.

Furthermore, it is crucial to extend our methodology to analyze alternative EFTQC algorithms rather than RFE towards solving practical problems. We have chosen to analyze the RFE algorithm as a proof-of-principle case study to demonstrate the viability of our methodology, for the RFE is analytically tractable, which comes at the price of a potential suboptimal performance. We note that, however, our methodology is readily applicable to more advanced versions of EFTQC algorithms, such as the QCELS algorithm, a variant of the phase estimation algorithm proposed in Ref. [19]. For example, the noise modeling and FT overhead estimation detailed in our paper can be directly transferable to the QCELS robustness analysis similar to that outlined in Ref. [21]. In addition, our noise modeling scheme allows for the substitution of depolarizing errors with alternative structured noise models, such as dephasing errors, with simple modifications to the statistics of the error channels modeled in Sec. IV. We expect that these alternative algorithms will also be found to reduce quantum resources in the early fault-tolerant regime.

We envision that by extending our methodology to encompass existing algorithms and future developments in the EFTQC regime, we can enable their comprehensive evaluation and enhance the practical utilization of early fault-tolerant quantum computers.

APPENDIX A: STATISTICS OF QUANTUM EXPECTATION VALUE

1. Expectation value $\mathbb{E}[\langle P \rangle]$

For a generic N-qubit Pauli operator P, the expected value of the quantum expectation value $\langle P \rangle$ over all possible Haar trajectory states sampled from a spherical 2-design is given by

$$\mathbb{E}[\langle P \rangle] = p_{\mathbf{0}} \mathbb{E}[\mathrm{Tr}[P|\psi_{\mathbf{0}}\rangle\langle\psi_{\mathbf{0}}|]] + \sum_{\mathbf{j}=\mathbf{1}}^{4^{ND}-1} p_{\mathbf{j}} \mathbb{E}[\mathrm{Tr}[P|\psi_{\mathbf{j}}\rangle\langle\psi_{\mathbf{j}}|]],$$

$$(A1)$$

$$= p_{\mathbf{0}} \langle\psi_{\mathbf{0}}|P|\psi_{\mathbf{0}}\rangle$$
$$+ \sum_{\mathbf{j}=\mathbf{1}}^{4^{ND}-1} p_{\mathbf{j}} \left(\frac{1}{|S^{2^N}|} \int_{\psi_{\mathbf{j}} \in S^{2^N}} \mathrm{Tr}[P|\psi_{\mathbf{j}}\rangle\langle\psi_{\mathbf{j}}|] d\psi_{\mathbf{j}} \right),$$

$$(A2)$$

$$= p_{\mathbf{0}} \langle\psi_{\mathbf{0}}|P|\psi_{\mathbf{0}}\rangle$$
$$+ \sum_{\mathbf{j}=\mathbf{1}}^{4^{ND}-1} p_{\mathbf{j}} \left(\mathrm{Tr}\left[P \frac{1}{|S^{2^N}|} \int_{\psi_{\mathbf{j}} \in S^{2^N}} |\psi_{\mathbf{j}}\rangle\langle\psi_{\mathbf{j}}| d\psi_{\mathbf{j}} \right] \right).$$

$$(A3)$$

Here we have distributed the integral within the Tr operator due to its linearity and arrived at the expression of an average N-qubit Haar state

$$\frac{1}{|S^{2^N}|} \int_{\psi_{\mathbf{j}} \in S^{2^N}} |\psi_{\mathbf{j}}\rangle\langle\psi_{\mathbf{j}}| d\psi_{\mathbf{j}} = I/2^N. \tag{A4}$$

It follows that the expectation value is

$$\mathbb{E}[\langle P \rangle] = p_{\mathbf{0}} \langle\psi_{\mathbf{0}}|P|\psi_{\mathbf{0}}\rangle + \sum_{\mathbf{j}=\mathbf{1}}^{4^{ND}-1} p_{\mathbf{j}} \mathrm{Tr}[PI/2^N], \tag{A5}$$

$$= p_{\mathbf{0}} \langle \psi_{\mathbf{0}} | P | \psi_{\mathbf{0}} \rangle + \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}} / 2^N \mathrm{Tr}[P], \quad \text{(A6)}$$

$$= p_{\mathbf{0}} \langle \psi_{\mathbf{0}} | P | \psi_{\mathbf{0}} \rangle. \quad \text{(A7)}$$

2. Variance Var[⟨P⟩]

Similarly, we can compute the variance of $\langle P \rangle$

$$\mathrm{Var}[\langle P \rangle] = \mathbb{E}[\langle P \rangle^2] - \mathbb{E}[\langle P \rangle]^2 = \mathbb{E}[\langle P \rangle^2] - p_{\mathbf{0}}^2 \langle \psi_{\mathbf{0}} | P | \psi_{\mathbf{0}} \rangle^2, \quad \text{(A8)}$$

where

$$\mathbb{E}[\langle P \rangle^2] = \sum_{\mathbf{i},\mathbf{j}=0}^{4^{ND}-1} p_{\mathbf{i}} p_{\mathbf{j}} \mathbb{E}[\mathrm{Tr}[P | \psi_{\mathbf{i}} \rangle \langle \psi_{\mathbf{i}} |] \mathrm{Tr}[P | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} |]] \quad \text{(A9)}$$

$$= \sum_{\mathbf{i},\mathbf{j}=0}^{4^{ND}-1} p_{\mathbf{i}} p_{\mathbf{j}} \left(\frac{1}{|S^{2^N}|^2} \int_{\psi_{\mathbf{i}} \in S^{2^N}} \int_{\psi_{\mathbf{j}} \in S^{2^N}} \mathrm{Tr}[P | \psi_{\mathbf{i}} \rangle \langle \psi_{\mathbf{i}} |] \right.$$

$$\left. \times \mathrm{Tr}[P | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} |] d\psi_{\mathbf{i}} d\psi_{\mathbf{j}} \right). \quad \text{(A10)}$$

