# The power of one qubit in quantum simulation algorithms

Polla, S.

# CHAPTER 3

# Error mitigation via verified phase estimation

## 3.1 Introduction

Error mitigation is likely essential for near-term quantum computations to realize valuable applications. State-of-the-art technology in superconducting qubits has recently pushed quantum computers beyond the capability of their classical counterparts [3] and enabled intermediate scale demonstrations of quantum algorithms for optimization [103, 104], quantum chemistry [59, 103, 105], and machine learning [106], with tens of qubits and hundreds of quantum gates. However, these experiments clearly reveal a noise barrier that needs to be overcome if such applications will ever scale to the classically intractable regime. In the long-term, a path towards this goal is known through quantum error correction [107–109]. Yet, the requirements to successfully error correct large-scale quantum applications [14, 71, 73, 110, 111] are still a few orders of magnitude above the current state-of-the-art, and will likely require many years to achieve. In the meantime, quantum applications research has focused on finding the elusive beyond-classical NISQ (noisy, intermediate-scale quantum) application [18], with the hope to accelerate the path to practical quantum computing. However without the resources to correct errors, one must develop strategies to mitigate the aforementioned noise barrier. Otherwise, the output of NISQ devices will be corrupted beyond usefulness for

algorithms significantly more complex than those already attempted.

Much of the attention in the NISQ era has been directed towards variational algorithms, with applications in optimization [55], chemistry and materials science [82], and machine learning [112, 113]. These shift much of the complexity of the algorithm to a classical outer loop involving many circuit repetitions, leaving the quantum computer with the task only of preparing quantum states and estimating expectation values of operators on said states. However, preparation circuits need to have significant depth to avoid being classically simulated [114]. Errors accumulated over this circuit quickly distort the prepared state to one different than was targeted. This has meant that most quantum experiments to date have had difficulty achieving standard accuracy benchmarks prior to applying error mitigation techniques [59, 103, 103, 105, 115]. However, accuracy improvements of orders of magnitude have been achieved with error mitigation in these experiments, suggesting there may yet be hope for NISQ.



The zoo of error mitigation techniques is large and varied. One may first attempt to design algorithms that are naturally noise-robust. For example, the optimization procedure in a variational algorithm makes the algorithm robust against control errors (e.g. over- or under-rotations when gates are applied) [82]. Also, subspace expansions of the variational quantum eigensolver (VQE) in materials science and chemistry correct errors that keep one within the desired subspace considered [116] or more generally through by approximate symmetry projection [117]. Given the ability to artificially introduce additional noise into a device, one can extrapolate from multiple experiments at different noise levels to a hypothetical noiseless experiment [118], which has shown promising results on real devices [119]. One may alternatively probabilistically compile circuits by inserting additional gates to average out or cancel out noise, given sufficient knowledge of the error model of the device [118, 120]. When classically post-processing partial state tomography data from an experiment, one may attempt to regularize the obtained results using reduced density matrix constraints [121]. Finally, one may mitigate errors that take a state outside of a symmetry-conserving subspace of a quantum problem, either by direct post-selection, or artificial projection of the estimated density matrix in post-processing, producing a 'symmetry-verified' state [29, 66, 117, 122]. Recent efforts have extended this protocol by introducing symmetries into problems to increase the range of errors that may be detected [123], which is analogous to the way quantum error correcting codes introduce engineered symmetries.

Ideally, we would prefer to go beyond verifying that a system's state remains within a target subspace and instead directly verify that the system's

state is the one we desire. This would result in reaching the information theoretic optimal limit of post-selected error mitigation in which one could completely mitigate the effect of all errors by repeating the experiment a number of times scaling inversely with the circuit fidelity (equivalent to the ability to perfectly detect errors). The fact that the circuit fidelity is expected to decrease exponentially in the gate complexity indicates that eventually we will still need error-correction; however, moving closer to this limit is certain to enable more powerful NISQ experiments.

In this work we develop a method for error mitigation of quantum phase estimation experiments, by verifying that the system returns to its initial state after the phase estimation step. We show that the set of experiments that pass this condition contain all the necessary information to perform quantum phase estimation. This yields a powerful error mitigation technique, as in most cases errors will not return the system to this initial state. Our techniques apply to variants of phase estimation which might involve post-processing on a single control qubit [124, 125], or when performing recently-developed control-free variants [27, 126]. We further develop it into a simple scheme for verified expectation value estimation by dividing a target Hamiltonian into a sum of fast-forwardable terms. This yields a simple, low-cost scheme for the measurement of expectation values, which may be immediately incorporated into the quantum step of a variational quantum algorithm. We study the mitigation power of this protocol in numerical simulations of small-scale experiments of free-fermion, transverse Ising, and electronic structure Hamiltonians. Verification is observed to mitigate all single (and even all double) errors throughout many of these simulations, as evidenced by a clear second (or third)-order sensitivity in our results to the underlying gate error rate. We observe in the best-case scenario case an up to $10,000$-fold suppression of error at physical error rates; this is not achieved for all systems studied, but verification is found to improve experimental error in all simulations performed. We find the error mitigation power to be highly system-, circuit-, and noise model-dependent. Finally, we study the measurement cost of this protocol in the presence of sampling noise, finding that it is comparable to standard partial state tomography techniques for energy estimation.

The outline of this paper is as follows. In Sec. 3.2, we give a pedagogical example of how one might verify the estimation of expectation values of an arbitrary Hamiltonian, by writing it as a sum of Pauli operators and performing (fast-forwarded) verified phase estimation on each individual term. In Sec. 3.3 we then derive the theory behind verified phase estimation itself, outline how it can mitigate errors, give algorithms for performing verified phase estimation with a single control qubit, or with access to a

reference state, and study the increased sampling noise cost. In Sec. 3.4, we extend these ideas to give algorithms for verified expectation value estimation, and derive the conditions under which one may perform verified estimation of multiple expectation values in parallel (i.e. using the same system register). In Sec. 3.5, we then implement these ideas, studying the mitigation power of verified expectation value estimation in a variety of systems and implementations developed earlier in the text ounder various noise models, and testing the convergence of the protocol under sampling noise.



## 3.2 Pedagogical example of verification protocol for expectation value estimation

In this section we outline a simple implementation of verified expectation value estimation of a target operator $H$ on a state $|\psi\rangle$, as a practical example of the more complicated methods to be found later in the text. The idea behind all verification protocols is to prepare $|\psi\rangle = U_{\mathrm{p}}|0\rangle$, indirectly estimate $\langle H\rangle$ via phase estimation, and then verify that we remain in $|\psi\rangle$ by uncomputing $|0\rangle = U_{\mathrm{p}}^{\dagger}|\psi\rangle$ and measuring in the computational basis. If $|\psi\rangle$ is not an eigenstate of $H$, the system may by shifted away from this state by the QPE unitary — i.e. even in the absence of error we do not expect the system to always pass verification. However, as we will show later in this work, the data required for phase estimation is contained entirely within the set of experiments that pass verification; we may effectively ignore any experiments that fail. This in turn allows us to ignore any errors that knock the system away from $|\psi\rangle$, making this a potent error mitigation scheme. We have constructed various implementations of this idea, which we will expand on in Sec. 3.3 and Sec. 3.4, and compare in Sec. 3.5. However, the most general protocols require relatively complicated circuits and classical post-processing. For clarity of exposition, in this section we focus on stepping through a simple protocol for the verification of expectation values, which avoids complex signal processing and circuity requirements. The protocol we describe will work for arbitrary $H$ and $|\psi\rangle$, and may often be a desirable choice for a real experiment. However, depending on the choice of $H$ and $|\psi\rangle$ and the noise model, other protocols described later in the text may be more optimal in terms of their mitigation power.

A process diagram for a simplified verified phase estimation protocol is given in Fig. 3.1. To begin, we write $H$ as a sum of fast-forwardable

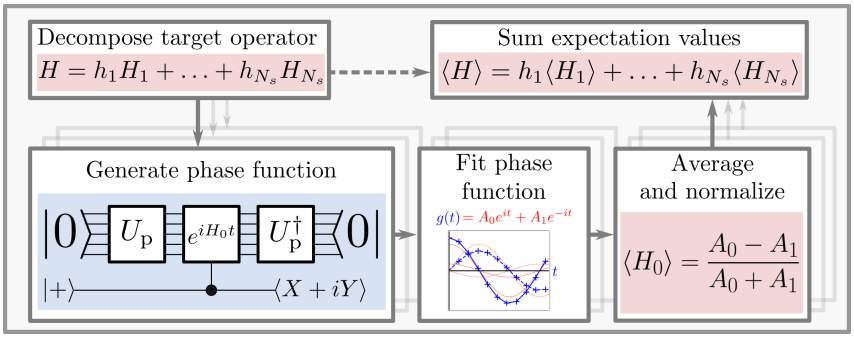


Figure 3.1: Process diagram of the protocol for verified estimation of the expectation value of a Hamiltonian on a state $|\psi\rangle = U_\mathrm{p}|\vec{0}\rangle$. Blue denotes circuits to be executed or data to be extracted from a quantum computer, red denotes signal details to be estimated via classical post-processing. The protocol proceeds as follows: (top-left) a complex Hamiltonian $H$ is split into a number of fast-forwardable summands $H_s$. The spectral function $g(t)$ of $|\Psi\rangle$ under time evolution of each piece is obtained (bottom-left) via verified, fast-forwarded phase estimation. In this example, a control qubit is used to extract the phase function via phase kickback. The resulting data is a weighted sum of oscillations with frequencies equal to the eigenvalues $E_j^{(s)}$ of the corresponding factor (bottom middle). This may be decomposed in a variety of classical post processing techniques to obtain estimations of the expectation values $\langle H_s \rangle$ depending on the type of $H_s$ chosen (bottom-right). Regardless of the method used, the expectation values must be normalized to obey Eq. 3.29, the last step in the verification process. As the expectation value is linear, the verified estimates of $\langle H_s \rangle$ obtained may be immediately summed together to give a verified estimate for $\langle H \rangle$ (top-right).

terms $H_s$ (multiplied by coefficients $h_s$)

$$H = \sum_{s=1}^{N_s} h_s H_s. \tag{3.1}$$

Here, by fast-forwardable, we mean that each $H_s$ is chosen such that time evolution $e^{iH_s t}$ may be implemented on a quantum register with the same number of gates for each value of $t$. Although fast-forwarding is forbidden for arbitrary $H$ [97], decomposition of any sparse, row-computable $H$ into

a linear combination of polynomially many fast-forwardable Hamiltonians is always possible [127]. For example, the $N$-qubit Pauli operators $P_i \in \mathbb{P}^N = \{\mathbb{1}, X, Y, Z\}^{\otimes N}$ form a basis for the set of all $N$-qubit operators and are themselves fast-forwardable; we take this decomposition for our simple example.

We then implement verified phase estimation (with a single control qubit) to estimate the expectation values $\langle\psi|H_s|\psi\rangle$. This involves evolving the system by $H_s$ conditional on a control qubit. (Circuits to implement this are well-known, see e.g. Ref. [128].) The conditional evolution encodes a phase function on the the control qubit. That is, if we write $X_\mathrm{c}$ and $Y_\mathrm{c}$ for the $X$ and $Y$ Pauli operators on this control qubit, we have following the conditional evolution that

$$\langle X_\mathrm{c}\rangle + i\langle Y_\mathrm{c}\rangle = A_0 e^{it} + A_1 e^{-it} =: g(t). \tag{3.2}$$

Here, $A_0$ and $A_1$ are the squared amplitudes of $|\psi\rangle$ in the eigenbasis of $H_s$ (which has known eigenvalues $\pm 1$). The expectation value $\langle X_\mathrm{c}\rangle$ may be estimated by measuring the control qubit $M$ times in the $x$-basis, counting the number of times $m_{x,0}$ or $m_{x,1}$ a 0 or 1 was seen, and approximating

$$\langle X_\mathrm{c}\rangle \approx \frac{m_{x,0} - m_{x,1}}{M}. \tag{3.3}$$

(A similar procedure may be performed for $Y$.) To verify this estimate, we uncompute the preparation of the system, and count the number $m_{x,0}^{(v)}$ ($m_{x,1}^{(v)}$) of measurements of 0 (1) on the control qubit when the uncomputed state on the system is returned to the initial $|0\rangle$ state. We then replace our estimation by

$$\langle X_\mathrm{c}\rangle \approx \frac{m_{x,0}^{(v)} - m_{x,1}^{(v)}}{M}. \tag{3.4}$$

(Note that we only replace the numerator, and not the denominator, of Eq. 3.3, which makes this not strictly post-selection — see Sec. 3.3.2 for more details.) The expectation value $\langle H_s\rangle$ is encoded within the phase function $g(t)$, and must be inferred from these estimates above. In our example protocol, this requires inferring the amplitudes $A_0$ and $A_1$ (as the eigenvalues $\pm 1$ are already known). These may be simply estimated by a two-parameter fit of Eq. 3.2 to the extracted values of $g(t)$.

As we show later in the text, in the absence of error Eq. 3.3 and Eq. 3.4 yield the same result (in the large $M$ limit). Errors tend to scatter the system into a state that fails verification. The primary effect this has on the estimator in Eq. 3.4 is to re-scale $g(t) \to p_\mathrm{ne} g(t)$ (where $p_\mathrm{ne}$ is the

probability of no error occurring). However, the converse is not true; states may fail verification due to the relative dephasing betwen the $|0\rangle$ and $|1\rangle$ eigenstates of $H_s$, and we cannot infer the value of $p_{\text{ne}}$ from a single point $g(t)$. Instead, we can infer the value of $p_{\text{ne}}$ from the normalization of the starting state $|\psi\rangle$. As our circuit is fast-forwarded, under reasonable noise assumptions $p_{\text{ne}}$ is independent of $t$, and this propagates immediately through the fit of Eq. 3.2: $A_0, A_1 \to p_{\text{ne}}A_0 =: \tilde{A}_0, p_{\text{ne}}A_1 =: \tilde{A}_1$. The normalization of $|\psi\rangle$ requires $A_0 + A_1 = 1$, and we may correct for this by estimating

$$\langle H_s \rangle = \frac{\tilde{A}_0 - \tilde{A}_1}{\tilde{A}_0 + \tilde{A}_1}. \tag{3.5}$$

Finally, as expectation values are linear, after repeating this procedure for all $H_s$ in Eq. 3.1, we may sum the result;

$$\langle H \rangle = \sum_s h_s \langle H_s \rangle. \tag{3.6}$$

Note that each $H_s$ will have different values of $A_0$, $A_1$, and $g(t)$ (we have avoided explicitly labeling the above for simplicity). In practice, the number of samples for estimation of each $\langle H_s \rangle$ should be varied to minimize the error in the final estimation of $\langle H \rangle$ (i.e. importance sampling on the $h_s$ coefficients).

## 3.3 Schemes for verified phase estimation

### 3.3.1 Review of single-control quantum phase estimation

Quantum phase estimation (QPE) refers to a family of protocols to learn eigenphases $e^{i\phi_j}$ of a unitary operator $U$. Equivalently, quantum phase estimation may be used to learn eigenvalues $E_j$ of a Hermitian operator $H$, as each such operator generates a unitary via exponentiation: $U = e^{iHt}$ [62]. (Such estimation requires limiting the size of $t$ to prevent aliasing - $e^{iE_j t} = e^{iE'_j t}$ if $E_j t = E'_j t + 2n\pi$, which makes estimation ambiguous.) The eigenvalues of $H$ and the eigenphases of $U$ are related by the same exponentiation and correspond to the same eigenstates $|E_j\rangle$ — if $H|E_j\rangle = E_j|E_j\rangle$, $U|E_j\rangle = e^{i\phi_j}|E_j\rangle$, and $\phi_j = E_j t$.

