Quantum Parameter Estimation for Early Fault-Tolerant Quantum Simulation

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. S. de Rijcke
volgens besluit van het college voor promoties
te verdedigen op vrijdag 22 mei 2026
klokke 11:30 uur

door

Alicja Dutkiewicz

geboren te Kraków, Polen
in 1997

Promotor:	Prof. Dr. C. W. J. Beenakker
Co-promotor:	Dr. T. E. O'Brien
Promotiecommissie:	Prof. Dr. S.J. van der Molen
	Prof. Dr. V. Dunjko
	Dr. J. Helsen (CWI, Amsterdam)
	Dr. S. Zhou (Perimeter Institute, Waterloo)
	Prof. Dr. R. Demkowicz-Dobrzański (Uniwersytet Warszawski)

An electronic version of this thesis can be found at https://openaccess.leidenuniv.nl/XXXX

On the cover:
An abstract interpretation of a quantum circuit — by Gina Borsje.

Contents

I. Introduction

The field of quantum computing is entering an exciting new phase: the era of early fault tolerance. After years of progress in building quantum processors, we are now witnessing the first devices capable of correcting their own errors at the logical level [1–4]. While remaining modest in scale, early fault-tolerant devices open the door to practical applications.

One of the most anticipated applications of quantum computing is the simulation of quantum systems beyond the reach of classical methods, with far-reaching implications for fields such as quantum chemistry [5], materials science [6], and high energy physics [7]. A complete simulation workflow consists of two essential steps: encoding the system of interest onto a quantum device, and extracting relevant information from that encoding. This thesis focuses on quantum parameter estimation (e.g. quantum phase estimation) as a key tool for extracting information in quantum simulation, with particular emphasis on algorithm design tailored to the constraints and opportunities of early fault-tolerant devices.

The rest of this chapter is structured as follows: Section I.1 introduces the foundations of quantum computing, including an overview of different hardware generations and corresponding algorithm design paradigms. Section I.2 discusses quantum simulation and how parameter estimation fits within this broader task. Section I.3 reviews the basic concepts of quantum parametric estimation theory. Section I.4 provides an overview of quantum phase estimation algorithms, which form the basis for the new techniques developed in this thesis. Finally, section I.5 summarizes the contributions of this thesis.

I.1. Quantum computation

A quantum computer is a device that processes information according to the laws of quantum mechanics. While a classical computer stores information using bits, which can take values 0 or 1, a quantum computer uses qubits — two-level quantum systems whose states are described as complex linear combinations of the basis states $|0\rangle$ and $|1\rangle$. A classical register of n bits can represent exactly one of the possible 2^n bitstrings at any given time. In contrast, a quantum register can exist in a probabilistic superposition of all

possible bitstrings (called *computational basis states*), represented as a vector in a 2^n-dimensional Hilbert space,

$$|\psi\rangle = \alpha_0 |0\ldots00\rangle + \alpha_1 |0\ldots01\rangle + \cdots + \alpha_{2^n-1} |1\ldots11\rangle \equiv \sum_{x=0}^{2^n-1} \alpha_x |x\rangle . \quad \text{(I.1)}$$

This description encodes the probabilistic nature of quantum measurements: according to the Born rule, the coefficients α_x determine the probability $P(x) = |\alpha_x|^2$ of observing a bitstring x if the system is measured in the computational basis.

Specifying a general state of n qubits in Eq. (I.1) requires 2^n complex numbers to be specified, which makes naive classical simulation infeasible and underpins the expectation that quantum computers can outperform classical ones for specific tasks. This concept, known as *quantum advantage*, serves as a computational counterpart to Bell's inequalities, which reveal non-classical correlations in physical systems [8–10]. A famous example is Shor's factoring algorithm [11], which provides a polynomial-time solution to a problem conjectured to require exponential classical resources. However, its large-scale demonstration remains far beyond the reach of current hardware [12, 13]. The first experimental claims of computational quantum advantage [14–16] were made in the context of quantum random circuit sampling [17, 18], where quantum devices generate outputs from randomly chosen circuits that are believed to be hard to simulate classically. More recent demonstrations have instead targeted physically motivated problems, whose solutions could be verified by another quantum computer [19–22]. However, these results remain the subject of debate, as classical algorithms continue to improve [18, 23–28]. The search for practical, unambiguous demonstrations of quantum advantage continues.

Quantum computation is performed by applying unitary operations to a quantum register initialized in a fiducial state (typically $|00\ldots0\rangle$). In the quantum circuit model, these operations are built from a sequence of quantum gates, which are the quantum analogs of logical gates in classical circuits. Examples include the Pauli matrices,

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad \text{(I.2)}$$

For instance, the X gate acts as a quantum analogue of the classical NOT operation, flipping the basis states $|0\rangle$ and $|1\rangle$. Building more general circuits

relies on universal gate sets. A common choice includes the Hadamard gate

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, \tag{I.3}$$

T gate

$$T = \begin{pmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{pmatrix}, \tag{I.4}$$

and CNOT gate

$$= \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}. \tag{I.5}$$

An example of a quantum circuit is shown in Fig. I.1. The Solovay–Kitaev theorem ensures that any target unitary can be approximated to arbitrary precision using a finite sequence of gates from such a universal set [29]. The goal of quantum algorithm design is to construct quantum circuits that solve the computational tasks in the least amount of time, which typically translates to the smallest possible number of gates.

At the end of a quantum computation, information is extracted by measuring the quantum state. The outcome of measurement is fundamentally probabilistic, with the probability of each outcome dictated by the Born rule. Quantum algorithms employ different measurement strategies to extract information from these measurement outcomes, depending on the computational task. Direct computational basis measurements sample from quantum probability distributions, while expectation value estimation usually requires statistical averaging across many copies of a quantum state. This is required

Figure I.1.: Example of a simple quantum circuit preparing a Bell state. A Hadamard gate H creates a superposition on the first qubit, followed by a CNOT gate entangling the two qubits. Applying the gates from left to right yields the state

$$\text{CNOT}\ (H \otimes \mathbb{1})\ |00\rangle = \text{CNOT}\ \frac{|00\rangle + |10\rangle}{\sqrt{2}} = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

because measurement is destructive: once observed, the state irreversibly collapses into the measured state. Thus, each measurement provides only a single random sample from the underlying quantum distribution, and expectation value estimation suffers from an inherent statistical noise. The combination of statistical noise and destructive measurement means that quantum algorithms must be designed to extract maximum classical information from minimum measurement overhead, a principle that guides the algorithmic techniques developed throughout this thesis.

The basic building blocks discussed above are used to construct quantum algorithms, but what counts as an efficient algorithm varies across generations of quantum devices. In the remainder of this section, we describe the different computational regimes characterized by how the device handles noise. We start by describing how noise affects quantum computation in section I.1.1. Section I.1.2 introduces *quantum error correction* as the ultimate solution to noise, which enables *fault-tolerant quantum computation*. Since implementation of these techniques remains out of reach, present-day *NISQ* devices instead employ *error mitigation* to reduce the effects of noise through data post-processing (sec. I.1.3). Finally, section I.1.4 introduces the intermediate regime of *early fault-tolerance*, which is central to this thesis.

I.1.1. Noise in quantum computers

Real quantum computers, like all physical systems, are inevitably subject to noise and imperfections. A common way to characterize this noise is through the error rate p, which denotes the probability that a given quantum gate operation introduces a fault. Assuming each gate has an independent error probability p, the probability of correctly executing a circuit with d gates decays exponentially as $(1-p)^d$. The exponential accumulation of errors poses a fundamental threat to quantum computation, potentially eliminating any quantum advantage unless errors can be actively corrected [30, 31].

Quantum computers face a broader spectrum of errors than classical devices. In classical computers, the primary error mechanism is the bit flip, that changes a 0 to a 1 or vice versa. However, quantum computers experience not only bit-flip errors (corresponding to unwanted X operations), but also phase-flip errors (unwanted Z operations) that alter the relative phases between computational basis states. One common simplified model is *local depolarizing noise*, where each qubit independently undergoes a random X, Y, or Z error with equal probability $p/3$. Mathematically, this is equivalent to replacing each qubit with a completely random classical state (*maximally mixed state*) with probability $4p/3$ [29]. This contrasts with *global depolarizing noise*, where the entire quantum register is replaced with a completely random state, which corresponds to treating all possible errors – from single-qubit flips to complex

multi-qubit correlations — as equally likely.

Because of the broad landscape of quantum errors described above, quantum error correction (QEC) faces several fundamental challenges that have no analog in classical computing. First, QEC must protect against both bit-flip and phase-flip errors, requiring more sophisticated codes than classical systems that only handle bit flips. Second, the no-cloning theorem [29, 32] prohibits copying unknown quantum states, eliminating the straightforward redundancy strategies used in classical error correction. Third, any measurement used to diagnose errors irreversibly alters the quantum state, creating a tension between detecting errors and preserving information. Fourth, quantum states are encoded in continuous complex amplitudes rather than discrete bit strings, allowing small perturbations to accumulate gradually over time. Finally, current quantum devices exhibit error rates on the order of 10^{-3} to 10^{-2} per gate operation [1, 4, 33], orders of magnitude higher than classical error rates of approximately 10^{-17} [29]. As a result, new techniques have to be developed to achieve fault-tolerant quantum computation.

I.1.2. Quantum error correction and fault-tolerance

The basic idea of QEC is to encode a logical qubit into multiple physical qubits such that errors on a small subset of the physical system can be detected and corrected without disturbing the logical information. Partial measurements known as syndrome measurements are performed during the computation to detect the presence and type of error, without collapsing the encoded logical state. The ability of a code to detect and correct errors is quantified by its code distance d, which is the minimum number of physical errors that cause an undetectable logical error. A code of distance d can detect any $d-1$ individual errors, and correct up to $(d-1)/2$ errors [29]. Crucially, the threshold theorem implies that if the physical error rate is below a threshold value, increasing the code distance allows for exponential suppression of logical error rates [29], restoring the potential for exponential quantum advantage with imperfect hardware.

The first QEC code, introduced by Shor in 1995, used 9 physical qubits to encode a single logical qubit that is robust to arbitrary single-qubit errors with $d = 3$ [29, 34]. Subsequent work led to more efficient and general codes, with surface codes [35, 36], color codes [37], and low-density parity check (LDPC) codes [38] being leading candidates for future scalable quantum architectures [39]. Each code family involves trade-offs between factors such as the number of physical qubits per logical qubit, the threshold error rate, the complexity of syndrome measurements, and the ability to implement logical gates efficiently—considerations that are central to designing practical quantum architectures.

While QEC enables the protection and recovery of static quantum information, achieving fault-tolerant quantum computation requires an additional layer of control: logical operations must be implemented in a way that prevents small physical errors from propagating into uncorrectable logical faults. One approach is to use transversal gates, which apply gate operations independently across the physical qubits of a code block. These gates naturally suppress the spread of single-qubit errors, but the Eastin-Knill theorem shows that no quantum error-correcting code can implement a universal set of logical gates using only transversal operations [40]. In many practical codes, the Clifford group or a subgroup can be implemented transversally, but non-Clifford gates such as the T gate (Eq. (I.4)) cannot. (The Clifford group is the set of unitary rotations that map tensor products of Pauli operators I, X, Y, Z to single tensor products of Pauli operators.) Unfortunately, the Gottesman–Knill theorem implies that circuits composed solely of Clifford gates can be efficiently simulated classically [29, 41], so realizing a quantum computational advantage requires fault-tolerant access to non-Clifford resources like the T gate.

A widely used approach to access non-Clifford resources is magic state injection, where the desired non-Clifford gate is realized via gate teleportation, consuming a special resource state known as a magic state. A standard approach to producing these states is magic state distillation, in which many noisy copies are processed to distill a smaller number of high-fidelity ones [42]. In common fault-tolerant architectures, this process dominates the computational cost, with around a third of the qubits used for magic state factories [12, 43]. Recently, a new technique called magic state cultivation has been introduced [44], which challenges the assumption that T gates dominate the computational cost. While asymptotically less efficient than distillation, this method allows for implementing T gates with the cost comparable to CNOT gates for error rates as low as 10^{-9}.

Estimating the runtime of fault-tolerant quantum algorithms is challenging and heavily dependent on implementation details such as hardware architecture, physical error rates, and code design, making direct algorithm comparisons difficult. As a result, a widely used, architecture-independent proxy for algorithmic cost is the T-count, the number of T gates in the logical circuit. Despite recent advances such as cultivation, T-count remains a standard heuristic for evaluating and optimizing fault-tolerant algorithms. The typical design approach is to minimize T-count at the logical circuit level and then choose error correction parameters, such as code distance, to ensure the total probability of logical failure across the computation remains negligible, typically below 1%.

Recent experiments have marked a critical milestone: the first demonstra-

tions of QEC on physical devices. For the first time, hardware platforms have achieved physical error rates below the surface code threshold [1], entering a regime where increasing the code distance d guarantees exponential suppression of logical errors. This exponential suppression has been demonstrated for surface codes with $d = 3, 5, 7$ on superconducting qubits [1, 2] and neutral atom arrays [3], as well as for Knill's Carbon code [45] with $d = 3, 4$ on trapped-ion platforms [4]. These implementations have produced logical qubits with error rates lower than any of the underlying physical qubits. Moreover, the first computations with logical qubits have been performed, involving up to 48 logical qubits and 48 non-Clifford gates [3].

Despite steady experimental progress, large-scale fault-tolerant quantum computing remains far beyond current capabilities. Estimates suggest that practically relevant quantum algorithms will require millions of physical qubits [12, 13, 36, 43, 46], while the largest quantum processor today contains just over 1,000 qubits [47]. This motivates the search for alternative strategies that can be applied to make use of current quantum devices before full-scale fault tolerance becomes practical.

I.1.3. NISQ and error mitigation

While the long-term vision of quantum computing relies on QEC, its large scale implementation remains infeasible on today's *noisy, intermediate-scale quantum* (NISQ) devices [48]. Current hardware platforms contain on the order of 100 qubits, with typical error rates around 10^{-3}. Despite these limitations, NISQ devices have enabled proof-of-principle demonstrations of quantum advantage [14–16, 19–22] and inspired extensive research into how to extract useful computational results under noise.

The NISQ cost model is fundamentally different from the fault-tolerant regime. In the fault-tolerant regime, Clifford gates are typically considered cheap, and the dominant cost measure is the T-count. By contrast, in NISQ devices single-qubit rotations with continuous parameters are often easy to implement, while two-qubit gates are more prone to errors. Moreover, without QEC, the fidelity of a quantum circuit decays exponentially with its depth. Consequently, NISQ algorithm design focuses on minimizing qubit count and circuit depth, measured using hardware-specific gate sets. A common strategy is *circuit division*: performing many repetitions of short quantum circuits, whose measurement results are combined classically.

While QEC remains infeasible, an alternative approach to noise known as quantum error mitigation (QEM) has been developed [49, 50]. Rather than correcting or detecting individual errors, QEM techniques aim to reduce the bias in measured expectation values by modifying the circuit or measurement procedure and combining the results in post-processing. This is achieved

by running many additional, modified quantum circuits (with error rate the same or higher than original) and using classical processing to reconstruct an estimate closer to the ideal outcome. A fundamental limitation is the bias–variance tradeoff: although mitigation reduces bias, it amplifies the variance, leading to an increase in the number of samples required. The sampling overhead is generally exponential in the circuit depth and noise rate [50]. As a result, QEM provides a powerful tool for extracting more accurate results from near-term devices, but does not in itself enable scalable quantum computation.

To illustrate this approach, consider an example QEM technique - probabilistic error cancellation (PEC) [49, 51] (that is a basis of a new error mitigation technique developed in Chapter III). The central idea of PEC is to express the ideal quantum operation as a linear combination of noisy operations that the device can implement. Suppose we wish to estimate the expectation value $\langle\psi| O |\psi\rangle$ for a state $|\psi\rangle$ which is prepared by a circuit $\mathcal{U}^{\text{ideal}} = \prod_i \mathcal{U}_i$ composed of ideal gates $\mathcal{U}_i$ that cannot be implemented directly. If we can decompose each ideal gate $\mathcal{U}_i$ as a linear combination of noisy operations $\mathcal{B}_j$ that can be implemented on the device as

$$\mathcal{U}_i = \sum_j a_{i,j} \mathcal{B}_j, \tag{I.6}$$

we can decompose the target circuit as

$$\mathcal{U}^{\text{ideal}} = \sum_k \beta_k \mathcal{E}_k, \tag{I.7}$$

where $\mathcal{E}_k$ are sequences of noisy gates $\mathcal{B}_j$ that can be executed on the noisy device, and β_k can be calculated from the coefficients $a_{i,j}$ of the gate decomposition. Then, we can obtain an unbiased estimate of the target value $\langle\psi| O |\psi\rangle$ by sampling noisy circuits $\mathcal{E}_k$ with probabilities $\propto |\beta_k|$, measuring the observable O on the resulting state, and averaging the results. This estimator is unbiased given exact knowledge of the noise model, but incurs a sampling overhead of $(\sum_k |\beta_k|)^2$.

I.1.4. Early fault-tolerance

Recent experimental demonstrations of QEC mark a turning point: quantum computing is beginning to transition out of the NISQ era and into a new regime, known as *early fault tolerance*. It has also been referred to as the *megaquop era* [52], reflecting the expectation that this generation of quantum devices will be capable of executing quantum circuits involving around a

million error-corrected operations. In this regime, QEC becomes feasible, but the limited device size constrains both the number of logical qubits and the achievable logical error rates. For example, a device with 10^4 physical qubits could support approximately 400 logical qubits using a surface code of distance 6 [53], or around 40 logical qubits with distance 9 [54]. These constraints define a qualitatively new computational regime, distinct from both NISQ, which lacks QEC altogether, and full fault tolerance, where logical noise can be ignored.

Early fault tolerant cost model inherits certain design principles from the NISQ era, but adapting NISQ strategies is not straightforward. As in NISQ, algorithm design prioritizes constant-factor efficiency over asymptotic optimality, favouring methods with reduced qubit overhead and shallower circuits [55–67]. However, in contrast to NISQ, where circuit depth is measured in terms of native hardware gates, the relevant cost metric in early fault tolerance is the number of error-correction cycles. In particular, single-qubit gates with continuous parameters — often considered free in NISQ — can now incur a significant cost, as they must be applied to logical qubits encoded via QEC.

The presence of residual logical errors has motivated research on the combination of QEC and QEM [54, 68–71]. The logical error rate in early devices can be tuned by adjusting QEC parameters, giving rise to a new trade-off: investing more resources into QEC to reduce logical errors, or tolerating higher logical noise and compensating with QEM [54, 68, 72]. Moreover, QEM was originally developed to improve the accuracy of expectation value estimates, whereas fully fault-tolerant algorithms typically avoid expectation values altogether. To bridge this gap, new early fault-tolerant algorithms have been proposed that rely on expectation values [55–57, 61, 62, 73–75], making them more compatible with QEM. In parallel, QEM methods are being extended beyond expectation values to other computational tasks [68, 76, 77].

A central goal of early fault tolerance is to determine the minimal hardware requirements for quantum advantage by identifying relevant problems that are classically hard yet solvable with limited quantum resources. The most compelling quantum computing applications identified to date require 10^3-10^4 logical qubits and 10^{10} non-Clifford gates, calling for fully fault-tolerant devices with over 10^6 physical qubits and error rates below 10^{-4} [12, 13, 43, 46]. In contrast, recent proposals suggest that useful applications could be realized with just 10^2–10^3 logical qubits and 10^4 non-Clifford gates, allowing for quantum advantage on early fault-tolerant hardware with around 10^4 physical qubits and error rates 10^{-3}-10^{-4} [53, 54, 67, 71, 78]. In the next section, we turn to quantum simulation as a leading candidate for such early demonstrations.

I.2. Quantum simulation

The idea that quantum computers should be used to simulate physical systems was first proposed in the early 1980s by Yuri Manin [79], and independently by Richard Feynman [80]. They argued that simulating quantum mechanics with classical computers is inherently inefficient, due to the exponential growth of the Hilbert space with system size. But if nature is quantum mechanical, then perhaps the most natural way to simulate it is by building computers that follow the same rules — quantum computers. Forty years later, quantum simulation remains one of the most promising applications for emerging quantum computers, with applications in chemistry [5, 6], material science [6] and high-energy physics [7, 81].

Yet the original intuition that the exponential growth of Hilbert space makes classical simulation infeasible is too naive. Many physically relevant quantum systems can in fact be simulated efficiently with classical heuristics. A key challenge, then, is to precisely identify problems for which classical methods fail and where quantum computers might achieve a provable or empirical speedup. In this section, we discuss promising targets for quantum simulation (Sec. I.2.1), and how they can be encoded on quantum a computer (Secs. I.2.2 and I.2.3).

I.2.1. Targets of quantum simulation

Most physical properties of a quantum-mechanical system that we wish to estimate via quantum simulation are defined by, and accessed through, the system's Hamiltonian H. This is a Hermitian operator that describes the time dynamics of a physical system's state $|\psi\rangle$ according to Schrödinger's equation:

$$\frac{d}{dt}|\psi\rangle = -iH|\psi\rangle\,. \tag{I.8}$$

Since H acts on the Hilbert space whose dimension grows exponentially with the system size n, specifying a general Hamiltonian in principle requires exponentially many parameters. However, physically relevant Hamiltonians are local, and therefore can be efficiently described and simulated. In particular, if each term involves at most k particles, H is said to be k-local and contains $\mathcal{O}(n^k)$ terms. In geometrically local systems – for example, lattice models with nearest-neighbor interactions – the number of terms in the Hamiltonian is linear in n. This makes the Hamiltonian a good entry and exit point for a quantum computer, as it can be described by far fewer parameters than the dynamics that it generates.

The task of quantum simulation is to extract physical properties of a system given a description of its Hamiltonian H. This thesis focuses on learning the

eigenvalues of H, as they directly determine many key physical properties, and are generally hard to compute classically, since naive approaches would require diagonalizing an exponentially large matrix. The most prominent and well-studied example is the ground-state energy – the lowest eigenvalue of H – which determines molecular properties and the rates of chemical reactions. In addition, excited-state energies enable simulation of spectroscopy experiments and photochemical reactions, while energy derivatives provide access to vibrational frequencies, equilibrium geometries, nuclear dynamics and quantities such as polarizability [5]. Beyond the spectrum, other quantities of interest include properties of eigenstates [58, 82] or thermal states [83], time dynamics [84], and correlators [19, 85]. Though the techniques explored in this thesis can play a role in estimating these properties [86], our focus will primarily be on spectral estimation.

A particularly important example of a Hamiltonian whose spectral properties are of practical interest is the electronic structure Hamiltonian [5, 87], which models electrons moving in a fixed external potential. It is central to quantum chemistry, as it governs the behavior of electrons in molecules and materials. In atomic units, it takes the form

$$H = -\frac{1}{2}\sum_a \nabla_a^2 + \sum_{a<b} \frac{1}{|\mathbf{r}_a - \mathbf{r}_b|} - \sum_{a,A} \frac{Z_A}{|\mathbf{r}_a - \mathbf{R}_A|}, \tag{I.9}$$

where the first term describes the kinetic energy of the electrons, the second the Coulomb repulsion between them, and the third the electron-nucleus interaction. Once a particular single-particle orbital basis has been chosen, the Hamiltonian can be written in second quantized form as

$$H = \sum_{i,j} t_{i,j} c_i^\dagger c_j + \sum_{i,j,k,l} V_{i,j,k,l} c_i^\dagger c_j^\dagger c_k c_l. \tag{I.10}$$

Accurate estimation of the low-lying energies of this Hamiltonian has long been suggested as a “killer application" of quantum computing [5, 6, 43, 46, 87–91], as well as used as a benchmark for early fault-tolerant algorithms [53, 55, 59, 64, 92, 93].

While the electronic-structure Hamiltonian provides an industry-relevant benchmark, quantum advantage could be achieved earlier in the context of theoretical physics by simulating condensed matter toy models. These capture the essential quantum features of materials, while requiring far fewer resources than simulating real molecules or solids of comparable size. They avoid the main overhead of simulating realistic electronic structure problems – encoding all parameters of Eq. (I.10) into the quantum computer [43, 89, 94] – because they can be described with only a few parameters. For example, the

Fermi-Hubbard model [57, 61, 67, 73, 94]

$$H = -t \sum_{\langle i,j \rangle} \sum_{\sigma \in \{\uparrow,\downarrow\}} (c^\dagger_{i,\sigma} c_{j,\sigma} + c^\dagger_{j,\sigma} c_{i,\sigma}) + U \sum_i c^\dagger_{i,\uparrow} c_{i,\uparrow} c^\dagger_{i,\downarrow} c_{i,\downarrow} \tag{I.11}$$

is a simplified model for interacting electrons in solids, and becomes classically intractable in the intermediate regime where $t \sim U$ [95]. Other frequently considered models include uniform electron gas [84, 96, 97], and spin systems such as the Ising [21, 22, 56, 60, 61, 73, 98, 99], Heisenberg [66, 100], and dipolar spin [101] Hamiltonians. These models have also been extensively studied using classical methods [24–27, 95, 102], making them well-characterized and compelling targets for the first demonstrations of quantum advantage. A crucial part of this effort is identifying the regimes where classical simulations fail to reach the required precision.

Achieving genuine quantum advantage requires understanding the strengths of the classical competition. For example, whether quantum algorithms can provide an asymptotic speedup for molecular ground-state calculations remains an open question [103]. Many quantum systems of practical interest can, and routinely are, simulated efficiently using classical heuristics. Quantum chemistry methods including density functional theory (DFT) and coupled-cluster theory, as well as more general approaches such as perturbation and mean field theory, quantum Monte Carlo, and tensor network methods exploit symmetries, locality, and low entanglement to bypass brute-force Hilbert space enumeration [5, 87]. In general, systems with strong correlations are expected to be classically hard, and thus represent promising targets for quantum simulation, including catalysts, transition-metal complexes, and high-temperature superconductors [5]. Interestingly, exponential quantum advantages have also recently been demonstrated for certain non-interacting systems [104, 105]. Meanwhile, classical algorithms continue to advance rapidly, often challenging quantum advantage claims [14, 20, 22] shortly after their publication [24–28]. As a result, the notion of quantum advantage remains a moving target, shaped by the continual progress of both classical and quantum methods – the latter of which we describe in the remainder of this section.

I.2.2. Digital quantum simulation pipeline

In contrast to analog quantum simulation, where a controllable quantum system is engineered to mimic the dynamics of the target system, digital quantum simulation employs a universal, programmable quantum computer to reproduce those dynamics through a sequence of quantum gates, as described in section I.1. Many quantum simulation pipelines follow the structure

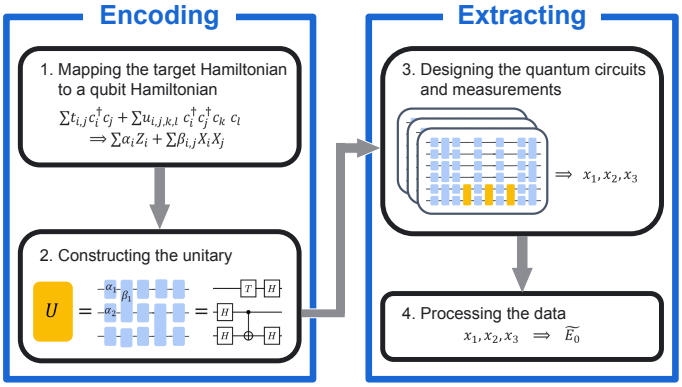


Figure I.2.: Flowchart for quantum simulation as considered in this thesis. First, the target system is mapped to qubits. Secondly, a quantum circuit U implementing the encoded dynamics, and a state preparation circuit U_ψ are constructed. Thirdly, higher-level quantum circuits that make black-box calls to U and U_ψ are designed such that the measurement outcomes x contain information about the target property. Finally, the measurement outcomes are classically processed to give an estimate of the target property. The first two steps are discussed in sections I.2.2 and I.2.3, and the latter two in sections I.3 and I.4.

presented in Fig. I.2: (i) mapping the target system to qubits, (ii) preparing an appropriate quantum state, (iii) implementing its time evolution under the encoded Hamiltonian, and (iv) extracting the desired physical properties. In this section we briefly review the first two steps, before turning to a more detailed discussion of step (iii) in section I.2.3 and step (iv) in sections I.3 and I.4.

The first step in digital quantum simulation is to represent the target Hamiltonian in a form suitable for implementation on qubits. Systems with continuous degrees of freedom must first be discretized by selecting and truncating an appropriate basis – for example, plane waves up to a cutoff frequency. The resulting finite-dimensional Hamiltonian is then typically expressed as a weighted sum of simpler operators,

$$H = \sum_{j=1}^{L} h_j P_j, \tag{I.12}$$

where $h_j \in \mathbb{R}$ are coefficients and P_j are easily implementable operators, such

as tensor products of Pauli matrices. This representation serves as the starting point for constructing a quantum circuit that encodes the corresponding time-evolution operator. The choice of decomposition and circuit design must be considered jointly, because different circuit constructions are sensitive to different Hamiltonian features — such as the number of terms L, or the one-norm of the parameter vector h, known as the λ factor,

$$\lambda = \sum_{j=1}^{L} |h_j|. \tag{I.13}$$

Methods for compiling these decompositions into quantum circuits are discussed in section I.2.3.

Realizing the decomposition of Eq. (I.12) requires choosing how the physical degrees of freedom are represented on qubits. For lattice spin systems, the mapping is straightforward: each spin degree of freedom can be represented directly by a qubit. By contrast, for fermionic or higher-dimensional systems, several strategies exist [5]. For example, in electronic structure simulations, one can use the second quantized approach, where each qubit represents occupation of one fermionic mode [43], or first-quantized approach, where the state of each electron is encoded into a separate qubit register [97, 106]. Constructing efficient mappings is an active field of research. Recent developements include simulation of exponentially many free fermions with only polynomial resources [104, 105], and significant reductions in λ factor for chemistry simulation by rewriting the Hamiltonian using tensor hyper-contraction (THC) [43, 94] or as a sum of squares (SOSSA) [107]. Progress along these lines has led to dramatic improvements – for instance, reducing the estimated runtime for simulating the FeMoco cluster in nitrogenase, a canonical benchmark for quantum chemistry, from centuries [90] to mere hours [107].

While much recent progress has focused on optimizing Hamiltonian encodings, an equally important challenge is the preparation of suitable quantum states. The complexity of this task is still under debate: some analyses suggest that state preparation will not be the dominant bottleneck [106, 108–110], whereas others argue it could fundamentally limit the prospects for quantum advantage [103]. Common approaches include preparing Hartree–Fock or mean-field reference states, using adiabatic or variational methods to approximate ground or excited states [6, 87], and employing filtering and amplitude amplification to improve overlap with the target eigenstates [60, 111, 112]. In this thesis, we follow the common practice of treating state preparation as a separate component. However, it plays a crucial role, since different assumptions about available states and their preparation cost translate into different estimation problems (sec. I.4.1) and determine which algorithms are

most effective for extracting the desired properties.

The ultimate goal of quantum simulation is to extract physically meaningful properties, examples of which were given in Sec. I.2.1. Once the Hamiltonian and initial state preparation have been encoded into quantum circuits, the resulting unitary operations can be treated as black boxes. We are left with the task of inference: we must learn about the encoded parameters by querying these quantum black boxes. This brings us to the domain of quantum estimation, which is the focus of section I.3 and I.4. Before that, however, we examine in more detail how the qubit Hamiltonian is compiled into an executable quantum circuit.

I.2.3. Hamiltonian simulation

In this section, we describe how to construct the unitary operator that encodes the Hamiltonian and serves as the black box in the estimation algorithms discussed later in this thesis. Several approaches have been developed for this task over the past three decades, two broad families have emerged: product formulas, which are often favored in the NISQ and early fault-tolerant regimes due to their lower qubit count and constant factors [6, 53, 59, 64, 66, 67, 87, 93, 113], and block-encoding methods, which achieve asymptotically optimal scaling and are expected to underpin large-scale fault-tolerant quantum simulation [6, 43, 46, 87, 89, 94, 97, 114].

Product formulas aim to approximate the time evolution operator e^{-iHt} by decomposing it into a product of unitaries that can be implemented directly. They rely on splitting the evolution into m small time steps, yielding approximations that become accurate in the limit $m >> 1$. The simplest case, first-order Trotter formula [113], relies on approximating $e^{i(A+B)} = e^{iA}e^{iB}$, which is true iff $[A, B] = 0$, so errors are on the order of the size of the commutator, and is given by:

$$e^{-it\sum_{j=1}^{L} h_j P_j} = \left(\prod_{j=1}^{L} e^{-ih_j P_j \frac{t}{m}}\right)^m + \mathcal{O}\left(\frac{1}{m^2}\right). \tag{I.14}$$

Each term in the product is a Pauli rotation and can easily be implemented both in the NISQ [115] and fault-tolerant regimes [116]. The approximation error scales as $\|U - e^{-iHt}\| = \mathcal{O}(Lt^2/m)$ [113]. More advanced techniques achieve better error scaling by using higher order expansions [113]. Alternatively, randomized product formulas [59, 117, 118] yield error bounds that depend on λ instead of L.

An alternative approach is block-encoding, which embeds the Hamiltonian

into a sub-block of a larger unitary matrix

$$V = \begin{bmatrix} H/\lambda & \cdot \\ \cdot & \cdot \end{bmatrix}. \tag{I.15}$$

Typically, V is constructed using the linear combination of unitaries (LCU) technique. This involves building two circuits P and S, so that $V = (P^\dagger \otimes \mathbb{1})S(P \otimes \mathbb{1})$. P creates a quantum state on a control register with amplitudes proportional to the coefficients h_j

$$P|0\rangle = \frac{1}{\sqrt{\lambda}} \sum_{j=1}^{L} \sqrt{|h_j|}\,|j\rangle \tag{I.16}$$

while S conditionally applies Pauli terms P_j based on the control register

$$S = \sum_{j=1}^{L} |j\rangle\langle j| \otimes \mathrm{sgn}(h_j) P_j. \tag{I.17}$$

Then we can check that the upper left block of V is

$$(\langle 0| \otimes \mathbb{1})V(|0\rangle \otimes \mathbb{1}) = (\langle 0| P^\dagger \otimes \mathbb{1})S(P|0\rangle \otimes \mathbb{1}) \tag{I.18}$$

$$= \frac{1}{\lambda} \sum_{j=1}^{L} h_j P_j = \frac{H}{\lambda}. \tag{I.19}$$

However, accessing H implemented in this way is only successful if the control register is measured in $|0\rangle$. Instead, one can construct a qubitization operator $W = V(R \otimes \mathbb{1})$, where $R = 2|0\rangle\langle 0| - \mathbb{1}$ acts on the control register [114]. The unitary W has eigenvalues related to the spectrum of the Hamiltonian, so it can be used directly to estimate the energies. For other applications, quantum signal processing allows implementing functions of the Hamiltonian by interleaving V with other gates; in particular time evolution e^{-iHt} can be implemented with optimal cost [114].

Both approaches ultimately yield a unitary operation that can be called as a subroutine by higher-level quantum algorithms. The choice between them depends on the available hardware and the cost model: product formulas minimize constant factors and are often best suited to early devices, while block-encoding achieves optimal asymptotic scaling at the expense of more complex circuit structure. Having established how the Hamiltonian can be encoded as a unitary, we now turn to how such unitaries are used to infer physical quantities — the subject of the next section on quantum estimation.

I.3. Quantum parametric estimation

Quantum parametric estimation studies how to optimally extract classical information from parametrized quantum objects. It provides a rigorous framework to understand the fundamental limits of estimation and to identify strategies that can achieve quantum advantage. While this thesis focuses on estimation problems arising in quantum simulation, the framework is equally relevant to quantum sensing applications where one aims to learn physical properties of a sample. In this section, we review the basic theory of classical and quantum estimation, and illustrate it through the canonical example of phase estimation.

I.3.1. Classical estimation theory

The goal of statistical estimation theory is to infer the value of an unknown parameter $\theta \in \mathbb{R}$ from observed data drawn from a parameter-dependent probability distribution $p_\theta(x)$. This is achieved by defining an estimator — a new random variable that is a function $\tilde{\theta}[x]$ of the observed outcomes x, and serves as a guess for the true parameter value. The quality of an estimator is typically measured in terms of two properties: its bias b,

$$b = \mathbb{E}[\tilde{\theta}[x] | x \sim p_\theta(x)] - \theta, \tag{I.20}$$

that quantifies systematic error, and root-mean-square (RMS) error ϵ,

$$\epsilon^2 = \mathbb{E}[(\tilde{\theta}[x] - \theta)^2 \mid x \sim p_\theta(x)], \tag{I.21}$$

that quantifies overall estimation uncertainty. Ideally, one aims to construct an estimator that is unbiased ($b = 0$) or at least consistent ($b \to 0$ with increasing resources), while ensuring that the error ϵ remains below acceptable thresholds.

The fundamental limit on estimation accuracy is determined by how much information about θ is contained in the probability distribution $p_\theta(x)$. This is captured by the Fisher information

$$\mathcal{I} = \mathbb{E}[(\partial_\theta \log p_\theta(x))^2 \mid x \sim p_\theta(x)]. \tag{I.22}$$

If the data $x = (x_1, x_2, \ldots, x_N)$ consists of N independent, identically distributed samples then $p_\theta^{(N)}(x) = \prod_{j=1}^N p_\theta(x_j)$ and the total Fisher Information becomes $\mathcal{I}^{(N)} = N\mathcal{I}$, where $\mathcal{I}$ is the Fisher information of a single sample.

The Fisher information sets a fundamental lower bound on how efficiently one can estimate θ. The Cramér–Rao bound states that the RMS error of any

unbiased estimator is lower bounded by the inverse of the Fisher information:

$$\epsilon^2 \geq (\mathcal{I})^{-1}. \tag{I.23}$$

This generalises to multiparameter settings ($\theta \in \mathbb{R}^d$), where the Fisher information becomes a matrix. If this matrix is full rank, its inverse bounds the covariance matrix of any unbiased estimator. If it is not full rank, only certain linear combinations of the parameters — those corresponding to non-zero eigenvalues of the matrix — can be estimated from data sampled from p_θ.

The Cramér–Rao bound is asymptotically achievable: the maximum likelihood estimator saturates it in the limit $N \to \infty$, with $\epsilon^2/N \xrightarrow{N\to\infty} (\mathcal{I})^{-1}$. Precision guarantees for finite datasets require a more refined analysis. Nevertheless, the linear scaling of the Fisher information with the number of samples implies that achieving precision ϵ requires $\Omega(\epsilon^{-2})$ samples. This is referred to as the *standard limit* [119]. In the next section, we will see how quantum strategies can surpass this classical limit.

I.3.2. Quantum estimation theory

Quantum estimation theory extends the classical framework to scenarios where data are obtained by measuring quantum systems. Instead of starting from $p_\theta(x)$, the goal is to estimate a parameter θ encoded in a quantum state $|\psi_\theta\rangle$. To extract information from the quantum state, a measurement must be chosen — for example, a single qubit can be measured in the X, Y, or Z basis. Each measurement choice yields a different probability distribution of outcomes, characterized by a different Fisher information. A key objective in quantum estimation is to identify the measurement that maximizes the information gained about θ.

The maximum achievable Fisher information over all possible measurements on the state $|\psi_\theta\rangle$ is quantified by the *quantum Fisher information* (QFI) [120, 121]. For pure states, the QFI can be computed as [121]

$$\mathcal{I}_Q = 4(\langle\dot{\psi}_\theta|\dot{\psi}_\theta\rangle - |\langle\psi_\theta|\dot{\psi}_\theta\rangle|^2), \tag{I.24}$$

where $|\dot{\psi}_\theta\rangle = \partial_\theta |\psi_\theta\rangle$. Since $\mathcal{I}_Q$ provides an upper bound on the classical Fisher information $\mathcal{I}$ obtainable from any measurement, we obtain the quantum Cramér-Rao bound [121]

$$\epsilon^2 \geq (\mathcal{I}_Q)^{-1}, \tag{I.25}$$

which lower-bounds the variance of any unbiased estimator based on measurements of $|\psi_\theta\rangle$.

To apply the quantum estimation framework in practice, one must specify

how the quantum state $|\psi_\theta\rangle$ is generated. In quantum metrology, this is typically done by initializing a probe in a known state $|\psi_0\rangle$ and allowing it to interact with the system of interest, modeled by a parameter-dependent quantum channel Λ_θ. The aim is to design estimation protocols that achieve a target precision with minimal resource cost, such as the number of photons in an optical interferometry setup, or the T-count in a fault-tolerant quantum algorithm. A common abstraction is to model the total resource cost as the number N of applications of Λ_θ [119].

Different metrological strategies are possible [119], as presented in Fig. I.3. *Parallel* strategies apply N copies of the channel simultaneously — $\Lambda_\theta^{\otimes N}$ is applied to a potentially entangled initial state of N probes. In contrast, in *sequential* strategies a single probe undergoes Λ_θ^N — N consecutive applications of Λ_θ. A more powerful variant allows for interleaving the studied channel Λ_θ with other operations $\mathcal{V}_j$, applying $\prod_{j=1}^{N}(\mathcal{V}_j\Lambda_\theta)$ to the probe. More complex strategies involving ancillary systems entangled with the probe can be considered [122] but are beyond the scope of this thesis.

Quantum metrology can outperform classical strategies by exploiting quantum features. For example, in parallel strategies, using product input states yields QFI that scales linearly with the number of channel uses N, matching the classical limit. However, with a carefully chosen entangled initial state, the QFI can scale as $\mathcal{I}_Q^{(N)} \propto N^2$. This allows for reaching precision ϵ in the estimate with cost N scaling as $\Omega(\epsilon^{-1})$. This optimal scaling is known as the *Heisenberg limit*, reflecting its connection to the Heisenberg uncertainty principle. Designing estimation protocols that achieve this quadratic improvement over the classical case is a central goal in the field.

Whether the Heisenberg limit is achievable depends on the estimation task and the metrological setup. Achieving this limit may require sequential strategies with intermediate control, rather than parallel or simple sequential protocols. This can be the case even for the estimation of a single parameter encoded in a unitary channel [123] and becomes even more relevant in noisy [122, 124] or multiparameter [125] problems. In fact, noise can make the Heisenberg limit fundamentally unattainable, even with the most general measurement strategies [126]. Finally, even in settings where Heisenberg scaling is attainable, the achievable precision can differ by a constant factor from the quantum Cramér-Rao bound [127, 128]. These considerations highlight the importance of metrological design and careful analysis of the impact of noise. We now illustrate these ideas through a concrete example of quantum phase estimation.

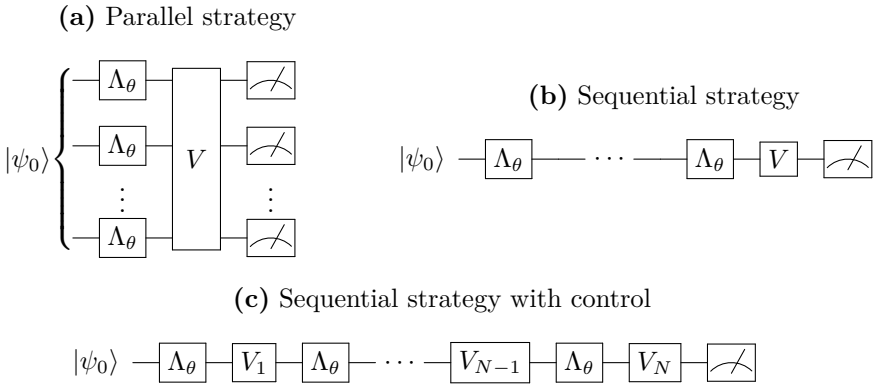


Figure I.3.: Three quantum metrology strategies for estimating a parameter θ encoded in a quantum channel Λ_θ considered in this thesis.
(a) *Parallel* — a (potentially entangled) initial state of multiple probes is evolved under parallel applications of Λ_θ, followed by a collective measurement.
(b) *Sequential* — a single probe undergoes repeated applications of Λ_θ.
(c) *Sequential with control* — control operations V_j are interleaved with applications of Λ_θ.

I.3.3. A case study in quantum metrology: phase estimation

To illustrate the concepts introduced above, we consider a canonical example of quantum parameter estimation: phase estimation. This task involves estimating an unknown phase θ encoded in the single-qubit unitary

$$U_\theta = e^{-i\theta Z/2}. \tag{I.26}$$

When this unitary is applied to an initial state $|\psi_0\rangle$, it produces the state $|\psi_\theta\rangle = U_\theta |\psi_0\rangle$. The QFI associated with this state can be calculated by substituting $|\dot{\psi}_\theta\rangle = -\frac{i}{2} Z |\psi_\theta\rangle$ into Eq. (I.24) to be

$$\mathcal{I}_Q = 4(\frac{-i^2}{4} \langle\psi_\theta| Z^2 |\psi_\theta\rangle - |\langle\psi_\theta| \frac{i}{2} Z |\psi_\theta\rangle|^2) \tag{I.27}$$

$$= 1 - |\langle\psi_0| Z |\psi_0\rangle|^2. \tag{I.28}$$

The QFI is maximized when $\langle\psi_0| Z |\psi_0\rangle = 0$, which is achieved by the equal superposition state $|\psi_0\rangle = |+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Measuring $|\psi_\theta\rangle$ in the X

basis yields outcomes ± 1 with probabilities

$$p(\pm 1|\theta) = \frac{1 \pm \cos\theta}{2}. \tag{I.29}$$

Inserting that into Eq. I.22 gives classical Fisher information for this measurement

$$\mathcal{I} = \frac{2}{1+\cos\theta} \times \frac{\sin^2\theta}{4} + \frac{2}{1-\cos\theta} \times \frac{\sin^2\theta}{4} \tag{I.30}$$

$$= \frac{\sin^2\theta}{2} \times \frac{1-\cos\theta+1+\cos\theta}{1-\cos^2\theta} = 1. \tag{I.31}$$

This matches the QFI, showing that X-basis measurement is optimal for this state.

As the number N of applications of U_θ is increased, one can achieve either the standard quantum limit $\mathcal{I}^{(N)} \propto N$ or the Heisenberg limit $\mathcal{I}^{(N)} \propto N^2$ depending on the choice of the experimental setup. Simply repeating the single-qubit experiment described above N times yields $\mathcal{I}^{(N)} = N$. Similarly, using an initial product state $|\psi_0\rangle = |+\rangle^{\otimes N}$ in the parallel strategy yields a state $|\psi_\theta\rangle = U_\theta^{\otimes N} |+\rangle^{\otimes N}$ with $\mathcal{I}_Q^{(N)} = N$. In contrast, an entangled initial state $\sqrt{2}\,|\psi_0\rangle = |0\rangle^{\otimes N} + |1\rangle^{\otimes N}$ becomes $\sqrt{2}\,|\psi_\theta\rangle = |0\rangle^{\otimes N} + e^{iN\theta}\,|1\rangle^{\otimes N}$ leading to QFI

$$\mathcal{I}_Q^{(N)} = 4\left(\left|\frac{iNe^{iN\theta}}{\sqrt{2}}\,|1\rangle^{\otimes N}\right|^2 - \left|\frac{\langle 0|^{\otimes N} + \langle 1|^{\otimes N}}{\sqrt{2}}\,\frac{iNe^{iN\theta}}{\sqrt{2}}\,|1\rangle^{\otimes N}\right|^2\right) \tag{I.32}$$

$$= 4(\frac{N^2}{2} - \frac{N^2}{4}) = N^2, \tag{I.33}$$

and achieving the Heisenberg limit. Similarly, the sequential strategy with initial state $|\psi_0\rangle = |+\rangle$ gives $|\psi_\theta\rangle = U_\theta^N |+\rangle = (|0\rangle + e^{iN\theta}\,|1\rangle)/\sqrt{2}$ with QFI

$$\mathcal{I}_Q^{(N)} = 4\left(\frac{-iN^2}{2}\,\langle 1|1\rangle - \left|\frac{\langle 0| + \langle 1|}{\sqrt{2}}\,\frac{iNe^{iN\theta}}{\sqrt{2}}\,|1\rangle\right|^2\right) \tag{I.34}$$

$$= 4(\frac{N^2}{2} - \frac{N^2}{4}) = N^2, \tag{I.35}$$

again achieving the Heisenberg limit. The corresponding outcome probabilities for X-basis measurements are $p(\pm 1|\theta) = \frac{1\pm\cos N\theta}{2}$, yielding classical Fisher information $\mathcal{I}^{(N)} = N^2$ and Cramér-Rao bound $\epsilon \geq N^{-1}$.

Unlike many asymptotic analyses, in phase estimation even constant pref-

actors in precision bounds are well understood. Ref. [127] showed that the Heisenberg limit $\epsilon \geq 1/N$ cannot be reached exactly, and derived a tighter bound $\epsilon \geq \pi/N$. This bound is asymptotically saturable in the parallel setting using a specially tailored entangled state known as the sine-state, originally proposed in Ref. [129]. Sequential strategies can achieve the Heisenberg scaling [130, 131], but with a worse constant factor [132].

To illustrate how noise limits quantum advantage, consider the sequential strategy outlined above under depolarizing noise (introduced in Section I.1.1). Suppose that with each application of $U(\theta)$, the quantum state is replaced by the maximally mixed state with probability η. The output distribution becomes

$$p(\pm 1|\theta) = (1-\eta)^N \frac{1 \pm \cos N\theta}{2} + (1-(1-\eta)^N)\frac{1}{2} \tag{I.36}$$

$$= \frac{1 \pm (1-\eta)^N \cos N\theta}{2} \tag{I.37}$$

and the corresponding Fisher information satisfies

$$\mathcal{I}^{(N)} = N^2(1-\eta)^{2N} \frac{\sin^2 N\theta}{1-(1-\eta)^{2N}\cos^2 N\theta} \leq N^2(1-\eta)^{2N}. \tag{I.38}$$

This exponential decay of Fisher information with N implies that multiple shorter experiments with fewer channel uses can outperform a single long one – a principle that guides the phase estimation algorithms developed in this thesis. Before presenting these new algorithms, the following section reviews the standard quantum phase estimation algorithms on which they are built.

I.4. Quantum phase estimation algorithms

In the context of quantum computing, Quantum Phase Estimation (QPE) refers to the task of estimating eigenphases of a given unitary operator. It plays a central role in quantum simulation — especially for estimating eigenenergies [5] — and appears in broader applications beyond simulation [11, 133]. QPE also serves as a benchmark problem for assessing resources needed for quantum advantage with fully fault-tolerant quantum computers [43, 46, 88, 90, 94], and early fault-tolerant implementations [53, 55, 64, 72, 100].

Phase estimation protocols generally consist of two interconnected components: the design of a quantum circuit and the construction of a classical estimator. The quantum circuit determines what measurement data is collected, while the classical estimator specifies how this data is processed to produce an estimate of the target phase. These two components are often

linked by a fundamental trade-off in complexity. A sophisticated quantum circuit can perform quantum information processing that simplifies the classical inference task, sometimes reducing it to a direct readout of measurement results. Conversely, simpler circuits may require more advanced classical estimators or larger volumes of measurement data to achieve comparable accuracy. In this section, we discuss two broad families of phase estimation algorithms: those that aim to simplify classical processing (Section I.4.2) and those that prioritize simpler quantum circuits (Section I.4.3). We begin, however, by outlining common variants of the QPE problem these algorithms aim to solve.

I.4.1. Variants of the QPE problem

QPE refers to a family of problems centered on estimating the eigenphases of a unitary operator U, typically assuming access to a controlled version of U (denoted c-U), an initial state $|\psi\rangle$, and a target precision. The task is to design quantum circuits and classical post-processing to estimate one or more eigenphases of U. Performance is typically measured in terms of *total query cost* - the number of uses of c-U. In near-term or early fault-tolerant settings, one also cares about the *query depth* - the maximal number of uses of c-U in a single circuit [55, 56, 61, 134]. The specific formulation of the QPE problem depends on assumptions about the input state and the desired form of the output; we now outline several common variants.

The textbook version of QPE [29] considers an idealized setting in which the input is an eigenstate $|\psi\rangle = |\phi\rangle$ of U, satisfying $U|\phi\rangle = e^{i\phi}|\phi\rangle$. This case is mathematically equivalent to the phase estimation problem discussed in section I.3.3 due to *phase kickback*: if the control qubit is in state $|\chi\rangle = \alpha|0\rangle + \beta|1\rangle$, and we apply controlled U with the register prepared in the eigenstate $|\phi\rangle$, then

$$\text{c-}U|\chi\rangle \otimes |\phi\rangle = \alpha|0\rangle \otimes |\phi\rangle + \beta|1\rangle \otimes U|\phi\rangle \tag{I.39}$$

$$= (\alpha|0\rangle + e^{i\phi}\beta|1\rangle) \otimes |\phi\rangle \tag{I.40}$$

$$\equiv (e^{-i\frac{\phi}{2}Z}|\chi\rangle) \otimes |\phi\rangle\,. \tag{I.41}$$

However, practical quantum simulation applications rarely provide access to exact eigenstates. For a general initial state $|\psi\rangle$, if $U = \sum_j e^{i\phi_j}|\phi_j\rangle\langle\phi_j|$ is the spectral decomposition of U, applying c-U generally entangles the control qubit with the system register

$$\text{c-}U(\alpha|0\rangle + \beta|1\rangle) \otimes |\psi\rangle = \sum_j \langle\phi_j|\psi\rangle\,(\alpha|0\rangle + e^{i\phi_j}\beta|1\rangle) \otimes |\phi_j\rangle\,. \tag{I.42}$$

This no longer yields a single-qubit rotation, necessitating consideration of alternative problem variants.

Depending on the assumptions made about the input state $|\psi\rangle$ and the estimation objective, different versions of the QPE problem can be defined. The most common version is *ground state energy estimation* [55–57, 61, 112, 135]. In this setting, $U = e^{iH}$, and the goal is to estimate the lowest eigenvalue of H under the assumption that the initial state has a sufficient overlap with the ground state, and given a lower bound on the spectral gap (the energy difference between the ground state and first excited state). Another variant is *multi-phase estimation* [62, 73, 75, 136], which assumes that the initial state has significant overlap with only a small number of eigenstates and aims to estimate all corresponding eigenphases. Alternatively, in a sampling variant of the problem, the goal is to sample from the spectrum of U with probabilities given by $|\langle\phi_j|\psi\rangle|^2$ [94, 137]. Finally, one may attempt to develop an approximate model for the spectral distribution $A(E) = \sum_j \delta(E - E_j)|\langle\phi_j|\psi\rangle|^2$, which is known as the quantum eigenvalue estimation problem (QEEP) [138].

Furthermore, QPE problems differ in the precision targets used to assess performance. In this thesis, we use the root-mean-square error ϵ as the primary error metric (Eq. (I.21)). Another common choice is to require that the estimate lies within ϵ' of the true value with probability at least $1 - \delta$ [55–57, 61]. These two metrics can be related: for instance, Chebyshev's inequality implies $\delta \leq (\epsilon/\epsilon')^2$. Moreover, since the eigenphases lie on a circle, the maximum possible error is bounded by π, yielding the inequality $\epsilon^2 \leq (1 - \delta)\epsilon'^2 + \delta\pi^2$. In multi-phase estimation, it is common to require that each eigenphase is estimated to within the same error ϵ, either with high probability [62] or in expectation [75, 136].

I.4.2. QFT-based QPE algorithms

Quantum phase estimation algorithms can be broadly grouped into two classes, depending on how they balance quantum circuit complexity with classical post-processing. This section focuses on the class of algorithms that prioritize coherent quantum processing: they use deep, structured circuits with minimal classical post-processing. These algorithms achieve the optimal information-theoretic scaling [139] and are expected to be standard approach in fully fault-tolerant quantum simulation [43, 46, 88, 90, 94]. As shown in Ref. [139], any algorithm that achieves this optimal performance can be expressed using a three-step circuit structure involving control state preparation, phase encoding, and a quantum Fourier transform (QFT) measurement. Therefore, we refer to this family as QFT-based QPE algorithms. This class of algorithms corresponds to the parallel strategies in section I.3.2 (Fig. I.3).

The structure of the QFT-based QPE circuit is illustrated in Fig. I.4. First,

some state $|\chi\rangle = \sum_x \alpha_x |x\rangle$ is prepared on the n-qubit control register using a unitary V. The second step applies a controlled unitary of the form

$$\sum_{x=0}^{2^n-1} |x\rangle\langle x| \otimes U^x, \tag{I.43}$$

which can be implemented using $2^n - 1$ controlled powers of U [29]. This maps the joint input state $|\chi\rangle \otimes |\phi\rangle$ to

$$\left(\sum_x \alpha_x e^{i\phi x} |x\rangle\right) \otimes |\phi\rangle\,, \tag{I.44}$$

encoding the phase information into the control register. Finally, the control state is measured in the Fourier basis via the inverse quantum Fourier transform followed by computational basis measurement.

Different control states $|\chi\rangle$ can be used depending on the task. The textbook algorithm [29] uses the Hadamard transform $V = H^{\otimes n}$ to prepare an equal superposition control state $|\chi\rangle = 2^{-n/2} \sum_{x=0}^{2^n-1} |x\rangle$. The optimal-variance sine state (Sec.I.3.3) can also be implemented efficiently [94] to improve precision. Other control states have also been proposed and analyzed in various settings [134, 140–142].

Importantly, the circuit in Fig. I.4 implements a projective measurement in the eigenbasis of U. If the system register is initialized in a state $|\psi\rangle = \sum_j a_j |\phi_j\rangle$ where $|\phi_j\rangle$ are the eigenstates of U, the state before the

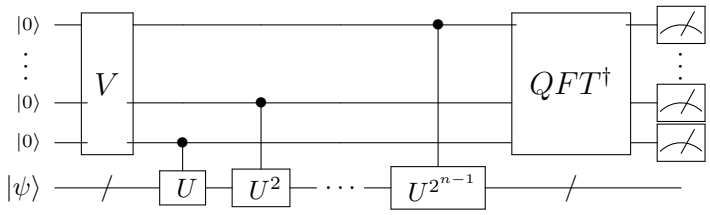


Figure I.4.: QFT-based QPE circuit. The system register is prepared in state $|\psi\rangle$. First, the unitary V is used to prepare the control register in a state $|\chi\rangle$. Then, controlled U are used to encode the phase estimation into the control register. Finally, the control register is measured in the Fourier basis. The system register is then projected onto the corresponding eigenstate of U.

computational basis measurement becomes

$$\sum_j a_j \left|\tilde{\phi}_j\right\rangle \otimes |\phi_j\rangle\,, \tag{I.45}$$

where $\left|\tilde{\phi}_j\right\rangle$ is a state on the control register that is sharply peaked near the n-bit approximation of $\phi_j/2\pi$ [29]. The probability of measuring a $\pi/2^n$-precise estimate of ϕ_j is approximately $|a_j|^2 = |\langle\phi_j|\psi\rangle|^2$, and upon measurement, the system register collapses to the corresponding eigenstate $|\phi_j\rangle$.

The classical post-processing for this class of algorithms is minimal. Each run of the circuit returns a bitstring x such that, with high probability, $2\pi x/2^n$ approximates one of the eigenphases with precision $\sim \pi/2^n$. This directly solves the sampling version of the phase estimation problem, or the full estimation problem when the initial state is an eigenstate. When targeting the ground state energy, typically an upper bound is assumed and samples above this threshold are discarded. The procedure is repeated until a low-energy sample is observed. Recent works have proposed averaging the lowest few outcomes over multiple runs [134] or using maximum-likelihood techniques [76] to guarantee precision with limited circuit depth.

QFT-based QPE algorithms achieve the optimal query complexity. To achieve precision ϵ, one typically uses $n = \Theta(\log(\epsilon^{-1}))$ control qubits and circuits of query depth $2^n = \Theta(\epsilon^{-1})$, leading to the optimal query cost $\Omega(\epsilon^{-1})$ [29, 129]. To ensure at least one sample of the target eigenphase ϕ_0, the number of measurements must scale as $\Omega(|\langle\phi_0|\psi\rangle|^{-2})$, though this can be improved to $\Omega(|\langle\phi_0|\psi\rangle|^{-1})$ using eigenstate amplification techniques [112, 135].

In early fault-tolerant and NISQ regimes, the standard approach of increasing circuit depth and control width with desired precision becomes impractical. Instead, one can fix the number of control qubits and limit circuit depth, improving precision by increasing the number of samples. This relaxes hardware requirements but increases the total query complexity to $\Omega(\epsilon^{-2})$. This trade-off has been largely unexplored in QFT-based algorithms, with only a few exceptions [76, 134]. A different class of phase estimation algorithms, however, is explicitly designed to operate in this regime. We turn to these "single-control" protocols next.

I.4.3. Single-control QPE algorithms

The second broad class of quantum phase estimation (QPE) algorithms reduces circuit complexity by shifting part of the computational effort to classical postprocessing. These algorithms are based on circuits where controlled-U is applied repeatedly using a single control qubit – hence we refer to them

as *single-control* QPE algorithms. By construction, they are well suited to NISQ and early fault-tolerant regimes: they minimize the number of logical qubits required and allow a flexible trade-off between circuit depth and the number of measurements. This class is analogous to the sequential metrology strategies introduced in Section I.3.2 (Fig. I.3).

Typical single-control QPE algorithms rely on the Hadamard test circuit shown in Fig. I.5. This circuit starts by preparing an equal superposition $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ on the controlled qubit, followed by k applications of the controlled unitary U. This prepares a state

$$\text{c-}U^k |+\rangle \otimes |\psi\rangle = \sum_j \langle\phi_j|\psi\rangle \frac{|0\rangle + e^{i\phi_j k} |1\rangle}{\sqrt{2}} \otimes |\phi_j\rangle . \tag{I.46}$$

Then, the control qubit is typically measured in the X or Y basis. The expected value of the X measurement can be calculated as

$$\sum_j |\langle\phi_j|\psi\rangle|^2 \frac{\langle 0| + e^{-i\phi_j k} \langle 1|}{\sqrt{2}} X \frac{|0\rangle + e^{i\phi_j k} |1\rangle}{\sqrt{2}} = \sum_j |\langle\phi_j|\psi\rangle|^2 \cos(k\phi_j), \tag{I.47}$$

and similarly the expected value of Y is $\sum_j |\langle\phi_j|\psi\rangle|^2 \sin(k\phi_j)$. Repeated measurements in both basis can be used to estimate a signal function

$$g(k) = \sum_j |\langle\phi_j|\psi\rangle|^2 e^{ik\phi_j} \tag{I.48}$$

from which the phases ϕ_j can be estimated via classical Fourier analysis over multiple values of k.

The choice of query depth k plays a crucial role. The Hadamard test

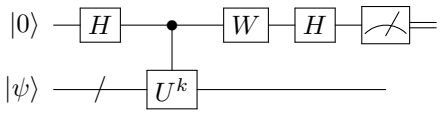


Figure I.5.: The Hadamard test circuit for U^k. The control qubit is prepared in the $|+\rangle$ state using a Hadamard gate, then used to control k applications of U on the system register prepared in state $|\psi\rangle$. Then, the control qubit is measured, typically, in the X basis (by setting $W = \mathbb{1}$) or Y-basis ($W = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$).

corresponds exactly to the simple sequential strategy in Section I.3.3, and when $|\psi\rangle$ is an eigenstate, the Fisher information is equal to k^2. To achieve Heisenberg scaling, one must allow large powers $k \propto \epsilon^{-1}$. In scenarios with bounded query depth $k \leq K$ and M repetitions, the Fisher information is limited by MK^2, leading to standard scaling of the query cost as ϵ^{-2}. Since the circuit with query depth k only allows to learn the phases ϕ_j modulo $\frac{2\pi}{k}$, circuits with different powers k are used to resolve this ambiguity. QPE algorithms can interpolate between high-depth circuits with Heisenberg scaling and low-depth circuits with standard scaling — but the performance remains at least a constant factor above the optimal limit [132].

Various signal processing techniques can be used to extract eigenphases from Hadamard test measurements. The first single-control algorithm by Kitaev [143] used k values equal to the powers of 2 up to ϵ^{-1} to read out the binary expansion of the phase from least to most significant bit. This method required adaptive measurements: the choice of the final gate W in Fig.I.5 depended on previous measurement outcomes. Ref. [131] was the first to show how to obtain the Heisenberg limit without adaptivity. In recent years, new methods have been developed for ground state energy estimation with arbitrarily small overlap $|\langle\phi_0|\psi\rangle|$ between the initial state and the ground state. These techniques use random values of k, sampled from carefully designed distributions, and achieve a query cost of $\Omega(|\langle\phi_0|\psi\rangle|^{-4}\epsilon^{-1})$ [55, 57, 61]. A parallel line of work has extended single-control QPE to multiphase estimation [62, 73, 75, 136]. Significant effort has been devoted to reducing the maximum query depth [55, 61, 73], and to ensuring robustness in the presence of noise [74, 99, 144].

More recently, a new class of single-control QPE algorithms for ground state estimation has emerged that goes beyond the Hadamard test [55, 60, 112, 145]. These approaches correspond to the sequential with control strategies of Section I.3.2 (Fig. I.3), and make use of quantum signal processing and related techniques to implement approximations of functions of H, where $U = e^{iH}$. In particular, approximating the step function allows one to locate the ground state energy via a binary search procedure. These algorithms also interpolate between low-depth and Heisenberg-scaling regimes, and improve the scaling of the query cost in the overlap to $\Omega(|\langle\phi_0|\psi\rangle|^{-2})$.

I.5. Outline of this thesis

This thesis introduces four algorithms for quantum parameter estimation problems, developed with early fault-tolerant quantum simulation in mind. Chapters II, III, IV focus on phase estimation under early fault-tolerance constraints, such as limited qubit counts, general initial states, and the

presence of residual logical noise. Chapter V shifts focus to multiparameter estimation beyond phase estimation, addressing the problem of learning the coefficients of a many-body Hamiltonian from its time dynamics.

Chapter II: Heisenberg-limited quantum phase estimation of multiple eigenvalues of few control qubits

Chapter II introduces a new algorithm for single-control multi-phase estimation. It combines Heisenberg-limited Robust Phase Estimation [74], previously limited to single-phase estimation, with time-series signal processing techniques for resolving multiple frequencies [136, 138]. A central contribution is an adaptive scheme for selecting query depths k in the Hadamard test to avoid aliasing and ensure identifiability of all relevant eigenphases. The resulting protocol achieves Heisenberg-limited scaling for estimating multiple phases in the single-control setting.

Chapter III: Error mitigation and circuit division for early fault-tolerant quantum phase estimation

Chapter III studies the performance of the optimal-variance sine-state QPE circuit in the presence of noise. We optimize the circuit depth for a given global depolarising noise rate and derive the sample complexity needed for high-precision estimation. We introduce a new error mitigation technique tailored to phase estimation, extending Probabilistic Error Cancellation reviewed in Section I.1.3 to maximum likelihood estimation. We provide detailed resource estimates (number of physical qubits and time) needed for executing our protocol for quantum chemistry and condensed matter models, and show that our scheme can halve the required number of physical qubits.

Chapter IV: Accurate ground state energy estimation with noise and imperfect state preparation

Chapter IV generalises the results of Chapter III to a more realistic setting where a perfect eigenstate is not available. We propose a new estimator for ground state energy estimation based on rejection sampling and approximate maximum likelihood, compatible with data generated by both QFT-based and quantum signal processing QPE circuits. By building on the error mitigation technique of the previous chapter, the estimator inherits robustness to circuit-level noise, and additionally achieves robustness to imperfect eigenstate preparation. We support our theoretical analysis with a small-scale numerical demonstration, showing that the method performs well in the simultaneous presence of noise and low ground state overlap, while requiring only a modest

number of samples. These features make the proposed approach a promising candidate for practical early fault-tolerant quantum phase estimation.

Chapter V: The advantage of quantum control for many-body Hamiltonian learning

Chapter V addresses the task of estimating coupling parameters in a many-body Hamiltonian from time evolution experiments. We analyze this problem in several experimental scenarios: with and without the ability to apply quantum control during evolution. Our main result is a characterization of when quantum control enables a quadratic improvement in the total experiment time required. In particular we prove that, in the absence of control, learning is limited to standard quantum scaling for thermalizing systems. We also provide a Heisenberg-limited algorithm for the continuous control setting, where tunable terms can be added to the system Hamiltonian during the evolution. This establishes a fundamental role for control in enabling efficient learning of many-body quantum systems.

II. Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits

II.1. Introduction

For quantum computers to overcome the 50^+ year head start in research and development enjoyed by their classical competition, quantum algorithms must eke out every inch of quantum speedup over their classical counterparts. Quantum phase estimation (QPE) of a unitary operator U, a BQP-complete problem [137], plays a central or support role in many promising quantum applications [11, 115, 133].

However, not all flavours of quantum phase estimation are equally powerful. To perform QPE with accuracy (root-mean-square) error δ, initial implementations of quantum phase estimation [29, 146] used a $O(\log(\delta^{-1}))$-qubit control register, and required computation time scaling as $T = O(\delta^{-2})$, known as the Sampling Noise Limit, when contributions from outlying (unlikely) data are taken into account [131]. Much work has been undertaken over the succeeding years to improve QPE estimation rates to the theoretically optimal Heisenberg limit $T = O(\delta^{-1})$ [139, 147, 148] and reduce the control overhead.

The requirement to perform applications of U conditional on a large entangled control register is technically challenging and has strong coherence requirements. It has been known for a long time [149] that the control register in QPE can be replaced by a single qubit using classical feedback and re-preparation of the control qubit, also known as iterative QPE [150]. For estimating the phase of a single eigenstate, assuming the preparation of this eigenstate, the sampling noise limit can thus simply be achieved using a single control qubit. In [131] iterative-QPE was extended to achieve the Heisenberg limit $T = O(\delta^{-1})$. This Heisenberg limit can be shown, via Cramer-Rao bounds [131], to be a lower bound on the cost of phase estimation, assuming one cannot fast-forward the unitary U [151]. This type of estimation have additionally demonstrated a relative robustness to er-

ror [74, 152], which is of interest to NISQ applications. Other analyses of QPE use maximum-likelihood [153], or Bayesian [152, 154] inference.

The requirement to prepare eigenstates of the unitary U is not possible for most applications. It is well known that the 'textbook' QPE algorithm succeeds for any initial state, i.e. the output is always an accurate estimate of one of the eigenphases of U [29]. However, the performance of few-ancilla QPE on starting states that are not eigenstates has only been examined recently. In [136] it was demonstrated numerically that one may infer single eigenvalues from mixed or superposed initial eigenstates using standard classical signal processing techniques [155]. Under some additional constraints on the system, it was recently found that these techniques could be performed in the absence of any control qubits or the need to apply controlled unitary operations [156, 157], a further significant saving. Due to the need to 'densely sample' the phase function $g(k) = \sum_j A_j e^{i\phi_j k}$, and a lack of optimization of the classical post-processing techniques, Ref. [136] only achieved sampling-noise-limited scaling, but not Heisenberg-limited scaling. By dense sampling we mean that we draw samples from $g(k)$ with k a sequence of integers, $k = 0, 1, \ldots, K$ (as opposed to, say, choosing $k = 2^d$, i.e. exponentially increasing which is used in textbook QPE and iterative QPE). By a clever adjustment of the quantum phase estimation problem to target estimation of the spectral function, Eq. (II.3), of the input state, Ref. [138] was able to prove rigorous results, with bounds that were subsequently improved in Ref. [158]. Still, these results fall short of reaching "the Heisenberg limit for the problem of estimating multiple phases", however that should be defined.

In this work, we demonstrate single-control qubit quantum phase estimation at a so-called Heisenberg limit. We extend the methods used in Refs. [74, 131] that obtain Heisenberg-limited scaling for single eigenphases to the multiple-phase setting by the use of a multi-order scheme and phase matching subroutines between different orders. We show that to make this phase matching unambiguous requires the sampling scheme to be adaptive, i.e. the next choice for k of $g(k)$ depends on the current phase estimates. At each order the multi-order algorithm requires input from a dense phase estimation method: for a given order k we use samples from $g(k\mathsf{k})$ with $\mathsf{k} = 0, 1, \ldots, K$. As we require the freedom to choose k a real number, to be applicable to a completely general U we must invoke the quantum singular value transformation of Ref. [159], which requires $O(1)$ additional control bits. Using the time-series or QEEP analysis of Ref. [138] as classical processing subroutine, we are able to obtain a rigorous proof of Heisenberg-limited scaling of our multi-order scheme. Using the matrix pencil method analysed in Ref. [136], as such a dense subroutine, we are able to show numerical results consistent with the Heisenberg limit, with a performance improvement over

the time-series analysis results.

In essence, our paper is concerned with what choices of k in $g(k)$ and what classical processing are needed to enable Heisenberg-limited scaling, i.e. scaling which minimizes the total number of applications T of (controlled) U (which we refer to as the quantum cost) given a targeted error δ with which to estimate multiple eigenvalue phases of U present in some input state $|\Psi\rangle$. It can thus be viewed as purely solving a problem of classical signal processing. This does not mean that such questions are trivial: for example, the question of how to estimate phases if one is allowed to only get single samples from $g(k)$ for a set of randomly chosen k relates to the dihedral hidden subgroup problem in quantum information theory [160]. We note that other work, based on a Monte Carlo extension of [138], achieving Heisenberg scaling (up to polylog factors) was recently presented in [57]. Another recent work also demonstrated numerical evidence for Heisenberg-limited phase estimation using Bayesian methods [161]. We also note that the information-theoretically optimal method in [153] which picks random k in $g(k)$ has a classical processing cost which is linear in the quantum cost T, and Theorem 1 in [153] can be converted to bound the mean-squared-error in estimating a single phase, see comments below Theorem II.1 in Section II.2.2. In principle this information-theoretic method can be extended to the case of n_ϕ eigenvalue phases, but the classical processing cost will be exponential in n_ϕ as one iterates over the possible values of the phases, while our Algorithm II.2 has a polynomial (but superlinear) quantum and classical cost in terms of the number of phases n_ϕ.

II.1.1. Outline

We begin in Sec. II.2 by defining a few mathematical objects. We separate the quantum part of a phase estimation problem, namely sampling of the phase (or signal) function $g(k)$ in Eq. (II.2) given an input unitary U and input state $|\Psi\rangle$ in Definition II.2 through running some quantum circuits, and the classical processing of samples from $g(k)$ to extract the eigenvalue data of U. In Sec. II.2.1 we state and prove some needed properties of the distance between phases. In Sec. II.2.2 we prove several Cramer-Rao bounds on the scaling of the error versus the total quantum cost for the estimation of a single eigenvalue phase, Theorem II.1. We state the previous result on getting Heisenberg-limited scaling for a single eigenvalue phase in Algorithm II.1. Table II.1 contains a glossary of the symbols used in this paper.

In Sec. II.3 we properly define a multi-eigenvalue phase estimation problem (Def. II.4). We discuss algorithms that can be used to extract multiple phases from densely-sampled signal $g(k)$. We state error bounds satisfied by the output of Alg. II.4 (Lemma II.3) that is used as the fixed-order subroutine in

our final Algorithm II.2 which achieves Heisenberg scaling.

In Section II.4 we present our Heisenberg-limited algorithm. We discuss a critical aliasing problem to be solved which occurs when estimating multiple eigenvalues. We show that an adaptive choice for k in $g(k)$ can solve this issue and we prove that such adaptive choice always exists in Lemma II.4. In Lemma II.5 and Lemma II.6 we prove some properties about Algorithm II.2 which will be helpful in proving the final Theorem II.2.