We note that here the summation of cross-terms where $\mathbf{i} \neq \mathbf{j}$ within Eq. (A10) evaluates to 0. Summing the rest of the terms where $\mathbf{i} = \mathbf{j}$, excluding the noiseless term $p_{\mathbf{0}}^2 \langle \psi_{\mathbf{0}} | P | \psi_{\mathbf{0}} \rangle^2$, we get

$$\mathrm{Var}[\langle P \rangle] = \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2 \mathbb{E}[\mathrm{Tr}[P | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} |]^2], \quad \text{(A11)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2 \left\{ \frac{1}{|S^{2^N}|} \int_{\psi_{\mathbf{j}} \in S^{2^N}} \mathrm{Tr}[(P \otimes P) | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} | \right.$$

$$\left. \otimes | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} |] d\psi_{\mathbf{j}} \right\}, \quad \text{(A12)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2 \left\{ \mathrm{Tr} \left[(P \otimes P) \frac{1}{|S^{2^N}|} \int_{\psi_{\mathbf{j}} \in S^{2^N}} | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} | \right. \right.$$

$$\left. \left. \otimes | \psi_{\mathbf{j}} \rangle \langle \psi_{\mathbf{j}} | d\psi_{\mathbf{j}} \right] \right\}, \quad \text{(A13)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2 \mathrm{Tr} \left[P \otimes P \frac{I \otimes I + \mathrm{SWAP}}{2 \mathrm{Tr}[(I \otimes I + \mathrm{SWAP})/2]} \right], \quad \text{(A14)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2 \mathrm{Tr} \left[\frac{P \otimes P + (P \otimes P) \mathrm{SWAP}}{\mathrm{Tr}[I \otimes I + \mathrm{SWAP}]} \right], \quad \text{(A15)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2 \left(\mathrm{Tr} \left[\frac{P \otimes P}{\mathrm{Tr}[I \otimes I + \mathrm{SWAP}]} \right] \right.$$

$$\left. + \mathrm{Tr} \left[\frac{(P \otimes P) \mathrm{SWAP}}{\mathrm{Tr}[I \otimes I + \mathrm{SWAP}]} \right] \right), \quad \text{(A16)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} \frac{p_{\mathbf{j}}^2}{\mathrm{Tr}[I \otimes I + \mathrm{SWAP}]} \mathrm{Tr}[(P \otimes P) \mathrm{SWAP}], \quad \text{(A17)}$$

$$= \sum_{\mathbf{j}=1}^{4^{ND}-1} \frac{p_{\mathbf{j}}^2}{\mathrm{Tr}[I \otimes I + \mathrm{SWAP}]} \mathrm{Tr}[P^2], \quad \text{(A18)}$$

$$= \frac{2^N}{\mathrm{Tr}[I \otimes I + \mathrm{SWAP}]} \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2, \quad \text{(A19)}$$

$$= \frac{2^N}{\mathrm{Tr}[I \otimes I] + \mathrm{Tr}[\mathrm{SWAP}]} \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2, \quad \text{(A20)}$$

$$= \frac{2^N}{2^{2N} + 2^N} \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2, \quad \text{(A21)}$$

$$= \frac{1}{2^N + 1} \sum_{\mathbf{j}=1}^{4^{ND}-1} p_{\mathbf{j}}^2. \quad \text{(A22)}$$

For the single-qubit depolarizing error channel proposed in Eq. (13) with error probability r, the variance expression becomes

$$\mathrm{Var}[\langle P \rangle] = \frac{1}{2^N + 1} \left[\left(\sum_{\mathbf{j}=0}^{4^{ND}-1} p_{\mathbf{j}}^2 \right) - p_{\mathbf{0}}^2 \right], \quad \text{(A23)}$$

$$= \frac{1}{2^N + 1} \{ [(1 - r)^2 + 3(r/3)^2]^{NDk} - (1 - r)^{2NDk} \}. \quad \text{(A24)}$$

Figure 8 illustrates the standard deviation of $\langle P \rangle$, $\sigma_{\langle P \rangle} := \sqrt{\mathrm{Var}[\langle P \rangle]}$ as a function of depths $D \times k$ and single-qubit depolarizing rate r, considering number of logical qubits $N = 1, 10, 100$. We observe that $\sigma_{\langle P \rangle}$ remains relatively small overall for $N = 100$, while it is comparatively larger for $N = 1$ and 10. We note that $\sigma_{\langle P \rangle}$ is particularly large for low-depth and high-error settings. In the extreme case of $N = 1$, $\sigma_{\langle P \rangle}$ reaches a concerning high magnitude $\sim 10^{-2}$, which is too large to be overlooked in our algorithmic error model.

Nonetheless, we opted to exclude the random fluctuation term stemming from variance in our algorithmic error model throughout our analysis. This decision aligns with the objective of our study, which focuses on examining the performance of devices in the early fault-tolerant regime. Specifically, we are interested in the parameter regime characterized by $r = p_{\mathrm{logical}} \lesssim 10^{-2}$, $N \gtrsim 100$, and $D \times k \gtrsim 1000$. In this regime, the standard deviation $\sigma_{\langle P \rangle}$ remains sufficiently small, justifying our assumption to omit it from the algorithmic error model. We point out that the performance of the RFE algorithm under the NISQ setting, where $p_{\mathrm{logical}} > 10^{-2}$, $N < 100$, and $D \times k < 1000$, falls outside the scope of our study and merits further investigation.