In the single-control variant of QPE, the phases $\phi_j$ are learnt by imprinting them on a control qubit — a process known as phase kickback. Any unitary $U$ may be implemented as a (perhaps approximate) quantum

circuit on a quantum 'system' register, but quantum mechanics tells us that $e^{i\phi}|\psi\rangle \equiv |\psi\rangle$ for all pure states $|\psi\rangle$ and numbers $\phi \in \mathbb{R}$. This implies that if the system register were prepared in the pure state $|E_j\rangle$ and $U$ applied, we would not be able to infer the phase $\phi_j$ from the resulting state $e^{i\phi_j}|E_j\rangle \equiv |E_j\rangle$. However, a relative phase $\phi$ between two states, $\frac{1}{\sqrt{2}}(|\psi_1\rangle + e^{i\phi}|\psi_2\rangle)$, is a physical observable that may be detected. Such detection may be achieved by acting the unitary $U$ conditional on the control qubit being in the state $|1\rangle$ (and doing nothing when the control qubit is in the state $|0\rangle$). This is commonly written as the 'controlled' unitary $\mathrm{C}-U$. When $\mathrm{C}-U$ acts on a system register prepared in an eigenstate $|E_j\rangle$ and a control qubit prepared in the state $(|0\rangle + |1\rangle)/\sqrt{2}$, the global state evolves to

$$\mathrm{C}-U\frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)|E_j\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\phi_j}|1\rangle)|E_j\rangle. \tag{3.7}$$

We see that the eigenphase $e^{i\phi_j}$ from the system register is kicked back onto the control qubit, while the system register itself remains unchanged. We may estimate this eigenphase $e^{i\phi_j}$ by repeatedly performing the QPE protocol, measuring the control qubit in the $X$ or the $Y$ basis, and recording the number of single-shot readouts of 1 and 0. In the Hamiltonian case, from this estimate one may immediately infer $\frac{1}{it}\mathrm{Arg}(e^{i\phi_j}) = E_j \mod 2\pi t$. The error in the estimation of $E_j$ decreases with $t$; asymptotically optimal protocols need to balance this against the ambiguity modulo $2\pi t$ by repeating the estimation at multiple values of $t$ [129–131]. In terms of estimating the eigenphases $e^{i\phi_j}$ of a unitary $U$, this optimization requires repeating the above procedure for $\mathrm{C}-U^k$ at varying points $k$.

Often, one does not prepare an eigenstate $|E_j\rangle$, but instead prepares a starting state

$$|\psi_\mathrm{s}\rangle = \sum_j a_j|E_j\rangle. \tag{3.8}$$

Applying $\mathrm{C}-U^k$ to such a state no longer leaves it unchanged, but instead entangles it with the control qubit. This produces the combined state (on the system+control register)

$$|\Psi(k)\rangle = \mathrm{C}-U^k\frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)|\psi_\mathrm{s}\rangle \tag{3.9}$$

$$= \sum_j \frac{a_j}{\sqrt{2}}(|0\rangle + e^{ik\phi_j}|1\rangle)|E_j\rangle. \tag{3.10}$$

When one has instead performed controlled time evolution (via the unitary $\mathrm{C}-e^{iHt}$), one may instead write

$$|\Psi(t)\rangle = \mathrm{C}-e^{iHt}\frac{1}{\sqrt{2}}\left(|0\rangle+|1\rangle\right)|\psi_{\mathrm{s}}\rangle \tag{3.11}$$

$$= \sum_j a_j \frac{1}{\sqrt{2}}(|0\rangle + e^{iE_j t}|1\rangle)|E_j\rangle. \tag{3.12}$$

The sum over $j$ in the above equation looks problematic, but it turns out that the eigenphases $\phi_j$ (or eigenvalues $E_j$) remain encoded on the control qubit, in a sum weighted by the norm square $A_j := |a_j|^2$ of the initial amplitudes $a_j$. To be precise, one may trace over the system register to obtain the reduced density matrix of the control qubit

$$\rho_{\mathrm{c}}(t) = \mathrm{Trace}_{\mathrm{sys}}\left[|\Psi(t)\rangle\langle\Psi(t)|\right] \tag{3.13}$$

$$= \frac{1}{2}\begin{pmatrix} 1 & g(t) \\ g^*(t) & 1 \end{pmatrix}, \tag{3.14}$$

with $g(t)$ the phase function of $|\psi_{\mathrm{s}}\rangle$ under $H$

$$g(t) = \sum_j A_j e^{iE_j t}. \tag{3.15}$$

Estimates of $g(t)$ may be obtained as an expectation value

$$g(t) = 2\,\mathrm{Trace}_{\mathrm{c}}\left[\rho_{\mathrm{c}}(t)|0\rangle\langle 1|\right] \tag{3.16}$$

$$= \mathrm{Trace}_{\mathrm{c}}\left[\rho_{\mathrm{c}}(t)X\right] + i\mathrm{Trace}_{\mathrm{c}}\left[\rho_{\mathrm{c}}(t)Y\right], \tag{3.17}$$

of the Pauli operators $X$ and $Y$. Measuring these expectation values requires rotating the control qubit into the $x$- or $y$-basis, reading it out, and averaging the output over many repetitions (or shots) of the experiment.

For a unitary operator $U$ one may obtain an equivalent phase function

$$g(k) = \sum_j A_j e^{ik\phi_j}, \tag{3.18}$$

by estimating

$$g(k) = 2\,\mathrm{Trace}_{\mathrm{c}}\big[\rho_{\mathrm{c}}(k)|0\rangle\langle 1|\big] \tag{3.19}$$

$$= \mathrm{Trace}_{\mathrm{c}}\big[\rho_{\mathrm{c}}(k)X\big] + i\mathrm{Trace}_{\mathrm{c}}\big[\rho_{\mathrm{c}}(k)Y\big], \tag{3.20}$$

$$\rho_{\mathrm{c}}(k) = \mathrm{Trace}_{\mathrm{sys}}\big[|\Psi(k)\rangle\langle\Psi(k)|\big], \tag{3.21}$$

with $|\Psi(k)\rangle$ defined in Eq. 3.10. The tomography to extract these expectation values is the same as described in the previous paragraph.

Information about the eigenvalues $E_j$ and amplitudes $A_j = |a_j|^2$ may be inferred classically from estimates of $g(t)$ at multiple values of $t$. When these are estimated sufficiently well, the expectation value of the Hamiltonian may be calculated

$$\langle H\rangle = \sum_j A_j E_j. \tag{3.22}$$

Inference of the amplitudes $A_j$ from $g(t)$ to error $\epsilon$ takes asymptotically time $\Omega(\epsilon^{-2})$ on a quantum device, even when the eigenvalues $E_j$ are already known [1]. By propagating variances, this implies equivalent convergence in the estimation of expectation values via Eq. 3.22. One need not resolve all $2^N$ eigenvalues of an $N$-qubit operator in order to evaluate Eq. 3.22. Time-series analysis methods [125] or integral methods [132] produce a coarse-grained approximation to the spectrum that may be averaged over to obtain expectation values with similar convergence rates. Alternatively, for simple operators with highly-degenerate spectrum (e.g. Pauli operators), curve fitting will be sufficient to extract the required data (as described in Sec. 3.2)[2].

### 3.3.2 Verifying a phase estimation experiment

As the data from single-control quantum phase estimation is accumulated entirely on the control qubit, one would be tempted to throw the system register away (or rather, reset the register and begin anew). In the absence of error correction this temptation grows larger; noise levels in near term devices are high enough that coherent states of more than a few qubits degrade over the course of any reasonably-sized algorithm to within a few percent fidelity to the target state — if not less [103]. However, even when

[1] This may be calculated via Cramer-Rao bounds as the derivative $\frac{\partial g(t)}{\partial A_j}$ is bounded as a function of $t$, which is not true for the derivative $\frac{\partial g(t)}{\partial E_j}$.

[2] The minimum number of points on the curve that require fitting is determined by the number of eigenvalues and amplitudes that need fitting.

corrupted, the information contained within the system register is valuable, as one can use this information to diagnose potential errors in the data to be read from the control qubit. For instance, in the presence of global symmetries of the Hamiltonian, one could imagine mitigating errors that do not commute with this symmetry via symmetry verification [29, 122, 123]. In verifying these symmetries, we are in effect projecting the system into a subspace of the global Hilbert space which contains the information we desire. One could imagine constructing ever-smaller Hilbert spaces, which trades circuit complexity for error-detection power. It turns out that the limit of this construction is achievable: instead of measuring one or more symmetries on the system register, we can instead verify that it has returned to its initial state $|\psi_\mathrm{s}\rangle$. (This is similar to the echo-type measurements made in randomized benchmarking [133] or quantum Hamiltonian learning [134].)



Assuming $|\psi_\mathrm{s}\rangle$ is prepared from the computational basis state $|0\rangle$ by a preparation unitary $U_\mathrm{p}$, this measurement may be achieved by applying $U_\mathrm{p}^\dagger$, and reading out each qubit in the computational basis. One would expect such a measurement to distort the phase function $g(t)$, but this is not so, as we may expand the trace in Eq. 3.13 to show that

$$\begin{aligned}&\mathrm{Trace}_\mathrm{c}\big[\rho_\mathrm{c}(t)|0\rangle\langle 1|\big]\\&\qquad=\mathrm{Trace}_\mathrm{c}\Big[\langle\psi_\mathrm{s}|\Psi(t)\rangle\langle\Psi(t)|\psi_\mathrm{s}\rangle|0\rangle\langle 1|\Big].\end{aligned}\tag{3.23}$$

Here, the left-hand side of the equation is the expectation value of $\rho_\mathrm{c}(t)$ regardless of the state of the control register, and the right-hand side is the (non-normalized) expectation value of $\rho_\mathrm{c}(t)$ on verified experiments only. The lack of normalization means this is not a post-selection technique; instead one assumes that the contribution of states which fail verification to the final estimation of $g(t)$ is zero. (By contrast, states that pass verification either contribute $+1$ or $-1$ to the estimation of $g(t)$.)

We can make a physical argument why Eq. 3.23 holds and verification should not affect the estimation of $g(t)$ in the absence of noise. Let us decompose the reduced density matrix on the control qubit

$$\rho_\mathrm{c}=\rho_\mathrm{c}^{(\mathrm{v})}+\rho_\mathrm{c}^{(\mathrm{f})},\tag{3.24}$$

into the ensemble of states $\rho_\mathrm{c}^{(\mathrm{v})}$ that have passed verification, and those that have failed. When the control qubit is in the $|0\rangle$ state, the system register is not evolved, so in the absence of noise the state will pass verification every time. This implies that a verification failure in the absence of noise projects

the control qubit into the $|1\rangle$ state; $\rho_c^{(f)} = |1\rangle\langle 1|$. As $\mathrm{Trace}[|1\rangle\langle 1|0\rangle\langle 1|] = 0$, this fraction of states on average contributes nothing to the estimate of $g(t)$. In other words,

$$\mathrm{Trace}[\rho_c|0\rangle\langle 1|] = \mathrm{Trace}[\rho_c^{(v)}|0\rangle\langle 1|] = g(t). \tag{3.25}$$

Note that post-selecting (i.e. keeping only the experimental data where verification was passed) would instead prepare the state $\rho_c^{(v)}/\mathrm{Trace}[\rho_c^{(v)}]$. This will not yield the desired result, as

$$\frac{\mathrm{Trace}[\rho_c^{(v)}(X+iY)]}{\mathrm{Trace}[\rho_c^{(v)}]} = \frac{g(t)}{1+|g(t)|^2} \tag{3.26}$$

which is not equal to $g(t)$ unless $|\psi_s\rangle$ is an eigenstate of $e^{iHt}$ (in which case $\rho_c^{(v)} = \rho_c$). (Moreover, this rescaling can be up to a factor 2 in the absence of noise, and the spectrum of this new function is significantly different to the original.) To give some intuition, one can imagine phase estimation on a mixed state in two steps: performing phase estimation on individual states to generate a set of signal functions $e^{iE_j t}$, and then summing and returning the weighted result $g(t)$. The set of states that fail verification, $\rho^{(f)}$, captures the relative dephasing between these states, which cannot be ignored when attempting to recover this result. Instead, an explicit protocol for the measurement of a single $g(t)$ within verified single-control phase estimation takes the form of Algorithm 3.1. We consider the increased sampling cost in the presence of error in Sec. 3.3.3.

### 3.3.3 Why verification mitigates errors

The mitigation power from verification is based on the relative size of the Hilbert spaces in which the states which have passed verification and states which have failed verification, $\rho = \rho^{(v)} + \rho^{(f)}$, live. If we define the Hilbert spaces in which the two ensembles live $\mathcal{H}^{(v)}$ and $\mathcal{H}^{(f)}$ respectively, we have $\dim[\mathcal{H}^{(v)}] = 2$, while $\dim[\mathcal{H}^{(f)}] = 2^{N+1} - 2$. An error that occurs during the circuit is then likely to scatter the system into the set of rejected states. As an extreme example, the probability that a completely random error (i.e. an error that scatters all states to a random state) at any point in the circuit will yield a state in $\mathcal{H}^{(v)}$ can be immediately calculated to be $2/(2^{N+1} - 2) \sim 2^{-N}$. This includes errors during preparation of $|\psi_s\rangle$ by the unitary $U_p$ and the inversion of $U_p^\dagger$ to perform the verification itself. As we are not post-selecting on the verification output $g(t)$ is still affected by this shift, but the distortion may be accounted for in classical

**Algorithm 3.1:** Single-control VPE

**Input:** Circuits to implement $U_\mathrm{p}, U_\mathrm{p}^\dagger$ and controlled time evolution $e^{iHt}$.
Number of repetitions $M$ of measurement in the $x$ and $y$ basis.

**Output:** An estimate of $g(t)$ with variance $O(\frac{1}{M})$ in both the real and imaginary part.

1 Prepare classical initial variables $g^x = 0$, $g^y = 0$;
2 Prepare the system register in a starting state $|\psi_\mathrm{s}\rangle = U_\mathrm{p}|0\rangle$ and the control qubit in the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$;
3 Simulate time evolution $e^{iHt}$ conditional on the control qubit;
4 Apply the inverse circuit $U_\mathrm{p}^\dagger$ to the system register;
5 Rotate the control qubit into the $X$ or $Y$ basis and measure it to obtain a number $m \in [0, 1]$;
6 If all qubits in the system register read 0, increment the relevant variable $g^x$ or $g^y$ by $(-1)^m$;
7 Repeat steps 2-6 $M$ times in the $X$ basis and $M$ times in the $Y$ basis, and estimate $g(t)$ by $\frac{g^x}{M} + i\frac{g^y}{M}$

post-processing. In this simple noise model the effect of noise is then to replace the estimate of $g(t)$ by

$$g_\mathrm{err}(t) = p_\mathrm{ne}(t)g(t) + \mathcal{O}(2^{-N}p_\mathrm{err}(t)), \tag{3.27}$$

where $p_\mathrm{ne}(t)$ and $p_\mathrm{err}(t)$ are the probabilities of no error or some error occurring, respectively. (In App. 3.A we derive the specific requirements for this to be the case.) Assuming that errors occur at a constant rate as a function of the circuit depth, and all scatter the system outside $\mathcal{H}^{(\mathrm{v})}$, for fast-forwardable Hamiltonians $p_\mathrm{ne}(t) = p_\mathrm{ne}$, and

$$g_\mathrm{err}(t) = p_\mathrm{ne}g(t) = \sum_j (p_\mathrm{ne}A_j)e^{iE_jt}. \tag{3.28}$$

This can be seen as a uniform damping of each squared amplitude $A_j$ to $A'_j = p_\mathrm{ne}A_j$. Such damping may be corrected for classically as we know $|\psi_\mathrm{s}\rangle$ is normalized

$$\sum_j A_j = 1, \tag{3.29}$$

and so we may estimate

$$A_j = \frac{A'_j}{\sum_j A'_j}. \tag{3.30}$$

Depending on the classical signal processing method used, one may not obtain estimates of all $A'_j$ and $E_j$, but may instead directly calculate $\sum_j A'_j E_j$ and $\sum_j A'_j$. For example, one could use $g_{\mathrm{err}}(0) = \sum_j A'_j$ as such a reference point. For non-fast-forwardable Hamiltonian, assuming again that errors occur at a constant rate throughout the circuit and that all scatter the system outside $\mathcal{H}^{(\mathrm{v})}$, we have

$$g_{\mathrm{err}}(t) = e^{-t/\tau_{\mathrm{err}}} g(t) = \sum_j A_j e^{i(E_j + i/\tau_{\mathrm{err}})t}. \tag{3.31}$$

This can be seen to be an imaginary shift to the eigenvalues $E_j \to E_j + i\tau_{\mathrm{err}}$. It can be corrected for in signal processing of the phase function by taking only the real parts of the $E_j$ eigenvalues.