Thus in Theorem II.2 we prove Algorithm II.2 achieves Heisenberg-limited scaling, which is the main result of this work. In Sec. II.5 we numerically compare this rigorous implementation to an implementation using the matrix pencil method, used in Ref. [136], for which we are unable to find a rigorous proof of Heisenberg scaling. We finish the paper with a discussion, Section II.6.

II.2. The classical and quantum tasks of phase estimation

One may separate quantum phase estimation into the extraction of a signal which consists of oscillations at eigenvalue frequencies ϕ_j at a chosen time k, and the processing of this signal to resolve the frequencies. Let us first define the following:

Definition II.1 (Signal or Phase Function, and Spectral Function)
Let $U \in \mathsf{U}(2^N)$ be an N-qubit unitary operator, and $|\Psi\rangle \in \mathbb{C}^{2^N}$ an N-qubit state. We label the eigenstates $|\phi_j\rangle$ of U by their phase — $U|\phi_j\rangle = e^{i\phi_j}|\phi_j\rangle$. We can decompose $|\Psi\rangle$ in terms of these eigenstates,

$$|\Psi\rangle = \sum_j a_j |\phi_j\rangle, \tag{II.1}$$

and write the overlap $A_j := |a_j|^2, \sum_j A_j = 1$. We define the phase function, – also called the signal–, $g(k)$ for $k \in \mathbb{R}$ of a state $|\Psi\rangle$ under U as

$$g(k) = \sum_j A_j e^{ik\phi_j}. \tag{II.2}$$

The spectral function $A(\phi)$ is defined as

$$A(\phi) = \sum_j A_j \delta(\phi - \phi_j). \tag{II.3}$$

Symbol	Term	Description
U	Unitary	The unitary whose eigenphases we wish to estimate.
T	Quantum cost	The total number of applications of controlled U over the course of the phase estimation algorithm.
ϕ_j	Phase	A number $\phi_j \in [0, 2\pi)$ such that the jth eigenvalue of U is $e^{i\phi_j}$.
A_j	Overlap	The overlap of the input state and the jth eigenstate of U; see Def. II.1.
$g(k)$	Phase function / Signal	See Def. II.1.
d	Order	Running index for the order of estimation. At each order we construct new phase estimates $\tilde{\phi}_j^{(d)}$ using new data and the previous estimates $\tilde{\phi}_j^{(d-1)}$.
$\tilde{\phi}_j^{(d)}$	Estimate	Estimate of ϕ_j obtained at the dth order; see step 4d of Alg. II.2.
k_d	Exponent	At order d we perform the QEEP subroutine for $V = U^{k_d}$.
κ_d	Multiplier	Defines $k_d = \kappa_d k_{d-1}$.
$\theta_j^{(d)}$	Phase	Eigenphase of U^{k_d}.
$\tilde{\theta}_j^{(d)}$	Estimate	Estimate of $\theta_j^{(d)}$; see step 4b of Alg. II.2.
ϵ	Single order error	Error parameter used for the QEEP subroutine at each order; see Alg. II.4.
δ	Final error	A bound for standard deviation of the final estimates; see Def. II.4.
p_d	Confidence bound	The probability with which the QEEP subroutine at order d succeeds; see Eq. (II.28).
A	Overlap bound	We wish to estimate the phases for which the overlap is $A_j > A$; see Def. II.4.

Table II.1.: Glossary of symbols used in this chapter.

Note that $\int_0^{2\pi} d\phi A(\phi) = 1$, *and* $g(k) = \int_0^{2\pi} d\phi e^{ik\phi} A(\phi)$*; i.e. the phase function sets the Fourier coefficients of the spectral function.*

Note that one may change seamlessly between the description of a unitary U and its eigenvalues and a Hermitian operator H and its eigenvalues using the transform $U = e^{iHt}$ for an appropriate choice of t. Note that since $g(-k) = g^*(k)$ we can restrict ourselves to $k \geq 0$.

One may consider algorithms estimating $g(k)$ at integer $k \in \mathbb{Z}^+$, which require the quantum circuits using controlled-U^k in Fig. II.1 with $k \in \mathbb{Z}^+$. In our final Alg. II.2 we will however use $k \in \mathbb{R}^+$ (in practice $k \in \mathbb{Q}^+$). In order to implement U^k, we can write $k = \lfloor k \rfloor + \alpha$, and we can simulate U^k in time $O(k)$ if we can simulate U^α in time independent of k. The accuracy of this fractional query to U^α can be independent of the final error in our phase estimation (as long as it is sufficiently small). Simulating U^α is not a significant issue for Hamiltonian simulation methods such as Trotter decompositions [113], which allow simulation of e^{iHt} for arbitrary $t \in \mathbb{R}$. If we instead have access to a circuit implementation of a unitary U, or a block-encoding of a Hamiltonian H, we can implement a fractional query of U^α via the quantum singular value transform [159, Corollary 34]. The circuit to implement U^α to error ϵ requires $\mathcal{O}(1)$ additional ancilla qubits, and $\mathcal{O}(\Delta_{\max}^{-1} \log(1/\epsilon))$ implementations of controlled-U (where $\Delta_{\max}$ is the largest gap in the spectrum of U) [1]. We will assume in this work that our states have support on at most n_ϕ phases, and so we can bound $\Delta \geq \pi/n_\phi$. The cost of implementing U^α will thus not be a significant part of the cost to implement U^k under the assumption $k >> A_0^{-1} \log(1/\epsilon)$. To simplify our remaining analysis, we assume herein that we can implement U^k at a total quantum cost k for all $k \in \mathbb{R}_+$.

The following task summarizes the quantum subroutine for phase estimation which is to be executed with the quantum circuits in Fig. II.1:

Definition II.2 (Phase Function Estimation, PFE)
Let U be an N-qubit unitary operator and $|\Psi\rangle$ an N-qubit quantum state. Assume

1. *A quantum circuit implementation of U (conditional on a control qubit) and,*

2. *A quantum circuit that prepares $|\Psi\rangle$.*

[1] The $\Delta_{\max}$ dependence in our circuit comes from the requirement in [159, Corollary 34] that our unitary have spectrum on $[-\pi + \Delta_{\max}, \pi - \Delta_{\max}]$. Though this will not immediately be the case, we are free to rotate the spectrum of our unitary to satisfy this requirement, as long as the spectral gap exists. We also only require to consider those eigenstates with support on our initial state in this algorithm; eigenstates with zero weight can be adjusted as part of implementing the quantum singular value transformation without affecting the phase function $g(k)$.

Given a $k \in \mathbb{Z}^+$, error $\epsilon > 0$, and confidence $0 < p \leq 1 \in \mathbb{R}$, PFE *outputs an estimate $\tilde{g}(k)$ of the phase function $g(k)$ of $|\Psi\rangle$ under U, with $\mathbb{P}(|\tilde{g}(k) - g(k)| \leq \epsilon) \geq 1 - p$ with quantum cost $T = M|k|$ where M is the number of repetitions of both experiments in Fig. II.1 and $M = \Theta(|\ln(1-p)|\epsilon^{-2})$ via a Chernoff bound. Our assumption[2] that the cost of implementing U^k for $k \in \mathbb{R}^+$ scales as $O(k)$ implies that the above scaling for the cost of phase function estimation holds when $k \in \mathbb{R}^+$.*

Note that estimating the quantum cost of the subroutine in Def. II.2 as linear in k is consistent with general no-fast forwarding statements [151] which state that for general Hamiltonians one cannot implement $U^t = \exp(itH)$ in time sub-linear in t. It is expected that phase function estimation is hard to do efficiently on a classical computer as it allows one, via classical post-processing, to sample from the eigenvalue distribution from the input state which can be reformulated as a BQP-complete problem [137].

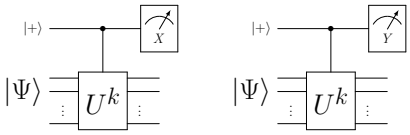


Figure II.1.: For an input state $|\Psi\rangle = \sum_j a_j |\phi_j\rangle$ and initial ancilla state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ the probability for the ancilla measurement outcome to be ± 1 is $\mathbb{P}(\pm 1) = \frac{1}{2}\sum_j A_j[1 \pm \cos(k\phi_j)]$ (left circuit measuring in the X-basis) and $\mathbb{P}(\pm 1) = \frac{1}{2}\sum_j A_j[1 \mp \sin(k\phi_j)]$ (right circuit measuring in the Y-basis).

The quantum subroutine for PFE proceeds by executing the circuits in Fig. II.1 for the given k. The control qubit is prepared in the $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ state, and is used to control k applications of the unitary U on the system register prepared in $|\Psi\rangle$. The reduced density matrix of the control qubit then takes the form

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & g(k) \\ g^*(k) & 1 \end{pmatrix}. \tag{II.4}$$

The phase function $g(k)$ is extracted by state tomography of the control qubit: one estimates the real and imaginary parts of $g(k) = g^{(r)}(k) + ig^{(i)}(k)$ by M

[2] In practice, the cost of phase function estimation for $k \in \mathbb{R}^+$ will scale as $M(\lfloor k \rfloor + O(n_\phi \log(1/\epsilon))$ when using quantum signal processing techniques. One can confirm that the effect this has on our result is to effectively increase the cost of calling the QEEP subroutine, Alg. II.4, from $O(\epsilon^{-6})$ to $O(\epsilon^{-6}\log(1/\epsilon))$. This will only change the prefactor of our Heisenberg-limited phase estimation algorithm (Alg. II.2), and does not prevent it achieving the Heisenberg limit.

repetitions of the circuit in Fig. II.1 and measurements of the control qubit in the X- or Y-basis respectively. We will ignore any dependence of phase function estimation on N.

II.2.1. Phase distance on the circle

Quantum phase estimation describes a series of protocols to estimate the eigenphases ϕ_j. As these eigenvalues are defined on the circle $[0, 2\pi)$, we need a notion of distance which respects this periodicity:

Definition II.3
For $x \in \mathbb{R}$ we define the distance $|\cdot|_T \in [0, \pi]$ as

$$|x|_T := \min_{m\in\mathbb{Z}}(|\Delta|), \text{ with } x = \Delta + 2\pi m, \quad \text{with } \Delta \in [-\pi, \pi). \tag{II.5}$$

Clearly, the distance obeys the triangle inequality: for $x_1, x_2 \in \mathbb{R}$

$$|x_1 + x_2|_T \leq |x_1|_T + |x_2|_T. \tag{II.6}$$

The following Lemma addresses a technical issue in the proof of performance of Algorithm II.2. For integer $k \in \mathbb{Z}^+$, we have $|kx|_T = \min_{m\in\mathbb{Z}} |k\Delta + 2\pi m| = k \min_{m\in\mathbb{Z}} |\Delta + \frac{2\pi m}{k}|_T$, implying Eq. (II.8) below directly. However, for $k \in \mathbb{R}^+$ (or rational numbers $k \in \mathbb{Q}^+$) we need to specify a range of x for which such a statement holds, that is:

Lemma II.1
Suppose $k > 1$ and $\theta, \phi \in [0, 2\pi)$. If

$$\frac{\pi}{k} \leq \phi \leq \frac{\pi(2\lfloor k\rfloor - 1)}{k}, \tag{II.7}$$

we have for any θ

$$\min_{n\in\{0,\ldots,\lfloor k\rfloor-1\}} \left|\phi - \frac{\theta}{k} - \frac{2\pi n}{k}\right|_T = \frac{1}{k}\left|k\phi - \theta\right|_T. \tag{II.8}$$

Proof. Let $x = \phi - \frac{\theta}{k}$, then Eq. (II.7) and $\theta \in [0, 2\pi)$ imply that

$$-\pi \leq kx \leq 2\pi\lfloor k\rfloor - \pi. \tag{II.9}$$

and thus $kx = \Delta + 2\pi m$ with $m \in \{0, \ldots \lfloor k\rfloor - 1\}$ and $\Delta \in [-\pi, \pi)$ and $|k\phi - \theta|_T = |\Delta|$. Hence $x = \frac{\Delta}{k} + \frac{2\pi m}{k}$ with $m \in \{0, \ldots \lfloor k\rfloor - 1\}$, implying Eq. (II.8) where the minimum can be achieved by $m = n$. □

II.2.2. Limits for single-eigenvalue phase estimation

For the special case of estimating a single eigenvalue phase, the Cramér-Rao theorem can be used to lower bound the quantum cost T to learn the phase with accuracy δ, known as the Heisenberg limit. The problem of estimating multiple phases ϕ_j, in the presence of unknown overlaps A_j, is not easily amenable to such Fisher information analysis as it is a multi-parameter estimation problem. However, it can be expected that the cost T of this task is at least as high as that of single phase estimation, hence it is of interest to review these bounds here. The following theorem also proves a dense signal limit which sits in between Heisenberg and sampling noise scaling:

Theorem II.1 (The Heisenberg, Dense Signal and Sampling Limits)
The (root-mean-square) error δ of an estimator $\tilde{\phi}$ of the eigenvalue phase ϕ employing the circuits in Fig. II.1 on a eigenstate of U is always lower bounded as

$$\text{Heisenberg Limit}: \delta \geq cT^{-1}, \tag{II.10}$$

where T is the quantum cost of implementing the circuits. If we choose to use only quantum circuits with $k = 1$, the sampling noise limit holds:

$$\text{Sampling Noise Limit}: \delta \geq T^{-1/2}. \tag{II.11}$$

If we choose circuits with $k = 0, 1, \ldots, K$ with a fixed number of repetitions M for each circuit we are bound by a so-called 'dense signal' limit:

$$\text{Dense Signal Limit}: \delta \geq cT^{-3/4}. \tag{II.12}$$

In these statements c is some constant.

Proof. Let $\tilde{\phi}$ be an estimator of ϕ which is inferred from the data $\mathbf{x}$. Here the data $\mathbf{x}$ is the string of outcomes of the ancilla qubit measurements for all the experiments using the left and right circuits in Fig. II.1. We have

$$\delta^2 = \sum_{\mathbf{x}} \mathbb{P}(\mathbf{x}|\phi)(\phi - \tilde{\phi}(\mathbf{x}))^2 \geq I^{-1}(\phi), \tag{II.13}$$

by the Cramér-Rao theorem [162, 163] where the Fisher information is defined as

$$I(\phi) = \sum_{\mathbf{x}} \mathbb{P}(\mathbf{x}|\phi) \left\{ \frac{\partial}{\partial \phi} \ln \left[\mathbb{P}(\mathbf{x}|\phi) \right] \right\}^2. \tag{II.14}$$

Thus $I(\phi)$ limits the information we may learn about ϕ given a dataset $\mathbf{x}$ drawn from $\mathbb{P}(\mathbf{x}|\phi)$ and we can calculate $I(\phi)$. Let M_k^r be the number of experiments, using the circuit with the X measurement, and M_k^i be the

number of experiments using the circuit with the Y measurement with a certain chosen k. The Fisher information for all independent experiments together is additive, i.e. $I(\phi) = \sum_k [M_k^r I(\phi|k,r) + M_k^i I(\phi|k,i)]$ with $I(\phi|k,r)$ and $I(\phi|k,i)$ the Fisher information of a single experiment and $\sum_k$ is the sum over the chosen set of ks. For a single experiment we can calculate, using Eq. (II.14) and the probability for the output bit given in Fig. II.1, that $I(\phi|k,r) = I(\phi|k,i) = k^2$ and thus

$$I(\phi) = \sum_k k^2 (M_k^r + M_k^i). \tag{II.15}$$

At the same time the total quantum cost of all experiments is

$$T = \sum_k k (M_k^r + M_k^i). \tag{II.16}$$

The key insight here is that the relative dependence on k is different between T and I and this implies that the trend of the number of experimental runs $M_k^{r/i}$ as a function of k will affect the maximum rate of estimation. If we choose only $k = 1$ we see that $\delta^2 \geq \frac{1}{T}$ which is the sampling noise limit.

The biggest value for $I(\phi)$ for a given T is obtained when we choose a single largest possible $k = K$ so that $T = KM_K$ and $I(\phi) = K^2 M_K = T^2/M_K$ with $M_K = M_K^r + M_K^i$. This implies a Heisenberg limit, i.e. $\delta \geq cT^{-1}$ where c is some constant depending on M_K. If we however make the 'dense signal' choice, that is, $M_k^r = M_k^i = M$ for $k = 1, 2 \ldots K$, then

$$I(\phi) = \frac{M}{3} K(K+1)(1+2K), \tag{II.17}$$

while the total quantum cost is $T = MK(K+1)$. Increasing K with M fixed, we have to leading order in K that $\delta \geq I(\phi)^{-\frac{1}{2}} = \sqrt{\frac{3}{2}} M^{-\frac{1}{2}} K^{-\frac{3}{2}} = \sqrt{\frac{3}{2}} M^{\frac{1}{4}} T^{-\frac{3}{4}}$, which is the dense signal limit.

□

Remark: We note that a randomized version of the dense signal choice can potentially scale in near-Heisenberg-limited fashion. In this method, one would draw k at random from $1, \ldots, K$ and repeat this S times to generate random variables $k_1, \ldots, k_S$ and repeat experiments with fixed M for each such k_i. With the right choice of $S \times M = \text{polylog}(K)$, one can argue, using the Cramer-Rao lower bound analysis above, that the expected error $\mathbb{E}(\delta) \geq \frac{\text{polylog}(\mathbb{E}(T))}{\mathbb{E}(T)}$ where $\mathbb{E}(T)$ is the expected quantum cost. The algorithm in [153] uses such strategy with $M = 1$. Clearly, the sampling noise limit can be achieved by

choosing $k = 1$ in the circuits of Fig. II.1. However, one can ask whether the dense signal limit or the Heisenberg limit can also be achieved, in particular when we demand that the classical post-processing is computationally efficient, meaning that this processing is polynomial in the quantum cost T. For the dense signal limit one needs a classical method to process the estimates of $g(k)$ at $k = 1, \ldots, K$ to estimate $\tilde{\phi}$. Using perturbation theory in the noise, the matrix pencil method has been claimed to achieve this for a single eigenvalue [164].

Achieving the Heisenberg limit is non-trivial due to phase aliasing: the phase function $g(k)$ obtained by the experiments using U^k remains invariant if the phase ϕ is shifted by $\frac{2\pi}{k}$. This implies that a strategy of estimating ϕ from a single point $g(k)$ at large k will fail unless ϕ is already known to sit within a window of width $\frac{2\pi}{k}$. This issue is circumvented by sampling $g(k)$ at multiple orders $k = 2^d$ to 'gradually zoom in' on ϕ. To get Heisenberg scaling, one lets the number of samples M and thus the confidence, to depend on the order, so that the most significant bits of ϕ are determined with the highest confidence. Methods for doing this were first introduced in Ref. [131], and improved in Ref. [74] for the purpose of gate calibration. Here we state the result:

Algorithm II.1 (Heisenberg Algorithm For Single Eigenvalue Phase [74, 131]) *Given an targeted error $\delta > 0$, and numbers $\alpha, \gamma \in \mathbb{Z}^+$. The Heisenberg algorithm which outputs an estimate $\tilde{\phi}$ for ϕ proceeds as follows:*

1. *Fix $d_f = \lceil \log_2(1/\delta) \rceil$.*
2. *For $d = 0, 1 \ldots, d_f$:*
   a) *Use the* PFE *subroutine, Def. II.2, circuits in Fig. II.1, to obtain an estimate $\tilde{g}(k)$ of $g(k)$ for $k = 2^d$ using $M_d = \alpha + \gamma(d_f + 1 - d)$ repetitions of both experiments.*
   b) *Compute $\tilde{\theta}^{(d)} = \mathrm{Arg}[\tilde{g}(2^d)] \in [0, 2\pi)$*
   c) *If $d = 0$, set $\tilde{\phi}^{(0)} = \tilde{\theta}^{(0)}$.*
   d) *Else, set $\tilde{\phi}^{(d)}$ to be the unique value in the interval $[\tilde{\phi}^{(d-1)} - \frac{\pi}{2^d}, \tilde{\phi}^{(d-1)} + \frac{\pi}{2^d})$ (with periodic boundaries) such that*

$$2^d \tilde{\phi}^{(d)} = \tilde{\theta}^{(d)} \mod 2\pi. \tag{II.18}$$

3. *Return $\tilde{\phi} = \tilde{\phi}^{(d_f)}$ as an estimate for ϕ.*

It was proven in [74] that for some choices of α and γ the root-mean-square error δ on the final estimate $\phi^{(d_f)}$ is at most cT^{-1} for a constant c and total cost $T = 2\sum_{d=0}^{d_f} 2^d M_d$, thus reaching the Heisenberg limit.

One might consider the effect of experimental noise on these limits. The algorithm given in [74] was proven to be robust against noise that affected the estimation of any $g(k)$ by no more than $\frac{1}{\sqrt{8}}$. However, realistic noise tends to scale with the circuit depth, eventually breaking this bound. In the presence of a uniform depolarizing channel, it is possible to extend the above calculation of Fisher information to optimize the recovery of a single phase ϕ, however in the limit that $\delta \to 0$ only the sampling-noise limit can be obtained:

Lemma II.2
The (root-mean-square) error δ of an estimator $\tilde{\phi}$ of the eigenvalue phase ϕ employing the circuits in Fig. II.1 on an eigenstate of U in the presence of a pure depolarizing channel with a fixed failure probability $p = e^{-1/\tau}$ per iteration of U is bounded as

$$\delta \geq \frac{\tau^{-\frac{1}{2}} T^{-\frac{1}{2}}}{2e}. \tag{II.19}$$

Proof. A pure depolarizing channel sends the off-diagonal element of the reduced density matrix in Eq. (II.4) to $p^k g(k)$. This adjustment can be propagated directly through to the Fisher information (Eq. (II.15)), which becomes

$$I(\phi) = \sum_k e^{-2k/\tau} k^2 (M_k^r + M_k^i), \tag{II.20}$$

while the total quantum cost (Eq. II.16) remains the same. Let us consider the relative contribution to I versus the contribution to T of a single choice of k; if we write $I(\phi) = \sum_k I_k$ and $T = \sum_k T_k$, we have $I_k/T_k = ke^{-2k/\tau}$. Differentiating w.r.t. k and setting equal to zero yields

$$\frac{d}{dk}\frac{I_k}{T_k} = e^{-2k/\tau} - \frac{2k}{\tau} e^{-2k/\tau} = 0 \tag{II.21}$$

$$\to k = \frac{\tau}{2}. \tag{II.22}$$

Optimizing $I(\phi)$ with respect to T thus requires setting $k = \frac{\tau}{2}$ and increasing $M_k = M_k^r + M_k^i$, which yields

$$I(\phi) = \frac{\tau^2}{4e} M_k, T = \frac{\tau}{2} M_k. \tag{II.23}$$

Substituting this into the Cramér-Rao bound $\delta \geq I(\phi)^{-\frac{1}{2}}$ yields the desired result. □

Remark: Note that as we fix $T \sim \tau$ in the above, we in effect have Heisenberg-limited scaling in τ. However, if we treat τ as a constant this is

only the sampling noise limit as defined above. This result holds only for the simplest-possible noise case; more complicated noise is difficult to analyse, but numerical results show it may prevent estimation beyond some minimum value using standard techniques [136]. We assume herein that all circuits are noiseless (and will not use Lemma II.2 in the rest of this work).

II.3. Defining the task of multiple-eigenvalue phase estimation

In this section we define the goal of estimating multiple eigenvalue phases of some unitary U. When the input state $|\Psi\rangle$ is supported on multiple eigenstates, choosing a single k does not suffice, simply since knowing Eq. (II.2) at a single point k does not give a unique solution $\{A_j, \phi_j\}$ [136]. A simple way to circumvent this problem is thus to estimate 'densely': estimating $g(k)$ for all integers $0 \leq k \leq K$ would allow us to fit up to $O(K)$ (ϕ_j, A_j) pairs. However, this does not saturate the Heisenberg limit as shown in Theorem II.1, hence we need to come up with a different method.

Separate from this, the full eigenspectrum of an arbitrary N-qubit unitary U has up to 2^N unique values, making it impossible to describe in polynomial time in N. In addition, the spectral content of the input state $|\Psi\rangle$ could be very dense, with many eigenvalues clustered together instead of separated by gaps, and the overlap for these eigenvalues, A_j, could be sharply concentrated or uniformly spread. To deal with general input states, Ref. [138] thus formulated the quantum eigenvalue estimation problem (QEEP): instead of estimating individual phases, the focus is on estimating the spectral function $A(\phi)$ in Eq. (II.3) with some resolution. We recall the precise definition of this problem in App. II.A, Def. II.5. In our case, we focus on the case where the initial state only has a non-zero overlap with a small number n_ϕ of eigenvectors of U, and we want to estimate eigenphases corresponding to each of them. Here is our precise definition of the problem to be solved:

Definition II.4 (Multiple eigenvalue estimation problem)
Fix an error bound $\delta > 0$, an overlap bound $A > 0$. For a unitary U and state $|\Psi\rangle$, we assume that $A_j > A$ for exactly n_ϕ phases ϕ_j and $A_j = 0$ for all other phases so that $n_\phi \leq A^{-1}$. Let $g(k) = \sum_j A_j e^{ik\phi_j}$ be the phase function in Def. II.1 and assume access to the PFE quantum subroutine in Def. II.2 for any $k \in \mathbb{R}^+$, generating data $\mathbf{x}$. The task is to output a set $\{\tilde{\phi}_l\}$ of n_ϕ or fewer estimates of the phases $\{\phi_j\}$ such that, if we take the closest estimate

$\tilde{\phi}_j^{(\text{closest})}(\mathbf{x})$ *of each phase* ϕ_j *given the data* $\mathbf{x}$*,*

$$\tilde{\phi}_j^{(\text{closest})}(\mathbf{x}) = \arg\min_{\tilde{\phi}_l}(|\tilde{\phi}_l(\mathbf{x}) - \phi_j|_T), \tag{II.24}$$

the accuracy error

$$\delta_j = \sqrt{\sum_{\mathbf{x}} \mathbb{P}\left(\mathbf{x}|\{\phi_l, A_l\}\right) \left|\tilde{\phi}_j^{(\text{closest})}(\mathbf{x}) - \phi_j\right|_T^2}, \tag{II.25}$$

is bounded by $\delta_j \leq \delta$ *for all* $j = 1, \ldots, n_\phi$*.*

Remark: Def. II.4 allows us the freedom to assign a single estimate to multiple phases when calculating the final mean-squared-error.

II.3.1. Methods of dense signal phase estimation

Achieving the Heisenberg limit for multiple eigenvalues requires solving the problem considered in Def. II.4 with a total quantum cost $T = O(\delta^{-1})$. We intend to accomplish that goal with a multi-order estimation scheme. At each order d, we will use a data processing method to estimate multiple eigenphases $\theta_j^{(d)}$ of U^{k_d} to within some error ϵ, from data generated by PFE (analogous to step 2b of Alg. II.1). (We will stitch the estimates of the phases $\theta_j^{(d)}$ together to give Heisenberg-limited estimates of the corresponding ϕ_j in a manner similar to step 2d of Alg. II.1.) We can offload the estimation of $\theta_j^{(d)}$ to a subroutine; we will show later that we can afford a subroutine with superlinear scaling in ϵ as the final error δ in our multi-order scheme can be made arbitrarily small even at fixed ϵ. In this section we discuss the two subroutines that we will consider in this work: the matrix pencil method (Alg. II.3) first studied for QPE in Ref. [136], and the 'time series analysis' proposed in Ref. [138] to solve the QEEP problem mentioned above.

We detail our implementation of the matrix pencil method in Alg. II.3; this is a well-known algorithm in signal processing, that is known to achieve the dense-sampling limit for a single eigenvalue [163, 164]. However, bounding the error of the matrix pencil method in estimating many phases typically requires a minimal gap Δ between these phases, $\Delta = \min_{i \neq j} |\phi_i - \phi_j|_T$, and that we query the PFE to obtain estimates of $g(k)$ at $k > \frac{1}{\Delta}$. This is a proven necessary condition to estimate multiple ϕ_j to error $\epsilon \leq c\Delta$ [165] for some constant c. In our case, we need to allow for the case where we are estimating two phases $\theta_0^{(d)}$, $\theta_1^{(d)}$ to an error $\epsilon \geq |\theta_0^{(d)} - \theta_1^{(d)}|$. In principle this is not forbidden by the result of [165] (and numerical simulation confirms that this indeed works), but we do not know of a formal statement about the scaling

of the matrix pencil method in this situation. Instead, we opt for a different method for the purposes of forming a rigorous proof of the Heisenberg limit, and test the matrix pencil method in numerics only.

Ref. [138] proved rigorously the QEEP can be solved from the densely sampled signal $g(k)$ generated with the PFE in Def. II.2 using a 'time-series analysis' algorithm. We review the results of Ref. [138] in detail in App. II.A. However, the solution to the QEEP is an estimation $\tilde{A}(\phi)$ of a discretization of the spectral function $A(\phi)$ (Eq. (II.3)) rather than a set of estimates $\{\tilde{\theta}_j^{(d)}\}$. To use the time-series analysis algorithm as a subroutine in our multi-order phase estimation algorithm then requires converting from one form to the other. This is achieved by Alg. II.4, the Conservative QEEP Eigenvalue Extraction algorithm. This algorithm is designed so that its output fulfills the following guarantees whenever the time-series analysis algorithm succeeds (defined as $\|\tilde{A}(\phi) - A(\phi)\|_1 \leq \epsilon$)

Lemma II.3
Fix a confidence bound $0 < p < 1$, an overlap bound A, a number of phases $n_\theta < \frac{1}{A}$, and an error bound $0 < \epsilon < \frac{A}{3}$. Let $g(k) = \sum_j A_j e^{ik\theta_j}$ be the phase function for a unitary V, with $A_j > A$ for exactly n_θ phases θ_j, and $A_j = 0$ for all other phases. Let $\{\tilde{\theta}_l\}$ be a set of estimates of $\{\theta_j\}$ generated by Alg. II.4 with error ϵ and confidence bound p. With probability at least p, the following statements are true:

1. *For each phase θ_j with $A_j > 0$, there exists at least one estimate $\tilde{\theta}_l$ such that*
$$|\theta_j - \tilde{\theta}_l|_T \leq 2\epsilon.$$

2. *For each estimate $\tilde{\theta}_l$ there exists at least one phase θ_j with $A_j > 0$ such that*
$$|\theta_j - \tilde{\theta}_l|_T \leq 2\epsilon.$$

3. *The number of estimates $|\{\tilde{\theta}_l\}| \leq n_\theta$.*

See App II.A for a proof. The total quantum cost of Alg. II.4 is inherited directly from the cost of the time series analysis algorithm (as it involves no additional quantum circuitry), which is $O(\epsilon^{-6}|\log(1-p)|)$.

II.4. Multiple eigenvalues: multi-order estimation and the phase matching problem

To achieve Heisenberg-limited scaling for multiple phases, we combine the dense signal algorithms which can resolve multiple phases in the previous section with the single-phase Heisenberg limited algorithm, Algorithm II.1, which achieves the correct scaling.

A natural way to achieve such combination is to estimate phases $\{\theta_j^{(d)}\}$ of $V = U^{2^d}$ for multiple orders $d = 0, 1 \ldots, d_f$ via a dense signal method (e.g. Alg. II.4), and then combine them in the same manner as in Algorithm II.1.

If we would manage to get an estimate of $\theta_j^{(d)} = \phi_j 2^d$ at each order d with error ϵ *and* be able to combine these estimates in an unequivocal manner, then reaching the Heisenberg limit for multiple phases may be feasible. Note that the error in the final d_f^{th} estimate in this case would be $\delta \sim \epsilon/2^{d_f}$ with ϵ in Algorithm II.4. One could thus achieve arbitrarily small δ for fixed ϵ by making d_f arbitrarily large. This allows us to use (possibly non-optimal) routines such as the QEEP algorithm since the scaling with ϵ does not propagate to a scaling in δ for a sufficiently small ϵ.

However, the combination of phase information at different orders in the case of multiple phases may not be feasible when we use $V = U^{2^d}$ for increasing d as in Algorithm II.1. The reason is that if we have multiple phases, the previous order estimates provide sets of 'ballpark' intervals and it may not clear or unambiguous which interval to choose in order to convert a new estimate $\tilde{\theta}_j^{(d)}$ to an updated $\tilde{\phi}_j^{(d)}$ (as in step 2d of Algorithm II.1). We would like the choice of the next order to be such that this 'matching with a previous estimate' can be done unambiguously.

For this, we will estimate the multiple eigenphases of $V = U^{k_d}$ for $k_d = \prod_{d'=1}^{d} \kappa_{d'}$ with $\kappa_d \geq 2$ a, possibly non-integer, multiplier. For this algorithm to have a means of associating each dth order estimate $\tilde{\theta}_j^{(d)}$ with a single previous-order estimates $\tilde{\phi}_j^{(d-1)}$, we use an adaptive strategy for choosing the next multiplier κ_d in Alg. II.2. That is, the algorithm will determine a κ_d based on the estimates $\tilde{\phi}_j^{(d-1)}$ from the previous round. Although this scheme requires some classical processing of the experimental data before the experiment is finished, it is not an adaptive scheme in the same sense as iterative QPE [150], as we do not require feedback within the coherent lifetime of a single experiment.

The generalization from using U^{2^d} to U^{k_d} for $k_d \in \mathbb{R}^+$ presents one small additional complication. In order to prove bounds on the estimation at

each order we will require invoking Lemma II.1. However, this requires that our phases ϕ_j satisfy Eq. (II.7) (unless $k_d \in \mathbb{Z}^+$). If a phase ϕ_j does not satisfy Eq. (II.7), one can construct a situation where two corresponding estimates $\tilde{\phi}_j^{(d)}$ are found on either side of the branch cut at 2π, and where we cannot guarantee that our algorithm would choose the 'correct' one (without knowledge of the hidden ϕ_j). To solve this issue, we note that one may shift the phases of U by a constant χ by performing phase estimation on $Ue^{-i\chi}$ instead of U. This need not even be done on the quantum device, as one simply multiplies estimates of $g(k)$ by $e^{-ik\chi}$. As we assume the existence of only n_ϕ phases, we can always find some $Ue^{i\chi}$ with phases in some window $[\phi_{\min}, \phi_{\max})$ with $\phi_{\min} \geq \frac{\pi}{k}$, $\phi_{\max} \leq \frac{\pi(2\lfloor k \rfloor - 1)}{k}$ when $k \geq 3n_\phi$. This will allow us to invoke Lem. II.1 to match estimates of eigenphases of $Ue^{i\chi}$ and estimates of eigenphases of $(Ue^{i\chi})^k$ as we require. We also note that the above issue can be circumvented when $U = e^{iHt}$ by a suitable choice of t.

II.4.1. Heisenberg-limited algorithm for multiple phases

We now describe our Heisenberg-limited phase estimation algorithm. This algorithm targets a final error $\delta = O(\delta_c)$, where δ_c is a fixed input to the algorithm itself (We will calculate the constant of proportionality in the proof of Theorem II.2). The Heisenberg limit will be achieved by making this δ_c smaller while keeping the error ϵ of the phase extraction subroutine, Alg. II.4, constant.

Algorithm II.2
[Adaptive multi-order phase estimation algorithm] We assume access to the conservative QEEP eigenvalue extraction algorithm, Alg. II.4 for a unitary $V = U^k$ (for arbitrary $k \in \mathbb{R}^+$), and an initial state $|\Psi\rangle$. Fix a final error δ_c, an overlap bound A, a number of phases $n_\phi \leq A^{-1}$, and error parameters ϵ_0 and ϵ bounded as

$$\epsilon_0 \leq \epsilon_{\text{crit},0} \equiv \frac{2\pi}{300 n_\phi^4} \tag{II.26}$$

and

$$\epsilon \leq \epsilon_{\text{crit}} \equiv \frac{2\pi}{300 n_\phi^2}. \tag{II.27}$$

Let the confidence parameter p_d be

$$p_d = 1 - e^{-\alpha} \left(\frac{k_d \delta_c}{\pi} \right)^\gamma, \tag{II.28}$$

given some real numbers $\alpha > 0$, $\gamma > 2$ and k_d to be chosen below. The algorithm proceeds as follows:

1. *Let $d = 0$ and $k_{d=0} = 1$. Use Alg. II.4 to find a set of first estimates $\{\tilde{\phi}_j^{(0)}\}$ of eigenvalues of U with error parameter ϵ_0 in Eq.* (II.26)*, overlap bound A, and confidence $p_{d=0}$ in Eq.* (II.28)*. If this set is empty or has more than n_ϕ elements, return $\{0\}$ (this is a failure mode).*

2. *Find the point $\zeta \in [0, 2\pi)$ defined by*

$$\zeta = \underset{\zeta' \in [0,2\pi]}{\arg\max} \min_j \left| \tilde{\phi}_j^{(0)} - \zeta' \right|_T , \tag{II.29}$$

i.e. ζ is the midway point in the largest gap between the phase estimates $\tilde{\phi}_j^{(0)}$. Let

$$d_\zeta = \min_j \left| \tilde{\phi}_j^{(0)} - \zeta \right|_T , \tag{II.30}$$

i.e. d_ζ is half the size of the largest gap. Shift the unitary $U \to Ue^{-i(\zeta + \frac{1}{2}d_\zeta - 8\epsilon_0)}$, $\tilde{\phi}_j^{(0)} \to \tilde{\phi}_j^{(0)} - \zeta - d_\zeta/2 + 8\epsilon_0 \mod 2\pi$.

3. *Choose $\kappa_1 = k_1$ with $k_1 \in [3n_\phi, 3n_\phi + 1]$ such that for all $\tilde{\phi}_j^{(0)} \neq \tilde{\phi}_l^{(0)}$, either*

$$|\tilde{\phi}_j^{(0)} k_1 - \tilde{\phi}_l^{(0)} k_1|_T > 4\epsilon_0(1 + k_1). \tag{II.31}$$

or

$$|\tilde{\phi}_j^{(0)} - \tilde{\phi}_l^{(0)}|_T < \frac{\pi}{k_1}. \tag{II.32}$$

4. *While $k_d < \frac{2\epsilon}{\delta_c}$:*

   a) *Set $d \to d + 1$.*

   b) *Use Alg. II.4 to find a set of estimates $\{\tilde{\theta}_l^{(d)}\}$ of eigenvalues of $V = U^{k_d}$ with error parameter ϵ in Eq.* (II.27)*, overlap bound A, and confidence p_d in Eq.* (II.28)*.*

   c) *If there exists some $\tilde{\phi}_j^{(d-1)}$ such that*

$$\min_l |k_d \tilde{\phi}_j^{(d-1)} - \tilde{\theta}_l^{(d)}|_T > 2\epsilon(1 + \kappa_d), \tag{II.33}$$

   or there exists some $\tilde{\theta}_l^{(d)}$ such that

$$\min_j |k_d \tilde{\phi}_j^{(d-1)} - \tilde{\theta}_l^{(d)}|_T > 2\epsilon(1 + \kappa_d), \tag{II.34}$$

   or the number of estimates $|\{\tilde{\theta}_l^{(d)}\}| > n_\phi$, return $\{\tilde{\phi}_j^{(d-1)} + \zeta + d_\zeta/2 - 8\epsilon_0 \mod 2\pi\}$. This is a failure mode.

d) If not, for each $\tilde{\theta}_l^{(d)}$*, find the estimate* $\tilde{\phi}_j^{(d-1)}$ *and an integer* $n \in [0, k_d)$ *which minimizes*

$$|\tilde{\phi}_j^{(d-1)} - (\tilde{\theta}_l^{(d)} + 2\pi n)/k_d|_T, \tag{II.35}$$

and set $\{\tilde{\phi}_l^{(d)}\} = \{(\tilde{\theta}_l^{(d)} + 2\pi n)/k_d\}$.

e) If any $\tilde{\phi}_j^{(d)} \in [0, \frac{\pi}{k_d}) \cup (\frac{\pi(2\lfloor k_d \rfloor - 1)}{k_d}, 2\pi]$*, return* $\{\tilde{\phi}_j^{(d-1)} + \zeta + d_\zeta/2 - 8\epsilon_0 \mod 2\pi\}$*. This is a failure mode.*

f) Choose the multiplier $\kappa_{d+1} \in [2, 3]$ *such that for all* $\tilde{\phi}_j^{(d)} \neq \tilde{\phi}_l^{(d)}$*, either*

$$|\tilde{\phi}_j^{(d)} k_d \kappa_{d+1} - \tilde{\phi}_l^{(d)} k_d \kappa_{d+1}|_T > 4\epsilon(1 + \kappa_{d+1}). \tag{II.36}$$

or

$$|\tilde{\phi}_j^{(d)} - \tilde{\phi}_l^{(d)}|_T < \frac{\pi - 2\epsilon(1 + \kappa_{d+1})}{k_d \kappa_{d+1}}, \tag{II.37}$$

and set $k_{d+1} = k_d \kappa_{d+1}$.

5. Return $\{\tilde{\phi}_j^{(d)} + \zeta + d_\zeta/2 - 8\epsilon_0 \mod 2\pi\}$.

In principle the first few orders d could be skipped given accurate prior knowledge of our phases ϕ_j. However, as the largest circuits are executed at the latter d values, this will only change the constant factor of the algorithm, rather than the asymptotic scaling with δ.

In the rest of this section, we prove that Alg. II.2 can achieve the Heisenberg limit. The first use of the QEEP subroutine requires a potentially smaller error parameter (ϵ_0, bounded by Eq. (II.26)) than subsequent uses (where ϵ needs to be only bounded by Eq. (II.27)). (Invoking Alg. II.4 requires that the error parameters ϵ and ϵ_0 are at most $A/3 \leq 1/(3n_\phi)$, which is fulfilled by both bounds.) This relates to a technical issue: we require $k_1 \geq 3n_\phi$ in order for Lemma II.7 in Appendix II.C and thus Lemma II.1 to apply. For later rounds $k_d \geq 3n_\phi$ automatically, and the multiplier κ_d is no longer constrained, which indirectly allows us to relax the region of valid choices for ϵ. The first step in proving the performance of Algorithm II.2 is to show that the multipliers can be chosen in the first (step 3 in Alg. II.2), and subsequent rounds (step 4f in Alg. II.2), which obey the desired conditions. This is accomplished by the following Lemma which is proved in Appendix II.B.

Lemma II.4
Let $\{\tilde{\phi}_j^{(0)}\} \in [0, 2\pi)$ *be a set of at most* n_ϕ *phases. Assuming Eq.* (II.26)*, for a randomly chosen* $k_1 \in [3n_\phi, 3n_\phi + 1]$ *with probability at least* 1/2*,*

either Eq. (II.31) *or Eq.* (II.32) *holds for all* $\tilde{\phi}_j^{(0)} \neq \tilde{\phi}_l^{(0)}$. *Fix a* k_d. *Let* $\{\tilde{\phi}_j^{(d)}\} \in [0, 2\pi)$ *be a set of at most* n_ϕ *phases. Assuming Eq.* (II.27)*, for a randomly chosen* $\kappa_{d+1} \in [2, 3]$ *with probability at least* $3/4$*, either Eq.* (II.36) *or Eq.* (II.37) *holds for all* $\tilde{\phi}_l^{(d)} \neq \tilde{\phi}_j^{(d)}$.

Remarks: The probability with which a multiplier can be found which obeys the desired property is rather arbitrary in this Lemma and can be increased by choosing a smaller ϵ. Note that it is easy to verify whether for a randomly chosen multiplier the desired conditions hold or not. The validity of this Lemma importantly does not depend on whether the phase estimates are actually accurate, it only depends on the number of phases n_ϕ. In practice, we do not generate a random multiplier κ_{d+1} through this Lemma, but simply exhaustively search for a valid κ_{d+1} starting at the maximal value, see Section II.5.

The reason to adaptively choose the multiplier κ_{d+1} for $d = 0, 1, \ldots$ is that two (estimated) phases in principle need to lead to separate estimates at the next order: this is expressed in Eq. (II.36). An exception to this occurs when the (estimated) phases are still close enough, as in Eq. (II.37), so that their next-order refined estimates could merge at the next order, see Fig. II.2. Phase estimates can thus split and merge over the multiple orders. They split when sufficient accuracy is available at the next order to distinguish them, they can stay or are allowed to merge when such accuracy is not yet needed at the given order.

In what follows below we will assume, just for simplicity of the proof, that the error parameter ϵ is bounded by $\epsilon_{\text{crit},0}$ in Eq. (II.26) for all rounds, and ϵ is the same for all rounds, including the first one.

II.4.2. Bounding the error with and without failures

In this section we state and prove the two key intermediate lemmas, Lemma II.5 and Lemma II.6 on our way towards proving that Alg. II.2 reaches the Heisenberg-limit. Together, these lemmas allow us to bound the error in Alg. II.2, –assuming that the phase extraction subroutine succeeds for the first d rounds–, to within $O(\epsilon/k_d)$.

These Lemmas deal with the issue of 'aliasing' or the correct matching of new estimates with older estimates which is solved by the specific choice of κ_{d+1} in step 4f of Alg. II.2, see also Fig. II.2. It is important to note that there is no 1-1 relation between these estimates and the actual phases since the number of estimates is at most the number of phases.

Let d_f be the last order executed in Alg. II.2, i.e. the last order for which we go through step 4b, construct the estimates $\{\tilde{\theta}_l^{(d_f)}\}$ and pass the tests at step

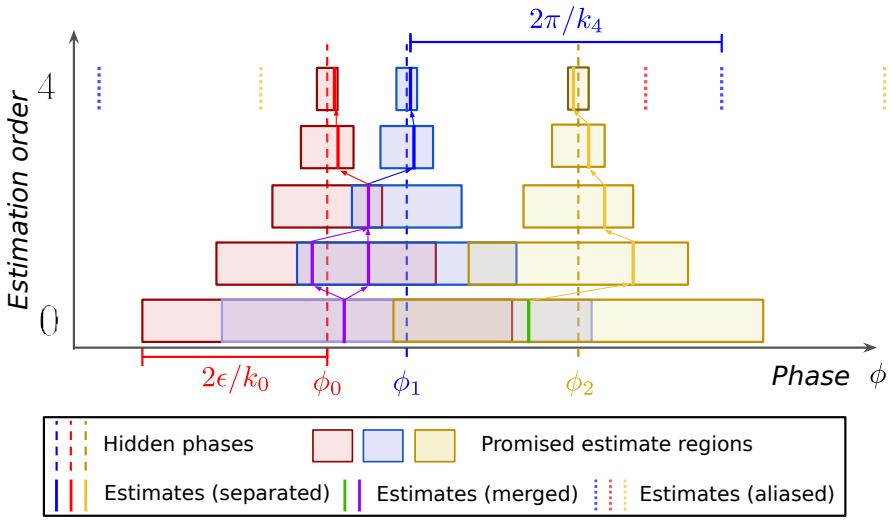


Figure II.2.: Schematic of the execution of Alg. II.2 to estimate three hidden phases, ϕ_0, ϕ_1 and ϕ_2 (dashed lines). The algorithm progresses from bottom to top as the estimation order d increases. At each order d, the phase extraction subroutine (Alg. II.4) promises to return an estimate $\tilde{\theta}^{(d)}_{\to j}$ of each $\theta^{(d)}_j = k_d\phi_j$ mod 2π that corresponds to an estimate $\tilde{\phi}^{(d)}_{\to j}$ (solid lines) lying within the promised estimate region about ϕ_j (coloured boxes). By matching phases at subsequent orders (arrows), the algorithm is able to converge to an ever-more accurate estimate of each ϕ_j. The phase extraction subroutine only promises that each region will contain at least one phase (and that the total number of estimates at each order is bounded by n_ϕ) - when two regions overlap, the subroutine may merge the phases to give a single estimate (green and purple lines). Estimates at subsequent orders may continue to separate and even re-merge until the regions separate, at which point the algorithm promises with high confidence a precise estimate for each hidden phase. The estimate $\tilde{\phi}^{(d)}_{\to j}$ at each order is only known mod $(2\pi)/k_d$, leading to a set of potential aliases (dotted lines at $d = 4$) for each phase. We do not know a priori which alias is correct, and must rely on the fact that the true estimate $\tilde{\phi}^{(d)}_{\to j}$ needs to be close to a previous estimate $\tilde{\phi}^{(d-1)}_{\to j}$. By carefully choosing each k_d, we can guarantee that no alias will satisfy this condition (so long as Alg. II.4 succeeds), and our phase matching will be unambiguous.

4c and 4e and output $\{\tilde{\phi}^{(d_f)}_l\}$. When none of the failure modes is encountered, d_f is set by the first k_d such that $k_d \geq \frac{\epsilon}{\delta_c}$ (since the next $k_{d+1} \geq 2\epsilon/\delta_c$ as

$\kappa_{d+1} \geq 2$). Since $\kappa_d \geq 2$, we observe that

$$d_f \leq \log_2 \left(\frac{2\epsilon}{\delta_c} \right). \tag{II.38}$$

In Corollary II.1 we argue that when the QEEP subroutines, Alg. II.4, succeed up to order d_f, we indeed never exit via these failure modes.

Lemma II.5
If each invocation of the QEEP subroutine, Alg. II.4, succeeds in Alg. II.2 up to order d_f, then in this last round d_f in step 4d it holds that

- *(Property 1a) For every phase ϕ_j there exists an estimate $\tilde{\phi}_l^{(d_f)}$ such that*
$$|\phi_j - \tilde{\phi}_l^{(d_f)}|_T \leq 2\epsilon/k_{d_f}.$$
- *(Property 1b) For every estimate $\tilde{\phi}_l^{(d_f)}$ there exists a phase ϕ_j such that*
$$|\phi_j - \tilde{\phi}_l^{(d_f)}|_T \leq 2\epsilon/k_{d_f}.$$

Proof. We prove this Lemma by induction. Consider the first round $d = 0$ ($k_{d=0} = 1$), i.e. step 1 of Alg. II.2. If the QEEP subroutine, Alg. II.4, succeeds (with probability p_0) then Lemma II.3 holds, namely for each ϕ_j there exists an estimate $\tilde{\phi}_l^{(0)}$ such that

$$|\phi_j - \tilde{\phi}_l^{(0)}|_T \leq 2\epsilon. \tag{II.39}$$

and for each estimate $\tilde{\phi}_l^{(0)}$ there exists at least one ϕ_j such that

$$|\phi_j - \tilde{\phi}_l^{(0)}|_T \leq 2\epsilon. \tag{II.40}$$

Hence the statement to be proven holds at $d = 0$. Now consider step 2 of Alg. II.2 and invoke Lemma II.7 for which the assumptions are fulfilled by Eqs. (II.39),(II.40). This implies that with the choice of $k_1 \geq 3n_\phi$ in step 3 in Alg. II.2 the shifted phases and their 0th-order estimates obey the technical condition in Lemma II.1 and we can use Eq. (II.8). In the next steps we work with these shifted phases but for simplicity we don't use any new notation and refer to them as ϕ_j and estimates $\tilde{\phi}_j^{(d)}$ etc.

Now assume the statement to be proven holds at order d i.e. let $\{\tilde{\phi}_l^{(d)}\}$ be a set of at most n_ϕ estimates of the phases $\{\phi_j\}$ with

- (Assumption 1a) For every phase ϕ_j there exists an estimate $\tilde{\phi}_l^{(d)}$ such

that

$$|\phi_j - \tilde{\phi}_l^{(d)}|_T \leq 2\epsilon/k_d.$$

- (Assumption 1b) For every estimate $\tilde{\phi}_l^{(d)}$ there exists a phase ϕ_j such that

$$|\phi_j - \tilde{\phi}_l^{(d)}|_T \leq 2\epsilon/k_d.$$

Note that these assumptions certainly imply that one can apply Lemma II.1 to the estimates $\tilde{\phi}_l^{(d)}$. That is, given that the real phases ϕ_j are $2\epsilon/k_d$ close to these estimates and that the (shifted) ϕ_j obey Eq. (II.109), it implies that Eq. (II.8) can be used with $k \geq 3n_\phi$ (which is the case for all rounds $d \geq 1$).

We consider the QEEP subroutine, Alg. II.4, with a given choice of κ_d obeying the conditions in step 3 (for $d = 1$) and step 4f (for higher d), executed in step 4b with confidence p_d. In the math below we refer to conditions on $\kappa_{d>1}$, namely Eq. (II.36) and Eq. (II.37), but the conditions on κ_1 in Eq. (II.31) and Eq. (II.32) are of identical form (so we don't make a separate argument for the $d = 0 \to d = 1$ induction step).

Let thus $\{\tilde{\theta}_l^{(d+1)}\}$ be a set of estimates of the eigenphases $\{\theta_j^{(d+1)}\}$ of $U^{k_{d+1}}$ corresponding to the set $\{\phi_j\}$, that is,

$$\theta_j^{(d+1)} = k_{d+1}\phi_j \mod 2\pi. \tag{II.41}$$

and $k_{d+1} = k_d\kappa_{d+1}$. By assuming that Alg. II.4 succeeds we can invoke Lemma II.3, namely

- (Assumption 2a) For every phase $\theta_j^{(d+1)}$ there exists an estimate $\tilde{\theta}_l^{(d+1)}$ such that

$$|\theta_j^{(d+1)} - \tilde{\theta}_l^{(d+1)}|_T \leq 2\epsilon.$$

- (Assumption 2b) For every estimate $\tilde{\theta}_l^{(d+1)}$ there exists a phase $\theta_j^{(d+1)}$ such that

$$|\theta_j^{(d+1)} - \tilde{\theta}_l^{(d+1)}|_T \leq 2\epsilon.$$

To prove the induction step, we thus need to show that the set $\tilde{\phi}_j^{(d+1)}$ generated by step 4d of Alg. II.2 satisfies the following two properties

- (Property 1a) For every phase ϕ_j there exists an estimate $\tilde{\phi}_l^{(d+1)}$ such that

$$|\phi_j - \tilde{\phi}_l^{(d+1)}|_T \leq 2\epsilon/k_{d+1}.$$

- (Property 1b) For every estimate $\tilde{\phi}_l^{(d+1)}$ there exists a phase ϕ_j such that

$$|\phi_j - \tilde{\phi}_l^{(d+1)}|_T \leq 2\epsilon/k_{d+1}.$$

First consider Assumption 2a. Assumption 2a implies that for every phase ϕ_j there exists a $\tilde{\theta}_l^{(d+1)}$ such that

$$|k_{d+1}\phi_j - \tilde{\theta}_l^{(d+1)}|_T \leq 2\epsilon. \tag{II.42}$$

In this proof we will use the label $l =\!\!\rightarrow j$ for this $\tilde{\theta}_l^{(d+1)}$ associated with ϕ_j. Thus, also using Eq. (II.8), Eq. (II.42) is equivalent to

$$\min_{n\in\{0,\ldots,\lfloor k_{d+1}\rfloor-1\}} |\phi_j - (\tilde{\theta}_{\rightarrow j}^{(d+1)}+2\pi n)/k_{d+1}|_T \leq \frac{2\epsilon}{k_{d+1}}, \tag{II.43}$$

with $n_{j,\rightarrow j}^{\text{ideal}}$ the integer which achieves the minimum, i.e.

$$n_{j,\rightarrow j}^{\text{ideal}} = \operatorname*{arg\,min}_{n\in\{0,\ldots,\lfloor k_{d+1}\rfloor-1\}} \Big\{ |\phi_j - (\tilde{\theta}_{\rightarrow j}^{(d+1)} + 2\pi n)/k_{d+1}|_T \Big\}. \tag{II.44}$$

Similarly, by Assumption 1a, there is some $\tilde{\phi}_{l=\rightarrow j}^{(d)}$ which is $2\epsilon/k_d$-close to ϕ_j, again using a label which shows this association.

Consider the optimization in Eq. (II.35) at step 4d in Alg. II.2. We define

$$\xi_l = \min_j \xi_{j,l}, \tag{II.45}$$

$$\xi_{j,l} \equiv \left|\tilde{\phi}_j^{(d)} - (\tilde{\theta}_l^{(d+1)} + 2\pi n_{j,l})/k_{d+1}\right|_T \tag{II.46}$$

with

$$n_{j,l} = \operatorname*{arg\,min}_{n\in\{0,\ldots,\lfloor k_{d+1}\rfloor-1\}} \Big\{ \left|\tilde{\phi}_j^{(d)} - (\tilde{\theta}_l^{(d+1)} + 2\pi n)/k_{d+1}\right|_T \Big\}, \tag{II.47}$$

$$n_l = \operatorname*{arg\,min}_{n_{j,l}} \xi_{j,l}. \tag{II.48}$$

The goal is thus to prove that for each ϕ_j, using the corresponding $\tilde{\theta}_{\rightarrow j}^{(d+1)}$, we have $n_{\rightarrow j} = n_{j,\rightarrow j}^{\text{ideal}}$ which directly implies Property 1a.

We can bound using Eq. (II.6) and then Eq. (II.8), Assumptions 1a and 2a and the optimality of $n_{\rightarrow j,\rightarrow j}$,

$$\xi_{\rightarrow j,\rightarrow j}$$

$$
\begin{aligned}
&= \left|\tilde{\phi}^{(d)}_{\to j} - (\tilde{\theta}^{(d+1)}_{\to j} + 2\pi n_{\to j,\to j})/k_{d+1}\right|_T \\
&\leq \left|\tilde{\phi}^{(d)}_{\to j} - \phi_j\right|_T \\
&\quad + \left|\phi_j - (\tilde{\theta}^{(d+1)}_{\to j} + 2\pi n^{\text{ideal}}_{j,\to j})/k_{d+1}\right|_T \\
&\leq \frac{2\epsilon}{k_d} + \frac{1}{k_{d+1}}\left|k_{d+1}\phi_j - \tilde{\theta}^{(d+1)}_{\to j}\right|_T \\
&= \frac{2\epsilon}{k_d} + \frac{1}{k_{d+1}}\left|\theta^{(d+1)}_j - \tilde{\theta}^{(d+1)}_{\to j}\right|_T \\
&\leq \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}}.
\end{aligned} \tag{II.49}
$$

Now if Eq. (II.36) holds for some other $m \neq\!\!\to j$, we claim on the other hand that

$$
\xi_{m,\to j} > \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}}, \tag{II.50}
$$

hence matching $\theta^{(d+1)}_{\to j}$ with such $\tilde{\phi}^{(d)}_m$, with $m \neq\!\!\to j$ is non-optimal and will not be chosen in the Algorithm. To indeed see that Eq. (II.36) implies Eq. (II.50), we can calculate

$$
\begin{aligned}
\frac{4\epsilon(1+\kappa_{d+1})}{k_{d+1}} &< \frac{1}{k_{d+1}}\left|k_{d+1}\tilde{\phi}^{(d)}_{\to j} - k_{d+1}\tilde{\phi}^{(d)}_m\right|_T \\
&\leq \frac{1}{k_{d+1}}\left|k_{d+1}\tilde{\phi}^{(d)}_{\to j} - \tilde{\theta}^{(d+1)}_{\to j}\right|_T \\
&\quad + \frac{1}{k_{d+1}}\left|\tilde{\theta}^{(d+1)}_{\to j} - k_{d+1}\tilde{\phi}^{(d)}_m\right|_T \\
&\leq \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}} + \min_{n\in\{0,\ldots,\lfloor k_{d+1}\rfloor-1\}}\Big\{ \\
&\qquad \left|\tilde{\phi}^{(d)}_m - (\tilde{\theta}^{(d+1)}_{\to j} + 2\pi n)/k_{d+1}\right|_T\Big\}
\end{aligned} \tag{II.51}
$$

$$
\begin{aligned}
&\Rightarrow \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}} < \\
&\min_{n\in\{0,\ldots,\lfloor k_{d+1}\rfloor-1\}}\left|\tilde{\phi}^{(d)}_m - (\tilde{\theta}^{(d+1)}_{\to j} + 2\pi n)/k_{d+1}\right|_T \\
&\qquad = \xi_{m,\to j}.
\end{aligned} \tag{II.52}
$$

Here we have used that Eq. (II.8) holds for the estimate $\tilde{\phi}^{(d)}_m$.

Alternatively, for those $m \neq\!\!\to j$ for which Eq. (II.37) holds, we claim that

$$
n_{m,\to j} = n_{\to j,\to j}, \tag{II.53}
$$

hence for those $\tilde{\phi}_m^{(d)} \neq \tilde{\phi}_{\to j}^{(d)}$ the algorithm produces a single new estimate equal to $(\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1}$.

To see that Eq. (II.37) implies Eq. (II.53) indeed, note that it is sufficient to prove that

$$\left|\tilde{\phi}_m^{(d)} - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1}\right|_T < \frac{\pi}{k_{d+1}}, \tag{II.54}$$

as one can then show that for $n' \neq n_{\to j,\to j} \in \{0, \ldots, \lfloor k_{d+1} \rfloor - 1\}$ that

$$\begin{aligned}
\frac{2\pi}{k_{d+1}} &\leq \left|\frac{2\pi}{k_{d+1}}(n_{\to j,\to j} - n')\right|_T \\
&= |(\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1} \\
&\qquad - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n')/k_{d+1}|_T \\
&\leq \left|(\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1} - \tilde{\phi}_m^{(d)}\right|_T \\
&\qquad + \left|\tilde{\phi}_m^{(d)} - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n')/k_{d+1}\right|_T \\
&< \frac{\pi}{k_{d+1}} + \left|\tilde{\phi}_m^{(d)} - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n')/k_{d+1}\right|_T \qquad \text{(II.55)} \\
\Rightarrow \frac{\pi}{k_{d+1}} &< \left|\tilde{\phi}_m^{(d)} - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n')/k_{d+1}\right|_T, \qquad \text{(II.56)}
\end{aligned}$$

so $n_{\to j,\to j}$ is optimal. Using Eq. (II.6), Eq. (II.37) and Eq. (II.60), we can prove Eq. (II.54) since

$$\begin{aligned}
&\left|\tilde{\phi}_m^{(d)} - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1}\right|_T \\
&\leq \left|\tilde{\phi}_m^{(d)} - \tilde{\phi}_{\to j}^{(d)}\right| \\
&\quad + \left|\tilde{\phi}_{\to j}^{(d)} - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1}\right|_T \\
&< \frac{\pi - 2\epsilon(1+\kappa_{d+1})}{k_{d+1}} + \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}} \\
&\qquad = \frac{\pi}{k_{d+1}}. \qquad \text{(II.57)}
\end{aligned}$$

We have thus shown that for each ϕ_j, there is a $\tilde{\theta}_{\to j}^{(d+1)}$, such that step 4d will output $(\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1}$ with $n_{\to j,\to j}$ defined in Eq. (II.48), related to the previous order estimate $\phi_{\to j}^{(d)}$ which was already close to $\phi_j^{(d)}$. The last step is to show that $n_{\to j,\to j} = n_{j,\to j}^{\text{ideal}}$ using Property 1a. It holds that

$$|\phi_j - (\tilde{\theta}_{\to j}^{(d+1)} + 2\pi n_{\to j,\to j})/k_{d+1}|_T$$

$$\begin{aligned}
&\leq |\phi_j - \tilde{\phi}^{(d)}_{\to j}|_T \\
&\quad + |\tilde{\phi}^{(d)}_{\to j} - (\tilde{\theta}^{(d+1)}_{\to j} + 2\pi n_{\to j,\to j})/k_{d+1}|_T \\
&\leq \frac{2\epsilon}{k_d} + \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}} < \frac{\pi}{k_{d+1}},
\end{aligned} \tag{II.58}$$

where we used that $4\epsilon(\kappa_{d+1}+1) < \pi$. Indeed for $d > 1$, $\epsilon < \frac{\pi}{16}$ ($\kappa_{d+1} \leq 3$) and for $d = 0$, $\epsilon < \frac{\pi}{4(3n_\phi+2)}$ ($\kappa_1 \leq 3n_\phi + 1$) given the upper bounds on ϵ in Eqs. (II.27) and (II.26). This implies through the same argument as in Eq. (II.56) that $n^{\text{ideal}}_{j,\to j}$ achieving the minimum in Eq. (II.43) equals $n_{\to j,\to j}$ and hence we obtain Property 1a.

Now let's prove Property 1b. Given a $\tilde{\theta}^{(d+1)}_l$, let $\phi_{j=\to l}$ be the real phase for which, by Assumption 2b, it holds that

$$\begin{aligned}
&|k_{d+1}\phi_{\to l} - \tilde{\theta}^{(d+1)}_l|_T \\
&\qquad = k_{d+1}|\phi_{\to l} - (\tilde{\theta}^{(d+1)}_l + 2\pi n^{\text{ideal}}_{\to l,l})/k_{d+1}|_T \\
&\qquad\qquad \leq 2\epsilon.
\end{aligned} \tag{II.59}$$

To prove Property 1b, we need to show that $n_l = n^{\text{ideal}}_{\to l,l}$ with n_l defined in Eq. (II.48). Let also $\tilde{\phi}^{(d)}_{\to j}$ be the previous order $2\epsilon/k_d$-close estimate to $\phi_{j=\to l}$ by Assumption 1a. The idea is that $\tilde{\phi}^{(d)}_{\to j} = \tilde{\phi}^{(d)}_{\to(\to l)}$ is matched to $\tilde{\theta}^{(d+1)}_l$ in the optimization step of the algorithm, so that this leads to a better estimate for the phase $\phi_{j=\to l}$.

Given a $\tilde{\theta}^{(d+1)}_l$ we can deduce, as before, that

$$\begin{aligned}
\xi_{\to j,l} &= \left|\tilde{\phi}^{(d)}_{\to j} - (\tilde{\theta}^{(d+1)}_l + 2\pi n_{\to j,l})/k_{d+1}\right|_T \\
&\leq \left|\tilde{\phi}^{(d)}_{\to j} - \phi_{\to l}\right|_T \\
&\quad + \left|\phi_{\to l} - (\tilde{\theta}^{(d+1)}_l + 2\pi n^{\text{ideal}}_{\to l,l})/k_{d+1}\right|_T \\
&\leq \frac{2\epsilon}{k_d} + \frac{1}{k_{d+1}}\left|k_{d+1}\phi_{\to l} - \tilde{\theta}^{(d+1)}_l\right|_T \\
&= \frac{2\epsilon}{k_d} + \frac{1}{k_{d+1}}\left|\theta^{(d+1)}_{\to l} - \tilde{\theta}^{(d+1)}_l\right|_T \\
&\leq \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}}.
\end{aligned} \tag{II.60}$$

Using previous arguments, all other $\xi_{m,l}$ are either larger or give the same integer $n_{\to j,l}$ and thus $n_l = n_{\to j,l}$. In addition, we can bound, using this

equality and Assumption 1a

$$\begin{aligned}|\phi_{\to l} - (\tilde{\theta}_l^{(d+1)} + 2\pi n_{\to j,l})/k_{d+1}|_T \\ \leq |\phi_{\to l} - \tilde{\phi}_{\to j}^{(d)}|_T + |\tilde{\phi}_{\to j}^{(d)} - (\tilde{\theta}_l^{(d+1)} + 2\pi n_{\to j,l})/k_{d+1}|_T \\ \leq \frac{2\epsilon}{k_d} + \frac{2\epsilon(1+\kappa_{d+1})}{k_{d+1}} \\ < \frac{\pi}{k_{d+1}}, \quad \text{(II.61)}\end{aligned}$$

implying that $n_l = n_{\to j,l} = n_{\to l,l}^{\text{ideal}}$ as desired. □

Algorithm II.2 has a few failure modes, namely steps 1, 4c and 4e where we exit and return an estimate of lower order. Arguments in Lemma II.5 show that these failure modes are only encountered when the QEEP subroutine, Alg. II.4, fails at some order. 'Failure' here is not complete failure; regardless of whether the algorithm fails, it will return a set of estimates $\tilde{\phi}_j$, and the error in these estimates will contribute to Eq. II.25. To achieve the Heisenberg limit we must make sure that both the probability of failure is small, and that the estimates $\tilde{\phi}_j$ from a failed instance of the algorithm still lie close to the true values ϕ_j to minimize their contribution to Eq. (II.25). The probability p_d with which the QEEP subroutine succeeds at the dth order is bounded by the parameters α, γ, given as an input to Alg. II.2 (Eq. II.28). The probability that this achieves for up to and including the dth order is bounded by $\prod_{d'=0}^{d} p_{d'}$. It is crucial for the success of our algorithm that we do not encounter any exit modes other than those mentioned above, which we can now prove given the machinery developed in Lem. II.5.

Corollary II.1
If each invocation of the QEEP subroutine, Alg. II.4, succeeds in Alg. II.2, we never exit at step 1, 4c or 4e.

Proof. Consider step 1 of Alg. II.2 applying Alg. II.4 which obeys Lemma II.3, showing that success of Alg. II.4 implies that the number of phases is at most n_ϕ. By assumption there is at least one phase with $A_j > 0$, and hence success means that Alg. II.4 cannot return the empty set due to statement 1. of Lemma II.3. Hence if Alg. II.4 succeeds we do not exit at step 1 of Alg. II.2. Now consider step 4c of Alg. II.2: again success of Alg. II.4 implies that the number of estimates does not exceed n_ϕ. Consider Eq. (II.33) and Eq. (II.34); we wish to show that these will not hold if Alg. II.4 succeeds up to order $d-1$. If Alg. II.4 succeeds up to order $d-1$, the phase estimates $\tilde{\phi}_j^{(d-1)}$ obey Eq. (II.8) for k_d, hence the condition in Eq. (II.34) equals, for

each $\tilde{\theta}_l^{(d)}$

$$\min_j \min_n |\tilde{\phi}_j^{(d-1)} - (\tilde{\theta}_l^{(d)} + 2\pi n)/k_d|_T > \frac{2\epsilon(1+\kappa_d)}{k_d}, \tag{II.62}$$

and we argued previously, via induction, that this does not happen when Alg. II.4 succeeds up to order d, as $\min_j \xi_{j,l}$ is upper-bounded as in Eq. (II.60) for all $d' \leq d$. Similarly, Eq. (II.33) implies the existence of a $\tilde{\phi}_j^{(d-1)}$ with

$$\min_l \min_n |\tilde{\phi}_j^{(d-1)} - (\tilde{\theta}_l^{(d)} + 2\pi n)/k_d|_T > \frac{2\epsilon(1+\kappa_d)}{k_d}. \tag{II.63}$$

which can not happen due to the success of Alg. II.4 which implies the bound in Eq. (II.49). Consider lastly step 4e which exits if the current dth order estimates do not lie in the region for which Eq. (II.7) holds with given k_d. We have argued in Lemma II.5 that, assuming success of the subroutines implementing Alg. II.4, that Eq. (II.7) holds for the phase estimates at all orders. □

Now let us consider failures of the QEEP subroutine, Alg. II.4, which do not lead to exiting. Let's imagine that the first failure occurs at some order d_0. Now we want to make sure that continuing with higher orders after such failure still leads to an error of order $\sim \epsilon/k_{d_0-1}$, even though the failure (or any subsequent failure) is not detected.

To show this, we check that if Alg. II.2 exits at some later round, namely during $d = d_f + 1$ and outputs estimates $\tilde{\phi}_j^{(d_f)}$ that these will be sufficiently close to the estimates right before failure, that is, the set of phases $\tilde{\phi}_j^{(d_0-1)}$.

Then, by Lem. II.5, these estimates will also be sufficiently close to the true phases ϕ_j.

Lemma II.6
Let Alg. II.2 exit at order $d = d_f + 1$ and let the QEEP subroutine, Alg. II.4, of step 4b first fail at $d = d_0 \leq d_f + 1$. For each ϕ_j, there will be an estimate $\tilde{\phi}_l^{(d_f)}$, produced at step 4d in Alg. II.2 which satisfies

$$\left|\phi_j - \tilde{\phi}_l^{(d_f)}\right|_T \leq \frac{14\epsilon}{k_{d_0-1}}. \tag{II.64}$$

Vice-versa, for each estimate $\tilde{\phi}_l^{(d_f)}$ there exists a phase ϕ_j such that

$$\left|\phi_j - \tilde{\phi}_l^{(d_f)}\right|_T \leq \frac{14\epsilon}{k_{d_0-1}}. \tag{II.65}$$

Proof. Since each QEEP subroutine, Alg. II.4, in Alg. II.2 succeeds up to order $d_0 - 1$, Lemma II.5 guarantees that

- (Property 1a) For every phase ϕ_j there exists an estimate $\tilde{\phi}_l^{(d_0-1)}$ such that
$$|\phi_j - \tilde{\phi}_l^{(d_0-1)}|_T \leq \frac{2\epsilon}{k_{d_0-1}}.$$
- (Property 1b) For every estimate $\tilde{\phi}_l^{(d_0-1)}$ there exists a phase ϕ_j such that
$$|\phi_j - \tilde{\phi}_l^{(d_0-1)}|_T \leq \frac{2\epsilon}{k_{d_0-1}}.$$

Then, since the algorithm does not exit at step 4c through Eqs. (II.33) or step 4e for any order $d = d_0, \dots d_f$, for each estimate $\tilde{\phi}_l^{(d-1)}$ we can associate some $\tilde{\theta}_{m_l}^{(d)}$ that satisfies

$$\begin{aligned}\frac{1}{k_d}\left|k_d\tilde{\phi}_l^{(d-1)} - \tilde{\theta}_{m_l}^{(d)}\right|_T = & \\ \min_{n\in\{0,\dots,\lfloor k_{d+1}\rfloor-1\}} & \left|\tilde{\phi}_l^{(d-1)} - (\tilde{\theta}_{m_l}^{(d)} + 2\pi n)/k_d\right|_T \\ & \leq \frac{2\epsilon(1+\kappa_d)}{k_d},\end{aligned} \tag{II.66}$$

where the second equality follows from being allowed to apply Eq. (II.8) (which is validated by passing the test at step 4e). This implies that in step 4d of Alg. II.2 at round d, for a given $\tilde{\theta}_{m_l}^{(d)}$, the optimization of ξ_{n,m_l} over n will pick the integer n_{l,m_l}, i.e. the integer associated with matching $\tilde{\theta}_{m_l}^{(d)}$ with $\tilde{\phi}_l^{(d-1)}$. Next, similar as in the proof of Lemma II.5, we can consider the possibility of matching to other estimates $\tilde{\phi}_k^{(d-1)} \neq \tilde{\phi}_l^{(d-1)}$. Since κ_d is chosen in step 4f of Alg. II.2, we claim that either Eq. (II.36) holds, in which case

$$\begin{aligned}\min_{n\in\{0,\dots,\lfloor k_d\rfloor-1\}} & \left|\tilde{\phi}_k^{(d-1)} - \frac{2\pi n + \tilde{\theta}_{m_l}^{(d)}}{k_d}\right|_T \\ & > \frac{2\epsilon}{k_d}(1+\kappa_d),\end{aligned} \tag{II.67}$$

hence this ξ_{k,m_l} is not optimal, or that Eq. (II.37) holds, in which case

$$n_{l,m_l} = \underset{n\in\{0,\dots,\lfloor k_d\rfloor-1\}}{\arg\min} \left|\tilde{\phi}_k^{(d-1)} - \frac{2\pi n + \tilde{\theta}_{m_l}^{(d)}}{k_d}\right|_T = n_{k,m_l}. \quad \text{(II.68)}$$

The proofs of these claims are exactly the same as in the proof of Lemma II.5, i.e. using Eqs. (II.54), (II.56),(II.57).