APPENDIX B: ALGORITHM PERFORMANCE

In this section, we detail the analytical derivation of the performance upper bound of the RFE algorithm in the presence of exponential decay noise with strength λ. Rather than presenting these results in a theorem-proof format, we choose a more narrative presentation to facilitate broader accessibility. Based on the success condition $|\hat{f}_j|^2 < |\hat{f}_{j*}|^2$ for all j with $|j - \Theta| > 1$, where $\Theta = J\theta/2\pi$, as discussed in Sec. V A, the
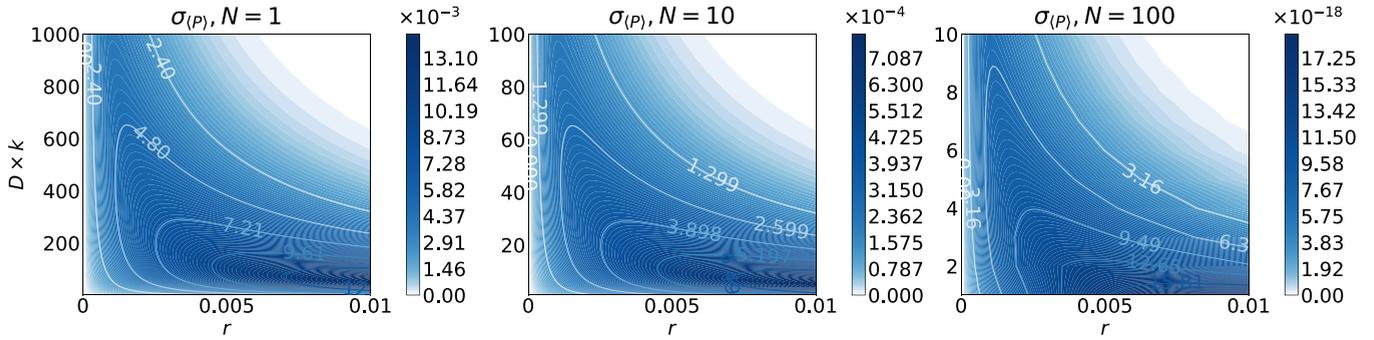
FIG. 8. Standard deviation of $\langle P \rangle$ as a function of $D \times k$ and r is plotted on a log scale for $N = 1, 10, 100$.

algorithm failure probability can be upper bounded using the union bound

$$\Pr(\text{fail}) \leqslant \sum_{|j-\Theta| \geqslant 1} \Pr(\text{fail}_j), \tag{B1}$$

$$\leqslant (J-2) \max_j \Pr(\text{fail}_j), \tag{B2}$$

$$\leqslant J \max_j \Pr(\text{fail}_j). \tag{B3}$$

We will upper bound this worst-case failure probability using Chebyshev's inequality.

In previous analyses of the RFE algorithm [10], we had analyzed the likelihood that every Fourier estimate $\hat{f}_j$ was within some distance of its mean f_j. This approach does not capture the performance of the algorithm in the regime of small K, as the correlation between Fourier estimates, $\hat{f}_j$ and $\hat{f}_{j'}$, is largely overlooked. As a result, the success probability bound becomes too loose to capture the actual performance scaling of the algorithm.

To account for the correlation among the Fourier estimates, we will consider a change of variables,

$$\hat{c}_{j,j*} \equiv \hat{f}_j + \hat{f}_{j*} \tag{B4}$$

$$\hat{d}_{j,j*} \equiv \hat{f}_j - \hat{f}_{j*}. \tag{B5}$$

The motivation for this choice is the fact that $\hat{d}_{j,j*}$ captures correlation between $\hat{f}_j$ and $\hat{f}_{j*}$ and therefore will have small variance for nearby $j, j*$ when K is small compared to J. In contrast, ignoring the correlation between $\hat{f}_j$ and $\hat{f}_{j*}$ and treating their variances separately, as was done in Ref. [10], leads to an overestimation of the algorithm failure probability. We will define the expectation values of these quantities to be

$$c_{j,j'} \equiv f_j + f_{j'} \tag{B6}$$

$$d_{j,j'} \equiv f_j - f_{j'}. \tag{B7}$$

The sufficient condition for success can be expressed as

$$\text{Re}(\hat{c}_{j,j*}\hat{d}_{j,j*}^*) < 0 \text{ for all } j \text{ with } |j - J\theta/2\pi| > 1 \tag{B8}$$

using the following relationship:

$$|\hat{f}_j|^2 - |\hat{f}_{j*}|^2 = \text{Re}[(\hat{f}_j + \hat{f}_{j*})(\hat{f}_j - \hat{f}_{j*})^*]$$

$$= \text{Re}(\hat{c}_{j,j*}\hat{d}_{j,j*}^*). \tag{B9}$$

This condition also holds for $c_{j,j*}$ and $d_{j,j*}$, which are the expected values of $\hat{c}_{j,j*}$ and $\hat{d}_{j,j*}$, respectively. It is the sta-

tistical fluctuations from the finite sampling that can cause the condition to fail. As the number of samples is increased, the estimates $\hat{c}_{j,j*}$ and $\hat{d}_{j,j*}$ will be expected to increasingly concentrate about their means. We can quantify this concentration with the following version of the central limit theorem [41],

$$\Pr\left(\left|\frac{1}{M}\sum_{i=1}^M \hat{q}_i - \mathbb{E}\hat{q}\right| > t\right) \leqslant 4\exp[Mt^2/8\text{Var}(\hat{q})], \tag{B10}$$

where $\hat{q}$ is any continuous random complex variable. We use the Chebyshev inequality to upper bound the likelihood that the estimator deviates from its *ideal* expected value:

$$\Pr(|\hat{c} - \mathbb{E}\hat{c}| > \chi) \leqslant 4\exp[-M\chi^2/8\text{Var}(\hat{c})]$$