The above analysis is not necessarily true for simulation of an arbitrary Hamiltonian under a realistic noise model. In particular, if the instantaneous state during simulation is a near-eigenstate of the error model, then the correction in Eq. 3.27 may be as large as $\mathcal{O}(1)$ instead of $\mathcal{O}(2^{-N})$. In App. 3.A we study this in more detail, and specify the conditions under which errors will distort the results of verified phase estimation.

### Sampling costs

The error mitigation from verification comes at the cost of increasing the number of samples require to estimate $g(t)$. Assuming all errors fall outside the verified subspace, estimating $g(t)$ to precision $\epsilon$ requires estimating $g_{\mathrm{err}}(t)$ to precision $p_{\mathrm{ne}}\epsilon$. To obtain $g^x$ in Alg. 3.1 (and equivalently for $g^y$) we average over a set of $M$ experimnetal outputs that may take the values $\{-1, 0, 1\}$. Let us define the $i$th experimental output $g_i^x$, and we have

$$P(g_i^x = \pm 1) = \frac{1}{2} p_{\mathrm{ne}}(1 \pm g^x), \tag{3.32}$$

$$P(g_i^x = 0) = 1 - p_{\mathrm{ne}}. \tag{3.33}$$

Our estimate of the noisy $g_{\mathrm{err}}(t)$ is then given by

$$\mathrm{Re}[g_{\mathrm{err}}(t)] = P(g_i^x = 1) - P(g_i^x = -1). \tag{3.34}$$

As each experiment is IID, the variance on our estimates of these probabilities is

$$\begin{aligned}\mathrm{Var}[P(g_i^x = \pm 1)] &= \frac{1}{M}\frac{1}{2}p_{\mathrm{ne}}(1 \pm g^x) \\ &\quad \times \left(1 - \frac{1}{2}p_{\mathrm{ne}}(1 \pm g^x)\right), \qquad (3.35)\\ \mathrm{Cov}[P(g_i^x = 1), &P(g_i^x = -1)] \qquad (3.36)\\ &= -\frac{1}{4M}p_{\mathrm{ne}}^2\left(1 - [g^x]^2\right). \qquad (3.37)\end{aligned}$$

Propagating variances obtains

$$\mathrm{Var}\left[\mathrm{Re}\left[g_{\mathrm{err}}(t)\right]\right] = \frac{1}{M}p_{\mathrm{ne}} - \frac{1}{M}p_{\mathrm{ne}}^2[g^x]^2. \qquad (3.38)$$

We may then bound the requirements to estimate $g_{\mathrm{err}}(t)$ to variance $\epsilon^{-2}p_{\mathrm{ne}}^{-2}$ by

$$M \geq \epsilon^{-2}p_{\mathrm{ne}}^{-1}. \qquad (3.39)$$

This is exactly what one would expect from an actual post-selection technique (i.e. where $Mp_{\mathrm{ne}}$ samples were used to estimate $g(t)$). We remind the reader that $p_{\mathrm{ne}}$ here is the probability of no error occurring over the entire circuit. As one should expect for an error mitigation technique, this in turn grows exponentially with the size of the circuit required to implement $e^{iHt}$ or $U_{\mathrm{p}}$. In a simple model, if the error per qubit per moment is $p$ (i.e. assuming qubit decay is more dominant than gate noise in the model), an $N$-qubit circuit of depth $d$ would have

$$p_{\mathrm{ne}} = (1-p)^{Nd}, \qquad (3.40)$$

and thus the number of shots required to estimate (the real or imaginary part) of $g(t)$ would scale as

$$M \sim (1-p)^{-Nd}\epsilon^{-2}. \qquad (3.41)$$

This is not to be ignored; verification requires at least doubling the size of the circuit, which if $p_{\mathrm{ne}} = 0.01$ (as has been reported [3] and mitigated successfully [103] in previous experiments) will increase the measurement count by a factor of 100. Some of the methods presented in this work involve increasing the circuit depth by factors of up to 14, which will be impractical for large experiments without further circuit optimization.

### Control noise

An important realistic error to consider in QPE is error on the control qubit. This keeps the system within the verified subspace, and so is not captured by the above analysis. However the effect of many common error channels may still be mitigated by verification. For example, let us assume that the circuit decomposition of $\mathrm{C}-U$ involves the control qubit performing only single-qubit gates and controlled operations on the rest of the circuit (which is typically the case). In this case, one may show that the effect of a depolarizing channel of strength $\lambda$

$$R_{\mathrm{depol}}[\rho] = (1 - \frac{3\lambda}{4})\rho + \frac{\lambda}{4}(X\rho X + Y\rho Y + Z\rho Z), \tag{3.42}$$

acting on the control qubit at any point in the circuit, sends the final state of the system to

$$(1-\lambda)\rho_{\mathrm{ne}} + \lambda\rho_{\mathrm{err}}, \tag{3.43}$$

where $\rho_{\mathrm{ne}}$ is the state in the absence of error, and

$$\mathrm{Trace}[\langle\psi_s|\rho_{\mathrm{err}}|\psi_s\rangle|0\rangle\langle 1|] = 0. \tag{3.44}$$

In this case, the (noisy) estimate of $g(t)$ is sent to $(1-\lambda)g(t)$, and expectation values and eigenvalues may be recovered via the same analysis as in Sec. 3.3.3. However, the above analysis will not hold for a more general noise model, and schemes such as randomized compiling [135] may be required to unbias the estimate of $g(t)$. An example of this biasing effect is if an amplitude-damping channel

$$\begin{aligned} R_{\mathrm{ampdamp}}[\rho] =& (1-\lambda)\rho + \frac{\lambda}{2}(Z+I)\rho(Z+I) \\ &+ \frac{\lambda}{2}(X+iY)\rho(X-iY), \end{aligned} \tag{3.45}$$

is present on the control qubit between the final measurement pre-rotation and readout in the computational basis. Left unchecked, this will shift the estimate of $g(t)$ to

$$g_{\mathrm{err}}(t) = (1-\lambda)g(t) + \lambda. \tag{3.46}$$

In addition to damping the true signal $g(t)$, this additive signal presents as a 0-energy eigenvalue in the spectrum of $g(t)$. This will not be accounted for by naive renormalization of $\langle H\rangle$ as outlined in Alg. 3.3; the estimation protocol will instead estimate $(1-\lambda)\langle H\rangle$. Though this could be corrected in post-processing, we suggest that a more stable mitigation is to flip the

$|0\rangle$ and $|1\rangle$ states on the control qubit for 50% of experiments. This may be compiled into the final pre-rotation, and does not increase the total sampling cost of the experiment (only half as many samples need to be taken at each pre-rotation setting for the same accuracy). We observe similar biases on bitflip noise channels which tend to decay the real and imaginary parts of $g(t)$ asymmetrically. This may be compensated for in turn by compiling a $\frac{\pi}{4}$ $Z$-rotation on the intial control qubit state, and uncompiling it in the final prerotation. (One can see that this commutes with all gates in the circuit). For the noise models studied numerically in this text we have found either one or both of the above compilation schemes sufficient to mitigate control error. More complicated noise models may required more complicated compilation schemes; extending the above will be an interesting task for future work. In particular, the above analysis does not apply to correlated two-qubit noise during operations between the control qubit and the rest of the system.



### 3.3.4 Verified control-free phase-estimation

As was recently demonstrated in Ref. [27], the control qubit may be removed from a QPE experiment if we have the ability to prepare an alternative reference eigenstate $|\psi_\mathrm{r}\rangle$ of the Hamiltonian $H$ (with $\langle\psi_\mathrm{s}|\psi_\mathrm{r}\rangle = 0$). For example, in the electronic structure problem in quantum chemistry the number-conserving Hamiltonian has the vacuum as a potential reference state. (A similar situation was considered in Ref. [136] for the purposes of random gap estimation, but estimating single eigenvalues $E_j$ from this class of experiments is somewhat awkward.) This was also recently considered as an extension to the well-known robust QPE scheme [137], requiring both $|\psi_\mathrm{r}\rangle$ and $|\psi_\mathrm{s}\rangle$ to be eigenstates of the system [126]. Note that $|\psi_\mathrm{r}\rangle$ need not necessarily be a zero-energy eigenstate of $H$, though the corresponding eigenenergy $E_r$ should be known to high accuracy. In this case, one need prepare the correlated state $\frac{1}{\sqrt{2}}(|\psi_\mathrm{s}\rangle + |\psi_\mathrm{r}\rangle)$, and perform uncontrolled time evolution, and finally measure the off-diagonal element $|\psi_\mathrm{s}\rangle\langle\psi_\mathrm{r}|$. This is shown in the circuit (Fig. 3.2). Evaluating the circuit provides an estimate of

$$\begin{aligned}\mathrm{Trace}\Big[U(|\psi_\mathrm{r}\rangle + |\psi_\mathrm{s}\rangle)(\langle\psi_\mathrm{r}| + \langle\psi_\mathrm{s}|)U^\dagger|\psi_\mathrm{r}\rangle\langle\psi_\mathrm{s}|\Big] \\ = e^{-iE_r t}g(t),\end{aligned} \tag{3.47}$$

and the additional phase may be subtracted in post-processing.

The protocol for verified control-free phase estimation does not differ

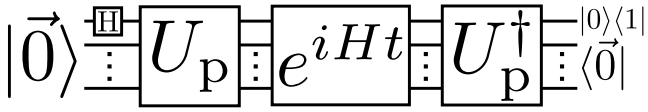


Figure 3.2: Quantum circuit for control-free verified phase estimation. The preparation unitary $U_\mathrm{p}$ is defined in Eq. 3.48. The first gate in the circuit is a Hadamard gate (Roman H) on the top-most qubit (labeled the target qubit in text), which should not be confused with the Hamiltonian $H$.

significantly from the single-control case. Besides the loss of the control qubit and removal of control from the time evolution circuit, we also now require our preparation circuit to prepare the starting state $\frac{1}{\sqrt{2}}(|\psi_\mathrm{s}\rangle + |\psi_\mathrm{r}\rangle)$. We assume that this is achieved by first applying a Hadamard gate to a single target qubit in the system register, placing the system in the state $\frac{1}{\sqrt{2}}(|0\rangle + |\vec{1}_\mathrm{T}\rangle)$. (Here we use the notation $|\vec{1}_\mathrm{T}\rangle$ for the basis state where the target qubit is in the $|1\rangle$ state and all other qubits are in $|0\rangle$.) Then, the desired preparation may be achieved by a preparation unitary $U_\mathrm{p}$ which performs the mapping

$$U_\mathrm{p}|0\rangle \mapsto |\psi_\mathrm{r}\rangle, \qquad U_\mathrm{p}|\vec{1}_\mathrm{T}\rangle \mapsto |\psi_\mathrm{s}\rangle. \tag{3.48}$$

(We use the same notation as for the single-control unitary on purpose, as under the association $|0\rangle|\psi_\mathrm{s}\rangle \leftrightarrow |\psi_\mathrm{r}\rangle$ and $|1\rangle \leftrightarrow |\psi_\mathrm{s}\rangle$ one may see the two are equivalent.) With this definition, estimation of $|\psi_\mathrm{r}\rangle\langle\psi_\mathrm{s}|$ may be achieved by inverting $U_\mathrm{p}$, as

$$|\psi_\mathrm{r}\rangle\langle\psi_\mathrm{s}| = U_\mathrm{p}|0\rangle\langle\vec{1}_\mathrm{T}|U_\mathrm{p}^\dagger. \tag{3.49}$$

In particular, after inversion, the reduced density matrix of the target qubit contains the desired phase function $g(t)$, and the verification consists of checking whether all other qubits are measured into 0. The resulting control-free protocol is fully written out in Algorithm 3.2. The analysis of Sec. 3.3.3 is identical for the control-free case, with the absence of the issue of control noise, as is the analysis of Sec. 3.3.3. However, we note that at the beginning and the end of any experiment, single-qubit noise on the target qubit behaves similarly to control qubit noise. This necessitates averaging over multiple initial and final rotations of the target qubit to prevent bias in the estimation of $g(t)$.

The above analysis implies that the algorithms studied in Ref. [27, 136] should be amenable to verification immediately as well. It also provides

**Algorithm 3.2:** Control-free VPE

**Input:** Circuits to prepare a superposition of $|\psi_\mathrm{s}\rangle$ and $|\psi_\mathrm{r}\rangle$, invert the preparation, and implement time evolution $e^{iHt}$.
Number of repetitions $M$ and $M$ of measurement in the $x$ and $y$ basis.
The reference eigenstate energy $E_r$

**Output:** An estimate of $g(t)$ (Eq. 3.47) with variance $O(\frac{1}{M})$ in both the real and imaginary part.

1 Prepare classical initial variables $g^x = 0$, $g^y = 0$ ;
2 Prepare the system register in a starting state $\frac{1}{\sqrt{2}}(|\psi_\mathrm{s}\rangle + |\psi_\mathrm{r}\rangle) = U_\mathrm{p}\frac{1}{\sqrt{2}}(|0\rangle + |\vec{1}_\mathrm{T}\rangle)$ ;
3 Apply the unitary $U^k$, (or equivalently simulate time evolution $e^{iHt}$) ;
4 Apply the inverse circuit $U_\mathrm{p}^\dagger$ to the system register ;
5 Rotate the target qubit into the $X$ or $Y$ basis and measure it to obtain a number $m \in 0, 1$ ;
6 Measure all other qubits, and if they all read out 0, increment the relevant variable $g^x$ or $g^y$ by $(-1)^m$ ;
7 Repeat steps 2-6 $M$ times in the $X$ basis and $M$ times in the $Y$ basis, and estimate $g(t)$ by $e^{iE_r t}(\frac{g^x}{M} + i\frac{g^y}{M})$.

some additional explanation for the error-robustness observed in the robust phase estimation of Ref. [126].

## 3.4 Verified expectation value estimation

In many circumstances, one wishes not to know the eigenvalues of a Hermitian operator $H$, but instead its expectation value $\langle H\rangle$ under a specified state $|\Psi\rangle$. For instance, in a variational quantum eigensolver [82], one prepares a state $|\Psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle$ dependent on a set of classical input parameters $\vec{\theta}$, then measures the expectation value $E(\vec{\theta}) = \langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle$. This is then optimized over $\vec{\theta}$ in a classical outer loop, with the optimized state $|\Psi(\vec{\theta}_\mathrm{opt})\rangle$ hopefully a good approximation of the true ground state $|E_0\rangle$. In quantum variational algorithms it is typical that $\langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle$ is estimated by means of partial state tomography [66, 67, 138]. However, noise in the preparation unitary $U(\vec{\theta})$ causes an errant state $\rho_\mathrm{err}(\vec{\theta}) \neq$

**Algorithm 3.3:** Verified expectation value estimation

**Input:** (Noisy) circuits to implement $U_\mathrm{p}$, $U_\mathrm{p}^\dagger$ and controlled time evolution $e^{iHt}$.
A set of $t$ values.
Number of repetitions $M$ of measurement in the $x$ and $y$ basis (that can be $t$-dependent).
A method for classical signal processing (e.g. a curve fitting algorithm).