Now let us prove Eq. (II.64). Given a phase ϕ_j, we can use Property (1a) to find an associated estimate $\tilde{\phi}_{\to j}^{(d_0-1)}$ within $2\epsilon/k_{d_0-1}$. Then for this estimate let $\theta_{m_{\to j}}^{(d_0)}$ be the matched estimate in the next round for which Eq. (II.66) holds, so that the round produces a new estimate $\tilde{\phi}_{\to j}^{(d_0)} = \frac{\tilde{\theta}_{m_{\to j}}^{(d_0)} + 2\pi n_{\to j,m_{\to j}}}{k_{d_0}}$ (which we label with $\to j$ again) for which

$$|\tilde{\phi}_{\to j}^{(d_0-1)} - \tilde{\phi}_{\to j}^{(d_0)}|_T \leq \frac{2\epsilon}{k_{d_0}}(1+\kappa_{d_0}). \quad \text{(II.69)}$$

Then again for $\tilde{\phi}_{\to j}^{(d_0)}$ there exists some matching $\theta_{m_{\to j}}^{(d_0+1)}$ etc. and this generates a series of estimates $\tilde{\phi}_{\to j}^{(d)}$ up to order d_f. For a given ϕ_j we can then bound, using this series of estimates and $\kappa_d \geq 2$ for all d,

$$\begin{aligned}
\left|\tilde{\phi}_j^{(d_f)} - \phi_j\right|_T &\leq \left|\tilde{\phi}_{\to j}^{(d_0-1)} - \phi_j\right|_T \\
&\quad + \sum_{d=d_0}^{d_f} \left|\tilde{\phi}_{\to j}^{(d-1)} - \tilde{\phi}_{\to j}^{(d)}\right|_T \\
&\leq \frac{2\epsilon}{k_{d_0-1}} + \sum_{d=d_0}^{d_f} \frac{2\epsilon(1+\kappa_d)}{k_d} \\
&= \frac{2\epsilon}{k_{d_0-1}} \left(1 + \sum_{d=d_0}^{d_f} \frac{1+\kappa_d}{\kappa_{d_0}\kappa_{d_0+1}\dots\kappa_d}\right) \\
&\leq \frac{2\epsilon}{k_{d_0-1}} \left(1 + \sum_{n=0}^{\infty} \frac{3}{2^n}\right) \\
&= \frac{14\epsilon}{k_{d_0-1}}. \quad \text{(II.70)}
\end{aligned}$$

Now let's prove Eq. (II.65) and start with an estimate $\tilde{\phi}_l^{(d_f)}$ which was obtained from some $\tilde{\theta}_l^{(d_f)}$ matched with a previous estimate $\tilde{\phi}_l^{(d_f-1)}$ (just for convenience we again use the same label) such that $|\tilde{\phi}_l^{(d_f)} - \tilde{\phi}_l^{(d_f-1)}|_T \leq \frac{2\epsilon}{k_{d_f}}(1+\kappa_{d_f})$, using that we do not exit through Eq. (II.34). Then again for this previous estimate $\tilde{\phi}_l^{(d_f-1)}$ we can repeat the argument and create a sequence of estimates up to $\tilde{\phi}_l^{(d_0-1)}$. For the last estimate, we invoke Property (1b), namely that there is a nearby ϕ_j. Then we can upperbound for this ϕ_j: $|\tilde{\phi}_l^{(d_f)} - \phi_j|_T \leq |\tilde{\phi}_l^{(d_0-1)} - \phi_j|_T + \sum_{d=d_0}^{d_f} |\tilde{\phi}_l^{(d)} - \tilde{\phi}_l^{(d-1)}|_T$ etc., exactly as in Eq. (II.70), leading to Eq. (II.65). □

II.4.3. Algorithm II.2 achieves the Heisenberg limit

We have seen that the success of the QEEP subroutines in Alg. II.2 leads to an error scaling as $\epsilon/k_{d_f} \sim \delta_c$. Now we must choose the success probability p_d of these subroutines in Eq. (II.28), depending on α, γ so that the total mean-squared-error is bounded by some $\delta^2 = O(\delta_c^2)$ while the quantum cost $T = O(\delta^{-1})$. We note that the next theorem contains no logarithmic factors in δ^{-1}, as in [57], but achieves pure Heisenberg scaling.

Theorem II.2
Algorithm II.2 solves the multiple eigenvalue estimation problem in Def. II.4 with accuracy error δ and total quantum cost $T = O(\delta^{-1})$, given A, n_ϕ and a fixed ϵ_0 and ϵ obeying Eqs. (II.27) *and* (II.26)*, and some choice for the constants $\alpha > 0$ and $\gamma > 2$.*

Remarks: Note that the dependence on the number of phases n_ϕ is not made explicit in the statement of this Theorem, but this dependence will be polynomial in n_ϕ, not necessarily a very low-order polynomial. This dependence comes through the choice for ϵ_0 (and ϵ) via Eq. (II.26) (resp. Eq. (II.27)) which sets the error and thus the running time of the QEEP Algorithm II.3.

Proof. Our proof is motivated by the analysis in [74] for a single phase ϕ leading to Theorem II.1. The idea is to bound the mean-squared-error in the final estimation of ϕ by summing over error contributions at each order d at which the phase extraction subroutine may fail (with probability $1 - p_d$).

In our case the multipliers κ_d (and k_d) at each order are not fixed (as in Theorem II.1) but depend on phase estimates at previous orders and thus measurement data at previous orders. Our confidence parameter p_d in Eq. (II.28), which determines the number of repeats of experiments, and hence the cost, in Alg. II.4, depends on k_d and is thus a random variable depending

on previous measurement data. All measurement data are denoted by $\mathbf{x}$ and thus we have random variables $k_d(\{\kappa_{d'}(\mathbf{x})\}_{d'=1}^{d})$ and $p_d(\{\kappa_{d'}(\mathbf{x})\}_{d'=1}^{d})$.

Consider the mean-squared-error δ_j^2 for the jth phase ϕ_j in Eq. (II.25) in Definition II.4. We have three error contributions to consider given a choice for the random variable k_d.

1. With probability $1-p_0$ the subroutine Alg. II.4 in Alg. II.2 fails at step 1 ($d=0$). In this case, as we always return *some* estimate, δ_j is bounded for all j by π.

2. With probability at most $(1-p_{d_0})\prod_{d=0}^{d_0-1} p_d \leq 1-p_{d_0} = e^{-\alpha}\left(\frac{k_{d_0}\delta_c}{\pi}\right)^{\gamma}$, the subroutine Alg. II.4 in Alg. II.2 fails for the first time at some order $1 \leq d_0 \leq d_f$, and the algorithm proceeds in any way afterwards (by possibly exiting or not). In this case, Lemma II.6 bounds δ_j for all j by $\frac{14\epsilon}{k_{d_0-1}}$ or Lemma II.5 bounds δ_j for all j by $\frac{2\epsilon}{k_{d_0-1}} \leq \frac{14\epsilon}{k_{d_0-1}}$.

3. With probability less than $\prod_{d=0}^{d_f} p_d < 1$ the subroutine Alg. II.4 in Alg. II.2 succeeds up to the final round d_f, and Lemma II.5 implies that $\delta_j \leq 2\epsilon/k_{d_f} \leq 2\delta_c$ for all j as $k_{d_f} \geq \epsilon/\delta_c$.

We can now bound the mean-squared-error as a sum over the above three contributions weighted by their relevant overlaps:

$$\begin{aligned}
\delta_j^2 &\leq (1-p_0)\pi^2 + 4\delta_c^2 + \\
&\quad + \sum_{\mathbf{x}} \mathbb{P}(\mathbf{x}) \left[\sum_{d_0=1}^{d_f-1} \mathbb{P}(\kappa_1,\ldots,\kappa_{d_0}|\mathbf{x})(1-p_{d_0}(\{\kappa'_d\}_{d'=1}^{d_0}))\left[\frac{14\epsilon}{k_{d_0-1}(\{\kappa'_d\}_{d'=1}^{d_0-1})}\right]^2\right] \\
&= \pi^2 e^{-\alpha}\left(\frac{\delta_c}{\pi}\right)^{\gamma} + \sum_{\mathbf{x}} \mathbb{P}(\mathbf{x}) \left[\sum_{d_0=1}^{d_f-1} \mathbb{P}(\kappa_1,\ldots,\kappa_{d_0}|\mathbf{x}) e^{-\alpha}\left(\frac{k_{d_0}\delta_c}{\pi}\right)^{\gamma} \frac{196\epsilon^2}{k_{d_0-1}^2}\right] + 4\delta_c^2 \\
&\leq \pi e^{-\alpha}\left(\frac{\delta_c}{\pi}\right)^{\gamma} + 0.15 \times \delta_c^2 \sum_{\mathbf{x}} \mathbb{P}(\mathbf{x})\mathbb{P}(\kappa_1,\ldots,\kappa_{d_0}|\mathbf{x}) \sum_{d_0=1}^{d_f-1}\left(\frac{k_{d_0}\delta_c}{\pi}\right)^{\gamma-2} + 4\delta_c^2.
\end{aligned} \tag{II.71}$$

Here we have removed the dependency of k_{d_0} and k_{d_0-1} on the previous multipliers for notational simplicity. For $d=1$ we have $e^{-\alpha}\kappa_{d_0}^2 196(\epsilon/\pi)^2 = e^{-\alpha}k_1^2 196(\epsilon_0/\pi)^2 \leq \frac{196\times 16\times 4}{(300)^2} \leq 0.15$ due to Eq. (II.26). For $d>1$, $e^{-\alpha}\kappa_{d_0}^2 196(\epsilon/\pi)^2 \leq \frac{9\times 196\times 4}{(300)^2} = 0.08$, due to Eq. (II.27).

To evaluate the middle term, we write $k_{d_0} = k_{d_f}\frac{k_{d_0}}{k_{d_f}}$, and note that as $k_d = \prod_{d'=1}^{d}\kappa_{d'}$, we have $\frac{k_{d_0}}{k_{d_f}} \leq 2^{d_0-d_f}$ as the multiplier $\kappa_d \geq 2$. As $k_{d_f} \leq \frac{2\epsilon}{\delta_c} < \frac{\pi}{2\delta_c}$,

we have

$$\sum_{d_0=1}^{d_f-1}\left(\frac{k_{d_0}\delta_c}{\pi}\right)^{\gamma-2} \leq \frac{1}{2^{\gamma-2}}\sum_{d_0=1}^{d_f-1}(2^{\gamma-2})^{(d_0-d_f)} \leq \frac{2^{4-\gamma}}{2^\gamma-4}, \tag{II.72}$$

where the last inequality holds since $\gamma > 2$. By letting the upper bound be independent of the κ_ds, we can remove the dependence on $\mathbf{x}$ in Eq. (II.71), using that $\sum_{\mathbf{x}} \mathbb{P}(\mathbf{x})\mathbb{P}(\kappa_1,\ldots,\kappa_{d_0}|\mathbf{x}) = 1$. This yields a final bound on δ_j of

$$\delta_j^2 \leq \delta_c^2\left[\pi^{1-\gamma}e^{-\alpha}\delta_c^{\gamma-2} + \frac{0.15\times 2^{4-\gamma}}{2^\gamma-4} + 4]\right]. \tag{II.73}$$

As $\gamma > 2$, this scales as δ_c^2 as $\delta_c \to 0$.

Let us now calculate the cost of executing Alg. II.2 in terms of the number of unitary applications. Again this depends on the choice of multiplier κ_d at each step. Let us fix a sequences of k_ds, and let d_f be the final round of estimation in this algorithm, i.e. the final round for which we invoked the quantum subroutine in Alg. II.4. At each order d we use $V^{\mathsf{k}} = U^{k_d\mathsf{k}}$, where $\mathsf{k} = 0, 1, \ldots, K$, with $2M_d$ samples where K is a function of ϵ as in Theorem II.3. The cost of each experiment is $k_d\mathsf{k}$.

We can calculate

$$\begin{aligned} T &= \sum_{d=0}^{d_f}\sum_{\mathsf{k}=1}^{K}(2M_d k_d \mathsf{k}) \\ &= \sum_{d=0}^{d_f} M_d k_d K(K+1). \end{aligned} \tag{II.74}$$

The QEEP algorithm in Theorem II.3 requires $M_d = \tilde{O}(|\ln(1-p_d)|\epsilon^{-4})$ with p_d in Eq. (II.28) and $K = \tilde{O}(\epsilon^{-1})$.

We may bound

$$\begin{aligned} T &\leq \tilde{O}(\epsilon^{-6})\sum_{d=0}^{d_f} k_d\left|-\alpha+\gamma\ln\left(\frac{k_d\delta_c}{\pi}\right)\right| \\ &= \tilde{O}(\epsilon^{-6})\sum_{d=0}^{d_f} k_d\left[\alpha-\gamma\ln\left(\frac{k_d\delta_c}{\pi}\right)\right]. \end{aligned} \tag{II.75}$$

We again bound $k_d = \frac{k_d}{k_{d_f}} k_{d_f} \leq 2^{d-d_f} \frac{\pi}{2\delta_c}$, which yields

$$\begin{aligned} T &\leq \delta_c^{-1} \tilde{O}(\epsilon^{-6}) \Bigg[\frac{\pi}{2} \\ &\quad \times \sum_{d=0}^{d_f} 2^{d-d_f} \left[\alpha - \gamma (d - d_f - 1) \ln(2) \right] \Bigg] \\ &\leq \delta_c^{-1} \tilde{O}(\epsilon^{-6}) (\alpha + 2\gamma \ln(2)). \end{aligned} \tag{II.76}$$

Combining our bounds then yields

$$\delta \leq \delta_c \left[e^{-\alpha} \pi^{1-\gamma} + 4 + \frac{0.15 \times 2^{4-\gamma}}{2^\gamma - 4} \right]^{\frac{1}{2}} \tag{II.77}$$

$$\begin{aligned} &\leq T^{-1} \left[e^{-\alpha} \pi^{1-\gamma} + 4 + \frac{0.15 \times 2^{4-\gamma}}{2^\gamma - 4} \right]^{\frac{1}{2}} \\ &\qquad \times \tilde{O}(\epsilon^{-6}) [\alpha + 2\gamma \ln(2)] \\ &= O(T^{-1}), \end{aligned} \tag{II.78}$$

which is the Heisenberg limit. Note that a dependence on the number of phase n_ϕ enters the scaling of δ via ϵ which needs to be bounded as in Eqs. (II.26) and (II.27). □

II.5. Numerical implementation

Thm. II.2 requires using the QEEP algorithm (Theorem II.3 and Algorithm II.4) in order to obtain provable bounds. Instead of analytic bounds, we now turn to a numerical demonstration, giving the opportunity to implement and test Algorithm II.2 with a few modifications. We test the algorithm using two different sub-routines, one based on the matrix pencil method [164], and one based on the QEEP time-series analysis of Theorem II.3, as described in Algorithm II.4. Code to implement all simulations can be found at `https://github.com/alicjadut/qpe`.

To improve the practical performance of Alg. II.2, we make the following two small changes. Firstly, instead of choosing κ_{d+1} in step 4f in the ranges declared in Lem. II.4, we choose the largest κ_{d+1} consistent with Eq. (II.36) and Eq. (II.37) for all $\tilde{\phi}_j^{(d)}, \tilde{\phi}_l^{(d)}$. We note that the maximum such κ_{d+1} is bounded above by $\frac{\pi}{2\epsilon} - 1$, as the left-hand side of Eq. (II.36) is bounded above by 2π and the left-hand side of Eq. (II.37) is bounded below by 0. (In practice,

tighter bounds can be found by checking the boundaries of the regions $R_{jl}^{(n)}$ defined in Eq. (II.92), and we find the largest possible κ_{d+1} by iterating backwards through these boundaries till a gap is found.) Secondly, as the bounds for ϵ and ϵ_0 in Lem. II.4 are rather loose, and our performance scales rather badly in both, we choose the largest $\epsilon = \epsilon_0$ that allows all simulations to find a value of $\kappa_{d+1} > 2$ at each order.

When using the matrix pencil processing subroutine, we follow the implementation described in Ref. [136]:

Algorithm II.3
The matrix pencil method takes as input estimates of the phase function $g(\mathsf{k}) = \sum_j A_j e^{i\mathsf{k}\theta_j}$ *for a unitary* V *at points* $\mathsf{k} = 0, 1, \dots K$ *and an overlap bound* A*, and proceeds as follows:*

1. *Construct the* $L_K \times (2K - L_K + 1)$ *Hankel matrices* $G^{(0)}$, $G^{(1)}$, *where* $G^{(a)}_{i,j} = g(i + j + a - K)$ *for* $i \in \{0, 1, ...L_K - 1\}, j \in \{0, 1, ..., 2K - L_K\}$, $a = 0, 1$, *with* $L_K = \lfloor (K + 1)/2 \rfloor$, *and using* $g(-\mathsf{k}) = g^*(\mathsf{k})$.

2. *Construct the* $L_K \times L_K$ *shift matrix* T *by least-squares minimization of the matrix* 2*-norm* $\|TG^{(0)} - G^{(1)}\|$.

3. *Calculate the eigenvalues of* T, $\lambda_j = |\lambda_j| e^{i\tilde{\theta}_j}$ *and from there the phase estimates* $\tilde{\theta}_j$.

4. *Calculate the overlap estimates* $\tilde{A}_j$ *by least-squares minimization of the vector* 2*-norm* $\|BA - g\|$, *where* B *is the* $(K + 1) \times L_K$ *matrix*

$$B_{k,j} = \lambda_j^{\mathsf{k}}, \tag{II.79}$$

   and $g = [g(0), ...g(K)]^T$.

5. *Return the phase estimates* $\tilde{\theta}_j$ *for which the corresponding overlap estimate* $\tilde{A}_j \geq A$.

To use this algorithm as a subroutine in Alg. II.2 (in place of Alg. II.4), we implement it on the matrix $V = U^{k_d}$, which requires implementing $V^{\mathsf{k}} = U^{\mathsf{k}k_d}$ for a range of integer k on a quantum device.

To isolate the performance of the estimation routine from the generation of the signal itself, we do not test our protocols on data generated from simulating or approximating a particular unitary. Instead, we test the ability of the protocols to estimate $n_\phi = 2$ and $n_\phi = 4$ randomly-chosen phases $\phi_j \in [0, 2\pi]$ when sampling from the true phase function $g(k)$. We take all phases with equal weight — $A_j = 1/n_\phi$. We simulate the sampling from $g(k)$ in Algorithm II.3 or Algorithm II.3 for some $V = U^{k_d}$ by simulating the

readout of a control qubit with the reduced density matrix of Eq. (II.4). (In practice this would be generated by the quantum circuit in Fig. II.1.) We first draw M_d i.i.d. samples from the two Bernoulli distributions

$$\mathbb{P}_k^r(+1) = \frac{1}{2}\sum_{j=1}^{n_\phi} A_j(1+\cos(\theta_j k)), \tag{II.80}$$

$$\mathbb{P}_k^i(+1) = \frac{1}{2}\sum_{j=1}^{n_\phi} A_j(1-\sin(\theta_j k)), \tag{II.81}$$

where $\theta_j = k_d\phi_j \mod 2\pi$ are the eigenvalues of V. Then, we return the fraction of $+1$s drawn as estimates for the real and imaginary parts of $g(k)$ respectively. This is then repeated at all points $k = k_d\mathsf{k}$ for $\mathsf{k} = 0, 1, \ldots, K$. Following the discussion in Sec. II.2.2 and using the notation from Eq. (II.74), we sum the total quantum cost for the algorithm over all requested $g(k)$ queries; $T = 2\sum_d \sum_{\mathsf{k}=1}^{K} \mathsf{k} k_d M_d$. (We ignore the sub-leading correction from the final term in Eq. (II.74) as this will not affect the scaling of our result.) For the signal length K and number of points M_d to sample each $g(k)$ at, we follow the bounds given in Ref. [138] (both when using the QEEP and matrix pencil subroutines):

- signal length: $K = \lceil 0.1L\ln^2 L\rceil$, with $L = \lceil \frac{2\pi}{\epsilon}\rceil$ the number of bins used in the QEEP subroutine (Def. II.5).
- number of measurements of each circuit: $M_d = \left\lceil |\ln(1-p_d)|\,\epsilon^{-4}\right\rceil$.

Here, p_d is given in Eq. (II.28) for a given k_d. This equation requires fixing a choice of α and γ — across all experiments we take $\alpha = 2$ and $\gamma = 2.1$.

To demonstrate that our methods achieve the Heisenberg limit, in Fig. II.3 we plot the error as a function of the total quantum cost for a set of simulations using the methods described above. Each simulation draws a different set of n_ϕ random phases, and a final error $\delta_c \in [10^{-5}, 10^{-2}]$ (for each choice of δ_c we use the same 50 sets of phases). We plot the error for each phase estimate separately in Fig. II.3 (i.e. each simulation corresponds to n_ϕ points in the plot). As both the total quantum cost and the error is different between simulations, we bin all experiments within a range of T values, and calculate the root mean square error and root mean square total quantum cost. This gives a good approximation to the accuracy error defined in Def. II.4 for the restricted data set used.

For the QEEP subroutine, we observe a clear fit of the data (blue points) to a $\delta \sim T^{-1}$ trend, as expected from Thm. II.2, but with a rather large constant factor; we find $T \sim 10^{10}\delta^{-1}$ for estimating 2 phases and $T \sim 10^{15}\delta^{-1}$ for estimating 4. Further optimization of the QEEP algorithm for these

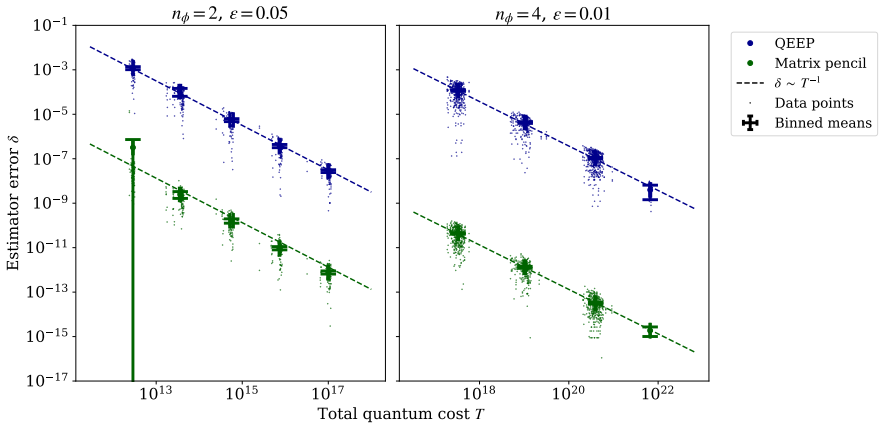


Figure II.3.: Convergence of the algorithm with total quantum cost T. Phase estimates were obtained with either QEEP (blue) or matrix pencil (green) subroutines with parameters described in the text. Individual points show the error $\delta = |\tilde{\phi}_j - \phi_j|_T$ on individual phases in each simulation. This data is binned in the x-axis, and for each bin a root-mean-square error and standard deviation (error bars) are plotted in the x- and y-direction. Dotted lines show a fit of these means to $\delta \sim T^{-1}$.

purposes may yet improve this constant factor. However, as the methods of Ref. [138] were not designed for estimating individual phases, it may be expected that this method performs somewhat badly for this purpose, so we have not pursued this further.

Simulations using the matrix pencil subroutine outperform simulations using the QEEP subroutine by a factor of $10^4 - 10^6$, and clearly demonstrate Heisenberg-limited scaling $\delta \sim T^{-1}$ as well. We take this result instead of an analytic proof as strong numerical evidence for Heisenberg-limited scaling when a version of Alg. II.2 is constructed using the matrix pencil method as a subroutine. We notice that the error in two phase estimates in the first bin of the matrix pencil method for $n_\phi = 2$ is significantly above the remainder of the population (by about a factor 100×), which blows up our error bars for this bin. Further investigation shows that the two phases in question are from the same simulation, and separated by only 1.5×10^{-4}. By contrast, for the simulation in question (at $d = 1$) our algorithm sets $k_1 K \sim 3 \times 10^3 < (1.5 \times 10^{-4})^{-1}$ (where $k_1 K$ is the largest value of the phase function $g(k)$ sampled during this simulation). This implies that our signal lies within the region where improving our estimation accuracy by increasing

the number of shots M_d is exponentially hard [165]. In latter simulations with these two phases we see that our estimation error regresses to similar results as all other estimates.

II.6. Conclusion

In this work we studied Heisenberg-limited quantum phase estimation using a single control qubit. In this form of phase estimation, we rely on classical signal processing to extract eigenvalue data from the phase function $g(k)$ in Eq. (II.2).

It has been an open question whether these methods can achieve the Heisenberg limit in the case of multiple phases: Ref. [57] answered this question up to log factors with a Heisenberg-limited Monte Carlo algorithm, providing a sampling of the spectral function $A(\phi)$ in Eq. (II.3) from which to estimate the phases. In this work we also answered this question in the affirmative exactly with a new adaptive multi-order phase estimation algorithm, for which we prove Heisenberg scaling if the algorithm uses a QEEP phase extraction subroutine. We numerically show the performance of this algorithm, also when instead of using a QEEP subroutine, one uses the matrix pencil method to extract phase estimates from the phase function $g(k)$. This result complements the previous work discussed in the introduction by closing the question of whether there exists a gap between classical post-processing of phase function data and fully quantum phase estimation.

In obtaining our results we encountered at least two details of quantum phase estimation that we have not seen discussed in the literature. The first is the dense signal limit, Eq. (II.12) in Thm. II.1: sampling $g(k)$ at *all* integer point $k = 1, \ldots, K$ is sub-optimal regardless of what method is used to process the data. However, we also briefly argued that by picking points among $k = 1, \ldots, K$ at random one may go beyond this, and one could interpret this as allowing the results obtained in [57] in which such randomized choices for k are taken.

The second is the need for adaptive choices of k_d to solve the phase matching problem. It is unclear to us how far this problem extends; although Lemma II.4 provides a practical solution, others may still exist. Another open question with respect to Algorithm II.2 is whether one can remove the need to choose real-valued multipliers κ_d and restrict to integer choices. Restricting $\kappa_d \in \mathbb{N}$ would significantly simplify some technical issues, i.e. the applicability of Lemma II.1 and the need for shifting phases in step 2 of Alg. II.2, but we don't know how to prove a version of Lemma II.4 for $\kappa_d \in \mathbb{N}$. In fact, we don't know whether there is a fundamental difference in performance between only using data obtained with integer k in $g(k)$ versus data obtain with real-valued,

–in practice rational–, k in $g(k)$.

We have assumed in our problem description, Def. II.4, that the spectrum in the input state is discrete consisting of n_ϕ phases with amplitudes above some cut-off. In practice this condition may not be fulfilled and thus studying the performance of the algorithms on more typical spectra induced by many-body Hamiltonians and easy-to-prepare input states will be of interest. The scaling of our (provable) Heisenberg-limited algorithm in n_ϕ is also rather poor [$O(n_\phi^{24})$ as described], as we have not attempted to optimize this aspect. This should in principle be immediately reducible to $O(n_\phi^3\delta^{-1} + n_\phi^6)$ under the assumption that the matrix pencil method continues to achieves the dense sampling limit when estimating multiple eigenvalue, however we do not know a proof of this. In principle linear scaling with n_ϕ should be achievable (or even sub-linear if the methods of [112] could be applied in this setting); optimizing this is a clear target for future work.

A direction for future research is to make this algorithm efficient in practice (i.e. improve the parameter dependence and the practical run time) or devise yet-different Heisenberg-scaling algorithms and examine their performance in the presence of experimentally-noisy signals $g(k)$.

II.A. Phase extraction subroutine

For our Heisenberg-limited algorithm, at each order d we need to extract eigenphases of $V = U^{k_d}$ from the signal generated by PFE. We require that the phase estimates satisfy the promises of Lem. II.3. This is achieved by Alg. II.4, which is based on the solution to QEEP of Ref. [138].

In this appendix, we first summarize the results of Ref. [138] by giving a precise definition of the QEEP (Def. II.5) and performance guarantees of the time-series analysis (Def. II.3). Then we describe the Conservative QEEP Eigenvalue Extraction algorithm (Alg. II.4) and give the proof of Lem. II.3.

II.A.1. The Quantum Eigenvalue Estimation Problem

Definition II.5 (Quantum Eigenvalue Estimation Problem, QEEP)
Let $A(\phi)$ be the spectral function defined in Eq. (II.3) *for a unitary U and $|\Psi\rangle$. Given is an error parameter $\epsilon > 0$, a confidence parameter $1 \geq p > 0$, and a set of non-negative (approximate indicator) functions $f^l(\phi)$ for $\phi \in [0, 2\pi)$ for $l = 0, \ldots, L-1$, $L = \lceil \frac{2\pi}{\epsilon} \rceil$, where $f^l(.)$ has support on only the interval bin*

$$\mathcal{V}_l = [(l-1)\epsilon, (l+1)\epsilon]_T, \tag{II.82}$$

and $f^l(\phi) + f^{(l-1)}(\phi) = 1$ *for all* $\phi \in \mathcal{V}_l \cap \mathcal{V}_{l-1}$. *Assuming access to the* PFE *subroutine, Def. II.2, the goal is to output an approximation* $\tilde{b}_l$ *for* $l = 0, \ldots, L-1$ *to the integral*

$$b_l = \int_0^{2\pi} d\phi A(\phi) f^l(\phi), \tag{II.83}$$

which satisfies

$$\sum_{l=0}^{L-1} |\tilde{b}_l - b_l| \leq \epsilon, \tag{II.84}$$

with probability at least p.

Note that the bins $\mathcal{V}_l$ have width 2ϵ and overlap on a region of width ϵ and $\sum_{l=0}^{L-1} b_l = 1$.

Theorem II.3 (QEEP Algorithm [138])
One can solve the QEEP problem in Definition II.5 with $\{f^l(.)\}$ *a set of 'bump' functions*

$$\begin{aligned} f^l(\phi) =& \frac{2a}{\epsilon} \int_{l\epsilon - \frac{\epsilon}{2}}^{l\epsilon + \frac{\epsilon}{2}} d\phi' \\ & \times \exp\left\{ - \left[1 - \frac{4}{\epsilon^2}(\phi - \phi')^2 \right]^{-1} \right\}, \end{aligned} \tag{II.85}$$

with normalization constant $a \approx 2.252$, *using* PFE *in Def. II.2 with* $k = 0, 1, \ldots, K$ *with* $K = O\left(\epsilon^{-1} \ln^2(\epsilon^{-1})\right)$ *and* $M = \tilde{O}(|\ln(1-p)|\epsilon^{-4})$ *for each* $k = 1, 2 \ldots, K$. *The total quantum cost for* U *is then bounded as* $T = O(MK^2) = \tilde{O}(|\ln(1-p)|\epsilon^{-6})$.

We note that the approximate indicator functions $f^l(.)$ in Eq. (II.85) are designed to have a quickly decaying Fourier series, which is required to achieve polynomial-time scaling. We also refer the reader to Ref. [158], which has extended the result by relaxing the requirement that $f^l(\phi) + f^{(l-1)}(\phi) = 1$ on the interval $\mathcal{V}_l \cap \mathcal{V}_{l-1}$. The idea of this QEEP algorithm is as follows. Since $b_l = \sum_j A_j f^l(\phi_j)$, using Eq. (II.3), periodically extending $f^l(\phi)$ beyond $[0, 2\pi)$ and Fourier decomposing gives $b_l = \sum_j A_j \sum_{k\in\mathbb{Z}} e^{ik\phi_j} \tilde{f}^l(k) = \sum_{k\in\mathbb{Z}} g(k)\tilde{f}^l(k)$. At the same time, the fact that $f^l(.)$ is an indicator function ensures that $b_l \approx \sum_{\phi_j \in \mathcal{V}_l} A_l$. Thus knowledge of $g(k)$ and the Fourier coefficients $\tilde{f}^l(k)$ for a range of k allows one to estimate the weights b_l. The requirement to estimate the spectral function to within a 1-norm ϵ, Eq. (II.84), is very stringent, hence the scaling of T with error ϵ is quite costly, $T = O(\epsilon^{-6})$. It is

possible that one can improve the scaling by re-examining the analysis in [138].

II.A.2. Conservative QEEP Eigenvalue Extraction

Algorithm II.4 (Conservative QEEP Eigenvalue Extraction)
Fix an overlap bound A, an error bound $0 < \epsilon < \frac{A}{3}$, and a confidence bound $0 < p < 1$. Assume access to a QEEP Algorithm II.3 for a unitary V. The algorithm proceeds as follows:

1. *Use the QEEP subroutine with error ϵ, Alg. II.3, and confidence p to generate an estimate $\tilde{b}_l$ for b_l as defined in Eq.* (II.83).
2. *Construct the set*
$$S = \{l \in \{0, \ldots, L-1\} | \tilde{b}_l \geq A/3\}. \tag{II.86}$$
3. *Find the smallest $l \in \{0, \ldots, L\}$ with $l \notin S$ and call it $l_{\min}$.*
4. *For $l' = l_{min}, \ldots, (l_{min} + L - 1) \bmod L$:*
   a) if $l' \in S$ and $(l' - 1) \bmod L \in S$, remove l' from S.
5. *Return the set $\{\tilde{\theta}_l = l\epsilon\}_{l \in S}$ as a set of estimates of eigenphases of V.*

Let us now motivate Algorithm II.4. One may identify phases with sufficient probability in the output of the QEEP algorithm as the bins l with $b_l > b_{\text{cutoff}}$ with b_l in Eq. (II.83). Then to convert this into an estimate of a phase θ_j, for each such bin above cut-off we could output the estimate ϵl, i.e. in the middle of the corresponding bin $\mathcal{V}_l$.

We calculate appropriate values of b_{cutoff} and the QEEP error ϵ to guarantee that we output an estimate for each θ_j with $A_j > A$ with a provable confidence, and to guarantee no estimate in the absence of any θ_j. Def. II.5 states that when there exists such a $\theta_j \in \mathcal{V}_l \cap \mathcal{V}_{l-1}$, we are guaranteed with confidence p that $\tilde{b}_l + \tilde{b}_{l-1} + \epsilon > A$ (as $b_l + b_{l-1} > A$). To guarantee that at least one of $\tilde{b}_l$ or $\tilde{b}_{l-1}$ is larger than b_{cutoff} with the same confidence, we thus require $b_{\text{cutoff}} \leq (A - \epsilon)/2$.

Similarly, Def. II.5 states that when there exists no $\theta_j \in \mathcal{V}_l$ with $A_j > 0$, we are guaranteed with confidence p that $b_l < \epsilon$. To prevent a spurious estimate in this case, we require $b_{\text{cutoff}} \geq \epsilon$. Solving this to maximise ϵ (which minimizes the cost of the QEEP routine) yields $b_{\text{cutoff}} = \epsilon = A/3$ which is what we use in Alg. II.4.

A further small complication exists in solving the problem posed in Def. II.4: we require that one outputs at most n_ϕ phases. This will be satisfied if we

can guarantee at most one phase estimate per θ_j with $A_j > A$, as we know there are $n_\theta = n_\phi$ such estimates. As the bins $\mathcal{V}_l$ (Def. (II.82)) overlap, a phase θ_j may participate in up to two neighbouring bins — corresponding to amplitudes $b_l, b_{l+1} > A/3$. To ensure that this does not result in two estimates being generated, one could take all contiguous sets $b_{l_1}, b_{l_1+1}, \ldots, b_{l_2} > A/3$ and prune away every second index l. In doing this we need to respect the periodicity of the $\mathcal{V}_l$: $\mathcal{V}_{L-1} \cap \mathcal{V}_0 \neq \emptyset$, so these contiguous sets may wrap around the circle. Pruning every second index when starting from the middle of one of these contiguous sets may result in two neighbouring $l, l+1$ being removed, which is not what we desire. Instead, in the following pseudocode, after generating the set of all l with sufficient b_l, we find the first gap (in l) between these regions (corresponding to the first $b_l < A/3$). We then iterate (from this point $l_{\min}$ to $L-1$ and then from 0 to $l_{\min}$) over the b_l, and remove each l from our set if $b_{l-1} > A/3$ and $l-1$ was not itself removed.

II.A.3. Proof of Lemma II.3

Lemma II.3 relates the performance of Alg. II.4 to the performance of the QEEP subroutine. We now prove this Lemma. Note that $\{\theta_j\}$ is an ordered list of n_ϕ phases, while $\{\tilde{\theta}_l\}$ is an ordered list (that we will show contains at most n_ϕ phases).

Proof of Lemma II.3 Our proof follows by showing that the output from Alg. II.4 satisfies these statements whenever Eq. (II.84) holds. This yields our confidence bound as Eq. (II.84) holds with probability p by Def. II.5.

To see that statement 2 holds when Eq. (II.84) is satisfied, note that if $\mathcal{V}_l$ contains no phases θ_j with $A_j > A$ and Eq. (II.84) is satisfied, $\tilde{b}_l < \epsilon < A/3$, and l will not be added to the set S in Alg. II.4. This implies that when Eq. (II.84) holds, if $l \in S$ there exists some $\theta_j \in \mathcal{V}_l$ with $A_j > A$, in which case $|\theta_j - \tilde{\theta}_l|_T = |\theta_j - \epsilon l|_T \leq 2\epsilon$.

To see that statement 3 holds when Eq. (II.84) is satisfied, note that $\theta_j \in \mathcal{V}_l \cap \mathcal{V}_{(l+1) \bmod L}$ for exactly one l (and $\theta_j \notin \mathcal{V}_m$ for $m \neq l, (l+1) \bmod L$). Then, when Eq. (II.84) holds, $\tilde{b}_l + \tilde{b}_{(l+1) \bmod L} > 2A/3$, so $\max(\tilde{b}_l, \tilde{b}_{(l+1) \bmod L}) > A/3$, and either l, $(l+1) \bmod L$ or l and $(l+1) \bmod L$ will be added to the set S during step 2 of Alg. II.4 for each phase θ_j. Then, step 4 of Alg. II.4 will remove $(l+1) \bmod L$ if l remains in S, so each phase θ_j can contribute to only one final estimate $\tilde{\theta}_l = l\epsilon$, and the number of estimates is bounded from above by the number of phases.

To see that statement 1 holds, we use the point in the previous paragraph that, when Eq. (II.84) is satisfied, each phase θ_j with $A_j > A$ adds either l, $(l+1) \bmod L$, or l and $(l+1) \bmod L$ to the set S during step 2. of Alg. II.4. Then in step 4. of Alg. II.4, l is removed from S only if $(l-1) \bmod L$ remains

in S, and $(l+1) \bmod L$ is removed from S only if l remains in S. This implies that the distance from θ_j to an estimate is bounded by

$$\max_{\theta_j \in \mathcal{V}_l \cap \mathcal{V}_{(l+1)}} \max_{l'=l-1,l,l+1 \bmod L} |\theta_j - \epsilon l'|_T = \\ = \epsilon(l+1) \bmod L - \epsilon(l-1)) \bmod L = 2\epsilon, \quad \text{(II.87)}$$

as required. □

II.B. Proof of Lemma II.4

Let us first prove the existence of $\kappa_{d+1} \in [2, \kappa_{\max}]$ with $\kappa_{\max} = 3$ that satisfies our conditions (the proof for k_1 is similar), for small enough ϵ in Eq. (II.27). Note that Eq. (II.27) implies

$$\epsilon \leq \frac{\pi}{300} \approx 0.01. \quad \text{(II.88)}$$

Given some pair $\tilde{\phi}_j^{(d)} \neq \tilde{\phi}_l^{(d)}$, $j < l$, let $\Delta_{j,l} = |\tilde{\phi}_j^{(d)} - \tilde{\phi}_l^{(d)}|$ and let $R_{j,l}$ be a set of κ_{d+1} such that neither Eq. (II.37) nor Eq. (II.36) holds for the chosen phases, that is,

$$R_{j,l} = \Big\{ \kappa_{d+1} \in [2, \kappa_{\max}] : \\ |\Delta_{j,l}|_T \geq \frac{\pi - 2\epsilon(1+\kappa_{d+1})}{k_d \kappa_{d+1}} \\ \wedge |k_d \kappa_{d+1} \Delta_{j,l}|_T \leq 4\epsilon(1+\kappa_{d+1}) \Big\}. \quad \text{(II.89)}$$

We call $\cup_{j,l} R_{j,l}$ the forbidden region and want to show that we can choose a value for $\kappa_{d+1} \in [2,3]$ outside this forbidden region if ϵ is sufficiently small. We do this by bounding the size of the forbidden region above and showing that this is smaller than the region $[2,3]$, leaving room to choose κ_{d+1}.

Note that $R_{j,l}$ is nonempty only if

$$k_d \Delta_{j,l} \geq k_d |\Delta_{j,l}|_T \\ \geq \frac{\pi - 2\epsilon(1+\kappa_{d+1})}{\kappa_{d+1}} \\ \geq \frac{\pi - 2\epsilon(1+\kappa_{\max})}{\kappa_{\max}}$$

$$> \frac{73\pi}{225}, \tag{II.90}$$

using Eq. (II.88).

We may write the set $R_{j,l}$ as

$$\begin{aligned} R_{j,l} = &\left[\max\left(2, \frac{\pi - 2\epsilon}{k_d|\Delta_{j,l}|_T + 2\epsilon}\right), \kappa_{\max}\right] \\ &\cap \bigcup_{n\in\mathbb{N}} R_{j,l}^{(n)}, \end{aligned} \tag{II.91}$$

where $R_{j,l}^{(n)}$ is the set of κ_{d+1} for which

$$\begin{aligned} |k_d\kappa_{d+1}\Delta_{j,l}|_T &= \left|k_d\kappa_{d+1}\Delta_{j,l} - 2\pi n\right| \\ &\leq 4\epsilon(1+\kappa_{d+1}), \end{aligned} \tag{II.92}$$

for some $n \in \mathbb{N}$.

Solving this equation for κ_{d+1} yields

$$R_{j,l}^{(n)} = \left[\frac{2\pi n - 4\epsilon}{k_d\Delta_{j,l} + 4\epsilon}, \frac{2\pi n + 4\epsilon}{k_d\Delta_{j,l} - 4\epsilon}\right]. \tag{II.93}$$

The size of each interval $R_{j,l}^{(n)}$ can then be calculated

$$\left|R_{j,l}^{(n)}\right| = \frac{8\epsilon(2n\pi + k_d\Delta_{j,l})}{k_d^2\Delta_{j,l}^2 - 16\epsilon^2}. \tag{II.94}$$

We can bound

$$\begin{aligned} |R_{j,l}| &\leq \left|\bigcup_{n=1}^{n_{\max}} R_{j,l}^{(n)}\right| \leq \sum_{n=1}^{n_{\max}} \left|R_{j,l}^{(n)}\right| \\ &= \frac{8\epsilon}{k_d^2\Delta_{j,l}^2 - 16\epsilon^2}\Big(\pi n_{\max}(n_{\max}+1) \\ &\qquad + k_d\Delta_{j,l}n_{\max}\Big). \end{aligned} \tag{II.95}$$

Here, $n_{\max} = n_{\max}(j,l)$ is the largest index of a set $R_{j,l}^{(n)}$ in Eq. (II.92) for

which $\kappa_{d+1} \in [2, \kappa_{\max}]$. Since $\kappa_{d+1} \leq 3$, Eq. (II.92) implies that

$$n_{\max} \leq \frac{3(k_d \Delta_{j,l} + 4\epsilon) + 4\epsilon}{2\pi}, \tag{II.96}$$

Now using Eq. (II.90) and Eq. (II.88) gives

$$\begin{aligned} \left|R_{j,l}\right| &\leq \frac{1}{1-(4\epsilon/k_d\Delta_{j,l})^2} \\ &\times \frac{30\epsilon(864000\epsilon^2 + 248160\epsilon\pi + 11899\pi^2)}{5329\pi^3} \\ &\leq \frac{3\epsilon(864000\epsilon^2 + 248160\epsilon\pi + 11899\pi^2)}{532\pi^3} \\ &\leq 0.24, \end{aligned} \tag{II.97}$$

where the last inequality used Eq. (II.88). As there are $n_\phi \geq 2$ phases the length of the total forbidden region $\cup_{j,l} R_{j,l}$ is bounded from above by $\frac{n_\phi^2}{2} |R_{j,l}|$. We want to this interval to be, say, at most $1/4$, so that by choosing κ_{d+1} randomly we have a 75% change of not landing in the forbidden interval. For larger n_ϕ we thus should use

$$\epsilon \leq \epsilon_{\text{crit}} = \frac{2\pi}{300 n_\phi^2}, \tag{II.98}$$

leading to Eq. (II.27).

We now repeat the above approach for the special case of finding the multiplier in the first round $\kappa_1 = k_1$. Consider thus $k_1 \in [3n_\phi, \kappa_{\max}]$ with $\kappa_{\max} = 3n_\phi + 1$. The key difference here is that there is a stricter lower bound on this multiplier $\kappa_{\max}$ so that ϵ needs to be chosen smaller, depending on n_ϕ, namely we choose

$$\epsilon_0 \leq \epsilon_{\text{crit},0} = \frac{2\pi}{300 n_\phi^4}, \tag{II.99}$$

as expressed in Eq. (II.26).

Given some pair $\tilde{\phi}_j^{(0)} \neq \tilde{\phi}_l^{(0)}$, $j < l$, and again let $\Delta_{j,l} = |\tilde{\phi}_j^{(0)} - \tilde{\phi}_l^{(0)}|$. Then, let $R_{j,l}$ be the set of k_1 such that neither Eq. (II.37) nor Eq. (II.36) holds for the chosen phases. That is,

$$R_{j,l} = \Big\{ k_1 \in [3n_\phi, \kappa_{\max}] : \\ |\Delta_{j,l}|_T \geq \frac{\pi - 2\epsilon_0(1+k_1)}{k_1}$$

$$\wedge |k_1 \Delta_{j,l}|_T \leq 4\epsilon_0(1+k_1)\Big\}. \tag{II.100}$$

We again call $\cup_{j,l} R_{j,l}$ the forbidden region and want to show that we can choose a value for $k_1 \in [3n_\phi, 3n_\phi + 1]$ outside this forbidden region, assuming that ϵ_0 is small enough. Note that the logic of the first few inequalities in Eq. (II.90) still holds in this new calculation, leading to

$$\begin{aligned}\Delta_{j,l} &\geq \frac{\pi - 2\epsilon_0(1+\kappa_{\max})}{\kappa_{\max}} \\ &\geq \frac{\pi(1 - \frac{2+3n_\phi}{75 n_\phi^4})}{1+3n_\phi},\end{aligned} \tag{II.101}$$

where the second inequality used Eq. (II.99) and the value for $\kappa_{\max}$. Note that for large n_ϕ this allows $\Delta_{j,l}$ to decrease like $\sim 1/n_\phi$, while previously $\Delta_{j,l}$ was lowerbounded by a constant, Eq. (II.90).

This time, we may write the set $R_{j,l}$ as

$$R_{j,l} = \left[\max\left(3n_\phi, \frac{\pi - 2\epsilon_0}{|\Delta_{j,l}|_T + 2\epsilon_0}\right), \kappa_{\max}\right] \cap \bigcup_{n\in\mathbb{N}} R_{j,l}^{(n)},$$

where $R_{j,l}^{(n)}$ is the set of k_1 satisfying

$$\left|k_1\Delta_{j,l} - 2\pi n\right| \leq 4\epsilon_0(1+k_1), \tag{II.102}$$

for some $n \in \mathbb{N}$. Solving this equation for k_1 yields

$$R_{j,l}^{(n)} = \left[\frac{2\pi n - 4\epsilon_0}{\Delta_{j,l} + 4\epsilon_0}, \frac{2\pi n + 4\epsilon_0}{\Delta_{j,l} - 4\epsilon_0}\right]. \tag{II.103}$$

with length

$$\left|R_{j,l}^{(n)}\right| = \frac{8\epsilon_0(2n\pi + \Delta_{j,l})}{\Delta_{j,l}^2 - 16\epsilon_0^2}. \tag{II.104}$$

We can then bound

$$\begin{aligned}|R_{j,l}| &\leq \left|\bigcup_{n=n_{\min}}^{n_{\max}} R_{j,l}^{(n)}\right| \leq \sum_{n=n_{\min}}^{n_{\max}} \left|R_{j,l}^{(n)}\right| \\ &= \frac{8\epsilon_0}{\Delta_{j,l}^2 - 16\epsilon_0^2}\left[\pi(n_{\max}^2 - n_{\min}^2) + (\Delta_{j,l}+\pi)(n_{\max}-n_{\min}) + 2\pi n_{\min} + \Delta_{j,l}\right].\end{aligned} \tag{II.105}$$

Here, $n_{\max} = n_{\max}(j,l)$ and $n_{\min} = n_{\min}(j,l)$ are the largest and smallest indices of sets $R_{j,l}^{(n)}$ in Eq. (II.102) for which $k_1 \in [3n_\phi, 3n_\phi + 1]$. Finding the minimal and the maximal value for n in Eq. (II.102) given the bounds on k_1 gives

$$\frac{3n_\phi(\Delta_{j,l} - 4\epsilon_0) - 4\epsilon_0}{2\pi} \leq n_{\min} \leq n_{\max} \leq \frac{(3n_\phi + 1)(\Delta_{j,l} + 4\epsilon_0) + 4\epsilon_0}{2\pi}. \quad \text{(II.106)}$$

As there are n_ϕ phases, the length of the total forbidden region $\cup_{j,l} R_{j,l}$ is bounded from above by $\frac{n_\phi^2}{2}|R_{j,l}|$. By plugging the bound for ϵ_0 in Eq. (II.99) and the bound for $\Delta_{j,l}$ in Eq. (II.101) together into Eq. (II.105), one can verify that

$$|R_{j,l}|\frac{n_\phi^2}{2} \leq \frac{-16-168n_\phi-621n_\phi^2+552n_\phi^4+14400n_\phi^5+44550n_\phi^6+40500n_\phi^7+73125n_\phi^8+270000n_\phi^9+202500n_\phi^{10}}{450n_\phi^3(-4-9n_\phi-100n_\phi^3-150n_\phi^4+1875n_\phi^7)}, \quad \text{(II.107)}$$

which can be verified to be less than 0.5 for all n_ϕ. Hence a random choice for k_1 in the interval $[3n_\phi, 3n_\phi + 1]$ gives at least a 50% chance to not land in the forbidden region. □

II.C. Range of shifted phase estimates

In this Appendix we prove that when the unitary U is shifted as in step 2 of Alg. II.2, and as long as the output of the phase extraction subroutine (Alg. II.4) meets the promises given in Lem. II.3, all phase estimates of U^{k_d} will lie in the region for which Lem. II.1 holds. This allows us to invoke Lem. II.1 as required during Lem. II.5 and Lem. II.6. Note that if we were to shift the spectrum such that the middle of the largest gap would sit at 0, we would do $U \to Ue^{-i\zeta}$. However, the 'stay-away-from-the-boundary' condition of Lem. II.1 is not symmetric, hence we shift by a different amount which also depends on the error ϵ_0 in the phase estimates.

Lemma II.7
Let $\{\phi_j\}$ be the list of eigenphases of unitary U, and let n_ϕ, $\{\tilde{\phi}_l^{(0)}\}$, ζ, d_ζ be as defined in steps 1 and 2 of Alg. II.2. Assume:

- *(Assumption 1a) For every phase ϕ_j, there exists an estimate $\tilde{\phi}_l^{(0)}$ such that $|\phi_j - \tilde{\phi}_l^{(0)}|_T \leq 2\epsilon_0$.*

- *(Assumption 1b) For every estimate $\tilde{\phi}_l^{(0)}$, there exists a phase ϕ_j such that $|\phi_j - \tilde{\phi}_l^{(0)}|_T \leq 2\epsilon_0$.*

Then for all $k \geq 3n_\phi$, for the estimated shifted eigenphases $\{\tilde{\varphi}_l^{(0)}\}$ it holds that

$$\frac{\pi}{k} + 16\epsilon_0 \leq \tilde{\varphi}_l^{(0)} \leq \frac{\pi(2\lfloor k \rfloor - 1)}{k} - 16\epsilon_0. \tag{II.108}$$

which implies Eq. (II.8) *for $\tilde{\varphi}_l^{(0)} = \phi$. In addition, the eigenphases $\{\varphi_j\}$ of the shifted unitary $Ue^{-i(\zeta + d_\zeta/2 - 8\epsilon_0)}$ satisfy*

$$\frac{\pi}{k} + 14\epsilon_0 \leq \varphi_j \leq \frac{\pi(2\lfloor k \rfloor - 1)}{k} - 14\epsilon_0. \tag{II.109}$$

which again implies Eq. (II.8) *for $\varphi_j = \phi$.*

Proof. Let

$$\tilde{\varphi}_l^{(0)} = \left(\tilde{\phi}_l^{(0)} - \zeta - \frac{d_\zeta}{2} + 8\epsilon_0\right) \bmod 2\pi. \tag{II.110}$$

Let us first show that

$$\tilde{\varphi}_l^{(0)} \in \left[\frac{\pi}{2n_\phi} + 8\epsilon_0, 2\pi - \frac{3\pi}{2n_\phi} + 8\epsilon_0\right]. \tag{II.111}$$

By definition of ζ (as the midway point in the largest gap) and d_ζ (as half the largest gap) we have

$$\left(\tilde{\phi}_l^{(0)} - \zeta - \frac{d_\zeta}{2}\right) \bmod 2\pi \in \left[\frac{d_\zeta}{2}, 2\pi - \frac{3d_\zeta}{2}\right]. \tag{II.112}$$

We have $d_\zeta \geq \pi/n_\phi$ (with equality corresponding to n_ϕ uniformly distributed estimates). By Eq. (II.26) it follows that

$$\epsilon_0 \leq \frac{\pi}{48n_\phi}, \tag{II.113}$$

and thus $\epsilon_0 < \frac{d_\zeta}{16}$, leading to Eq. (II.111). By the assumptions the shifted phases φ_j lie within $2\epsilon_0$ from the estimates $\tilde{\varphi}_l^{(0)}$. Thus for each φ_j there exists a $\bar{\varphi}_l^{(0)}$ such that

$$\begin{aligned}
\varphi_j - 14\epsilon_0 &\geq \tilde{\varphi}_l^{(0)} - 16\epsilon_0 \\
&\geq \frac{\pi}{2n_\phi} - 8\epsilon_0 \\
&\geq \frac{\pi}{2n_\phi} - \frac{8\pi}{48n_\phi}
\end{aligned}$$

$$= \frac{\pi}{3n_\phi} \geq \frac{\pi}{k}. \tag{II.114}$$

and

$$\begin{aligned} \varphi_j + 14\epsilon_0 &\leq \tilde{\varphi}_l^{(0)} + 16\epsilon_0 \\ &\leq 2\pi - 3\left(\frac{\pi}{2n_\phi} - 8\epsilon_0\right) \\ &\leq 2\pi - \frac{3\pi}{k} \\ &\leq \frac{\pi(2\lfloor k \rfloor - 1)}{k}. \end{aligned} \tag{II.115}$$

where we have used Eq. (II.111), Eq. (II.113), $k \geq 3n_\phi$ and $\lfloor k \rfloor > k - 1$. This implies Eq. (II.109) and also Eq. (II.108). □

III. Error mitigation and circuit division for early fault-tolerant quantum phase estimation

III.1. Introduction

Despite significant experimental progress in recent years [1–3, 166], quantum computers still require many orders of magnitude more qubits to achieve advantage for practically relevant computational problems [43, 90, 94]. While quantum error correction (QEC) promises to enable fault-tolerant quantum computing with a cost that scales polylogarithmically in the size of the computation, the overhead required for the most practically realizable schemes is large. For example, the combination of the 2-dimensional surface code with magic state distillation [36, 167–169] is capable of universal quantum computation, but may require hundreds or thousands of physical qubits per logical qubit to ensure the success of the computation. This overhead has prompted a push for work on *early fault-tolerance* [54, 56–61, 64, 67, 71, 72, 78, 92, 99, 170–172], that is, the search for computational frameworks and trade-offs that enable us to take advantage of the machinery of quantum error correction without incurring its full cost.

Initial explorations of early fault-tolerance have focused along a few lines. A fist line of research has aimed to reduce costs for small and medium-sized problems, by making tailored algorithm design choices. For example, using Trotter-based simulation instead of methods with better asymptotic scaling has been shown to offer lower costs for smaller systems [53, 66, 67, 71]. Another approach has focused on adapting quantum algorithms to achieve a low depth per circuit. This is motivated by the observation that a given error rate p, determines the maximum number of allowable operations $G \approx -\log(1-p)^{-1} \approx p^{-1}$ before the exponential decay of circuit fidelity sets in. Several works have accomplished this by trading off reduced circuit depth for an increased number of circuit repetitions [56, 56–65, 99, 173]. A third approach has been to use error mitigation techniques – originally designed to

handle physical noise in near-term quantum devices – in order to reduce the noise-induced bias without reducing circuit depth, albeit at the expense of many circuit repetitions [68, 69, 71].

Quantum error mitigation (EM) [50] has seen increased interest in recent years, due to the development of experimental platforms which are increasingly powerful, yet remain far from enabling fault-tolerant computing. Mitigating errors caused by physical noise avoids the large constant factor requirements of fault tolerance, but results in an exponential overhead with increasing circuit depth and error rate [23, 30, 174–177]. A wide range of EM techniques have been conceived [49–51, 157, 178–181, 181, 182] and implemented in practice [183–188], demonstrating up to thousandfold-reductions in noise bias [187]. While these methods cannot replace scalable error correction, they have been suggested as a last mile solution in fault-tolerant quantum algorithms, relaxing the demands on a QEC protocol by enabling a computation to tolerate $\mathcal{O}(1)$ errors per circuit repetition [50, 54, 68, 69, 72]. Some approaches specifically address scenarios where the error rates vary between qubits or operations [68, 69, 71, 171], with a particular emphasis on the non-Clifford operations that are expected to be the most costly to implement fault-tolerantly. However, most EM protocols have focused on expectation value estimation, as opposed to more general quantum computation; extending beyond this was labeled an outstanding challenge in Ref. [50].

Quantum phase estimation [143], a workhorse of quantum computing, aims to compute an eigenvalue of a unitary matrix given its circuit implementation and an initial state with sufficiently high overlap with the target eigenstate. Traditional phase estimation techniques [29] use a control register which, after accumulating phase information via controlled time-evolution, undergoes the quantum Fourier transform (QFT) to yield a state with high density on binary representations of the desired phase. These *QFT-based QPE* algorithms achieve the Heisenberg limit, the information-theoretic optimum for phase estimation, including constant factors [139]. It is possible to remove the quantum Fourier transform, either by direct dequantization [132, 143], or by analysing the outputs of independent circuits that use a single control qubit via signal processing techniques [57, 59, 60, 62, 74, 75, 136, 138]. (Indeed, under some constraints on U phase estimation without any control qubits is possible [157, 189–191].) These *single-control QPE* techniques are of particular interest in early-fault tolerance [57, 58, 60, 62, 64, 71, 72, 100, 170, 172], as these protocols can tolerate a constant amount of error per circuit [74, 136], avoiding the need for a high overall success probability across all circuit runs. However, it remains to understand how this reduction in circuit depth and increase in error tolerance propagates through quantum error correction to a trade-off between physical qubit count and wall-clock time (of the full

estimation protocol). Additionally, besides Refs. [134, 192], there has been little exploration of whether QFT-based QPE techniques can be divided into smaller circuits akin to single-control methods. To the best of our knowledge, no prior work applied circuit-division to make QFT-based QPE robust to noise.

In this chapter, we develop a framework that jointly optimizes quantum error correction, mitigation, and algorithm design for early fault-tolerant quantum computation. Specifically, we study the trade-off between running fewer, larger circuits at lower error rates (requiring more physical qubits per logical qubit) versus subdividing computations into many smaller circuits that can be executed at reduced code distances. Our work bridges two research directions that have so far been explored independently: the optimization of algorithms for early fault-tolerance and the integration of quantum error mitigation within fault-tolerant protocols. Within this framework, we propose a new noise-robust QPE algorithm, based on subdividing the QFT-based QPE circuit of Ref. [139] used in state-of-the-art fault-tolerant benchmarks [43, 94]. In the presence of global depolarizing noise, we prove that the algorithm converges to any target precision and quantify its cost as a function of noise strength. We compare our method to the noise-optimized robust phase estimation (RPE) scheme of Ref. [144], finding that our approach requires between 30% and 130% of the resources needed for RPE, depending on the target precision and noise strength. Then, in order to mitigate biases from more general noise models, we develop a regularised explicitly-unbiased maximum-likelihood estimator, representing the first extension of error mitigation techniques beyond expectation value estimation. This estimator, when combined with our QPE algorithm, guarantees convergence under arbitrary noise channels. For natural noise models based on local Pauli errors, we upper-bound the overhead of our method relative to the global-depolarizing noise case by a factor of $7\times$. Finally, we present concrete and comprehensive estimates of the the physical qubit count and wall-clock time for computing ground state energies of the two dimensional Hubbard model [94], and of various molecular Hamiltonians compressed with tensor hypercontraction [43]. Our estimates assume surface-code computing [167] with CCZ factories [169], and quantify how the previously discussed method trade increased computation time for reduced qubit counts. We find that the available gains are limited by the the error mitigation overhead, along with the large costs associated to magic state distillation; a $2\times$ reduction in qubit count comes with an increase in the total wall time of $2 - 100\times$.

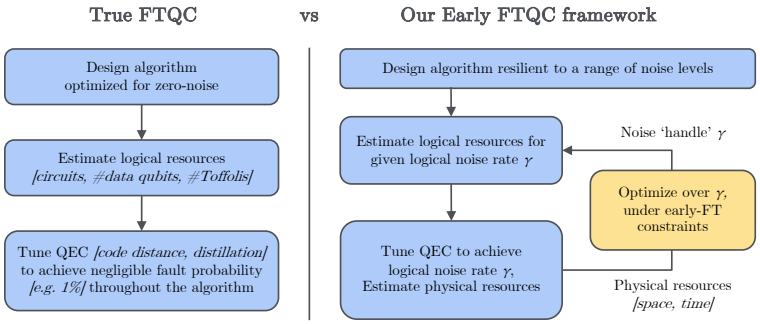


Figure III.1.: Illustration of our framework for designing algorithms for early fault-tolerance and how it differs from traditional fault-tolerant algorithm design and resource estimation. The standard approach (left) involves optimizing an algorithm in the absence of noise and then choosing quantum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

III.2. Summary of results

Our overall goal is to extend recent resource estimates for practical quantum phase estimation [43, 94] to the case of early-fault-tolerant quantum computers. We define early-fault tolerance as *the operational regime where we can use quantum error correction to correct all but* $\Omega(1)$ *errors, and we must design algorithms to be robust to the remainder (via e.g. quantum error mitigation).* In this setup, we retain the ability to spend more physical resources to reduce the logical error rate γ through quantum error correction. In contrast to standard fault-tolerant quantum algorithm design, we consider γ as an additional free parameter to optimize jointly with other parameters of the algorithm, subject to any constraint imposed by the device (e.g. number of available physical qubits). The comparison of the standard fault-tolerant algorithm design process with our early fault-tolerant framework is illustrated in Fig. III.1. We choose this *top-down* approach to early fault-tolerance (which is similar to the approach taken in Refs. [72]) instead of a *bottom-up* approach taken by some other works [54, 78] as this requires relatively few assumptions on the physical device used.

In order to adapt our quantum algorithm to a range of noise rates γ, we explore *circuit division* protocols that replace a single coherent execution of a quantum circuit consisting of G gates by the execution of M shorter circuits, each consisting of less than G/R gates (with $M > R$). The samples from these shorter circuits are then combined by a classical algorithm to recover the same result as the original quantum algorithm. Circuit division allows us to trade between executing a few large circuits at lower error rates γ (thus requiring more physical qubits per logical qubit), versus dividing these into many shorter circuits (which can tolerate higher γ to achieve the same success probability). Reducing the number of physical qubits required would enable a small early-fault-tolerant quantum computer to perform calculations that would otherwise be out of reach, at the cost of increasing the total run time of the calculation by an $\widetilde{\mathcal{O}}(M/R)$ factor. Note that this flavor of circuit division is different than divide-and-conquer approaches which have been proposed for e.g. optimization [193–195] and variational algorithms [196, 197]. These are not known to be noise-robust, and the overhead in most cases is exponential in the number of cuts. For a circuit division protocol to be practical, its overhead M/R needs to be reasonable. Furthermore, the classical post-processing should be robust to faults in multiple sub-circuits, as a union bound implies the probability of failure of at least one sub-circuit is larger than the probability of the original circuit failing.

Error mitigation can further be employed to boost the robustness of circuit-division protocols to the remaining sub-circuit faults. The overhead of EM typically depends on the sub-circuit fidelity F [50], which can be controlled either by limiting the sub-circuit depth or by running more error correction. We provide a general, back-of-the-envelop study of the trade-off between error mitigation and error correction in App. III.A; we find that combining the two can lead to a significant boost of the maximum logical circuit size that can be run on a given device, in a narrow regime of physical qubit count and error rate.

Single-control QPE variants are readily adaptable to circuit division. In the presence of global depolarizing noise, they are known to be robust to faults in the sub-circuits with a quadratic overhead $M \sim R^2$ [72, 99, 136]. As these methods rely on measuring expectation values of Hadamard-test-like circuits, they are naturally compatible with existing EM protocols. This has motivated a lot of interest in recent years [55–57, 60, 61, 75, 190], however little work has gone into testing the robustness of more traditional versions of QPE (that utilize a multi-qubit control register and the quantum Fourier transform).

Our first contribution is to provide a circuit-division variant of QFT-based QPE, *Maximum-likelihood Sin-state QPE* (MSQPE), that is robust to global

depolarizing noise. We give the algorithm in App. III.C.2, and prove the following result.

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

- $\pi\epsilon^{-1}$, *if* $\epsilon \gg \gamma$,
- $C\gamma\epsilon^{-2}$ *with* $C \approx 20$, *if* $\epsilon \ll \gamma$,

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

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

Thm. III.1 demonstrates that QFT-based QPE methods can be made robust to noise, thus allowing for fair comparison with single-control QPE variants. We optimize the hyperparameters of the (QFT-based) MSQPE algorithm (details in App. III.C.2), and compare it against the (single-control) RPE algorithm [74] with optimized parameters from Ref. [144] In Fig. III.2, we plot the ratio between the costs of the two algorithms to achieve the same target error ϵ, where the cost $\mathcal{T}_{\text{tot}}$ is measured as the total number of unitaries across all circuits used. We observe both methods perform similarly; MSQPE performs better when the noise rate γ is low compared to the target error ϵ, by a factor that increases for smaller γ. Vice versa, when the target error is small $\epsilon \ll \gamma$, the cost of RPE is about 30% smaller. In App. III.C.3 we expand on this comparison, and note that both estimators can likely be improved; the MSQPE estimator by optimizing the sin-state in the presence of noise, and the RPE estimator by numerically optimizing the RPE hyperparameters (the optimization in Ref. [144] aimed to achieve analytic bounds). However, across the range of errors and noise rates considered, the performance of both estimators falls within a factor $10\times$ of information-theoretic bounds, which places a limit on any further gains.

While the global depolarizing channel is a useful analytical tool, it fails to accurately describe the residual logical noise in fault-tolerant quantum computers. Previous work has shown that other natural noise models (in particular, local depolarizing noise) can result in a biased signal in single-control QPE methods [136], and thus a biased phase estimate. Consequently, more complex EM strategies will be required to address the general noise models expected to characterize residual logical error. EM introduces a

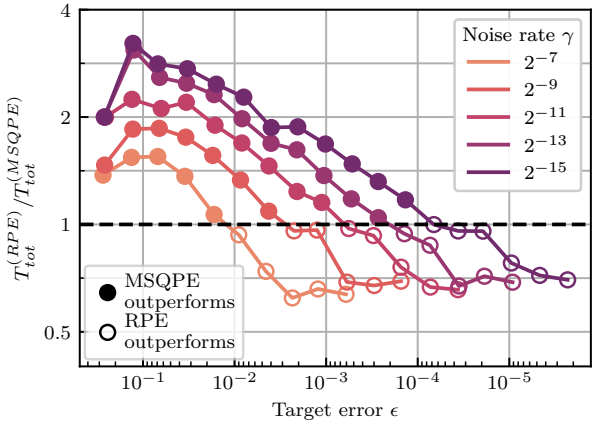


Figure III.2.: Ratio of the total executions times $\mathcal{T}_{\text{tot}}$ of the optimized robust phase estimation (RPE) algorithm of Ref. [144] and the maximum-likelihood sin-state quantum phase estimation (MSQPE) algorithm developed in this work across a range of target errors on the phase estimate ϵ and noise rates γ translating to an error rate per unitary of $p_{\text{err}} = (1 - e^{-\gamma})$. The dashed line at 1 denotes equivalent performance between the two methods; points above 1 give a region where the MSQPE estimator performs better, points below 1 show where RPE outperforms MSQPE.

circuit-repetition overhead that scales inverse-polynomially with the circuit fidelity F.

Our second contribution is to study how overhead associated with error mitigation (defined in terms of circuit repetitions) translates to an overhead in the cost of error-mitigated QPE (in terms of number of calls to a noisy implementation of the unitary operator). We find that this overhead can be suppressed logarithmically thanks to circuit division.

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

This result highlights a significant advantage, as it narrows the cost gap between EM methods with limited scope but low overhead (e.g., postselection) and more general yet costlier approaches (e.g., probabilistic error cancellation).

In Table. III.1, we summarize the circuit-repetition overheads C_{em} known for various methods, and we show how these translate to unitary-call overheads for phase estimation $C_{\mathrm{em}}^{\mathrm{QPE}}$.

	$C_{\mathrm{em}} =$	$C_{\mathrm{em}}^{\mathrm{QPE}}(\gamma, \epsilon) \leq$	
Postselection	ΛF^{-1}	$\Lambda \kappa e/\pi \cdot \gamma/\epsilon$	Symmetry verification, verified phase estimation*
Rescaling	ΛF^{-2}	$\Lambda \kappa e/\pi \cdot 2\gamma/\epsilon$	Echo verification*, direct rescaling under GDN
Explicit unbiasing	ΛF^{-4}	$\Lambda \kappa e/\pi \cdot 4\gamma/\epsilon$	Probabilistic error cancellation, virtual distillation

Table III.1.: Error mitigation overhead for expectation value estimation (C_{em}) and for quantum phase estimation ($C_{\mathrm{em}}^{\mathrm{QPE}}$) for different error mitigation techniques. Λ and κ are constants depending on the circuit, noise model and error mitigation approach, γ is the noise rate and ϵ is the target precision of phase estimation.
* The overhead of echo verification (a.k.a. verified phase estimation) depends on the assumptions on the circuit: if the cost of state preparation is negligible, its overhead is postelection-bounded; if the dominant cost comes from state preparation, it behaves like a rescaling technique due to the required doubling of the circuit depth. See App. III.D for a detailed discussion.

Our third contribution is to develop a general EM protocol for maximum-likelihood estimation. This is motivated by our intention to apply MSQPE in the presence of arbitrary noise. However, our protocol is more general; it can be applied to any estimation problem that would require samples from a noiseless quantum circuit, while only samples from a noisy device with a known error channel are available. Maximum-likelihood is a more general estimator than sample average, typically used for expectation value estimation. To the best of our knowledge, prior EM schemes have exclusively targeted expectation value estimation, leaving the extension of EM to other estimation tasks as an open problem identified in the literature [50].

Our protocol, detailed in App. III.E.1, is based on rewriting the distribution $P(x)$ that describes the outcomes x of a noiseless circuit as a non-convex combination of distributions $Q_j(x)$ that can be sampled on the noisy device:

$P(x) = \sum_j \alpha_j Q_j(x)$. This can be done by using the channel decomposition techniques from probabilistic error cancellation [49, 51]. Given a parameterized model $P(x|\phi)$ for the noiseless distribution, the expected value of the log-likelihood function $\bar{\mathcal{L}}(\phi)$ can then be expanded in terms of Q_j:

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

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

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

We then apply the EUMLE to MSQPE, obtaining a QFT-based QPE algorithm robust to arbitrary noise channels. To ensure the estimator has a bounded sampling cost, we introduce a further regularization which addresses the issue of potential zeros of the likelihood function (see App. III.E.2). Under local depolarizing noise at rate γ, the unitary-call cost of performing phase estimation to precision ϵ is found to be $\mathcal{T}_{\text{tot}} \leq 137\gamma/\epsilon^2$. This corresponds to a QPE error mitigation overhead of approximately $44\gamma/\epsilon$. Comparing to the cost of MSQPE with global depolarizing noise in Theorem III.1, we observe that the additional overhead of mitigating a more general noise channel is approximately $7\times$ (see App. III.E.4).

Finally, we turn to compiling MSQPE for a set of example problems to the surface code, for a range of residual logical noise rates γ. Our goal is to quantify the trade-off between number of physical qubits and total computation time required to estimate ground state energies to a given precision. The details of this resource estimates are given in App. III.F, and we provide open-source code to reproduce and extend our results [198].

To perform this resource analysis, we need to select a specific QEC architecture, but we expect that our framework can be used to extend any such architecture into the early-fault tolerant regime. We consider a quantum computer using a two-dimensional rotated surface code [36], that performs fault-tolerant Clifford gates using lattice surgery [199] and uses CCZ magic state injection and distillation to implement Toffoli gates [169]. Such an

architecture is popular because it can tolerate relatively high physical error rates, and because it can be implemented using a two-dimensional lattice of locally connected qubits [200, 201].

The target Hamiltonians we consider are (1) a set of $L \times L$-sites Fermi-Hubbard models and (2) electronic structure Hamiltonians for a selection of of small- to classically nontrivial molecules. In both cases, we estimate the ground state energy to precision ΔE by applying MSQPE to a qubitized walk operator $\mathcal{W}$ [114]. For the Fermi-Hubbard Hamiltonian, we choose model parameters ($t = 1, u = 4$) and target precision ($\Delta E = 10^{-2}$) following Ref. [94], reproducing a regime that is challenging for classical methods [95]. $\mathcal{W}$ is implemented following the methods of Ref. [94]. Note that qubitization of the Hubbard model does not require data uploading by quantum read-only memories (QROMs), and the only approximation error comes from the rotations used to prepare a two-qubit state encoding the Hamiltonian coefficients. The molecular models we consider span active space sizes between 6 and 26 spatial orbitals. All molecular Hamiltonians were compressed with tensor hypercontraction (THC) [43] after applying symmetry shifts [202] to further reduce the qubitization 1-norm λ. We aim for the chemically desirable $\Delta E = 10^{-3} E_\mathrm{h}$ accuracy of the final results, choosing the THC hyperparameters accordingly. See App. III.F.2 for computational details on the construction of the molecular Hamiltonians.