$$\Pr(|\hat{d} - \mathbb{E}\hat{d}| > \eta) \leqslant 4\exp[-M\eta^2/8\text{Var}(\hat{d})], \tag{B11}$$

where $\text{Var}(\hat{c})$ and $\text{Var}(\hat{d})$ are the single sample variances of $\hat{c}$ and $\hat{d}$. Thus, we have two free parameters (χ and η) to choose in a manner such that (1) $|\hat{c} - \mathbb{E}\hat{c}| \leqslant \chi$ and $|\hat{d} - \mathbb{E}\hat{d}| \leqslant \eta$ imply success (letting us upper bound the failure probability) and (2) the number of samples does not become too high. While it does not necessarily minimize the number of samples, we will choose the value of χ such that the upper bounds in Eq. (B11) become equal: $\chi = \eta\sqrt{\text{Var}(\hat{c})/\text{Var}(\hat{d})}$.

Next we aim to establish a sufficiently large value for η, while still ensuring that $|\hat{c} - \mathbb{E}\hat{c}| \leqslant \eta\sqrt{\text{Var}(\hat{c})/\text{Var}(\hat{d})}$ and $|\hat{d} - \mathbb{E}\hat{d}| \leqslant \eta$ imply success. This will be achieved by recasting the success condition. Note that the condition $\text{Re}(\hat{c}_{j,j*}\hat{d}_{j,j*}^*) < 0$ is independent of the magnitudes of $\hat{c}_{j,j*}$ and $\hat{d}_{j,j*}^*$; it only depends on the relative phase angle between these two complex numbers. Thus, an equivalent condition for algorithm success is that, for all j with $|j - J\theta/2\pi| > 1$, the phase angle formed between complex values $\hat{c}_{j,j*}$ and $\hat{d}_{j,j*}$ is not within $[-\pi/2, \pi/2]$. For the time being, we will ease the notation by dropping the $j, j*$ subscripts.

We use this phase-angle condition to establish a sufficient condition for success in terms of the allowable sizes of deviations from the mean. The largest possible angle formed between $\hat{c}$ and c is given by $\sin(\gamma) = |\hat{c} - c|/|c|$ and the largest possible angle formed between $\hat{d}$ and d is given by $\sin(\tau) = |\hat{d} - d|/|d|$. We also define the angle between c and d according to $|c||d|\cos(\alpha) = \text{Re}(cd^*)$. With these definitions, the smallest possible phase angle formed between $\hat{c}$ and

$\hat{d}$ is $\alpha - \gamma - \tau$, hence, the condition for success becomes

$$\pi/2 < \alpha - \gamma - \tau. \tag{B12}$$

We use the bounds of $x \leqslant \arcsin x \leqslant \pi x/2$ for $0 \leqslant x \leqslant 1$ to establish a sufficient condition for success that is amenable to using the Chebyshev inequality. First, we have that $\gamma + \tau \leqslant \frac{\pi}{2}(|\hat{c} - c|/|c| + |\hat{d} - d|/|d|)$. Next we have that $\cos(\alpha) = -\sin(\alpha - \pi/2)$, and so $\arcsin[-\mathrm{Re}(cd^*)/|c||d|] = \alpha - \pi/2$. Using the lower bound on arcsin, we then have $-\mathrm{Re}(cd^*)/|c||d| \leqslant \alpha - \pi/2$. From the chain of inequalities

$$-\mathrm{Re}(cd^*)/|c||d| < \alpha - \pi/2 < \gamma$$
$$+ \tau < \frac{\pi}{2}(|\hat{c} - c|/|c| + |\hat{d} - d|/|d|), \tag{B13}$$

we can establish the following implication:

$$\alpha - \pi/2 < \gamma + \tau$$
$$\Rightarrow$$
$$-\mathrm{Re}(cd^*)/|c||d| < \frac{\pi}{2}(|\hat{c} - c|/|c| + |\hat{d} - d|/|d|). \tag{B14}$$

The contrapositive of this statement gives our sufficient condition for algorithm success

$$-\mathrm{Re}(cd^*) > \frac{\pi}{2}(|c||\hat{d} - d| + |d||\hat{c} - c|)$$
$$\Rightarrow$$
$$\pi/2 < \alpha - \gamma - \tau. \tag{B15}$$

We can then set the maximal allowable value for η according to Eq. (B15). Defining $\rho = \sqrt{\mathrm{Var}(\hat{c})/\mathrm{Var}(\hat{d})}$, this is

$$\eta = -\frac{2}{\pi}\frac{\mathrm{Re}(cd^*)}{|c| + |d|\rho}$$
$$= \frac{2}{\pi}\frac{|\hat{f}_{j*}|^2 - |\hat{f}_j|^2}{|c| + |d|\rho}. \tag{B16}$$

Putting this all together, the probability of failure is upper bounded by

$$\Pr(\mathrm{fail}) \leqslant J \max_j \Pr(\mathrm{fail}_j), \tag{B17}$$

$$\leqslant J \max_j \Pr(|\hat{c} - c| > \chi \text{ or } |\hat{d} - d| > \eta), \tag{B18}$$

$$\leqslant J \max_j [\Pr(|\hat{c} - c| > \chi) + \Pr(|\hat{d} - d| > \eta)], \tag{B19}$$

$$\leqslant 8J \exp[-M\eta_{j',j*}^2/8\mathrm{Var}(\hat{d}_{j',j*})], \tag{B20}$$

where we have reintroduced the indices j', $j*$ for clarity and are using j' to indicate the index that realizes the maximization. Let $W(K, J, \lambda)$ be a parameterized upper bound on the quantity $\mathrm{Var}(\hat{d}_{j',j*})/\eta_{j',j*}^2$ that is independent of θ, $j*$, and j'. To ensure that $P(\mathrm{fail}) \leqslant \delta$, we can choose M to be a value such that

$$8J \exp\left[-M\eta_{j',j*}^2/8\mathrm{Var}(\hat{d}_{j',j*})\right] \leqslant \delta. \tag{B21}$$