**Output:** An estimate of $\langle H \rangle$.

1 Estimate $g_\mathrm{err}(t)$ for all given points $t$ using Alg. 3.1 to the chosen precision;
2 Obtain estimates for individual $E_j$ and $A'_j$ values via classical signal processing;
3 Estimate $\langle H \rangle$ as

$$\langle H \rangle = \frac{\sum_j A'_j E_j}{\sum_j A'_j}. \tag{3.51}$$

$|\Psi(\vec{\theta})\rangle\langle\Psi(\vec{\theta})|$ to be prepared and tomographed, propagating the preparation error directly to a final estimation error. The noise analysis in Sec. 3.3.3 extends to both the preparation and mitigation unitaries, so if verified phase estimation is used to provide estimates of eigenvalues and amplitudes, one may reconstruct

$$\langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle = \sum_j |a_j|^2 E_j, \tag{3.50}$$

and inherit the mitigation power of the verification protocol. This has the added advantage that control errors in the preparation circuit (which, being a repeated error, are not mitigated against) are able to be compensated for during the outer optimization loop of the VQE, as is well-known [82, 103]. Quantum phase estimation has previously been suggested as an alternative to partial state tomography for expectation value estimation, both to improve the rate of estimation [139], and to provide a witness for the presence of eigenstates of the Hamiltonian [140]. The verification protocols described in this work should be applicable to these methods as well. A general algorithm for verified expectation value estimation takes the form of Algorithm 3.3

One might worry that the sum in Eq. 3.50 is over an exponentially large number of eigenstates $|E_j\rangle$. However one need not resolve all eigenvalues $E_j$ in order to accurately estimate the expectation value $\langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle$; if eigenvalues within $\delta$ of each other are binned, the resulting expectation value will be accurate to within $\delta$. We may formalize this by considering the spectral function $g_{\mathrm{S}}$ of $|\psi_{\mathrm{s}}\rangle$ under $H$,

$$g_{\mathrm{S}}(E) = \sum_j A_j \delta(E - E_j). \tag{3.52}$$

This can be seen to be the Fourier transform of the phase function $g(t)$ [strictly, $g(t)$ is the inverse Fourier transform of $g_{\mathrm{S}}(E/2\pi)$], and a coarse-grained approximation may be obtained via time-series methods [125] or integral methods [132] with rigorous bounds on each. Numerically, we find signal processing methods such as Prony's method [124] also perform acceptably (see Sec. 3.5.4). For fast-forwardable Hamiltonians (such as Pauli operators), one often already knows the target eigenvalues of the problem. Furthermore, the eigenspectrum of these Hamiltonians is often highly degenerate, making simple curve fitting a practical (and attractive) alternative.

Instead of analysing the phase function at many points as described above, one may expand

$$\mathrm{Im}[g(t)] = \sum_j |a_j|^2 \sin(E_j t) \tag{3.53}$$

$$= t \sum_j |a_j|^2 E_j + \frac{1}{3} t^3 \sum_j |a_j|^2 E_j^3 + \mathcal{O}(t^5) \tag{3.54}$$

$$\frac{1}{t}\mathrm{Im}[g(t)] = \langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle + \mathcal{O}(t^2), \tag{3.55}$$

and simply estimate $\mathrm{Im}(g(t))$ for short times $t$. This is similar to the manner in which eigenphases are estimated in the WAVES protocol [140] (sans verification). In this case, the normalisation of the resulting amplitudes (Eq. 3.30) must be achieved by the condition that $g(0) = \sum_j A_j$, yielding

$$\langle H\rangle = \frac{\mathrm{Im}[g_{\mathrm{err}}(t)]}{t|g_{\mathrm{err}}(0)|} + \mathcal{O}(t^2). \tag{3.56}$$

### 3.4.1 Fast-forwarded and parallelized Hamiltonian decompositions

As expectation values are linear, we may estimate $\langle H \rangle$ by splitting it into multiple terms, estimating the expectation values of each term individually, and re-summing;

$$H = \sum_s H_s \to \langle H \rangle = \sum_s \langle H_s \rangle. \tag{3.57}$$

If individual $H_s$ may be simulated at lower circuit depth, this can reduce the accumulation of unmitigated errors, at the cost of requiring more simulation. This ability becomes especially useful if one chooses the $H_s$ to be fast-forwardable. Here, we define a fast-forwardable Hamiltonian $H_s$ as one for which a circuit implementation of $e^{iH_s t}$ has constant depth in $t$. The circuit depth required to simulate $e^{iHt}$ for arbitrary $H$ is bounded below as $\mathcal{O}(t)$ [97], but for certain operators this may be improved on [141]. For example, as the Pauli operators $\{\mathbb{1}, X, Y, Z\}^{\otimes N}$ are both fast-forwardable and form a basis for the set of $N$-qubit Hermitian operators, a set of $H_s$ terms may be taken from these to decompose an arbitrary Hamiltonian. As another example, given an instance of the electronic structure problem, one may attempt a low-rank factorization of the interaction operator into a sum of $\mathcal{O}(N)$ diagonalizable (and thus fast-forwardable) terms [142].

In order to speed up estimation of expectation values of multiple terms $H_s$ in a decomposed Hamiltonian $H = \sum_s H_s$, it may be possible perform the verified phase estimation step of each $H_s$ in parallel. For example, we can perform time evolution of $L$ multiple summands, each controlled by a different control qubit, in between the preparation and verification steps of a single instance. In the absence of verification, such parallelization will not affect the outcome of quantum phase estimation of any individual $H_s$, so long as all terms estimated in parallel commute. This follows immediately from the fact that the time evolution for one such term does not evolve the system between eigenspaces of another. This is complicated by the addition of verification, as the additional circuitry means that the system may evolve away from $|\psi_s\rangle$ despite a specific control qubit being in $|0\rangle$. In App. 3.B, we show that this gives rise to a set of spurious signals in the estimated phase function $g^{(s)}(t)$:

$$g_q^{(s)}(t) = \sum_{v,j,j'} B_{j,j'}^{(s)} e^{iF_{v,j,j'}^{(s)} t}. \tag{3.58}$$

Here, the ghost eigenvalues are

$$F^{(s)}_{v,j,j'} = E^{(s)}_j + \sum_{s' \neq s} v_{s'} \left( E^{(s)}_j - E^{(s')}_{j'} \right), \tag{3.59}$$

where $E^{(s')}_j$ are the true eigenvalues of the Hamiltonians $H_{s'}$, and $v$ is a $L$-bit vector written in binary (i.e. $v_s \in 0, 1$). The corresponding, $v$-independent amplitudes are

$$B_{j,j'} = \frac{1}{2^L} A_j A_{j'}. \tag{3.60}$$

Although this is a far more complicated signal than the standard phase function $g(t)$, we calculate in App. 3.B that it yields the same expectation value; i.e.

$$\sum_{v,j,j'} B_{j,j'} F^{(s)}_{v,j,j'} = \langle H_s \rangle. \tag{3.61}$$

This implies that verified parallel phase estimation may proceed in much the same way as the series protocol.

### 3.4.2 Comparison to other methods of error mitigation

Error mitigation techniques differ vastly, both in their cost to implement and their effectiveness against different forms of noise. This implies that care needs to be taken in a real experiment to choose the best mitigation technique (or combination of mitigation techniques) for the job. Though a comparison between multiple techniques in a realistic setting lies outside the scope of this work, we give some predictions here on how VPE might compare in performance to other mitigation techniques, and whether it might be possible to compare to different techniques. We can classify all error mitigation techniques that the authors know of into the following broad categories

- *Circuit design* — many forms of noise may be mitigated by careful design of a circuit to e.g. minimize crosstalk between simultaneous gates [103], cancel out Z over- or under-rotation (e.g. via echo pulses [143]), or optimize a circuit variationally to cancel out control parameter drift on a long timescale [82, 144]. (Whether or not this counts as error mitigation or calibration of the underlying quantum device is left to the reader to decide.) Depending on the source of noise these techniques may significantly reduce or even nullify its effect, which may be far more effective than VPE. On the other hand, noise sources such as T1 error

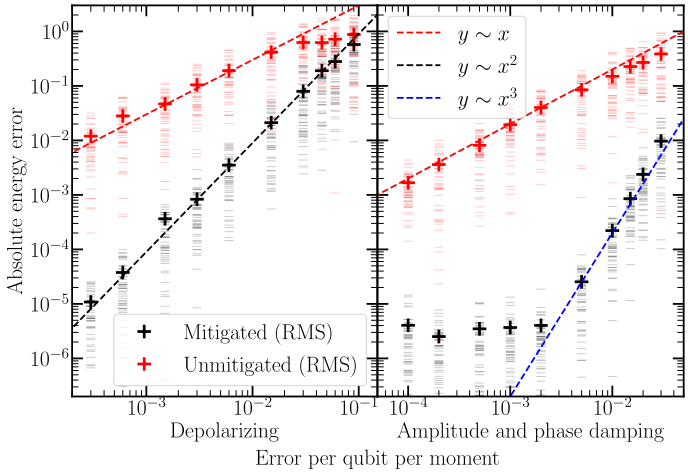


Figure 3.3: Mitigation of a 4-qubit Givens rotation circuit via verified phase estimation. (left) Error in estimation of random states in a free-fermion system (Eq. 3.63) under a uniform depolarizing channel. (right) Error in the same estimation, but this time under an amplitude and phase damping model. In both plots, the RMS error (crosses) is calculated over 50 different estimations for each error rate using either standard partial state tomography (red) or using verified control-free phase estimation. Individual data points (dashes) are additionally shown. For reference, dashed lines showing linear (red), quadratic (black), and cubic (blue) dependence on the gate error rate are plotted.

cannot be easily calibrated away (due to the associated photon loss); in these situations (where VPE performs quite well) these methods will have little effect. VPE is clearly compatible with any such techniques, as these consist of adjustments to the implementation of a given circuit rather than an algorithmic overhead.

- *Post-selection or verification techniques* — this class of techniques uses knowledge of the problem to restrict the state of the quantum device to within a small region of the $N$-qubit Hilbert space, often by leveraging symmetries of the Hamiltonian of the problem to be solved. VPE itself falls into this category, alongside symmetry verification [29, 67], and quantum subspace expansion techniques [116, 117]. The performance of these techniques is dependent on their ability to catch errors outside the allowed Hilbert space, so as the dimension of the Hilbert space for VPE

is only 2, we expect it to have greater mitigation power in general than these other techniques. (This can be observed in App. 3.G, where VPE shows an asymptotic improvement over symmetry verification in a small numerical simulation.) However, as the circuit depth of VPE is typically far longer than that of other post-selection or verification techniques (which can be achieved in some cases without any additional circuitry), the requirements on the number of measurements to overcome sampling noise will be significantly worse. As these techniques overlap in their effect on the quantum state, it is not particularly possible to combine them; instead one should choose the best trade-off between mitigation power and number of measurements.

- *Error extrapolation techniques* — assuming that one can artificially introduce noise into a system, these techniques rely on parameterizing the output $f$ of a quantum circuit as a function of a 'noise parameter' $f = f(\lambda)$, fitting a functional form, and extrapolating to $\lambda = 0$. The noise parameter can either be adjusted experimentally (e.g. by adjusting the wait time or detuning of an underlying gate) [118, 120] or algorithmically (e.g. by inverting noisy gates [145]). The mitigation power of such a technique depends on how well the noise can be tuned as a function of this single parameter, and how well one can pin down a functional form for $f(\lambda)$. This is not easily comparable to VPE, as the physical source of the mitigation is qualitatively significantly different. We expect that the relative performance will depend on the experiment and the hardware itself. In theory these methods could be combined with VPE (either by extrapolating the phase function or the VPE result). However, it is unclear whether the output of VPE will be more challenging to fit, reducing the effectiveness of the extrapolation.

- *Result extrapolation techniques* — instead of fitting the output $f$ of a quantum circuit to an artificial noise term, one can consider comparing the output of similar quantum circuits tailored to efficient classical simulation. This technique has been demonstrated experimentally in Refs. [103, 146], and proposed within a VQE setting (by tuning the parameters to points where the solution is known) [147]. In some sense VPE can be considered to be similar to these methods, with the $|\psi_\mathrm{s}\rangle|0\rangle$ or $|\psi_\mathrm{r}\rangle$ states providing an entangled reference state for the target evolution. However, this relationship is not completely clear, as VPE strictly relies on the coherence between the two states. Understanding this similarity is a clear avenue for future research. Regardless, VPE should be able to be combined with at least some of these techniques to provide yet more mitigation power.

- *Probabilistic cancellation techniques* — given knowledge of the true process maps of the gates being performed on a quantum device, one can in principle construct families of quantum circuits that when combined, yield a target noiseless result [118, 120]. However, these methods require much additional characterization of the device, which is a problem in systems with large amounts of drift. In principle given sufficient knowledge of the noise this method works perfectly, but at a greatly increased measurement cost, making it difficult to make a fair comparison in a theoretical setting. Testing this method against VPE in a real experiment would be an interesting target for future research.
- *Purification techniques* — as the output of a quantum algorithm is often ideally pure, these techniques attempt to reduce errors by mapping a noisy impure state to a purer one. This may be achieved e.g. for free fermion states via McWeeny purification [103], or for more general states via virtual distillation [148]. For more complex states the McWeeny process cannot be used, but it has proven remarkably effective when available. Virtual distillation and VPE appear to be remarkably similar in their increased measurement cost and their mitigation performance, as well as their circuit structure. Understanding this similarity and comparing the two in more detail is a clear avenue for future research.



## 3.5 Numerical Experiments

To investigate the mitigation capability of verified phase estimation, we first use it for expectation value estimation. To prepare states, we take different variational ansatze with randomly-drawn parameters. We compare the performance of verified and unverified circuits across multiple target Hamiltonians, noise strengths and noise models, to attempt to identify trends in the method. All simulations were executed using the Cirq quantum software development framework [102] and simulators therein. Hamiltonians and complex circuits were further generated using code from the OpenFermion [117] libraries. Except for when mentioned, the Cirq noise models were chosen to be a constant error rate per qubit per moment, where a moment is a period of the circuit where gates occur. Equivalently, this can be thought of as an error rate per qubit per gate, but including error on idling gates as well. The noise models considered are not as complex as those typically observed in experiment (which are typically highly non-uniform, and can include crosstalk and non-Markovianity alongside other effects), but we expect our results should provide a suggestion of the mitigation power of this method in a real quantum device.