We estimate the Toffoli costs of phase estimation on the qubitization walk operators $\mathcal{W}$ using Qualtran [203]. We consider using one or multiple CCZ factories [169] to produce magic states for implementing Toffoli gates, and We optimize the data-qubit code distances d and the two-stage distillation code distances d_0, d_1 to minimize the computational volume of $c\mathcal{W}$ while keeping the error rate bounded by γ. We then use the methods detailed earlier in the paper to optimize a robust estimator from oracle calls to the noisy $\mathcal{W}$, with a final target standard deviation of $\Delta\phi = \Delta E/\lambda$ (where λ is the qubitization 1-norm). (Note that this is different than requiring precision $\Delta\phi$ with a fixed success probability, which is a common choice in literature. We expand on this comparison in App. III.F.3)

In Fig. III.3, we plot the resulting estimates of wall-clock time and number of physical qubits required to execute our algorithm. Across the range of experiments considered, we observe that circuit splitting allows for a factor ~ 2 decrease in the qubit count at the cost of a factor ~ 10 increase in wall-clock time. All the curves demonstrate an "elbow-like" behaviour; reducing the number of physical qubits below a specific point incurs a significant penalty in total runtime. The effect becomes more pronounced for larger systems. This aligns with our back-of-the-envelope characterization of circuit division, where we observed there only is a thin region of error rate and number of available

qubits where significant circuit depth enhancements are possible by combining error correction with error mitigation. As anticipated, it is apparent that the simulation of chemistry is more resource-intensive than the simulation of the Hubbard model at the same state-space size. Our analysis makes this quantitative: compared to the $L = 5$ Hubbard model with 25 spatial modes, the 26 spatial orbital Co(salophen) molecule requires roughly four times more qubits and around one and a half orders of magnitude longer wall-clock time. This is due to the higher complexity of molecular Hamiltonians and the costs associated to the QROMs requried to upload the Hamiltonian coefficients for realizing the qubitization oracle.

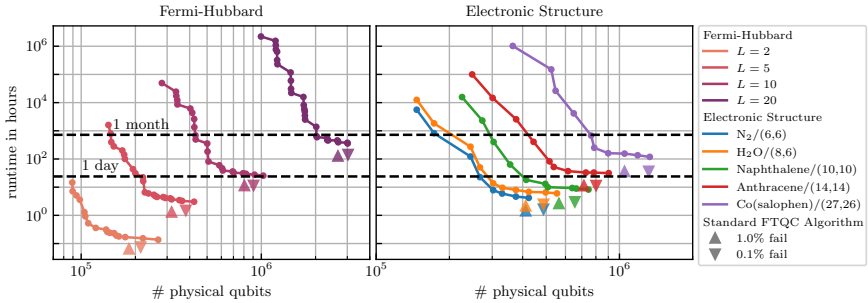


Figure III.3.: Physical costs for MSQPE applied to (left) the qubitized Fermi-Hubbard model Hamiltonian and (right) the active-space molecular Hamiltonians of some chemical systems. For the Fermi-Hubbard Hamiltonian, we choose the hopping parameter $t = 1$ to set the energy units and interaction strength $u/t = 4$. For the electronic structure Hamiltonians, the active spaces sizes (number of electrons and spatial orbitals) are notated in the legend. The target precision on the energy is $\Delta E = 10^{-2}$ for the Hubbard model and $\Delta_E = 10^{-3} E_\mathrm{h}$ for the chemical systems. The physical resources are estimated assuming computation in the surface code using CCZ resource states produced by a single factory; the size of surface code and CCZ factory parameters are chosen to minimize computation volume while keeping the error rate below a choosen γ, assuming physical error rate of 10^{-3}, a surface code clock cycle of 1 µs, and 50% routing overhead. The residual error is assumed to follow a global depolarizing noise model. Changing γ allows to trade between the number of physical qubits (used as resource for implementing better error correction) and total runtime. For comparison, we report the physical costs of the standard fault-tolerant single-circuit implementation of sin-state QPE, accepting a failure probability of 1% or 0.1% (triangles).

III.3. Discussion and Conclusion

Besides various technical contributions to phase estimation and error mitigation, the key focus of this work is to develop a framework that adapts fully fault-tolerant algorithms to the early-fault-tolerant regime and quantifies the savings achieved by trading error rates for physical qubit counts. Such a framework requires a full-stack treatment of quantum algorithm compilation, including a robust treatment of the output of a noisy quantum computer. Improvements in logical gates, QEC codes, or Hamiltonian simulation algorithms can be immediately tested within our code, which relies on Qualtran [203] for computing logical costs, in a plug-and-play fashion. The importance of this adaptability is highlighted by recent advances in T-gate cultivation [44], which will significantly reduce the resource estimates presented here and necessitate going beyond simple T-gate counting. Additionally, our approach allows for exploring alternative QEC architectures, such as those with long-range connectivity and using non-surface codes [3, 166]. As illustrated in Fig. III.3, with current methods, qubit counts can be reduced by a factor $2-3$, but with significant overheads of a factor 10-100 in runtime. This result, while somewhat sobering for early fault-tolerance research, stems from the fact that the gain in physical qubit count from increased logical circuit depth [Eq. (III.9)] is only logarithmic.

We would like this paper to be understood as an invitation to try and beat the provided early-FT estimates, and we can already suggest various improvements that could be achieved within our framework. In this work, we have focused on developing noise-robust post-processing for data sampled using the sine-state QPE circuit, which is optimal only in the noiseless case. Optimizing this circuit while taking noise into account could further improve MSQPE and bring its performance closer to the information-theoretic limit [204], potentially saving a factor ≈ 2 (see App. III.C.3). Additionally, applying the unary-iteration-based technique from Ref.[43] to the iterated qubitized walk operator could yield another $2\times$ reduction in logical costs. Furthermore, in App. III.E.5 we outline a filtering procedure for the EUMLE, which could reduce the error mitigation overhead towards the bound from postselection, potentially achieving a $4\times$ reduction in total cost (analogous to Thm. III.2). Additional savings could also be achieved in the compilation of our algorithms, by allowing higher algorithmic error. For instance, reducing the precision of rotation angles could lower physical resource requirements, though balancing this against energy estimation accuracy remains nontrivial.

We emphasize that the EUMLE introduced in this work has potential applications beyond QPE, in more general estimation problems. On the other hand, the EUMLE needs accurate characterization of the noise model in order to construct a quasiprobability decomposition. Investigating the

robustness of EUMLE against inaccurate noise characterization is a clear direction for future work. Similarly, further research is needed to characterize residual logical noise in specific fault-tolerant computation models. This characterization could be carried out analytically, by analyzing error correcting codes, decoders and distillation protocols, or experimentally, by extending noise characterization techniques such as gate set tomography [205–207] to fault-tolerant computation.

We have aimed to be as complete as possible in the resource estimate of MSQPE. However, we have neglected a significant complicating factor for ground state energy estimation: the preparation of a good initial state. Efficient protocols for preparing high-fidelity initial states are a topic of active research in the field of quantum simulation for chemistry [103, 109, 208, 209], and the cost of preparing a state $|\psi\rangle$ with overlap $a = |\langle\phi|\psi\rangle|$ with the ground state $|\phi\rangle$ can vary widely depending on the problem. If only an approximate ground state is available (i.e. $a \neq 1$) the energy can still be estimated, provided that the first excited state of the Hamiltonian is well separated from the ground state, by an energy gap Δ. This can be achieved in two ways: (1) energy-filtering and amplitude amplification, which allow to disill the exact ground state $|\phi\rangle$ [112, 135]; or (2) classical signal processing techniques that isolate the ground-state phase information [57, 61, 62, 75, 136]. While the first method integrates immediately with our algorithm, it requires long state-preparation circuits, with a cost scaling as $\mathcal{O}(\Delta^{-1}a^{-1})$. This cost needs to be added for each circuit sample, which penalizes circuit division. The second class of methods require only one preparation of $|\psi\rangle$ for each circuit sample, but introduce a sampling overhead scaling as $\mathcal{O}(a^{-2})$ and require QPE circuits of at least depth $\mathcal{O}(\Delta^{-1})$ [165]. Drawing inspiration from these methods, we propose a filtered EUMLE algorithm in App. III.E.5, which we anticipate will achieve similar performance; determining its precise overhead is left to future work. All the options described above establish a lower bound on the minimum circuit depth required for QPE of $\Omega(\Delta^{-1})$, which translates in our framework to a maximum tolerable error rate γ (and thus a minimum required number of qubits). If the error rate exceeds this threshold, the cost of estimating the ground state energy becomes exponential in ϵ [165]. Therefore, we suggest that early fault-tolerant algorithm research should prioritize optimizing initial state preparation and identifying applications where state preparation costs is reasonably low.

III.A. Designing algorithms for early fault tolerance

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

Some works on early fault-tolerance consider other limitations besides size and error rate that might impose limits on quantum error-correction [72]. For example, quantum error correction is designed to suppress uncorrelated errors, but early experimental evidence with superconducting qubits has demonstrated that large correlated errors can arise from cosmic ray impact events [2, 210]. Experimental progress has shown that this effect can be suppressed [211, 212], although unknown mechanisms still appear to set limitations on the performance of quantum error correction in state-of-the-art experiments [1]. In this work, we generally make the assumption that scalable quantum error correction is possible and that any confounding factors will be addressed by scientific and engineering progress. However, it is possible that the early fault-tolerant era will be prolonged by unforeseen difficulties, in which case the techniques we explore may prove especially useful.

There is a growing body of work focused on understanding and maximizing the power of early fault-tolerant quantum computers. One frequently-taken approach is to assume that some combination of techniques will allow us to execute much larger circuits than we can on today's noisy devices, but that practical concerns will still impose limits on the number of qubits and the circuit size. Works in this direction focus on applications, usually picking a simple computational task and exploring different design choices in order to minimize the resources required to perform the task in the absence of error. The focus on early fault-tolerance in these works appears mainly in the choice of the computational task, which might be a toy example designed to demonstrate algorithmic primitives [53], or a scientific application just beyond the reach of classical computation [66, 67]. When the constant factors are taken into account, and when applicable (finite) target precisions are fixed, it is common to find that the simpler methods with worse asymptotic scaling (such as Trotterized time evolution) are superior to their asymptotically optimal

counterparts (such as quantum signal processing) [66, 67].

Other works have focused on developing new algorithmic tools for the setting where we have large but finite upper bounds on the maximum circuit size [55–65]. These algorithms focus on reducing the number of gates, the circuit depth, or the number of ancillary qubits by classically combining the results from multiple separate circuit executions. Underlying many of these approaches is a trade-off between using deeper circuits to estimate some quantity at the Heisenberg limit and using more repetitions of a shallow circuit to estimate the same quantity at the shot-noise limit. For example, Refs. 55, 56, 61 all allow for ground state energy using a total runtime that scales with the precision ϵ as $\epsilon^{-\alpha}$ for an $\alpha \in [1, 2]$ that decreases as the maximum circuit depth is increased. Other works make a related trade-off, choosing between running many copies of a shorter circuit and postselecting on observing some rare event, or obtaining a quadratic speedup with amplitude amplification and the deeper circuits it requires. [60, 65].

III.A.1. Combining quantum error mitigation and quantum error correction

Early fault-tolerant quantum computers won't be able to arbitrarily suppress error rates with a negligible resource overhead, which makes it desirable to consider computations that can tolerate a non-zero error rate. This can be achieved either by developing algorithms that have some natural robustness to error [170, 172, 213], or by augmenting error correction with quantum error mitigation (EM) [54, 68–70]. Years of work in the NISQ era has produced a large body of work on algorithms robust to error and we do not attempt to review that literature here. For similar reasons, we limit our discussion of error mitigation to those works that explicitly explored the combination of quantum error correction and error mitigation. Broadly, error mitigation allows us to estimate noise-free expectation values at the expense of increasing the overall computation time exponentially in the number of expected errors [50]. For example, given a known noise model, probabilistic error cancellation lets us express the noise-free expectation value of an observable as a quasi-probability distribution over quantities sampled from modified versions of the original circuit [49]. In this case, the exponential overhead manifests as an increased variance, or, equivalently, as a larger number of samples required to estimate an expectation value to a fixed precision.

The combination of quantum error mitigation and quantum error correction can take a number of different forms. Motivated by the fact that fault-tolerant non-Clifford gates are particularly costly to implement, some works have focused on combining error-corrected Clifford gates with noisy non-Clifford

gates [68, 69, 71]. By using error mitigation to handle the noise in the non-Clifford gates, such schemes avoid the difficulties and overheads of, e.g., magic state distillation. Other works have explored the use of error mitigation for both Clifford and non-Clifford gates, potentially allowing the use of lower code distances and fewer qubits. In some cases, researchers have explicitly considered the optimization of quantum error correction parameters for use in conjunction with error mitigation [54, 68, 72].

To develop some intuition, we illustrate a toy example of how QEC and EM can be combined in Fig. III.4. In the top panel of this figure, we estimate the depth of the largest circuit that one could perform on 100 logical qubits for a variety of physical error rates and numbers of (physical) qubits. In this model, we demand an "effective sampling rate" of one sample per minute. In other words, the combination of error correction and error mitigation should emulate the effect of an idealized error-free quantum computer that executes the circuit and performs a measurement of a desired observable once per minute. We approximate the effective time per sample using the expression $N_{\text{parallel}}^{-1} D t_{\text{gate}} \gamma^2$, where N_{parallel} is the number of copies of the computation that we can perform in parallel given our physical resources, D is the circuit depth, t_{gate} is the time per gate, and γ^2 is the overhead required to mitigate the logical errors using probabilistic error cancellation. We optimize over the choice of whether or not to use error correction and the code distance d, finding the choice that allows for the largest D such that the effective time per sample is less than one minute.

For simplicity, in this example we assume a circuit composed entirely of CNOT gates executed on a superconducting quantum processor similar to the one described in Ref. [2]. We model the error suppression factor of the surface code using the expression

$$\Lambda = \frac{0.01}{p_{phys}}, \tag{III.2}$$

where 0.01 is an estimate of the surface code threshold and p_{phys} is a single parameter that describes the error rate of physical gate and measurement operations [167]. This means that we can estimate the error per surface code cycle as $0.1\Lambda^{-(d+1)/2}$, where the constant 0.1 is an estimate of the true prefactor. We can perform CNOT gates using lattice surgery using two qubits operating for $3\,d$ surface code cycles, so we can therefore estimate the logical error probability for a gate as

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

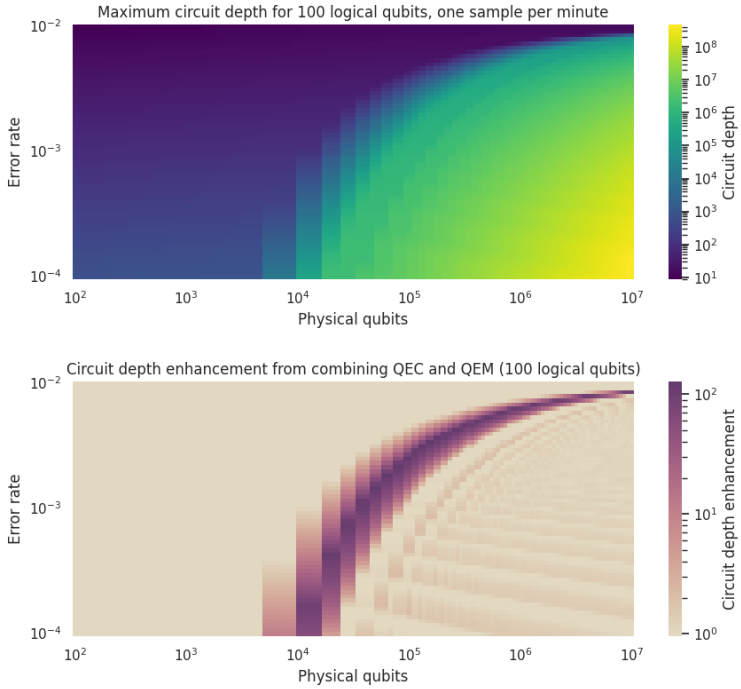


Figure III.4.: (top) Maximum circuit depth for a circuit with 100 logical qubits and an effective sampling rate of one sample per minute over a variety of physical error rates and qubit numbers. We assume a circuit composed entirely of CNOT gates, a raw two-qubit gate time of 20 nanoseconds, and a surface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

using the convention that $d = 1$ indicates that we are not using quantum error correction at all. Using a physical gate time of 20 nanoseconds and a

surface code cycle time of 970 nanoseconds [2], we therefore have

$$t_{\text{gate}} = \begin{cases} 20 \cdot 10^{-9} & \text{if } d = 1 \\ 3 \cdot d \cdot 970 \cdot 10^{-9} & \text{if } d > 1. \end{cases} \tag{III.4}$$

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

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

copies of a 100-logical qubit computation in parallel. Treating $p_{logical}$ as the probability of applying a two-qubit depolarizing channel, we can follow Ref. [49], finding that the overhead incurred by applying probabilistic error cancellation is given by

$$\gamma^2 = \frac{1 + 7/8 p_{logical}}{1 - p_{logical}}^{2G}, \tag{III.6}$$

where $G = 49.5D$ is the number of noisy gates in the circuit.

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

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

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

III.A.2. Circuit division in early fault-tolerance

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

multiple quantum circuits j of depth $D_j < D_{\max}$, given access to a classical post-processing method that is robust against the failure of some of these circuits. Perhaps surprisingly, such circuits exist for many problems. The main contribution of our paper is to explore this "circuit division" in the context of on such problem class — quantum phase estimation (QPE) — to combine this with optimized error mitigation techniques, and compile the resulting algorithms to a realistic physical implementation.

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

Definition III.1
A circuit division of a quantum algorithm $\mathcal{A}$ that requires a circuit of G gates corresponds to a set of M "sub-circuits" of length $\leq G/R$, and a classical post-processing routine which can be applied to the sub-circuits to solve the problem targeted by $\mathcal{A}$.

We target phase estimation routines, for which (under some problem constraints [165]), this division is allowed at a cost $M \sim R^2$. For a circuit with G gates, we can approximate the overall probability for the circuit to execute without error as $1-p_{circuit} = 1-(1-p_{logical})^G \approx e^{-Gp_{logical}}$, where $p_{logical}$ is the error rate per gate. Each logical qubit in the surface code requires $2d^2+1$ physical qubits (not accounting for overheads from routing or magic state distillation), where the code distance d is an odd number that we can choose freely. For a given physical error rate p_{phys}, we can approximate the error suppression factor of the surface code Λ as in Eq. III.2, which implies an error per logical qubit per cycle proportional to $\Lambda^{-((d+1)/2)}$. Without specifying a particular choice of gate and compilation into surface code operations, we can still make the approximation that $p_{logical} \propto d\Lambda^{-(d+1)/2}$ (since most gates require a number of cycles proportional to d and the error rates involved are small). The constant of proportionality is determined by the details of the error model and the particular choice of gates, but we can set it to 1 to obtain a qualitative understanding, yielding

$$1 - p_{circuit} \approx \exp\left(-Gd\Lambda^{-(d+1)/2}\right). \tag{III.8}$$

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

A protocol that replaces a single execution of a circuit with G gates by R^2 executions of circuits with G/R gates increases the overall time required to obtain a solution by a factor of R. If we demand a constant error rate per circuit and neglect logarithmic terms, then Eq. III.8 suggests that we should

take $G \propto \Lambda^{(d+1)/2}$. To obtain a concrete estimate, let us consider $\Lambda = 10$ and $G = 10^6$, which implies that we need a code distance of $d = 11$. If we let d' be the code distance required for a circuit with $\frac{G}{R}$ gates, then we have $d - d' \approx 2 \log_\Lambda R$. The number of physical qubits required per logical qubit is approximately proportional to d^2, so taking $N'_{phys} \propto (d')^2$ and $N_{phys} \propto d^2$, we can approximate the savings in terms of the number of physical qubits as

$$\frac{N'_{phys}}{N_{phys}} \approx 1 - \frac{4 \log_\Lambda R}{d} + \frac{4 \log_\Lambda^2 R}{d^2}. \tag{III.9}$$

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

III.B. Background on quantum phase estimation

The term 'quantum phase estimation' (QPE) refers to a family of computational problems related to estimating eigenphases ϕ_j of a unitary operator U, $U |\phi_j\rangle = e^{i\phi_j} |\phi_j\rangle$, to some target precision ϵ. In this work, we target the estimation of a single, specific eigenvalue ϕ; e.g. if $U = e^{iHt}$ for a Hamiltonian H, one might target the lowest eigenvalue of H. (Note that much recent work on QPE considers the estimation of multiple eigenvalues, see e.g. Refs. [57, 75, 112, 136, 138].) Following common practices in metrology [94], we define 'precision' to mean a bound on the Holevo error of the error in our estimator $\widetilde{\phi}$:

$$\epsilon_H = \sqrt{\mathbb{E}[4 \sin^2(\frac{\widetilde{\phi} - \phi}{2})]} \tag{III.10}$$

(as opposed to the computer science literature [29] where one typically requires $P(|\phi - \widetilde{\phi}| \leq \epsilon) \geq (1 - p)$ for fixed p). This metrological definition is slightly stricter [74] as it requires bounding the tail of the distribution of the estimator $\widetilde{\phi}_j$.

The problem of determining ground states is known to be hard even with a quantum computer (i.e. QMA-hard) for even 2-local Hamiltonians [214]), but it reduces to a problem solvable by a quantum computer (i.e. one in BQP) when one assumes access to an initial state $|\psi\rangle$ which has large overlap $\langle\psi|\phi\rangle$ with the target eigenstate [137, 215]. In this work, we further require that

our target eigenvalue ϕ is separated from other eigenvalues ϕ_j (at least those for which $|\langle\phi_j|\psi\rangle| > 0$) by a gap $\min_j |\phi - \phi_j| =: \Delta \gg \epsilon$. This is necessary by information-theoretic constraints [165] which set Δ^{-1} as a lower bound on the required evolution by controlled-U in order to resolve an eigenvalue (without an overhead that grows exponentially in the number of nearby eigenvalues). This in turn places an upper bound on the amount by which we can divide a circuit, which will later coincide with a minimum fidelity requirement on a noisy implementation of U. Under this assumption, one can project the starting state $|\psi\rangle$ onto the eigenstate $|\phi\rangle$ at a cost proportional to $\Delta^{-1}|\langle\psi|\phi\rangle|^{-1}$ [112, 135]. In this work we absorb this cost onto the cost of initial state preparation.

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

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

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

Existing quantum algorithms for phase estimation can be divided into two classes: *Quantum Fourier Transform (QFT)-based* methods (also called *parallel* or *entanglement-based*) and *single-control based* methods (also called *sequential*, *iterative* or *single-control*). In the first class, the phase estimate is obtained from a single run of a circuit with multiple control qubits. Improving the precision requires increasing the dimension of the control register and the circuit depth. In the second class, the phase is extracted by classically processing expectation values of multiple simpler circuits. Better precision can be achieved either by increasing the depth of the circuits, or the number of samples used to estimate the expectation values. We describe both classes in more details in the following 2 sections.

III.B.1. QFT-based QPE

In this approach we obtain the first n bits of the binary expansion of the phase from a single measurement of an additional n-qubit register. The circuit consists of three stages: first, a probe state is prepared; then, the register is used to control the U; finally QFT. Using n control qubits yields precision of $\epsilon = \mathcal{O}(2^{-n})$ [29].

Possibly the most well-known variant of phase estimation is that presented in Nielsen and Chuang [29], based on previous work from [216]; this is sometimes known in the field as 'textbook' QPE. This algorithm uses two registers, a $\log_2(K)$-qubit control register and a system register. The first step of the algorithm is preparing a "probe state" on the control register (in the textbook algorithm this is done by the Hadamard transform, preparing $\frac{1}{\sqrt{k}}\sum_k^K |k\rangle$). The second step is applying a repeated time-evolution oracle U^k on the system register, with k being controlled by the control register state. Finally, an inverse quantum Fourier transform is applied before measuring the control register in a computational basis. The output of this measurement is a binary expansion of the phase estimate.

It has been shown that, in the absence of noise, the estimator constructed following this algorithm is optimal in terms of accuracy when there is no prior information on the phase to estimate [139]. The probe state that optimizes the Holevo variance of the estimator, however, is not the uniform superposition $\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1} |k\rangle$ but rather the state

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

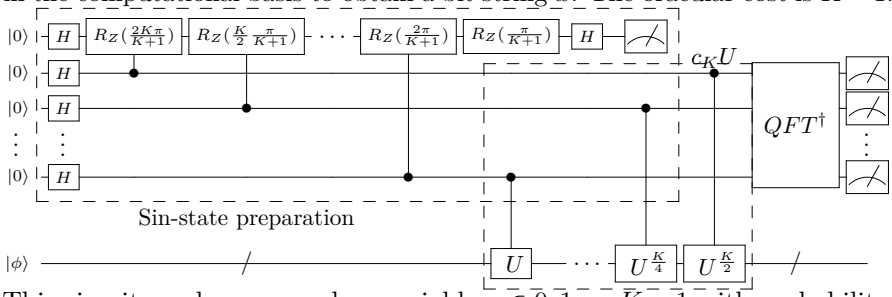
which can be prepared with cost $\mathcal{O}(\log K)$ [94] (see Fig. III.5). The Holevo error of this estimator is $\tan\left(\frac{\pi}{K+1}\right) \approx \pi/K$ [217]. This algorithm has been the most utilized in fault-tolerant application compilation research due to this optimality[43, 89, 94].

Algorithm III.1 (Single circuit sin-state phase estimation algorithm)
Input: target precision ϵ_t, oracle access to initial state $|\phi\rangle$ and cU.

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

[Note that it is also possible to choose $K = \left\lceil \arctan\left(\frac{\pi}{\epsilon_t} - 2\right)\right\rceil \neq 2^n$. This is

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



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

$$P^{(SinQPE)}(x|\phi) = \frac{\sin^2 \frac{\pi}{K+1}}{K(K+1)} \frac{1+\cos[(K+1)(\phi - 2\pi\frac{x}{K})]}{(\cos(\phi - 2\pi\frac{x}{K}) - \cos\frac{\pi}{K+1})^2}. \tag{III.12}$$

discussed later in Fig. III.7 where we explicitly construct the sin-state QPE circuit for general control dimension $K \in \mathbb{N}$ with a log-log overhead.]

We distinguish here between the "target precision" $\epsilon_{\rm t}$ (which is the number given to the algorithm that defines the input parameter Δ), and the actual precision ϵ of the algorithm. As we will see in Sec. III.C, these parameters are no longer the same in the presence of noise.

Theorem III.4 ([217])
Alg. III.1 with $\epsilon_{\rm t} = \epsilon$ solves Problem III.3 with cost $\mathcal{T}_{\rm tot} = \mathcal{T} = K - 1 \approx \pi/\epsilon$, the lowest possible $\mathcal{T}_{\rm tot}$.

III.B.2. Single-control QPE

The $|\log(\epsilon)|$-qubit control register is not essential for a quantum speedup in quantum phase estimation. This has been known since it's conception; Kitaev's original presentation of quantum phase estimation [143] used only a single control qubit, a semi-classical quantum Fourier transform using only one control qubit was presented in Ref. [218] around the same time, and though Shor's original algorithm [219] used a control register, this was simplified

shortly after [220–223]. These original algorithms relied on performing the QFT sequentially; reading bits of ϕ one-by-one via the single control qubit whilst partially projecting the system register into the corresponding eigenstate $|\phi\rangle$. This is unnecessary when $|\phi\rangle$ is already prepared in an eigenstate, in which case the measurement does not affect the system register. Discarding and re-preparing the system register separates the algorithm into classical post-processing of a series of single-qubit Hadamard tests, which can be advantageous in the presence of noise. However, as an optimal probe state is no longer prepared, these single-control QPE algorithms do not obtain the Heisenberg limit [130]. (This is not an issue for NP algorithms such as factoring as one can classically confirm the result and repeat till success.) This problem was solved up to a $\log^*$ factor by the maximum likelihood algorithm of Ref. [153], and later up to a constant factor by the robust phase estimation algorithm (RPE) of Ref. [74], and the Bayesian phase estimation algorithm of Ref. [224]. The RPE algorithm, which repeats lower-order Hadamard tests multiple times to increase the confidence in more significant bits of the phase ϕ was further optimized Ref. [225] and Ref. [144]. The latter paper provided an analytic optimization of the hyperparameters of the RPE algorithm that we will use in this work, proving a separation of around a factor 25 from the strict Heisenberg limit $\mathcal{T}_{\text{tot}} = \pi/\epsilon$. Indeed, this limit cannot be achieved with only a single control qubit; it was shown in Ref. [132] that this requires two control qubits (and can be achieved with such).

The replacement of the control register by a single qubit, and the circuit depth reduction from going to a single qubit has generated much interest in single-control QPE methods for NISQ or early-fault-tolerant research. This was made stronger by the demonstration that one can remove control qubits in some cases [189], which leads to a natural error mitigation strategy via verification of the initial state [157]. However, the restriction to start with an eigenstate is impractical for quantum simulation applications, unless one allows access to state preparation methods such as those in Refs. [112, 135] (which require additional control registers). In Ref. [136], one of the authors demonstrated that a) single-control methods do not require an initial state, and b) that one has a freedom of choice to trade between shorter QPE circuits and more repetitions (with a quadratic overhead as one is reduced to the sampling-noise limit). This freedom of choice is ultimately bounded by the gap between eigenenergies, which was shown earlier in Ref. [165]; the problem of phase estimation of a continuous (or near-continuous) spectrum was formalised as the quantum eigenvalue estimation problem in Ref. [138]. Further work based on Bayesian [226] and integral transform [158] among other methods appeared, however the Heisenberg limit was not achieved with single-control methods without eigenstate access till Ref. [57] and separately

in Ref. [75]. This was optimized further in Refs. [55, 61, 62].

For future use, we now state the RPE algorithm as taken from Ref. [144].

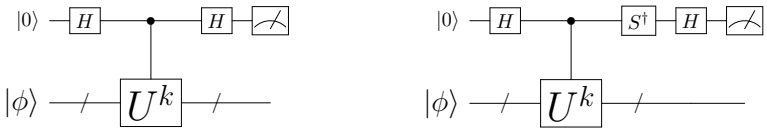


Figure III.6.: Hadamard test circuits with exponent k. Fist, a $|+\rangle$ state is prepared on the control qubit, and state $|\phi\rangle$ is prepared on the system register. Then, cU is applied k times. Finally, control qubit is measured in the X basis to obtain a single bit x. The procedure is repeated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

$$P^{(HT)}((-1)^x + i(-1)^y|\phi) = \frac{1 + (-1)^x \cos(k\phi)}{2} \frac{1 + (-1)^y \sin(k\phi)}{2}. \quad \text{(III.13)}$$

Algorithm III.2 (Robust Phase Estimation (RPE))
Input: number of orders J, vector of numbers of samples $\vec{M} \in \mathbb{N}^J$.

1. *For $j = 0, 1, 2, 3, ..., J-1$:*
   a) *Fix $k = 2^j$.*
   b) *Run M_j repetitions of the circuits of Fig. III.6 with exponent k to obtain a set of bits $\{Z_i\}$.*
   c) *Calculate the average $\bar{Z}$ of measurement results $\{Z_i\}$.*
   d) *Compute $\widetilde{\theta}^{(j)} = \mathrm{Arg}[\bar{Z}] \in [0, 2\pi)$.*
   e) *If $j = 0$, set $\widetilde{\phi}^{(0)} = \widetilde{\theta}^{(0)}$. Else, set $\widetilde{\phi}^{(j)}$ to be the unique value in the interval $[\widetilde{\phi}^{(j-1)} - \frac{\pi}{k}, \widetilde{\phi}^{(j-1)} + \frac{\pi}{k})$ (with periodic boundaries) such that*
   $$k\,\widetilde{\phi}^{(j)} = \widetilde{\theta}^{(j)} \mod 2\pi. \quad \text{(III.14)}$$
2. *Return $\widetilde{\phi} = \widetilde{\phi}^{(J-1)}$.*

In the above algorithm we do not fix the values of the maximum order J, nor the number of repeat measurements at each order M_j. This is because these numbers will change in the presence of global depolarizing noise in the next section. However, in the noiseless setting, near-optimal choices of M_j and J are known thanks to Ref. [74].

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

It was proven in Ref. [74] that for a range of α, β, RPE achieves the Heisenberg limit up to a constant multiplicative factor; $\mathcal{T}_{\text{tot}} \leq c\epsilon^{-1}$. Near-optimal values for these constant factors were found and bounds tightened in Ref. [144] yielding the following result:

Theorem III.5 ([144])
Fix target precision $\epsilon_{\text{t}} > 0$*. With input parameters* $\Delta = 0.409\epsilon_{\text{t}}$*,* $\alpha = 4.0835$*,* $\beta = 11$*, RPE solves Def. III.3 for* $\epsilon_{\text{t}} = \epsilon$ *with oracular cost* $\mathcal{T}_{\text{tot}} \leq 24.26\pi\epsilon^{-1}$ *and maximal depth* $\mathcal{T}_{max} \leq 2\Delta^{-1} \leq 5\epsilon^{-1}$*.*

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

III.B.3. Information theoretic bounds

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

$$\epsilon_H^{(\text{alg})}(\mathcal{T}_{\text{tot}}) = \sqrt{\int_0^{2\phi} \frac{d\phi}{2\pi} \int_0^{2\pi} d\widetilde{\phi} P^{(alg)}(\widetilde{\phi}|\phi) 4\sin^2\left(\frac{\widetilde{\phi} - \phi}{2}\right)}. \tag{III.15}$$

We want to optimize this and find the minimum error that can be achieved at a fixed cost $\mathcal{T}_{\text{tot}}$, but this is not practical as we do not have a simple form of $P^{(\text{alg})}$ for an arbitrary algorithm. Instead, one can fix a choice of quantum circuits (e.g. repetitions of SinQPE with different control dimensions K, or repetitions of single Hadamard tests on U^k for different k), and bound the performance of any classical algorithm on the classical output distributions $\vec{x}$ of these circuits. The Cramer-Rao theorem [227, 228], bounds the Holevo from below by the inverse of the Fisher information

$$\epsilon_H^2 \geq \frac{1}{\mathcal{I}(\phi)}, \qquad \mathcal{I}(\phi) = \int dx P(\vec{x}|\phi) \left[\frac{d\log P(\vec{x}|\phi)}{d\phi}\right]^2. \tag{III.16}$$

For the Hadamard test, it turns out that $\mathcal{I}(\phi)$ is independent of ϕ [75], however this is not true for SinQPE. To compensate for this, it is common to take

the Bayesian Cramer-Rao bound [227, 229], where the Fisher Information is replaced by the average over all possible phases

$$\epsilon_H^2 \geq \frac{1}{\bar{\mathcal{I}}}, \qquad \bar{\mathcal{I}} = \int_0^{2\phi} \frac{d\phi}{2\pi} \mathcal{I}(\phi) \tag{III.17}$$

We will invoke this definition and the Cramer-Rao bound for fixed probability distributions later in this work, in order to bound the error in a QPE algorithm from above. As the Fisher information is additive, we can use invoke these results on a single circuit at a time and add their results. We will find that for all our considered algorithms, the lowest cost is achieved by running an increasing number M of circuits with depth $\mathcal{T}_*$ that maximises the specific Fisher Information $\bar{\mathcal{I}}(\mathcal{T})/\mathcal{T}$ [75, 204],

$$\epsilon_H^2 \leq \frac{1}{\bar{\mathcal{I}}(\mathcal{T})M} = \frac{\mathcal{T}}{\bar{\mathcal{I}}(\mathcal{T})\mathcal{T}_{\text{tot}}} \leq \frac{\mathcal{T}_*}{\bar{\mathcal{I}}(\mathcal{T}_*)\mathcal{T}_{\text{tot}}}. \tag{III.18}$$

In this work we fix our quantum circuits and optimise over the parameters of the probability distributions, but in some cases it is possible to optimize the Fisher information over all choices of measurements on all states generated using a fixed oracle. This yields the channel Fisher information, which characterises the black box in which the parameter ϕ is encoded. For the noiseless version of QPE, Problem III.3, the channel Fisher information can be taken to yield the ultimate bound on the error: the *Heisenberg limit* $\epsilon \geq \frac{\pi}{\mathcal{T}_{\text{tot}}}$ [230]. This is saturated by the probability distribution $P^{(SinQPE)}$ (Eq. III.12) [139]. For a noisy oracle subject to global depolarizing noise with rate γ (see App. III.C), the channel Fisher information can be bounded above, yielding $\epsilon \geq \sqrt{\frac{\gamma}{\mathcal{T}_{\text{tot}}}}$ [75].

III.C. Quantum phase estimation with global depolarizing noise

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

Global depolarizing noise (GDN) [29, 136, 181, 231, 232] is a stochastic noise model where with some probability p a random global unitary is applied to the system, replacing the system's state with the maximally mixed state.

This maps our controlled unitary U_c to the quantum channel $\widetilde{\mathcal{U}}_\gamma^{(GDN)}$, defined by

$$\widetilde{\mathcal{U}}_\gamma^{(GDN)}[\rho] = e^{-\gamma} U_c \rho U_c^\dagger + (1 - e^{-\gamma})\rho_{\mathbb{1}}. \tag{III.19}$$

This definition is exact as global depolarizing noise process commutes with any subsequent unitary operations. This allows us to calculate the effect of repeated GDN channels interleaved by arbitrary operations as

$$(\widetilde{\mathcal{U}}_\gamma^{(GDN)})^k[\rho] = e^{-\gamma k}(U_c)^k \rho (U_c^\dagger)^k + (1 - e^{-\gamma k})\rho_{\mathbb{1}} = \widetilde{\mathcal{U}}_{k\gamma}^{(GDN)}[\rho] \tag{III.20}$$

We see that this yields an exponential decay in circuit fidelity regardless of the unitary; with probability $1 - e^{-\gamma k}$ the system yields the maximally-mixed state, and all useful information about the target problem (e.g. in our case information about the phase ϕ) is lost. This makes global depolarizing noise a particularly simple noise model to analyse, though it comes with an exponential decay in fidelity that cannot be mitigated or even error corrected without an overhead that is exponential in k [50]. This is in contrast to E.g. *local* i.i.d. depolarizing noise on individual qubits. These do not commute with unitary operations, and so a simple form like Eq. (III.20) cannot be found, but this noise model allows error correction when error rates are below the fault-tolerant threshold [167].

In this appendix we optimize and compare two different variants of QPE robust to global depolarizing noise. One of these will be based on the optimized RPE algorithm of Ref. [144], while for the other we develop a robust version of sin-state QPE, using maximum-likelihood estimation. The main outcome of this analysis is a black box that takes in an error ϵ and fixed noise rate γ, and tells the number of calls to the oracle $\mathcal{T}_{\text{tot}}$ that would be required to solve

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

This will be later used in App. III.F to optimize surface code parameters according to the framework defined in Fig. III.1.

III.C.1. Single-control QPE in the presence of noise

If we directly apply RPE (Alg. III.3) in the presence of depolarising noise, the (Holevo) error ϵ_H in the estimate will plateau as the target precision ϵ_t is decreased, in spite of the growing cost. This is because the algorithm requires running ever-deeper circuits, whose output converges exponentially quickly to uniform noise. In Ref. [74], it was shown that the RPE algorithm

is robust to the probabilities $p_j(Z)$ being corrupted by an arbitrary noise term of strength no greater than $\frac{1}{\sqrt{8}}$. For global depolarizing noise this holds for all distributions p_j till $j = j_{\max} = \lfloor \log_2[\frac{\ln(2/\sqrt{8})}{\gamma}] \rfloor$, which estimates ϕ with high confidence to error $\epsilon \leq \frac{2\pi}{2^j} \sim \mathcal{O}(\gamma)$. The algorithm will work with reduced confidence slightly beyond here, but will quickly start to fail. To adapt the algorithm to the presence of noise, we can instead stop at some maximum depth around $\mathcal{T}_{\max} \sim 2^{j_{\max}}$, and repeat these last circuits many times to increase precision at the sampling noise limit (as was suggested in Refs. [75, 144, 204]). To adapt RPE to GDN we identify two distinct regimes: firstly when $\epsilon \gg \gamma$, in which case the number of samples are rescaled to ensure the final error is within the required precision, and secondly when $\epsilon \lessapprox \gamma$, in which case the maximum depth is fixed to some $\mathcal{T}_*$, and most samples are taken at this maximum depth.

As discussed in section III.B.3, the optimal maximal depth $\mathcal{T}_*$ can be quantified by specific Fisher Information [75, 204]. In the case of the Hadamard test with global depolarising noise, the Fisher Information $\bar{\mathcal{I}}_\gamma(\mathcal{T}) = e^{-2\gamma\mathcal{T}}\mathcal{T}^2$ can easily be calculated by substituting the probability distribution from Eq. (III.13) into Eq. (III.17). The depth that maximises $\bar{\mathcal{I}}_\gamma(\mathcal{T})/\mathcal{T}$ is $\mathcal{T}_* = 1/2\gamma$ [75]. Asymptotically, the best performance we can get is given by Eq. (III.18)

$$\mathcal{T}_{\text{tot}} \geq \frac{2\mathcal{T}_*}{\bar{\mathcal{I}}_\gamma(\mathcal{T}_*)\epsilon^2} = \frac{4e\gamma}{\epsilon^2}. \tag{III.21}$$

In practice, this bound will not be saturated, as we need to allocate a number of shots for circuits with lower depths in order to determine the most-significant bits of the phase with sufficiently high confidence. In Ref. [144], the authors optimize the number of measurements M_j of Alg. III.2 to account for global depolarizing noise. This results in a well-defined algorithm, whose cost scales asymptotically as $\mathcal{T}_{\text{tot}} \sim \mathcal{O}(\epsilon^{-2}\gamma)$ (we will later fit the prefactor numerically):

Algorithm III.4
[RPE with global depolarizing noise [144]] Fix a target precision ϵ_{t}, and define the constants

$$J = \lfloor \log_2(1/\epsilon_{\text{t}}) \rfloor, \quad \beta = 11 \quad \textit{if} \quad \epsilon_{\text{t}} > \gamma \tag{III.22}$$

$$J = \lfloor \log_2(1/\gamma) \rfloor, \quad \beta = 11\frac{\gamma^2}{\epsilon_{\text{t}}^2} \quad \textit{if} \quad \epsilon_{\text{t}} < \gamma \tag{III.23}$$

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

Several alternative single-control QPE algorithms to RPE have been pro-

posed for the early-FT regime [55, 57, 61] and studied in the presence of GDN [99, 172]. In Ref. [64] the costs of surface-code implementations of several algorithms were analysed, comparing the performance of various estimators under GDN. The study found that when the initial state is an eigenstate, as in our case, the total runtime is minimized by either RPE or the quantum complex exponential least-squares algorithm of Ref. [61]. For simplicity, we choose RPE with the parameters of Ref. [144] as the benchmark to compare to.

III.C.2. Multi-circuit QFT-based QPE

Similar to the single-control case, in the presence of noise a sin-state QPE circuit will lose information as it grows arbitrarily deep. This effect is even more drastic than for RPE, because a single measurement will be used; if we try to apply Algorithm III.1 using a noisy oracle, the actual error ϵ will become larger for decreasing target error ϵ_t (instead of plateauing). This can be seen from Eq. (III.19); the final measurement will return an estimate with the same error as the noiseless algorithm with probability $e^{-\gamma\mathcal{T}}$, and a fully random estimate with error $\mathcal{O}(1)$ with probability $1-e^{-\gamma\mathcal{T}}$, yielding a Holevo variance

$$\epsilon_H^2(\gamma,\mathcal{T}) = e^{-\gamma\mathcal{T}}\tan^2\left(\frac{\pi}{\mathcal{T}+2}\right) + 2\,(1-e^{-\gamma\mathcal{T}}). \tag{III.24}$$

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

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

$$P_\gamma^{(\text{SinQPE})}(x|\phi) = e^{-\mathcal{T}\gamma}P^{(\text{SinQPE})}(x|\phi) + (1-e^{-\mathcal{T}\gamma})\frac{1}{\mathcal{T}+1} \tag{III.25}$$

where $P^{(\text{SinQPE})}$ is the distribution of the noiseless circuit (Eq III.12). An asymptotically optimal estimator is then to perform maximum-likelihood

estimation on the joint noisy distributions $P_\gamma^{(\mathrm{SinQPE})}(x|\phi)$ taken from different runs. (This is slightly different, but related to the maximum-likelihood estimation performed on single-control QPE data by Ref. [61, 153].) The maximum-likelihood routine can be performed classically with classical cost $\mathcal{O}(\epsilon^{-2})$; due to the constant overheads of quantum computationl, we assume this cost is negligible compared to the total cost of the quantum algorithm $\mathcal{T}_{\mathrm{tot}} \sim \mathcal{O}(\epsilon^{-2})$. We numerically find that for the sin-state circuit with GDN, $\arg\max_{\mathcal{T}} \bar{\mathcal{I}}(\mathcal{T})/\mathcal{T} \approx \gamma^{-1}$. It remains to choose what to do for larger target errors.

Algorithm details

There are 3 regimes characterised by the target precision ϵ relative to the noise strength γ. When $\epsilon \gg \gamma$, the effects of noise are negligible and the optimal strategy is the same as for the noiseless case - run a single circuit of sufficient depth, such that the variance of the output is $\leq \epsilon^2$. As the depth increases, the output becomes more and more influenced by noise till a maximal depth $\mathcal{T}_{\max} \sim \mathcal{O}(\gamma^{-1})$ beyond which running larger depths is never efficient. Instead, asymptotically small errors can be achieved by repeating this single circuit multiple times. In this regime the best possible precision can be achieved by processing the circuit samples through maximum likelihood estimation. Maximum likelihood estimation is only optimal in the limit of a large number of samples; as a consequence, in the intermediate regime ($\gamma \approx \epsilon$), we will need to run a large enough number of samples for circuits of depth smaller than for the asymptotic regime, effectively interpolating circuit depth and number of shots between the two edge cases.

To optimize our measurement strategy for the intermediate regime, we first choose a maximal depth $\mathcal{T}_1$ for any single experiment. With a small number of samples, the error is characterised by the variance of the circuit output probability distribution. We need to find $\mathcal{T}_1$ such that when the total resources are $\mathcal{T}_1$ the expected error is smaller when running a circuit with depth $\mathcal{T}_1$ once, then if the circuit of depth $\mathcal{T}_1/2$ is repeated twice. In the first case, the expected error is simply $\epsilon(\mathcal{T}_1, \gamma)$. In the second case, we get two independent samples, each coming from the target distribution with probability $p = e^{-\gamma \mathcal{T}_1/2}$. There are three possibilities to consider for this second case: if both samples come from the target distribution, the output estimate $\widetilde{\phi}$ is their average, and the error is $\approx \frac{1}{\sqrt{2}} \frac{\pi}{T1/2}$; if one of the samples comes from the target distribution, the error is either $\frac{\pi}{\mathcal{T}_1/2}$ or 2; if both samples come from the noise distribution, the output is random and the error is 2. Summing over these possibilities for the second case, we find the

threshold depth $\mathcal{T}_1$ by solving

$$p^2\frac{\pi^2}{\mathcal{T}_1^2}+(1-p^2)\times 2 \le p^2\times\frac{4\pi^2}{2\mathcal{T}_1^2}+2p(1-p)\times\big(\frac{1}{2}\times\frac{4\pi^2}{\mathcal{T}_1^2}+\frac{1}{2}\times 2\big)+(1-p)^2\times 2. \tag{III.26}$$

One can check this condition is equivalent to

$$\frac{\pi^2}{\mathcal{T}_1^2}\le 2\frac{1-p}{1+3p/4}=8\frac{1-e^{-\gamma\mathcal{T}_1/2}}{4-3e^{-\gamma\mathcal{T}_1/2}}=2\gamma\mathcal{T}_1+O(\gamma^2\mathcal{T}_1^2) \tag{III.27}$$

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

We have not yet chosen the depth $\mathcal{T}_{\max}$ for the small error regime. Again, it is best to choose the depth $\mathcal{T}_{\max}=\mathcal{T}_2$ that maximizes the specific Fisher Information $\mathcal{T}_2=\arg\max_{\mathcal{T}}\mathcal{I}(\mathcal{T};\gamma)/\mathcal{T}$. Integrating $\mathcal{I}(\mathcal{T};\gamma)$ numerically, we verify that the speficic Fisher information is maximized at depth $\mathcal{T}_2=\gamma^{-1}$. In the limit of many samples M (in practice, we assume the estimator has converged for $M>M_2=100$), the variance of the MLE is $(\bar{\mathcal{I}}M)^{-1}$ so to achieve $\epsilon\ll\gamma$ the number of samples required is $M=1/\bar{\bar{\mathcal{I}}}_\gamma(\mathcal{T}_2)\epsilon^2$ leading to the total cost $\mathcal{T}_{\text{tot}}=\mathcal{T}_2M=\mathcal{T}_2/\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)\epsilon^2$.

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

$$\epsilon^2\approx p_F\times 2+(1-p_F)\frac{1}{\bar{\mathcal{I}}M}. \tag{III.28}$$

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

the signal or by the noise; in that case MLE is forced to randomly pick one, and the estimator for two samples reduces to that with one sample, and the probability of success is equal to the fidelity $e^{-\gamma\mathcal{T}}$. We assume that the probability of failure becomes negligible when the number of shots exceeds $M_2 = 100$.

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

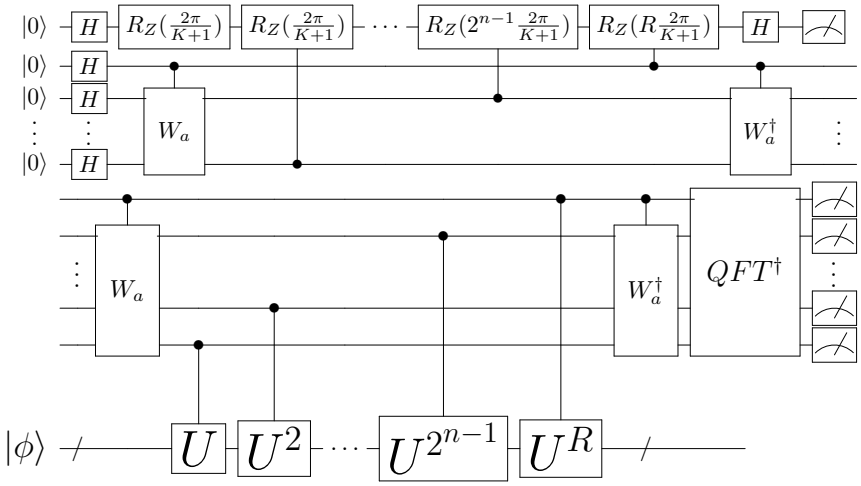


Figure III.7.: Implementation of sin-state QPE circuit for $K \in \mathbb{N}$. Let $n = \lfloor \log_2(K) \rfloor$, $R = K - 2^n$. If $R = 0$ (K is a power of 2), then implement the circuit from Fig. III.5. Otherwise, the $c_K U$ operation can be realised using a control register of $n+1$ qubits in the following way: First, the most significant qubit controls an addition gate $W_a = \sum_{j=0}^{2^{n-1}-a-1} |j+a\rangle\langle j|$ for $a = 2^n - R$ in the remaining register (W_a can be constructed with up to n Toffoli gates and as many auxiliary qubits [233, 234]), and then R applications of U_c. Then, the remaining register controls a $c_{2^n}U$. Finally, the addition of a is uncomputed. We can apply the same trick to the controlled rotations in the preparation of the sin-state to prepare $|s_K\rangle$. QFT of order K can be realised with cost $\widetilde{\mathcal{O}}(\log(K))$ [235, 236]. The oracular cost is $K-1$, as in the case of $K = 2^n$ (Fig. III.5).

Algorithm III.5

[Multi-circuit sin-state QPE with GDN] Input: Target precision ϵ_t, noise

strength γ.

1. *Let the threshold parameters be $\mathcal{T}_1 = \lfloor \sqrt[3]{2\pi^2/3\gamma} \rfloor$, $\epsilon_1 = \epsilon_H(\gamma, \mathcal{T}_1)$ according to Eq. III.24, and $\mathcal{T}_2 = \lfloor \gamma^{-1} \rfloor$, $M_2 = 100$ $\epsilon_2 = 1/\sqrt{\bar{\mathcal{I}}_\gamma(\mathcal{T}_2) M_2}$.*

2. *Choose the number of shots M and circuit depth $\mathcal{T}$:*
   a) *If $\epsilon_\mathrm{t} \geq \epsilon_1$: $\mathcal{T} = \lceil \frac{\pi}{\arctan(\epsilon_\mathrm{t})} - 2 \rceil$, $M = 1$*
   b) *If $\epsilon_\mathrm{t} \leq \epsilon_2$: $\mathcal{T} = \mathcal{T}_2$, $M = 1/\bar{\mathcal{I}}(\mathcal{T}_2, \gamma)\epsilon_\mathrm{t}^2$.*
   c) *If $\epsilon_1 < \epsilon_\mathrm{t} < \epsilon_2$: $\mathcal{T}, M = \arg\min_{\mathcal{T}\in\mathbb{N}, M\in\mathbb{N}}(\mathcal{T} \cdot M)$ under the constraint*

$$2e^{-\gamma\mathcal{T}M/2} + (1 - e^{-\gamma\mathcal{T}M/2})\frac{1}{\bar{\mathcal{I}}_\gamma(\mathcal{T})M} \leq \epsilon_\mathrm{t}^2. \tag{III.29}$$

3. *Run M shots of the SinQPE circuit ((Fig. III.7)) with control dimension $K = \mathcal{T} + 1$ to obtain M samples $\{x_j\}$.*

4. *Return $\arg\min_{\widetilde{\phi}} L(\{x_j\}, \widetilde{\phi})$.*

As before, we again define this algorithm in terms of a parameter ϵ_t that allows us to fix the circuit parameters, but does not necessarily form a bound on the final Holevo error. However, we find numerically that the true Holevo error bound ϵ lies very close to this target.

Performance guarantees

Here we state and prove the formal version of Theorem III.1 in the main text.

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

1. $\lim_{\epsilon\to 0} \lim_{\gamma\to 0} \mathcal{T}_\mathrm{tot}\epsilon = \pi$ *(i.e. $\gamma \ll \epsilon$)*

2. $\lim_{\gamma\to 0} \lim_{\epsilon\to 0} \mathcal{T}_\mathrm{tot}\epsilon^2/\gamma = C$ *(i.e. $\epsilon \ll \gamma$), where $C \approx 20$ is a constant of proportionality.*

Proof. We start by proving statement 1., by noticing that for $\gamma \to 0$ we recover the noiseless case, and Algorithm III.5 behaves identically to Algorithm III.1, as described in Theorem III.4. Specifically, as $\gamma \to 0$ for a fixed ϵ, Alg. III.5 prescribes a single circuit with depth $\mathcal{T} = \lceil \frac{\pi}{\arctan(\epsilon)} - 2 \rceil$, and the total cost is

$\mathcal{T}_{\text{tot}} = \mathcal{T}$. The error ϵ is given by the Holevo variance of the noisy distribution, Eq. (III.24).

$$\epsilon = \sqrt{e^{-\gamma\mathcal{T}_{\text{tot}}} \tan^2\left(\frac{\pi}{\mathcal{T}_{\text{tot}}+2}\right) + 2\left(1-e^{-\gamma\mathcal{T}_{\text{tot}}}\right)}. \tag{III.30}$$

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

$$\lim_{\epsilon\to 0} \mathcal{T}_{\text{tot}}\epsilon = \lim_{\mathcal{T}_{\text{tot}}\to\infty} \mathcal{T}_{\text{tot}} \times \tan\left(\frac{\pi}{\mathcal{T}_{\text{tot}}+2}\right) = \pi. \tag{III.31}$$

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

$$\lim_{\epsilon\to 0} \mathcal{T}_{\text{tot}}\epsilon^2 = \frac{\mathcal{T}_2}{\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)}. \tag{III.32}$$

It remains to calculate the limit $\lim_{\gamma\to 0} \mathcal{T}_2/\gamma\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)$. For brevity of notation, let $f_\mathcal{T}(\Delta\phi)$ be such that the probability distribution in Eq. III.12 can be written as

$$P_{0,\mathcal{T}}(x|\phi) = f_\mathcal{T}\left(\phi - \frac{2\pi x}{\mathcal{T}+1}\right), \tag{III.33}$$

$$f_\mathcal{T}(\Delta\phi) = \frac{\sin^2\frac{\pi}{\mathcal{T}+2}}{(\mathcal{T}+1)(\mathcal{T}+2)} \frac{1+\cos[(\mathcal{T}+2)\Delta\phi]}{\left[\cos(\Delta\phi) - \cos\left(\frac{\pi}{\mathcal{T}+2}\right)\right]^2}. \tag{III.34}$$

The fidelity of the circuit depth $\mathcal{T}$ under GDN with strength γ is $e^{-\gamma\mathcal{T}}$, and the associated probability distribution is $P_{\gamma,\mathcal{T}}(x|\phi) = e^{-\gamma\mathcal{T}}P_{0,\mathcal{T}}(x|\phi) + (1 - e^{-\gamma\mathcal{T}})/(\mathcal{T}+1)$. The Fisher Information is by definition [Eq. (III.17)]

$$\bar{\mathcal{I}}_\gamma(\mathcal{T}) = \int_0^{2\pi} \frac{d\phi}{2\pi} \sum_{x=0}^{\mathcal{T}} \frac{e^{-2\gamma\mathcal{T}} f'^2_\mathcal{T}(\phi - 2\pi\frac{x}{\mathcal{T}+1})}{e^{-\gamma\mathcal{T}} f_\mathcal{T}(\phi - 2\pi\frac{x}{\mathcal{T}+1}) + (1-e^{-\gamma\mathcal{T}})/(\mathcal{T})} \tag{III.35}$$

$$= \sum_{x=0}^{\mathcal{T}} \int_{-\pi-2\pi\frac{x}{\mathcal{T}+1}}^{\pi-2\pi\frac{x}{\mathcal{T}+1}} \frac{ds}{2\pi} \frac{e^{-2\gamma\mathcal{T}} f'^2_\mathcal{T}(s)}{e^{-\gamma\mathcal{T}} f_\mathcal{T}(s) + (1-e^{-\gamma\mathcal{T}})/(\mathcal{T}+1)} \tag{III.36}$$

$$= (\mathcal{T}+1)\int_{-\pi}^{\pi}\frac{ds}{2\pi}\frac{e^{-2\gamma\mathcal{T}}f_{\mathcal{T}}'^2(s)}{e^{-\gamma\mathcal{T}}f_{\mathcal{T}}(s)+(1-e^{-\gamma\mathcal{T}})/(\mathcal{T}+1)} \tag{III.37}$$

$$= (\mathcal{T}+1)e^{-\gamma\mathcal{T}}\int_{-\pi}^{\pi}\frac{ds}{2\pi}\frac{f_{\mathcal{T}}'^2(s)}{f_{\mathcal{T}}(s)+(e^{\gamma\mathcal{T}}-1)/(\mathcal{T}+1)}, \tag{III.38}$$

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

$$\bar{\mathcal{I}}_\gamma(\mathcal{T}) = e^{-\gamma\mathcal{T}}\times(\mathcal{T}+2)(\mathcal{T}+1)\times\int_{-(\mathcal{T}+2)\pi}^{(\mathcal{T}+2)\pi}\frac{1}{2\pi}\frac{(\mathcal{T}+2)^{-2}f_{\mathcal{T}}'^2(\frac{s}{\mathcal{T}+2})ds}{f_{\mathcal{T}}(\frac{s}{\mathcal{T}+2})+(e^{\gamma\mathcal{T}}-1)/(\mathcal{T}+1)}. \tag{III.39}$$

Using L'Hôpital's rule, we can show that in the limit $\mathcal{T}\to\infty$, the function $f_{\mathcal{T}}$ converges pointwise to

$$f(s) = \lim_{\mathcal{T}\to\infty} f_{\mathcal{T}}(\frac{s}{\mathcal{T}+2}) = (2\pi)^2\frac{1+\cos s}{(\pi^2-s^2)^2}, \tag{III.40}$$

and the derivative converges as

$$\frac{1}{\mathcal{T}+2}f_{\mathcal{T}}'\left(\frac{s}{\mathcal{T}+2}\right)\xrightarrow{\mathcal{T}\to\infty} f'(s). \tag{III.41}$$

Therefore, in the limit $\gamma\to 0$, where the fidelity is $e^{\gamma\mathcal{T}_2},\to e$ the integrand converges pointwise to

$$j(s) = \frac{1}{2\pi}\frac{f'^2(s)}{f(s)} \tag{III.42}$$

$$= \frac{1}{(s^2-\pi^2)^2}\left[(1-\cos s)+\frac{16s^2}{(s^2-\pi^2)^2}(1+\cos s)+\frac{8s}{s^2-\pi^2}\sin s\right], \tag{III.43}$$

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

$$|j(s)| \le \frac{1}{(s^2-\pi^2)^2}\left|2+2\frac{16s^2}{(s^2-\pi^2)^2}+\frac{8s}{s^2-\pi^2}\right| \tag{III.44}$$

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

$$\frac{\gamma}{\mathcal{T}_2}\bar{\mathcal{I}}_\gamma(\mathcal{T}_2)\xrightarrow{\mathcal{T}\to\infty} e^{-1}\int_{-\infty}^{+\infty}j(s)ds \equiv C^{-1}. \tag{III.45}$$

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

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

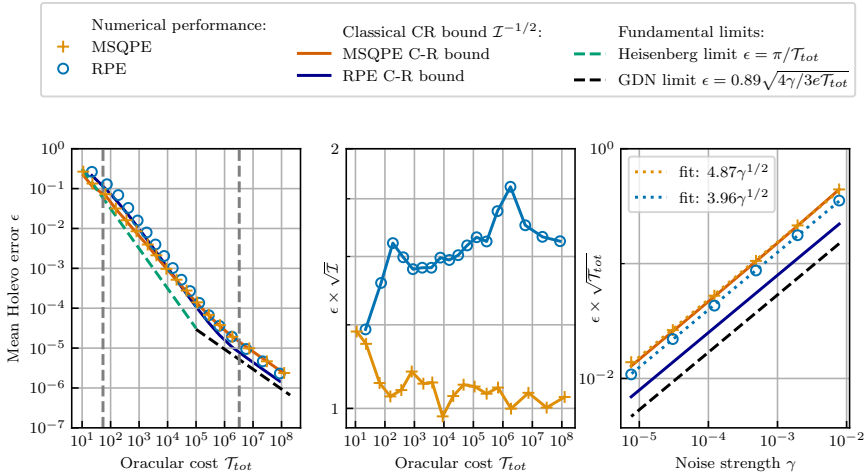


Figure III.8.: Comparison of RPE (Alg. III.4) and MSQPE (Alg. III.5) in the presence of global depolarising noise. (Left) Numerical mean Holevo error ϵ as a function of the total cost $\mathcal{T}_{\text{tot}}$ for a fixed noise rate $\gamma = 2^{-15}$. The vertical dashed lines represent $\mathcal{T}_{\text{tot}}(\epsilon_1)$ and $\mathcal{T}_{\text{tot}}(\epsilon_1)$ separating different regimes in Alg. III.5. The solid lines correspond to the Cramer-Rao bound $\epsilon \geq \bar{\mathcal{I}}^{-1/2}$, where $\bar{\mathcal{I}}$ is the Fisher Information of all the quantum circuits used. (Center) Numerical error ϵ relative to the minimal error allowed by the Cramer-Rao bound $\bar{\mathcal{I}}^{-1/2}$. (Right) Performance in the limit $\epsilon \to 0$ as a function on the noise strength γ. The blue circles and orange pluses are obtained by fitting $\epsilon \propto \mathcal{T}_{\text{tot}}^{1/2}$ to the numerical errors ϵ in the small error regime for different noise rates γ, and the solid lines by fitting $\bar{\mathcal{I}}^{-1/2} \propto \mathcal{T}_{\text{tot}}^{1/2}$. The dotted lines represent the least-squares fit to these points. Error bars in all cases are insignificant.

We compare numerically the MSQPE algorithm against RPE across a range of target precisions ϵ_{t} and error rates γ. For RPE, we choose the optimized parameters from Ref. [144], fixing the target precision $\epsilon_{\text{t}} = 2^{-j}$ to give optimal performance for the algorithm. (As currently defined in Alg. III.4, to target $2^{-(j-1)} > \epsilon_{\text{t}} > 2^{-j}$ has the same cost as estimating at a precision 2^{-j}, though this could be improved upon - e.g. by extending Alg. III.2 in a similar way as in Alg. III.7.) To calculate the Holevo error $\epsilon_H^{(\text{RPE})}$ achieved by RPE, we run 1000 simulations of Alg. III.4 targeting a different random phase on

$[0, 2\pi)$. This yields the orange pluses in Fig. III.8. With the Holevo errors $\epsilon_H^{(\mathrm{RPE})}$ from RPE determined, we use this as a target error $\epsilon_\mathrm{t} = \epsilon_H^{(\mathrm{RPE})}$ for MSQPE. We run 1000 simulations resulting in the blue circles in Fig. III.8. We observe that the mean Holevo error $\epsilon_H^{(\mathrm{MSQPE})}$ for MSQPE is very close to the target precision ϵ_t, which validates our choice of error model used to set the parameters of the algorithm (Eq. (III.28)). For comparison, we plot the Cramer-Rao bounds from the simulated quantum device data that is fed into either algorithm (orange for RPE and blue for SinQPE). We additionally state the standard Heisenberg limit and the global depolarizing noise limit calculated in Ref. [204] (see App. III.B). To construct Fig. III.2, we repeat the above calculation for different noise rate $\gamma = 2^{-m}$, and plot the relative total cost $\mathcal{T}_\mathrm{tot}$ of RPE vs MSQPE at a fixed $\epsilon_H^{(\mathrm{RPE})} \approx \epsilon_H^{(\mathrm{MSQPE})}$.

We observe that the MSQPE achieves the target precision as expected; the orange and blue points on Fig. III.8 [left] have nearly-identical y-values, but differ in the cost $\mathcal{T}_\mathrm{tot}$. The numerical error for the MSQPE is very close to the classical Cramer-Rao bound for the data taken, whilst the error of RPE is about 50% above its Cramer-Rao bound (Fig. III.8 middle panel). However, the Cramer-Rao bounds for the two algorithms are different (as they take different data), so this does not directly translate into a performance advantage of MSQPE at all regimes in Fig. III.2. In the regime of large target error ϵ_t, the noise is not significant, and both algorithms perform as in the noiseless case, achieving near Heisenberg scaling $\epsilon_H \propto 1/\mathcal{T}_\mathrm{tot}$ (Fig. III.8[left]). However, MSQPE has a better scaling constant; it satisfies the Heisenberg limit $\epsilon_H \approx \pi/\mathcal{T}_\mathrm{tot}$, while RPE has a small gap $\epsilon_H \approx 5\pi/\mathcal{T}_\mathrm{tot}$. This matches known results for the noiseless case [132, 139]. The advantage of MSQPE continues in the intermediate regime, and is more significant for smaller noise strengths γ (as can be seen in Fig. III.2). In the limit of small errors both algorithms asymptotically show $\epsilon_H \propto \sqrt{\gamma/\mathcal{T}_\mathrm{tot}}$ as expected. In this regime, RPE outperforms the MSQPE by a factor of ≈ 1.3. Note that MSQPE continues to closely follow the Cramer-Rao bound in this case, which implies that the performance loss is driven by the sin-state being suboptimal in the presence of noise, rather than the classical estimation routine performing badly. For MSQPE, the scaling factor is ≈ 4.9 (which is close to the analytical result in Thm. III.6, as $\sqrt{C} \approx 4.6$), and for RPE ≈ 4.0 - slightly worse than $\sqrt{4e} \approx 3.3$ bound from the specific Fisher Information in Eq. (III.21). Both are significantly ($2\times$) above the fundamental limit $0.89\sqrt{4\gamma/3e} \approx 1.7$ [204].

III.D. Error mitigation overhead for QPE

The depolarizing noise model is a useful toy model to understand the idealized performance of a quantum algorithm (especially in terms of the variance cost), but it fails to capture the bias present in a typical noise model for a quantum device [50]. Over recent years, many error mitigation techniques have emerged, promising a range of fidelity overheads and unbiasing ability. In this section we consider the effect realistic noise and error mitigation techniques would have on the results in App. III.C. Our primary result is to incorporate a fidelity overhead of F^α directly into our $\mathcal{T}$ optimization for QPE. We find that we can trade an increased α (typically $\alpha = 1$, 2, or 4) against $\mathcal{T}$ to yield a multiplicative overall cost $\mathcal{T}_{\text{tot}} \to \alpha\mathcal{T}_{\text{tot}}$. This suggests the overhead for explicitly unbiased methods such as probabilistic error cancellation [49, 51] may be affordable, but it also suggests that improving error mitigation will yield little further gains for QPE.

Error mitigation theory typically focuses on expectation value estimation tasks: measuring to precision ϵ an expectation value $\langle\psi|O|\psi\rangle$ on a state $|\psi\rangle$ produced by a given noiseless quantum circuit [50]. This can theoretically be obtained running the ideal circuit M_1 times. Only having access to a noisy device, an error mitigation scheme attempts to reproduce the same estimate by running a larger number M_F of shots of noisy quantum circuits. The error mitigation overhead $C_{\text{em}}(F) = \frac{M_F}{M_1}$ depends on the fidelity $F = \langle\psi|\,\rho\,|\psi\rangle$ to which the original circuit can be run on the noisy device (where $|\psi\rangle$ is the state that would be produced by a noiseless run of the logical circuit and ρ is the mixed state actually produced by the noisy run on a device).

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

Error mitigation techniques can be classified based on the overhead scaling as a function of F. If errors can be detected deterministically and associated

runs discarded, error mitigation can be achieved with overhead $C_{\text{em}} \sim F^{-1}$. This is known as *postselection bound*, and no error mitigation technique can perform better than this. An example of an error mitigation scheme whose overhead scales as $C_{\text{em}} \propto F^{-1}$ is symmetry verification [179, 180], and another is echo verification when state preparation is cheap compared to measurement [157]. Postselection schemes are typically problem dependent and can only mitigate part of the errors. A more general set of error mitigation schemes relies on measuring a noisy signal and processing it classically to remove the bias, which is estimated analytically under specific assumptions on the problem and noise model. This is particularly effective for global depolarizing noise [237], or when the effect of the noise can be simulated effectively [238]. Due to error propagation in the required rescaling, the overhead for these techniques is $C_{\text{em}} \sim F^{-2}$: we call this *rescaling bound*. Note that echo verification and virtual distillation match this bound when the state preparation is expensive compared to measurement, as the volume of the noisy circuits that need to be run is double than the original circuit [157, 181]. Finally, a last set of error mitigation techniques prescribes to measure explicitly the bias on the signal caused by circuit errors, and subtract it from the biased signal. This is the case of probabilistic error cancellation (PEC), which can completely remove the bias from an observable expectation value by rewriting the noiseless signal as a non-convex combination of signals from noisy "basis" circuits. In the case of local stochastic noise, the overhead of PEC is $C_{\text{em}} \propto F^{-4}$ [49–51]. This can be interpreted as a cost F^{-2} due to the necessity of explicitly measuring the bias and F^{-2} due to the necessity of rescaling of the resulting signal after subtracting the bias.

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

$$\mathcal{U}(\rho) = e^{-\gamma} cU\rho cU^\dagger + (1 - e^{-\gamma})\mathcal{N}(\rho), \tag{III.46}$$

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

Theorem III.7 (Error mitigation overhead for QPE)
Assume an unbiased error mitigation scheme for expectation values with overhead $C_{em}(F) = \Lambda F^{-\nu}$. Then we can realize an error-mitigated single-control QPE protocol that estimates ϕ to an accuracy ϵ with an overhead $C_{em}^{QPE}(\gamma, \epsilon) \propto \nu\gamma/\epsilon$ using the noisy implementation $\mathcal{U}$ of U in Eq. III.46.

Proof. In the absence of noise, we know from Theorem III.4 that the optimal QPE algorithm (sin-state QPE - Alg. III.1) estimates ϕ with a precision ϵ in time $\mathcal{T}_{\text{tot}}^{\text{opt}} = \frac{\pi}{\epsilon}$. From Theorem III.5 we know that, using RPE, we can estimate ϕ to precision $\widetilde{\epsilon} \leq \epsilon$ using noiseless circuits of depth bounded by $\mathcal{T} = \frac{b}{\widetilde{\epsilon}} \approx \frac{b}{\epsilon}$ for a cost $\mathcal{T}_{\text{tot}}^{\text{RPE}}(\widetilde{\epsilon}) = \frac{a}{\widetilde{\epsilon}} \approx \frac{a}{\epsilon}$ (for algorithm-specific constants $a = 5$ and b analytically bounded by $b < 24.26\pi$ and numerically found to be $b \approx 5\pi$). We can recover an estimate to precision ϵ by repeating $\frac{\widetilde{\epsilon}^2}{\epsilon^2}$ times this RPE algorithm to precision $\widetilde{\epsilon}$ and averaging the results; the maximum depth for the RPE circuits remains unchanged, and the total time becomes $\mathcal{T}_{\text{tot}}^{\text{SC}}(\mathcal{T}, \epsilon) = \frac{a\widetilde{\epsilon}}{\epsilon^2} = \frac{\kappa}{\mathcal{T}\epsilon^2}$ (where we defined the overall constant $\kappa := ab$). As the single-control QPE circuits measure expectation values, we can apply error mitigation to recover the noiseless signal. The fidelity of these circuits is bounded by $F \geq e^{-\gamma\mathcal{T}}$, so the error mitigation overhead is bounded by $\mathcal{C}_{\text{em}}(\mathcal{T}, \gamma) \leq \Lambda e^{\nu\gamma\mathcal{T}}$. The QPE error mitigation overhead is then calculated as

$$\mathcal{C}_{\text{em}}^{\text{QPE}}(\mathcal{T}, \gamma, \epsilon) = \frac{\mathcal{T}_{\text{tot}}^{\text{SC}}(\mathcal{T}, \epsilon)\, \mathcal{C}_{\text{em}}(\mathcal{T}, \gamma)}{\mathcal{T}_{\text{tot}}^{\text{opt}}(\epsilon)} \leq \frac{\kappa\Lambda}{\mathcal{T}\pi\epsilon} e^{\nu\gamma\mathcal{T}}$$

$$\xrightarrow[\text{over } \mathcal{T}]{\text{minimize}} \quad C_{\text{em}}^{\text{QPE}}(\gamma, \epsilon) \leq \frac{\kappa\Lambda e}{\pi} \frac{\nu\gamma}{\epsilon}, \quad \text{(III.47)}$$

with the bound minimized by the choice of $\mathcal{T} = \frac{1}{\gamma\nu}$. □

III.E. Explicit unbiasing of multi-control QPE

In sin-state QPE, the estimation task cannot be reduced to sample averaging, as the estimate of the phase is not an expectation value. In this section, we extend explicit unbiasing techniques to a more complex data processing scheme: maximum likelihood estimation. Extending error mitigation techniques beyond expectation value estimation is identified as a relevant open problem in literature [50].

III.E.1. The explicitly-unbiased maximum-likelihood estimator

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

$$\mathcal{U} = \sum_j \alpha_j \mathcal{B}_j. \tag{III.48}$$

For stochastic Pauli channels (which can represent a model for logical noise in surface code even when the underlying physical noise is coherent [239]), the decomposition procedure is known. In particular, for a circuit of volume A and local depolarizing noise with constant rate per volume element r, the decomposition has 1-norm $|\boldsymbol{\alpha}|_1 = e^{2rA} = F^{-2}$ where $F = e^{-rA}$ is the circuit fidelity. (The cost of standard PEC for mitigating expectation values is then $C_{\text{em}} = |\boldsymbol{\alpha}|_1^2 = F^{-4}$.)