We can then solve for such a value of M as

$$-M\eta_{j',j*}^2/8\mathrm{Var}(\hat{d}_{j',j*}) \leqslant \log(\delta/8J), \tag{B22}$$

$$M \geqslant \frac{8\mathrm{Var}(\hat{d}_{j',j*})}{\eta_{j',j*}^2}\log(8J/\delta). \tag{B23}$$

To achieve a high probability of success, it is also sufficient to set M to be greater than a value that is larger than the right-hand side above. Let $W(K, J, \lambda)$ be a function satisfying

$$8W(K, J, \lambda)\log(8J/\delta) \geqslant \frac{8\mathrm{Var}(\hat{d}_{j',j*})}{\eta_{j',j*}^2}\log(8J/\delta). \tag{B24}$$

Then, setting

$$M \geqslant 8W(K, J, \lambda)\log(8J/\delta) \tag{B25}$$

is sufficient to ensure success with high probability. As described in the main text, K and J are set as functions of λ and ϵ, given by

$$J \leftarrow 2\pi/\epsilon, \tag{B26}$$

$$K \leftarrow \max\{\lfloor[(1/\lambda)^{-1} + (2\pi/\epsilon)^{-1}]^{-1}\rfloor, 2\}, \tag{B27}$$

where we see that K is a harmonic average of J and $1/\lambda$ and $K \leqslant J$. This will make the number of measurements M an implicit function of ϵ and λ.

We now establish a function W that satisfies $W \geqslant \frac{\mathrm{Var}(\hat{d})}{\eta^2}$. From Eq. (B16), we can establish the upper bound,

$$\frac{\sqrt{\mathrm{Var}(\hat{d})}}{\eta} = \frac{\pi}{2}\frac{\sigma_{\hat{d}}|c| + \sigma_{\hat{c}}|d|}{|f_{j*}|^2 - |f_j|^2}, \tag{B28}$$

$$\leqslant 2\pi\frac{\sigma_{\hat{d}} + |d|}{|f_{j*}|^2 - |f_j|^2}, \tag{B29}$$

where in the second line we have used the fact that both $\sigma_{\hat{c}} = \sqrt{\mathrm{Var}(\hat{c})}$ and $|c|$ are upper bounded by 4. This follows from the fact that for a single sample estimate, $\hat{c}$ has magnitude at most 4 as it is the sum of two quantities that have magnitude at most 2. Since $\mathrm{Var}(\hat{d}) = \mathbb{E}(|\hat{d}|^2) - d^2$ is positive, we have that $\mathbb{E}(|\hat{d}|^2) \geqslant \mathrm{Var}(\hat{d})$ and $\mathbb{E}(|\hat{d}|^2) \geqslant |d|^2$. This lets us further upper bound $\sqrt{\frac{\mathrm{Var}(\hat{d})}{\eta^2}}$ as

$$\frac{\sqrt{\mathrm{Var}(\hat{d})}}{\eta} \leqslant 4\pi\frac{\sqrt{\mathbb{E}(|\hat{d}|^2)}}{|f_{j*}|^2 - |f_j|^2}. \tag{B30}$$

Using $\omega = e^{-i2\pi/J}$, $\mathbb{E}(|\hat{d}|^2)$ is computed to be

$$\mathbb{E}(|\hat{d}|^2) = 4\mathbb{E}[|e^{-i\phi}z(\omega^{jk} - \omega^{j*k})|^2], \tag{B31}$$

$$= 8\{1 - \mathbb{E}\mathrm{Re}[\omega^{k(j*-j)}]\}, \tag{B32}$$

$$= 8 - \frac{4}{K}\left[\frac{1 - \omega^{K(j*-j)}}{1 - \omega^{j*-j}} + \frac{1 - \omega^{-K(j*-j)}}{1 - \omega^{-j*+j}}\right], \tag{B33}$$

$$= 8 - 8\cos\left[\pi(K-1)\frac{j*-j}{J}\right]\frac{\sin(\pi K\frac{j*-j}{J})}{K\sin(\pi\frac{j*-j}{J})}. \tag{B34}$$

Defining $x = \pi(j_* - j)/J$, we establish a function upper bounding the quantity above

$$\mathbb{E}(|\hat{d}|^2) = 8\left\{1 - [\cos(Kx)\cos(x) + \sin(Kx)\sin(x)]\frac{\sin(Kx)}{K\sin(x)}\right\},$$
(B35)

$$= 8\left[1 - \frac{\sin(2Kx)}{2K\tan(x)} - \sin^2(Kx)/K\right],$$
(B36)

$$\leqslant 8\left[1 - \frac{\sin(2Kx)}{2K\tan(x)}\right],$$
(B37)

$$\leqslant 8\left[1 - \frac{\sin(2Kx)}{2K\tan(x)}\right],$$
(B38)

$$\leqslant \frac{32}{3}\{1 - \exp[-4(Kx)^2/7]\},$$
(B39)

where the last inequality is established by using numerical software and exploiting the fact that $K \geqslant 2$. Observe that this function is monotonically increasing in x. We rewrite $x = (\theta - 2\pi j/J)/2 - (\theta - 2\pi j_*/J)/2$. Using the fact that $|\theta - 2\pi j_*/J| \leqslant \pi/J$ and that the above function is monotonic in x, we obtain an upper bound by setting $|x| = \pi/J + |\theta - 2\pi j/J|/2$,

$$\mathbb{E}(|\hat{d}|^2) \leqslant \frac{32}{3}\left\{1 - \exp\left[-\frac{4K^2}{7}(\pi/J + |\theta - 2\pi j/J|/2)^2\right]\right\}$$
$$\equiv Q(K, J; j).$$
(B40)