### 3.5.1 Givens rotation circuits for free-fermion Hamiltonians

We first test the mitigation ability of the verification protocol on an instance of a "Givens rotation circuit" of the form developed for implementing rotations of single-particle fermionic basis functions in [149]. This circuit takes the form

$$U(\vec{\theta}) = \exp\left(i \sum_{j,l} \vec{\theta}_{j,l} c_j^\dagger c_l\right), \tag{3.62}$$

where $c_j^\dagger$ and $c_j$ are the creation and annihilation operators for a fermion on site $j$, and $\theta_{j,l} = \theta_{l,j}$. Such a circuit is classically simulatable, but it is a critical piece of infrastructure in quantum computing applications for quantum chemistry [66, 71, 73, 103, 142]. It is also low depth: it may be decomposed exactly by a sequence of matchgates [150], with optimal compilation in a circuit depth of exactly $N$. When acting on a $N$-qubit register prepared in the state $\prod_{n=0}^{N_f-1} X_n|0\rangle$, this may prepare an arbitrary ground state of a free-fermion Hamiltonian with $N_f$ particles by an appropriate choice of $\vec{\theta}$. In this work, we take a simple free-fermion Hamiltonian as an example - namely a one-dimensional chain:

$$H = -t \sum_j c_j^\dagger c_{j+1} + \text{h.c.} \tag{3.63}$$

Such a Hamiltonian may be diagonalized,

$$H = V^\dagger \sum_\alpha \epsilon_\alpha c_\alpha^\dagger c_\alpha V, \tag{3.64}$$

where $V$ here takes the same form as in Eq. 3.62. This decomposition allows immediately for the fast-forwarding of time evolution, as

$$e^{iHt} = V^\dagger e^{it\sum_\alpha \epsilon_\alpha c_\alpha^\dagger c_\alpha} V \tag{3.65}$$

$$= V^\dagger \prod_\alpha e^{it\epsilon_\alpha c_\alpha^\dagger c_\alpha} V. \tag{3.66}$$

As the Givens rotation circuits conserve particle number, the vacuum $|0\rangle$ may be used as a reference state for control-free verified estimation. A superposition of this reference state and starting state $U(\vec{\theta}) \prod_{n=1}^{N_f} X_n|0\rangle$

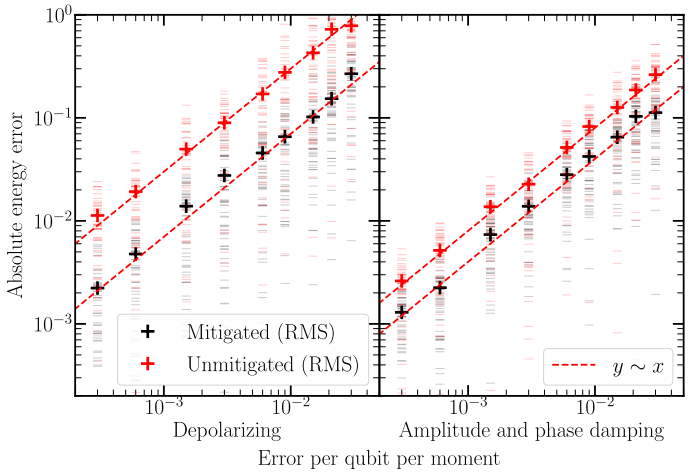


Figure 3.4: Mitigation of a 4-qubit VHA circuit via verified phase estimation. (left) Error in estimation of the energy of random states generated by the quantum approximate optimization ansatz in the critical phase of the transverse-field Ising model (Eq. 3.70) under a uniform depolarizing channel. (right) Error in the same estimation, but this time under an amplitude and phase damping model. In both plots, the RMS error (crosses) is calculated over 50 different estimations for each error rate (with randomly-chosen ansatz parameters) using either standard partial state tomography (red) or using verified control-free phase estimation. Individual data points (dashes) are additionally shown. For reference, dashed lines showing linear (red) dependence on the gate error rate are plotted.

may be prepared by acting the Givens rotation circuit on the GHZ state

$$|\mathrm{GHZ}_{N_f}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + \prod_{n=1}^{N_f} X_n|0\rangle\right), \tag{3.67}$$

which may itself be prepared by e.g. a chain of CNOT gates:

$$|\mathrm{GHZ}_{N_f}\rangle = \prod_{j=N_f-1}^{1} \mathrm{CNOT}_{j-1,j}\mathrm{H}_0|0\rangle. \tag{3.68}$$

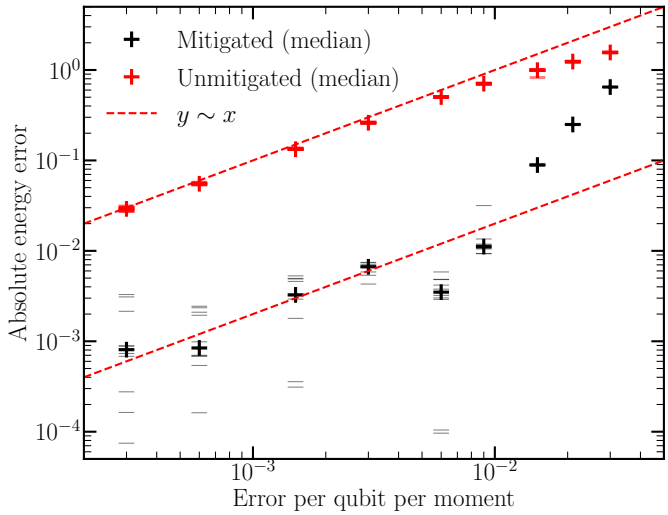


Figure 3.5: Error in estimating the ground state energy of a 4-site transverse-field Ising model (Eq. 3.70) by variational optimization of a VHA ansatz. The resulting expectation values are measured either by verified single-control phase estimation (black) or taken directly from the simulated state state (red). We plot the median (crosses) of the absolute energy error over 10 optimization attempts, each starting from a different initial point. Individual errors are plotted behind (faint dashes). Guide lines showing a linear dependence are additionally plotted (red dashed lines).

Note here the backwards product that runs left-to-right (i.e. the CNOT gate between qubit 1 and qubit 0 is executed first). Following the definitions in Sec. 3.3.4 for verified control-free phase estimation, we can write the complete preparation unitary as

$$U_{\mathrm{p}} = U(\vec{\theta}) \prod_{j=N_f-1}^{1} \mathrm{CNOT}_{j-1,j}. \tag{3.69}$$

Then, as the product of two Givens rotation circuits is itself a Givens rotation circuit [149], we may compile $VU(\vec{\theta}) = U(\vec{\theta}')$ and implement this in a single Givens rotation circuit.

The complete VPE circuit for this circuit consists of the GHZ preparation,

a single Givens rotation, a set of single-qubit z-rotations, uncomputing the Givens rotation, uncomputing the GHZ preparation, and measurement in the $X$ or $Y$ basis. The resulting circuit for verified phase estimation is more than twice the length of the circuit required for the unmitigated VQE. We assume here that the VQE tomography does not require any additional overhead, and directly estimate the expectation value from the simulated density matrix. For verified phase estimation, we extract the phase function from the simulated density matrix, and then process it to estimate expectation values using Prony's method. In order to not bias the final readout (which can lead to significant error in estimation), we average the rotation into the $X$ and $Y$-bases over both $+\pi/2$ and $-\pi/2$ rotations 3.3.3. To simplify the analysis here, we do not include additional sampling noise. In Fig. 3.3, we plot the RMS error for two error models over a range of noise models and strengths. For each noise model and at each strength we sample 50 random choices for the initial parameters $\vec{\theta}$ (and set $t = 1$ in Eq. 3.63). In the presence of a uniform single-qubit depolarizing channel (Fig. 3.3, left), we see that the verified error displays a clear $\epsilon \sim p^2$ trend (where $\epsilon$ is the error in the final estimation, and $p$ is the error per qubit per moment). This implies that the effect of all single errors in this noise model are suppressed by the error mitigation (or fortuitously cancel), but that pairs of errors near to each other in time may affect results. Under the effect of an amplitude and phase damping channel (Fig. 3.3, right), the suppression is even starker; we see a clear $\epsilon \sim p^3$ trend till the error drops to below $10^{-5}$, providing up to four orders of magnitude gain in precision. Below $10^{-5}$ the error plateaus. This is due to numerical stability issues with Prony's method, and not a fundamental limit of the procedure [3]. This level of estimation error only becomes relevant after $> 10^{10} p_{\mathrm{err}}^{-2}$ individual shots have been taken (with $p_{\mathrm{err}}$ the probability of an error over the entire circuit). As such, we expect this to not be relevant for most experiments. The lower error rate makes some sense: amplitude damping errors can only ever reduce the number of excitations in the circuit, and so by themselves can never return to a state with non-zero overlap with $|\psi_{\mathrm{s}}\rangle$. However, the precise mode for the leading contribution to the error rate is still somewhat unclear.

[3] Proof of this can be found in e.g. Fig. 3.9, where phase fitting obtained a $\sim$ 100-fold reduction in this noise floor, which is typical for all simulations performed thus far.

### 3.5.2 The variational Hamiltonian ansatz for the transverse-field Ising model

We next attempt the verification of a completely different model and ansatz. The transverse-field Ising model (TFIM) is a well-known spin system, with Hamiltonian

$$H = J_z \sum_j Z_j + J_x \sum_j X_j X_{j+1}, \tag{3.70}$$

where we take the sum $j+1$ modulo $N$ (i.e. periodic boundary conditions). In one dimension, this model has a critical phase when $J_z = J_x$, making this a simple model to study interesting quantum phenomena. Exact ground states of this model may be found by the variational Hamiltonian ansatz (VHA) [54] for any values of $J_x$ and $J_z$ [151]. The VHA consists of alternating the Ising model and transverse field terms $p$ times, with at each layer $p$ the amount of time to be treated as a free variable:

$$U(\vec{\theta}) = \prod_p e^{i\theta_{p,Z} \sum_j Z_j} e^{i\theta_{p,X} \sum_j X_j X_{j+1}}. \tag{3.71}$$

(Note that for this given model the VHA is equivalent to the quantum alternating operator ansatz of Ref. [55].) The TFIM does not have any simple eigenstates, and nor does the VHA, so simple methods of control-free verified phase estimation are not available. Instead, we attempt single-control verified phase estimation. To lower the error incurred during the circuit, we perform VPE in series for every term in Eq. 3.70. Unfortunately, verification works significantly less well in this setting, as is shown in Fig. 3.4. For both noise models considered, we see a clear $\epsilon \sim p$ trend with $\epsilon$ the energy error in the final result, and $p$ the error per qubit per moment. This suggests that errors that map the noiseless state into one with nontrivial overlap with the verified density matrix are dominant in this circuit. Regardless, we note that verification does provide an $\sim 8$-fold improvement in error rate over the unmitigated circuit, despite the verification circuit requiring one additional qubit and being three times as long. This result is lessened in the presence of amplitude and phase damping noise, till the point where the mitigation only improves estimation by a factor of 2.

Variational optimization is well-known to mitigate certain types of coherent noise (e.g. coherent parameter drift) [82, 144]; it also appears to provide some mitigation of incoherent noise when in combination with

verified phase estimation. In Fig. 3.5, we perform a variational outer loop over the circuit studied in Fig. 3.4. Although the $\epsilon \sim p$ behaviour appears to roughly remain in the latter half of the optimization, the gain from error mitigation improves from $2-8$x to around 50x, a significant improvement. We note that the optimization is no longer variationally bound - below about $10^{-2}$ error per qubit per moment, the results are scattered relatively evenly on either side of the true value. By contrast, in the absence of sampling noise partial state tomography results will always be variationally bound. We suspect this result may be due to the fact that slightly different circuits need to be run to measure different terms, yielding an 'effective state' that lies slightly outside the positive cone of allowed physical quantum states. Though this effect does not appear to be particularly severe in this case, further study may be needed to see it does not become an issue in larger experiments.



### 3.5.3 Fermionic swap networks for electronic structure Hamiltonians

As a final system for simulation, we move to studying the ability to verify molecular hydrogen on four qubits using a fermionic swap network. This ansatz was first studied in [149]; it consists of a network of two-qubit fermionic simulation gates, which take the form

$$U_{\text{fsim}}(\theta, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & i\sin(\theta) & 0 \\ 0 & i\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix}. \tag{3.72}$$

The parameters $\theta$ and $\phi$ are then left free to be optimized during the circuit. Molecular hydrogen is a simple example of the full electronic structure Hamiltonian, which takes the form

$$H = \sum_{i,j} t_{i,j} c_i^\dagger c_j + \sum_{i,j,k,l} V_{i,j,k,l} c_i^\dagger c_j^\dagger c_k c_l. \tag{3.73}$$

Solving this Hamiltonian for mid-to-large system sizes ($\sim 60+$ qubits) with strong interactions is a key target application for quantum computers [64, 71, 73].

We study three different methods for verified expectation value estimation of the electronic structure Hamiltonian. Following a transformation from fermionic to qubit operators, Eq. 3.73 we first consider a decomposi-

tion over single Pauli operators for single-control VPE, as was performed for the transverse-field Ising model in Sec. 3.5.2. However, in order to perform control-free VPE on these terms, we require a reference state. Individual fermionic terms in Eq. 3.73 are number-conserving, so the fermionic vacuum is a good reference state for these, but this is not the case for individual Pauli terms. To circumvent this problem, we split Eq. 3.73 into fermionic terms (summed with their Hermitian conjugate), and decompose these into Pauli operators. (One can check that the resulting Pauli operators commute, and so their time evolution may be easily fast-forwarded.) The VPE circuits in both of the above methods are $3-4$ times the depth of the original VQE.

Alternatively, by performing a low-rank factorization of the Coulomb operator, we may write $H$ in the form [142]

$$H = \sum_{i,j} t'_{i,j} c_i^\dagger c_j + \sum_l U_l^\dagger \left[ \sum_\alpha E_\alpha^{(l)} c_\alpha^\dagger c_\alpha \right]^2 U_l \tag{3.74}$$

$$= H^{(0)} + \sum_l H^{(l)}, \tag{3.75}$$

Where the $U_l$ are single-particle basis changes that may be implemented via Givens rotation circuits. Each such term in this factorization is fast-forwardable. $H^{(0)}$ is a free-fermion Hamiltonian and may be simulated via the methods discussed earlier in this section. The interacting factors $H^{(l)}$ may also be diagonalized by diagonalizing the single-particle $t_{i,j}^{(l)}$ matrices. One finds

$$e^{iU_l^\dagger H^{(l)} U_l t} = U_l^\dagger \prod_{\alpha\beta} E^{itE_\alpha E_\beta c_\alpha^\dagger c_\alpha c_\beta^\dagger c_\beta} U_l, \tag{3.76}$$

which may be easily implemented on superconducting hardware, as $e^{itE_\alpha E_\beta c_\alpha^\dagger c_\alpha c_\beta^\dagger c_\beta}$ is realised by a C-Phase gate. All of the above Hamiltonians, as well as the fermionic swap network itself, conserve particle number, and so we may again use the vacuum as a reference state for verified control-free quantum phase estimation. We do not consider the single-control version for comparison in this case. The resulting circuit is over ten times as long as the VQE itself, as we are unable to compile the final basis rotation into the ansatz.

The mitigation power of VPE differs vastly between the different choices of decomposition used, and the different noise models chosen. In Fig. 3.6, we plot the effect of mitigating depolarizing, and amplitude and phase

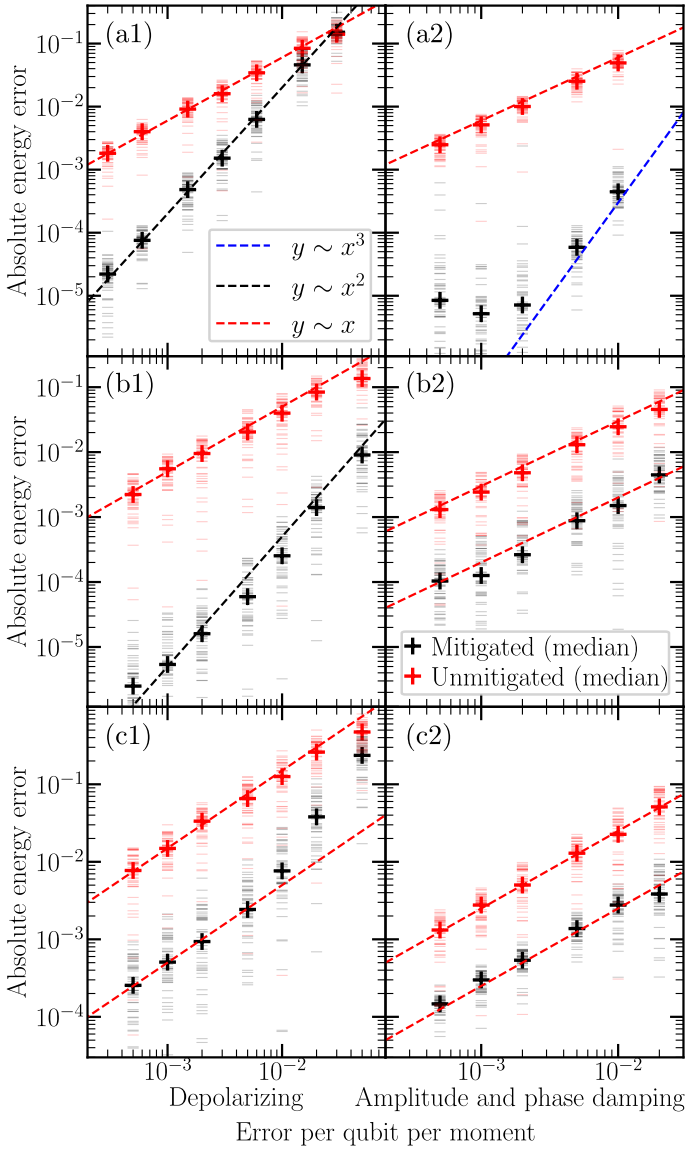


Figure 3.6: Mitigation of a 4-qubit fermionic swap network via verified phase estimation. (Continues on following page)