In multi-circuit sin-state QPE, we are interested in estimating the phase ϕ given a set of M measurements $\{x_i\}_{i=1,\dots,M}$. In the ideal, noiseless case, these are i.i.d. samples obtained from the SSQPE circuit (see Fig. III.5) with distribution $x \sim P(x) = \text{Tr}\{\Pi_x \mathcal{U}[\rho_0]\}$, where $\{\Pi_x\}$ is the projector of the control register onto the bit-string x and ρ_0 is the input state. We can decompose this distribution using Eq. (III.48) as

$$P(x) = \sum_j \alpha_j Q_j(x) \quad , \quad Q_j(x) = \text{Tr}\{\Pi_x \mathcal{B}_j[\rho_0]\}. \tag{III.49}$$

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

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

$$\hat{\phi} = \arg\max_\phi \mathcal{L}(\phi) \quad , \quad \mathcal{L}(\phi) = \frac{1}{M}\sum_{i=1}^{M} \log P(x_i|\phi). \tag{III.50}$$

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

$$\lim_{M\to\infty} \mathcal{L}(\phi) = \bar{\mathcal{L}}(\phi) = \mathbb{E}\big[\log P(x|\phi)\big| x \sim P(x)\big], \tag{III.51}$$

which we can decompose through Eq. (III.49) as

$$\bar{\mathcal{L}}(\phi) = \sum_j \alpha_j \, \mathbb{E}\big[\log P(x|\phi) \big| x \sim Q_j(x)\big] \tag{III.52}$$

$$= |\boldsymbol{\alpha}|_1 \, \mathbb{E}\left[\operatorname{sgn}(\alpha_j) \, \mathbb{E}[\log P(x|\phi)|x \sim Q_j(x)] \Big| j \sim \frac{|\alpha_j|}{|\boldsymbol{\alpha}|_1}\right] \tag{III.53}$$

$$= |\boldsymbol{\alpha}|_1 \, \mathbb{E}\Big[\operatorname{sgn}(\alpha_j) \log P(x|\phi) \Big| j, x \sim \widetilde{Q}(j,x)\Big] \tag{III.54}$$

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

$$\widetilde{\mathcal{L}}(\phi) = \frac{|\boldsymbol{\alpha}|_1}{\widetilde{M}} \sum_{i=1}^{\widetilde{M}} \operatorname{sgn}(\alpha_{j_i}) \log P(y_i|\phi). \tag{III.55}$$

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

Definition III.6
[Explicitly unbiased maximum likelihood estimator] Given (1) a parameterized model $P(x|\phi)$ for the distribution of x obtained from a noiseless quantum circuit, (2) a non-convex decomposition of $P(x|\phi^*) = \sum_j \alpha_j Q_j(x)$ in terms of distributions $Q_j(x)$ that can be sampled using a noisy quantum device; the explicitly unbiased maximum likelihood estimator (EUMLE) of the phase ϕ^* is defined as $\widetilde{\phi} = \arg\max_\phi \widetilde{\mathcal{L}}(\phi)$, where $\widetilde{\mathcal{L}}(\phi)$ in Eq. (III.55) *is constructed by sampling $\widetilde{M}$ points $\{(j_i, x_i)|j_i \sim |\alpha_{j_i}|/|\boldsymbol{\alpha}|_1, x_i \sim Q_{j_i}(x_i)\}_{i=1,\dots,M}$.*

The EUMLE is a consistent estimator, i.e. in large samples limit $\widetilde{M} \to \infty$ it converges to the true value of ϕ. In the absence of regularization ($c = 0$), $\widetilde{\mathcal{L}}_c(\phi) = \widetilde{\mathcal{L}}(\phi) \to \bar{\mathcal{L}}(\phi)$ asymptotically matches the log-likelihood of the noiseless data Eq. (III.51). Here we prove the consistency of the EUMLE by showing that, even for $c > 0$, the maximum of the likelihood for $\widetilde{M} \to \infty$ remains the same. We also prove the stronger property of asymptotic normality and recover an expression for the asymptotic variance of the estimator.

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

Proof. For this proof we follow a strategy similar to the one use to prove asymptotic normality of the maximum likelihood estimator in Theorem 10.1.12 of [229]. First, we Taylor-expand $\tilde{\mathcal{L}}'(\phi)$ (the primes are derivatives w.r.t. ϕ) around the true value $\phi^* = \arg\min_\phi \tilde{\mathcal{L}}(\phi)$:

$$\tilde{\mathcal{L}}'(\phi) = \tilde{\mathcal{L}}'(\phi^*) + (\phi - \phi^*)\tilde{\mathcal{L}}''(\phi^*) + \mathcal{O}((\phi - \phi^*)^2). \tag{III.56}$$

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

$$\tilde{\phi} - \phi^* = \frac{-\tilde{\mathcal{L}}'(\phi^*)}{\tilde{\mathcal{L}}''(\phi^*)}. \tag{III.57}$$

The denominator is

$$\begin{aligned}
\tilde{\mathcal{L}}''(\phi^*) &= \frac{|\boldsymbol{\alpha}|_1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} \operatorname{sgn}(\alpha_{j_i}) \left[\partial_\phi^2 \log P(y_i|\phi)\right]_{\phi^*} && \text{(III.58)}\\
&\xrightarrow[\tilde{M}\to\infty]{\text{p}} |\boldsymbol{\alpha}|_1 \, \mathbb{E}\big[\operatorname{sgn}(\alpha_j) \left[\partial_\phi^2 \log P(y|\phi)\right]_{\phi^*} \big| j, y \sim \tilde{Q}(j,y)\big] && \text{(III.59)}\\
&= \sum_j \alpha_j \int \mathrm{d}Q_j(y) \; \left[\partial_\phi^2 \log P(y|\phi)\right]_{\phi^*} && \text{(III.60)}\\
&= \mathbb{E}\big[\left[\partial_\phi^2 \log P(y|\phi)\right]_{\phi^*} \big| y \sim P(y)\big] && \text{(III.61)}\\
&= \mathbb{E}\big[\underbrace{\left[\partial_\phi \log P(y|\phi)\right]_{\phi^*}^2}_{[s(y)]^2} \big| y \sim P(y)\big] =: \mathcal{I}(\phi^*) && \text{(III.62)}
\end{aligned}$$

where the convergence in probability $\xrightarrow{\text{p}}$ is justified by the law of large numbers, and $\mathcal{I}(\theta^*)$ is the Fisher information of an ideal sample x from the noiseless distribution $P(x)$. For brevity of further notation, we define the score of the probability model $s(x) := [\partial_\phi \log P(x|\phi)]_{\phi=\phi^*}$. By the central limit theorem, the numerator

$$-\tilde{\mathcal{L}}'(\phi^*) = -\frac{|\boldsymbol{\alpha}|_1}{\tilde{M}} \sum_{i=1}^{M'} \operatorname{sgn}(\alpha_{j_i}) s(y_i) \tag{III.63}$$

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

totically normal with mean 0 and variance

$$\sigma^2 = \frac{|\boldsymbol{\alpha}|_1^2}{\widetilde{M}} \frac{\mathbb{E}\big[s^2(y)\big|y \sim \tilde{Q}(y)\big]}{\mathbb{E}\big[s^2(y)\big|y \sim P(y)\big]^2}. \tag{III.64}$$

□

III.E.2. Regularizing the EUMLE

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

$$\widetilde{\mathcal{L}}_c(\phi) = \frac{|\boldsymbol{\alpha}|_1}{\widetilde{M}} \sum_{i=1}^{\widetilde{M}} \operatorname{sgn}(\alpha_{j_i}) \log P_c(y_i|\phi) + \frac{c}{|\mathcal{D}|} \sum_{x\in\mathcal{D}} \log P_c(x|\phi), \tag{III.65}$$

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

For periodic distributions with continuous domain $\mathcal{D} = [0, 2\pi)$ where $P(x|\phi) = P(x - \phi \mod 2\pi)$ holds, the second term is not dependent on ϕ and can be neglected. Such a distribution can be obtained from QPE adding a random offset to the phase being estimated [240, 241]. Here, however, we introduce a general regularized EUMLE which uses this modified likelihood and prove its asymptotic normality.

Definition III.7
[regularized EUMLE] The regularized Explicitly Unbiased Maximum Likelihood Estimator (rEUMLE) of the phase ϕ^* is derived from the EUMLE (definition III.6); it requires the same inputs plus an additional regularization constant c, it uses the same samples from the quantum device, and is defined as $\tilde{\phi} = \arg\max_\phi \widetilde{\mathcal{L}}_c$ using the regularized log likelihood Eq. III.65.

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

$$\sigma^2 = \frac{|\boldsymbol{\alpha}|_1^2}{\widetilde{M}} \frac{\mathbb{E}\big[s_c^2(y)\big|y \sim \widetilde{Q}(y)\big]}{\mathbb{E}\big[s_c^2(y)\big|y \sim P(y)\big]^2}, \tag{III.66}$$

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

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

$$\begin{aligned}\mathcal{L}_c(\phi) &\equiv \lim_{\widetilde{M}\to\infty} \widetilde{\mathcal{L}}_c(\phi) && \text{(III.67)}\\ &= |\boldsymbol{\alpha}|_1 \,\mathbb{E}\left[\operatorname{sgn}(\alpha_j)\log P_c(x|\phi) \;\middle|\; (j,x)\sim \frac{|\alpha_j|}{|\boldsymbol{\alpha}|_1} Q_j(x)\right] + \\ &\quad + \frac{c}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} \log P_c(x|\phi) && \text{(III.68)}\\ &= \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} P(x)\log P_c(x|\phi) + c\log P_c(x|\phi) && \text{(III.69)}\\ &= \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} P_c(x)\log P_c(x|\phi). && \text{(III.70)}\end{aligned}$$

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

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

$$\mathcal{L}_c''(\phi^*) = -\frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} \frac{[\partial_\phi P(x|\phi)]^2_{\phi=\phi^*}}{P_c(x|\phi^*)} < 0. \qquad \text{(III.72)}$$

The proof of asymptotic normality then follows the same steps as the proof of Thm. III.8 using the regularized versions of the likelihood $\widetilde{\mathcal{L}}_c$ and score s_c. □

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

$$\mathcal{C}_{\text{EUMLE}} = \frac{\widetilde{M}}{M} = |\boldsymbol{\alpha}|_1^2 \frac{\mathbb{E}\left[s_c^2(y)\middle|y\sim\widetilde{Q}(y)\right]}{\mathbb{E}\left[s_c^2(x)\middle|x\sim P(x)\right]} \frac{\mathbb{E}\left[s_0^2(x)\middle|x\sim P(x)\right]}{\mathbb{E}\left[s_c^2(x)\middle|x\sim P(x)\right]}. \qquad \text{(III.73)}$$

Here, the term $\mathbb{E}[s_0^2(x)|x\sim P(x)]$ is the Fisher information of the noiseless distribution. The overhead consists of three factors: 1. the rescaling constant $|\boldsymbol{\alpha}|_1^2$, 2. the ratio between the average (regularized) score of and the importance-sampled noisy distribution and the noiseless distribution, and 3. the overhead

due to the loss of information caused by regularization. The regularization constant c ensures term 2 remains bounded, but it also causes term 3 to increase, so it should be chosen to optimize the product of the two. Quantifying $\mathcal{C}_{\text{EUMLE}}$ requires additional assumptions about the noise model and the consequent basis circuit decomposition, and about the specific model $P(x|\phi)$.

III.E.3. Applying the regularized EUMLE to sin-state QPE

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

Theorem III.10 (EUMLE for Sin QPE)
We consider the rEUMLE (Def. III.7) applied to the SinQPE probability distribution [Eq. (III.12)*], which can be expressed with a PEC decomposition [Eq.* (III.49)*] with one-norm $|\boldsymbol{\alpha}|_1$. For any constant c, there exists a threshold depth $\mathcal{T}_{\text{thr}}$ and a constant B such that, for any SinQPE circuit with depth $\mathcal{T} > \mathcal{T}_{\text{thr}}$, the variance of the EUMLE σ^2 is bounded as*

$$\sigma^2 \leq B\frac{|\boldsymbol{\alpha}|_1^2}{M\mathcal{T}^2}, \tag{III.74}$$

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

Proof. The goal of this theorem is to bound the variance of the rEUMLE Eq. III.66 for the sin-state QPE outcome probability distribution (from Theorem III.6)

$$P_{\mathcal{T}}(x|\phi) = f_{\mathcal{T}}\left(\phi - \frac{2\pi x}{\mathcal{T}+1}\right), \tag{III.75}$$

$$f_{\mathcal{T}}(\Delta\phi) = \frac{\sin^2\frac{\pi}{\mathcal{T}+2}}{(\mathcal{T}+1)(\mathcal{T}+2)}\frac{1+\cos[(\mathcal{T}+2)\Delta\phi]}{\left[\cos(\Delta\phi) - \cos\left(\frac{\pi}{\mathcal{T}+2}\right)\right]^2}. \tag{III.76}$$

In order to bound σ^2, we calculate asymptotic bounds for $\mathbb{E}\left[s_c^2(y)\middle|y \sim P(y)\right]$ (bounded from below as $\Omega(\mathcal{T})^2$) and for $\mathbb{E}\left[s_c^2(y)\middle|y \sim \widetilde{Q}(y)\right]$ (bounded from above as $\mathcal{O}(\mathcal{T}^2)$). This allows us to bound σ^2 for sufficiently large $\mathcal{T}$.

We first show that $\frac{1}{T^2}\mathbb{E}\left[s_c^2(y)\middle|y \sim P(y|\phi)\right]$ converges to a constant a in the

$\mathcal{T} \to \infty$ limit. Substituting the definition of the probability we can write

$$\frac{1}{\mathcal{T}^2}\,\mathbb{E}\big[s_c^2(y)\big|y\sim P(y|\phi)\big] = \sum_{x=0}^{\mathcal{T}} \frac{\mathcal{T}^{-2} f_{\mathcal{T}}'\left(\phi-\frac{2\pi x}{\mathcal{T}+1}\right)^2}{\left[f_{\mathcal{T}}\left(\phi-\frac{2\pi x}{\mathcal{T}+1}\right)+c\right]^2} f_{\mathcal{T}}\left(\phi-\frac{2\pi x}{\mathcal{T}+1}\right). \tag{III.77}$$

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

$$\begin{aligned}\frac{1}{\mathcal{T}^2}\,\mathbb{E}\big[s_c^2(y)\big|y\sim P(y|\phi)\big] &\to \frac{\mathcal{T}+1}{2\pi}\int_0^{2\pi} ds\, \frac{\mathcal{T}^{-2} f_{\mathcal{T}}'(\phi-s)^2}{\left[f_{\mathcal{T}}(\phi-s)+c\right]^2} f_{\mathcal{T}}(\phi-s)\\ &= \frac{\mathcal{T}+1}{2\pi}\int_{-\pi}^{\pi} ds\, \frac{\mathcal{T}^{-2} f_{\mathcal{T}}'(s)^2}{\left[f_{\mathcal{T}}(s)+c\right]^2} f_{\mathcal{T}}(s) =\\ &= \frac{\mathcal{T}+1}{2\pi(\mathcal{T}+2)}\int_{-(\mathcal{T}+2)\pi}^{(\mathcal{T}+2)\pi} ds'\, \frac{\mathcal{T}^{-2} f_{\mathcal{T}}'\left(\frac{s'}{\mathcal{T}+2}\right)^2}{\left[f_{\mathcal{T}}\left(\frac{s'}{\mathcal{T}+2}\right)+c\right]^2} f_{\mathcal{T}}\left(\frac{s'}{\mathcal{T}+2}\right)\end{aligned} \tag{III.78}$$

where, in the second step, we used the periodicity and evenness of $f_{\mathcal{T}}$ to remove the dependence on ϕ and shift the integration bounds, and in the third step we performed a change of variable $s' = (\mathcal{T}+2)s$. The $\mathcal{T}\to\infty$ limit of this integral can be calculated following the same steps as in the proof of Thm. III.6. Using the convergence property $\lim_{\mathcal{T}\to\infty} f(s/(\mathcal{T}+2)) = f(s)$ with $f(s)$ in Eq. (III.40), we can write

$$\frac{1}{\mathcal{T}^2}\,\mathbb{E}\big[s_c^2(y)\big|y\sim P(y|\phi)\big] \xrightarrow{\mathcal{T}\to\infty} \int_{-\infty}^{\infty} \frac{ds}{2\pi}\frac{f'(s)^2}{\left[f(s)+c\right]^2} f(s) =: a. \tag{III.79}$$

This integral converges to a finite constant a, as the integrand (independent on $\mathcal{T}$) can be bounded by $j(s)$ defined in Eq. (III.42), which can be bounded by $\mathcal{O}(|s|^{-4})$ [see Eq. (III.44)].

Secondly, we show $\lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}^2}\,\mathbb{E}\big[s_c^2(y)\big|y\sim\widetilde{Q}(y)\big]$ is bounded by bounding $\lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}}|s_c(x)|$ for any x. By definition, the score is,

$$s_c(x) = [\partial_\phi \log P_c(x|\phi)]_{\phi=\phi^*} = \frac{f_{\mathcal{T}}'(\frac{2\pi x}{\mathcal{T}+1}-\phi)}{f(\frac{2\pi x}{\mathcal{T}+1}-\phi)+c}, \tag{III.80}$$

and we can bound it as $|s_c(x)| \le |f_{\mathcal{T}}'(\frac{2\pi x}{\mathcal{T}+1}-\phi)|/c$. As per Eq. (III.41), the limit function $f'(s) = \lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}}f'(s/(\mathcal{T}+2))$ is continuous and $f'(s)\xrightarrow{s\to\pm\infty}0$ so f' is bounded by a constant $b := \sup_s |f'(s)|$. Combining all the steps, we

get the bound

$$\lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}^2}\,\mathbb{E}\big[s_c^2(y)\big|y\sim\widetilde{Q}(y)\big] \le \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}^2}\max_x |s_c^2(x)| \tag{III.81}$$

$$\le \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}^2}\frac{\max_s |f'_{\mathcal{T}}(s)|^2}{c^2} \le \frac{b^2}{c^2}. \tag{III.82}$$

Combining the above limits,

$$\lim_{\mathcal{T}\to\infty} \mathcal{T}^2 \frac{\mathbb{E}\big[s_c^2(y)\big|y\sim\widetilde{Q}(y)\big]}{\mathbb{E}\big[s_c^2(y)\big|y\sim P(y)\big]^2} \le \frac{b^2}{a^2c^2}. \tag{III.83}$$

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

$$\sigma^2 = \frac{\mathbb{E}\big[s_c^2(y)\big|y\sim\widetilde{Q}(y)\big]}{\mathbb{E}\big[s_c^2(y)\big|y\sim P(y)\big]^2}\frac{|\boldsymbol{\alpha}|_1^2}{\widetilde{M}\mathcal{T}^2} \le B\frac{|\boldsymbol{\alpha}|_1^2}{\widetilde{M}\mathcal{T}^2}. \tag{III.84}$$

Numerically, we find that the limits for $\mathcal{T}\to\infty$ act as bounds in the correct direction: for any $\mathcal{T}$, $\max f'_{\mathcal{T}} \le (\mathcal{T}+2)\max f'$, and $\mathbb{E}\big[s_c^2(y)\big|y\sim P(y)\big] \ge a(\mathcal{T}+2)^2$. We find the value of $b\approx 0.16$. For an example choice of $c=0.4$, we find $a\approx 0.01$. Hence, the bound on the variance Eq. (III.74) is satisfied for any depth $\mathcal{T}$ for $B = b^2/a^2c^2 \approx 12.6$. □

III.E.4. Overhead for QPE with local Pauli noise

To quantify the EUMLE overhead in quantities that we can relate to the rest of this work, we assume a local stochastic noise channel with uniform error rate that is locally invertible [49–51]. (This is the case for local Pauli channels with constant and qubit-independent error rate, e.g. local depolarizing noise.) Under this assumption, Eq. (III.48) can be written

$$\mathcal{U} = \frac{1}{F}\mathcal{B}_0 + \sum_{j>0}\alpha_j\mathcal{B}_j\,, \quad |\boldsymbol{\alpha}|_1 = \frac{1}{F^2} \tag{III.85}$$

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

It remains to optimize $\mathcal{T}$, in an equivalent manner to Thm. III.7. To achieve precision ϵ using a circuit of depth $\mathcal{T}$, we need to choose the number

of samples $\widetilde{M} = \sigma^2/\epsilon^2$. Using Theorem III.10, we can write

$$\mathcal{T}_{\text{tot}} = \mathcal{T}\widetilde{M} = \mathcal{T}\frac{\sigma^2}{\epsilon^2} \leq \frac{B|\boldsymbol{\alpha}|_1^2}{\epsilon^2\mathcal{T}}. \tag{III.86}$$

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

$$\mathcal{T}_{\text{tot}} \leq \frac{Be^{-4\gamma\mathcal{T}}}{\epsilon^2\mathcal{T}} \tag{III.87}$$

This is optimized at $\mathcal{T} = \frac{1}{4\gamma}$, yielding

$$\mathcal{T}_{\text{tot}}^{(\text{Pauli noise})} = 4Be\gamma\epsilon^{-2} \approx 137\gamma\epsilon^{-2}. \tag{III.88}$$

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

III.E.5. Improving the overhead by filtering

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

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

Given a guess ϕ_{guess} for the phase we are attempting to estimate and a phase distance Δ within which we expect to find the true phase, we define

the filtering region $[\phi_{\text{guess}} - \Delta, \phi_{\text{guess}} + \Delta]$ and the relative indicator function

$$R(x) = [1 \text{ if } |x - \phi_{\text{guess}}| < \Delta, \text{ else } 0]. \tag{III.89}$$

We start from a set of measurements $\{(j_i, y_i)\}_{i=1,\dots,\widetilde{M}}$ sampled from the same distribution $\widetilde{Q}(j, y)$ defined in the section above, obtained by importance-sampling j from the PEC decomposition Eq. (III.49). We then reject all samples that lie outside the filtering region, obtaining a subset of the data

$$\mathcal{S} = \Big\{(j_i, y_i) \in \{(j_i, y_i)\}_{i=1,\dots,\widetilde{M}} \Big| y_i \in [\phi_{\text{guess}} - \Delta, \phi_{\text{guess}} + \Delta]\Big\}. \tag{III.90}$$

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

$$\widetilde{\mathcal{L}}_R(\phi) = \frac{|\boldsymbol{\alpha}|_1}{|\mathcal{S}|} \sum_{(j_i, y_i) \in \mathcal{S}} \text{sgn}(\alpha_{j_i}) \log P_R(y_i|\phi). \tag{III.91}$$

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

Let us formalize the above:

Definition III.8 (Filtered EUMLE)
Given the same assumptions as Definition III.6, as well as an indicator function $R(x)$ (taking values in $\{0, 1\}$), the filtered explicitly unbiased maximum likelihood estimator (filtered EUMLE) of the phase ϕ is defined as $\widetilde{\phi} \sim \arg\max_\phi \widetilde{\mathcal{L}}_R(\phi)$ where $\widetilde{\mathcal{L}}_R(\phi)$ [Eq. (III.91)] is constructed by sampling $\widetilde{M}$ points $\{(j_i, y_i)|j_i \sim |\alpha_{j_i}|/|\boldsymbol{\alpha}|_1, y_i \sim Q_{j_i}(y_i)\}_{i=1,\dots,M}$ and constructing a subset $\mathcal{S}$ [Eq. (III.90)] by rejecting all points where $R(y_i) = 0$.

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

$$\lim_{\widetilde{M}\to\infty} \widetilde{\mathcal{L}}_R(\phi) = \mathcal{N}_R \, \mathbb{E}\big[\log P(x|\phi)R(x)/\mathcal{N}_R(\phi)\big|x \sim P(x)R(x)/\mathcal{N}_R\big] \tag{III.92}$$

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

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

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

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

III.F. Surface-code resource estimate and optimization

In this section, we aim to benchmark our circuit-division QPE method and get an idea of the runtime overhead implied by a limitation on the number of available physical qubits. To do this, we estimate the physical resources for running fixed-precision circuit-division sin-state QPE on qubitized block-encodings of Fermionic Hamiltonian. In particular, we consider the 2-dimensional Fermi-Hubbard Hamiltonian [94] and a selection of molecular Hamiltonians for in small to medium active spaces, using a block encoding based on the tensor hypercontraction factorization [43]. We assume free and perfect state preparation – including a cost for state preparation would increase the circuit division overhead as the state needs to be prepared for every run; imperfect state preparation would require filtering or a different analysis of data

processing. We also assume that the residual noise (after error correction) follows the global depolarizing model. This allows us to apply the results from App. III.C.2 and choose the optimal depth $\mathcal{T}$ and the number of samples M as a function of the error rate γ as detailed in Algorithm III.5. In practice, in the presence of a more realistic noise model, the techniques proposed in Sec. III.E would need to be applied, resulting in an additional overhead which depends strongly on the assumptions about the noise model.

In the following sections, we describe our procedure for estimating the logical qubit and Toffoli gate count of the QPE circuits for the target problems. In order to translate these to a physical resource estimate, we must make some assumptions about the error correction and magic state distillation model used. We consider a rotated surface code, with Clifford gates performed via lattice surgery [167] and CCZ2T magic state factories [169] used to produce CCZ states (as our algorithm only requires Toffoli gates as non-Clifford resources). We consider both using a single factory, to minimize the number of physical qubits (as in [94]) and using two factories to reduce the runtime. We assume 50% of routing overhead [169] on top of the physical qubits used for data encoding and magic state distillation. We assume a physical gate error rate of $p = 0.001$ and a surface code cycle time of 1 µs [167]. The code distances for data qubits encoding d and for the two levels of distillation d_0 and d_1 are then optimized (by grid search) to minimize the physical computation volume while keeping the total error probability per circuit below $1 - e^{-\gamma\mathcal{T}}$.

III.F.1. Cost estimates for the qubitized Fermi-Hubbard Hamiltonian

We consider quantum phase estimation of the qubitized walk operator for the 2-dimensional Fermi-Hubbard model. We consider a model with $L \times L$ sites, for a total of $2L^2$ spin-orbitals, and interaction strength $u/t = 4$ (with the hopping parameter t setting the energy dimensions). This choice of interaction strength is common in resource estimates literature, as it represents the most challenging regime to treat at scale with classical algorithms [95]. We compute the Toffoli cost of the controlled-walk operator $\mathcal{W}$ using Qualtran [203] and following the procedure of [94]. We assume the rotation angles in the PREPARE step are encoded using $\beta = 10$ bits and performed with the phase-gradient technique (we neglect the cheap one-time preparation of the phase-gradient resource state). We separately compute the Toffoli cost of sin-state preparation and quantum Fourier transform on $\lceil \log_2(\mathcal{T} + 1) \rceil$ qubits, also using Qualtran. We estimate the number of logical data qubits

analytically as

$$\overbrace{2L^2}^{\text{spin-orbitals}} + \overbrace{\lceil \log_2(\mathcal{T}+1) \rceil}^{\text{QPE control qubits}} + \overbrace{\lceil 12 + 3\log_2(2L^2) \rceil}^{\substack{\text{qubitization ancillas} \\ \text{Eq. (64) of [94]}}} + \overbrace{\beta + 1}^{\substack{\text{controlled} \\ \text{phase gradient}}} \tag{III.93}$$

The walk operator has the same spectrum as $\pm \arccos H/\lambda$, with $\lambda/t = 8L^2$ the qubitization 1-norm. We fix the goal of estimating the energy to a precision $\Delta E/t = 0.01$, which can be achieved by requiring a precision on the eigenphase of $\mathcal{W}$ of $\epsilon = \Delta E/\lambda = 0.01/8L^2$ [94].

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

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

We observe that the curves representing the computational space-time tradeoff (in the right panel of Fig. III.9) have an "elbow-like" shape. We expect our method will lead to the most significant advantage around the elbow of the curve, as further increasing the number of available qubits affects the total runtime only marginally, while further reducing the computational time requires unreasonable space overhead. This elbow point represents the regime in which is most advantageous to combine error correction and error mitigation, similar to the enhancement regime discussed in section III.A.1 and represented in the bottom panel Fig. III.4. We find that, across system sizes, the elbow point corresponds to the algorithm prescribing around $M \approx 15$ to 40 samples, i.e. to the middle regime in Alg. III.5, and to the regime where

MSQPE slightly outperforms RPE (see Fig. III.2).

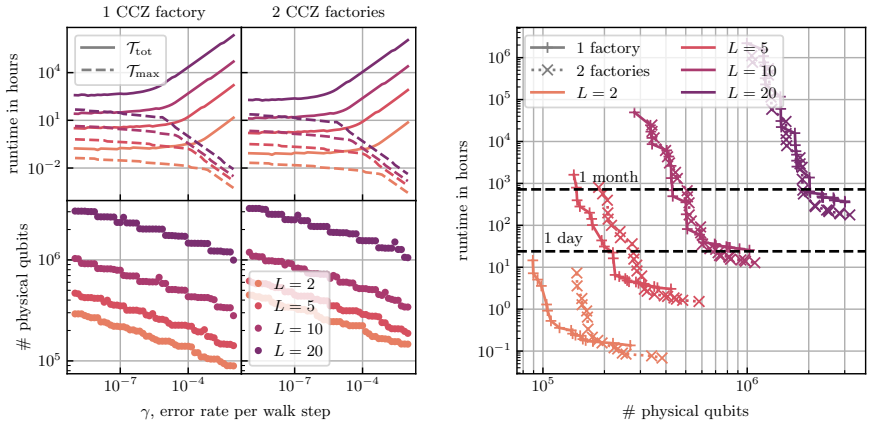


Figure III.9.: Physical costs for MSQPE applied to the qubitized Hubbard model Hamiltonian, with a target precision on the phase of $\epsilon = 10^{-2}/\lambda$ (where λ is the qubitization 1-norm). (Left) runtime and qubit costs as a function of the error rate per step of the qubitized walk γ. (Right)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

III.F.2. Cost estimates for Electronic Structure Calculations using THC

We compute the quantum resources for a number of molecular systems of small to medium size to illustrate how the higher complexity of molecular Hamiltonians and the more stringent precision requirements affect the required

quantum resources and run-times. All quantum chemistry calculations are performed with the help of PySCF [242]. For H_2O, N_2, naphthalene, and anthracene the active spaces were constructed either from restricted Hartree-Fock (RHF) orbitals or with AVAS [243] as implemented in PySCF. The CAS(27, 26) of the Co(salophen) complex was determined with the help of the Active Space Finder (ASF) [244] software developed by HQS Quantum Simulations and Covestro (see Table III.2 for details).

The resulting active spaces Hamiltonians were first symmetry-shifted as proposed in Ref. [202]. This means finding shifts a_1 and a_2 and modifying the molecular Hamiltonian H by adding powers of the total particle number operator N_e according to

$$H' = H + a_1 N_e + a_2 N_e^2. \tag{III.94}$$

This leaves the eigenvectors and the spectrum of the Hamiltonian inside the particle number sectors intact up to a particle number dependent offset that can be easily subtracted classically but reduces the norm of the coefficients of the Hamiltonian. Concretely, one picks a_1 as the median of the eigenvalues of the one-body part of the Hamiltonian and a_2 as the median of a certain diagonal of the two-body tensor [202].

The shifted Hamiltonian H' of each molecule was then factorized with the tensor hypercontraction (THC) procedure with L2 regularization with the penalty parameter ρ_{THC} as implemented in OpenFermion [245], leading to a λ factor denoted by λ'. Symmetry shifting leaves the coupled cluster singles doubles (CCSD) energy unaffected, but does cause very small changes in the triples correction contained in CCSD(T). It was thus more convenient to chose the THC rank based on the constraint that the CCSD energy error, ΔE_{CCSD}, stayed below 1 mEh. Consistent with what was reported in [209], we also observed roughly a factor of two reduction of the lambda value due to symmetry shifting.

Table III.2.: Orbital selection method, selected active spaces, and THC factorization parameters and results for the molecular systems used in the resource estimates. The AVAS [243] and ASF [244] methods were used for some systems.

Molecule	Active Space	Selection Method	THC Rank	ρ_{THC}	λ'	ΔE_{CCSD} [mEh]
N_2	CAS(6, 6)	around HOMO-LUMO	30	10^{-4}	4.66	0.62
H_2O	CAS(8, 6)	AVAS (H $1s$, O $2s2p$)	30	10^{-4}	6.97	0.63
naphthalene	CAS(10, 10)	AVAS (C $2p_z$)	45	10^{-5}	6.45	0.67
anthracene	CAS(14, 14)	AVAS (C $2p_z$)	80	10^{-5}	12.12	0.19
Co(salophen)	CAS(27, 26)	ASF default settings	100	10^{-5}	34.58	0.44

The logical resources for obtaining QPE estimates with $\epsilon = 10^{-3}/\lambda'$ preci-

sion (see Fig. III.10) were then computed with Qualtran [203], from the THC tensors. The logical circuit for the controlled walk operator is constructed using the algorithmic primitives included in Qualtran, and from that both the number of logical qubits and Toffoli gates are computed (assuming that rotations in PREPARE and SELECT are implemented to 10-bit and 16-bit precision, respectively [43]). The logical cost for the controlled walk operator is converted into physical costs for Algorithm III.5 as a function of the residual noise rate γ, under the same assumptions made for the Hubbard model.

III.F.3. Comparison to the fully fault-tolerant algorithm

For the sake of comparison in Fig. III.3 we also estimated the resources required for running the standard fault-tolerant single-circuit sin-state QPE algorithm. The standard fault tolerant algorithm is defined by choosing an acceptable failure probability δ as a free parameter. We fix the logical circuit to depth $\mathcal{T} = \mathcal{T}_{\text{tot}} = \lambda\pi/\Delta E$ to ensure a phase estimate to the desired precision in the absence of errors. The surface code and distillation distances are then chosen to optimize computational volume while keeping the total error probability below δ, under the same assumptions as above. The resulting requirements are reported in Fig. III.3 for $\delta = 1\%$ (▲) and $\delta = 0.1\%$ (▼), together with the requirements for the circuit-division algorithm.

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

Note that, in most cases, the cost of phase estimation circuits on a qubitized walk operator can be improved by an additional factor of ≈ 2 by using the unary iteration circuit to control the reflect operator as described in [43]. This improvement is not considered in this work, as it would require a subtle modification of the parameter selection procedure for MSQPE. Nevertheless, we expect this would yield an improvement by about a factor 2 in total runtime.

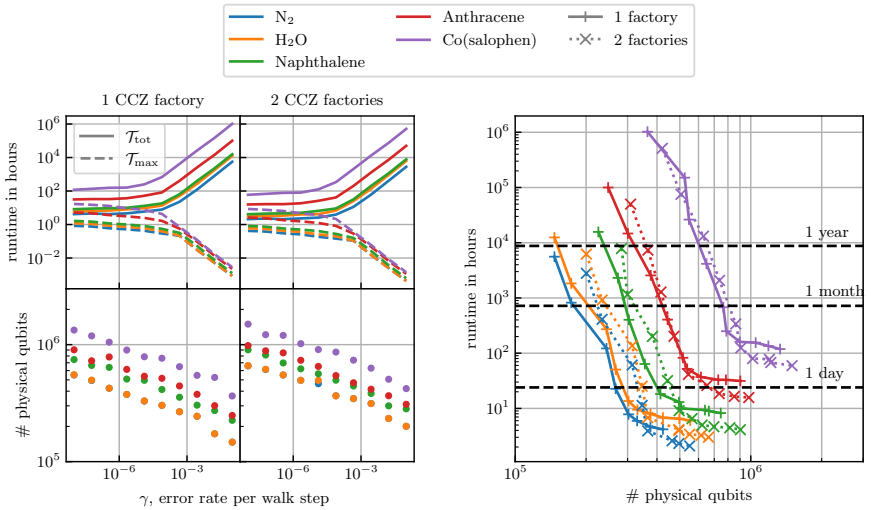


Figure III.10.: Physical costs for MSQPE applied to the electronic structure Hamiltonian factorized with THC, with a target precision on the phase of $\epsilon_t = 10^{-3}/\lambda'$ (where λ' is the symmetry-shifted THC lambda value). (Left) runtime and qubit costs as a function of the error rate γ per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

IV. Accurate ground state energy estimation with noise and imperfect state preparation

IV.1. Introduction

Quantum hardware is quickly approaching the era of early fault tolerance [1–4], where error correction is possible, but not perfect. Experiments must continue to be designed to tolerate imperfections, and classical post-processing must be exploited to make efficient use of limited gate and qubit counts. This also requires continued development of reliable error mitigation techniques to overcome residual bias and deliver accurate results. Initial experimental demonstrations of quantum error correction [1, 2, 4] and logical gates [3] have already made progress into early fault-tolerance, alongside early theoretical works outlining what is possible within this era [54, 56–61, 64, 67, 71, 72, 78, 92, 99, 170–172].

A natural target for early-fault-tolerant quantum computing is quantum phase estimation (QPE). This is a foundational computational task that underpins many applications in quantum simulation [43, 46, 88, 90, 94, 246], and more broadly in quantum information processing [11, 133]. QPE targets estimating the eigenvalue $e^{i\phi_0}$ of a unitary U[1] given access to a circuit implementing U (or a controlled version thereof) and an initial state $|\psi\rangle$ that overlaps with the eigenstate $|\phi_0\rangle$. To perform phase estimation, one can use the Hadamard test [55, 57, 61, 75] or other methods [225, 247] to estimate $\langle U^k \rangle$, and process this classically at multiple points k to infer the spectrum of U. Alternatively, one can coherently accumulate phase on a multi-qubit quantum register, and perform the quantum Fourier transform, which samples from a distribution that is peaked around the eigenphases of U [29, 88, 248]. It is possible to interpolate between these two methods [76, 132, 134], which

[1] In practice, most applications target a specific eigenvalue E_0 of a Hamiltonian H and implement unitaries e.g. $U = e^{-iHt}$ (Trotter-based) or $U = e^{i \arccos(H/\lambda)}$ (Qubitization-based) with $t^{-1}, \lambda \geq \|H\|$.

becomes relevant in the early-fault-tolerant setting where arbitrarily long circuit depths cannot be afforded. A third method uses quantum signal processing (QSP) circuits to implement block-encodings of some function of the Hamiltonian $f(H - \mathbb{1}x)$, which allows sampling from a distribution similarly peaked around eigenphases of U [56, 60, 112, 135, 145].

The classical post-processing of any of the above methods is a crucial piece of an early-FT QPE algorithm. One must compensate here for both the presence of experimental noise and imperfect state preparation ($a_0 := |\langle\phi_0|\psi\rangle|^2 \ll 1$), which can otherwise bias the estimation of ϕ_0. Significant recent work has gone into optimizing Hadamard-test-based QPE in the presence of imperfect state preparation, using matrix pencil [75, 136], cumulative distribution function [57, 59], and maximum likelihood methods [61, 62, 73]. These methods have been shown in some cases to be robust to small amounts of noise [74, 99, 190], and can be error mitigated using standard techniques [50] due to their intermediate estimation of expectation values $\langle\psi|e^{iHt}|\psi\rangle$. The same is not true for QFT-based or QSP-based QPE algorithms, as these do not work with expectation values. In previous work [76], we demonstrated that QFT-based QPE algorithms could be adapted to handle global depolarizing and circuit-level noise, but under the assumption of access to a perfect eigenstate. (This used a maximum-likelihood framework that is immediately extensible to QSP-based QPE methods.) Separately, Ref. [134] constructed a bias-free estimator for QFT-based QPE with imperfect initial states, but in the absence of noise. This leaves a gap in the literature to combine both sources of imperfection.

IV.1.1. Summary of key results

In this work, we consider the classical task of learning an eigenphase ϕ_0 of a unitary from phase estimation data generated by a quantum computer. We consider quantum phase estimation methods that sample from a distribution $p(x)$ peaked around $x = \phi_0$. In the presence of multiple eigenstates and experimental noise, this distribution becomes distorted and difficult to model; we set ourselves the task of optimally inferring a target phase ϕ_0 given samples from a distribution p with this distortion present. (We are concerned in this work in the case where experimental noise is large enough that one cannot afford the circuit depths required to purify the state and estimate phases at the Heisenberg limit, and must instead run short QPE circuits and repeatedly sample from this distribution $p(x)$.) We focus on the situation where ϕ_0 is isolated within a region $\mathcal{D}$, which allows us to bound the contributions from spurious eigenvalues $\phi_{j\neq 0}$. The key contribution of this work is to optimize the classical post-processing of samples from the distribution p in this situation

Definition IV.1 (Def. IV.8, informal)
Fix a unitary U, and assume that there exists a known region $\mathcal{D}$ containing a single target eigenphase ϕ_0. Let $p(x)$ be the output distribution from a phase estimation routine for U on a quantum computer. The classical task of phase estimation is, given M samples from $p(x)$, to construct an estimator $\tilde{\phi}_0$ of ϕ_0. The performance of this estimator is measured by the bias

$$b = \mathbb{E}[\tilde{\phi}[x]|x \sim p(x)] - \phi_0 \tag{IV.1}$$

and variance

$$\epsilon^2 = \mathbb{E}[(\tilde{\phi}[x] - \phi_0 - b)^2|x \sim p(x)]. \tag{IV.2}$$

(In Sec. IV.3.1 and Sec. IV.3.2 we explain how samples from $p(x)$ can be efficiently generated by a quantum computer.)

The above definition shifts the problem of near-term phase estimation to the slightly more abstract problem of extracting features of data drawn from a complex distribution. In Sec. IV.4.1 we solve this task by filtering out only the data that lies in $\mathcal{D}$ (yielding a filtered distribution P), and fitting this data with a parametrized model $Q(x|\phi)$ that ignores all eigenphases other than ϕ_0. This fitting procedure is known (among other names) as *moment projection* [249–251], so we call the resulting estimator the "moment projection phase estimator". We propagate the error in the simplifying assumption $P(x \in \mathcal{D}) \approx Q(x|\phi_0)$ to an error in the estimation of our phase:

Lemma IV.1 (Lemma IV.4, informal)
Assume U is a unitary with single eigenphase ϕ_0 in a known interval $\mathcal{D}$. Let P be an output distribution from a phase estimation circuit of U confined to the interval $\mathcal{D}$, and let $Q(x|\phi)$ be a model distribution parametrized by $\phi \in \mathcal{D}_\phi \subset \mathbb{R}$. Given M samples $\{x_j\}$ from P, define the moment projection estimator as the choice of ϕ that maximizes the likelihood $l(\phi|\{x_j\}) = \sum_{j\,;\,x_j \in \mathcal{D}} Q(x_j|\phi)$. To lowest order in the model error $h(x) = Q(x|\phi_0) - P(x)$, in the $M \to \infty$ limit this has variance $\epsilon^2 = \mathcal{I}_0^{-1} M^{-1} + \mathcal{O}(\|h\|_1)$, and bias $b = \|h\|_1 \mathcal{I}_0^{-1} S + \mathcal{O}(\|h\|_1^2)$, where $S = \max_{x \in \mathcal{D}} [\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}$ is the maximum of the score, and $\mathcal{I}_0$ is the Fisher information of $Q(x|\phi)$ at $\phi = \phi_0$.

The moment projection estimator defined above is able to reliably estimate a single phase in the presence of spurious neighbours, but only in the absence of experimental noise. To consider a more realistic setting, we follow the unbiasing procedure of Ref. [76]. This samples from a quasiprobability distribution as in PEC, but performs likelihood estimation instead of expectation value estimation on the output. This requires regularizing the likelihood function, as the quasiprobability samples often have large negative (i.e. highly unlikely) values that only slowly average out. However, naive regularization of the

moment projection estimator introduces a large bias. We solve this by adding the regularization term to the distribution P itself (this is the same as adding a small amount of global depolarizing noise)

Theorem IV.1 (Theorem IV.4, informal)
Assume U is a unitary with single eigenphase ϕ_0 in a known interval $\mathcal{D}$, and fix a regularization constant $c \geq 0$. Let P be the output distribution from a phase estimation circuit of U confined to the interval $\mathcal{D}$, let $Q(x|\phi)$ be a model distribution parametrized by $\phi \in \mathcal{D}$, and let $Q_c(x_j|\phi) = Q(x|\phi) + c$ be the regularized (non-normalized) model distribution. Assume the ability to write down a quasiprobability distribution

$$\sum_a \alpha_a P_a = P, \tag{IV.3}$$

and sample from the distributions P_a using a noisy quantum device. Given M samples $\{x_j\}$ from the P_a distributed with probability $\alpha_a/\|\alpha\|_1$, define the Noise-unbiased moment projection Estimator (NME) of ϕ_0 as the value that maximises the quasi-likelihood

$$\ell(\phi|\{x_j, a_j\}) = \frac{\|\alpha\|_1}{M} \sum_{j=1}^{M} \operatorname{sgn}(\alpha_{a_j}) \log[Q_c(x_j|\phi)] + c \int_{\mathcal{D}} \mathrm{d}x \, \log[Q_c(x|\phi)]. \tag{IV.4}$$

To lowest order in the error $h(x) = P(x) - Q(x|\phi_0)$, the NME has bias $\|h\|_1 S_c \mathcal{I}_c^{-1} + \mathcal{O}(\|h\|_1^2)$ and variance $\|\alpha\|_1^2 \epsilon^2 = S_c^2 \mathcal{I}_c^{-2} + \mathcal{O}(\|h\|_1)$, where $S_c = \max_{x\in\mathcal{D}}[\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi_0}$ is the maximum of the score, and $\mathcal{I}_c$ is the Fisher information of $Q_c(x_j|\phi)$ at $\phi = \phi_0$.

The results thus far are completely general and hold for any type of distribution that could be generated when performing phase estimation on a quantum computer. To obtain analytic resource requirements, we explicitly calculate the above bounds for a model with a Gaussian model distribution $Q(x|\phi) \sim e^{-(x-\phi)^2/2\sigma^2}$, and a linear combination of Gaussians for the true distribution $P(x)$. (We show how this distribution can be generated using phase estimation in Lemma IV.2 following [141].) We then make connection to the wider phase estimation literature by replacing our promise interval with an initial guess of ϕ_0 and a promise of a gap Δ to other eigenvalues, and obtain the following result for the number of calls to the (controlled) unitary U.

Theorem IV.2 (Theorem IV.5)
Let U be a unitary with spectral gap Δ around a target state ϕ_0 ($\forall_{j>0}|\phi_j - \phi_0| > \Delta$). Assume oracle access to a controlled version of U, and an initial state

$|\psi\rangle$ such that $|\langle\phi_0|\psi\rangle| > \eta$. Further assume an initial estimate ϕ_{guess} of ϕ_0 such that $|\phi_0 - \phi_{guess}| < \Delta/3$. Then, one can produce an estimate $\tilde{\phi}$ of ϕ_0 with RMS error ϵ using $M = O(\eta^{-1}t^{-2}\epsilon^{-2})$ samples of a phase estimation circuit, where each circuit requires $t = \Omega(\Delta^{-1}\log^{-1/2}(\Delta\epsilon^{-1}\eta^{-1}))$ calls to the unitary U, and the total number of calls $T = O(\eta^{-1}t^{-1}\epsilon^{-2})$.

In the absence of any noise, Theorem IV.2 recovers the Heisenberg limit when $t \sim \epsilon^{-1}$. To go beyond the noiseless assumption, in Sec. IV.4.5 we consider moment projection in the case where a circuit is affected by global depolarizing noise. This corresponds to a uniform probability distribution, and the optimal mitigation strategy is to simply incorporate the noise into our model function $Q(x|\phi)$. This yields a similar theorem to the above, but with an additional noise cost.

Theorem IV.3 (Theorem IV.7, informal)
Let U be a unitary with spectral gap Δ around a target state ϕ_0 ($\forall_{j>0}|\phi_j - \phi_0| > \Delta$). Assume oracle access to a controlled version of U with global depolarizing noise $e^{-\gamma}$ per call, and an initial state $|\psi\rangle$ such that $|\langle\phi_0|\psi\rangle| = a_0$. Further assume an initial estimate ϕ_{guess} of ϕ_0 such that $|\phi_0 - \phi_{guess}| < \Delta/3$. Then, one can produce an estimate $\tilde{\phi}$ of ϕ_0 with RMS error ϵ using $M = O(a_0^{-1}t^{-2}\epsilon^{-2})$ samples of a noisy phase estimation circuit, where each circuit requires $t = \Omega(\Delta^{-1}(\gamma\Delta^{-1} + \log(a_0^{-1}\epsilon^{-1})))$ calls to the unitary U. The total number of calls to U to execute the algorithm is $T = O(\epsilon^{-2}t^{-1}e^{-2\gamma t}a_0^{-2})$. Minimizing T as a function of t at fixed γ yields a cost $T = \Theta(\gamma\epsilon^{-2}a_0^{-2})$.

In Sec. IV.5 we test our estimators numerically in the presence of local depolarizing noise, on a toy phase estimation problem of a 4-qubit Ising model with up to 10 ancilla qubits. Beyond confirming our analytic results, this provides a clear implementation for the phase estimation practitioner that wishes to use these estimators. We observe that the moment projection estimator assuming global depolarizing noise performs surprisingly well, and often outperforms explicit unbiasing especially at low values of M. This is because the difference in output distribution between local and global depolarizing noise is not so large, and fitting the functional form of the noise will always be preferable to cancelling it via a quasiprobability distribution. This suggests future improvements to the NME via better modelling of the noisy distribution may yet be achievable.

IV.2. Definitions

In this section we define the phase estimation problem that we will focus on solving in this work (Def. IV.8). This splits phase estimation as a whole into

quantum and classical subroutines [75], which in our case interface via the distribution that a the quantum computer provides samples from Def. IV.3. (We defer the discussion of how these samples are obtained to Section IV.3.) We modify this distribution in Def. IV.5 by adding noise, and in Def. IV.7 by filtering (via rejection) to an interval, and in Def. IV.4 give a specific Gaussian example (which we will use throughout this work).

Quantum phase estimation takes as input a unitary U and initial state $|\psi\rangle$. The output of QPE depends on the decomposition of this state $|\psi\rangle$ in the eigenbasis of U.

Definition IV.2 (Spectral distribution)
Let U be a unitary operator with eigenbasis $U|\phi_j\rangle = e^{i\phi_j}|\phi_j\rangle$, $\phi_j \in [0, 2\pi)$. The spectral distribution of a state $|\psi\rangle$ in the eigenbasis of U is the function

$$a(x)\Big(= a_{U,|\psi\rangle}(x)\Big) := \sum_j a_j \delta(x - \phi_j), \qquad a_j := |\langle\phi_j|\psi\rangle|^2. \tag{IV.5}$$

The normalization of the state $|\psi\rangle$ ensures that $a(x)$ is a normalized probability distribution:

$$\int_0^{2\pi} a(x)\mathrm{d}x = \sum_j |\langle\phi_j|\psi\rangle|^2 = 1. \tag{IV.6}$$

In this work, we consider variants of QPE which use a quantum computer to provide samples from a distribution approximating $a(x)$. Perfectly sampling from $a(x)$ is in general not possible; instead, one typically approximates the delta functions $\delta(x - \phi_j)$ in Eq. (IV.5) by convolving with a so-called kernel function $f(x) \approx \delta(x - \phi_j)$

Definition IV.3 (Kernel function)
A kernel function is a non-negative normalized function $f : [0, 2\pi) \to [0, \infty)$; $\int_{-\pi}^{\pi} f(x)\mathrm{d}x = 1$. Given such a function, the smoothed spectral distribution of a state $|\psi\rangle$ in the eigenbasis of U is

$$[f * a](x) := \sum_j a_j f(x - \phi_j), \tag{IV.7}$$

where $a(x)$ is the spectral distribution (Def. IV.2).

Various kernel functions have been explored in the quantum phase estimation literature, in particular the Fejer kernel [29], sine window [43, 76, 129], cosine taper [140], DPSS taper [252], the Kaiser window [142], and the Gaussian kernel [56, 134, 141]. In this work we focus on Gaussian kernels due to

their ease of manipulation, but our techniques can be readily adapted to any kernel function with exponentially-decaying tails and controllable width.

Definition IV.4 (Gaussian kernel function)
A Gaussian kernel function of width $\sigma > 0$ and precision $\delta \geq 0$ is a kernel function $f_\sigma(x)$ for which approximates a Gaussian in the interval $(-\pi, \pi)$ such that

$$\sup_{x \in (-\pi,\pi)} \left| f_\sigma(x) - \frac{e^{-x^2/2\sigma^2}}{\int_{-\pi}^{\pi} e^{-x^2/2\sigma^2} \mathrm{d}x} \right| < \delta, \tag{IV.8}$$

The Gaussian kernel function $f_\sigma(x)$ can be constructed by polynomial approximation, with only a logarithmic overhead in the approximation precision [56]. We will ignore the details of this approximation in this work, and assume $f_\sigma \propto e^{-x^2/2\sigma^2}$.

In the absence of noise, the quantum computer targets sampling from a smoothed spectral distribution $[f * a_{U,|\psi\rangle}]$. The circuits required for this distribution typically have a depth proportional to the inverse of width σ of the kernel f (which is a consequence of the no-fast-forward theorem [151]). This becomes more complicated in an early fault-tolerant or NISQ setting. Under a stochastic noise model (where noise is treated as a series of discrete events that either occur or do not), the noisy probability distribution can be rewritten as a convex combination of the noiseless distribution $a(x)$ and a distribution $u(x)$ of all cases where a noise event happened, weighted by the fidelity $F \in (0, 1)$:

Definition IV.5 (Noisy distribution)
Let U be a unitary, $|\psi\rangle$ be a state, and f be a kernel function (Def. IV.3). In the presence of stochastic noise, let F be the circuit fidelity (the probability of no noise event occuring), and let $u(x) : [0, 2\pi) \to [0, \infty)$ be the distribution sampled from the quantum computer in the event that at least one noise event occurs. The noisy distribution is then the function

$$p(x) = F \cdot [f * a](x) + (1 - F) \cdot u(x), \tag{IV.9}$$

where $[f * a](x)$ is the distribution of $|\psi\rangle$ in the eigenbasis of U with kernel function f in the absence of noise (Def. IV.3).

Typically the circuit fidelity F is exponentially small in the circuit depth T. As we expect $T \propto \sigma^{-1}$, one can assume a form $F = e^{-\gamma/\sigma}$ for some decoherence rate γ. Optimizing QPE in the presence of noise trades thus requires trading between deep circuits with small σ (making estimation easier), and shallow circuits with large F [76].

The classical subroutine of a QPE algorithm takes the samples output from the quantum subroutine, and processes them to recover information about

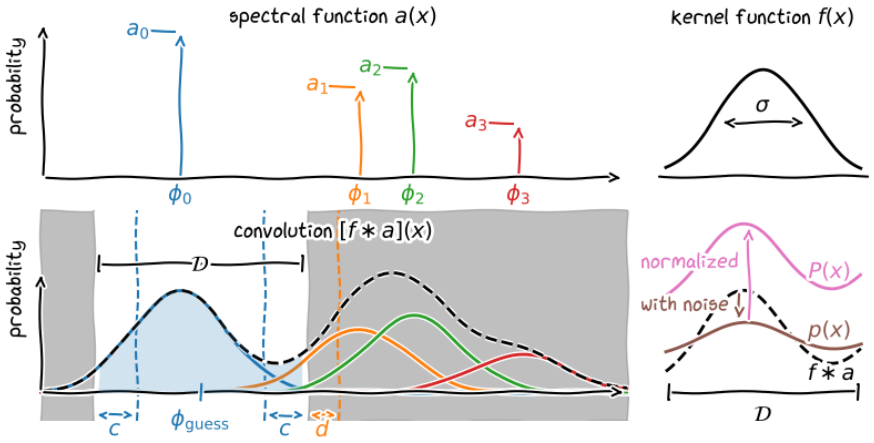


Figure IV.1.: Schematic plot illustrating the construction of the signal distribution of a filtered QPE experiment. **(Top left)** A unitary operator U and initial state $|\psi\rangle$ define the spectral distribution $a(x)$ in Def. IV.2 – a sum of Dirac deltas centered at the eigenphases ϕ_j of U with amplitudes a_j. **(Top right)** We define a kernel function $f(x)$ (Def. IV.3), with a shape and width σ depending on the details of the circuit. **(Bottom left)** The QPE circuits can sample from a distribution obtaining by convolution of a and f – the distribution in Def. IV.3, shown by the dashed black line. (The colored lines indicate the contributions of each eigenphase ϕ_j to the total distribution.) The filtering procedure discards all samples outside of a filtering region $\mathcal{D}$, which is chosen assuming that the phase of interest ϕ_0 is within the interval (farther than an *inner buffer* distance c from the interval edges) and all other phases are outside the interval (farther than an *outer buffer* distance d from the interval edges) – see Def. IV.6. The unnormalized distribution of filtered outcomes in the absence of noise is highlighted in blue. **(Bottom right)** Adding noise to the distribution $f * a$ yields the distribution $p(x)$ (Def. IV.5). Normalizing this distribution within the filtering interval $\mathcal{D}$ we obtain the filtered noisy distribution $P(x)$ (Def. IV.7).

the phases $\{\phi_j\}$. In this work we consider the estimation of a specific single phase, ϕ_0, which is identified by the promise of a region $\mathcal{D}$ in which it alone exists. This is in contrast to methods which attempt to estimate all phases in the problem simultaneously [62, 73, 75, 136], or to estimate a discretized form of the spectral distribution $a(x)$ itself [138], or to prepare the ground state itself [112, 135]. In practice, this promise is a reasonable assumption for e.g. the ground state energy of a gapped Hamiltonian, where $\mathcal{D}$ could be estimated via classical means or a lower-cost QPE method.

Definition IV.6 (Promise interval)
Let U be a unitary with eigendecomposition $U|\phi_j\rangle = e^{i\phi_j}|\phi_j\rangle$, $\phi_j \in [0, 2\pi)$. A promise interval $\mathcal{D}$ for an eigenphase ϕ_0 with inner buffer c and outer buffer d is a connected [2] subset of $[d, 2\pi - d)$ centred around ϕ_{guess} that satisfies the following two properties:

1. $\phi_0 \in [\phi_{\text{guess}} - \frac{|\mathcal{D}|}{2} + c, \phi_{\text{guess}} + \frac{|\mathcal{D}|}{2} - c]$
2. $\forall j \neq 0, \quad \phi_j \notin [\phi_{\text{guess}} - \frac{|\mathcal{D}|}{2} - d, \phi_{\text{guess}} + \frac{|\mathcal{D}|}{2} + d]$

In words, we require a buffer zone of width $c + d$ around the edges of $\mathcal{D}$, such that no phases lie within this buffer. The existence of a promise interval $\mathcal{D}$ with inner buffer c and outer buffer d implies a gap $\Delta \geq c + d$ between ϕ_0 and any other phase, and a promise interval can always be constructed given an ϵ-accurate estimate of ϕ_0 (i.e. $|\phi - \phi_{\text{guess}}| < \epsilon$) and the promise of a gap $\Delta > 2\epsilon + c + d$ (choosing a promise interval of size $|\mathcal{D}| = 2\epsilon + 2c$). For simplicity, in this work we will fix the inner buffer size $c = \frac{|\mathcal{D}|}{6}$ as a fraction of the interval width $|\mathcal{D}|$.

To use the promise interval $\mathcal{D}$ to optimize our estimation of ϕ_0, we will filter the noisy distribution $p(x)$ to lie within $\mathcal{D}$ only. This yields a new distribution, that can be sampled from by sampling from $p(x)$ and rejecting samples from outside $\mathcal{D}$:

Definition IV.7 (Filtered noisy distribution)
Let $p(x)$ be a noisy distribution (Def. IV.5) for unitary U, state $|\psi\rangle$, kernel function f and fidelity F, and let $\mathcal{D} \subset [0, 2\pi)$. The filtered noisy distribution $P(x)$ is the normalized distribution with support on $\mathcal{D}$ defined as

$$P(x) = \begin{cases} \frac{p(x)}{\int_{\mathcal{D}} p(x)\mathrm{d}x} & x \in \mathcal{D} \\ P(x) = 0 & x \notin \mathcal{D} \end{cases}. \tag{IV.10}$$

[2] To avoid unnecessary complications, we do not consider promise intervals that wrap around the circle. In the case where this would occur (i.e. if ϕ_0 is near 0 or 2π), one can trivially shift all phases by a constant to yield a connected promise interval $\mathcal{D}$.

Given oracular access to some filtered noisy distribution $P(x)$, we measure the performance of our classical estimation of ϕ_0 by bounding the bias and standard deviation of the constructed estimator $\tilde{\phi}_0$. This is a common metric used in the quantum metrology community [129, 144]. It differs slightly from the confidence interval formalism commonly used in computer science [29], however the two can be related to each other with at most a logarithmic overhead in the error probability through Chebyshev's inequality. We are now ready to state the phase estimation problem considered in this work.

Definition IV.8 (Classical and quantum subroutines of QPE)
Let P be a filtered noisy distribution (Def. IV.7) for given U, $|\psi\rangle$, f, F, $\mathcal{D}$; and assume $\mathcal{D}$ is a promise interval (Def. IV.6) for ϕ_0, with buffers c and d. The classical subroutine of the quantum phase estimation algorithm, given M samples from P, constructs an estimator $\tilde{\phi}$ for ϕ_0, with bias

$$b = \mathbb{E}[\tilde{\phi}[x]|x \sim P(x)] - \phi_0 \tag{IV.11}$$

and variance

$$\epsilon^2 = \mathbb{E}[(\tilde{\phi}[x] - \phi_0 - b)^2|x \sim P(x)] \tag{IV.12}$$

The quantum subroutine of a QPE algorithm is to generate samples from the distribution $P(x)$ given $\mathcal{D}$, f, F, a circuit implementation of U, and copies of $|\psi\rangle$.

In this work, for the sake of recovering simpler constant factors and making the proofs clearer, we restrict to a version of the above problem, but with the distribution $p(x)$ fixed to be a Gaussian:

Definition IV.9 (Classical QPE subroutine with Gaussian kernels)
Def. IV.8 with $f = f_\sigma$ Gaussian kernel function with variance σ^2 and $\delta = 0$, and inner buffer $c = |\mathcal{D}|/6$.

The techniques used in our proofs easily extend to smaller inner buffers and any kernel function that vanishes exponentially in x/σ [i.e. $f_\sigma(x) \sim o\left(\frac{1}{exp(x/\sigma)}\right)$], but the resulting expressions for bias and variance will have different constant factors.

IV.3. Background

Due to its BQP-completeness [137] and use as a subroutine in various quantum algorithms [11, 88, 133], much prior work has focused on optimizing phase estimation in various settings. In this section, we describe the various methods for constructing the quantum subroutine for quantum phase estimation,

followed by a review of the state of the art in constructing the classical subroutine (following our division of QPE into two subroutines in Def. IV.8).

A large body of work exists on quantum phase estimation that has a different quantum-classical interface to the one considered in this work, namely single-control QPE and related methods [57, 74, 75, 246]. These methods still have a well-defined split into quantum and classical subroutines, but here the quantum computer provides estimates of expectation values $\langle\psi|e^{iHt}|\psi\rangle$ instead of samples from a smoothed spectral distribution $f * a(x)$ (Def. IV.3). This allows these methods to access standard error mitigation techniques for expectation values [50]. However, it was shown in Ref. [132] that single control methods converge slower in estimation in the absence of noise. Furthermore, the ability to filter noisy data (as studied in this work) allows phase estimation to tolerate higher levels of noise, analogous to the difference in fidelity cost between postselection and rescaling [76]. Thus, extending error mitigation techniques from single-control to QFT-based and QSP-based methods is clearly of relevance for early-fault-tolerant phase estimation.

IV.3.1. QFT-based phase estimation

Quantum phase estimation was first studied as a subroutine in Shor's factoring algorithm [11]. Here, the quantum algorithm uses application of U^k controlled on the kth basis state of a control register prepared in some initial state $\sum_{k=0}^{K-1} b_k|k\rangle$, to generate

$$(c_K - U)\sum_{k=0}^{K-1} b_k|k\rangle|\psi\rangle = \sum_{k=0}^{K-1}\sum_{j} b_k a_j e^{ik\phi_j}|k\rangle|\phi_j\rangle. \tag{IV.13}$$

Here, $(c_K - U) = \oplus_{k=0}^{K-1}(|k\rangle\langle k| \otimes U^k)$ is the unitary U controlled by the entire quantum register. The algorithm proceeds by performing the quantum Fourier transform on the control register, and reading out the result. The cost of executing a single shot of $c_K - U$ is proportional to the maximum number of calls K to the unitary; for Shor's algorithm this is logarithmic, however in quantum simulation this is bounded below by the no-fast-forward theorem [151] to be worst-case linear in K.

A large body of work in phase estimation has focused on the optimization of the control register state. Originally, the b_k values were chosen to be a uniform superposition across $K = 2^{n_a}$ qubits ($b_k = 2^{-n_a}$) [216], which was popularized as the 'textbook phase estimation' due to its appearance in Ref. [29]. Following the quantum Fourier transform, measurement in the computational basis sample bitstrings $\tilde{x}$ from a convolution of the spectral

function with a Fejer kernel as kernel function

$$p^{(QPEA)}(\tilde{x}) = \sum_j a_j \frac{2\pi}{K} f_K^{(\text{Fejer})}\left(\frac{2\pi}{K}\tilde{x} - \phi_j\right), \quad \tilde{x} \in \{0, 1, ..., K-1\}, \tag{IV.14}$$

$$\frac{2\pi}{K} f_K^{(\text{Fejer})}(x) = \frac{1}{K^2}\frac{1-\cos(Kx)}{1-\cos(x)}. \tag{IV.15}$$

Textbook phase estimation has the advantage of having simple state preparation, and yielding exact eigenvalues given the promise that $2^{n_a}\phi_j/(2\pi) \in \mathbb{N}$. However, the Fejer kernel has a suboptimal width σ (as a function of K); $\sigma \sim K^{-1/2}$, as $\epsilon^{-1} = \mathcal{O}(\sigma^{-1})$, this implies that a classical estimator constructed from this data cannot achieve the Heisenberg limit (variance $\epsilon^2 \sim K^{-2}$). Refs. [131] improved on this by careful choice of the control state amplitudes c_k, such that the resulting kernel functions achieved tighter widths σ. Namely, Ref. [94, 129] uses a sine kernel to achieve an optimal standard deviation, while Ref. [127, 142] uses a Kaiser window to achieve optimal confidence-probability bounds. Gaussian kernels are also considered in Refs. [134, 141], as they allow for easy analysis when multiple samples are involved. Circuit constructions for sine states and Kaiser window states are known [94, 142]. More generally, as the cost of phase estimation for non-fast-forwardable unitaries grows linearly in the control register Hilbert space size, constructing even arbitrary control initial states should not be a significant factor in the overall cost of phase estimation.

In their standard definition, QFT-based QPE algorithms sample discrete variables. However these can be easily be adapted to the continuous description of the distributions we gave in Defs. IV.3, IV.5 and IV.7 through the random-phase technique [253, 254]. This technique consists in classically sampling a phase ϕ_{ref} uniformly at random in the interval $[0, 2\pi)$ before every circuit run, and implementing the QPE circuit on the modified unitary $e^{i\phi_{\text{ref}}}U$ (this implies a very small additive overhead, logarithmic in the precision of the classical variable). The reference phase is then added the output of the quantum circuit, yielding a random variable $x = \frac{2\pi}{K}\tilde{x} - \phi_{\text{ref}}$ with continuous support in $[0, 2\pi)$. For instance, the resulting distribution for textbook QPE becomes

$$p^{(QPEA)}(x) = \sum_j a_j f_K^{(\text{Fejer})}(x - \phi_j), \qquad x \in [0, 2\pi), \tag{IV.16}$$

$$f_K^{(\text{Fejer})}(x) = \frac{1}{2\pi K}\frac{1-\cos(Kx)}{1-\cos(x)}. \tag{IV.17}$$

More generally, if the circuit samples bitstrings $\tilde{x} \in 0, 1, \ldots, K-1$ with probability $p(\tilde{x}) = \sum_j a_j \frac{2\pi}{K} f(\frac{2\pi}{K}\tilde{x} - \phi_j)$ for some kernel function f, then the random phase technique will modify this to $p(\tilde{x}|\phi_{\text{ref}}) = \sum_j a_j \frac{2\pi}{K} f(\frac{2\pi}{K}\tilde{x} - (\phi_j + \phi_{\text{ref}}))$ and yield samples $x \in [0, 2\pi)$ distributed as

$$\begin{aligned} p(x) &= \sum_{\tilde{x}=0}^{K-1} \int_0^{2\pi} \frac{\mathrm{d}\phi_{\text{ref}}}{2\pi} p(\tilde{x}|\phi_{\text{ref}})\ \delta\left(x - \left(\frac{2\pi}{K}\tilde{x} - \phi_{\text{ref}}\right)\right) \\ &= \sum_{\tilde{x}=0}^{K-1} \int_0^{2\pi} \frac{\mathrm{d}\phi_{\text{ref}}}{2\pi} \frac{2\pi}{K} \sum_j a_j f\left(\frac{2\pi}{K}\tilde{x} - (\phi_j + \phi_{\text{ref}})\right)\ \delta\left(x - \left(\frac{2\pi}{K}\tilde{x} - \phi_{\text{ref}}\right)\right) \\ &= \frac{1}{K} \sum_{\tilde{x}=0}^{K-1} \sum_j a_j f\left(x - \phi_j\right) = \sum_j a_j f\left(x - \phi_j\right) = (a * f)(x). \end{aligned} \tag{IV.18}$$

To sample from $p(x)$ with a Gaussian kernel function of width σ (Def. IV.4), we must prepare the QPE control register in a quantum state $\sum_{k=0}^{K-1} b_k |k\rangle$ whose computational basis amplitudes b_k approximate a Gaussian distribution. The random phase technique (Eq. (IV.18)) then smoothens out the discrete measurement grid, allowing us to sample from the (continuous) convolved distribution $f * a(x)$ (Def. IV.3). We can bound the Gaussian tails to precision δ by exploiting the Fourier duality of the discrete sampling errors analysed in [141]. This requires a register dimension (and thus maximum evolution time) scaling as $K = \mathcal{O}(\sigma^{-1}\sqrt{\log(\delta^{-1})})$. Neglecting the one-off cost of preparing the state on the $n = \log_2(K)$ control qubits (which is subdominant to the cost of applying the K controlled unitaries), we obtain the following result as a direct consequence of Ref. [141], Theorem16:

Lemma IV.2 (Gaussian kernel synthesis)
One can prepare the window state and sample from the continuous phase distribution $p(x)$ of a Gaussian kernel function f_σ (Def. IV.4) to precision $\delta > 0$ using a preparation circuit of depth $\mathcal{O}\left(\sigma^{-1}\sqrt{\log(\delta^{-1})}\right)$.

In the presence of noise this distribution will change as per Def. IV.5. In order to get to the filtered distribution of Def. IV.7, we neglect the samples that lie outside of the given promise interval $\mathcal{D}$. To obtain M samples from the filtered distribution P, we need to run the quantum subroutine $M' > M$ times, yielding an average sample overhead $\mathbb{E}[M']/M = 1/\int_{\mathcal{D}} p(x)$. For kernel functions with fast decaying tails, such as the Gaussian kernel f_σ (Def. IV.4), this overhead is approximately a_0^{-1}. We will discuss this more in detail in section IV.4.3.

In the absence of noise, and given an initial eigenstate $|\psi\rangle = |\phi_0\rangle$, the optimal strategy for phase estimation involves a single-shot readout of an

estimate of ϕ_0 from the control register [29]. This renders complicated classical post-processing unnecessary, as one cannot optimize further over a single estimate. As a mixed state can be purified to the ground state using a circuit of depth $a_0^{-1/2}\Delta^{-1}$ [112, 135](with Δ the gap to the first excited state), less focus has been traditionally given to the classical QPE subroutine. However, in the presence of noise, one cannot afford the depth of such circuits,. In Ref. [134], QFT-based phase estimation was studied in the absence of noise, using a mean estimator on a subset of lowest-energy outcomes. This yielded a bound on the cost of estimation of circuit depth $T = O(1/\Delta)$ and number of repetitions $M = O(1/\epsilon^2)$. In Ref. [76], we studied QFT-based phase estimation of an eigenstate in the presence of general noise and global depolarizing noise, finding that optimal phase estimation occurs at circuit fidelities $\sim 1/e \sim 30\%$. However, no works have yet studied the realistic phase estimation context, with non-eigenstate starting states and noise.

IV.3.2. QSP-based phase estimation

Recently a new class of algorithms to estimate eigenvalues emerged, which use a completely different quantum subroutine from the QFT-based or Hadamard-test-based QPE algorithms [60, 145, 255]. Given a target unitary U, these algorithms rely on variants of quantum signal processing [114, 159, 256] to construct block-encodings of polynomial functions $h(U)$:

$$W_{h,U} = \begin{bmatrix} h(U) & \cdot \\ \cdot & \cdot \end{bmatrix}, \quad \langle 0| W_{h,U} |0\rangle = h(U), \tag{IV.19}$$

with $|h(e^{i\phi})| < 1 \, \forall \phi \in [0, 2\pi)$. Applying $W_{h,U}$ on $|0\rangle |\psi\rangle$ and measuring the control qubit will yield 0 with probability

$$p^0_{h,U,\psi} = \langle\psi| \langle 0| W^\dagger_{h,U} |0\rangle\langle 0| W_{h,U} |0\rangle |\psi\rangle = \langle\psi| h^\dagger(U) h(U) |\psi\rangle , \tag{IV.20}$$

and 1 otherwise. Sampling from these binary-test circuits with an appropriately-chosen set of functions $\{h\}$ allows to extract information about the phases of U.

Following the approach of [56], we aim to reconstruct the smoothed spectral distribution $f * a(x)$ of Def. IV.3 with kernel f by choosing a set of h_x such that $h_x(e^{i\phi}) = \sqrt{f(x-\phi)}/\max_\phi \sqrt{f(\phi)}$, with x taking values in the interval $[0, 2\pi)$. The sample probability $p^0_{f(x-\phi),U,\psi}$ will be proportional to $f * a(x)$:

$$p^0_{h_x,U,\psi} = \frac{1}{\max_\phi f(\phi)} \sum_j a_j f(x - \phi_j). \tag{IV.21}$$

Given access to these binary samples with probabilities, we can obtain M samples from the smoothed spectral distribution $[f * a](x)$ through rejection sampling, with the following steps: (1) sample x at random in $[0, 2\pi)$, (2) run the binary-test circuit with $h = h_x$; if the outcome is 0 accept the sample x (with probability $p^0_{h_x,U,\psi}$) and (3) repeat from 1 until M samples are accepted. We call M' the total number of repetitions, i.e. the total number of quantum circuits ran in order to obtain M accepted samples. The sampling overhead is equal to the inverse of the expected acceptance probability

$$\frac{\mathbb{E}[M']}{M} = \left[\frac{1}{2\pi}\int_0^{2\pi} p^0_{h_x,U,\psi}\mathrm{d}x\right]^{-1} = 2\pi \max_\phi f(\phi). \tag{IV.22}$$

In this setting, we can naturally implement filtering (Def. IV.7) by changing step (1) of rejection sampling, choosing x uniformly at random in $\mathcal{D}$ rather than in $[0, 2\pi)$. The sampling overhead is then reduced to

$$\frac{\mathbb{E}[M']}{M} = \left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} p^0_{h_x,U,\psi}\mathrm{d}x\right]^{-1} = \frac{|\mathcal{D}|\ \max_\phi f(\phi)}{\int_{\mathcal{D}}[f * a](x)\mathrm{d}x}. \tag{IV.23}$$

In the case of a Gaussian kernel function f_σ (Def. IV.4), the maximal value $\max_\phi f(\phi) = (\int_{-\pi}^{\pi} e^{-x^2/2\sigma^2}\mathrm{d}x)^{-1}$ is proportional to σ^{-1}, while $\int_{\mathcal{D}}[f * a](x)\mathrm{d}x \xrightarrow{\sigma\to 0} a_0$ because only ϕ_0 is in the promise interval. The sampling overhead is $\propto a_0^{-1}\sigma^{-1}|\mathcal{D}|$, with an additional factor of $\sigma^{-1}|\mathcal{D}|$ comparing to the QFT-based method.