Next we lower bound the denominator of Eq. (B30). From Eq. (18), we calculate

$$|f_j|^2 = \frac{e^{-(K-1)\lambda}}{K^2}\left\{\frac{\cosh(K\lambda) - \cos[K(\theta - 2\pi j/J)]}{\cosh(\lambda) - \cos(\theta - 2\pi j/J)}\right\}.$$
(B41)

Defining $y_* = \theta - 2\pi j*/J$, we have that $|y_*| \leqslant \epsilon/2 = \pi/J$. Note that the above function, for $j = j*$, is symmetric about $y_* = 0$ and thus depends only on $|y_*|$. Using the fact that $K \leqslant J$, on the range $0 \leqslant |y_*| \leqslant \pi/J$, the above function is monotonically decreasing (as observed numerically). Thus, it achieves its minimum at $|y_*| = \pi/J$. From this we lower bound $|f_{j*}|^2$ as

$$|f_{j*}|^2 \geqslant \frac{e^{-(K-1)\lambda}}{K^2}\left[\frac{\cosh(K\lambda) - \cos(\pi K/J)}{\cosh(\lambda) - \cos(\pi/J)}\right] \equiv R(K, J, \lambda).$$
(B42)

Last, we upper bound the square magnitude of the Fourier coefficient of the nonadjacent frequencies, $|f_j|^2$, where $|J\theta/2\pi - j| \geqslant 1$. To simplify the analysis, we define

$$m(x; K, \lambda) := \frac{e^{-(K-1)\lambda}}{K^2}\left[\frac{\cosh(K\lambda) - \cos(x)}{\cosh(\lambda) - \cos(x/K)}\right],$$
(B43)

such that $|f_j|^2 = m(K(\theta - 2\pi j/J); K, \lambda)$. We will establish that

$$g(x; K, \lambda) = \left(1 - r\left[1 - \tanh^2\left(\frac{\lambda}{2}\right)\right]\right.$$
$$\left. \times \left\{1 - \left[\cos^2\left(\frac{\pi x}{K}\right)\right]^{\frac{K^2}{4}}\right\}\right)m(0; K, \lambda)$$

is an upper bound of $m(x; K, \lambda)$ when $r = 0.89$. Consider two cases: (1) $K = 2$ and (2) $K > 2$. In the case of $K = 2$, through use of trigonometric identities, it can be shown that g is an exact expression for m when $r = 1$. By setting $r = 0.89 < 1$, g can only increase given that r multiplies a positive-valued function. Therefore, setting $r = 0.89$ leads to g being an upper bound for m in the case of $K = 2$. In the case of $K > 2$, we will establish that $1 - [\cos^2(\frac{\pi x}{K})]^{\frac{K^2}{4}}$ is a *lower bound* of

$$\frac{1 - g(x; K, \lambda)}{r[1 - \tanh^2(\lambda/2)]}.$$
(B44)

First, we use the fact that the above function is monotonically decreasing in $\lambda > 0$ and the fact that λ is upper bounded, to set λ to this value and establish a new λ-independent function of K and x that is a lower bound. Since, as set in the main text, $K = \max\{10\lfloor\frac{1}{10(2\lambda + 3/2J)}\rfloor, 2\} \geqslant \frac{1}{2\lambda + 3/2J} \geqslant \frac{1}{2\lambda}$, we have $\lambda \geqslant 1/2K$. Since Eq. (B44) is monotonically decreasing in λ, setting λ to its maximum value of $1/2K$ yields a lower bound for Eq. (B44) (on the valid domain of λ) that is only a function of K and x. It can then be numerically established that for all integer $K \geqslant 2$, we have

$$\frac{1 - g(x; K, \lambda)}{r[1 - \tanh^2(\lambda/2)]} \geqslant \frac{1 - g(x; K, 1/2K)}{r[1 - \tanh^2(1/4K)]}$$
$$\geqslant 1 - \cos^{K^2/2}(\pi x/K).$$
(B45)

This gives us the desired functional upper bound on $|f_j|^2$ as

$$|f_j|^2 \leqslant \frac{e^{-(K-1)\lambda}}{K^2}\left[\frac{\cosh(K\lambda) - 1}{\cosh(\lambda) - 1}\right]$$
$$\times \left(1 - 0.89\left[1 - \tanh^2\left(\frac{\lambda}{2}\right)\right]\right.$$
$$\left. \times \{1 - \cos^{K^2/2}[(\theta - 2\pi j/J)/2]\}\right) \equiv S(K, J, \lambda; j),$$
(B46)

and this function $S(K, J, \lambda; j)$ decreases monotonically away from $j = J\theta/2\pi$ on the interval $-J/2 \leqslant j - J\theta/2\pi \leqslant J/2$, which will be used later.