Figure 3.6: (continued) Three different VPE protocols are explored — a low-rank factorization (top row — a1, a2), a control-free number-conserving Pauli decomposition (middle row — b1, b2), and a single-control Pauli decomposition (bottom row — c1, c2). Details of all decompositions are given in the text. The low-rank factorization was studied for the $H_2$ Hamiltonian at the equilibrium bond distance with a swap network of depth 4, while the other two models were studied at a bond distance of 2 Angstrom with a swap network of depth 6. All protocols are tested under depolarizing (left column) and amplitude and phase damping (right column) noise models. In all plots, the median error (crosses) is calculated over 50 different estimations for each error rate using either standard partial state tomography (red) or using verified control-free phase estimation. Individual data points (dashes) are additionally shown. For reference, dashed lines showing linear (red), quadratic (black) and cubic (blue) dependence on the gate error rate are plotted.



damping channels, using the three decompositions described above. We see that control-free [Fig. 3.6(a1-b2)] VPE typically outperforms single-control VPE [Fig. 3.6(c1,c2)], despite the single-control VPE circuits being in all cases smaller (due to the lack of coherent state preparation). Under a depolarizing noise model, both control-free VPE implementations [Fig. 3.6(a1, b1)] demonstrate a second-order sensitivity to the physical qubit error rate, consistent with the previous results in Fig. 3.3. In this case, the Pauli decomposition clearly outperforms the low-rank factorization, which we attribute to the large reduction ($\sim 2-3\times$) in total circuit depth. However, although the low-rank factorization repeats the third-order sensitivity to amplitude and phase damping seen in Fig. 3.3 [Fig. 3.6(a2)], this is not observed in the Pauli decomposition case [Fig. 3.6(b2)]. We investigate this further in App. 3.F, and find that this first-order error can be traced back to the verified estimation of a single term — the two-body interaction term. We attribute this to the fact that the time evolution circuit for this term breaks number conservation (which is not the case for any other term in the sum), which makes it more susceptible to amplitude damping noise. Understanding this feature in detail, and determining whether better circuit optimizations exist, are clear targets for future research. In any case, all three implementations of VPE studied show at least an order of magnitude improvement compared to partial state tomography, and in some cases up to three orders of magnitude improvement, demonstrating the power of this technique.

### 3.5.4 Sampling costs

In a realistic experiment, direct estimation of any expectation value requires repeatedly re-preparing the target state and measuring in an appropriate basis to accumulate statistics on the probability of seeing a given 0 or 1 measurement. In verified phase estimation, this repetition must be performed instead on the control qubit (for single-control) or target qubit (for control-free) to accumulate the phase function. Re-preparation is necessary between subsequent measurements, as such a measurement collapses the global wavefunction, erasing the information about the probability to be estimated. This implies that each repetition carries substantial cost, and the rate of convergence of error estimation is a critical bottleneck in any variational algorithm. Although one might expect quantum phase estimation to speed up this estimation (which has been proposed previously [139]), this is only the case when one is estimating eigenvalues of the target Hamiltonian in a specific QPE instance. We wish to divide up our Hamiltonian for fast-forwarding purposes, and in most cases the resulting terms will not be simultaneously diagonalizable, so no set of mutual eigenstates will exist; instead, the results of Sec. 3.3.3 will hold. Furthermore, as our expectation value estimation requires to sum over multiple different amplitudes, we should not expect this to improve over the cost of partial state tomography (which requires non-commuting terms to be measured on separate preparations of the state). The error in expectation value estimation will further depend on the type of classical post-processing used.

In Fig. 3.7, we compare the convergence of two types of classical post-processing to that of standard partial state tomography. We perform this simulation on the 4-spin VHA-TFIM system studied in Fig. 3.4 and Fig. 3.5, on a representative point in the spectrum (the error-free variational minimum). We do not perform any measurement grouping or parallelization strategies for either method, and instead report our results as a function of the number of measurements per Pauli operator. The first method (green) assumes knowledge about the eigenvalues of the fast-forwarded Hamiltonians, in which case one need only fit the amplitudes, while the second (blue) first estimates the eigenvalues using Prony's method before fitting the amplitudes to the resulting signal. (We compensate for the presence of spurious phases in Prony's method by a slight adjustment described in App. 3.C.) All methods of estimation are seen to converge at a rate $\epsilon \sim M^{-1/2}$, where $\epsilon$ is the estimation error and $M$ is the number of samples taken.

We see that using the prior knowledge of the phases gives a significant

advantage in convergence, with the resulting error rate being almost an order of magnitude worse when using Prony's method. This advantage persists in the presence of a depolarizing channel (1% error rate), although the convergence of all methods flattens as they approach the sampling-noise-free estimation value. We note that both classical post-processing methods converge to the same result here, as expected. It is unclear whether the good overlap between the unverified circuit and the phase fitting method is due to them both achieving a lower bound for convergence or just coincidence. Further investigation here would be a good target for future work. The addition of noise makes convergence more costly. This increase can be bounded below by removing the fraction of experiments where at least one error has occurred (as we are at best effectively removing these results). Confirming this trend would also be a good target for future work.



## 3.6 Conclusion

In this work, we presented a new method for error mitigation, based on verification of the system register in a single-control quantum phase estimation routine. We further extended this method to a scheme for verification of control-free quantum phase estimation. By writing a complex Hamiltonian as a sum of fast-forwardable parts and using this technique to estimate the expectation value of each part, this becomes a powerful error mitigation tool for near-term experiments such as variational algorithms. Errors that take the system away from the small verified subspace do not affect the mitigated QPE results (at the cost of requiring additional repetitions of the circuit). We performed numerical studies of this error mitigation capability of the verification protocol on three different systems, finding the suppression of all single depolarizing errors when a Givens rotation circuit or a fermionic swap network prepare random states of a small fermionic system. The suppression is further magnified in the presence of amplitude and phase damping, resulting in a gain of up to four orders of magnitude in accuracy. For a simulation of the transverse-field Ising model the error suppression is less pronounced. However, we find that variational optimization improves the error mitigation to a gain in accuracy of about 50-fold. We further demonstrated that the combination of variational optimization and verification mitigates against constant control error (which is not naturally mitigated by the verification itself). However, we found that the choice of post-processing technique in the classical post-processing may affect the estimation error by a factor of 10

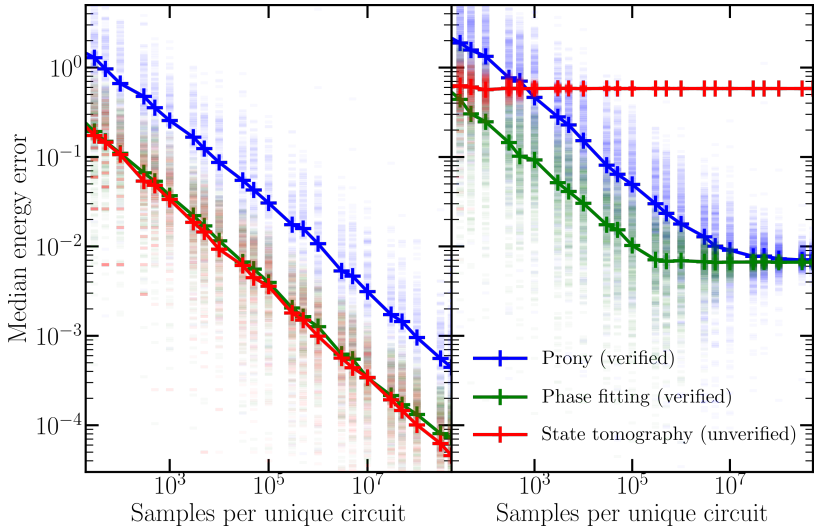


Figure 3.7: Convergence of the estimation of a single point in a 4-site transverse-field Ising model with the number of samples taken, using verified phase estimation processed either with Prony's method (blue) or by fitting known phases to the phase function (green), or standard partial state tomography (red) on individual Pauli terms. (Left) convergence in the absence of error. (Right) convergence in the presence of 1% depolarizing error per qubit per moment. In each subfigure we plot the median energy error (crosses and lines) over 200 simulations, which are plotted themselves behind (faint dashes).

in the presence of sampling noise.

Though verified phase estimation as presented already appears to be one of the most powerful error mitigation techniques available to NISQ-era quantum computing, further avenues for optimization exist. The wide range of possible options for verification, how to divide the Hamiltonian, and classical post-processing method all provide metaparameters which we have not yet determined how to optimize for any specific problem. Furthermore, circuits which quickly scramble errors would appear to make verification more reliable. Whether this observation can be used for meaningful optimization is a clear target for future work. Similarly, as errors need to have the instantaneous state as a near-eigenstate to not fail verification, the errors that verified phase estimation is most-susceptible to

must commute, and could potentially be corrected with a classical error correcting code. As these codes require much less overhead than full-blown QEC, this may be a practical method to ensure universal suppression of single-qubit errors. Future work could also investigate whether verified phase estimation may be combined efficiently with other error mitigation techniques. More generally, it would be timely to benchmark the zoo of error mitigation techniques against one another, and determine which combination of techniques works best in a range of situations.

## 3.A Error analysis



Let us formalize the ideas outlined in Sec. III.C of the main text by considering how the verified and unverified Hilbert spaces $\mathcal{H}^{(\mathrm{v})}$ and $\mathcal{H}^{(\mathrm{f})}$, and the verified and unverified ensembles $\rho^{(\mathrm{v})}$ and $\rho^{(\mathrm{f})}$ within them, evolve over the course of a noisy quantum circuit. (We remind the reader here that $\rho^{(\mathrm{v})}$ and $\rho^{(\mathrm{f})}$ are not normalized, hence our use of the word 'ensemble' rather than 'state'.) We will then attempt to provide some mechanisms for the observed scaling laws in Sec. V of the main text. At the end of the VPE circuit, the verified Hilbert space $\mathcal{H}^{(\mathrm{f})}$ is spanned by the two verified basis states. In single-control VPE, these are $|0\rangle|0\rangle$ and $|1\rangle|0\rangle$, while in control-free VPE these are $|0\rangle$ and $|\vec{1}_{\mathrm{T}}\rangle$. Let us label these $|0_{\mathrm{v}}\rangle$ and $|1_{\mathrm{v}}\rangle$ respectively, and then we may define the verified Hilbert space as

$$\mathcal{H}^{(\mathrm{v})} = \mathrm{Span}\{|0_{\mathrm{v}}\rangle, |1_{\mathrm{v}}\rangle\}, \tag{3.77}$$

and the verified ensemble as

$$\rho^{(\mathrm{v})}(t) = P_{\mathrm{v}}\rho(t)P_{\mathrm{v}}, \quad P_{\mathrm{v}} = |0_{\mathrm{v}}\rangle\langle 0_{\mathrm{v}}| + |1_{\mathrm{v}}\rangle\langle 1_{\mathrm{v}}|. \tag{3.78}$$

The system state $\rho$ here is the state at the end of the VPE circuit, let us now consider how the system evolves to get here. This evolution is not a function of the simulated time $t$, as we may use entirely different circuits to estimate the phase function $g(t)$ and $g(t')$. Instead, we must frame the evolution of the state on the quantum device over the course of the VPE circuit in terms of the device time $\tau$. That is, let us fix $t$, and assume that the circuit that implements $U = e^{iHt}$ is split into a set of discrete moments $U(\tau)$ (with the last moment occuring at time $\tau_{\max}$),

$$U = \prod_{\tau=0}^{\tau_{\max}} U(\tau), \tag{3.79}$$

where each moment consists of a set of gates acting in parallel

$$U(\tau) = \prod_i U(\tau)_i. \tag{3.80}$$

This is how circuits are represented in the cirq quantum programming framework [102], and is a good way of approximating the behaviour of a real quantum circuit.

To best understand how noise and verification work together, we must move to the interaction picture, or rather a rotating reference frame. In the Schr odinger picture, the system begins entirely within $\mathcal{H}^{(\mathrm{v})}$, as in all cases it is initialized in $|0_\mathrm{v}\rangle = |0\rangle$ and immediately rotated to $\frac{1}{\sqrt{2}}(|0_\mathrm{v}\rangle + |1_\mathrm{v}\rangle)$. It then evolves out of $\mathcal{H}^{(\mathrm{v})}$ as we prepare, evolve, and un-prepare the system, even in the absence of error. However, for us it is more helpful to consider the states that will be rotated into $\mathcal{H}^{(\mathrm{v})}$ at the end of the circuit. This may be achieved by re-defining the verified basis states in the reference frame

$$|0_\mathrm{v}\rangle \to \left(\prod_{\tau'>\tau} U(\tau')\right)^{-1} |0_\mathrm{v}\rangle \tag{3.81}$$

$$|1_\mathrm{v}\rangle \to \left(\prod_{\tau'>\tau} U(\tau')\right)^{-1} |1_\mathrm{v}\rangle. \tag{3.82}$$

(This a slightly non-standard choice of reference frame, as we are shifting backwards in time from the final state, rather than forwards in time from the initial state, but it makes our error analysis far easier.) In the absence of error, this is the Heisenberg picture: our system remains in the state

$$\rho = |\rho\rangle\langle\rho| \tag{3.83}$$

$$|\rho\rangle = \frac{1}{\sqrt{2}}|0_\mathrm{v}\rangle + \frac{g(t)}{\sqrt{2}}|1_\mathrm{v}\rangle + \sqrt{\frac{1-|g(t)|^2}{2}}|\rho^{(\mathrm{f})}\rangle, \tag{3.84}$$

throughout the entire circuit. Here $|\rho^{(\mathrm{f})}\rangle$ is the fraction of the state that will eventually fail verification —

$$\rho^{(\mathrm{f})} = |\rho^{(\mathrm{f})}\rangle\langle\rho^{(\mathrm{f})}|. \tag{3.85}$$

(Recall here that $|\rho^{(f)}\rangle \neq 0$ even in the absence of noise, but this fraction of the state does not contribute to the phase function — $\langle\rho^{(f)}|X_c + iY_c|\rho^{(f)}\rangle =$

0.) We may project our system at any device time $\tau$ into the verified Hilbert space via Eq. 3.78, but with the basis states in their rotating reference frame (Eq. 3.81 and Eq. 3.82).