In a pre-print version of [56], the authors proposed an algorithm (Algorithm 2 in [255]) to estimate ϕ_0 using samples from a (filtered) Gaussian distribution (Def. IV.4). The authors generated data $\{x_j\}$ for this using the QSP-based circuits described in this section, and constructed a classical estimator by taking an average of the accepted samples; $\tilde{\phi} = \langle x_j\rangle_{x_j\in\mathcal{D}}$. This algorithm further uses an adaptive choice of the interval $\mathcal{D}$ and Gaussian width σ to achieve arbitrarily low bias. Throughout this work we will use the mean estimator suggested here without these adaptive updates as an estimator to compare our results to.

Relative to the QFT-based implementation, QSP-based phase estimation techniques carry an additional sampling overhead of $\sigma^{-1}|\mathcal{D}|$. In principle this could be reduced by an adaptive choice of $\mathcal{D}$, and QSP-based circuits have lower requirements for ancilla qubits (1 as opposed to $\mathcal{O}(\sigma^{-1})$). For simplicity we do not consider the overhead from the QSP scheme (nor do we consider adaptive updates of $\mathcal{D}$) further in this work. However, the estimators designed in this work can be applied immediately to samples generated by the QSP-based rejection-sampling scheme above.

IV.4. Results

In this work, we construct a classical estimator for QPE (the "filtered moment projection phase estimator") in two steps. The key idea here is that a sample from the smoothed spectral distribution that falls within the filtering interval is, with high likelihood, caused by the eigenvalue of interest ϕ_0. Thus, we can fit the samples from the filtered distribution $P(x)$ [Def. IV.7] with a simple model $Q(x|\phi)$ that consider a single eigenvalue ϕ. The small amount of samples due to eigenvalues other than ϕ_0 (i.e. spurious phases), will however produce a small amount of bias in the resulting estimator. We bound both the bias and the variance of the resulting estimator. We calculate these bounds in Sec. IV.4.1 for a generic moment projection estimator, under the assumption that $P(x)$ is close to $Q(x|\phi_0)$, but without assuming any specific form for $P(x)$ and $Q(x|\phi)$.

In practice, the effect of noise on the outcome distribution is far more complex than global depolarizing noise, and can't be modelled with an explicit functional form. Instead, we can mitigate the noise using explicit unbiasing, a method we developed in previous work [76]. In section IV.4.2 we show that the moment projection estimator also works with explicit unbiasing.

To calculate the scaling of these bounds with quantities like the cost of execution on a quantum device, we must first fix a family of parametrized distributions. We achieve this by ignoring all phases except for the target phase ϕ_0, after which a distribution naturally occurs from the chosen kernel function (Def. IV.3) that we sample data from. In Sec. IV.4.4 and Sec. IV.4.5 we focus on a Gaussian kernel (Def. IV.4) in the case of no noise and global depolarizing noise respectively. In Lemma IV.7 and Lemma IV.8, we estimate the first-order constant factor terms for both cases, and in Theorem IV.5 and Theorem IV.7), we propagate this to costs in a standard phase estimation model.

IV.4.1. Moment projection estimator

When a stochastic phenomenon producing samples $x \sim P(x)$ can be modelled exactly with a parametrized distribution – i.e. there exists $Q(x|\phi)$ which matches the true distribution $Q(x|\phi_0) = P(x)$ for some *true value* of the parameter $\phi = \phi_0$ – the asymptotically optimal estimator for ϕ_0 is obtained by likelihood maximisation. In quantum phase estimation, this is the case if we are promised the initial state is the eigenstate $|\psi\rangle = |\phi_0\rangle$, thus the spectral distribution $a(x) = \delta(x - \phi_0)$. Under the knowledge of the noise distribution $u(x)$ we can then fully model the distribution $p(x)$ (Def. IV.5) by

$$q(x|\phi) = Ff(x - \phi) + (1 - F)u(x) \tag{IV.24}$$

We explored this setting in a previous work [76], further relaxing the assumption that $u(x)$ is known and defining an explicitly-unbiased maximum-likelihood estimator based on probabilistic error cancellation circuits.

To extend these techniques to the case of a more complicated spectral distribution, we propose to give up exactly modelling $P(x)$. Instead, we fit the same single-phase model $Q(x|\phi)$ to $P(x)$ only within a promise interval $\mathcal{D}$ (Def. IV.6), where we know that $f(x-\phi_0)$ is the main contributor to $P(x)$. The estimator we choose is the maximiser of the likelihood of the model $Q(x|\phi)$. However, as we do not expect $Q(x|\phi_0) = P(x)$ exactly this is not a canonical *maximum-likelihood estimator*. Instead, this estimator is known in information geometry [249] and machine learning [250] as *moment projection*, *M-projection* or *reverse-KL minimization* [251]. These names derive from the observation that $Q(x|\phi_0)$ is an orthogonal projection $P(x)$ onto the manifold defined by $Q(x|\phi)$, in a geometry defined by the reverse Kullback-Leibler (KL) divergence $D_{\mathrm{KL}}(P(x)\|Q(x|\phi)) = \int P(x)\log\frac{P(x)}{Q(x|\phi)}\mathrm{d}x$. We use the term "moment projection" going forward

Definition IV.10 (Moment projection estimator)
Let $Q(x|\phi)$ be a model distribution parametrized by $\phi \in \mathcal{D}_\phi$, and let $\{x_j\}_{j=1,\dots,M}$ be M independent samples distributed according to $P(x)$. The moment projection estimator is defined as

$$\tilde{\phi} = \arg\max_{\phi\in\mathcal{D}_\phi} \ell(\phi|\{x_j\}) \tag{IV.25}$$

$$\ell(\phi|\{x_j\}) = \frac{1}{M}\sum_j \log Q(x_j|\phi) \tag{IV.26}$$

We want to apply this estimator to the case where $P = Q(x|\phi_0) + h(x)$, with $|h(x)|$ sufficiently small. The discrepancy between $Q(x|\phi_0)$ and $P(x)$ will result in a bias in the estimator $\tilde{\phi}$ of ϕ_0, i.e. $\phi^* := \lim_{M\to\infty}\mathbb{E}[\tilde{\phi}] \neq \phi_0$. We want to study this in the $M\to\infty$ limit; let us first obtain functional forms for ϕ^* and the variance ϵ^2/M of the M-projection estimator for an arbitrary family of distributions $Q(x|\phi)$.

Lemma IV.3 (Asymptotic distribution of the moment projection estimator)
Let $Q(x|\phi)$ be a model distribution on $\mathcal{D}$ parametrized by $\phi \in \mathcal{D}_\phi \subset \mathbb{R}$. Let $\tilde{\phi}$ be the moment projection estimator defined in Def. IV.10, using M independent samples $x \in \mathcal{D}$ drawn according to P. Assume that $\log Q(x|\phi)$ is twice continuously differentiable in ϕ, that the expectations of its first and second derivatives exist under P, and that the minimizer of $D_{\mathrm{KL}}(P(x)||Q(x|\phi))$ is unique and lies in the interior of $\mathcal{D}_\phi$. Then, in the limit $M\to\infty$, the estimator is asymptotically Gaussian: the random variable $\sqrt{M}(\tilde{\phi}-\phi^*)$ converges in

distribution to $\mathcal{N}(0, \epsilon^2)$*, where*

$$\phi^* = \arg\min_{\phi} D_{\mathrm{KL}}(P(x)||Q(x|\phi)) = \arg\max_{\phi} \int_{\mathcal{D}} \mathrm{d}x\, P(x)[\log Q(x|\phi)], \tag{IV.27}$$

$$\epsilon^2 = \frac{\int_{\mathcal{D}} \mathrm{d}x\, P(x)[(\partial_\phi \log Q(x|\phi^*))^2]}{\left(\int_{\mathcal{D}} \mathrm{d}x\, P(x)[\partial_\phi^2 \log Q(x|\phi^*)]\right)^2}. \tag{IV.28}$$

This result is a relatively standard application of the central limit theorem; we prove this for completeness in App. IV.A.

The main result of this section is a bound on the closeness of the mean ϕ^* and variance σ of the moment projection estimator in the case where $P(x) = Q(x|\phi_0) + h(x)$ for some "target" ϕ_0, with $\|h\|_1 = \int_{\mathcal{D}} |h(x)|\mathrm{d}x$ sufficiently small. In the phase estimation case, ϕ_0 is the underlying phase we are trying to estimate, and the bias $\phi^* - \phi_0 \neq 0$ emerges from our incomplete modelling of the target distribution $P(x) \neq Q(x|\phi_0)$ [Def. IV.7]. The resulting bias can be bounded proportionally to the norm of the model error $\|h\|$, whilst the variance can be linked back to the Fisher information of $Q(x|\phi_0)$.

Lemma IV.4 (Moment projection estimator for distributions close to the model)
Under the same assumptions as in Lemma IV.3, let ϕ_0 *be a hidden target parameter such that*

$$P(x) = Q(x|\phi_0) + h(x), \tag{IV.29}$$

where the deviation h *is small in the* 1*-norm*

$$\|h\|_1 = \int_{\mathcal{D}} |h(x)|\mathrm{d}x. \tag{IV.30}$$

Assume furthermore that $\partial_\phi^2 D_{\mathrm{KL}}(P(x)||Q(x|\phi)) > 0$ *for* $\phi \in [\phi^*, \phi_0]$*. Then the asymptotic bias* b *and variance* ϵ^2/M *in Lemma IV.3 satisfy*

$$b = \frac{\int_{\mathcal{D}} \mathrm{d}x\, h(x)[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}}{\mathcal{I}_0} + O(\|h\|_1^2) \tag{IV.31}$$

$$\leq \|h\|_1 \frac{\max_{x\in\mathcal{D}}[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}}{\mathcal{I}_0} + O(\|h\|_1^2), \tag{IV.32}$$

$$\epsilon^2 = \frac{1}{\mathcal{I}_0} + O(\|h\|_1), \tag{IV.33}$$

where

$$\mathcal{I}_0 = \int_{\mathcal{D}} \mathrm{d}x\, Q(x|\phi_0)\, [\partial_\phi \log Q(x|\phi)]^2_{\phi=\phi_0} \tag{IV.34}$$

is the Fisher information of $Q(x|\phi_0)$.

We prove this lemma in appendix IV.A.
In practice, the assumptions of this theorem are satisfied if $Q(x|\phi)$ is a reasonable model for $P(x)$. This in turn requires that:

1. $Q(x|\phi)$ is close to P in a single region around ϕ_0, and for values of ϕ far from ϕ_0 the distributions are very different [this ensures a well-defined global minimum of $D_{\mathrm{KL}}(P(x)||Q(x|\phi))$]

2. $Q(x|\phi)$ is smooth [this ensures that $D_{\mathrm{KL}}(P(x)||Q(x|\phi))$ is convex around in a finite region around the optimum ϕ^*]

3. $\|h\|$ is small enough [this ensures ϕ_0 and ϕ^* are close enough, and both contained in the convex region $\mathcal{D}_\phi$].

IV.4.2. Explicit unbiasing for moment projection

The moment projection estimator requires a model $Q(x|\phi)$ of the output distribution; in the case of general noise, this model is not known. One possibility is to approximate the effect of general noise with a simplified model, e.g. one that assumes global depolarizing noise. This will generally result in an estimation error (bias) due to the incorrect modelling of the noise. Though this error may not be terribly large, we desire a method that can provably remove the bias from noise in the asymptotic resource limit. For expectation value estimation, this is achievable via probabilistic error cancellation (PEC) [49, 51], which expands the target expectation value as a linear combination of expectation values that can be estimated on a noisy device. In Ref. [76], we extended the PEC approach to QFT-based phase estimation, by constructing an explicitly-unbiased maximum likelihood estimator (EUMLE). The EUMLE writes the output distribution of the inaccesible (noiseless) QPE circuit as a quasi-probabilistic sum $P(x) = \sum_a \alpha_a P_a(x)$, where the P_a are the output distributions of accessible (noisy) circuits. From this, one can derive a likelihood function to optimize over, giving an estimate of the phase that is asymptotically bias-free whenever the noisy decomposition is correct. In this section, we extend this result to a result that uses the full moment projection estimator.

Care needs to be taken when implementing the explicitly unbiasing procedure, as sampling from the distributions P_a can yield data that has near zero probability to be sampled from $Q(x|\phi_0)$. This causes any estimation of ϕ_0 to

be dominated by the cost of obtaining sufficient samples to cancel this effect out; the variance of the resulting estimator becomes decoupled from the width σ of the kernel function. This was circumvented in Ref. [76] by regularization; one adds a small spurious constant to $Q(x|\phi_0) \to Q_c(x|\phi_0) = Q(x|\phi_0) + c$. This works when our range of estimation covers the full circle, however here the smaller interval $\mathcal{D}$ reintroduces a bias. To solve this problem, we could consider adding a small amount of global depolarizing noise to the data itself $P \to P_c(x) = P + c$, so that this regularization term properly models the data [3]. However, this noisy distribution comes with a large variance term. To circumvent this, instead of randomly sampling the noise, we can add the expected contribution to the likelihood itself.

Definition IV.11 (Noise-unbiased M-Projection Estimator (NME))
Let $\alpha \in \mathbb{R}^d$ be a vector of real coefficients and let $\{P_a(x)\}_{a=1}^d$ be probability distributions on $\mathcal{D} \subset \mathbb{R}$. Let $Q(x|\phi)$ be a model distribution parametrized by $\phi \in \mathcal{D}_\phi$. Fix a regularization constant $c \geq 0$, and define $Q_c(x_j|\phi) = Q(x|\phi) + c$ Let $\{(x_j, a_j)\}_{j=1,...,M}$ be M independent samples generated by first sampling $a_j \in \{1, ..., d\}$ with probability $P(a_j) = |\alpha_{a_j}|/\|\alpha\|_1$, and then sampling x_j from $P_{a_j}(x)$. The noise-unbiased moment projection estimator estimates

$$\tilde{\phi} = \arg\max_{\phi \in \mathcal{D}_\phi} \ell(\phi|\{x_j, a_j\}), \tag{IV.35}$$

$$\ell(\phi|\{x_j, a_j\}) = \frac{\|\alpha\|_1}{M} \sum_{j=1}^{M} \operatorname{sgn}(\alpha_{a_j}) \log[Q_c(x|\phi)(x_j)] + c \int_{\mathcal{D}} \mathrm{d}x \log[Q_c(x|\phi)]. \tag{IV.36}$$

As one might expect, the NME is not significantly different from the EUMLE of Ref. [76], and one can derive similar results to that work.

Lemma IV.5 (Asymptotic distribution of NME)
Let $Q(x|\phi)$ be a model distribution on $\mathcal{D}$ parametrized by $\phi \in \mathcal{D}_\phi \subset \mathbb{R}$ and $\{P_a(x)\}_{a=1}^d$ be sampleable distributions on $\mathcal{D}$. Fix a regularization constant $c > 0$, and define the regularized distributions $Q_c(x|\phi) = Q(x|\phi) + c$, $P_c(x) = \sum_{a=1}^d \alpha_a P_a(x) + c$. Let $\tilde{\phi}$ be the estimator defined in Def. IV.11, using M independent samples generated as described in Def. IV.11. Assume that $\log Q_c(x|\phi)$ is twice continuously differentiable in ϕ, that the expectations of its derivatives exist under each distribution P_a, and that the minimizer of $D_{\mathrm{KL}}(P_c(x)||Q_c(x|\phi))$ is unique and lies in the interior of $\mathcal{D}_\phi$. Then, in the limit $M \to \infty$, the estimator is asymptotically Gaussian: $\sqrt{M}(\tilde{\phi} - \phi^*)$

[3] $Q_c(x_j|\phi)$ and $P_c(x)$ are not normalized; we add the effect of the regularization directly to the likelihood rather than sampling from either distribution, so normalizing these distributions is unnecessary.

converges in distribution to $\mathcal{N}(0, \epsilon^2)$*, where*

$$\phi^* = \arg\max_{\phi} \int_{\mathcal{D}} \mathrm{d}x\, P_c(x)[\log Q_c(x|\phi)], \tag{IV.37}$$

$$\epsilon^2 \leq \|\alpha\|_1^2 \frac{\int_{\mathcal{D}} \mathrm{d}x\, R(x)[\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi_*}}{\left(\int_{\mathcal{D}} \mathrm{d}x\, P_c(x)[\partial_\phi^2 \log Q_c(x|\phi)]_{\phi=\phi^*}\right)^2}. \tag{IV.38}$$

Here, $R(x)$ *is the marginal distribution* $R(x)(x) = \frac{1}{\|\alpha\|_1} \sum_{a=1}^{d} |\alpha_a| P_a(x)$*.*

Theorem IV.4 (NME for distributions close to the model)
Under the same assumptions as in Lemma IV.5, let ϕ_0 *be a hidden target parameter such that*

$$\sum_{a=1}^{d} \alpha_a P_a(x) = Q(x|\phi_0) + h(x), \tag{IV.39}$$

where the deviation h *is small in the* 1*-norm*

$$\|h\|_1 = \int_{\mathcal{D}} |h(x)| \mathrm{d}x. \tag{IV.40}$$

Assume further that $\partial_\phi^2 D_{\mathrm{KL}}(P(x)||Q(x|\phi)) > 0$ *for* $\phi \in [\phi^*, \phi_0]$*. Then the asymptotic bias* $b = |\phi^* - \phi_0|$ *and variance* ϵ^2/M *in Lemma IV.5 satisfy*

$$b \leq \|h\|_1 \frac{\max_{x\in\mathcal{D}}[\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi_0}}{\mathcal{I}_c} + O(\|h\|_1^2), \tag{IV.41}$$

$$\epsilon^2 \leq \|\alpha\|_1^2 \frac{\max_{x\in\mathcal{D}}[\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi_0}}{\mathcal{I}_c^2} + O(\|h\|_1), \tag{IV.42}$$

where

$$\mathcal{I}_c = \int_{\mathcal{D}} \mathrm{d}x\, Q_c(x|\phi_0)\, [\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi_0}\,, \tag{IV.43}$$

is the Fisher information of $Q_c(x|\phi_0)$ *[Eq.* (IV.34)*].*

We prove Lemma IV.5 and Theorem IV.4 in Appendix IV.A.
While the EUMLE in Ref. [76] was compatible with PEC alone, NME is compatible with a larger class of error mitigation methods. This is because EUMLE required the quasiprobability to exactly match the model, i.e. $\sum_{a=1}^{d} \alpha_a P_a(x) = Q(x|\phi_0)$, and therefore the quasiprobability distribution needed to reconstruct the noiseless probability $P_0(x)$ exactly. Instead, since NME allows for a small mismatch between the quasiprobability $\sum_{a=1}^{d} \alpha_a P_a(x)$

and the model $Q(x|\phi_0)$, it suffices to have an approximate linear decomposition $\sum_{a=1}^{d} \alpha_a P_a(x) \approx P_0(x)$. This can be achieved by any linear quantum error mitigation technique, including PEC but also methods such as zero-noise extrapolation or symmetry verification [257], which reconstruct mitigated quantities as linear combinations of measurements from noisy circuits. This can be treated as an additional contribution to h. The asymptotic bias in Eq. (IV.41) will then contain contributions from both the modelling error (the mismatch between the noiseless distribution P_0 and $Q(x|\phi_0)$) and the imperfect $\sum_{a=1}^{d} \alpha_a P_a$ reconstruction of P_0 produced by the QEM procedure. We can separate these contributions to the error via the triangle inequality,

$$\|h(x)\|_1 = \|\sum_{a=1}^{d} \alpha_a P_a(x) - Q(x|\phi_0)\| \tag{IV.44}$$

$$\leq \|\sum_{a=1}^{d} \alpha_a P_a(x) - P_0(x)\| + \|P_0(x) - Q(x|\phi_0)\|. \tag{IV.45}$$

IV.4.3. Moment projection for phase estimation

In the previous section we focused on how the error in fitting a distribution with an imperfect parametric model propagates to a bias in the estimate of the parameter. In this section, we apply these tools to QPE. First, we prove that the underlying sampling input (Def. IV.8) can be generated by QFT-based QPE techniques, and that we can construct an appropriate promise interval $\mathcal{D}$ under standard assumptions. Then, we define estimators for noiseless and noisy phase estimation, to which we apply the results of the previous section.

Lemma IV.6 (Generating filtered QPE samples)
Let U be a unitary with target phase ϕ_0. Assume an initial guess ϕ_{guess} of ϕ_0 accurate up to $\Delta/3$, for $\Delta \geq \min_{j\neq 0} |\phi_j - \phi_0|$. Then, given the ability to sample from $p(x)$ (Def. IV.5) for U, a state $|\psi\rangle$ with ground state overlap $a_0 > 0$, and a kernel function f using t calls to a circuit implementation of U, one can construct a promise interval $\mathcal{D}$ and generate M samples from $P(x)$ that satisfies the conditions of Def. IV.8 (i.e. one can execute the quantum subroutine), using on average Mt/P_A calls to a circuit implementation of U, where P_A is a lower bound on the probability of accepting a sample given by

$$P_A = Fa_0 \int_{\phi_{guess}-\Delta/2}^{\phi_{guess}+\Delta/2} f(x-\phi_0)\mathrm{d}x + (1-F)\frac{|\mathcal{D}|}{2\pi}$$

$$\geq Fa_0 \int_{-\Delta/3}^{\Delta/3} f(x)\mathrm{d}x + (1-F)\frac{|\mathcal{D}|}{2\pi} \tag{IV.46}$$

Proof. First, we will construct a promise interval that satisfies the properties in Def. IV.6. Let $\mathcal{D}$ be an interval of size $|\mathcal{D}| = \Delta$ centered around ϕ_{guess}, i.e. $\mathcal{D} = [\phi_{guess} - \Delta/2, \phi_{guess} + \Delta/2]$. By assumption $|\phi_0 - \phi_{guess}| < \Delta/3$, and the first condition in Def. IV.6 is satisfied for an inner buffer $c = \Delta/6$. Using triangle inequality, and the assumption that $\forall_{j>0}|\phi_j - \phi_0| > \Delta$, we have

$$|\phi_j - \phi_{guess}| \geq |\phi_j - \phi_0| - |\phi_0 - \phi_{guess}| \geq \frac{2}{3}\Delta \tag{IV.47}$$

Therefore the second condition in Def. IV.6 is satisfied with an outer buffer $d = |\phi_1 - \phi_{guess}| - |\mathcal{D}|/2 \geq \Delta/6)$.

To obtain M samples within the promise interval $\mathcal{D}$, we need to get $M' \geq M$ samples from $p(x)$. We can write probability of accepting each sample as

$$P_{accept} = F \int_{\mathcal{D}} \mathrm{d}x \quad \left[(a * f)(x) + (1 - F)\frac{1}{2\pi}\right] \tag{IV.48}$$

$$= F \left[\sum_j a_j \int_{\mathcal{D}} f(x - \phi_j)\mathrm{d}x\right] + (1 - F)\frac{|\mathcal{D}|}{2\pi}. \tag{IV.49}$$

Since f is positive, we can bound it from below by neglecting the contributions of $\phi_{j\neq 0}$ as

$$P_{accept} \geq F a_0 \int_{\mathcal{D}} f(x - \phi_0)\mathrm{d}x + (1 - F)\frac{|\mathcal{D}|}{2\pi}. \tag{IV.50}$$

Again using positivity of f, we can bound the integral in the expression above by an integral over a subset $[\phi_0 - \Delta/3, \phi_0 + \Delta/3] \subset \mathcal{D}$, yielding the desired bound. □

With Lemma IV.6 in hand, we construct estimators for phase estimation to which fit the assumptions of Lemma IV.4. As described in Sec. IV.3, these depend on a choice of kernel function f (Def. IV.3); we leave this free for now, but will consider Gaussian kernels in the rest of this section (Sections IV.4.4 and IV.4.5). In the noiseless case, the probability distribution $P(x)$ is a combination of contributions from different phases, which we approximate by a model $Q(x|\phi)$ which assumes a single phase. The mismatch $P(x) - Q(x|\phi) = h(x)$ then comes from the signal due to residual phases; within the promise interval this contribution is small, which we investigate in the next sections. Here, we introduce the models $Q(x|\phi)$ and related moment projection phase estimators, with different assumptions about noise. Starting from the noiseless case:

Definition IV.12
[filtered moment projection Phase Estimator (fMPE), noiseless case] Let f be

a kernel function (Def. IV.3), $\mathcal{D}$ be a promise interval (Def. IV.6), $p(x)$ be the distribution in Def. IV.5 with kernel function f, fidelity $F = 1$ and ground state overlap a_0, and $P(x)$ be the corresponding filtered distribution on $\mathcal{D}$ (Def. IV.7). In the noiseless case, the fMPE is moment projection estimator (Def. IV.10) with samples $x \sim P(x)$ and a model distribution

$$Q(x|\phi) = \frac{f(x-\phi)}{\int_{\mathcal{D}} f(x-\phi)\mathrm{d}x}. \tag{IV.51}$$

If instead the circuit is affected by global depolarising noise, we can model it exactly by adding a constant noise level of $(1-F)\frac{1}{2\pi}$ to the model for $p(x)$ and normalizing properly after filtering. This yields a new estimator, where again the mismatch h only comes from the spurious phases:

Definition IV.13
[fMPE, assuming global depolarising noise] Let f be a kernel function (Def. IV.3), $\mathcal{D}$ be a promise interval (Def. IV.6), $p(x)$ be a noisythe distribution in Def. IV.5 with kernel function f, fidelity F and ground state overlap a_0, and $P(x)$ be the corresponding filtered distribution on $\mathcal{D}$ (Def. IV.7). Assuming global depolarising noise, fMPE is moment projection estimator (Def. IV.10) with samples $x \sim P(x)$ and a model distribution

$$Q(x|\phi) = \frac{Fa_0 f(x-\phi) + (1-F)\frac{1}{2\pi}}{Fa_0 \int_{\mathcal{D}} f(x-\phi)\mathrm{d}x + (1-F)\frac{|\mathcal{D}|}{2\pi}}. \tag{IV.52}$$

In Figure IV.2 we give a schematic representation of this estimator.

The case of general noise cannot be simply modelled by adding a term in $Q(x|\phi)$, as modelling general noise would amount to simulating the full quantum circuit. Instead, we use the explicitly-unbiased moment projection estimator introduced above:

Definition IV.14
[filtered Noise-unbiased moment projection Phase Estimator (fNMPE)] Let f be a kernel function (Def. IV.3), $\mathcal{D}$ be a promise interval (Def. IV.6). Let the coefficients $\alpha \in \mathbb{R}^d$ and probability distributions on $\mathcal{D}$ $\{P_a(x)\}_{a=1}^d$ be such that $P(x) = \sum_{a=1}^d \alpha_a P_a(x)$ is the distribution in Def. IV.7 with kernel function f, fidelity $F = 1$ and ground state overlap a_0. The fNMPE is NME (Def. IV.11) with samples $(a, x) \sim \frac{|\alpha_a|}{\|\alpha\|_1} P_a(x)$ and the noiseless model distribution from Eq. (IV.51).

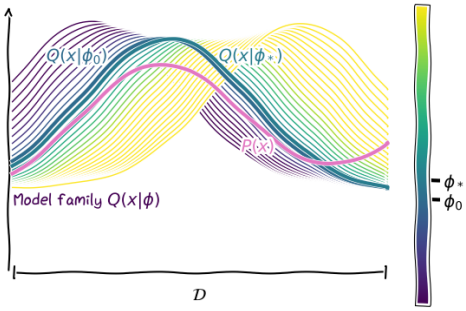


Figure IV.2.: Schematic representation of the fMPE (Def. IV.13): among a family of models $Q(x|\phi)$ [Eq. (IV.52), colored as per the colorbar based on the value of the parameter ϕ] supported on $\mathcal{D}$, the one that minimizes the inverse KL divergence with $P(x)$ (Def. IV.7) is $Q(x|\phi^*)$. The target distribution $P(x)$ is the final result of the filtered-QPE scheme represented in Fig. IV.1. We also highlight the model distribution for the ideal value of $\phi = \phi_0$: $Q(x|\phi_0)$ – the difference between ϕ_0 and the optimal parameter ϕ_* is the bias of the fMPE.

IV.4.4. Gaussian kernel and no noise

In order to apply Lemma IV.4 to a real phase estimation problem, we need to define a noise model and a kernel function. We first consider the noiseless setting, and choose a Gaussian kernel. This allows us to directly compare our moment projection estimator to the sample mean estimator of [255]. For simplicity of notation, we use the following shorthand for the Gaussian and its integral on the filtering interval $\mathcal{D}$

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \tag{IV.53}$$

$$G_\sigma(\phi) = \int_{\mathcal{D}} g_\sigma(x-\phi)\mathrm{d}x. \tag{IV.54}$$

Let us first give some intuition into how the M-projector differs from a simple mean estimator $\bar{x} = M^{-1}\sum_j x_j$. In the case where the domain of the filtering is $[-\infty, \infty]$, the sample mean estimator matches the moment projection estimator with $Q(x|\phi) = g_\sigma(x-\phi)$. Taking a finite filtering interval $\mathcal{D}$,

$$P(x) = \frac{\sum_j a_j g_\sigma(x-\phi_j)}{\sum_j a_j G_\sigma(x-\phi_j)}, \tag{IV.55}$$

modifies the model for the moment projection estimator to

$$Q(x|\phi) = \frac{g_\sigma(x-\phi)}{G_\sigma(\phi)} \text{ for } x \in \mathcal{D}. \tag{IV.56}$$

The objective function we have to maximize (log-likelihood ℓ of the model) is then

$$\ell(\phi|\{x_j\}) = -\frac{1}{M}\sum_j \frac{(x_j-\phi)^2}{2\sigma^2} - \log G_\sigma(\phi) \tag{IV.57}$$

Maximising this leads to a truncated normal distribution

$$\tilde{\phi} = \bar{x} + \sigma^2 \frac{g_\sigma(\tilde{\phi} - \phi_{guess} - \frac{|\mathcal{D}|}{2}) - g_\sigma(\tilde{\phi} - \phi_{guess} + \frac{|\mathcal{D}|}{2})}{G_\sigma(\phi)}. \tag{IV.58}$$

The second term is an additional correction due to the normalization term, which depends on ϕ but not on the samples. This clarifies explicitly the difference between the mean estimator and the moment projection estimator for a filtered Gaussian model.

If the probability distribution of samples matches the model $P(x) = Q(x|\phi_0)$ (i.e., for the case of QPE, if we do not have spurious phases), the moment projection estimator matches the maximum-likelihood estimator, which we know to be unbiased. This implies the bias of the mean estimator in this case is precisely the second term of Eq. (IV.58). This is exponentially small in σ^{-1}, as long as ϕ_0 is contained in the promise interval with inner buffer (Def. IV.6).

In the presence of spurious phases, the moment projection estimator also picks up a bias. We characterise the bias and variance of this estimator in the following lemma.

Lemma IV.7
Consider the fMPE (Def. IV.12) with a Gaussian kernel function $f_\sigma(x) \propto g_\sigma(x)$ (Def. IV.4 with $\delta = 0$) and ground state overlap a_0. Assume that $\mathcal{D}$ is a promise interval (Def. IV.6) with inner buffer $c = |\mathcal{D}|/6$ and outer buffer d, i.e. target phase ϕ_0 falls within the filtering interval $\mathcal{D} = [\phi_{guess} - |\mathcal{D}|/2, \phi_{guess} + |\mathcal{D}|/2]$,

$$|\phi_0 - \phi_{guess}| \leq |\mathcal{D}|/3, \tag{IV.59}$$

and further that all ϕ_j for $j \neq 0$ are sufficiently far from the filtering region

$$\min_{j\neq 0} \max_{x\in\mathcal{D}} |\phi_j - x| \geq d. \tag{IV.60}$$

Then, for any $\sigma \leq |\mathcal{D}|/6$, the asymptotic bias and variance of the estimator

in Lemma IV.3 satisfy

$$|\phi - \phi^*| \leq 4\sigma^2 g_\sigma(d) \times \frac{1-a_0}{a_0}\frac{|\mathcal{D}|}{d} + O\Big(\frac{|\mathcal{D}|^2\sigma^4}{d^2} g_\sigma(d)^2\Big) \tag{IV.61}$$

$$\lim_{M\to\infty} M \operatorname{Var}[\tilde{\phi}] \leq 2\sigma^2 + O(|\mathcal{D}| g_\sigma(d)). \tag{IV.62}$$

The proof of this theorem is given in appendix IV.B. Our choice of the buffer between the ground state eigenphase and interval edges [Eq. (IV.59)] is artificial; in principle this can be removed entirely without affecting the asymptotic scaling of our estimator with σ and a_0. The non-zero buffer between the spurious phases ϕ_j and the filtering interval $\mathcal{D}$ [Eq. (IV.60)] is necessary however to accommodate the fact that our model distribution $Q(x|\phi)$ explicitly does not consider any additional phases; when $d = 0$, the M-projection estimator remains biased for arbitrarily small σ.

In order to characterize the performance of the moment projection estimator beyond upper bounds and compare it to the mean estimator, we integrate numerically the first-order bias in Eq. (IV.31), ignoring the $O(|\mathcal{D}|^2 g_\sigma(d)^2)$ correction. We observe in Fig. IV.3 that the bias of the moment projection estimation is largely independent on the value of ϕ_0; the bias is only due to the contribution of the spurious phases and decreases as ϕ_1 is further from the interval $\mathcal{D}$, as predicted by the bound in Eq. (IV.61). In contrast, the mean estimator $\tilde{\phi} = \langle x_j \rangle_{x_j \in \mathcal{D}}$ picks up an additional bias that depends on the value of ϕ_0 [Eq. (IV.58)]. When this latter bias dominates, the moment projection estimator achieves a total bias that is exponentially smaller in σ^{-1} than the bias of the mean estimator. The improvement is especially evident for ϕ_1 farther from the filtering interval. The narrow region where the mean estimator has vanishing bias is due to a fortuitous cancellation of the positive bias coming from the spurious phases and the negative bias coming from the filtered-out samples from the ground phase.

At this point we have constructed an estimator that takes samples from a distribution $P(x)$ and fits a Gaussian model $Q(x|\phi)$ with a fixed distribution σ. This considers only the classical subroutine of phase estimation (Def. IV.8). To connect this to the standard phase estimation literature, we extend this to a quantum algorithm

Theorem IV.5 (cost of noiseless fMPE)
Let U be a unitary with spectral gap Δ around a target state ϕ_0 ($\forall_{j>0}|\phi_j - \phi_0| > \Delta$). Assume oracle access to a controlled version of U, and an initial state $|\psi\rangle$ such that $|\langle\phi_0|\psi\rangle| > \eta$. Further assume an initial estimate ϕ_{guess} of ϕ_0 such that $|\phi_0 - \phi_{guess}| < \Delta/3$. Then, the fMPE (Def. IV.12) using t calls to U per circuit produces an estimate $\tilde{\phi}$ with RMS error ϵ using $M = O(\eta^{-1}t^{-2}\epsilon^{-2})$ samples and $T = O(\eta^{-1}t^{-1}\epsilon^{-2})$ total calls to U, as long

as $t = \Omega(\Delta^{-1}\log^{-1/2}(\Delta\epsilon^{-1}\eta^{-1}))$.

Proof. By Lemma IV.2, we can generate samples from $p(x)$ with a Gaussian kernel function f_σ using $t = \Theta(\sigma^{-1})$ calls to U (we ignore the overhead of $\log(\delta)$). By Lemma IV.6, we can then generate samples from $P(x)$ with $\mathcal{D}$ that satisfies the assumptions of Lemma IV.7 with $|\mathcal{D}|, d = \Theta(\Delta)$. By Lemma IV.7, $t = \Omega(\Delta^{-1}\log^{-1/2}(\Delta\epsilon^{-1}\eta^{-1}))$ is enough to ensure that the bias is $\mathcal{O}(\epsilon)$. To ensure that the variance is $\mathcal{O}(\epsilon^2)$, we need $M = \Theta(t^{-2}\epsilon^{-2})$ samples. By Lemma IV.6, to generate M samples, we need $T = \mathcal{O}(\eta^{-1}Mt = \mathcal{O}(\eta^{-1}t^{-1}\epsilon^{-2})$ total calls to U. □

It remains to fix the number of calls to the unitary t in each circuit, however, in doing so we must maintain the large-M limit in which our results were obtained. In the early-FT setting, this is achieved as we fix t by the maximum depth allowable for a circuit (see Sec. IV.4.5 for more details), and achieve arbitrary precision by increasing the number of samples M. In the absence of experimental error, one can fix $t = c_t\epsilon^{-1}$ and recover the Heisenberg limit at sufficiently large M. However, one cannot for instance set $t \propto a_0^{-1}$ and take the limit $a_0 \to 0$, as this would lead to arbitrarily low M.

IV.4.5. Gaussian kernel and global depolarising noise

We now consider the performance of the moment projection estimator again using a Gaussian kernel, but this time under the presence of global depolarizing noise (GDN). Global depolarizing noise assumes that each error event is maximally scrambling; this can be modelled by (with probability p) replacing the quantum state with the maximally mixed state on N qubits; i.e. $\rho \to (1-p)\rho + pI$.

In the presence of GDN, the probability distribution $p(x)$ becomes (as per Def. IV.5)

$$p(x) = Fa * f_\sigma(x) + (1-F)\frac{1}{2\pi} \tag{IV.63}$$

where F is the circuit fidelity. To see that this is correct, one can repeat the calculation in Eq. (IV.18) modifying the probability of each circuit as $p(\tilde{x}|\phi_{\text{ref}}) \to Fp(\tilde{x}|\phi_{\text{ref}}) + (1-F)\frac{1}{K}$, as in the presence of GDN each bitstring is equally probable. The probability of accepting a sample is

$$\int_{\mathcal{D}} \mathrm{d}x\, p(x) = F\sum_j a_j \int_{\mathcal{D}} \mathrm{d}x f_\sigma(x-\phi_j) + (1-F)\frac{|\mathcal{D}|}{2\pi} \tag{IV.64}$$

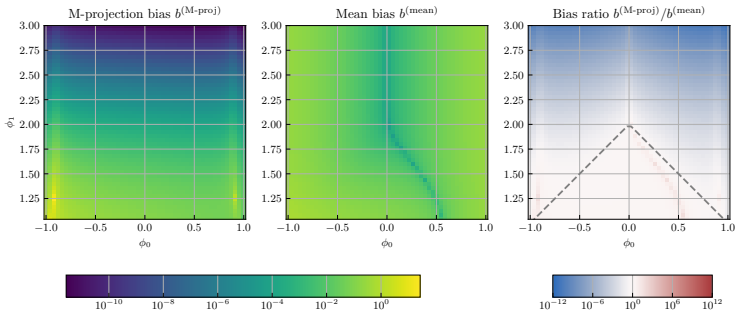


Figure IV.3.: Comparison of the bias of the moment projection and mean estimators for the distribution $P(x)$ in Eq. (IV.55), with phases ϕ_0, ϕ_1, amplitudes $a_0 = 1 - a_1 = 0.7$, filtering region $\mathcal{D} = [-1, 1]$, and kernel width $\sigma = 0.3$. (Left) The first order bias $b^{(\text{M-proj})}$ of the moment projection estimator [first term in Eq. (IV.31)]. (Center) The bias of the mean estimator $b^{(\text{mean})} = \left|\int_{\mathcal{D}} P(x)x\mathrm{d}x - \phi_0\right|$. (Right) The ratio of the two biases, $b^{(\text{M-proj})}/b^{(\text{mean})}$. The dashed lines represent $d = |\phi_0 - \phi_{guess}|$ and mark the regime in which the dominant source of the mean estimator's bias is the distance of ϕ_0 from the center of the interval [Eq. (IV.58)].

and the filtered distribution becomes

$$P(x) = \frac{F\sum_j a_j g_\sigma(x-\phi_j) + (1-F)\frac{1}{2\pi}\mathcal{M}_\sigma}{F\sum_j a_j G_\sigma(\phi_j) + (1-F)\frac{|\mathcal{D}|}{2\pi}\mathcal{M}_\sigma}. \tag{IV.65}$$

where $\mathcal{M}_\sigma = \int_{-\pi}^{\pi} g_\sigma(x)\mathrm{d}x = \mathrm{erf}\left(\frac{\pi}{\sqrt{2}}\sigma^{-1}\right)$ is the normalisation of the kernel function f_σ (Def. IV.4).

In the presence of global depolarizing noise, mean sampling incurs a large sampling overhead compared to the moment projection estimator. Ignoring terms exponentially small in σ for simplicity, the expected value of $P(x)$ is

$$E[x] \approx (1-w)\phi_0 + w\phi_{\text{guess}}, \tag{IV.66}$$

where the noise weight w is given by

$$w = \frac{(1-F)\frac{1}{2\pi}\mathcal{M}_\sigma}{Fa_0 + (1-F)\frac{|\mathcal{D}|}{2\pi}\mathcal{M}_\sigma}. \tag{IV.67}$$

To get an estimator with an exponentially small bias using the sample average $\bar{x}$, we can use a shifted and rescaled mean estimator:

$$\tilde{\phi} = \left(1 + \frac{1-w}{w}\right)\bar{x} - \phi_{guess}\frac{1-w}{w}. \tag{IV.68}$$

However, as $\sigma \to 0$, the variance of P remains constant, which means in turn that the variance of the mean estimator is constant.

In contrast, to extend our moment projection estimator to the setting with GDN, we can include the additional noise term in our model probability, and neglect the contribution of the other phases as before: $Q(x|\phi) \propto F\, a_0\, f_\sigma(x - \phi_j) + (1-F)\,\frac{1}{2\pi}$. We assume that F and a_0 are known; determining these from the data itself would be an interesting target for future work. As in the noiseless case, the bias decreases exponentially with decreasing σ, and the variance decreases as σ^2.

Lemma IV.8
Consider the moment projection phase estimator (Def. IV.13) with Gaussian kernel function f_σ (Def. IV.4 with $\delta = 0$), fidelity F and ground state overlap a_0. Assume that the $\mathcal{D}$ is a promise interval (Def. IV.6) with inner buffer $c = |\mathcal{D}|/6$ and outer buffer d. Then, for

$$\sigma \leq |\mathcal{D}| \min\left(\frac{1}{6}, \log^{-1/2}\left(\frac{1-F}{Fa_0}\right)\right) \tag{IV.69}$$

the asymptotic bias and variance of the estimator in Lemma IV.3 satisfy

$$|\phi - \phi^*| \leq 20\ \sigma^2 g_\sigma(d)\frac{1-a_0}{a_0}\left(1 + \frac{1-F}{Fa_0}\right)\frac{|\mathcal{D}|}{d} + \mathcal{O}(g_\sigma(d)) \tag{IV.70}$$

$$\lim_{M\to\infty} M\,\mathrm{Var}[\tilde{\phi}] \leq 10\ \sigma^2\left(1 + \frac{1-F}{Fa_0}\right)\left(1 + \frac{1}{2\pi}\frac{1-F}{Fa_0}\right) + \mathcal{O}(g_\sigma(d)^2). \tag{IV.71}$$

We provide the proof of Lemma IV.8 in Appendix IV.B. We can now convert the above result into the physical cost to execute phase estimation, assuming a fixed circuit fidelity F.

Theorem IV.6 (Cost of noisy fMPE with fixed fidelity)
Let U be a unitary with spectral gap Δ around a target state ϕ_0 ($\forall_{j>0}|\phi_j -$

$\phi_0| > \Delta$). *Assume oracle access to a controlled version of* U*, and an initial state* $|\psi\rangle$ *such that* $|\langle\phi_0|\psi\rangle| = a_0$*, and global depolarising noise with fixed circuit fidelity* F*. Further assume an initial estimate* ϕ_{guess} *of* ϕ_0 *such that* $|\phi_0 - \phi_{guess}| < \Delta/3$*. Then, the fMPE (Def. IV.12) using* t *calls to* U *per circuit produces an estimate* $\tilde{\phi}$ *with RMS error* ϵ *using* $M = O(a_0^{-1}t^{-2}\epsilon^{-2})$ *samples and* $T = O(\epsilon^{-2}t^{-1}F^{-2}a_0^{-2})$ *total calls to* U*, as long as* $t = \Omega(\Delta^{-1}\log^{-1/2}(\Delta\epsilon^{-1}a_0^{-2}F^{-1}))$.

Proof. As in the proof of Theorem IV.5, we can then generate samples from $P(x)$ with $\mathcal{D}$ that satisfies the assumptions of Lemma IV.8 with $|\mathcal{D}|, d = \Theta(\Delta)$. By Lemma IV.8, $t = \Omega(\Delta^{-1}\log^{-1/2}(\Delta\epsilon^{-1}a_0^{-2}F^{-1}))$ is enough to ensure that the bias is $\mathcal{O}(\epsilon)$. By Lemma IV.6 and Eq. (IV.176) to ensure that the variance is $\mathcal{O}(\epsilon^2)$, we need $M' = \Theta(\sigma^2\epsilon^{-2}F^{-2}a_0^{-2}) = \Theta(t^{-2}\epsilon^{-2}F^{-2}a_0^{-2})$ shots, so $T = M't = \Theta(\epsilon^{-2}t^{-1}F^{-2}a_0^{-2})$ total calls to U. □

Note that while in the noiseless case it was enough to assume a lower bound $\eta \leq a_0$, here we instead assume that both a_0 and F are known exactly. The difference is that in the noiseless setting these parameters do not enter the model $Q(x|\phi)$, and a lower bound on a_0 is only needed to choose the total number of samples M. In the present setting, by contrast, both a_0 and F appear directly in the model, since they determine the relative size of the signal term $Fa_0f(x-\phi_0)$ and the noise term $(1-F)/(2\pi)$. In principle both parameters could be estimated from the data itself; optimizing this would be an interesting task for future work.

So far in this work, we have treated the circuit fidelity and the width of the Gaussian g_σ as independent variables. However, to generate a distribution of width σ using either the methods described in Sec. IV.3.2 or Sec. IV.3.1 requires a circuit depth $t \sim \sigma^{-1}$. Under a typical noise model, the circuit fidelity decreases exponentially in the circuit depth t: $F = e^{-\gamma t}$ for some decay rate γ. This creates a trade-off in choosing the optimal circuit depth between limiting the onset of noise and achieving Heisenberg rather than sampling noise scaling, which we studied in detail in Ref. [76]. We now re-apply these methods to develop a QPE estimator in the presence of global depolarizing noise.

Theorem IV.7 (Cost of noisy fMPE with fixed noise rate)
Let U *be a unitary with spectral gap* Δ *around a target state* ϕ_0 *(*$\forall_{j>0}|\phi_j - \phi_0| > \Delta$*). Assume oracle access to a controlled version of* U*, and an initial state* $|\psi\rangle$ *such that* $|\langle\phi_0|\psi\rangle| = a_0$*, and global depolarising noise with circuit fidelity* $F = e^{-\gamma t}$ *for* t *uses of* U*. Further assume an initial estimate* ϕ_{guess} *of* ϕ_0 *such that* $|\phi_0 - \phi_{guess}| < \Delta/3$*. Then, the fMPE (Def. IV.12) using* t *calls to* U *per circuit produces an estimate* $\tilde{\phi}$ *with RMS error* ϵ *using*

$M = \mathcal{O}(a_0^{-1}t^{-2}\epsilon^{-2})$ samples and $T = \mathcal{O}(\epsilon^{-2}t^{-1}F^{-2}a_0^{-2})$ total calls to U, as long as $t = \Omega(\Delta^{-1}(\gamma\Delta^{-1} + \log(a_0^{-1}\epsilon^{-1})))$.

Proof. As in the proof of Theorem IV.5, we can then generate samples from $P(x)$ with $\mathcal{D}$ that satisfies the assumptions of Lemma IV.7 with $|\mathcal{D}|, d = \Theta(\Delta)$. By Lemma IV.8, to ensure that the bias is $\mathcal{O}(\epsilon)$, it is enough to take $t = \Omega(\Delta^{-1}(\gamma\Delta^{-1}+\log(a_0^{-1}\epsilon^{-1})))$ By Lemma IV.6 and Eq. (IV.176) to ensure that the variance is $\mathcal{O}(\epsilon^2)$, we need $M' = \Theta(\sigma^2\epsilon^{-2}e^{2\gamma t}a_0^{-2}) = \Theta(t^{-2}\epsilon^{-2}e^{2\gamma t}a_0^{-2})$ shots, so $T = M't = \Theta(\epsilon^{-2}t^{-1}e^{2\gamma t}a_0^{-2})$ total calls to U. □

Unlike in Theorems IV.5 and IV.6, where one can take $t = \Theta(\epsilon^{-1})$ and obtain Heisenberg-limited scaling $T = \Theta(\epsilon^{-1})$, such a choice is no longer optimal here, as the cost grows exponentially with t. Instead, as in Ref. [76], optimizing over t yields an optimal depth $t = \Theta(\gamma^{-1})$ and total cost $T = \Theta(\gamma\epsilon^{-2}a_0^{-2})$. Thus, the optimal circuit depth is set by the noise rate rather than the target precision. This generalizes the result $T = \Theta(\gamma\epsilon^{-2})$ of Ref. [76] to the case of imperfect initial states, with an additional multiplicative overhead of a_0^{-2}.

IV.5. Numerical demonstration

In this section, we implement and demonstrate the various estimators described throughout this text on a simplified problem for illustrative purposes. This allows us to show the potential phase estimation practitioner how the above theory can be used in practice. It furthermore allows us to test the effect of finite statistics on our estimators. The results derived above typically hold in the asymptotic limit in the number of samples, but the experimentalist is far more concerned over whether they need 5 or 5 million circuit samples to guarantee robustness. And finally, the results allow us to check the critical constant factors in our phase estimation performance.

For simplicity, in this section we consider a 1D Ising model on 4 qubits:

$$H = h\sum_{i=1}^{4} Z_i + t\sum_{i=1}^{3} Z_iZ_{i+1} \tag{IV.72}$$

with magnetic field $h = 0.27$, and coupling strength $t = -0.46$. This is chosen to give a well-separated ground state, $\phi_1 - \phi_0 = 1.46$, and we can choose a simple filtering interval $\mathcal{D} = [-\pi, -\pi/2]$ [Fig. IV.5] that contains only ϕ_0. We obtain an initial state by applying Y rotations with angle $\theta = 0.8\,\text{rad}$ to each qubit, yielding a ground state overlap $a_0 \approx 0.52$. In Fig. IV.5, we plot the spectral function $a(x)$ [Eq. (IV.5)], as well as its convolution with the Fejer kernel [Eq. (IV.17)] from a textbook QPE implementation [29].

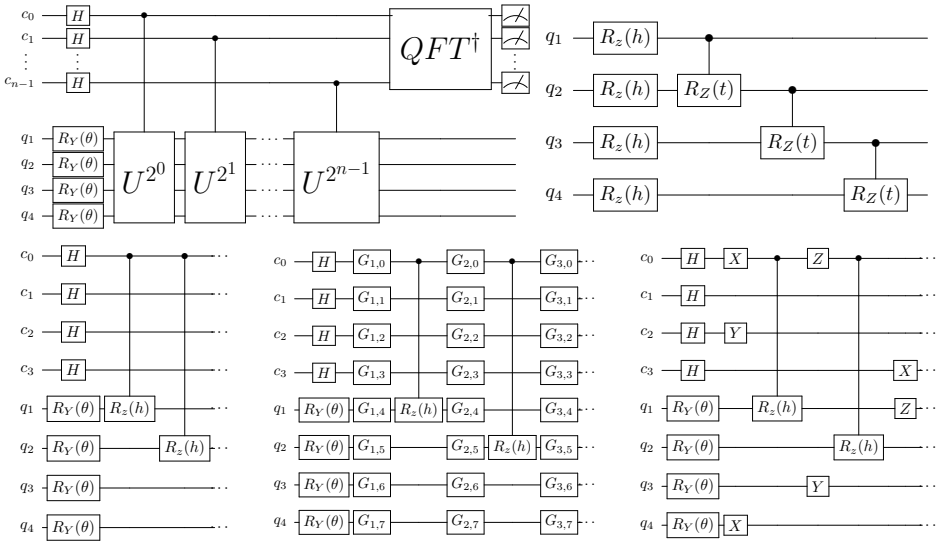


Figure IV.4.: Circuits used in this work for example numerical QPE implementation. *Top row:* (Left) Noiseless circuit (Right) Decomposition of the system unitary $U = \prod_j e^{-itZ_jZ_{j+1}} \prod_j e^{-ihZ_j}$. *Bottom row:* (Left) First layers of the noiseless circuit for $n = 4$ control qubits; (Centre) insertion of stochastic gates $G_{i,j}$ after each moment; (Right) one explicit noisy circuit realization obtained by sampling Pauli errors.

We implement the full QPE circuit (see Fig. IV.4, top row) using a quantum register of variable size (n between 4 and 10 control qubits) using cirq. For a simple noise model, we choose local depolarizing noise

$$\rho \to (1 - p_{\text{err}})\rho + \tfrac{p_{\text{err}}}{3}(X_j\rho X_j + Y_j\rho Y_j + Z_j\rho Z_j), \tag{IV.73}$$

applied on all qubits between each moment of gates (see Fig. IV.4, bottom row). We fix the error rate p_{err} such that the probability of no error is $(1 - p_{\text{err}})^{N_q \times N_{\text{d}}} = \frac{1}{e}$, where N_q is the total number of qubits (system plus control) and N_{d} is the circuit depth. The resultant ochre probability distribution $p(x)$ in Fig. IV.5 can be split into a rescaled noiseless distribution and the distribution coming from circuits where at least one error has occurred [Fig. IV.5, maroon]

$$p(x) = (1 - F)p_1(x) + F(a * f)(x), \tag{IV.74}$$

which we will use as the quasiprobability distribution when testing the fN-

MPE[4]. We observe that the distribution $p_1(x)$ in Fig. IV.5 still retains a peak around ϕ_0, implying that it contains significant phase estimation information as well. This makes sense, as the effective volume of our QPE circuit is likely small, and PEC schemes that make use of this [258] can likely be incorporated into the fNMPE as well. Pseudocode for the sampling routines used in the above is as follows:

Algorithm 1:

Filtered sampler

Input: Filtering interval $\mathcal{D} \subset [-\pi, \pi]$,
Number of circuit shots $M \in \mathbb{N}$.
Output: Samples X
for $j \leftarrow 1, \ldots, M$ **do**
  Sample noisy circuit C
  $P(x) \leftarrow |\langle x|\, C\, |0\rangle|^2$ `// Simulate` C
  Sample $x_j \sim P(x)$
  if $x_j \in \mathcal{D}$ **then**
    add x_j to X
return X

Algorithm 2:

Filtered PEC sampler

Input: Filtering interval $\mathcal{D} \subset [-\pi, \pi]$,
Fidelity $F \in (0, 1]$,
Number of circuit shots $M \in \mathbb{N}$.
Output: Samples X_0, X_1
$\alpha \leftarrow [F^{-1}, 1 - F^{-1}]$ **for** $j \leftarrow 1$ *to* M **do**
  Sample $a \sim \frac{|\alpha_a|}{\|\alpha\|_1}$
  if $a = 0$ **then**
    Sample noisy circuit C
    $P(x) \leftarrow |\langle x|\, C\, |0\rangle|^2$
      `// Simulate` C
    Sample $x \sim P(x)$
  else
    repeat
      Sample noisy circuit C
    until C contains at least one error;
    $P(x) \leftarrow |\langle x|\, C\, |0\rangle|^2$
      `// Simulate` C
    Sample $x \sim P(x)$
  if $x \in \mathcal{D}$ **then**
    add x to X_a
return X_0, X_1

With our schemes for sampling from QPE probability distributions given, we now give pseudocode implementations for each of the estimators given in the text. As we are using textbook quantum phase estimation, our model distribution $Q(x|\phi)$ for fMPE and fNMPE is the single phase distribution in Defs. IV.13, IV.14 with the Fejer kernel (Eq. (IV.17)).

[4]This decomposition ignores the overhead from constructing a real quasiprobability distribution, as in practice we do not have access to $p_1(x)$.

Algorithm 3:

Filtered PEC average

Input: Samples X_0, X_1 generated by Alg. 2

Output: Estimate $\tilde{\phi}$

$M_0 \leftarrow |X_0|$

$M_1 \leftarrow |X_1|$

return $\tilde{\phi} = \frac{\sum_{x\in X_0} x - \sum_{x\in X_1} x}{M_0 - M_1}$

Algorithm 4:

fMPE assuming GDN

Input: Samples X generated by Alg. 1, promise interval $\mathcal{D}$, number of control qubits $n \in \mathbb{N}$, fidelity $F \in (0, 1]$, overlap $a_0 \in (0, 1]$

Output: Estimate $\tilde{\phi}$

$q(x|\phi) \leftarrow F a_0 f_{2^n}^{(\text{Fejer})}(x - \phi) + (1 - F)\frac{1}{2\pi}$;

$Q(x|\phi) \leftarrow \frac{q(x|\phi)}{\int_{\mathcal{D}} q(x|\phi)\mathrm{d}x}$;

$\ell(\phi) \leftarrow \frac{1}{M}\sum_{x\in X} \log Q(x|\phi)$;

return $\tilde{\phi} = \arg\max \ell(\phi)$

Algorithm 5:

fNMPE

Input: Samples X_0, X_1 generated by Alg. 2, promise interval $\mathcal{D}$, number of control qubits $n \in \mathbb{N}$

Output: Estimate $\tilde{\phi}$

$q(x|\phi) \leftarrow f_{2^n}^{(\text{Fejer})}(x - \phi)$;

$Q(x|\phi) \leftarrow \frac{q(x|\phi)}{\int_{\mathcal{D}} q(x|\phi)\mathrm{d}x}$;

$\ell(\phi) \leftarrow \frac{1}{M}\sum_{x\in X_0} \log Q(x|\phi) - \frac{1}{M}\sum_{x\in X_1} \log Q(x|\phi)$;

return $\tilde{\phi} = \arg\max \ell(\phi)$

In Fig. IV.6, we show the bias and the variance of all three estimators on our Ising model QPE simulation, for a range of control register sizes n and numbers of circuit repetitions M (shots). (Note that we change the error rate with the number of qubits n in order to maintain a constant circuit fidelity $F = \frac{1}{e}$.) We observe that all estimators converge quickly in variance to the standard quantum limit $M^{-1/2}$ with different constant factors, and the moment projection estimators converge as 2^{-n} whilst the mean estimator has a constant variance in n, as expected from the prior sections. The moment projection estimator under a GDN assumption achieves this convergence to the standard quantum limit with fewer than 50 circuit repetitions [5], making it an attractive choice for early fault-tolerant QPE implementations. The bias of the mean estimator and the fMPE under a GDN assumption saturate as a function of M, which can be expected due to them not properly modeling experimental noise. However the fNMPE without regularization has 0 bias within the error bars of the simulation. (In practice we expect a small residual bias due to the imperfect initial state, but this is likely too small to be observed.) The bias trend with n is less pronounced. We see some evidence that that the fMPE and fNMPE have an exponentially decreasing bias while the mean estimator saturates. However, more simulation at larger system

[5] For the given starting state and circuit fidelity, filtering removes around 80% of the data in these simulations, and the likelihood maximization is performed on only around 10 data points at $M = 50$.

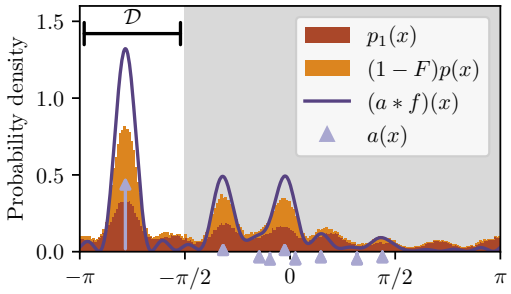


Figure IV.5.: Spectral distribution and QPE results on a 4-qubit Ising model, with $n = 4$ QPE control qubits. The arrows represent the eigenphases and corresponding amplitudes for the fast-forwarded unit-time evolution unitary of the considered model, i.e. the spectral function $a(x)$. The blue line shows the smoothed spectral function $f * a(x)$, with the kernel function corresponding to the continuous Fejér kernel expected from textbook QPE: $f(x) = f^{(\text{Fejer})}K(x)$ from Eq. (IV.17) with $K = 2^n$. The ochre histogram shows the samples obtained from simulating $100k$ shots of textbook QPE with circuit local depolarizing noise, with error probabilities tuned such that the total circuit fidelity is $F = e^{-1}$ [i.e. samples from the noisy QPE distribution $p(x)$, Eq. (IV.74)]. The maroon histogram represents the subset of samples coming from simulations where at least one noise event happened [i.e. samples from the pure noise distribution, $p_1(x)$, Eq. (IV.74)], which we use to simulate a simplified version of PEC/Explicit Unbiasing. The un-shaded area represents the filtering interval $\mathcal{D}$.

sizes is required for a convincing demonstration of this result, and we present this solely as an example of the implementation of the estimators from this work.

IV.6. Conclusion and outlook

In this work, we extended the classical processing schemes for quantum phase estimation that we introduced in Ref. [76] to the case where we cannot prepare an initial eigenstate. This utilized prior knowledge of a promise region containing only the eigenstate of interest, which allows us to post-select away data outside this region while avoiding the bias potentially introduced by this post-selection. In the absence of noise, our estimator improves over a naive mean estimation by correcting for the bias due to the choice of the filtering

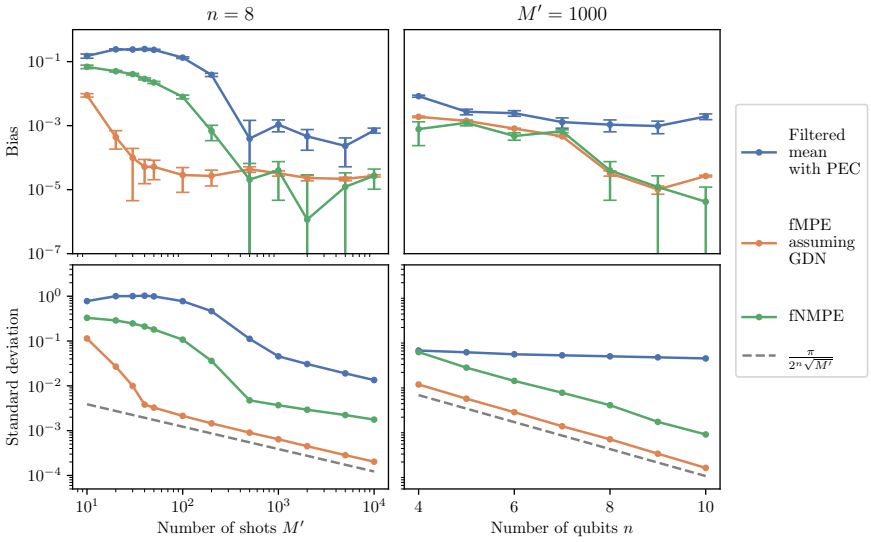


Figure IV.6.: Performance of estimators considered in this work for 4-qubit Ising model [estimators distinguished in legend]. The upper two plots show the estimator bias $|\tilde{\phi} - \phi_0|$, while the bottom two plots show the standard deviation. For the plots on the left, the number of control qubits n is fixed to 8, and the total number of shots taken on a quantum computer (i.e. number of samples before filtering) M' is varied. For the plots on the right, $M' = 1000$, and n is varied. Error bars are obtained by bootstrapping [259].

interval. We further extended our work to consider global depolarizing noise, which can be handled via a relatively simple change to the model distribution $Q(x|\phi)$ in our estimator. The resulting estimator remains asymptotically normal. Furthermore, for Gaussian kernels with variance σ, it achieves a bias exponentially small in σ^{-1} and a variance scaling as σ^2. We combined the estimators developed in this work with the explicit unbiasing scheme of Ref. [76], and tested both schemes in an illustrative example of QPE on an Ising model. We observe that the moment projection estimator is surprisingly noise-robust even in the absence of explicit unbiasing, and requires < 50 shots to overcome small-sample-statistics effects and converge to an expected asymptote (at least for the parameters chosen).

Our new estimator solves a key roadblock to quantum phase estimation in early fault tolerance, where high-overlap state preparation is likely unfeasible and noise remains a concern. However, it remains to test the performance

of the moment projection estimator on realistic noise models and phase estimation problems, and to project performance at beyond-classical system sizes. This is especially relevant given the observed noise-robustness, which suggests that the overhead for explicit unbiasing may not be necessary across a range of target error rates. One can further consider learning a few phases within the promise region $\mathcal{D}$ instead of just one, to allow for excited state energy estimation, or ground state estimation of gapless systems. It may also be worth re-optimizing the choice of kernel function in the presence of noise (both within the QFT-QPE and QSP-QPE framework), as previous optimality results do not hold in the presence of noise. And finally, it remains to compile the results of this work to make resource estimates for phase estimation of real problems of interest, in fault-tolerant architectures being pursued today.

IV.A. Proofs of Lemmas IV.3, IV.5, IV.4, and Theorem IV.4

IV.A.1. Proof of Lemmas IV.3 and IV.5

In this section we prove Lemma IV.5. Note that this also proves Lemma IV.3 by fixing $d = 1$, $\alpha = 1$, and $c = 0$ (in which case $P_c(x) = R(x) = P(x), Q_c(x|\phi) = Q(x|\phi)$).

The estimate $\tilde{\phi}$ satisfies the stationary point condition

$$\ell'(\tilde{\phi}|\{x_j, a_j\}) = 0, \tag{IV.75}$$

where $\tilde{\phi}$ and ℓ are defined in Def. IV.11. Expanding this equation around ϕ^* (as defined in Lemma IV.5) gives

$$\ell'(\tilde{\phi}|\{x_j, a_j\}) = \ell'(\phi^*|\{x_j, a_j\}) + (\tilde{\phi} - \phi^*)\ell''(\phi^*|\{x_j, a_j\}) + O((\tilde{\phi} - \phi^*)^2), \tag{IV.76}$$

which implies

$$\tilde{\phi} - \phi^* = -\frac{\ell'(\phi^*|\{x_j, a_j\})}{\ell''(\phi^*|\{x_j, a_j\})} + O((\tilde{\phi} - \phi^*)^2). \tag{IV.77}$$

We now analyze the numerator and denominator of the expression above in the limit $M \to \infty$. First, we consider the denominator. Using the definition of $\ell(\phi^*|\{x_j, a_j\})$,

$$\ell''(\phi^*|\{x_j, a_j\}) =$$

$$\frac{\|\alpha\|_1}{M}\sum_{j=1}^{M}\mathrm{sgn}(\alpha_{a_j})[\partial_\phi^2\log Q_c(x_j|\phi)]_{\phi=\phi^*}+c\int_{\mathcal{D}}\mathrm{d}x\,[\partial_\phi^2\log Q_c(x|\phi)]_{\phi=\phi^*}. \tag{IV.78}$$

The pairs (x_j, a_j) are independent samples generated according to the procedure in Def. IV.11, so in the limit $M\to\infty$, by the law of large numbers the mean above converges to its expectation value

$$\begin{aligned}\mathbb{E}[\ell''(\phi^*|\{x_j,a_j\})] &= \|\alpha\|_1\sum_{a=1}^{d}\frac{|\alpha_a|}{\|\alpha\|_1}\int_{\mathcal{D}}\mathrm{d}x\,P_a(x)\mathrm{sgn}(\alpha_a)[\partial_\phi^2\log Q_c(x|\phi)]_{\phi=\phi^*}\\ &\quad+c\int_{\mathcal{D}}\mathrm{d}x\,[\partial_\phi^2\log Q_c(x|\phi)]_{\phi=\phi^*} &\text{(IV.79)}\\ &=\int_{\mathcal{D}}\mathrm{d}x\sum_{a=1}^{d}\alpha_a\,P_a(x)[\partial_\phi^2\log Q_c(x|\phi)]_{\phi=\phi^*}\\ &\quad+c\int_{\mathcal{D}}\mathrm{d}x\,[\partial_\phi^2\log Q_c(x|\phi)]_{\phi=\phi^*} &\text{(IV.80)}\\ &=\int_{\mathcal{D}}\mathrm{d}x\,P_c(x)[\partial_\phi^2\log Q_c(x|\phi)]_{\phi=\phi^*} &\text{(IV.81)}\\ &\equiv\bar{\ell}''(\phi_*). &\text{(IV.82)}\end{aligned}$$

Here, $P_c(x)$ is the regularized distribution defined in Lemma IV.3, and $\bar{\ell}(\phi) := \int_{\mathcal{D}}\mathrm{d}x\,P_c(x)\log Q_c(x|\phi)$ is the objective function in the limit $M\to\infty$.

Next consider the numerator

$$\begin{aligned}\ell'(\phi^*|\{x_j,a_j\}) = \frac{\|\alpha\|_1}{M}\sum_{j=1}^{M}\mathrm{sgn}(\alpha_{a_j})[\partial_\phi\log Q_c(x_j|\phi)]_{\phi=\phi^*}\\ +c\int_{\mathcal{D}}\mathrm{d}x\,[\partial_\phi\log Q_c(x|\phi)]_{\phi=\phi^*}.\end{aligned} \tag{IV.83}$$

This is again a mean of independent random variables, and it converges to

$$\begin{aligned}\mathbb{E}[\ell'(\phi^*|\{x_j,a_j\})] &= \|\alpha\|_1\sum_{a=1}^{d}\frac{|\alpha_a|}{\|\alpha\|_1}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{sgn}(\alpha_a)P_a(x)[\partial_\phi\log Q_c(x|\phi)]_{\phi=\phi^*}\\ &\quad+c\int_{\mathcal{D}}\mathrm{d}x\,[\partial_\phi\log Q_c(x|\phi)]_{\phi=\phi^*} &\text{(IV.84)}\\ &=\int_{\mathcal{D}}\mathrm{d}x\,P_c(x)[\partial_\phi\log Q_c(x|\phi)]_{\phi=\phi^*} &\text{(IV.85)}\end{aligned}$$

$$= \left[\partial_\phi \int_{\mathcal{D}} \mathrm{d}x\, P_c(x) \log Q_c(x|\phi)\right]_{\phi=\phi^*} \tag{IV.86}$$

$$= \bar{\ell}'(\phi_*). \tag{IV.87}$$

By definition ϕ^* is the maximum of $\bar{\ell}(\phi)$, and satisfies the stationary point condition $\bar{\ell}'(\phi_*) = 0$. Therefore the expectation value of the numerator vanishes. To calculate the variance, we note that the second term of Eq. (IV.83) is not a random variable, and so

$$\mathrm{Var}[\ell'(\phi^*|\{x_j, a_j\})] = \mathrm{Var}\Big[\frac{\|\alpha\|_1}{M} \sum_{j=1}^{M} \mathrm{sgn}(\alpha_{a_j})[\partial_\phi \log Q_c(x_j|\phi)]_{\phi=\phi^*}\Big] \tag{IV.88}$$

$$= \frac{1}{M}\mathrm{Var}\left[\|\alpha\|_1 \mathrm{sgn}(\alpha_a)[\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi^*}\right] \tag{IV.89}$$

$$= \frac{1}{M}\Bigg(\mathbb{E}\Big[\Big(\|\alpha\|_1 \mathrm{sgn}(\alpha_a)[\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi^*}\Big)^2\Big] - \mathbb{E}\Big[\|\alpha\|_1 \mathrm{sgn}(\alpha_a)[\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi^*}\Big]^2\Bigg). \tag{IV.90}$$

We can evaluate the second term using Eq. (IV.84) above, which yields

$$\mathrm{Var}[\ell'(\phi^*|\{x_j, a_j\})] = \frac{\|\alpha\|_1^2}{M}\Bigg(\int_{\mathcal{D}} \mathrm{d}x\, R(x)\, [\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi^*} - c^2\left[\int_{\mathcal{D}} \mathrm{d}x\, [\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi^*}\right]^2\Bigg) \tag{IV.91}$$

$$\leq \frac{\|\alpha\|_1^2}{M} \int_{\mathcal{D}} \mathrm{d}x\, R(x)\, [\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi^*} \tag{IV.92}$$

and so by the central limit theorem,

$$\sqrt{M}\, \ell'(\phi^*|\{x_j, a_j\}) \tag{IV.93}$$

converges in distribution to a Gaussian with mean 0 and variance

$$\leq \|\alpha\|_1^2 \int_{\mathcal{D}} \mathrm{d}x\, R(x) [\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi^*}. \tag{IV.94}$$

Now we combine these results with Eq. (IV.77). Since the maximizer of $\int_{\mathcal{D}} \mathrm{d}x\, P_c(x) \log Q_c(x|\phi)$ is unique and lies in the interior of D_ϕ, the estimator converges to $\tilde{\phi} \to \phi^*$ as $M \to \infty$ and the Taylor expansion becomes valid. Then, applying Slutsky's theorem, $\sqrt{M}(\tilde{\phi} - \phi^*)$ converges in distribution to

a Gaussian with mean 0 and variance

$$\epsilon^2 \leq \|\alpha\|_1^2 \frac{\int_{\mathcal{D}} \mathrm{d}x\, R(x)[\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi^*}}{\left(\int_{\mathcal{D}} \mathrm{d}x\, P_c(x)[-\partial^2_\phi \log Q_c(x|\phi)]_{\phi=\phi^*}\right)^2}. \tag{IV.95}$$

□

IV.A.2. Proof of Lemma IV.4

We first analyze the bias in the moment projection estimator (Def. IV.10).