Putting the previous bounds together, we have that

$$\frac{\sqrt{\mathrm{Var}(\hat{d})}}{\eta} \leqslant 4\pi\frac{\sqrt{\mathbb{E}(|\hat{d}|^2)}}{|f_{j*}|^2 - |f_j|^2}$$
$$\leqslant 4\pi\frac{\sqrt{Q(K, J; j)}}{R(K, J, \lambda) - S(K, J, \lambda; j)},$$
(B47)

where Q, R, and S are defined in Eqs. (B40), (B42), and (B46), respectively. This overall function is monotonically decreasing in $|j - J\theta/2\pi|$, and thus setting this value to its smallest allowed value of $|j - J\theta/2\pi| = 1$ (the nonadjacent condition) leads to an upper bound that is independent of j_* and j,

$$\frac{\mathrm{Var}(\hat{d})}{\eta^2} \leqslant 16\pi^2\frac{Q(K, J; J\theta/2\pi + 1)}{[R(K, J, \lambda) - S(K, J, \lambda; J\theta/2\pi + 1)]^2}$$
$$\equiv W(K, J, \lambda).$$
(B48)

From this we are able to establish the final bound on the sufficient number of measurements to ensure an accurate estimate with high probability,

$$M \geqslant 128\pi^2 \log(8J/\delta)$$
$$\times \frac{Q(K, J; J\theta/2\pi + 1)}{[R(K, J, \lambda) - S(K, J, \lambda; J\theta/2\pi + 1)]^2}, \quad \text{(B49)}$$

where Q, R, and S are defined in Eqs. (B40), (B42), and (B46), respectively. We note that, although θ appears in the arguments of Q and S, this cancels with the θ in the definition, making the overall expression independent of θ.

APPENDIX C: COMPARISON TO COST OF STANDARD QUANTUM PHASE ESTIMATION

In this section, we analyze the cost of the standard QPE algorithm as described in Ref. [42]. This allows us to compare the fault-tolerant overhead associated with the traditional approach to the frugal approach that we have taken using the RFE algorithm.

Based on Ref. [42], the standard QPE algorithm can achieve an ϵ-accurate estimation with probability greater than $1 - \delta'$ by using $n = \lceil \log_2(\frac{1}{\epsilon}) \rceil + \lceil \log_2(\frac{1}{2\delta} + \frac{1}{2}) \rceil$ ancillary qubits and performing $2^{n+1} - 1$ c-U operations in the circuit. We note that here we have ignored the cost of the quantum Fourier transform operations, as in most cases it is negligible compared to the rest of the circuit. While this analysis assumes that the circuit is implemented perfectly, in practice, there will be some implementation failure probability.

The standard approach for analyzing the failure rate of quantum algorithms is to upper bound the failure rate as

$$\delta \leqslant \delta_{\text{alg}} + \delta_{\text{imp}}. \quad \text{(C1)}$$

Here the implementation failure probability δ_{imp} accounts for any uncorrected errors within the fault-tolerant protocols, contributing to the overall algorithm failure probability δ alongside the standard algorithm failure probability δ_{alg}. This assumption holds reasonably well in terms of asymptotic scaling, with the fault-tolerant overhead exhibiting a logarithmic dependency on the inverse failure rate. However, in the context of early fault-tolerant systems where resources are scarce, a more economical allocation of resources becomes advantageous.

Given a target total failure rate of δ, the allocation of the error budget between the algorithm failure and the implementation failure becomes a crucial decision. Though not optimal, a nearly optimal approach is to evenly distribute an error budget of $\delta/2$ to both failure modes. Consequently, this establishes a lower bound on the required number of ancilla qubits,

$$n \geqslant \left\lceil \log_2\left(\frac{1}{\epsilon}\right) \right\rceil + \left\lceil \log_2\left(\frac{1}{\delta} + \frac{1}{2}\right) \right\rceil. \quad \text{(C2)}$$

Similarly, the number of c-U operations in the circuit satisfies the lower bound

$$\#(c - U) \geqslant \frac{1}{\epsilon}\left(\frac{1}{\delta} + \frac{1}{2}\right) - 1. \quad \text{(C3)}$$

We will only slightly overcount resources by dropping the -1 above. In order to ensure that $\delta_{\text{imp}} \leqslant \delta/2$, each c-U component must fail with a probability no greater than

$$\delta_{c-U} \leqslant \frac{\delta}{2} \frac{1}{2^{n+1} - 1} \leqslant \frac{\delta}{2} 2^{-(n+1)} = \frac{\epsilon\delta^2}{1 + \delta}. \quad \text{(C4)}$$

We note that here the implementation error solely arises from QEC failure. As a reminder, the logical error rate for a surface code is assumed by the model $p_{\text{logical}} = Ae^{-Bd}$ (as discussed in Sec. IV A), where A and B are constants from empirical observation of the surface code scaling and d is the distance of the surface code. Assuming that each logical c-U operation includes $\sim ND$ physical operations, where N is the number of qubits and D the circuit depth, the failure probability of one logical c-U is then

$$\delta_{c-U} \leqslant ND p_{\text{logical}}. \quad \text{(C5)}$$

To ensure that the c-U logical failure rate in Eq. (C5) is bounded by the maximal acceptable failure probability as required by the algorithm from Eq. (C4), we can therefore bound the logical error rate p_{logical},

$$p_{\text{logical}} \leqslant \frac{\epsilon\delta^2}{ND(1 + \delta)}, \quad \text{(C6)}$$

whereby we can then bound the minimal distance require to reach the desired success probability δ for the traditional QPE algorithm

$$d \geqslant \frac{1}{B} \ln\left[\frac{NAD(1 + \delta)}{\epsilon\delta^2}\right]. \quad \text{(C7)}$$

In the case of a high physical error rate with $A = 0.5$ and $B = 1.6$, achieving a success probability of $\delta = 10^{-2}$ and precision of $\epsilon = 10^{-3}$ for a QPE problem instance of $N = 100$ and $D = 1000$ requires a minimal distance of 18 with the traditional algorithm. This corresponds to approximately 64 800 physical qubits, currently surpasses the capabilities of both NISQ and EFTQC devices.

To illustrate this, in Fig. 7, we compare the runtime, measured in the number of error correction cycles, between QPE and RFE as a function of the required physical qubits $2Nd^2$. It is important to note that while RFE has a higher runtime upper bound, it offers a trade-off between runtime and the required distance or number of physical qubits in the EFTQC regime (as indicated by the shaded green region). This trade-off becomes particularly valuable for upcoming devices that can accommodate a few thousand qubits and support error-correcting codes of moderate distances.