Noise may be added to the above by treating it as a perturbation and switching to the interaction picture. Without loss of generality, we may say the effect of noise is to shift the unitary $U(\tau)$ at each moment

$$U(\tau) \to R_{U(\tau)}(1 - p_\tau + p_\tau E_\tau), \tag{3.86}$$

where $R_U$ is the process map associated with a unitary $U$

$$R_U[\rho] = U^\dagger \rho U, \tag{3.87}$$

$E_\tau$ is the process map associated with errors during the moment $\tau$, and $p_\tau$ is the probability of any such errors occurring. In the interaction picture, the action of the circuit has been shifted into our basis states, and $U(\tau) = 1$. So, we may write our final state in the presence of error as

$$\begin{aligned}\rho^{(\mathrm{err})} =& \frac{1}{\mathcal{N}} \Bigg\{ \rho + \sum_{\tau_0 < \tau} p'_{\tau_0} E_{\tau_0}[\rho] \\ &+ \sum_{\tau_1 < \tau_0 < \tau} p'_{\tau_0} p'_{\tau_1} E_{\tau_0}\left[E_{\tau_1}[\rho]\right] + \dots \Bigg\},\end{aligned} \tag{3.88}$$

where $\mathcal{N}$ is a normalization coefficient

$$\mathcal{N} = \prod_\tau \frac{1}{1 - p_\tau}, \tag{3.89}$$

and $p'_\tau$ are the rescaled probabilities

$$p'_\tau = \frac{p_\tau}{1 - p_\tau}. \tag{3.90}$$

If desired, one can recognise this also as a discrete form of the well-known time-ordered integrals in quantum mechanics — a time-ordered sum

$$\rho^{(\mathrm{err})} = \frac{1}{\mathcal{N}} \mathcal{T} \exp\left( \sum_\tau p'_\tau E_\tau \right)[\rho], \tag{3.91}$$

where $\mathcal{T}$ is the time-ordering operator. Our projection onto the verified subspace is linear, so we may consider it on each of the individual terms

in the sum. Assuming $p'(\tau)$ is small for all $\tau$, the first-order corrections to $\rho^{(v)}$ occur from errors $E_\tau$ during a single timestep. These corrections take the form

$$P_{\mathrm{v}} p'_\tau E_\tau[\rho] P_{\mathrm{v}} = p'_\tau \begin{pmatrix} p_{0,\tau} & \frac{1}{2} g_\tau^{(\mathrm{err})}(t) \\ \frac{1}{2} g_\tau^{(\mathrm{err})\dagger}(t) & p_{1,\tau} \end{pmatrix}, \tag{3.92}$$

where

$$p_{0,\tau} = \langle 0_{\mathrm{v}} | E_\tau[\rho] | 0_{\mathrm{v}} \rangle \tag{3.93}$$

$$p_{1,\tau} = \langle 1_{\mathrm{v}} | E_\tau[\rho] | 1_{\mathrm{v}} \rangle \tag{3.94}$$

$$g_\tau^{(\mathrm{err})}(t) = \langle 1_{\mathrm{v}} | E_\tau[\rho] | 0_{\mathrm{v}} \rangle \tag{3.95}$$

The off-diagonal element in this matrix gives the contribution to the phase function $g(t)$

$$g(t) \to \frac{1}{\mathcal{N}} g(t) + \frac{1}{\mathcal{N}} \sum_\tau p'_\tau g_\tau^{(\mathrm{err})}(t). \tag{3.96}$$

One may generalize this to higher-order terms. For example, the second-order contribution to the error takes the form

$$\frac{1}{\mathcal{N}} \sum_{\tau_0 < \tau_1} p'_{\tau_0} p'_{\tau_1} \langle 1_{\mathrm{v}} | E_{\tau_1}[E_{\tau_0}[\rho]] | 0_{\mathrm{v}} \rangle \tag{3.97}$$

The mitigation power from verification requires two conditions: that the dependence of the normalization $\mathcal{N}$ on the simulated time $t$ is simple, and that the off-diagonal error contributions (Eq. 3.95) are small. We expect both conditions to often be the case. The positivity of $E_\tau[\rho]$ implies that

$$g_\tau^{(\mathrm{err})} + p_{0,\tau} + \sum_{n=1}^{2^N-2} \langle \rho_n^{(\mathrm{f})} | E_\tau[\rho] | 0_{\mathrm{v}} \rangle \leq 1, \tag{3.98}$$

where $|\rho^{(\mathrm{f})}\rangle$ is an appropriately chosen basis for $\mathcal{H}^{(\mathrm{f})}$. On average all terms are equally-weighted, so $g_\tau^{(\mathrm{err})} \sim 2^{-N}$. As such, negligible $g_\tau^{(\mathrm{err})}$ should be the norm rather than the exception; we need reason to expect that the error channel $E_\tau$ will not scatter us out of the verified subspace. If $g_\tau^{(\mathrm{err})} = 0$, the effect of $E_\tau$ on $g(t)$ occurs via the damping by $\mathcal{N}$, which itself may depend on $t$. However, $\mathcal{N}$ depends only on the rate at which errors occur, and is oblivious to their specific form. If a Hamiltonian is fast-forwardable, $e^{iHt}$ may be implemented in time constant in $t$. Then assuming a constant error rate per moment of the circuit, our phase function is dampened by a

constant amount,

$$g^{\mathrm{err}}(t) = \frac{1}{\mathcal{N}} g(t), \tag{3.99}$$

which may be corrected for by renormalization (Eq.30 of the main text). If a Hamiltonian is not fast-fowardable, $e^{iHt}$ must take real time $\tau_{\max}(t) = \mathcal{O}(t)$ to simulate to constant error. Assuming this is the case, and that we have a constant error rate per moment of the circuit, the damping from each possible error $E_\tau$ is multiplicative, and our estimation takes the form

$$g^{\mathrm{err}}(t) = e^{-t/T_1} g(t). \tag{3.100}$$

Here, $T_1$ is defined as the (simulated) time $t$ over which enough errors $E$ have accumulated that

$$\frac{1}{\mathcal{N}(\tau)}(E) = e^{-1}. \tag{3.101}$$

This constant damping may be considered an imaginary shift to the eigenvalues $E_j$; $E_j \to E_j + \frac{1}{T_1}$. It may be removed by classical postprocessing techniques [124, 130, 137]. However, the shrinking of the signal increases the sampling requirements to estimate $g(t)$ exponentially in $t$.

Although random error channels are exponentially suppressed by verification (following Eq. 3.98), realistic error models are biased, and may apply undesired phases to $g_\tau^{\mathrm{err}}(t)$ instead of setting it to 0. The density matrix in Eq. 3.92 is not normalized, but it must be positive, which implies

$$|g_\tau^{(\mathrm{err})}|^2 < \frac{p_{0,\tau} p_{1,\tau}}{p_{0,\tau} + p_{1,\tau}}. \tag{3.102}$$

This means that errors must either fail to scatter *both* $|0_{\mathrm{v}}\rangle$ and $|1_{\mathrm{v}}\rangle$, or rotate between these states and the failed state $|\rho^{(f)}\rangle$. When control-free methods are used, $|0_{\mathrm{v}}\rangle$ is separated from $|1_{\mathrm{v}}\rangle$ and $|\rho^{(f)}\rangle$ by highly non-local excitations, which are non-physical error channels. However, when single-control methods are used, $|0_{\mathrm{v}}\rangle$ is coupled to $|1_{\mathrm{v}}\rangle$ and $|\rho^{(f)}\rangle$ by control qubit errors. These control qubit errors deform the Bloch sphere defined by $|0_{\mathrm{v}}\rangle = |0\rangle|0\rangle$ and

$$g(t)|1_{\mathrm{v}}\rangle + \sqrt{1 - |g(t)|^2}|\rho^{(f)}\rangle = |0\rangle|\vec{1}\rangle. \tag{3.103}$$

When this deformation is asymmetric around the z-axis, or a rotation, $g(t)$ may be quickly corrupted[4]. However, symmetric noise (such as a

[4] We have observed this for instance due to $T_1$ decay on the control qubit between final rotation and readout. However, we can correct for this easily by measuring in

depolarizing channel, or $T_1$ or $T_2$ channels during the bulk of the circuit) can be seen to simply dampen $g(t)$ in an identical manner to $\mathcal{N}$. That is, the dampening will depend only on the rate at which these errors occur. Such dampening will be cancelled by renormalization, as observed in Fig. 3.9.

Errors that do not rotate between $|0_\mathrm{v}\rangle$ and $|1_\mathrm{v}\rangle$, but still contribute non-trivially to $g_\tau^{(\mathrm{err})}(t)$ to first order must have both $|0_\mathrm{v}\rangle$ and $|1_\mathrm{v}\rangle$ as approximate eigenstates of the error channel. This suggests a reason why control-free VPE is more noise-robust to noise than single-control VPE: the starting and reference states are very different when looked at locally, which makes it less likely that a single local error will have both states as near-eigenstates. It also suggests a reason why we might expect the suppression of errors to only second-order: if the same error occurs in subsequent moments (in a local frame), and the basis states $|0_\mathrm{v}\rangle = |0_\mathrm{v}(\tau)\rangle$ have not evolved significantly between these moments, the second error will almost (but not completely) cancel out the first, driving the system back into the verified subspace in an uncorrectable manner. This implies that a circuit which more quickly scrambles the basis states $|0_\mathrm{v}\rangle$ and $|1_\mathrm{v}\rangle$ between moments should be less susceptible to error than one where the states evolve slowly. Understanding the dynamics of these noisy circuits in more detail is a clear target for future work.

## 3.B Effect of parallelizing QPE

In this appendix we investigate the phase function obtained during the parallel estimation of multiple commuting Hamiltonians, and demonstrate that the resulting expectation values from this estimation are not affected by the parallelization and verification process combined. Let us consider the case where we have two commuting Hamiltonians $H_0$, $H_1$. In this case, we may write a simultaneous eigenbasis $|E_j\rangle$ for both Hamiltonians — $H_b|E_j\rangle = E_j^{(b)}|E_j\rangle$. Let $|\psi_\mathrm{s}\rangle = \sum_j a_j|E_j\rangle$, and we may calculate the controlled-time-evolved global state $|\Psi(t)\rangle$ to be

$$\sum_j a_j(|0\rangle + e^{iE_j^{(0)}t}|1\rangle)(|0\rangle + e^{iE_j^{(1)}t}|1\rangle)|E_j\rangle. \tag{3.104}$$

the opposite basis 50% of the time

Tracing out control qubit 1 obtains the following reduced density matrix for the system + control qubit 0

$$\sum_{j,j'} a_j a_{j'}^* \left[1 + e^{i\left(E_j^{(1)} - E_{j'}^{(1)}\right)}\right] \times \left[\left(|0\rangle + e^{iE_j^{(0)}t}|1\rangle\right)|E_j\rangle\langle E_{j'}|\left(\langle 0| + e^{-iE_{j'}^{(0)}t}\langle 1\right)|\right]. \tag{3.105}$$

The issue here then comes from this additional factor $\left[1 + e^{i\left(E_j^{(1)} - E_{j'}^{(1)}\right)}\right]$ at the front. Note that (as we should expect) this goes away upon tracing out the system register, as the trace over $|E_j\rangle\langle E_j'|$ yields (dropping all additional terms in the above expression).

$$\sum_l \langle E_l|E_j\rangle\langle E_{j'}|E_l\rangle = \delta_{j,j'} \tag{3.106}$$

However, post-selection implies that we take the expectation value with regards to $|\psi_\mathrm{s}\rangle$, obtaining

$$\langle\psi_\mathrm{s}|E_j\rangle\langle E_{j'}|\psi_\mathrm{s}\rangle = a_j^* a_{j'}. \tag{3.107}$$

The off-diagonal element of the control qubit density matrix can then be found to be

$$\frac{1}{2}\sum_{j,j'} |a_j|^2|a_{j'}|^2\left(1 + e^{i(E_j^{(1)} - E_{j'}^{(1)})t}\right)e^{iE_j^{(0)}t} \tag{3.108}$$

$$= \frac{1}{2}\sum_j |a_j|^2 e^{iE_j^{(0)}t} + \frac{1}{2}\sum_j |a_j|^2 e^{i(E_j^{(1)} + E_j^{(0)})t}\sum_{j'} |a_{j'}|^2 e^{-iE_{j'}^{(1)}t}. \tag{3.109}$$

One can see that this is a linear combination of products of the phase functions of $H_0$, $H_1$, and $H_0 + H_1$. In theory the eigenvalues $E_j^{(0)}$ and amplitudes squared $|a_j|^2$ are still present in this function, and could be extracted via classical postprocessing. However, the $\frac{1}{2}$ coefficient implies we need 4 times as many single-shot experiments for the estimation of $|a_j|^2$ to the same error (compared to a standard QPE experiment for $H_0$). Extending this to $L > 2$ summands, the off-diagonal for the $s$th control

qubit can be written:

$$\frac{1}{2^L}\sum_{j,j'}|a_j|^2|a_{j'}|^2 e^{iE_j^{(s)}t}\prod_{s'\neq s}\left(1+e^{i(E_j^{(s')}-E_{j'}^{(s')})t}\right), \tag{3.110}$$

and we see that the signal corresponding to 'just' $g(t)$ is exponentially small. However, all is not lost. Inspecting the form of Eq. 3.110, we see that we may expand this as a sum of $2^L J^2$ separate (possibly degenerate) spurious energies $F^{(s)}_{v,j,j'}$, indexed by a $L$-bit binary integer $v$ and the original $j$ and $j'$ indices

$$F^{(s)}_{v,j,j'} = E_j^{(s)} + \sum_{s'\neq s} v_{s'}(E_j^{(s')} - E_{j'}^{(s')}), \tag{3.111}$$

with corresponding ($v$-independent) spurious amplitudes

$$B_{j,j'} = \frac{1}{2^L}|a_j|^2|a_{j'}|^2. \tag{3.112}$$

(Note that as stated these energies are automatically at least doubly-degenerate as $v_s$ does not appear in the equation for $F^{(s)}_{v,j,j'}$.) If we then calculate the weighted average of the $F^{(s)}_{v,j,j'}$ (which is what we would do if we processed the signal as if the parallelization had not occurred), we find

$$\begin{aligned}\sum_{v,j,j'} B_{j,j'}F^{(s)}_{v,j,j'} =& \frac{1}{2^L}\sum_{v,j,j'}|a_j|^2|a_{j'}|^2 E_j^{(s)} \\ &+ \frac{1}{2^L}\sum_{v,j,j'} B_{j,j'}\sum_{s'\neq s} v_{s'}E_j^{(s')} \\ &- \frac{1}{2^L}\sum_{v,j,j'} B_{j,j'}\sum_{s'\neq s} v_{s'}E_{j'}^{(s')}.\end{aligned} \tag{3.113}$$

As $j$ and $j'$ are just dummy indices, and as $B_{j,j'} = B_{j',j}$, the last two terms cancel, and as $\sum_v = 2^L$ and $\sum_{j'}|a_{j'}|^2 = 1$, we have

$$\sum_{v,j,j'} B_{j,j'}F^{(s)}_{v,j,j'} = \sum_j |a_j|^2 E_j^{(s)} = \langle H_s\rangle. \tag{3.114}$$

This implies that expectation values may be extracted via parallel verified phase estimation, even though the signal itself may be significantly more complex. For the case of Pauli $H_s$ operators, the spectrum $F^{(s)}_{v,j,j'}$ is highly

degenerate — it is the set of odd integers $\{-2L+1, -2L+3, \ldots, 2L-3, 2L-1\}$. (This parallels the spectrum of a spin-$\frac{2L-1}{2}$ operator, which one might not expect following Hund's rules for the combination of spin-$\frac{1}{2}$ systems, which is curious.) This must be taken into account when signal processing by amplitude-fitting, as one would otherwise miss components of the energy. However, the overhead for this is only linear in the number of simultaneously-estimated terms.

## 3.C Compensation for spurious eigenvalues due to sampling noise



When quantum phase estimation is used to estimate eigenvalues as well as amplitudes to sum together to give an expectation value (Eq. 53 of the main text), finite sampling noise introduces a small bias to this estimation that may be cancelled. This bias does not come from the QPE itself. The sampling noise has a white spectrum which is invariant under a Fourier transform, so classical post-processing of a noisy spectrum yields a set of spurious eigenvalue/amplitude pairs evenly distributed around the circle. However, in order to evaluate Eq. 53 of the main text, we have to make a branch cut in this circle. The resulting terms then average to bias the signal by a term $\Delta_{\text{bias}} = \langle H \rangle - \overline{\langle H \rangle}$ towards the center of the resulting region. (Here, $\langle H \rangle$ is the true expectation value, and $\overline{\langle H \rangle}$ that measured naively.) For example, if we assume all eigenvalues $E_j \in [-\pi, \pi]$, this biases the signal towards zero. This bias is dependent on both the number of steps $K$, and the number of samples $M$ used in the QPE process. Numerically, we find (Fig. 3.8):

$$\Delta_{\text{bias}} = -\langle H \rangle \times (K-2)^{\frac{1}{2}} M^{-\frac{1}{2}}. \tag{3.115}$$

Inverting this obtains

$$\langle H \rangle = \overline{\langle H \rangle} \left[ 1 + (K-2)^{\frac{1}{2}} M^{-\frac{1}{2}} \right]^{-1}, \tag{3.116}$$

which is used in the estimation in Sec. V.D.