By definition ϕ^* maximizes $\bar{\ell}(\phi) = \int_{\mathcal{D}} \mathrm{d}x\, P(x) \log Q(x|\phi)$, and therefore satisfies the stationary point condition

$$0 = \bar{\ell}'(\phi^*) = \int_{\mathcal{D}} \mathrm{d}x (Q(x|\phi_0) + h(x)) \left[\partial_\phi \log Q(x|\phi)\right]_{\phi=\phi^*} \tag{IV.96}$$

When $h(x) = 0$, $\phi^* = \arg\max_\phi D_{KL}(Q(x|\phi_0)||Q(x|\phi)) = \phi_0$. Therefore for small $\|h\|$ we can expand the above equation around ϕ_0:

$$0 = \int_{\mathcal{D}} \mathrm{d}x (Q(x|\phi_0) + h(x)) \big([\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0} + \\ + (\phi^* - \phi_0)[\partial^2_\phi \log Q(x|\phi)]_{\phi=\phi_0} \big) + O((\phi^* - \phi_0)^2) \tag{IV.97}$$

Rearranging this formula obtains

$$(\phi^* - \phi_0) = -\frac{\int_{\mathcal{D}} \mathrm{d}x\, (Q(x|\phi_0) + h(x))[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}}{\int_{\mathcal{D}} (Q(x|\phi_0) + h(x))[\partial^2_\phi \log Q(x|\phi)]_{\phi=\phi_0}} + O((\phi^* - \phi_0)^2). \tag{IV.98}$$

Then, since $Q(x|\phi)$ is normalised, we have that

$$\int_{\mathcal{D}} \mathrm{d}x\, Q(x|\phi_0)[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0} = \int_{\mathcal{D}} \mathrm{d}x\, [\partial_\phi Q(x|\phi)]_{\phi=\phi_0}(x) = 0, \tag{IV.99}$$

so the numerator is

$$\int_{\mathcal{D}} \mathrm{d}x\ h(x)[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0} \leq \|h\|_1 \max_x [\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}. \tag{IV.100}$$

We expand the denominator of Eq. (IV.98) to lowest order in $\|h\|_1 =$

$\int_{\mathcal{D}} |h(x)| \mathrm{d}x$ as

$$\int \mathrm{d}x\ Q(x|\phi_0)[-\partial_\phi^2 \log Q(x|\phi)]_{\phi=\phi_0} = \int \mathrm{d}x\ Q(x|\phi_0)\left(\left(\frac{[\partial_\phi Q(x|\phi)]_{\phi=\phi_0}}{Q_{\phi_0(x)}} \right)^2 - \frac{[\partial_\phi^2 Q(x|\phi)]_{\phi=\phi_0}}{Q_{\phi_0(x)}} \right) \tag{IV.101}$$

$$= \int \mathrm{d}x\ Q(x|\phi_0)[\partial_\phi \log Q(x|\phi)]^2_{\phi=\phi_0} - \partial_\phi^2 \left[\int \mathrm{d}x\ Q(x|\phi_0) \right]_{\phi=\phi_0} \tag{IV.102}$$

$$= \mathcal{I}_0. \tag{IV.103}$$

Here, the second term of the second-to-last-line evaluates to zero again by the normalisation of $Q(x|\phi)$ (Eq. (IV.99)). We can simplify

$$|\phi^* - \phi_0| \leq \|h\|_1 \frac{\max_x [\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}}{\mathcal{I}_0} + O(\|h\|^2), \tag{IV.104}$$

where $\mathcal{I}_0$ is the Fisher information of Q with respect to ϕ, which is exactly Eq. (IV.32)

We now bound the variance ϵ^2,

$$\epsilon^2 = \|\alpha\|_1^2 \frac{\int \mathrm{d}x\, P(x)(\partial_\phi \log Q(x|\phi^*))^2}{\left(\int \mathrm{d}x\, P(x)[-\partial_\phi^2 \log Q(x|\phi^*)] \right)^2}. \tag{IV.105}$$

Since $\phi^* = \phi_0 + O(\|h\|_1)$ we may replace ϕ^* with ϕ_0 up to $O(\|h\|_1)$ corrections.

$$\epsilon^2 = \|\alpha\|_1^2 \frac{\int \mathrm{d}x\, P(x)(\partial_\phi \log Q(x|\phi_0))^2}{\left(\int \mathrm{d}x\, P(x)[-\partial_\phi^2 \log Q(x|\phi_0)] \right)^2} + \mathcal{O}(\|h\|_1). \tag{IV.106}$$

As in Eq. (IV.101), the leading order of the denominator in $\|h\|_1$ is $\mathcal{I}_0^2$. We finally recover Eq. (IV.33) by writing $P(x) = Q(x|\phi_0) + \mathcal{O}(\|h\|_1)$ and collecting terms. □

IV.A.3. Proof of Theorem IV.4

Our proof for Theorem IV.4 follows similarly to the proof of Lemma IV.4; the calculation for the bias follows Eq. (IV.77)-(IV.104) under the substitutions $Q(x|\phi) \to Q_c(x|\phi)$, $P(x) \to P(x) + c$. However, our calculation for the

variance does not simplify in the same way, as we integrate the numerator over the marginal distribution $R(x)$, which we expect to be significantly different to $P_c(x)$. Instead, because $R(x)$ is a valid, normalized probability distribution, we can bound the numerator as

$$\int \mathrm{d}x\, R(x)[\partial_\phi \log Q_c(x|\phi)]_{\phi=\phi_0}]^2 \leq \max_x [\partial_\phi \log Q_c(x|\phi)]^2_{\phi=\phi_0}, \tag{IV.107}$$

and identify within the denominator

$$\int_{\mathcal{D}} \mathrm{d}x\, P_c(x)[-\partial^2_\phi \log Q_c(x|\phi)]_{\phi=\phi_*} = \mathcal{I}_c + \mathcal{O}(\|h\|_1), \tag{IV.108}$$

from which the result follows. □

IV.B. Proofs of Lemmas IV.7 and IV.8

In this appendix we calculate the application of the result of Lemma IV.4 to the case of a Gaussian kernel, both without noise and with global depolarizing noise. First, we note that for a Gaussian kernel function, the regularity assumptions in Lemma IV.4 are satisfied in both cases, so we can use the bounds in this lemma. Our target in both cases is to bound the bias $b = |\phi^* - \phi_0|$ [Eq. (IV.32)] and the variance $\lim_{M\to\infty} M\mathrm{Var}[\tilde{\phi}]$ [Eq. (IV.33)]. We achieve this bound piecewise. First, in App. IV.B.1 we bound the 1-norm of the the model deviation h from above by using the assumption on the distance of spurious phases from the filtering interval [Eq. (IV.60)]. Then, in App. IV.B.2 we bound the normalization constant that arises from cutting the Gaussian model distribution to a finite filtering interval, and its derivative with respect to ϕ_0, using the existence of gaps between the edges of the filtering interval and the target phase [Eq. (IV.59)]. Next, in we derive an upper bound on the score of the model distribution (App. IV.B.3) and a lower bound on the Fisher information (App. IV.B.4, IV.B.5). This is finally summarized in App. IV.B.6, where we derive the explicit forms of the bounds appearing in Lemma IV.7 and Lemma IV.8.

Before we begin, let us define some notation to be used in the rest of the section. The model distribution used to define the moment projection estimator is

$$Q(x|\phi) = \frac{q(x|\phi)}{\int_{\mathcal{D}} \mathrm{d}x\, q(x|\phi)} \tag{IV.109}$$

$$q(x|\phi) = Fa_0 g_\sigma(x-\phi) + (1-F)\frac{\mathcal{M}_\sigma}{2\pi}, \tag{IV.110}$$

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \tag{IV.111}$$

$$\mathcal{M}_\sigma = \int_{-\pi}^{\pi} g_\sigma(x)\mathrm{d}x = \mathrm{erf}\left(\frac{\pi}{\sqrt{2}}\sigma^{-1}\right). \tag{IV.112}$$

Note that the case of no noise is included by taking $F = 1$. The model $q(x|\phi)$ is defined so that $q(x|\phi)/\mathcal{M}_\sigma$ matches $p(x)$ in the case of $a_0 = 1$, and $\mathcal{N}/\mathcal{M}_\sigma$ is equal to P_A in Eq. (IV.46).

For brevity of notation, let $\mathcal{N}$ be the normalization of $Q(x|\phi_0)$

$$\mathcal{N} = \int_{\mathcal{D}} q(x|\phi_0)\mathrm{d}x, \tag{IV.113}$$

and let $(.)'$ denote derivative with respect to ϕ,

$$q'_{\phi_0}(x) = \left[\partial_\phi q(x|\phi)\right]_{\phi=\phi_0}, \tag{IV.114}$$

$$\mathcal{N}' = \left[\partial_\phi \int_{\mathcal{D}} q(x|\phi)\mathrm{d}x\right]_{\phi=\phi_0} = \int_{\mathcal{D}} q'_{\phi_0}(x)\mathrm{d}x. \tag{IV.115}$$

$$\tag{IV.116}$$

IV.B.1. Step 1: Bound on $\|h\|$

In order to bound the 1-norm of the model deviation, $h(x) = P(x) - Q(x|\phi_0)$, we calculate a bound on a related distribution, $\|u\|$, defined by

$$u(x) = F\sum_{j\neq 0} a_j g_\sigma(x - \phi_j) \tag{IV.117}$$

so the true probability distribution is

$$P(x) = \frac{q(x|\phi_0) + u(x)}{\mathcal{N} + \|u\|}. \tag{IV.118}$$

Using the triangle inequality (noting that $u(x), q(x|\phi_0) > 0$), we can bound $\|h\|$ as

$$\|h\| = \int_{\mathcal{D}} \mathrm{d}x \ |P(x) - Q(x|\phi_0)| \tag{IV.119}$$

$$= \int_{\mathcal{D}} \mathrm{d}x \ \left|\frac{q(x|\phi_0) + u(x)}{\mathcal{N} + \|u\|} - \frac{q(x|\phi_0)}{\mathcal{N}}\right| \tag{IV.120}$$

$$= \frac{\int_{\mathcal{D}} \mathrm{d}x \ |\mathcal{N}u(x) - \|u\|q(x|\phi_0)|}{\mathcal{N}(\mathcal{N} + \|u\|)} \tag{IV.121}$$

$$\leq \frac{\int_{\mathcal{D}} \mathrm{d}x \ (\mathcal{N}u(x) + \|u\| q(x|\phi_0))}{\mathcal{N}(\mathcal{N} + \|u\|)} \tag{IV.122}$$

$$= \frac{2\mathcal{N}\|u\|}{\mathcal{N}(\mathcal{N} + \|u\|)} \tag{IV.123}$$

$$\leq 2\frac{\|u\|}{\mathcal{N}}. \tag{IV.124}$$

Now, we use the assumption on the distance of the spurious phases $\phi_{j\neq 0}$ from the filtering region,

$$\min_{j\neq 0} \min_{x\in\mathcal{D}} |x - \phi_j| \geq d \tag{IV.125}$$

to bound

$$\|u\| = F \sum_{j\neq 0} a_j G(\phi_j). \tag{IV.126}$$

We can bound

$$G_\sigma(\phi_j) = \int_{\mathcal{D}} g_\sigma(x-\phi)\mathrm{d}x \leq \int_d^\infty g_\sigma(y)dy \leq \frac{\sigma^2}{d} g_\sigma(d), \tag{IV.127}$$

where the last inequality is obtained from standard bounds on Mill's ratio for a normal distribution. Then, substituting this into Eq. (IV.126) yields

$$\|u\| \leq F(\sum_{j\neq 0} a_j)\frac{\sigma^2}{d} g_\sigma(d) \tag{IV.128}$$

$$= F(1-a_0)\frac{\sigma^2}{d} g_\sigma(d). \tag{IV.129}$$

Combining the above with Eq. (IV.124) yields

$$\|h\| \leq 2\mathcal{N}^{-1} F(1-a_0)\frac{\sigma^2}{d} g_\sigma(d). \tag{IV.130}$$

IV.B.2. Step 2: Bounds on $\mathcal{N}$, $\mathcal{N}'$

In this step we bound the norm $\mathcal{N}$ and its derivative $\mathcal{N}'$ using the assumption that ϕ_0 is well inside the filtering region $\mathcal{D} = [\phi_{guess} - |\mathcal{D}|/2, \phi_{guess} + |\mathcal{D}|/2]$:

$$|\phi_{guess} - \phi_0| \leq \frac{|\mathcal{D}|}{3}. \tag{IV.131}$$

Using the definition of the Gaussian model $q(x|\phi)$ in Eq. (IV.110), we can

evaluate

$$\mathcal{N} = Fa_0 G_\sigma(\phi) + (1-F)\frac{\mathcal{M}_\sigma|\mathcal{D}|}{2\pi}, \tag{IV.132}$$

$$\mathcal{N}' = Fa_0(g_\sigma(\phi_{guess} - \phi_0 - |\mathcal{D}|/2) - g_\sigma(\phi_{guess} - \phi_0 + |\mathcal{D}|/2)). \tag{IV.133}$$

$$\tag{IV.134}$$

The normalization $\mathcal{N}$ is minimal when $|\phi_{guess} - \phi_0|$ is maximal, for $\phi_0 = \phi_{guess} + |\mathcal{D}|/3$, and

$$\begin{aligned}
\mathcal{N} &\geq Fa_0 G_\sigma(\phi) && \text{(IV.135)}\\
&\geq Fa_0 G_\sigma(\phi_{guess} + |\mathcal{D}|/3) && \text{(IV.136)}\\
&= Fa_0 \frac{\operatorname{erf}(\frac{5}{6}\frac{|\mathcal{D}|}{\sigma}) + \operatorname{erf}(\frac{1}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma})}{2}. && \text{(IV.137)}
\end{aligned}$$

Conversely, the derivative of the normalization reaches a maximum when $|\phi_{guess} - \phi_0|$ is maximal, for $\phi_0 = \phi_{guess} + |\mathcal{D}|/3$ (to convince oneself one can check that $\mathcal{N}'' < 0$ for $\phi_0 \in \mathcal{D}$). Thus, we can bound

$$\begin{aligned}
|\mathcal{N}'| &\leq Fa_0\left(g_\sigma\left(\frac{|\mathcal{D}|}{6}\right) - g_\sigma\left(\frac{5|\mathcal{D}|}{6}\right)\right) && \text{(IV.138)}\\
&\leq Fa_0 g_\sigma\left(\frac{|\mathcal{D}|}{6}\right) && \text{(IV.139)}\\
&= \frac{Fa_0}{\sigma} g_1\left(\frac{|\mathcal{D}|}{6\sigma}\right). && \text{(IV.140)}
\end{aligned}$$

We can also bound the norm from above as

$$\mathcal{N} \leq Fa_0 + (1-F)\frac{1}{2\pi} \tag{IV.141}$$

since $G_\sigma(\phi)$ and $\mathcal{M}_\sigma$ are both bounded by 1 for all values of ϕ and σ.

IV.B.3. Step 3: Bound on $\max_x \left[\partial_\phi \log Q(x|\phi)\right]_{\phi=\phi_0}$

In this step we bound $\left[\partial_\phi \log Q(x|\phi)\right]_{\phi=\phi_0}$ that appears in the numerator in the first order of the bias bound in Eq. (IV.32).

First, using notation introduced at the start of the section, we have

$$\left[\partial_\phi \log Q(x|\phi)\right]_{\phi=\phi_0} = \frac{q'_{\phi_0}(x)}{q(x|\phi_0)} - \frac{\mathcal{N}'}{\mathcal{N}}. \tag{IV.142}$$

Hence we can bound the numerator in Eq. (IV.32) as

$$\max_x |\partial_\phi \log Q(x|\phi)| \leq \max_x \left| \frac{q'_{\phi_0}(x)}{q(x|\phi_0)} \right| + \frac{|\mathcal{N}'|}{\mathcal{N}}. \tag{IV.143}$$

We can bound the second term in the above expression using Eq. (IV.140) and Eq. (IV.137) in Step IV.B.2 as

$$\frac{|\mathcal{N}'|}{\mathcal{N}} \leq \frac{1}{\sigma} \frac{g_1\left(\frac{|\mathcal{D}|}{6\sigma}\right)}{\frac{1}{2}\left(\mathrm{erf}(\frac{5}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma}) + \mathrm{erf}(\frac{1}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma})\right)}. \tag{IV.144}$$

As this is monotonically decreasing, for $\sigma \leq |\mathcal{D}|/6$ we can bound this as

$$\frac{|\mathcal{N}'|}{\mathcal{N}} \leq \frac{|\mathcal{D}|}{\sigma^2}\frac{1}{6}\frac{g_1(1)}{\frac{1}{2}\left(\mathrm{erf}(\frac{5}{\sqrt{2}}) + \mathrm{erf}(\frac{1}{\sqrt{2}})\right)} \lesssim 0.05\frac{|\mathcal{D}|}{\sigma^2}. \tag{IV.145}$$

To bound the first term, inserting the definition of the Gaussian $g_\sigma(\phi)$ into q yields

$$\left| \frac{q'_{\phi_0}(x)}{q(x|\phi_0)} \right| = \frac{|x-\phi_0|}{\sigma^2} \frac{g_\sigma(x-\phi_0)}{g_\sigma(x-\phi_0) + \frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}} \leq \frac{|x-\phi_0|}{\sigma^2}. \tag{IV.146}$$

Using assumption (IV.131), we can bound this as

$$\max_x \left| \frac{q'_{\phi_0}(x)}{q(x|\phi_0)} \right| \leq \frac{\max_x |x-\phi_0|}{\sigma^2} \leq \frac{5|\mathcal{D}|}{6\sigma^2}, \tag{IV.147}$$

leading to a final bound

$$\left| \max_x \left[\partial_\phi \log Q(x|\phi)\right]_{\phi=\phi_0} \right| \leq |\mathcal{D}|\sigma^{-2}. \tag{IV.148}$$

IV.B.4. Step 4: $\Omega(\sigma^{-2})$ bound on the Fisher Information $\mathcal{I}_0$

Next, we can calculate the Fisher information of the filtered distribution, $\mathcal{I}_0$ (see Eq. (IV.34)):

$$\mathcal{I}_0 = \int_{\mathcal{D}} \mathrm{d}x \;\; Q(x|\phi_0) \left[\partial_\phi \log Q(x|\phi)\right]^2_{\phi=\phi_0} \tag{IV.149}$$

$$= \int_{\mathcal{D}} \mathrm{d}x \; \frac{q(x|\phi_0)}{\mathcal{N}} \left(\frac{q'_{\phi_0}(x)}{q(x|\phi_0)} - \frac{\mathcal{N}'}{\mathcal{N}} \right)^2 \tag{IV.150}$$

$$= \int_{\mathcal{D}} \mathrm{d}x \left(\frac{q'_{\phi_0}(x)^2}{\mathcal{N} q(x|\phi_0)} - 2\frac{\mathcal{N}'}{\mathcal{N}^2} q'_{\phi_0}(x) + \frac{\mathcal{N}'^2}{\mathcal{N}^3} q(x|\phi_0) \right) \tag{IV.151}$$

$$= \frac{1}{\mathcal{N}} \left(\int_{\mathcal{D}} \mathrm{d}x \frac{q'_{\phi_0}(x)^2}{q(x|\phi_0)} - 2\frac{\mathcal{N}'^2}{\mathcal{N}} + \frac{\mathcal{N}'^2}{\mathcal{N}} \right) \tag{IV.152}$$

$$= \frac{1}{\mathcal{N}} \left(\int_{\mathcal{D}} \mathrm{d}x \frac{q'_{\phi_0}(x)^2}{q(x|\phi_0)} - \frac{\mathcal{N}'^2}{\mathcal{N}} \right). \tag{IV.153}$$

For the Gaussian model, we can bound the boundary term using Eq. (IV.140) and Eq. (IV.137) in Step IV.B.2 as

$$\frac{\mathcal{N}'^2}{\mathcal{N}} \leq \frac{F a_0}{\sigma^2} \frac{g_1^2\left(\frac{|\mathcal{D}|}{6\sigma}\right)}{\frac{1}{2}\left(\mathrm{erf}(\frac{5}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma}) + \mathrm{erf}(\frac{1}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma})\right)} \tag{IV.154}$$

and rewrite the integral in the first term as

$$\int_{\mathcal{D}} \mathrm{d}x \frac{q'_{\phi_0}(x)^2}{q(x|\phi_0)} = \frac{F^2 a_0^2}{\sigma^4} \int_{\phi_{guess}-|\mathcal{D}|/2}^{\phi_{guess}+|\mathcal{D}|/2} \frac{g_\sigma^2(x-\phi_0)(x-\phi_0)^2 \mathrm{d}x}{F a_0 g_\sigma(x-\phi_0) + (1-F)\frac{\mathcal{M}_\sigma}{2\pi}} \tag{IV.155}$$

$$= \frac{F a_0}{\sigma^4} \int_{\phi_{guess}-|\mathcal{D}|/2}^{\phi_{guess}+|\mathcal{D}|/2} \frac{g_\sigma(x-\phi_0)(x-\phi_0)^2 \mathrm{d}x}{1 + \frac{1-F}{F a_0}\frac{\mathcal{M}_\sigma}{2\pi}\frac{1}{g_\sigma(x-\phi_0)}} \tag{IV.156}$$

$$= \frac{F a_0}{\sigma^2} \int_{\frac{\phi_{guess}-\phi_0-|\mathcal{D}|/2}{\sigma}}^{\frac{\phi_{guess}-\phi_0+|\mathcal{D}|/2}{\sigma}} \frac{g_1(\xi)\xi^2 \mathrm{d}\xi}{1 + \sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{F a_0}\frac{1}{g_1(\xi)}}, \tag{IV.157}$$

To bound $\mathcal{I}_0$, notice that assumption in Eq. (IV.131) implies that either $[\phi_0 - |\mathcal{D}|/6, \phi_0 + |\mathcal{D}|/2] \subset \mathcal{D}$ or $[\phi_0 - |\mathcal{D}|/2, \phi_0 + |\mathcal{D}|/6] \subset \mathcal{D}$. Since the integrand in Eq. (IV.155) is always positive, we can bound the integral by an integral over a subset of $\mathcal{D}$,

$$\int_{\mathcal{D}} \mathrm{d}x \frac{q'_{\phi_0}(x)^2}{q(x|\phi_0)} \geq \frac{F a_0}{\sigma^2} \int_{-\frac{|\mathcal{D}|}{6\sigma}}^{\frac{|\mathcal{D}|}{2\sigma}} \frac{g_1(\xi)\xi^2 d\xi}{1 + \sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{F a_0}\frac{1}{g_1(\xi)}}. \tag{IV.158}$$

This yields a lower bound on the Fisher information:

$$\mathcal{N}\mathcal{I}_0 \geq \frac{Fa_0}{\sigma^2}\left[\int_{-\frac{|\mathcal{D}|}{6\sigma}}^{\frac{|\mathcal{D}|}{2\sigma}} \frac{g_1(\xi)\xi^2 d\xi}{1+\sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}\frac{1}{g_1(\xi)}} - \frac{g_1^2\left(\frac{|\mathcal{D}|}{6\sigma}\right)}{\frac{1}{2}\left(\mathrm{erf}(\frac{5}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma})+\mathrm{erf}(\frac{1}{6\sqrt{2}}\frac{|\mathcal{D}|}{\sigma})\right)}\right]. \tag{IV.159}$$

IV.B.5. Step 5: Upper bound on $(\mathcal{I}_0)^{-1}$

In order to use the lower bound in Eq. (IV.159) to upper bound $(\mathcal{I}_0)^{-1}$, we need to make sure that the RHS is positive.

For $F = 1$, the RHS is positive for $\sigma < |D|/2$, and the term in brackets is monotonically increasing to $\int_{-\infty}^{\infty} g_1(\xi)\xi^2 d\xi = 1$. For any $\sigma < |D|/6$ it is $\gtrsim 0.5$, so we get

$$\frac{1}{\mathcal{N}\mathcal{I}_0} \leq 2\frac{\sigma^2}{a_0} \quad \text{for } \sigma \leq |\mathcal{D}|/3, F = 1. \tag{IV.160}$$

For $F < 1$: assume $\sigma < |D|/6$, then we can again bound the integral by the integral of the subset $[-1,1] \subset [-\frac{|\mathcal{D}|}{6\sigma}, \frac{|\mathcal{D}|}{2\sigma}]$:

$$\int_{-\frac{|\mathcal{D}|}{6\sigma}}^{\frac{|\mathcal{D}|}{2\sigma}} \frac{g_1(\xi)\xi^2 d\xi}{1+\sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}\frac{1}{g_1(\xi)}} \geq \int_{-1}^{1} \frac{g_1(\xi)\xi^2 d\xi}{1+\sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}\frac{1}{g_1(\xi)}}. \tag{IV.161}$$

Within the subset, $|\xi| < 1$ and so $g_1(\xi) \geq g_1(1) = (2\pi e)^{-1/2}$, so we can further bound

$$\int_{-a/6}^{a/2} \frac{g_1(\xi)\xi^2 d\xi}{1+\sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}\frac{1}{g_1(\xi)}} \geq \int_{-1}^{1} \frac{g_1(\xi)\xi^2 d\xi}{1+\sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}\frac{1}{g_1(1)}} \tag{IV.162}$$

$$= \frac{\int_{-1}^{1} g_1(\xi)\xi^2 d\xi}{1+\sigma\frac{\mathcal{M}_\sigma}{2\pi}\frac{1-F}{Fa_0}\frac{1}{g_1(1)}} \tag{IV.163}$$

We can also bound $\sigma\mathcal{M}_\sigma$ in the denominator as $\leq |\mathcal{D}|/6$, and finally we get

$$\int_{\mathcal{D}} \mathrm{d}x \frac{q'_{\phi_0}(x)^2}{q(x|\phi_0)} \geq \frac{Fa_0}{\sigma^2}\frac{\int_{-1}^{1} g_1(\xi)\xi^2 d\xi}{1+\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\frac{1}{6g_1(1)}}. \tag{IV.164}$$

Inserting $\sigma < \mathcal{D}/6$ into the denominator of the boundary term yields

$$\mathcal{N}\mathcal{I}_0 \geq \frac{Fa_0}{\sigma^2}\left(\frac{\int_{-1}^{1} g_1(\xi)\xi^2 d\xi}{1+\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\frac{1}{6g_1(1)}} - \frac{g_1^2\left(\frac{|\mathcal{D}|}{6\sigma}\right)}{\frac{1}{2}\left(\mathrm{erf}(\frac{5}{\sqrt{2}})+\mathrm{erf}(\frac{1}{\sqrt{2}})\right)}\right). \tag{IV.165}$$

To ensure the RHS of the inequality above is positive, we take σ such that the negative term is at most half of the positive term,

$$\sigma \leq \frac{|\mathcal{D}|}{6}\log^{-1/2}\left(\frac{1+\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\frac{1}{6g_1(1)}}{\frac{\pi}{2}\left(\mathrm{erf}(\frac{5}{\sqrt{2}})+\mathrm{erf}(\frac{1}{\sqrt{2}})\right)\int_{-1}^{1} g_1(\xi)\xi^2 d\xi}\right). \tag{IV.166}$$

We get the final bound

$$\frac{1}{\mathcal{N}\mathcal{I}_0} \leq C_1\frac{\sigma^2}{Fa_0}\left(1+C_2\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\right) \text{ for } \sigma \leq \frac{|\mathcal{D}|}{6}\frac{1}{\sqrt{C_3+\log\left(1+C_2\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\right)}} \tag{IV.167}$$

with constants

$$C_1 = 10.1 \gtrsim \frac{1}{\frac{1}{2}\int_{-1}^{1} g_1(\xi)\xi^2 d\xi}, \tag{IV.168}$$

$$C_2 = 0.7 \gtrsim \frac{1}{6g_1(1)}, \tag{IV.169}$$

$$C_3 = 1.2 \gtrsim -\log\left(\frac{\pi}{2}\left(\mathrm{erf}(\frac{5}{\sqrt{2}})+\mathrm{erf}(\frac{1}{\sqrt{2}})\right)\int_{-1}^{1} g_1(\xi)\xi^2 d\xi\right). \tag{IV.170}$$

IV.B.6. Step 6: Final bound on the bias and variance

Now we are ready to combine all the steps above to get the bound on the bias. The first order in Eq. (IV.32) can first be bound using the bound on $\|h\|$ in Eq. (IV.130) from step IV.B.1 and the bound on $|\max_x[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}|$ in Eq. (IV.148) in step IV.B.3

$$\|h\|\frac{|\max_x[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}|}{\mathcal{I}_0} \leq 2F(1-a_0)\frac{|\mathcal{D}|}{d}\sigma^2 g_\sigma(d)\frac{1}{\mathcal{N}\mathcal{I}_0\sigma^2}. \tag{IV.171}$$

Then we can bound $(\mathcal{N}\mathcal{I}_0)^{-1}$ using Eq. (IV.160) and Eq. (IV.167) leading to the bounds

$$\|h\|\frac{|\max_x[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}|}{\mathcal{I}_0} \leq \frac{1-a_0}{a_0}\frac{|\mathcal{D}|}{d}\sigma^2 g_\sigma(d) \times 4 \qquad F=1, \tag{IV.172}$$

$$\|h\|\frac{|\max_x[\partial_\phi \log Q(x|\phi)]_{\phi=\phi_0}|}{\mathcal{I}_0} \leq \frac{1-a_0}{a_0}\frac{|\mathcal{D}|}{d}\sigma^2 g_\sigma(d) \times 2C_1\left(1+C_2\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\right) \qquad F<1. \tag{IV.173}$$

To get the variance bound in Lemmas IV.7 and IV.8, we combine Eq. (IV.160) and Eq. (IV.167) with Eq. (IV.141) to get

$$\frac{1}{\mathcal{I}_0} \leq 2\sigma^2 \qquad F=1, \tag{IV.174}$$

$$\frac{1}{\mathcal{I}_0} \leq C_1\sigma^2\left(1+C_2\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\right)\left(1+\frac{1-F}{Fa_0}\frac{1}{2\pi}\right) \qquad F<1. \tag{IV.175}$$

To bound the number of shots needed $M' = M/P_A$ in Theorem IV.7, it is useful to also have

$$\frac{1}{P_A\mathcal{I}_0} = \frac{\mathcal{M}_\sigma}{\mathcal{N}\mathcal{I}_0} \leq \frac{1}{\mathcal{N}\mathcal{I}_0} \leq C_1\frac{\sigma^2}{Fa_0}\left(1+C_2\frac{|\mathcal{D}|}{2\pi}\frac{1-F}{Fa_0}\right), \tag{IV.176}$$

as the variance is

$$\mathrm{Var}[\tilde{\phi}] \approx \frac{1}{M\mathcal{I}_0} = \frac{1}{M'P_A\mathcal{I}_0}. \tag{IV.177}$$

V. The advantage of quantum control in many-body Hamiltonian learning

The characterization of unknown systems is a critical topic in science and engineering. Quantum mechanical systems are governed by a Hamiltonian that determines the evolution of the system in time. In this setting, system characterization takes the form of learning parameters of the Hamiltonian from experimental data [123, 260–301]. Hamiltonian learning is of central interest for both applications of quantum technologies and explorations of quantum systems in nature; these include quantum metrology [260–263], molecular structure identification [300, 302, 303] quantum device benchmarking [270–272], and verification of quantum simulations [273, 274, 304]. A universal objective is to develop learning strategies that consume as few resources (for example, as little time) as possible. This optimization requires designing both a set of experiments and an algorithm to infer the Hamiltonian from the experimental data.

A central goal of Hamiltonian learning is the so-called Heisenberg limit. Arising from the quantum Fisher information [305–307], the Heisenberg limit stipulates that the error, ϵ, of any estimation of a Hamiltonian parameter scales at best with the inverse of the total experimental runtime, T [308]. Protocols that achieve the Heisenberg limit, such as Ramsey spectroscopy [309], gate-set tomography [310], and Floquet calibration [311], have become standard procedures across quantum technologies. Theoretical progress has laid rigorous foundations for Heisenberg-limited learning in single- and few-qubit systems, including theoretical bounds on the learnability and optimal measurement schemes in noiseless [123, 260–268] and noisy [126, 312] situations. These works tie in closely with work on Heisenberg-limited phase estimation [57, 74, 75, 131], and Heisenberg-limited unitary estimation [313].

Despite this progress, achieving Heisenberg-limited learning of *many-body* Hamiltonians remains a largely open direction. The past decade has seen an explosion of learning protocols for many-body Hamiltonians that utilize full [261, 267, 268, 275–286] or restricted [260, 272, 273, 291–294] tomographic access, access to eigenstates [280, 294–296] or thermal states [268, 282, 287–

289, 291] of the unknown Hamiltonian, or the ability to copy experimental quantum data to a trusted quantum register [269–271]. In most of these works, the estimation error scales as the inverse square root of the total evolution time, known as the standard quantum limit or shot noise limit, and the Heisenberg limit has neither been achieved nor proven impossible. Interestingly, the only work to achieve the Heisenberg limit thus far, Ref. [286], requires a high degree of *control* over the system to be learned. Specifically, the algorithm in Ref. [286] resembles dynamical decoupling, and requires interleaving ever-smaller steps of time evolution with large single-qubit rotations. (See also earlier related algorithms [275, 290].) By contrast, the standard quantum limit can be achieved with only the ability to prepare product states, perform time evolution, and measure in a product basis [289]. It is natural, and practical, to wonder whether a large degree of control is necessary to achieve Heisenberg-limited learning in the many-body setting.

In this work, we establish a rigorous separation in many-body Hamiltonian learning between systems with and without quantum control. We consider two forms of quantum control: continuous control, in which one can continuously time-evolve under a Hamiltonian that is the sum of both unknown terms and controlled known terms, and discrete control, in which one can interleave time-evolution under the unknown Hamiltonian with discrete quantum gates. We begin by providing a new algorithm for learning many-body Hamiltonians, which uses continuous quantum control to achieve Heisenberg-limited learning of any bounded-degree many-body Hamiltonian. Our algorithm calls an algorithm by Haah, Kothari and Tang (henceforth the "HKT algorithm") [289] as a subroutine, and augments this with quantum controls that simulate reversed time-evolution under a best estimate of the unknown Hamiltonian. The estimate is updated adaptively as the algorithm proceeds, leveraging techniques for phase estimation introduced in Ref. [74].

Complementary to this algorithm, we show that learning at the Heisenberg limit is *not* possible in the absence of quantum control for large classes of many-body Hamiltonians. We first show that, if there exist infinitesimal translations in parameter space that do not change the spectrum of the Hamiltonian, then the Hamiltonian parameters cannot be learned at the Heisenberg limit without quantum control. This result applies to several common classes of Hamiltonians, including qubits or fermions with local or non-local interactions, and even a single qubit evolving under a field along an unknown axis. We then provide an analogous no-go result for any many-body Hamiltonian that satisfies the eigenstate thermalization hypothesis (ETH), a widely validated conjecture for how closed many-body systems relax to effectively thermal states [314–317]. Both of our no-go theorems apply quite generally, and preclude any algorithm without sufficient quantum control, including adaptive

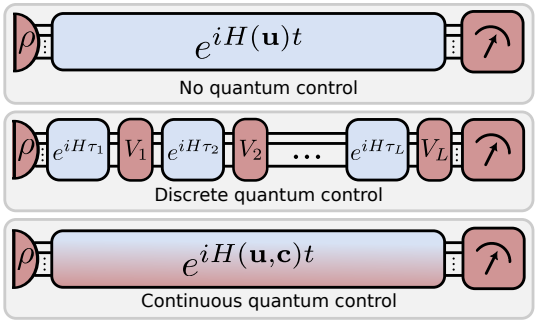


Figure V.1.: Schematic quantum circuit diagrams of the three experimental models that we consider in this work. Colors denote when the experimentalist has complete control (red), some control (light blue-red gradient), or no control (light blue) over the operation. We assume the experimentalist has the freedom to specify state preparation (half-ellipse) and measurement (dial symbol). Further, during time evolution (rectangles) we assume the freedom to specify the evolution times t/τ_l, and the instantaneous quantum operations V_l (in the discrete quantum control model), or the control parameters $\mathbf{c}$ (in the continuous quantum control model). In the discrete control model, we assume the quantum operations V_l are unitary but allow them to act on an arbitrary number of ancilla qubits (not shown), and we hide the dependence of H on the unknown parameters $\mathbf{u}$ for clarity.

algorithms, algorithms with access to a quantum memory, and algorithms that involve arbitrarily complex quantum or classical operations that do not involve the unknown Hamiltonian. This generality arises naturally because our proofs utilize bounds on the quantum Fisher information [305–307] and related quantities that we define for unitary quantum processes.

V.1. Background and problem definition

We consider the problem of learning a Hamiltonian,

$$H(\mathbf{u}) = \sum_{a=1}^{N_p} u_a P_a, \qquad |u_a| \leq 1 \tag{V.1}$$

where u_a are unknown parameters and P_a are k-local Pauli operators on N qubits, with $k = \mathcal{O}(1)$. We abbreviate $\mathbf{u} = (u_1, \ldots, u_{N_p})$. Our goal is to learn the parameters $\mathbf{u}$ from a system with dynamics governed by $H(\mathbf{u})$

by performing experiments on the aforementioned system. Each experiment consists of state preparation, time-evolution, and measurement. We assume that the time-evolution features some form of 'black-box access' to $H(\mathbf{u})$ that we describe below.

We consider three experimental settings (see Fig. V.1), which are distinguished by the presence of 'quantum control': the ability to manipulate the quantum state of the unknown system during time evolution. (These settings do not change our abilities regarding system preparation or measurement. In each setting, we consider either product state preparation and computational basis measurement – for our learning algorithm, Theorem V.1, or arbitrary state preparation and measurement – for our lower bounds, Theorems V.2, V.3.) First, we consider learning with 'no quantum control' (top, Fig. V.1), in which all experiments consist only of state preparation, a single instance of time evolution under the unknown Hamiltonian, $U = e^{-iH(\mathbf{u})t}$, and measurement. Unless otherwise stated, we assume the experimentalist can perform arbitrary state preparations and measurements, and evolve for an arbitrary time t (at a cost given in the next paragraph). Second, we consider experiments with 'discrete quantum control' (middle, Fig. V.1), in which time evolution may be interleaved with instantaneous quantum operations by the experimentalist (i.e. operations during which $H(\mathbf{u})$ does not act). This definition is appropriate for hybrid analog-digital quantum platforms, and for applications of Hamiltonian learning to verify digital quantum simulations [273, 274, 304]. In practice, the degree of discrete control may be limited (for example, by energetic considerations, pulse discretization, or noise), so we will set an upper bound, L, on the number of instantaneous operations in a single experiment [1]. Finally, we consider experiments with 'continuous quantum control' (bottom, Fig. V.1), in which control terms are added to the Hamiltonian itself:

$$H(\mathbf{u}, \mathbf{c}) = \sum_{a=1}^{N_p} (u_a + c_a) P_a, \qquad |u_a|, |c_a| \leq 1. \tag{V.2}$$

Compared to discrete quantum control, this model reflects the fact that in physical quantum systems the experimental control frequently has a bounded

[1] Let us elaborate on this point briefly. In many settings of interest, the Hamiltonian to be learned will be native to the system of interest, and it may not be possible to instantaneously halt the Hamiltonian during the implementation of a discrete quantum gate. Thus, the implementation of a digital gate might incur some error (e.g. if one attempts to implement the gate when the native Hamiltonian is still "on") or require some additional time (e.g. to transition the system between Hamiltonian and digital modes of operation). To capture these effects within a simple theoretical model, we assume that the total number of digital gates is bounded by some number L, and study the dependence of learning on L.

strength that renders instantaneous operations impossible. For simplicity, we assume the control parameters c_a are time-independent within a single experiment, but can be reset by the experimentalist between different experiments. In practice, experiments may also be able to modify c_a within the timescale of a single experiment. However, this limited degree of control will already be sufficient for our learning algorithm in Theorem V.1.

To quantify the performance of a Hamiltonian learning algorithm, we now define a cost model and a metric for success. The cost of generating data is taken to be the total experiment time

$$T = \sum_x t_x, \tag{V.3}$$

where x labels an individual experiment, and t_x is the duration of time-evolution under the unknown Hamiltonian in the experiment. This neglects any cost of state preparation and measurement. We quantify the accuracy of an estimator $\tilde{\mathbf{u}}$ of the unknown parameters $\mathbf{u}$ by requiring that the maximum root-mean-square (RMS) error be bounded as

$$\epsilon \leq \max_a \left[\overline{(\tilde{u}_a - u_a)^2}\right]^{1/2}, \tag{V.4}$$

where the overline denotes an expectation value over experimental outcomes.

The Heisenberg limit provides a universal bound, $T = \Omega(\epsilon^{-1})$, on the total experiment time T required to achieve a maximum RMS error ϵ [131, 262, 308]. (We provide a derivation of this via the quantum Cramer-Rao bound in Appendix V.A.3.) The bound passes through to other error metrics up to factors of $\sqrt{N_p}$ using a median-of-means approach (see Appendix V.A.2). The Heisenberg limit applies to any learning model, and does not prohibit stricter bounds being established in specific settings.

V.2. Results

In this work, we establish a gap between the learnability of Hamiltonians with and without quantum control. We present our results informally in this section, and establish rigorous statements and proofs in the Appendix.

Our first result is an algorithm to learn any local many-body Hamiltonian at the Heisenberg limit using continuous quantum control (Appendix V.B). Our algorithm builds off the the HKT algorithm introduced in Ref. [289], which learns a many-body Hamiltonian at the *standard quantum limit*, $T = \Omega(\epsilon^{-2})$, from experiments involving no quantum control. The HKT algorithm is standard quantum limited because each experiment involves evolution only

up to a maximum time t_c, which is constant in ϵ (see Appendix V.B.1 for details). This restriction is fundamental to the HKT algorithm, which involves approximating time-evolved operators by their Taylor series, $e^{iHt}P_a e^{-iHt} \approx P_a + it[H, P_a] + \mathcal{O}(t^2)$, which diverges at large t.

We surpass the standard quantum limit by introducing adaptive quantum controls to the HKT algorithm, as illustrated by the flowchart in Fig. V.2. Suppose that previous experiments allow us to form an estimate $\tilde{\mathbf{u}}$ of $\mathbf{u}$ with error less than ε. We then set our continuous quantum control parameters to the negative of the estimated Hamiltonian, $\mathbf{c} = -\tilde{\mathbf{u}}$. This yields time-evolution under the Hamiltonian,

$$H(\mathbf{u}, \mathbf{c}) = \varepsilon H(\mathbf{u}'), \qquad H(\mathbf{u}') = \sum_a u'_a P_a, \tag{V.5}$$

$$u'_a = (u_a - \tilde{u}_a)/\varepsilon, \tag{V.6}$$

where the new parameters obey $|u'_a| \leq 1$. We can now apply the HKT algorithm to the rescaled Hamiltonian H', by evolving under the original Hamiltonian H for a time t_c/ε. This improves the error in our estimate of $\mathbf{u}$ by a constant factor. Repeating this procedure in a similar fashion to protocols for Heisenberg-limited phase estimation [74, 75] yields our result:

Theorem V.1 (informal)
Consider a low-intersection Hamiltonian $H(\mathbf{u})$ with unknown parameters $\mathbf{u}$ and continuous quantum control as in Eq. (V.2). There exists an algorithm to learn $\mathbf{u}$ within maximum RMS error ϵ using a total experimental time $T = \mathcal{O}(\epsilon^{-1})$ and only product states and computational basis measurements.

(The term 'low-intersection' is a technical constraint inherited from the HKT algorithm that we describe in Appendix V.B.1.) Our algorithm bears some resemblance to that of Ref. [285]; however, by using the HKT algorithm as an inner loop, we improve the error bound exponentially in N and remove the need for entanglement with a large ancilla register.

Our remaining results place lower bounds, $T = \Omega(\epsilon^{-2})$, on Hamiltonian learning without control. Our first bound is particularly simple: suppose two Hamiltonians $H(\mathbf{u})$ and $H(\mathbf{u} + \delta\mathbf{u})$ are equal up to a unitary transformation. Then, they cannot be distinguished at the Heisenberg limit for $\epsilon \approx \|\delta\mathbf{u}\|_2$ without quantum control (Appendix V.C). This is based on a straightforward calculation from the quantum Fisher information which shows that the eigenvalues of $H(\mathbf{u})$ can be learned at the Heisenberg limit, but the eigenvectors cannot. Our Hamiltonian learning problem does not target the learning of eigenvalues or eigenvectors, but rather the unknown parameters $\mathbf{u}$. However, we find that if there exists a shift in the unknown parameters that leaves

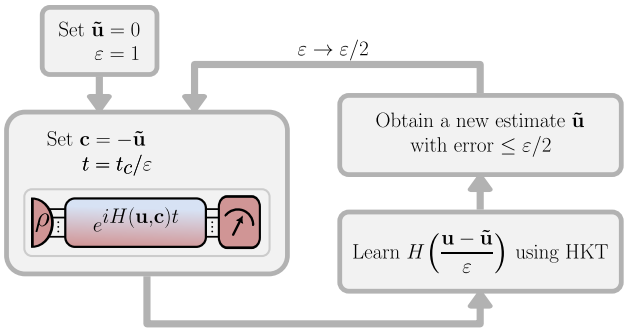


Figure V.2.: Flowchart of our Heisenberg-limited algorithm. The algorithm starts with an initial guess $\tilde{\mathbf{u}}$ for the unknown Hamiltonian parameters $\mathbf{u}$ accurate up to $\varepsilon = 1$, and adaptively adjusts the control parameters to refine the estimate. The HKT algorithm of Ref. [289] is used to learn a rescaled Hamiltonian (Eq. (V.6)) and then the estimate of $\mathbf{u}$ is updated based on the learned Hamiltonian. The algorithm continues iteratively until the error ε in the estimate is smaller than a specified threshold ϵ.

the eigenvalues unchanged, then the parameters cannot be learned at the Heisenberg limit:

Theorem V.2 (informal)
Consider a Hamiltonian $H(\mathbf{u})$ with unknown parameters $\mathbf{u}$, and either no quantum control or discrete quantum control with a constant number $L = \mathcal{O}(1)$ of interleaved operations. Suppose that there exists an infinitesimal transformation, $\mathbf{u} \to \mathbf{u} + \delta\mathbf{u}$, that leaves the spectrum of the Hamiltonian unchanged to leading order in $\delta\mathbf{u}$. Then learning the linear combination of parameters $\mathbf{u} \cdot (\delta\mathbf{u}/\|\delta\mathbf{u}\|)$ to within RMS error ϵ requires a total experimental time $T = \Omega(\epsilon^{-2})$

Theorem V.2 is completely general, in that it holds for arbitrary state preparation and measurement in the experiment. As stated, the theorem concerns the RMS error of a linear combination of parameters specified by $\delta\mathbf{u}$. If $\delta\mathbf{u}$ has support on only a single parameter, then this trivially lower bounds our original choice of error metric, the maximum RMS error over the individual parameters u_a [Eq. (V.4)]. More generally, the error in any linear combination lower bounds the maximum RMS error up to a factor of $\sqrt{1/N_p}$.

Hamiltonians that obey the conditions of Theorem V.2 are quite common in the natural world. For instance the single-qubit Hamiltonian,

$$H(\mathbf{u}) = u_x X + u_y Y + u_z Z, \tag{V.7}$$

has the same spectrum whenever $\|\mathbf{u}_1\|_2 = \|\mathbf{u}_2\|_2$. The angle of $\mathbf{u}$ can therefore not be learned at the Heisenberg limit. This point was established by other methods in Refs. [260, 266]; Theorem V.2 can be thus seen as a generalization of the single-qubit no-go theorems in these works to the general many-body case. Similarly, the fermionic Hamiltonian,

$$H(\mathbf{u}) = \sum_{ij} u^{(1)}_{ij} c^\dagger_i c_j + \sum_{ijkl} u^{(2)}_{ijkl} c^\dagger_i c^\dagger_j c_k c_l, \tag{V.8}$$

possesses orbits of $\mathbf{u}$ connected by Givens rotations [318]. This holds for both local and all-to-all connectivity, since Givens rotations may be chosen to preserve locality. Analogous spin Hamiltonians are discussed in Appendix V.C.

While the applicability of Theorem V.2 is broad, the conditions of the theorem can be circumvented by adding only a small amount of continuous quantum control or prior knowledge about the Hamiltonian (for example, by adding fixed non-adaptive background fields [123, 265, 266, 311]). Our second lower bound excludes Heisenberg-limited learning on a more robust physical context. Namely, we show that learning is bounded by the standard quantum limit, $T = \Omega(\epsilon^{-2})$, for any Hamiltonian that thermalizes via the eigenstate thermalization hypothesis (ETH) [314–317] (Appendix V.D). Thermalization describes the relaxation in time of local observables to their thermal values [316]. The connection between learning without quantum control and thermalization is natural, since both concern the late time dynamics of the unitary evolution $e^{-iH(\mathbf{u})t}$. Thermalization clearly precludes learning at the Heisenberg limit from local measurements, since they approach thermal distributions that are constant in time. However, it is not a priori clear whether this extends to more complex measurement strategies, such as those involving backwards time-evolution under an estimate of the Hamiltonian [270, 271]. Indeed, even thermalizing Hamiltonians exhibit coherent late time phenomena in their own eigenbasis, which suggests that some degree of Heisenberg limited learning may be possible.

We resolve these questions by invoking the ETH, a conjecture for how thermalization occurs in closed Hamiltonian systems. The ETH concerns the matrix elements of local operators in the Hamiltonian eigenbasis. It posits that (i) the diagonal matrix elements are equal to the expectation value of the operator in a corresponding thermal state, and (ii) the off-diagonal matrix elements have magnitude exponentially small in the system size, and are random up to certain macroscopic constraints [314–317]. We formulate these conditions precisely in Appendix V.D.1. The core tenets of the ETH have been thoroughly validated by extensive numerical studies across many Hamiltonians; see Ref. [316] for a comprehensive review. Taking the validity of ETH as an assumption, we establish the following theorem:

Theorem V.3 (informal)
Consider a Hamiltonian $H(\mathbf{u})$ with an extensive number, $N_p = \Omega(N)$, of unknown parameters $\mathbf{u}$, and either no quantum control or discrete quantum control with a constant number, $L = \mathcal{O}(1)$, of interleaved operations. Suppose that, within the states accessed by an experiment, the Hamiltonian obeys the ETH and that its thermal expectation values have bounded derivatives with respect to the temperature. Then learning the parameters $\mathbf{u}$ with maximum RMS error ϵ requires a total experimental time, $T = \Omega(\epsilon^{-2})$, as long as the time is sub-exponential in the system size, $T = o(exp(N))$.

The theorem holds for arbitrary state preparation and measurement, so long as the Hamiltonian is thermalizing on the states accessed during the experiment. The required condition on derivatives of thermal expectation values is quite loose, and is a standard feature of the thermodynamic limit. (In Appendix V.D.6, we prove it holds for geometrically local Hamiltonians whenever correlations decay exponentially [319, 320].) The restriction to sub-exponential times is natural, since at later times the system may exhibit non-thermal revivals that are not captured by the ETH.

Our proof of Theorem V.3 relies on several technical lemmas, which may be of interest for future work on multi-parameter Hamiltonian learning or rigorous analysis of thermalizing Hamiltonians. We begin by expressing the quantum Fisher information in terms of time-integrals over operators entering the Hamiltonian, which are then amenable to analysis via the ETH. We establish that the spectral norm of the contribution of off-diagonal matrix elements of these operators is bounded above by $\mathcal{O}(t^{1/2})$, and hence cannot lead to learning at the Heisenberg limit. Our proof leverages techniques from random matrix theory. Meanwhile, the diagonal matrix elements grow instead as $\mathcal{O}(t)$. However, we show that their contribution to the Fisher information matrix is low-rank when the derivatives of thermal expectation values are bounded. Intriguingly, this rank-deficiency is only a limitation for *many*-parameter learning. In keeping with our initial intuition, we find that Heisenberg-limited learning is possible for many-body Hamiltonians where only a constant number of parameters are unknown, using non-local measurement strategies (Appendix V.D.7).

V.3. Comparison to previous work

Models of quantum control: Prior studies on Hamiltonian learning have considered a diverse range of experimental scenarios, almost all of which are encapsulated by one of our three models of quantum control. Many works fall under our 'no quantum control' model [261–263, 267, 269, 270, 273, 275–278, 281, 283, 284, 287, 289, 291–293, 300, 301, 321], some of which include

additional, experimentally-motivated constraints that are not imposed in our work [273, 276, 281, 293]. Notably, our no control model encompasses previous proposals for "quantum-assisted" Hamiltonian learning [269–271, 282, 301], which nonetheless use only a single application of time-evolution under the unknown Hamiltonian. A comparatively fewer number of works fall under our discrete quantum control model [123, 266, 285, 286, 297]; these are referred to as "quantum-enhanced" experiments in Ref. [286]. Meanwhile, our continuous quantum control model encapsulates a number of additional works that use both time-independent [265, 272, 285, 299] and time-dependent continuous control [279, 294, 297, 298, 322, 323]. Some works have allowed evolving an entangled state over multiple copies of the system [260, 285], which are encapsulated within our discrete control model by allowing swap operations between the the system and an ancillary register [124]. Finally, a number of other works have focused on learning from access to steady states of Hamiltonians (e.g. eigenstates or thermal states) [268, 280, 288, 289, 291, 294–296]; while in principle steady states might be obtained by using time-evolution under the unknown Hamiltonian, this is largely orthogonal to the focus of our work.

Many works have investigated learning Hamiltonians with a single [123, 263, 265, 266, 275, 297] or few [260–262, 267, 269, 272, 273, 278, 279, 285, 290, 292, 294, 294, 296, 298–300, 322] unknown parameters. Here, we also use "few" to refer to algorithms whose scaling with the number of unknown parameters is either exponential [278, 279, 285] or unknown [261, 269, 272, 273, 290, 292, 294, 294, 296, 298, 300, 322]. Several of these works have also considered the benefits of quantum control in Hamiltonian learning [123, 260, 262, 263, 265–267, 275, 297, 299]. For the simplest such Hamiltonian, with $H(u) = uP$ and u unknown, learning corresponds to the well-known phase estimation problem. In this case, the Heisenberg limit can be reached with no quantum control [131], and thus quantum control yields no asymptotic advantage [297]. However, beyond this, there are known examples of Hamiltonians with a single unknown parameter where learning without quantum control is standard quantum limited [123]. Previous works have proven that the Heisenberg limit can be achieved in these scenarios by leveraging discrete [123] or continuous [265] quantum control. Some of these algorithms leverage approximate backwards time-evolution [123, 265], in a similar manner to our many-body learning algorithm.

Rigorous algorithms for learning many-body Hamiltonians in polynomial time have been demonstrated in Refs. [268, 270, 276, 277, 280, 281, 283, 284, 287–289, 289, 291, 291, 293, 321], all with precision scaling as the standard quantum limit. For other algorithms [261] the Heisenberg limit is anticipated, but has not been proven. Among these, the algorithm of Ref. [289] is crucial to

the present manuscript, as we use it as a subroutine in our Heisenberg-limited learning algorithm.

The only prior work to achieve the Heisenberg limit in many-body Hamiltonian learning was Ref. [286]. Their algorithm interleaves large single-qubit rotations in between short steps of time-evolution, in a manner similar to dynamical decoupling. In particular, the rate at which single-qubit gates are applied becomes ever more frequent with the desired precision ϵ, such that the total algorithm requires a number of interleaved operations $L = \Theta(\epsilon^{-2})$. Interestingly, this is quadratically worse than our lower bound, $L = \Omega(\epsilon^{-1})$. It remains an open question whether the algorithm in Ref. [286] can be strengthened to match this lower bound. In terms of experimental feasibility, for small ϵ the frequent interleaved operations in Ref. [286] may be difficult to apply in practice (depending, for example, on the power of single-qubit pulses relative to the strength of the unknown Hamiltonian). In comparison, our algorithm only uses control operations of the same strength as the unknown Hamiltonian. At the same time, these controls must be able to implement backwards time-evolution under a best estimate of the unknown Hamiltonian, which will require two-qubit interactions or gates. Thus, where each algorithm succeeds in practice will depend on the specific control set of the system of interest.

V.4. Discussion

The above results establish an advantage in many-body Hamiltonian learning for both continuous and discrete quantum control. For the former, the algorithm in Theorem V.1 establishes that any k-local Hamiltonian can be learned at the Heisenberg limit with continuous quantum control. This circumvents the no-go Theorems V.2 and V.3, and establishes an advantage for continuous quantum control over the no quantum control model in learning Hamiltonians where either of these theorems apply. Moreover, the total *number* of experiments needed to implement our algorithm is only $\mathcal{O}(\log \epsilon^{-1})$. This is exponentially fewer than all known learning algorithms with no quantum control. At the same time, our algorithm requires the ability to individually modify every term in the Hamiltonian, a relatively strong requirement for physical experiments. A clear question for future research is whether this can be relaxed, for example to experiments with control over only single-qubit terms or global combinations of terms. This is relevant in the case of e.g. NMR [300, 302, 303] where one has only control over a global magnetic field.

For discrete quantum control, our no-go theorems combine with the algorithm introduced in Ref. [286] to establish a separation in Hamiltonian learning as a function of the number L of interleaved quantum operations.

Ref. [286] establishes a Heisenberg-limited algorithm in the discrete quantum control setting that implements a dynamical decoupling scheme requiring $L = \Theta(T^2)$ interleaved unitaries. Meanwhile, our formal versions of Theorems V.2 and V.3 establish that L must scale at least linearly with the evolution time, $L = \Omega(T)$ (see Appendix V.C and V.D). Given recent work establishing a Heisenberg-limited unitary learning scheme [313], we expect that Heisenberg-limited Hamiltonian learning in the discrete quantum control model with $L = \Theta(T)$ can be achieved, making the bound tight. This would require repeatedly alternating forward time-evolution by e^{iHdt} and backward time evolution by $e^{-iH(\bar{u})dt}$ for dt constant in ϵ. This follows a different principle to Ref. [286] (term cancellation versus dynamical decoupling); these two principles correspond to the limits of optimal control for learning in the single-parameter case [265]. We think it an interesting open question whether a dynamical decoupling scheme with $L = \Theta(T)$ can be achieved; we speculate that this may indeed be possible by leveraging rigorous approximations for Floquet evolution with geometrically local Hamiltonians [324].

While each of our no-go theorems apply to Hamiltonians that are common throughout nature, Theorem V.3 is in some ways more robust than Theorem V.2. As we have remarked, even small modifications to the Hamiltonian, such as the addition of an external field of known value, can break the unitary equivalence necessary for Theorem V.2. However, Theorem V.3 cannot be so easily circumvented. For example, suppose that one has continuous quantum control, but chooses control parameters $\mathbf{c}$ such that each instance of the Hamiltonian $H(\mathbf{u}, \mathbf{c})$ satisfies the ETH. From the proof of Theorem V.3, one can show that $\Omega(N/\text{polylog}(N))$ unique choices of $\mathbf{c}$ will be required to learn at the Heisenberg limit. Our algorithm in Theorem V.1 circumvents this by choosing $\mathbf{c}$ such that the system does not thermalize until ever-later times in each successive experiment. We conjecture that similar methods might be more broadly required: that efficiently learning many-body Hamiltonians at the Heisenberg limit requires breaking, or delaying, ergodicity. Interestingly, in Appendix V.D.7 we provide strong arguments that Heisenberg-limit learning *can* be achieved in thermalizing Hamiltonians when only a constant $\mathcal{O}(1)$ number of parameters are unknown, using the quantum-assisted Hamiltonian learning algorithms of Refs. [269–271].

Interestingly, we can in fact understand our Theorem V.3 for thermalizing Hamiltonians in terms of Theorem V.2. This comes from an observation within the proof of Theorem V.3 that there exist extensively many linear combinations $\sum_a h_a P_a$ of local operators whose matrix elements are purely off-diagonal in the eigenbasis of the Hamiltonian. This implies that the spectrum of the Hamiltonian is invariant (to first order in perturbation theory) under the parameter shift, $H \to H + \lambda \sum_a h_a P_a$, which falls precisely under the

conditions of Theorem V.3. (To see this, recall that from perturbation theory the change in an eigenenergy E_j is given by $\lambda\langle E_j|\sum_a h_a P_a|E_j\rangle + \mathcal{O}(\lambda^2)$, which vanishes to first order when $\sum_a h_a P_a$ is off-diagonal.) The fact that the spectrum of a Hamiltonian obeying the ETH is independent of so many microscopic parameters is a stark illustration of the averaging effects of thermalization.

In summary, we have established separations in Hamiltonian learning between experiments that can control a system throughout time-evolution, and those that cannot. Our work begins with a rigorous formulation of three models of quantum control: no control, discrete, and continuous. These definitions encapsulate nearly all previous algorithms for Hamiltonian learning. Utilizing continuous quantum control, we introduce a new algorithm for learning any many-body Hamiltonian at the Heisenberg limit, which combines techniques from robust phase estimation [74] with the HKT algorithm [289]. Our algorithm joins with that of Ref. [286] among the first provable algorithms for learning many-body Hamiltonians at the Heisenberg limit. Complementary to this, we establish that the Heisenberg limit is impossible to achieve in general without quantum control. We identify two roadblocks—unitary equivalences between Hamiltonians, and thermalization via the eigenstate thermalization hypothesis—and show that either of these preclude nearly all learning strategies that do not interrupt time-evolution at, at least, a constant rate. We hope that these results provide a new lens through which to view existing Hamiltonian learning algorithms, and spark further refinements in algorithms to come.

V.A. Preliminaries

We begin in Appendix V.A.1 by providing a formal definition of a quantum experiment, quantum control, and the many-body Hamiltonian learning problem. We discuss the relations between various error bounds on Hamiltonian learning in Appendix V.A.2. In Appendix V.A.3, we establish several basic facts regarding Hamiltonian learning experiments, related to the Fisher information.

Throughout this work, we use $\mathcal{O}, \Theta, \Omega, o$ and ω following standard 'big-O' notation. We use $\|\cdot\|_2$ and $\|\cdot\|_\infty$ to refer to the standard vector 2- and ∞-norms over $\mathbb{C}$ and $\mathbb{R}$, and $\|\cdot\|_s$ to refer to the spectral norm of a matrix or operator. We use $\mathbb{P}^N = \{I, X, Y, Z\}^N$ to denote the set of all tensor products of Pauli matrices on N qubits. We use bold-font to imply that the object in question is a vector of dimension equal to the number of parameters N_p in our learning problem; depending on the context this may be a vector of real or imaginary numbers, or a vector of operators acting on the N-qubit

Hilbert space $\mathbb{C}^{2^N}$. We do not distinguish between operators on Hilbert space and matrices acting on $\mathbb{R}^{N_p}$; this is hopefully clear from context. We use $\langle\cdot\rangle_\rho$ to denote an expectation value of an observable on a quantum state ρ, and $\overline{(\cdot)}$ to denote the expectation value of a random variable over all possible measurement outcomes.

V.A.1. Hamiltonian learning as an oracle problem

In this work, we consider a quantum experiment to consist of three steps: state preparation, unitary evolution, and measurement. For a given experiment x, we denote the initial state as ρ_x. We denote the unitary evolution by U_x. Finally, a measurement is specified by a positive operator-valued measure (POVM), which corresponds to set of positive operators, $\{E_{x,m}\}$, that sum to the identity, $\sum_m E_{x,m} = \mathbb{1}$. The measurement produces an outcome m with probability,

$$S_{x,m} = \text{Trace}[E_{x,m} U_x \rho_x U_x^\dagger]. \tag{V.9}$$

which is returned to the experimentalist. In what follows, we will sometimes omit the experiment label x when it is clear from context.

In this work, we are concerned with learning the parameters $\mathbf{u}$ of a Hamiltonian $H(\mathbf{u})$. For specificity, we consider Hamiltonians that act on N qubits and take the following form:

$$H(\mathbf{u}) = \sum_{a=1}^{N_p} u_a P_a, \qquad u_a \in [-1,1], \qquad P_a \in \mathbb{P}^N, \qquad P_a \neq I. \tag{V.10}$$

where $u_a \in [-1,1]$ denotes an individual unknown parameter, N_p is the total number of unknown parameters, and $P_a \in \{\mathbb{1}, X, Y, Z\}^N$ is a Pauli operators on N qubits. We are particularly interested in learning many-body Hamiltonians, when N is large.

The data that we will learn $\mathbf{u}$ from is the outcomes of quantum experiments in which the unitary evolution U_x involves time evolution by $H(\mathbf{u})$. Before formally defining the Hamiltonian learning problem, we define three specific classes of such experiments that we will consider.

Definition V.1 (No quantum control)
A quantum experiment features no quantum control *if the unitary evolution is equal to a single application of time-evolution under the Hamiltonian,*

$$U = e^{-iH(\mathbf{u})t}, \tag{V.11}$$

for time t. Unless otherwise stated, we assume that state preparation and read-out can be performed in an arbitrary basis, and that the experimentalist

can freely choose t in each experiment.

Experiments with no quantum control are the most restrictive class we will consider.

We now consider two, idealized ways in which increased control by the experimentalist can enhance Hamiltonian learning experiments. First, the experimentalist may be able to insert discrete instantaneous quantum gates throughout time-evolution [286]:

Definition V.2 (Discrete quantum control)
A quantum experiment features discrete quantum control if the experimentalist can interleave time-evolution by a Hamiltonian $H(\mathbf{u})$ (Eq. V.10) with L instantaneous unitary operations V_l. That is,

$$U = V_L e^{-iH\tau_L} V_{L-1} e^{-iH\tau_{L-1}} \dots V_2 e^{-iH\tau_2} V_1 e^{-iH\tau_1}, \qquad H = H(\mathbf{u}) \tag{V.12}$$

for some times $\tau_1, \dots, \tau_L$. We assume that state preparation and read-out can be performed in an arbitrary basis, that the experimentalist can freely choose τ_l and V_l in each experiment, and that V_l may act on an arbitrary number of additional ancilla qubits.

We note that the use of unitary control operations allows a large amount of generality. For instance, an arbitrary mid-experiment POVM measurement can be implemented within the above definition as a unitary operation. Moreover, our definition also incorporates experiments involving up to L copies of a quantum memory, since the control operations can swap the qubits acted on by $H(\mathbf{u})$ with ancilla qubits. Our consideration of a finite number L of discrete operations is motivated by physical concerns, since any "instantaneous" operation will in practice require a non-zero amount of time and incur a non-zero amount of noise. The above definition includes experiments with no quantum control as the special case $L = 1$. Also, the inclusion of the final unitary V_L is not strictly necessary as it can be absorbed into the final measurement.

While the discrete quantum control model is natural for certain hybrid analog-digital quantum setups, in practice energetic restrictions often preclude the application of instantaneous control operations. For example, when evolution under the Hamiltonian $H(\mathbf{u})$ is native to the system and cannot be paused, the application of instantaneous control operations requires driving the system with a control field orders of magnitude stronger than the native Hamiltonian strength. To capture setups where only moderate control strength is possible, we consider the following alternate scenario. We suppose that the experimentalist can *continuously* modify time-evolution by augmenting the Hamiltonian with additional control parameters $\mathbf{c}$:

Definition V.3 (Continuous quantum control)
A quantum experiment features continuous quantum control if the unitary evolution is equal to a single application of time-evolution for time t under an augmented Hamiltonian,

$$U = e^{-iH(\mathbf{u},\mathbf{c})t}, \tag{V.13}$$
$$\text{where} \quad H(\mathbf{u},\mathbf{c}) = H(\mathbf{u}) + H(\mathbf{c}) = \sum_{a=1}^{N_p}(u_a + c_a)P_a,$$
$$|u_a|, |c_a| \leq 1.$$

Again, we assume that state preparation and read-out can be performed in an arbitrary basis, and that the experimentalist can choose $\mathbf{c}$ *and* t *in each experiment.*

We restrict the control parameters $\mathbf{c}$ to be time-independent, which turns out to be sufficient for our learning algorithm in Theorem 1 of the main text. We also assume that the control parameters encompass the same Pauli operators as the unknown parameters. This is an important and rather strict assumption that is necessary for our theorem. Exploring Hamiltonian learning under more restrictive forms of continuous control (e.g. when the control parameters are restricted to single-qubit operators) remains an interesting and potentially relevant future direction.

We can now formalize the Hamiltonian learning problem. We do so by treating each quantum experiment as a query to a probabilistic classical oracle.

Definition V.4 (Classical oracularization of quantum experiments)
A quantum experiment x involving a Hamiltonian $H(\mathbf{u})$ with unknown parameters $\mathbf{u}$ takes as input a classical description of a quantum state ρ_x, a classical description of a POVM $\{E_{x,m}\}$, and: a time t_x (if no quantum control), a list of L times $\tau_{x,l}$, and L unitary operations $V_{x,l}$ (if discrete quantum control), or a time t_x, and a list of control parameters $\mathbf{c}_x$ (if continuous quantum control). The oracle returns a measurement outcome m drawn from the probability distribution $S_{x,m}(t)$ [Eq. (V.9)].

The Hamiltonian learning problem is to learn $\mathbf{u}$ from calls to the classical oracle.

Definition V.5 (Hamiltonian learning problem)
Consider a Hamiltonian $H(\mathbf{u})$ with unknown parameters $\mathbf{u}$, and a desired error ϵ. The Hamiltonian learning problem is to construct an unbiased estimator $\tilde{\mathbf{u}}$

of $\mathbf{u}$ *such that the maximum root-mean-square error is less than* ϵ*,*

$$\max_a \left[\overline{(u_a - \tilde{u}_a)^2}\right]^{1/2} \leq \epsilon, \tag{V.14}$$

given access to an experiment oracle for $H(\mathbf{u})$*. We define the cost of a Hamiltonian learning algorithm as the total Hamiltonian evolution time,*

$$T = \sum_x t_x, \qquad (\text{where } t_x = \sum_l \tau_{x,l} \text{ for discrete quantum control}) \tag{V.15}$$

over all experiments performed.

V.A.2. Alternative error bounds

In the above definition, we have chosen the maximum RMS error as a metric for the performance of a learning algorithm. This is not the only choice possible, and different choices can lead to different costs to learn to the same ϵ. Popular performance metrics for the estimator $\tilde{\mathbf{u}}$ include:

1. Have maximum RMS error $\max_a \sqrt{\overline{(u_a - \tilde{u}_a)^2}} \leq \epsilon$ (the metric used here).
2. Have total RMS error $\sqrt{\sum_a \overline{(u_a - \tilde{u}_a)^2}} \leq \epsilon$.
3. With probability $\geq 1 - \delta$, satisfy $\|\mathbf{u} - \tilde{\mathbf{u}}\|_2 \leq \epsilon$.
4. With probability $\geq 1 - \delta$, satisfy $\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty \leq \epsilon$.

The RMS error $\sqrt{\overline{(u_a - \tilde{u}_a)^2}}$ is equivalent to the 2-norm of the standard deviation of the estimator $\tilde{u}_a$ under the assumption that this estimator is unbiased. Thus, metric 2 is equal to the 2-norm of the standard deviation of the vector estimator $\tilde{\mathbf{u}}$, and metric 1 is equivalent to the infinity norm of the standard deviation of $\tilde{\mathbf{u}}$.

Let us now outline the costs of converting between these various estimators. We summarize this in Table V.1. The contrapositive of a result "existence of estimator $\tilde{\mathbf{u}}$ with cost T $\rightarrow$ existence of estimator $\tilde{\mathbf{u}}'$ with cost T'" is the result "a bound T' on the cost of estimator $\tilde{\mathbf{u}}' \rightarrow$ a bound T on the cost of estimator $\tilde{\mathbf{u}}$", so we will only discuss the costs of generating one estimator from another. An estimator with $P(\|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty \leq \epsilon) > 1 - \delta$ is an estimator with $P(\|\tilde{\mathbf{u}} - \mathbf{u}\|_2 \leq N_p^{1/2}\epsilon) > 1 - \delta$, and an estimator with $P(\|\tilde{\mathbf{u}} - \mathbf{u}\|_2 \leq \epsilon) > 1 - \delta$ is an estimator with $P(\|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty \leq \epsilon) > 1 - \delta$ as

$$N_p^{-\frac{1}{2}} \|\mathbf{u} - \tilde{\mathbf{u}}\|_2 \leq \|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty \leq \|\mathbf{u} - \tilde{\mathbf{u}}\|_2. \tag{V.16}$$

Table V.1.: Conversion table between well-known estimators. Entry i, j gives the conversion from estimator j (defined in top row) to estimator i (defined in left-column). Equivalently, entry i, j gives the conversion from a bound on estimator i (defined in left column) to a bound on estimator j (defined in top row). As indicated in the top-left corner, we assume the cost of constructing the estimator for column j is T, and the cost of constructing the estimator for row i is T'.

Similar logic holds for the RMS error, as

$$\max_a \overline{(u_a - \tilde{u}_a)^2} \leq \sum_a \overline{(u_a - \tilde{u}_a)^2} \leq N_p \max_a \overline{(u_a - \tilde{u}_a)^2}. \tag{V.17}$$

As we have not declared for these estimators how T scales with ϵ, we cannot propagate these changes through to a change in T.

We now consider how to change from bounds on the RMS error to probability-confidence bounds and vice versa. Given an estimator $\tilde{\mathbf{u}}$ with total RMS error ϵ, Chernoff's bound implies we can take the mean of $\mathcal{O}(\log(1/\delta))$ repeated estimates to guarantee $P(\|\mathbf{u} - \tilde{\mathbf{u}}\|_2 \leq \epsilon) \geq 1 - \delta$. Conversely, given an estimator $\tilde{\mathbf{u}}$ with $P(\|\mathbf{u} - \tilde{\mathbf{u}}\|_2 \leq \epsilon) \geq 1 - \delta$, as $\|\mathbf{u} - \tilde{\mathbf{u}}\|_2 \leq N_p^{1/2}$, we can take a median over $\mathcal{O}(\log(N_p)\log(1/\epsilon)/\log(1/\delta))$ estimation copies to obtain an estimator with total RMS error 2ϵ. Similarly, given an estimator $\tilde{\mathbf{u}}$ with $P(\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty \leq \epsilon) \geq 1 - \delta$, I can take a term-wise median over $\mathcal{O}(\log(1/\epsilon)/\log(1/\delta))$ estimation copies to obtain an error with maximum RMS error ϵ. Then, given an estimator $\tilde{\mathbf{u}}$ with maximum RMS error ϵ, we can take a mean over $\mathcal{O}(\log(1/\delta))$ copies and for each a construct an estimator $\tilde{u}_a$ satisfying $P(|u_a - \tilde{u}_a| \leq \epsilon) > 1 - \delta$. But, these estimators can be correlated; in the worst case we can only bound $P(\max|u_a - \tilde{u}_a| \geq \epsilon) \leq \min(1, N\delta)$. To achieve $P(\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty \leq \epsilon) \leq 1 - \delta'$ then requires we set $\delta = \delta'/N$, which incurs an additional $\mathcal{O}(\log(N))$ cost. All other conversions in Tab. V.1 can be obtained by propagating through the results outlined above. One can see that our chosen metric — the max RMS error — is relatively weak (in that it is typically more costly to expand to another metric than vice-versa). This implies that algorithms to achieve a given ϵ are easier to construct, but bounding the cost of an algorithm to achieve accuracy ϵ is more challenging.

V.A.3. The Fisher information and the Cramer-Rao bound

The ability to infer the hidden parameters $\mathbf{u}$ from the experimental outcomes $\{m_x\}$ is bounded by how sensitive the probability distributions $S_{x,m}$ are to

changes in $\mathbf{u}$. The celebrated Cramer-Rao theorem solidifies this, by bounding the covariance matrix of any unbiased estimation of $\mathbf{u}$ by the inverse of the Fisher information matrix $\boldsymbol{\mathcal{I}}$ [297]. The Fisher information matrix quantifies the sensitivity of $S_{x,m}$ to $\mathbf{u}$, and has matrix elements:

$$\mathcal{I}_{ab} = \sum_x \mathcal{I}_{x,ab} = \sum_{m,x} S_{x,m} \left(\frac{\partial \log S_{x,m}}{\partial u_a} \right) \left(\frac{\partial \log S_{x,m}}{\partial u_b} \right) \tag{V.18}$$

$$= \sum_{m,x} \frac{1}{S_{x,m}} \left(\frac{\partial S_{x,m}}{\partial u_a} \right) \left(\frac{\partial S_{x,m}}{\partial u_b} \right). \tag{V.19}$$

where $a, b = 1, \ldots, N_p$ index the unknown parameters. We have used the fact that the Fisher information is additive to write it as a sum over experiments x, and then differentiated the logarithm. In the experiments considered in this work, the dependence of $S_{x,m}$ on $\mathbf{u}$ arises from time-evolution under the unknown Hamiltonian within U_x. The Cramer-Rao theorem establishes that this is the maximum amount of information that can be obtained from the set of experiments $\{x\}$ with probabilistic outcomes $\{m_x\}$, regardless of whether the experiments were performed adaptively or not.