In practice, as mentioned in Sec. V B of the main text, the actual runtime of RFE is expected to be significantly lower than the calculated upper bound. Additionally, we note that the runtime cost can be effectively reduced through parallelization across multiple devices of smaller sizes, leveraging the statistical nature of the RFE algorithm.

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. GSL Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (Lond.) **574**, 505 (2019).

[2] Y. Kim, A. Eddins, S. Anand *et al.*, Evidence for the utility of quantum computing before fault tolerance, Nature (Lond.) **618**, 500 (2023).

[3] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[4] C. Gidney and M. Ekerå, How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits, Quantum **5**, 433 (2021).

[5] J. J. Goings, A. White, J. Lee, C. S. Tautermann, M. Degroote, C. Gidney, T. Shiozaki, R. Babbush, and N. C. Rubin, Reliably assessing the electronic structure of cytochrome p450 on today's classical computers and tomorrow's quantum computers, Proc. Natl. Acad. Sci. USA **119**, e2203533119 (2022).

[6] I. H. Kim, Y.-H. Liu, S. Pallister, W. Pol, S. Roberts, and E. Lee, Fault-tolerant resource estimate for quantum chemical simulations: Case study on li-ion battery electrolyte molecules, Phys. Rev. Res. **4**, 023019 (2022).

[7] E. T. Campbell, Early fault-tolerant simulations of the Hubbard model, Quant. Sci. Technol. **7**, 015007 (2022).

[8] Y. Tong, Designing algorithms for estimating ground state properties on early fault-tolerant quantum computers, Quant. Views **6**, 65 (2022).

[9] A. Katabarwa, K. Gratsea, A. Caesura, and P. D. Johnson, Early fault-tolerant quantum computing, arXiv:2311.14814.

[10] R. Kshirsagar, A. Katabarwa, and P. D. Johnson, On proving the robustness of algorithms for early fault-tolerant quantum computers arXiv:2209.11322.

[11] G. Wang, D. E. Koh, P. D. Johnson, and Y. Cao, Minimizing estimation runtime on noisy quantum computers, PRX Quant. **2**, 010346 (2021).

[12] A. Katabarwa, A. Kunitsa, B. Peropadre, and P. Johnson, Reducing runtime and error in vqe using deeper and noisier quantum circuits, arXiv:2110.10664.

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

[14] R. Zhang, G. Wang, and P. Johnson, Computing ground state properties with early fault-tolerant quantum computers, Quantum **6**, 761 (2022).

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

[16] G. Wang, D. Stilck-França, R. Zhang, S. Zhu, and P. D. Johnson, Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision, Quantum **7**, 1167 (2023).

[17] T. Giurgica-Tiron, I. Kerenidis, F. Labib, A. Prakash, and W. Zeng, Low depth algorithms for quantum amplitude estimation, Quantum **6**, 745 (2022).

[18] G. Wang, D. S. França, G. Rendon, and P. D. Johnson, Faster ground state energy estimation on early fault-tolerant quantum computers via rejection sampling, arXiv:2304.09827.

[19] Z. Ding and L. Lin, Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with ap-

plications to ground-state energy estimation, PRX Quant. **4**, 020331 (2023).

[20] H. Ni, H. Li, and L. Ying, On low-depth algorithms for quantum phase estimation, Quantum **7**, 1165 (2023).

[21] Z. Ding, Y. Dong, Y. Tong, and L. Lin, Robust ground-stateenergy estimation under depolarizing noise (unpublished).

[22] A. Dalal and A. Katabarwa, Noise tailoring for robust amplitude estimation, New J. Phys. **25**, 023015 (2023).

[23] 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).

[24] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, in *Electron. Colloquium Comput. Complex.* (1995).

[25] 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).

[26] 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).

[27] N. J. Ross and P. Selinger, Optimal ancilla-free Clifford+T approximation of z-rotations, Quantum Info. Comput. **16**, 901 (2016).

[28] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (Lond.) **549**, 172 (2017).

[29] D. Horsman, A. G. Fowler, S. Devitt, and R. Van Meter, Surface code quantum computing by lattice surgery, New J. Phys. **14**, 123011 (2012).

[30] D. Litinski, A game of surface codes: Large-scale quantum computing with lattice surgery, Quantum **3**, 128 (2019).

[31] C. Chamberland and E. T. Campbell, Universal quantum computing with twist-free and temporally encoded lattice surgery, PRX Quant. **3**, 010331 (2022).

[32] M. E. Beverland, P. Murali, M. Troyer, K. M. Svore, T. Hoefler, V. Kliuchnikov, G. H. Low, M. Soeken, A. Sundaram, and A. Vaschillo, Assessing requirements to scale to practical quantum advantage, arXiv:2211.07629.

[33] 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).

[34] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[35] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[36] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[37] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, J. Math. Phys. **48**, 052104 (2007).

[38] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, Nat. Phys. **14**, 595 (2018).

[39] N. Yoshioka, T. Okubo, Y. Suzuki, Y. Koizumi, and W. Mizukami, Hunting for quantum-classical crossover in condensed matter problems, arXiv:2210.14109.

[40] M. Fellous-Asiani, J. H. Chai, R. S. Whitney, A. Auffèves, and H. K. Ng, Limitations in quantum computing from resource constraints, PRX Quant. **2**, 040335 (2021).

[41] S. Arora, Le c tur e Notes on COS 521: Advanced Algorithm Design (2015), https://www.cs.princeton.edu/courses/archive/fall15/cos521/.

[42] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. Roy. Soc. Lond. Ser. A: Math. Phys. Eng. Sci. **454**, 339 (1998).