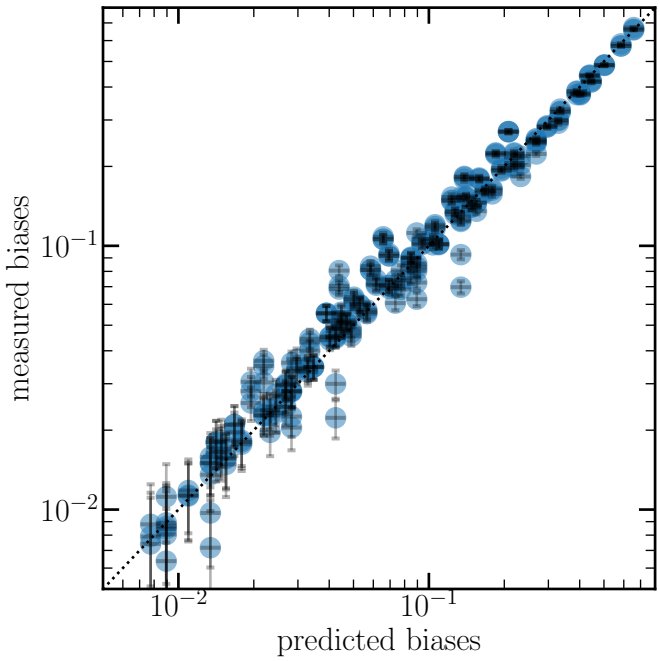


Figure 3.8: Predicted (Eq. 3.115) vs found bias from estimating expectation values using Prony's method.

## 3.D Demonstration of immunity to control noise in single-control VPE

One might expect that the discrepancy between the scaling of the error mitigation power of the control-free and single-control circuits seen throughout this work comes from accumulation of errors on the control qubit alone. In this appendix, we show that this is not the case. In Fig. 3.9, we see that removing all errors on the control qubit does little to reduce the total error rate (black crosses), whilst a model with noise only on the control qubit achieves an error limited by our use of Prony's method for post-processing. In App. 3.A we argue that the increased error suppression from control-free VPE comes from the large separation between reference and starting states. Errors will be removed by verification unless they maintain coherence between these states, which these error models fail to do.



## 3.E Use of a variational outer loop to mitigate constant unitary noise

One of the main uses of expectation values $\langle H \rangle$ in quantum experiments is to use them as a cost function in a variational outer loop. Optimizing the parameters $\vec{\theta}$ in a preparation unitary $U_{\mathrm{p}}(\vec{\theta})$ to minimize the expectation value of the prepared state $|\Psi(\vec{\theta})\rangle = U_{\mathrm{p}}(\vec{\theta})|0\rangle$ then gives an approximation for the true ground state of $H$. The variational optimization process is itself known to be robust against certain types of error [82, 144], in particular control errors. These occur when a signal meant to implement a gate $G(\theta)$ either drifts or is distorted and instead implements $G(\theta')$. As this error is often repeated throughout an experiment, i.e. every instance of $G$ is miscalibrated by a similar amount, it will be repeated throughout the experiment. Verification can only correct single errors, and as such is not targeted for this type of noise. By contrast, the dominant source of errors in a VQE are often the incoherent errors that verification is designed to target. As such, verification and variational optimization provide cumulative mitigation by targeting sources of error the other lets through.

To demonstrate the combined mitigation effects, we use verified control-free phase estimation of a Givens rotation circuit in the inner loop of a variational quantum eigensolver. In order to prevent oversimplifying the problem, we add a next-nearest-neighbour coupling and on-site potential

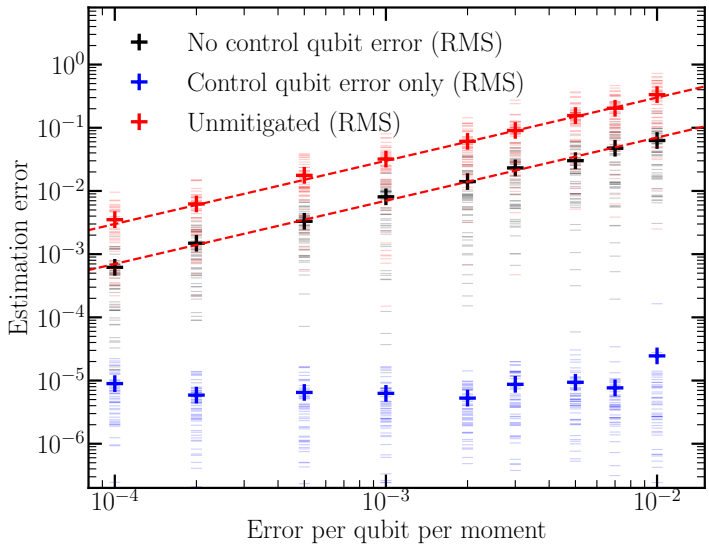


Figure 3.9: Mitigation of the same 4-qubit VHA circuit as in Fig. 4 of the main text, but with either depolarizing noise on only the system register (black) or only on the control qubit (blue). This is compared with the error in estimation using partial state tomography instead of VPE (red). For each dataset, the RMS error (crosses) is plotted over 50 different estimations for each error rate (with randomly-chosen ansatz parameters), and individual data points are plotted as dashes behind. For reference, dashed lines showing linear (red) dependence on the gate error rate are plotted.

to the Hamiltonian in Eq. 66 of the main text, yielding

$$H = H_1 + H_2 \tag{3.117}$$

$$H_1 = -t_1 \sum_{j=1}^{N} c_j^\dagger c_{j+1} + \text{h.c.} \tag{3.118}$$

$$H_2 = -t_2 \left( \sum_{j=1}^{N} c_j^\dagger c_{j+1} + \text{h.c.} + \sum_{j=1}^{N} c_j^\dagger c_j \right), \tag{3.119}$$

and estimate expectation values for $H_1$ and $H_2$ separately. Here, we again

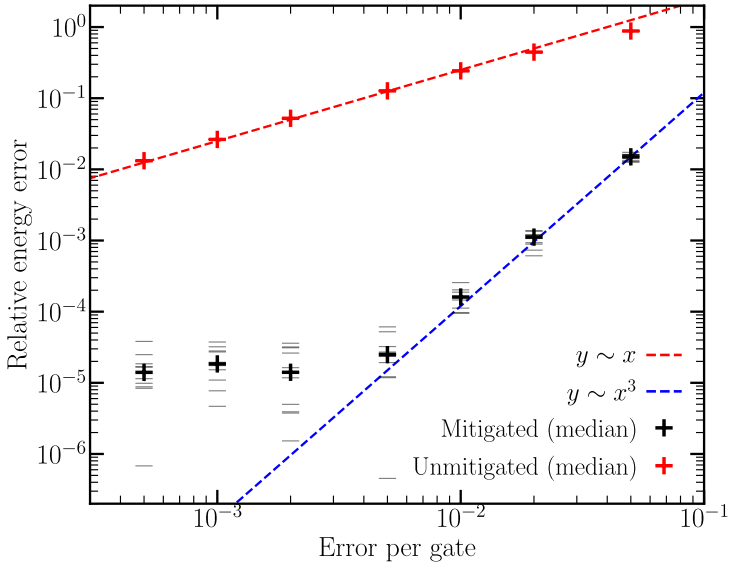


Figure 3.10: Error in estimating the ground state energy of a free-fermion system (Eq. 66 of the main text) of 4 fermions (on four qubits), using control-free verified phase estimation and a VQE. Noise model is a mixture of amplitude and phase damping and constant two-qubit control error (details in text). Median absolute errors for both verified estimation (black crosses) and standard partial state tomography (red crosses) are calculated over 10 different optimization attempts. Individual simulations are plotted behind (faint dashes) Each optimization started from a different parameter set and had different control rates set. Linear (red dashed) and cubic (blue dashed) lines are shown as guides.

take periodic boundary conditions for a $N = 4$-site system (i.e. all sums in indices are taken modulo 4), and fix $t_1 = 1$, $t_2 = 0.5$. This ensures that the ground state of the system is neither a ground state of $H_1$ or $H_2$ (in which case the compiled variational ansatz and basis rotation would cancel to become an identity circuit). For a simple model combining control error and incoherent noise, we fix $p$, draw a random offset $x_i \in [-\frac{p}{\pi}, \frac{p}{\pi}]$ for each two-qubit ISWAP gate, and decompose the variational circuit into $\text{ISWAP}^{1/2}$ gates. (Though not terribly well-known, the $\text{ISWAP}^{1/2}$ gate is a good native hardware gate for superconducting qubits, and decompositions of other gates into $\text{ISWAP}^{1/2}$ gates are known [3].) Then, throughout

the circuit, we implement $\mathrm{ISWAP}^{1/2+x_i}$ gates in place of $\mathrm{ISWAP}^{1/2}$ gate. We additionally add amplitude and phase damping noise at a rate $\frac{p}{2}$. In Fig. 3.10, we plot the result following optimization via the COBYLA algorithm implemented in scipy [152], in the absence of sampling noise. We see that the verification circuit is insensitive to the incoherent noise as expected, and behaves similarly to the effect of amplitude and phase damping alone (Fig. 3(right) of the main text).

## 3.F Term-wise comparison of VPE performance



To attempt to further understand the ability of VPE to mitigate errors, in this appendix we consider the effect of estimating different types of terms on the same preparation circuit. We consider the fermionic swap network used in Sec.V.C of the main text to prepare states for a $H_2$ Hamiltonian. When this was split into number-conserving Pauli operator sums (Fig.6(2a-2b) of the main text), different circuits had to be used to estimate individual terms. In Fig. 3.11, we show the result of estimating the expectation values of two of the individual terms used in the control-free Pauli operator decomposition under an amplitude-damping noise model (Fig.6(2b) of the main text). (Recall that this figure demonstrated first-order sensitivity to this error model, whilst the low-rank factorization demonstrated a third-order sensitivity to the same model.) We see that the $H_s = Z_0 Z_1$ term (left plot) shows the cubic dependence on error rate observed in previous amplitude-damping experiments, whilst the two-body scattering term (right plot)

$$H_s = X_0 Y_1 Y_2 X_3 + Y_0 X_1 X_2 Y_3 - X_0 X_1 Y_2 Y_3 - Y_0 Y_1 X_2 X_3, \tag{3.120}$$

does not. This two-body scattering term is the only term contributing to the first-order decay of the VPE estimation observed in Fig.6(b2) of the main text — all other terms in the decomposition display similar decay to Fig. 3.11(left). This indicates that the errors to which we are first-order sensitive occur during the circuit implementation of $e^{iH_s t}$, and not the state preparation. The circuit implementing $e^{iH_s t}$ for the two-body scattering term is the only such circuit that does not conserve number throughout. (Instead, this evolution is achieved in two steps: a basis transform of $XY, YX \to IZ, ZI$ on pairs of qubits, $ZZ$ rotations between the pairs and uncomputing, and then a basis transform of $XX, YY \to IZ, ZI$ on pairs of qubits, $ZZ$ rotations between the pairs, and uncomputing again.)

Finding decompositions of these circuits more amenable to VPE is a clear target for future work.

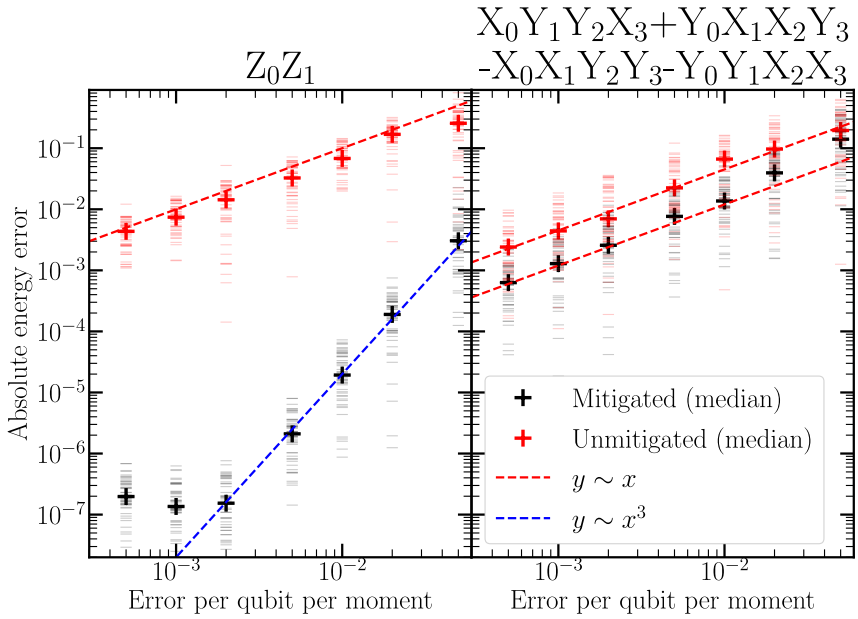


Figure 3.11: Expectation value estimation of two individual $H_s$ terms from the control-free number-conserving Pauli operator decomposition of the $H_2$ Hamiltonian studied in Fig.6 of the main text on states prepared by a fermionic swap network. The two terms here comprise part of the sum (Eq.6 of the main text) for the expectation value of Fig.6(2b) of the main text — but are studied here without pre-factors (i.e. $\|H_s\| = 1$). Each figure is labeled with the studied term, and guide-lines (dashed red and blue) are given to show observed scaling laws. Data presented is the median (crosses) over 50 individual data points (faint dashes) of the absolute error in estimation using VPE (black) and standard partial state tomography (red).

## 3.G Comparison to symmetry verification

In this section we present a comparison of verified phase estimation and symmetry verification on a depolarizing noise model, using the experiment

in Fig.6(b1) of the main text. In order to improve performance, we choose to verify on the number operator $\sum_i Z_i$, instead of the parity $\prod_i Z_i$. To perform symmetry verification we take the quantum state prepared by the circuit, and directly project this into the number-conserving space. (In a real experiment simultaneous readout of the number operator and all terms is possible [103], but requires a slight addition of circuitry, which would increase the final error slightly.) In Fig. 3.12, we observe that while symmetry verification reduces error by around an order of magnitude, it does not provide the same asymptotic improvement as VPE. We also note that VPE improves over symmetry verification at all error rates, despite having circuit over 3 times as deep. This is to be expected; as phase ($Z_i$) errors commute with the number operator, these cannot be detected by symmetry verification and so contribute at first order to the final error rate.



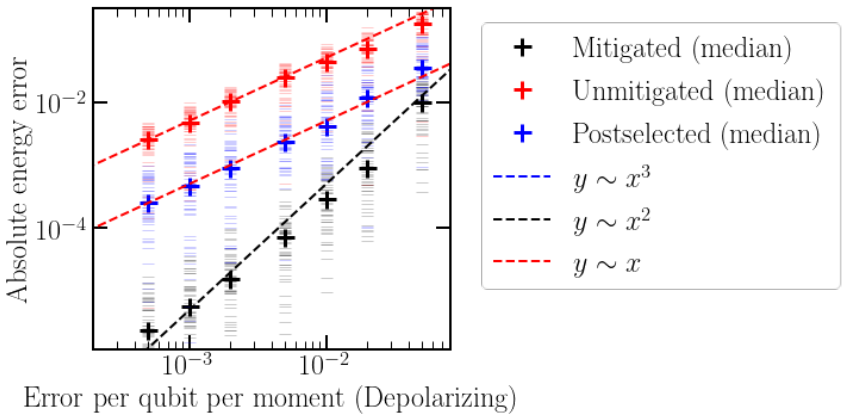


Figure 3.12: Comparison between verified phase estimation and symmetry verification. Both techniques are compared on the estimation of the expectation value of the electronic structure Hamiltonian for $H_2$ under a depolarizing noise model. Verified and unverified results are from the same simulated experiment as Fig.6(b1) of the main text. Although symmetry verification improves the energy error by around a factor 10, it still exhibits first-order scaling, as it cannot correct for phase (Z) errors during the experiment.