The classical Fisher information above depends upon the measurement bases $E_{m,x}$ of each experiment x. One can also ask the question: what is the maximum information extractable from the final state $U_x \rho_x U_x^\dagger$ of each experiment by any possible measurement. This is quantified by the quantum Fisher information [306, 307]. For the experiments considered in this work, we assume the initial state ρ_x is pure and the dependence on the unknown parameters enters only through the unitary evolution U_x. The quantum Fisher information matrix then takes the simple form [125]:

$$\mathcal{I}^{(Q)}_{x,ab} = 4 \left(\langle A_{x,a} A_{x,b} \rangle_{\rho_x} - \langle A_{x,a} \rangle_{\rho_x} \langle A_{x,b} \rangle_{\rho_x} \right), \tag{V.20}$$

where we denote $\langle \cdot \rangle_\rho \equiv \text{Trace}[(\cdot)\rho]$. (Note that maximal quantum Fisher information from a unitary process is always achieved for a pure state, and so the bounds we derive in this work will hold for mixed states ρ_x as well [125].) Here, we define a central object for learning unitary quantum processes, the Hermitian time-integrated perturbation operator:

$$A_{x,a} = -i U_x^\dagger \frac{\partial U_x}{\partial u_a}. \tag{V.21}$$

For convenience, we write the same equations in vector notation as:

$$\boldsymbol{\mathcal{I}}_x^{(Q)} = 4\left(\left\langle \boldsymbol{A}_x \boldsymbol{A}_x^T \right\rangle_{\rho_x} - \left\langle \boldsymbol{A}_x \right\rangle_{\rho_x} \left\langle \boldsymbol{A}_x^T \right\rangle_{\rho_x}\right), \qquad \boldsymbol{A}_x = -iU_x^\dagger \left(\boldsymbol{\nabla} U_x\right), \quad \text{(V.22)}$$

where the transpose T acts in parameter space (not Hilbert space).

The quantum Cramer-Rao theorem bounds the covariance matrix of any unbiased estimation of $\mathbf{u}$ in terms of the quantum Fisher information matrix, regardless of the choice of measurement bases $E_{m,x}$. We state it without proof:

Theorem V.4 (Quantum Cramer-Rao theorem [125])
Let $\{m_x\}$ be measurement outcomes of a set of quantum experiments x whose unitary evolution depends on an unknown parameter $\mathbf{u}$, and define the quantum Fisher information matrix $\boldsymbol{\mathcal{I}}^{(Q)} = \sum_x \boldsymbol{\mathcal{I}}_x^{(Q)}$ following Eq. (V.20). Suppose that $\tilde{\mathbf{u}}(\{m_x\})$ is an unbiased estimator of $\mathbf{u}$ from the outcomes $\{m_x\}$. Then the covariance matrix of the estimator,

$$\boldsymbol{\Sigma} = \overline{\left(\tilde{\mathbf{u}} - \mathbf{u}\right)\left(\tilde{\mathbf{u}} - \mathbf{u}\right)^T} \quad \text{(V.23)}$$

is lower bounded by the inverse quantum Fisher information matrix:

$$\boldsymbol{\Sigma} \geq [\boldsymbol{\mathcal{I}}^{(Q)}]^{-1}. \quad \text{(V.24)}$$

Here $\overline{(\cdot)}$ denotes the expectation value over measurement outcomes $\{m_x\}$.

The quantum Cramer-Rao theorem immediately allows us to lower bound the error metrics introduced previously. The root-mean-square error in an individual parameter u_a is lower bounded by the diagonal elements of the inverse Fisher information matrix:

$$\overline{\left(u_a - \tilde{u}_a\right)^2} \geq [\boldsymbol{\mathcal{I}}^{(Q)}]^{-1}_{aa}, \quad \text{(V.25)}$$

and the total root-mean-square error by its trace:

$$\sum_a \overline{\left(u_a - \tilde{u}_a\right)^2} \geq \mathrm{Trace}([\boldsymbol{\mathcal{I}}^{(Q)}]^{-1}). \quad \text{(V.26)}$$

Following the previous section, we note that the trace also lower bounds our error metric of choice, the maximum individual root-mean-square error ϵ, via:

$$\epsilon = \max_a \left(\overline{\left(u_a - \tilde{u}_a\right)^2}\right)^{1/2} \geq \left(\max_a [\boldsymbol{\mathcal{I}}^{(Q)}]^{-1}_{aa}\right)^{1/2} \geq \left(\frac{1}{N_p}\mathrm{Trace}([\boldsymbol{\mathcal{I}}^{(Q)}]^{-1})\right)^{1/2}. \quad \text{(V.27)}$$

We are now able to formally derive the Heisenberg limit for Hamiltonian learning:

Theorem V.5 (Heisenberg limit in Hamiltonian learning)
Consider a set of quantum experiments x with unitary evolution,

$$U_x = \mathcal{T}\left\{e^{-i\int_0^{t_x} ds\, H_x(s,\mathbf{u})}\right\}, \tag{V.28}$$

where $H_x(t,\mathbf{u})$ is an arbitrary time-dependent Hamiltonian with a bounded dependence on the unknown parameters $\mathbf{u}$,

$$\left\|\frac{\partial H_x(s,\mathbf{u})}{\partial u_a}\right\|_s \leq 1, \quad \forall\, a. \tag{V.29}$$

Then any algorithm that solves the Hamiltonian learning problem requires a total evolution time,

$$T \geq \Omega(\epsilon^{-1}), \tag{V.30}$$

where $T = \sum_x t_x$.

In the above, we allow a completely general dependence of the unitary evolution on $\mathbf{u}$ in order to capture both the discrete and continuous quantum control scenarios.

Proof— Using the definition of the parametric derivative of an operator exponential [325], we have

$$\frac{\partial U_x}{\partial u_a} = -i\int_0^{t_x} ds\, U_x(t_x,s)\frac{\partial H_x(s,\mathbf{u})}{\partial u_a}U_x(s,0), \tag{V.31}$$

where $U_x(t,s)$ denotes evolution from time s to time t. Observing Eq. (V.21), we can then write the time-integrated perturbation vector as follows:

$$A_{x,a} = -\int_0^{t_x} ds\, U_x^\dagger(s,0)\frac{\partial H_x(s,\mathbf{u})}{\partial u_a}U_x(s,0). \tag{V.32}$$

We can bound the spectral norm as $\|A_{x,a}\|_s \leq t_x$, using the triangle inequality and the assumption $\|\frac{\partial H_x(s,\mathbf{u})}{\partial u_a}\|_s \leq 1$. This bounds the diagonal elements of the inverse quantum Fisher information matrix as,

$$[\boldsymbol{\mathcal{I}}^{(Q)}]^{-1}_{aa} \geq \frac{1}{\mathcal{I}^{(Q)}_{aa}} \geq \frac{1}{4\sum_x t_x^2} \geq \frac{1}{4T^2}. \tag{V.33}$$

Applying the quantum Cramer-Rao bound (Theorem V.4) and Eq. (V.27)

yields,

$$T \geq \frac{1}{2\epsilon} \tag{V.34}$$

which is the Heisenberg limit. □

V.B. A Heisenberg-limited algorithm for low-intersection Hamiltonians with continuous quantum control

V.B.1. The Haah-Kothari-Tang algorithm for learning from short time evolution

In Ref. [289], Haah, Kothari and Tang described an algorithm to learn a Hamiltonian H from short-time evolution by e^{-iHt}, along with state preparation and measurement in an arbitrary product basis. We restate their result here in our formalism for ease of reading.

The HKT algorithm for finite times requires taking a Taylor expansion of $e^{-iHt}Pe^{iHt}$, which in turn requires that this expansion converges. This sets a maximum time, t_c, beyond which the algorithm no longer works. Ref. [289] calculates t_c, in the case that P_a is a Pauli operator with support on $\mathcal{O}(1)$ qubits, as follows. Let us define Supp(P) as the set of qubits that P acts non-trivially on, and let us further define

$$\mathfrak{d} = \max_{a=1,\ldots,N_p} \left| \left\{ b = 1, \ldots, N_p;\, b \neq a;\, \mathrm{Supp}(P_a) \cap \mathrm{Supp}(P_b) \neq \emptyset \right\} \right|, \tag{V.35}$$

which is the maximum degree in the dual graph to H. Ref. [289] gives a critical time $t_c = \mathcal{O}(\mathrm{Poly}(\mathfrak{d}^{-1}))$. Note that if H has maximum degree $d-1 = \max_q |\{a = 1, \ldots, N_p;\, q \in \mathrm{Supp}(P_a)\}|$, we can bound $d \leq \mathfrak{d} \leq k(d-1)$. It is typically assumed that $\mathfrak{d}$ is constant in N, in which case the Hamiltonian is called 'low-intersection', but we will not use this terminology in this work. With this fixed, Ref. [289] gives an algorithm (the HKT algorithm) for learning at finite times. We refer the reader to the original work for details of the algorithm, and state it in a form relevant to us here:

Theorem V.6 ([289] A.1, restated)
Consider the Hamiltonian learning problem for an unknown Hamiltonian $H(\mathbf{u})$, with $\|\mathbf{u}\|_\infty < 1$, with no quantum control (Definition V.1). There exists an algorithm (the HKT algorithm) that with probability $1-\delta$ can estimate $\mathbf{u}$ to

an error $\|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty \leq \epsilon$ *in total experimental time,*

$$T = \mathcal{O}\left(\frac{\mathrm{Poly}(\mathfrak{d})}{\epsilon^2} \log \frac{N_p}{\delta}\right), \tag{V.36}$$

where N_p *is the number of parameters, and* $\mathfrak{d}$ *is the maximum degree in the dual graph [Eq.* (V.35)*].*

The HKT algorithm does not achieve the Heisenberg limit, $T = \mathcal{O}(\epsilon^{-1})$, since each experiment involves time-evolution only for times $t \leq t_c$.

V.B.2. A Heisenberg-limited extension of the HKT algorithm using continuous quantum control

In this section, we extend the HKT algorithm described above to a new algorithm that works at the Heisenberg limit, and prove Theorem 1 of the main text. The idea is, as we improve our estimate of $\mathbf{u}$, to add a control term $\mathbf{c} \sim -\mathbf{u}$, so that the terms in Eq. (V.13) approximately cancel out. We can rescale the resulting Hamiltonian, allowing us to 'zoom in' on the error between our control and fixed parameters at the Heisenberg limit. This is similar in spirit to robust phase estimation [74] (and its multiple phase extension [75]), which achieves the Heisenberg limit by using estimates of $e^{i2^d\phi}$ to estimate ϕ to within an error of $\frac{\pi}{2^{d-1}}$. Doing this for $d = 0, 1, \ldots D$ avoids the aliasing error $e^{i2^d(\phi+2^{d-1}\pi)} = e^{i2^d\phi}$.

To achieve the Heisenberg limit requires we assume continuous quantum control of our system (Definition V.3). This allows us to 'zoom in' on our Hamiltonian: we use our control term to approximately cancel our fixed parameters, and rescale our time variable. This gives a new black box that we can plug into the HKT algorithm, but one which takes longer to query.

Lemma V.1
Assume we have an experiment oracle time-independent $H(\mathbf{u}, \mathbf{c})$ *with dynamic control, and an estimate* $\tilde{\mathbf{u}} \in [-1, 1]^{N_p}$*, such that* $\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty < \Delta$*. We can query a 'rescaled' experiment oracle with a Hamiltonian* $H'(\mathbf{u}')$*, where* $\mathbf{u}' = \frac{1}{\Delta}(\mathbf{u} - \tilde{\mathbf{u}})$ *and* $\|u'\|_\infty \leq 1$*. This incurs a rescaled cost* $\frac{t}{\Delta}$ *to query for an experiment with evolution time* t*.*

Proof—If $\|\mathbf{u} - \mathbf{c}\|_\infty \leq \Delta$, $\|\Delta^{-1}(\mathbf{u} - \mathbf{c})\|_\infty \leq 1$ and $\Delta^{-1}(\mathbf{u} - \mathbf{c}) \in [-1, 1]^{N_p}$. Then, fix input $\rho(0), \mathbf{0}, t, \{E_i\}$, for the new black box. From Eq. (V.13) we can see that $H(\mathbf{a}, \mathbf{b}) = H(\mathbf{a} + \mathbf{b}, 0)$ and $H(\lambda\mathbf{a}, 0) = \lambda H(\mathbf{a}, 0)$, which implies

we need to sample from

$$S_m(t) = \mathrm{Trace}\Big[E_m \exp\big(-it\Delta^{-1}H(\mathbf{u},-\mathbf{c},0)\big)\rho(0)\exp\big(it\Delta^{-1}H(\mathbf{u},-\mathbf{c})\big)\Big]. \tag{V.37}$$

We can obtain this sampling by querying the original black box with $\rho = \rho(0), \mathbf{c} = -\tilde{\mathbf{u}}, t = t/\Delta, \{E_i\} = \{E_i\}$, which has a cost $\frac{t}{\Delta}$. □

With this given, our algorithm proceeds as follows:

Algorithm V.1 (Heisenberg-limited extension of the HKT algorithm)
Input: Hamiltonian $H = H(\mathbf{u}, \mathbf{c})$ as given by Eq. (V.13)*, desired precision ϵ, confidence parameter $c \in (0, 1/24]$.*

1. *Let $D = \lceil \log_2(1/\epsilon) \rceil$.*
2. *Let $\tilde{\mathbf{u}}^{(0)} = \mathbf{0}$ be our 0th order estimate for $\mathbf{u}$.*
3. *For $d = 0, ..., D$:*
   a) *Fix $H^{(d)}(\mathbf{u}) = 2^d H(\mathbf{u}, -\tilde{\mathbf{u}}^{(d)})$.*
   b) *Apply the HKT algorithm (Theorem V.6) to the rescaled Hamiltonian $H^{(d)}$, with error bound 1/2 and error probability $\delta^{(d)} = \frac{c}{8^{D-d}}$. Define the output estimate $\mathbf{g}^{(d)}$.*
   c) *If $\|\mathbf{g}^{(d)}\|_\infty \geq 1$, set $\mathbf{g}^{(d)} = 0$.*
   d) *Set $\tilde{\mathbf{u}}^{(d+1)} = \tilde{\mathbf{u}}^{(d)} + 2^{-d}\mathbf{g}^{(d)}$.*
4. *Return $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^{(D+1)}$.*

To show that this algorithm achieves the Heisenberg limit as defined, we first bound the error in the d-th order estimate $\tilde{\mathbf{u}}^{(d)}$ assuming the first d runs of the HKT subroutine have succeeded (Lemma V.2). We then bound the error in the final estimate $\tilde{\mathbf{u}}$ under the same assumption (Lemma V.3). Finally, in Theorem V.7, we combine these results to show the final RMS error is $\leq \epsilon$, and calculate the cost of running Algorithm V.1 to prove the Heisenberg scaling.

Lemma V.2
Consider an application of Algorithm V.1 to an experiment oracle with unknown parameters $\mathbf{u}$. If for each $d' = 0, 1, \ldots, d-1$ the HKT algorithm subroutine in step 3b of the algorithm succeeds, then $\|\tilde{\mathbf{u}}^{(d)} - \mathbf{u}\|_\infty \leq 2^{-d}$.

Proof—We prove this result by induction in d. First we prove the case for $d = 1$. As $H^{(1)}(\mathbf{u}) = H(\mathbf{u})$ we always satisfy the input condition of the HKT algorithm $\|\mathbf{u}\|_\infty \leq 1$. We have $\tilde{\mathbf{u}}^{(1)} = g^{(1)}$, hence the success of the HKT algorithm implies $\|\tilde{\mathbf{u}}^{(1)} - \mathbf{u}\| \leq 1/2$.

Next, we prove that $d \to d+1$. Lemma V.1 implies that if $\|\tilde{\mathbf{u}}^{(d)} - \mathbf{u}\| \leq 2^{-d}$ we satisfy the input conditions to the HKT algorithm at round d. If the subroutine succeeds, the output will satisfy

$$\|\mathbf{g}^{(d)} - 2^d(\mathbf{u} - \tilde{\mathbf{u}}^{(d)})\| \leq \frac{1}{2}. \tag{V.38}$$

Multiplying both sides of this inequality by 2^{-d}, and substituting in the definition of $\tilde{\mathbf{u}}^{(d+1)}$ we find that $\|\tilde{\mathbf{u}}^{(d+1)} - \mathbf{u}\| \leq 2^{-(d+1)}$, as required. □

Lemma V.3
Consider an application of Algorithm V.1 to an experiment oracle with unknown parameters $\mathbf{u}$*. If for each* $d' = 0, 1, \ldots, d-1$ *the HKT algorithm subroutine in step 3b of the algorithm succeeds, then the final estimate* $\tilde{\mathbf{u}}$ *satisfies* $\|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty \leq 3 \times 2^{-d}$*.*

Proof— We will first show that the final estimate $\tilde{\mathbf{u}}$ is close to the d-th order estimate $\tilde{\mathbf{u}}^{(d)}$. Combining this observation with Lemma V.2 completes the proof.

We have defined the estimate $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^{(D+1)}$ recursively in step 3d of Algorithm V.1. We can expand this as

$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^{(D)} + \frac{\mathbf{g}^{(D)}}{2^D} = \tilde{\mathbf{u}}^{(D-1)} + \frac{\mathbf{g}^{(D-1)}}{2^{D-1}} + \frac{\mathbf{g}^{(D)}}{2^D} = \ldots = \tilde{\mathbf{u}}^{(d)} + \sum_{d'=d}^{D} \frac{\mathbf{g}^{(d')}}{2^{d'}}. \tag{V.39}$$

As in step 3c we ensure $\|\mathbf{g}^{(d')}\|_\infty \leq 1$ for all d', by the triangle inequality we have

$$\|\tilde{\mathbf{u}} - \tilde{\mathbf{u}}^{(d)}\|_\infty \leq \sum_{d'=d}^{D} 2^{-d'} \|\mathbf{g}^{(d')}\|_\infty \leq 2^{-d+1}. \tag{V.40}$$

We use the triangle inequality again to bound the error in the final estimate

$$\|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty \leq \|\tilde{\mathbf{u}} - \tilde{\mathbf{u}}^{(d)}\|_\infty + \|\tilde{\mathbf{u}}^{(d)} - \mathbf{u}\|_\infty. \tag{V.41}$$

Bounding the first term as in Eq. (V.40), and the second term as in Lemma V.2 yields the required result. □

Now, let us calculate the cost of implementing Algorithm V.1 and show that it scales at the Heisenberg limit.

Theorem V.7
Given access to an experiment oracle with continuous quantum control, Algorithm V.1 solves the learning problem (Definition V.5) to RMS error ϵ *at a cost* $T = \mathcal{O}(\epsilon^{-1})$*.*

Proof—We first show that the RMS error in the final estimate $\tilde{\mathbf{u}}$ is $\leq 2^{-D} \leq \epsilon$, where the second inequality follows from the definition of the final order D in step 1. We have to consider contributions to the error from two different scenarios: either the HKT subroutine in step 3b succeeds every time, or it fails for the first time at some order d. If the subroutine always succeeds, which happens with probability $\prod_{d'=0}^{D}(1-\delta^{(d')}) \leq 1$, by Lemma V.2 the final error is at most $2^{-(D+1)}$. With probability $\delta^{(d)}\prod_{d'=0}^{d-1}(1-\delta^{(d')}) \leq \delta^{(d)} = c/8^{D-d}$, the HKT subroutine succeeds for $d' = 0, 1, \ldots, d-1$ and fails at order d. Lemma V.2 bound the final error in this case by 3×2^{-d}. Combining all the possible cases, we can bound the squared error as

$$\overline{(\tilde{u}_a - u_a)^2} \leq 2^{-2(D+1)} + \sum_{d=0}^{D} c/8^{D-d} \times 3^2 \times 2^{-2d} \tag{V.42}$$

$$= 2^{-2(D+1)} + 9c\sum_{d=0}^{D} 2^{-3D+d}. \tag{V.43}$$

Summing the geometric series in the second term we get

$$\overline{(\tilde{u}_a - u_a)^2} \leq 2^{-2D}\left(\frac{1}{4} + 18c\right). \tag{V.44}$$

As we have assumed $c \leq 1/24$, the right hand side is $\leq 2^{-2D} \leq \epsilon^2$. The above equation holds for all a, which implies the desired bound on the maximum RMS error.

It remains to calculate the cost of executing the algorithm Let us write this as a sum $C = \sum_d C_d$ where C_d is the cost of the dth order call to the HKT algorithm Substituting in $\epsilon = 1/2$ and $p = 1 - c/8^{D-d}$, and recalling that the runtime of the dth order black-box experiment scales as 2^d, we have

$$C_d = \mathcal{O}\left(\mathrm{Poly}(\mathfrak{d})2^d \log\frac{8^{D-d}N}{c}\right). \tag{V.45}$$

We can then evaluate $C = \sum_d C_d$ by summing the geometric series,

$$\sum_d \mathrm{Poly}(\mathfrak{d})2^d \log\frac{8^{D-d}N}{c} = \mathrm{Poly}(\mathfrak{d})\sum_{d=0}^{D}\Big(\log(N/c) + (D-d)\log(8)\Big)2^d$$

$$= \mathrm{Poly}(\mathfrak{d})\left[\log(N/c)\Big(2^{D+1}-1\Big) + \log(8)\Big(2^{D+1} - D - 2\Big)\right], \tag{V.46}$$

and keeping only terms at leading order in 2^D, yields,

$$C = \sum_d C_d = \Theta\left(2^D \mathrm{Poly}(\mathfrak{d}) \log \frac{8N}{c}\right) = \Theta\left(\frac{\mathrm{Poly}(\mathfrak{d}) \log \frac{8N}{c}}{\epsilon}\right). \tag{V.47}$$

We have $C = \Theta(\epsilon^{-1})$as required. □

V.C. Absence of the Heisenberg limit under unitary invariance

In this section, we establish our first no-go theorem, showing that learning at the Heisenberg limit is impossible without quantum control, for certain Hamiltonians. We in fact establish a moderately stronger result: that learning at the Heisenberg limit is impossible even with discrete quantum control, unless the number L of interleaved unitaries scales with the total time T, $L = \Omega(T)$. The key result underpinning this section is a bound on the quantum Fisher information matrix as a function of the derivatives of eigenvectors and eigenvalues of the Hamiltonian. We state this for the discrete quantum control model with fixed L, and recall that this is equivalent to no quantum control when $L = 1$.

We begin by relating the quantum Fisher information to the eigenvalue decomposition of the Hamiltonian through the following Lemma.

Lemma V.4
Consider a quantum experiment involving a Hamiltonian, $H(\mathbf{u})$, in the discrete quantum control model with up to L interleaved unitaries. Write the eigenvalue decomposition of the Hamiltonian as $H(\mathbf{u}) = W(\mathbf{u})^\dagger D(\mathbf{u}) W(\mathbf{u})$, where W is unitary and D is diagonal. Then for any normalized $\mathbf{v} \in \mathbb{R}^{N_p}$, we have:

$$\mathbf{v}^T \boldsymbol{\mathcal{I}}^{(Q)} \mathbf{v} \leq \min\left(t\|\partial_{\mathbf{v}} D\|_s + 2L\|\partial_{\mathbf{v}} W\|_s \ ,\ t\|\partial_{\mathbf{v}} H\|_s\right)^2, \tag{V.48}$$

where $\boldsymbol{\mathcal{I}}^{(Q)}$ is the quantum Fisher information matrix for the experiment and $\partial_{\mathbf{v}} \equiv \sum_a v_a \frac{\partial}{\partial u_a}$.

Proof— Observing the definition of the quantum Fisher information matrix [Eq. (V.20)], we immediately have

$$\mathbf{v}^T \boldsymbol{\mathcal{I}}^{(Q)} \mathbf{v} = \langle A_{\mathbf{v}}^2 \rangle - \langle A_{\mathbf{v}} \rangle^2 \leq \langle A_{\mathbf{v}}^2 \rangle \leq \|A_{\mathbf{v}}\|_s^2. \tag{V.49}$$

Here we denote $A_{\mathbf{v}} = \sum_a v_a A_a$, with the operator A_a defined in terms of the time-evolution unitary in Eq. (V.21), $A_a = U(t,0)^\dagger \frac{\partial}{\partial u_a} U(t,0)$. Working in

the discrete quantum control model [Eq. (V.12)], we can apply the eigenvalue decomposition of the Hamiltonian to the unitary to give

$$U(t,0) = V_L W^\dagger e^{iD\tau_L} W V_{L-1} W^\dagger e^{iD\tau_{L-1}} W \dots V_2 W^\dagger e^{iD\tau_2} W V_1 W^\dagger e^{iD\tau_1} W. \tag{V.50}$$

Differentiating with respect to $\mathbf{v}$ yields

$$\partial_{\mathbf{v}} U(t,0) = \tag{V.51}$$
$$\sum_l U(t,t_l) V_l \Big({}_{i\tau_l W^\dagger \partial_{\mathbf{v}} D e^{i\tau_l D} W + \partial_{\mathbf{v}} W^\dagger e^{-i\tau_l D} W + W^\dagger e^{-i\tau_l D} \partial_{\mathbf{v}} W} \Big) U(t_{l-1},0),$$

where $t_l = \sum_{l' \le l} \tau_{l'}$ is the total duration of time-evolution before the l^{th} interleaved unitary. Since multiplication by a unitary operator does not change the spectral norm, we can bound the spectral norm of $A_{\mathbf{v}}$ by the spectral norms of the terms in parentheses. Applying the triangle inequality gives

$$\|A_{\mathbf{v}}\|_s = \Big\| \partial_{\mathbf{v}} U(t,0) \Big\|_s \le \sum_l \left(\tau_l \|\partial_{\mathbf{v}} D\|_s + 2\|\partial_{\mathbf{v}} W\|_s \right) \tag{V.52}$$
$$= t\|\partial_{\mathbf{v}} D\|_s + 2L\|\partial_{\mathbf{v}} W\|_s. \tag{V.53}$$

We can improve this bound by invoking a second upper bound on $\|A_{\mathbf{v}}\|_s$ and taking the minimum of the two. Note that the term in parentheses in Eq.(V.51) is equal to

$$\partial_{\mathbf{v}} e^{-iH(\mathbf{u})\tau_l} = -i \int_0^{\tau_l} ds\, e^{-iH(\mathbf{u})(\tau_l - s)} \partial_{\mathbf{v}} H e^{-iH(\mathbf{u})s}, \tag{V.54}$$

which is upper bounded by $\|\partial_{\mathbf{v}} e^{-iH(\mathbf{u})\tau_l}\|_s \le \tau_l \|\partial_{\mathbf{v}} H(\mathbf{u})\|_s$. Applying the triangle inequality to Eq. (V.51) gives

$$\|A_{\mathbf{v}}\|_s \le \|\partial_{\mathbf{v}} H(\mathbf{u})\|_s \sum_l \tau_l = t\|\partial_{\mathbf{v}} H(\mathbf{u})\|_s. \tag{V.55}$$

Inserting the minimum of the two bounds into Eq. (V.49) yields the final result. □

In the above lemma, the contribution of eigenvalues D to the quantum Fisher information grows quadratically in time, while the contribution of the eigenvectors W eventually saturates to a constant. Intuitively, this tells us that eigenvalues of the Hamiltonian can be learned at the Heisenberg limit, while eigenvectors cannot. We formalize this in the following theorem.

Theorem V.8
Consider the Hamiltonian learning problem (Definition V.5) for an unknown Hamiltonian $H(\mathbf{u})$ and desired error ϵ, and write the eigenvalue decomposition of the Hamiltonian as $H(\mathbf{u}) = W(\mathbf{u})^\dagger D(\mathbf{u}) W(\mathbf{u})$. Suppose that each experiment involves discrete quantum control with up to L interleaved unitaries per experiment, and that there exists a normalized vector $\mathbf{v} \in \mathbb{R}^{N_p}$ such that $\partial_{\mathbf{v}} D = 0$. Then learning the linear combination of parameters $\mathbf{u} \cdot \mathbf{v}$ to within RMS error ϵ requires a total time

$$T \geq \Omega\left(\epsilon^{\frac{-1}{1+\alpha}}\right), \quad \text{if} \quad L = \mathcal{O}(T^\alpha). \tag{V.56}$$

In particular, the combination cannot be learned at the Heisenberg limit unless $L = \Omega(T)$, i.e. $\alpha \geq 1$.

Note that the bound in Eq. (V.56) is superseded by the Heisenberg limit whenever $\alpha > 1$. Theorem 2 of the main text follows from this result by setting $L = \mathcal{O}(1)$, i.e. $\alpha = 0$, for which we have $T = \Omega(\epsilon^{-2})$.

Proof of Theorem V.8—For $\mathbf{v}$ satisfying the conditions of the theorem, application of Lemma V.4 to each experiment x yields

$$\mathbf{v}^T \boldsymbol{\mathcal{I}}^{(Q)} \mathbf{v} \leq \sum_x \min\left(2L\|\partial_{\mathbf{v}} W\|_s \ , \ t_x \|\partial_{\mathbf{v}} H\|_s\right)^2. \tag{V.57}$$

For fixed $T = \sum_x t_x$, this quantity is maximized by choosing $t_x \leq 2L\|\partial_{\mathbf{v}} W\|_s / \|\partial_{\mathbf{v}} H\|_s$ across $\lceil T/t_x \rceil$ experiments. Setting T to a multiple of t_x for convenience, this yields

$$\mathbf{v}^T \boldsymbol{\mathcal{I}}^{(Q)} \mathbf{v} \leq 4L^2 \|\partial_{\mathbf{v}} W\|_s^2 \left(\frac{T\|\partial_{\mathbf{v}} H\|_s}{2L\|\partial_{\mathbf{v}} W\|_s}\right) = 2L\|\partial_{\mathbf{v}} W\|_s \|\partial_{\mathbf{v}} H\|_s T. \tag{V.58}$$

Following the Cramer-Rao bound, Eq. (V.27), we have

$$\epsilon^2 \geq \mathbf{v}^T [\boldsymbol{\mathcal{I}}^{(Q)}]^{-1} \mathbf{v} \geq (\mathbf{v}^T \boldsymbol{\mathcal{I}}^{(Q)} \mathbf{v})^{-1}, \tag{V.59}$$

where the latter bound follows from the Cauchy-Schwarz inequality, as

$$1 = \mathbf{v}^T [\boldsymbol{\mathcal{I}}^{(Q)}]^{1/2} [\boldsymbol{\mathcal{I}}^{(Q)}]^{-1/2} \mathbf{v} \tag{V.60}$$

$$\leq \|[\boldsymbol{\mathcal{I}}^{(Q)}]^{1/2} \mathbf{v}\|_2 \|[\boldsymbol{\mathcal{I}}^{(Q)}]^{-1/2} \mathbf{v}\|_2 \tag{V.61}$$

$$= \sqrt{(\mathbf{v}^T \boldsymbol{\mathcal{I}}^{(Q)} \mathbf{v})(\mathbf{v}^T [\boldsymbol{\mathcal{I}}^{(Q)}]^{-1} \mathbf{v})}. \tag{V.62}$$

Combining with Eq. (V.58), we have

$$\epsilon^2 \geq (2L\|\partial_{\mathbf{v}} W\|_s \|\partial_{\mathbf{v}} H\|_s T)^{-1}. \tag{V.63}$$

Finally, we can re-arrange terms and set $L = \mathcal{O}(T^\alpha)$ to find

$$T \geq \left(2\|\partial_{\mathbf{v}} W\|_s \|\partial_{\mathbf{v}} H\|_s \epsilon^2\right)^{\frac{-1}{1+\alpha}}, \tag{V.64}$$

as desired. □

We conclude by outlining a more physical picture for where the above theorem applies (as was discussed briefly in the main text). Namely, suppose, as above, that the eigenvalues D of a Hamiltonian H are invariant under an infinitesimal parameter shift, $\mathbf{u} \to \mathbf{u} + \delta\mathbf{u}$, for some $\delta\mathbf{u}$. This implies the two Hamiltonians are unitarily equivalent: $H(\mathbf{u} + \delta\mathbf{u}) = U^\dagger(\delta\mathbf{u}) H(\mathbf{u}) U(\delta\mathbf{u})$. This is a one-to-one relationship, since unitary equivalence in turn implies that $\partial_{\delta\mathbf{u}} D = 0$. We can therefore state the above result in a more intuitive way: if we can unitarily transform between two nearby Hamiltonians, then we cannot distinguish between them at the Heisenberg limit. To formalize this, consider a orbit in parameter space along which the spectrum of a Hamiltonian remains invariant. The presence of such a orbit will preclude Heisenberg-limited learning of Hamiltonians lying along the orbit.

We mention two families of Hamiltonians that feature unitary equivalences—single-qubit Hamiltonians and interacting fermion Hamiltonians—in the main text. Here, we mention that unitary equivalence also holds for interacting spin Hamiltonians of the form:

$$H(\mathbf{u}) = \sum_i u_i^{(1)} Z_i + \sum_{\langle i,j \rangle} \left(u_{ij,xx}^{(2)} X_i X_j + u_{ij,xy}^{(2)} X_i Y_j + u_{ij,yx}^{(2)} Y_i X_j + u_{ij,yy}^{(2)} Y_i Y_j \right), \tag{V.65}$$

where $\langle \cdot, \cdot \rangle$ denotes nearest neighbour pairs on a grid. This possess orbits at any $\mathbf{u}$ that are unitarily-equivalent up to local Z-rotations. This can be extended to arbitrary local rotations by adding other 1- and 2-qubit terms on nearest neighbours.

V.D. Bounds on learning thermalizing Hamiltonians

In this section we formally state and prove Theorem 3 of the main text, establishing that Heisenberg-limited Hamiltonian learning is not possible without quantum control in systems that obey the eigenstate thermalization

hypothesis (ETH). We remark that, even on physical grounds, the existence of this result is not clear a priori. On the one hand, for typical initial states, time-evolution under a thermalizing Hamiltonian will cause the system to approach an equilibrium state where few-body observables are well approximated by a thermal density matrix. Since the equilibrium state is independent of time, this appears to preclude learning at the Heisenberg limit. On the other hand, all Hamiltonian systems can support long-lived coherences within the energy eigenbasis. The dependence of these coherences on Hamiltonian parameters could in principle open the door to Heisenberg-limited learning.

In what follows, we resolve these questions using the ETH. We begin in Appendix V.D.1 with a formal statement of the ETH, adapted from the standard but non-rigorous definitions in Refs. [314–317]. In Appendix V.D.2, we provide several preliminary definitions and results regarding low-rank approximations of the Fisher information matrix $\boldsymbol{\mathcal{I}}^{(Q)}$, as well as the operators A_a defined in Eq. (V.21), which will be useful later on. In particular, we prove that the existence of a sufficiently low-rank approximation forbids learning all $N_p = \Omega(N)$ Hamiltonian parameters at the Heisenberg limit. This motivates our main result in Appendix V.D.3: a proof of Theorem V.9, the formal version of Theorem 3 in the main text. Our proof strategy is to leverage the connection between the quantum Fisher information matrix and the operators A_a defined in Eq. (V.21). The operators A_a are equal to integrals over time-evolved local Hamiltonian terms, which can be treated via the ETH.

Our proof of Theorem V.9 builds upon two lemmas, which we prove separately in Appendixs V.D.4 and V.D.5. In Appendix V.D.4, we show that, if a Hamiltonian obeys the ETH, the spectral norm of the off-diagonal components of A_a are upper bounded by $\mathcal{O}(t^{1/2})$ and thus cannot contribute to learning at the Heisenberg limit. Our proof utilizes standard techniques from random matrix theory. In Appendix V.D.5, we show that the diagonal components of A_a are highly linearly dependent for different parameters a, which we capture formally using the notion of low-rank approximations introduced in Appendix V.D.2. This result relies on a physical assumption, namely that the temperature derivatives of thermal expectation values are bounded by a constant (i.e. do not diverge in the system size). In Appendix V.D.6, we prove that this assumption is satisfied in any geometrically local Hamiltonian with exponential decay of correlations for operators that enter the Hamiltonian. Exponential decay of correlations have in turn been established for one-dimensional Hamiltonians at any finite temperature [320] and higher-dimensional Hamiltonians at sufficiently large temperatures [319]. We expect our assumption to hold more broadly than is currently provable (e.g. in all-to-all interacting systems as well), as long as the states of interest are a non-zero temperature difference away from any finite temperature phase

transitions.

The above results concern learning many-body Hamiltonians with $N_p = \Omega(N)$ unknown parameters. In Appendix V.D.7, we conclude by discussing learning in thermalizing Hamiltonians with only a constant number, $N_p = \mathcal{O}(1)$, of unknown parameters. In this case, we show that Heisenberg-limited Hamiltonian learning may in fact be possible even without quantum control, but requires the ability to prepare and measure coherences between states with extensive differences in energy. Relatedly, we show that Heisenberg-limited learning is not possible if the initial states are related to product states by any finite depth quantum circuit.

V.D.1. The eigenstate thermalization hypothesis

We begin by formulating a precise statement of the eigenstate thermalization hypothesis. Thermalization is the common behaviour of observables in physical systems to relax towards a thermal average, regardless of their initial state. This is surprisingly difficult to predict in quantum mechanics, as the absolute values of a density matrix in the eigenbasis of a system's Hamiltonian H are invariant in time. A popular method to explain this has been the eigenstate thermalization hypothesis [314–317].

The eigenstate thermalization hypothesis is commonly stated as follows [316]. Consider a Hamiltonian H with energy eigenstates $H|E_j\rangle = E_j|E_j\rangle$, and a k-local operator V with $k = \mathcal{O}(1)$. We are interested in the matrix elements of V in the energy eigenbasis. The eigenstate thermalization hypothesis posits that:

(i) The diagonal matrix elements of V are given by their thermal expectation values[2]

$$\langle E_j|V|E_j\rangle = \langle V\rangle_{\beta(E_j)}\,, \tag{V.66}$$

where we define an inverse temperature, $\beta(E)$, for each energy via $\langle H\rangle_{\beta(E)} = E$, where $\langle\cdot\rangle_\beta$ denotes the trace with the thermal density matrix, $\rho_\beta = e^{-\beta H}/\mathrm{Trace}(e^{-\beta H})$. (Such a temperature exists for every energy E in the spectrum of H.)

(ii) The off-diagonal matrix elements of V are given by

$$\langle E_i|V|E_j\rangle = e^{-S(E_{ij})/2} f_V\Big(E_{ij}, \omega_{ij}\Big)\, r^V_{ij}, \quad i \neq j \tag{V.67}$$

[2]To be maximally precise, the diagonal matrix elements are expected to display small random variations $r^V_{ii} \sim \mathcal{O}(e^{-S/2})$ about their thermal values. However, these contributions have exponentially small spectral norm and are thus negligible for our analysis. Nevertheless, if desired, they are easily incorporated into our proof of Theorem V.9 by modifying the operators $A_{\mathrm{od},a}$ in Eq. (V.92) to include the diagonal fluctuations.

where $E_{ij} \equiv \frac{1}{2}(E_i + E_j)$ is the average energy, $\omega_{ij} \equiv E_i - E_j$ is the energy difference, $r^V_{ij} = (r^V_{ji})^*$ are random[3] complex numbers with zero mean and unit variance, and $S(E)$ and $f_V(E, \omega)$ are smooth functions defined below.

Statement (i) implies that eigenstates "look" thermal from the perspective of observables V. Numerical experiments have found that this indeed holds, up to corrections that are exponentially small in the system size (see [316] for a comprehensive review of numerical experiments). Statement (ii) implies that the off-diagonal matrix elements "look" like random complex numbers, up to the requirement that they reproduce the auto-correlation function of the operator V. Here, $e^{S(E)}$ is the density of states at energy E (which is expected to be exponential in the system size at any finite temperature). A straightforward calculation shows that the function $f_V(E, \omega)$ is related to the Fourier transform of the connected auto-correlation function $G^c(t; E_i)$ in an eigenstate $|E_i\rangle$ via [316]

$$|f_V(E_{ij}, \omega_{ij})|^2 = e^{S(E_j) - S(E_{ij})} \int_{-\infty}^{\infty} dt\, e^{-i\omega_{ij} t} G^c(t; E_i), \tag{V.68}$$

where we define

$$G^c(t; E_i) = \langle E_i | V(t) V(0) | E_i \rangle - \langle E_i | V | E_i \rangle^2, \tag{V.69}$$

and $V(t) = U(t, 0)^\dagger V U(t, 0)$. In previous literature it is common to equate the above auto-correlation function with its value in a thermal state [i.e. replacing $\langle E | (\cdot) | E \rangle \to \mathrm{Trace}((\cdot)\rho_{\beta(E)})$], as well as to approximate the difference in entropies using the inverse temperature via $S(E_j) - S(E_{ij}) \approx \beta(E_j)\omega_{ij}/2$. However, we will not need these simplifications in this work. Numerical studies have found that the magnitudes of off-diagonal matrix elements are indeed well-described by Eq. (V.67) (see again [316] for a review). However, several more recent works have pointed out that the coefficients r^V_{ij} are not in general i.i.d. random numbers; indeed, correlations between different coefficients are in fact necessary to capture the behavior of higher-point correlation functions in the theory [327–332]. After a time-scale $T_{\mathrm{rmt}} \sim \mathrm{poly}(N)$ that is polynomial in the system size, these correlation functions are expected to fully decay and

[3] As has been previously noted [326, 327], it is not immediately obvious how to interpret the randomness in r^V_{ij}, since the actual matrix elements of interest are defined for a specific, non-random Hamiltonian. An intuitive way to reconcile this difference is to imagine adding small, random perturbations to the Hamiltonian. If these perturbations are large compared to the level spacing but small compared to the time-scales of interest, then they can serve to randomize the matrix elements while leaving physical observables approximately unchanged. Since the level spacing is exponentially small in the system size, there is an exponentially large time window in which this picture can apply.

the coefficients to be treatable as i.i.d. random numbers [327, 333–335].

Motivated by these observations, in this work we adopt the following definition of ETH.

Definition V.6
Consider a Hamiltonian H acting on N qubits, with energy eigenstates $H|E_j\rangle = E_j|E_j\rangle$. We say that H satisfies the eigenstate thermalization hypothesis (ETH) over an inverse temperature range $[\beta_1, \beta_2]$, if the following conditions hold for all energy eigenvalues E_i, E_j such that $\beta(E_i), \beta(E_j) \in [\beta_1, \beta_2]$:

1. *For Pauli operators V that act on a constant number of qubits, time-integrated operators $V(t)$ are well-approximated by random matrix theory predictions,*

$$\left(\frac{1}{T}\int_0^T dt\, V(t)\right) = \left(\frac{1}{T}\int_0^T dt\, V_{\mathrm{rmt}}(t)\right) + \delta V, \quad \text{with} \quad \|\delta V\|_s \leq \frac{C}{T}, \tag{V.70}$$

   where V_{rmt} is defined via Eq. (V.66) and Eq. (V.67) by taking r_{ij}^V as i.i.d. normal random numbers with unit variance. Here, $C = \mathcal{O}(poly(N))$ is a constant independent of T.

2. *The system thermalizes in the conventional sense: namely, auto-correlation functions decay to their thermal values after a time-scale that is polynomial in the system size. We capture this via the following condition:*

$$\frac{1}{T}\int_0^T dt\, G^c(t; E_i) \leq \frac{C}{T}, \tag{V.71}$$

   for every $|E_j\rangle$. We take the constant $C = \mathcal{O}(poly(N))$ to be equal to that in Eq. (V.70) for convenience.

3. *The density of states is exponential in the system size, $e^{S(E)} = \Omega(exp(N))$.*

In conditions both 1 and 2, we find it more natural to state the requirements for thermalization in terms of averages over time-evolved operators. In condition 1, we note that the time-average leads to a *weaker* condition than the conventional formulation of ETH in Eqs. (V.66, V.67), which involves the operator without time-averaging. This modification enforces that any correlation function involving the approximated operator will be dominated by time differences of order T, where we expect the random matrix prescription to be appropriate for sufficiently large times[4] (i.e. $T \gtrsim T_{\mathrm{rmt}}$; see the discussion

[4]Several previous works [326] have addressed this issue in an alternate fashion, by only

above Definition V.6). We expect the corrections, δV, to this approximation to scale as at most, $\|V\|_s \lesssim T_{\text{rmt}}/T$. We therefore expect the constant C in the definition above to be proportional to the random matrix time scale, T_{rmt}. In condition 2, we expect the the connected auto-correlation function itself, $G^c(t; E_i)$, to be zero for nearly all times after the "thermalization time" [316, 336], which occurs before the random matrix time scale T_{rmt}. The average over time is useful to tolerate Poincare recurrences in the correlation function, which are expected to be exponentially rare in the system size.

Several recent extensions of the ETH go beyond the systems and time regimes captured by the above definition. Firstly, the above definition fails to describe thermalizing Hamiltonians that possess a conserved quantity besides the energy (for instance, a conserved particle number or a conserved spin parity). When the additional conserved quantities are mutually commuting, they can be accounted for straightforwardly by applying the ETH to each charge or parity sector independently; additional subtleties are encountered when the conserved quantities are non-commuting [337]. We expect on physical grounds that our results in the following sections will carry over to both cases, but we do not explore this here. Secondly, recent work has made tremendous progress towards capturing the correlations between matrix elements at times before T_{rmt} using non-Gaussian random matrix ensembles [332]. It would interesting and non-trivial to explore whether our results can be extended using this framework.

V.D.2. Low-rank matrix approximations

Similar to Appendix V.C, our eventual strategy for bounding learning in thermalizing Hamiltonians with no quantum control will utilize the quantum Fisher information matrix, $\boldsymbol{\mathcal{I}}^{(Q)}$ [Eq. (V.20)]. Recall that in order for all N_p parameters of the Hamiltonian to be learnable at the Heisenberg limit, all N_p eigenvalues of the Fisher information matrix must scale as $\Omega(T^2)$. In the subsequent sections, we will show that for Hamiltonians that obey the ETH nearly all eigenvalues of $\boldsymbol{\mathcal{I}}^{(Q)}$ are instead $\mathcal{O}(T)$, and only a sub-extensive number of the eigenvalues scale as $\Omega(T^2)$.

In this section, we set up a few preliminary definitions to capture this behavior. To maintain generality across different experimental initial states and discrete control operations, we will find it convenient to work with the vector of operators $\boldsymbol{A} = (A_1, \ldots, A_{N_p})$ defined in Eq. (V.21) instead of the quantum Fisher information matrix itself. For an initial state ρ, the vector $\boldsymbol{A}$

assuming that ETH holds within a narrow energy band $\omega_{ij} \leq 1/T_{\text{rmt}}$. Our definition has the benefit that it does not involve a sharp energy cut-off (which, when Fourier transformed, would involve integrating over times ranging up to the inverse level spacing).

upper bounds the Fisher information matrix via

$$\boldsymbol{\mathcal{I}}^{(Q)} \leq 4\langle \boldsymbol{A}\boldsymbol{A}^T\rangle_\rho, \tag{V.72}$$

which corresponds to dropping the second term in Eq. (V.22), since it is negative semidefinite. In thermalizing Hamiltonians, the aforementioned sub-extensive number of large Fisher information matrix eigenvalues will be inherited from linear dependencies among the operators A_a. We can see how this might occur with a simple example. Suppose that the operators A_a are identical for every parameter a up to a multiplicative constant, i.e. $A_a = b_a A$ for all a, or in vector form, $\boldsymbol{A} = \boldsymbol{b}A$. From Eq. (V.72) we have $\boldsymbol{\mathcal{I}}^{(Q)} \leq 4\langle A^2\rangle_\rho \boldsymbol{b}\boldsymbol{b}^T$, which is a rank-1 matrix with a single eigenvector, $\boldsymbol{b}$. While this example represents an extreme case, we will find that in thermalizing Hamiltonians, approximate linear dependencies among the A_a significantly restrict learning regardless of the initial state or discrete control operations.

We formalize the linear dependencies above via the notion of a 'low-rank' approximation of the vector of operators $\boldsymbol{A}$:

Definition V.7
Consider a vector of N_p operators, $\boldsymbol{A} = (A_1, \ldots, A_{N_p})$, and a set of R linearly independent vectors of complex numbers, $\{\boldsymbol{b}_\alpha \in \mathbb{C}^{N_p}\}_{\alpha=1,\ldots,R}$. We say that $\{\boldsymbol{b}_\alpha\}$ form a rank-R approximation of $\boldsymbol{A}$ with error δ, if $\boldsymbol{A}$ can be decomposed as follows:

$$\boldsymbol{A} = \sum_{\alpha=1}^{R} \boldsymbol{b}_\alpha B_\alpha + \boldsymbol{E}, \qquad \textit{i.e.} \qquad A_a = \sum_{\alpha=1}^{R} (b_\alpha)_a B_\alpha + E_a, \tag{V.73}$$

where $\{B_\alpha\}$ are a set of linearly independent operators, and $\boldsymbol{E}$ is a vector of operators that obeys

$$\|\boldsymbol{v}^T\boldsymbol{E}\|_s = \left\| \sum_a v_a E_a \right\|_s \leq \delta, \tag{V.74}$$

for any normalized vector $\boldsymbol{v} \in \mathbb{R}^{N_p}$, $\boldsymbol{v}^T\boldsymbol{v} = 1$. We say that the decomposition is orthonormal if the $\boldsymbol{b}_\alpha$ vectors are orthonormal.

Intuitively, the low-rank approximation captures the linearly dependencies among $A_1, \ldots, A_{N_p}$ up to deviations less than δ. From Eq. (V.72), the existence of a rank-R approximation of $\boldsymbol{A}$ immediately implies that the associated quantum Fisher information matrix has at most R eigenvalues greater than $4\delta^2$.

Note that we can apply the Gram-Schmidt procedure to orthonormalize the vectors $\boldsymbol{b}_\alpha$, by performing an appropriate linear transformation on the

operators B_α. We can also assume that the error term, $\boldsymbol{E}$, is orthogonal to the vectors $\boldsymbol{b}_\alpha$, i.e. $\boldsymbol{b}_\alpha^T \boldsymbol{E}' = 0$, without loss of generality. This follows from the following lemma:

Lemma V.5
Consider a set of vectors $\boldsymbol{A}$ with a rank-R orthonormal decomposition

$$\boldsymbol{A} = \sum_\alpha \boldsymbol{b}_\alpha B_\alpha + \boldsymbol{E}, \tag{V.75}$$

with error δ. The decomposition

$$\boldsymbol{A} = \sum_\alpha \boldsymbol{b}_\alpha B'_\alpha + \boldsymbol{E}', \qquad B'_\alpha = B_\alpha + \boldsymbol{b}_\alpha^T \boldsymbol{E}, \qquad \boldsymbol{E}' = \boldsymbol{E} - \sum_\alpha \boldsymbol{b}_\alpha \boldsymbol{b}_\alpha^T \boldsymbol{E}, \tag{V.76}$$

is also a rank-R orthonormal decomposition of $\{\boldsymbol{A}\}$ with error δ. Moreover, the latter decomposition obeys $\boldsymbol{b}_\alpha^T \boldsymbol{E}' = 0$ for all α.

Proof—One can immediately verify that Eq. (V.76) is a rank-R decomposition of $\{\boldsymbol{A}\}$, and that $\boldsymbol{b}_\alpha^T \boldsymbol{E}' = 0$. It remains to show that the adjustment $\boldsymbol{E} \to \boldsymbol{E}'$ does not increase the approximation error δ. To show this, we decompose an arbitrary vector $\boldsymbol{v}$ as $\boldsymbol{v} = \boldsymbol{v}_\parallel + \boldsymbol{v}_\perp$, where $\boldsymbol{v}_\parallel \in \text{span}\{\boldsymbol{b}_\alpha\}$ and $\boldsymbol{v}_\perp$ is orthogonal to $\text{span}\{\boldsymbol{b}_\alpha\}$. Since $\boldsymbol{b}_\alpha^T \boldsymbol{E}' = 0$, we have $\boldsymbol{v}^T \boldsymbol{E}' = \boldsymbol{v}_\perp^T \boldsymbol{E}' = \boldsymbol{v}_\perp^T \boldsymbol{E}$. Taking the infinity norm gives the desired approximation error bound, $\|\boldsymbol{v}^T \boldsymbol{E}'\|_s = \|\boldsymbol{v}_\perp^T \boldsymbol{E}\|_s \leq \delta$. □

Let us now extend the this framework to a learning protocol involving multiple experiments x, associated vectors of operators, $\boldsymbol{A}_x$, and a total quantum Fisher information matrix, $\boldsymbol{\mathcal{I}}^{(Q)} = \sum_x \boldsymbol{\mathcal{I}}_x^{(Q)}$ (see Appendix V.A for definitions of each quantity). Now, suppose that the vectors $\boldsymbol{A}_x$ in each experiment have a low-rank decomposition, and moreover, that the vectors $\boldsymbol{b}_\alpha$ are *shared* between the decompositions for each experiment. In this case, Eq. (V.72) again immediately implies that the total quantum Fisher information matrix has at most R eigenvalues greater than $4\sum_x \delta_x^2$, where the errors δ_x in each approximation may add in quadrature. This holds even if the operators $B_{\alpha,x}$ and error terms $\boldsymbol{E}_x$ vary from experiment to experiment.

With this intuition, let us extend the definition of a low-rank approximation from a single vector $\boldsymbol{A}$ to a set of vectors $\{\boldsymbol{A}_x\}$ corresponding to different experiments x.

Definition V.8
Consider a set of vectors of N_p operators, $\{\boldsymbol{A}_x\}$, and a set of R linearly independent vectors of complex numbers, $\{\boldsymbol{b}_\alpha \in \mathbb{C}^{N_p}\}_{\alpha=1,\ldots,R}$. We say that $\{\boldsymbol{b}_\alpha\}$ form a rank-R approximation of $\{\boldsymbol{A}_x\}$ with errors $\{\delta_x\}$, if $\{\boldsymbol{b}_\alpha\}$ is a

rank-R decomposition (Definition V.73) of $\boldsymbol{A}_x$ with error δ_x for each x. That is, there exists operators $B_{x,\alpha}$ and vectors of operators $\boldsymbol{E}_x$ such that

$$\boldsymbol{A}_x = \sum_\alpha \boldsymbol{b}_\alpha B_{x,\alpha} + \boldsymbol{E}_x, \tag{V.77}$$

with

$$\|\boldsymbol{v}^T \boldsymbol{E}_x\|_s = \left\| \sum_a v_a E_{x,a} \right\|_s \leq \delta_x, \tag{V.78}$$

for any normalized vector $\boldsymbol{v} \in \mathbb{R}^{N_p}$, $\boldsymbol{v}^T\boldsymbol{v} = 1$.

The preceding arguments now lead us to the following lemma:

Lemma V.6
Consider a set of quantum experiments x with unitary evolutions U_x, and define $\boldsymbol{A}_x$ for each experiment as in Eq. (V.22). Suppose that $\{\boldsymbol{b}_\alpha\}$ forms a rank-R approximation for $\{\boldsymbol{A}_x\}$ with errors $\{\delta_x\}$. Then the total RMS error of any resulting estimate of the parameters is lower bounded by

$$\epsilon \geq \frac{1}{2\delta}\sqrt{1 - R/N_p}, \tag{V.79}$$

with $\delta^2 \equiv \sum_x \delta_x^2$.

Proof—Inserting the low-rank decomposition of $\boldsymbol{A}$ into the bound in Eq. (V.72) and invoking Lemma V.5 to fix $\boldsymbol{b}_\alpha^T \boldsymbol{E}_x = 0$, we have

$$\begin{aligned}\boldsymbol{\mathcal{I}}^{(Q)} \leq 4\sum_{x=1}^{N_x} \left\langle \boldsymbol{A}_x \boldsymbol{A}_x^T \right\rangle_{\Psi_x} &\leq 4\sum_{x=1}^{N_x} \left(\sum_{\alpha,\beta} \boldsymbol{b}_\alpha \boldsymbol{b}_\beta^T \left\langle B_{x,\alpha} B_{x,\beta} \right\rangle_{\Psi_x} + \left\langle \boldsymbol{E}_x \boldsymbol{E}_x^T \right\rangle_{\Psi_x} \right) \\ &= 4\sum_{\alpha,\beta} \boldsymbol{b}_\alpha \boldsymbol{b}_\beta^T \sum_{x=1}^{N_x} \left\langle B_{x,\alpha} B_{x,\beta} \right\rangle_{\Psi_x} + \sum_{x=1}^{N_x} \left\langle \boldsymbol{E}_x \boldsymbol{E}_x^T \right\rangle_{\Psi_x}.\end{aligned} \tag{V.80}$$

Note that we are able to invoke Lemma V.5 simultaneously on all $\boldsymbol{A}_x$ as the orthogonalization with the error term only adjusts the operators $\boldsymbol{E}_x$ and $B_{x,\alpha}$ (which are unique to each $\boldsymbol{A}_x$), and not the shared vectors $\boldsymbol{b}_\alpha$. The first term has matrix rank of at most R, while the latter term obeys $\boldsymbol{v}^T \langle \boldsymbol{E}_x \boldsymbol{E}_x^T \rangle_{\Psi_x} \boldsymbol{v} \leq \delta_x^2$ for any $\boldsymbol{v} \in \mathbb{R}^{N_p}$. Summing over x, we see that $\boldsymbol{\mathcal{I}}^{(Q)}$ has at most R eigenvalues of magnitude greater than $4\delta^2$, with $\delta^2 = \sum_x \delta_x^2$. We can therefore bound the

trace of $[\boldsymbol{\mathcal{I}}^{(Q)}]^{-1}$ below by

$$\sum_a \left[\boldsymbol{\mathcal{I}}^{(Q)}\right]^{-1}_{aa} \geq \frac{N_p - R}{4\delta^2}, \tag{V.81}$$

which corresponds to taking the R eigenvalues of $\mathcal{I}$ that are greater than $4\delta^2$ to infinity, and the $N_p - R$ remaining eigenvalues to $4\delta^2$. Applying Eq. (V.27) then gives Eq. (V.79). □

We conclude this section by turning directly to Hamiltonian learning. We consider learning a time-independent Hamiltonian with L discrete quantum control operations (Definition V.2). The case of no quantum control (Definition V.1) corresponds to $L = 1$. Differentiating Eq. (V.12) via the product rule then yields

$$\boldsymbol{A}_x = -i \sum_l U_x^\dagger(t_{x,l-1}, 0) \boldsymbol{A}_H(\tau_{x,l}) U_x(t_{x,l-1}, 0), \tag{V.82}$$

$$\tag{V.83}$$

$$\boldsymbol{A}_H(t) = -i \int_0^t ds\, e^{iHs} \boldsymbol{P} e^{-iHs}, \tag{V.84}$$

where $U(t_{x,l-1}, 0) = \prod_{l'=1}^{l-1} \left(V_{l'} e^{iH\tau_{x,l'}}\right)$, $t_{x,l} = \sum_{l' \leq l} \tau_{x,l}$ is time evolution up till time $t_{x,l-1}$, and $\boldsymbol{P} = (P_1, \ldots, P_{N_p})$ is the vector of Pauli operators that enter H. Note that the operator $\boldsymbol{A}_H(\tau_{x,l})$ depends on the experiment x only through the times $\tau_{x,l}$.

We can now state the main result of this section:

Lemma V.7
Consider the Hamiltonian learning problem (Definition V.5), with discrete control of up to L interleaved unitaries. Suppose that the vectors $\{\boldsymbol{b}_\alpha\}$ form a rank-R approximation of $\boldsymbol{A}_H(t)$,

$$\boldsymbol{A}_H(t) = \sum_\alpha \boldsymbol{b}_\alpha A_\alpha(t) + \boldsymbol{E}(t) \tag{V.85}$$

with error $\delta_H = \sqrt{\mathfrak{a}t}$, for all times $t \geq 0$ with $\mathfrak{a}$ constant. Then, the maximum root-mean-square error is bounded below as

$$\epsilon \geq \frac{1}{2\sqrt{\mathfrak{a}LT}} \sqrt{1 - R/N_p}, \tag{V.86}$$

where T is the total evolution time. Hence, the parameters $\boldsymbol{u}$ cannot be learned at the Heisenberg limit if $L = o(T)$ and $R = o(N_p)$.

The above assumes that the vectors $\boldsymbol{b}_\alpha$ are time-independent, which we will find is indeed the case in the Hamiltonians we consider.

Proof of Lemma V.7—Inserting the low-rank approximation of $\boldsymbol{A}_H(\tau)$ into Eq. (V.82) gives a low-rank approximation for $\{\boldsymbol{A}_x\}$:

$$\boldsymbol{A}_x = \sum_{l,\alpha} \boldsymbol{b}_\alpha \left[U_x^\dagger(t_{x,l-1},0) A_\alpha(\tau_{x,l}) U_x(t_{x,l-1},0)\right] + \\ + \sum_l \left[U_x^\dagger(t_{x,l-1},0)\boldsymbol{E}(\tau_{x,l})U_x(t_{x,l-1},0)\right]. \quad \text{(V.87)}$$

The latter term has an operator norm of at most $\delta_x = \sum_l \sqrt{\mathfrak{a}\tau_{l,x}} \leq \sqrt{\mathfrak{a}Lt_x}$, where $t_x = \sum_l \tau_{x,l}$ is the total evolution time in measurement round x. Summing these error terms over x then yields $\delta^2 = \sum_x \delta_x^2 \leq \mathfrak{a}L \sum_x t_x = \mathfrak{a}LT$. Application of Lemma V.6 then gives Eq. (V.86). □

V.D.3. Formal statement and proof of Theorem V.9

We are now ready to state a formal version of Theorem 3 in the main text.

Theorem V.9
Consider the learning problem (Definition V.5) for an N-qubit Hamiltonian without continuous quantum control [Eq. (V.10)*] and up to L interleaved unitaries in the discrete quantum control model. Suppose that, within an inverse temperature range $[\beta_1, \beta_2]$, the Hamiltonian $H(\mathbf{u})$: (i) obeys the eigenstate thermalization hypothesis as in Definition V.6, and (ii) has thermal expectation values with bounded derivatives with respect to the inverse temperature*

$$\left|\frac{\partial_\beta^m \langle P_a\rangle_\beta}{m!}\right| \leq (m^\gamma B)^m, \qquad \gamma, B = \mathcal{O}(1) \quad \text{(V.88)}$$

for all operators P_a in the Hamiltonian. Then the root-mean-square error in the parameters $\mathbf{u}$ is lower bounded by

$$\epsilon \geq \frac{1}{2\sqrt{F^* LT}} \sqrt{1 - \frac{2B|\beta_2 - \beta_1| \log_2\left(\sqrt{N_p T/F^*}\right)^{\gamma+1}}{N_p}}, \quad \text{(V.89)}$$

for any learning protocol that can only access states within $[\beta_1, \beta_2]$. Here, $F^* = \mathrm{poly}(N)$ is constant in t, and the bound holds whenever the term inside the square root is positive. Thus, if $L = o(T)$ and $N_p = \Omega(N^c)$ for some $c > 0$, then the Heisenberg limit cannot be achieved for $T = o\big(\exp\big(N^{2c/(\gamma+1)}\big)\big)$.

In many-body Hamiltonian learning, the number of unknown parameters is

typically at least linear in the system size, $N_p = \Omega(N)$. In this case, the above theorem precludes Heisenberg-limited learning until times exponential in the system size. [See Appendix V.D.7 for a discussion of learning in thermalizing Hamiltonians with only a constant number of unknown parameters, $N_p = \mathcal{O}(1)$.] The required bound on thermal derivatives is satisfied by most physical systems away from finite temperature phase transitions (where the constant B diverges) [338]. In Appendix V.D.6, we leverage previous results [319, 320] to prove that this bound holds for all local one-dimensional Hamiltonians at any finite temperature, and local higher-dimensional Hamiltonians at sufficiently high temperatures.

We note that our lower bound contains a prefactor $1/\sqrt{F^*} \sim 1/\mathrm{poly}(N)$, and thus becomes weaker as the system size increases. This prefactor arises from our lenient definition of thermalization: We assume only that connected correlation functions decay to zero after some time polynomial in N, and place no restriction on their behavior before this time (see the second condition in Definition V.6). We expect that the prefactor can be improved by placing more stringent assumptions on the early-time behavior.

Proof of Theorem V.9—The bulk of the proof is relegated to the following two sections. Here we show how the results of these sections, Lemmas V.8 and V.9, lead to Theorem V.9.

We begin by using the eigenstate thermalization hypothesis (Definition V.6) to decompose the term $\boldsymbol{A}_H(t)$ in Eq. (V.82) into a sum of three components,

$$\boldsymbol{A}_H(t) = \int_0^t ds\, \boldsymbol{P}(s) = \boldsymbol{A}_{\mathrm{d}}(t) + \boldsymbol{A}_{\mathrm{od}}(t) + \delta\boldsymbol{A}(t), \tag{V.90}$$

where $\boldsymbol{P}$ is the vector of N_p Pauli operators that appear in the Hamiltonian. For each term above, we set all matrix elements outside of the energy range specified by $[\beta_1, \beta_2]$ to zero, which is allowed by assumption. The first term corresponds to the diagonal component in ETH,

$$A_{\mathrm{d},a}(t) = t \sum_i \langle P_a \rangle_{\beta(E_i)} |E_i\rangle\langle E_i| \,. \tag{V.91}$$

The second term corresponds to the off-diagonal component in ETH,

$$A_{\mathrm{od},a}(t) = \sum_{i<j} \left(e^{-S(E_{ij})/2} f^a(E_{ij}, \omega_{ij}) \frac{e^{i\omega_{ij}t} - 1}{i\omega_{ij}} r_{ij}^a \right) |E_i\rangle\langle E_j| + h.c., \tag{V.92}$$

where r_{ij}^a are i.i.d. normal random complex numbers with unit variance. Here we have performed the time integral, $\int_0^t ds\, e^{i\omega s} = (e^{i\omega t} - 1)/(i\omega)$. Finally, the third term is given by $\delta A_a(t) = t \cdot \delta P_a(t)$, where $P_a(t)$ quantifies the

deviation from ETH behavior in Definition V.6. This term is upper bounded by assumption from Definition V.6,

$$\|\boldsymbol{w}^T \delta \boldsymbol{A}(t)\|_s \leq C\sqrt{N_p}, \quad \boldsymbol{w}^T \boldsymbol{w} = 1, \tag{V.93}$$

where $C = \mathcal{O}(\text{poly}(N))$ is defined via Definition V.6 and is independent of t. This implies that the third term cannot contribute to parameter learning at the Heisenberg limit.

In Lemma V.8 of the following section, we show that the second term can be upper bounded using techniques from random matrix theory,

$$\|\boldsymbol{w}^T \boldsymbol{A}_{\text{od}}(t)\|_s \leq \sqrt{Ft}, \quad \boldsymbol{w}^T \boldsymbol{w} = 1, \tag{V.94}$$

where $F = \mathcal{O}(\text{poly}(N))$ is a constant independent of t (see Lemma V.8 for a definition). This implies that the second term cannot contribute to parameter learning at the Heisenberg limit.

In Lemma V.9 of Appendix V.D.5, we show that the first term, $\boldsymbol{A}_{\text{d}}(t)$, possesses a low-rank approximation with error δ composed of $R(\delta, t) = 2B|\beta_2 - \beta_1| \log_2(t\sqrt{N_p}/\delta)^{\gamma+1}$ *time-independent* vectors[5]. Combining this fact with the above bounds on the second and third term, we see that the entire operator vector $\boldsymbol{A}_H(t)$ possesses the same low-rank approximation with error

$$\delta^* = \delta + \sqrt{Ft} + \sqrt{N_p}C. \tag{V.95}$$

Since the latter two terms are $\mathcal{O}(\sqrt{t})$, there exists a constant F^* such that $\sqrt{F^* t/4} > \sqrt{Ft} + \sqrt{N_p}C$ for all $t \geq F$. We note that $F^* = \text{poly}(N)$ since $F = \text{poly}(N)$ and $N_p = \text{poly}(N)$ (the latter follows because there are at most $\binom{N}{k}3^k = \mathcal{O}(N^k)$ possible Pauli terms in a k-local Hamiltonian).

The results of the above paragraph demonstrate that $\boldsymbol{A}_H(t)$ satisfies the conditions of Lemma V.7, with error $\delta^* = \sqrt{F^* t}$ and rank $R(\delta^*/2, t)$. Invoking Lemma V.7 and inserting the definition of $R(\delta, t)$ above yields Eq. (V.89).

One can confirm that as long as Eq. (V.89) holds, the Heisenberg limit cannot be achieved unless L scales at least linearly in T. It remains to consider what would be required for this inequality to break down; i.e. for the term within the square root to become negative. Some quick algebra shows that the inequality holds whenever

$$\log_2(T) \leq \left[\frac{N_p}{2^{-\gamma} B|\beta_2 - \beta_1|}\right]^{\frac{1}{\gamma+1}} + \log_2(F^*) - \log_2(N_p). \tag{V.96}$$

[5] By 'time-independent' we mean e.g. that each vector used in the low-rank approximation of $\boldsymbol{A}_d(t)$ appears in the low-rank approximation of $\boldsymbol{A}_d(2t)$. However, the low-rank approximation of $\boldsymbol{A}_d(2t)$ will contain $\mathcal{O}(1)$ additional vectors than that of $\boldsymbol{A}_d(t)$.

We recall that F^* is constant in t and polynomial in N. Thus, if the number of unknown parameters scales with any non-zero power of the system size, $N_p = \Omega(N^c)$ with $c > 0$, then term in the square root is positive up to times exponentially large in a non-zero power of the system size, $T = o(\exp(N^{c/(\gamma+1)}))$. □

V.D.4. Upper bounding the spectral norm of the off-diagonal component

In this section, we bound the operator norm of the off-diagonal matrix elements via the following Lemma.

Lemma V.8
The vector of operators $\boldsymbol{A}_{\text{od}}$ defined in Eq. (V.70) obeys the inequality

$$\|\boldsymbol{w}^T \boldsymbol{A}_{\text{od}}\|_s \leq \sqrt{Ft}, \tag{V.97}$$

for any $\boldsymbol{w} \in \mathbb{R}^{N_p}$ with $\boldsymbol{w}^T\boldsymbol{w} = 1$, with probability double exponentially close to one in the system size. Here F is defined via

$$F = 8\sqrt{2} \max_i \int_{-t}^{t} ds\ (1 - |s|/t) \cdot G^c(s; E_i), \tag{V.98}$$

where $G^c(s; E_i)$ is the connected auto-correlation function [Eq. (V.69)] of the operator $\boldsymbol{w}^T\boldsymbol{P}$ in the energy eigenstate $|E_i\rangle$ at time s.

By condition 2 of Definition V.6 [Eq. (V.69)], we have $F = \mathcal{O}(\text{poly}(N))$.

Proof of Lemma V.8—Our proof uses several basic results borrowed from random matrix theory [339]. Let us begin by defining, for a given vector $\boldsymbol{w}$, the time-integrated perturbation operator $A \equiv \boldsymbol{w}^T\boldsymbol{A}_{od}$ and the associated sum of Pauli operators $P \equiv \boldsymbol{w}^T\boldsymbol{P}$. The off-diagonal matrix elements of A are normal random complex variables with mean zero and variance,

$$v_{ij} \equiv \mathbb{E}\left[|A_{ij}|^2\right] = e^{-S(E_{ij})}|f(E_{ij}, \omega_{ij})|^2 \frac{\sin(\omega_{ij}t/2)^2}{\omega_{ij}^2}, \tag{V.99}$$

where $|f|^2 = \sum_a |w_a|^2|f_a|^2$, and f_a describe the off-diagonal matrix elements of the individual Pauli operator P_a. The matrix A is thus drawn from an ensemble with probability measure dP given by

$$dP(A) = \left(\prod_{i<j} \frac{dA_{ij}}{\sqrt{2\pi v_{ij}}} \exp\left(-|A_{ij}|^2/2v_{ij}\right)\right). \tag{V.100}$$

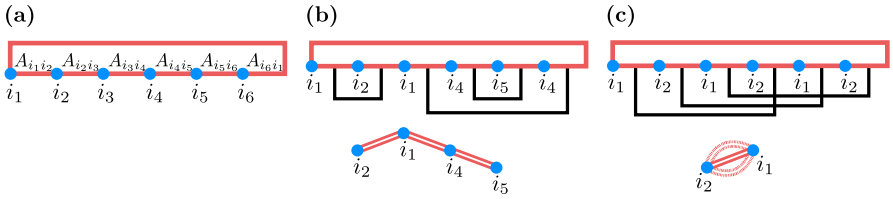


Figure V.3.: **(a)** Schematic of matrix multiplications and Wick contractions for evaluating $\mu'_{2m} = \mathrm{Trace}(A^{2m})$ with $m = 3$. Blue dots denote energy indices i_r for $r = 1, \ldots, 6$ which are summed over. Red lines denote matrix elements $A_{i_r i_{r+1}}$. **(b)** An example of a non-crossing Wick contraction. Top: The Wick contraction (black lines) pairs matrix elements $A_{i_p i_{p+1}}$ and $A_{i_q i_{q+1}}$ by enforcing the constraints $i_{p+1} = i_q$, $i_{q+1} = i_p$. For non-crossing pairings, these constraints leave $m+1$ independent indices to sum over (in this example, indices i_1, i_2, i_4, i_5). Bottom: The graph G defined by identifying all blue dots with the same index. Edges in G are denoted with a double line to signify that they carry a factor of the matrix element squared, $|A_{ij}|^2$. Since the contraction is non-crossing, the graph is a tree (see text for further discussion). **(c)** An example of a Wick contraction with crossings, with genus $g = 1$. Top: The constraints leave $m + 1 - 2g$ independent indices to sum over (in this example, indices i_1, i_2). Bottom: The graph G defined by identifying all blue dots with the same index. Edges that are not part of the spanning tree T (see text) are denote with striped double lines.

Note that the variance depends on the element indices i, j. The Gaussian unitary ensemble typically is defined via $v_{ij} = v/d$, where v is an arbitrary scaling factor and d is the matrix dimension.

Following established techniques in random matrix theory [339], the expected moments of A,

$$\mu'_{2m} \equiv \mathbb{E}\left[\mathrm{Trace}(A^{2m})\right] = \sum_{i_1,\ldots,i_{2m}} \mathbb{E}\left[A_{i_1 i_2} A_{i_2 i_3} \ldots A_{i_{2m} i_1}\right], \tag{V.101}$$

can be calculated as sums over Wick contractions:

$$\mu'_{2m} = \sum_{\pi} c_\pi = \sum_{\pi} \sum_{i_1,\ldots,i_{2m}} \left(\prod_{(p,q)\in\pi} \mathbb{E}\left[|A_{i_p i_q}|^2\right] \cdot \delta_{i_p, i_{q+1}} \cdot \delta_{i_q, i_{p+1}} \right), \tag{V.102}$$

where the contribution c_π of an individual Wick contraction π is defined via the second equality. See Fig. V.3 for a visual representation of the matrix multiplication and Wick contractions. Each contraction corresponds to a

partition π of the $2m$ copies of A into m pairs. Each pair $(p,q) \in \pi$ contributes a factor of $\mathbb{E}\left[|A_{i_p i_q}|^2\right] = v_{i_p i_q}$ to the contraction. Meanwhile, the matrix multiplications and trace enforce the conditions $\delta_{i_p, i_{q+1}}$ and $\delta_{i_q, i_{p+1}}$ for each p, q. Here, the addition should be performed modulo $2m$, to incorporate the condition δ_{i_{2m}, i_1} enforced by the trace.

In what follows, we will upper bound the sum over all Wick contractions using techniques from random matrix theory. We refer to [339] for a comprehensive introduction to the relevant random matrix techniques.

We begin by discussing only the Wick contractions that contribute to leading order in the inverse density of states, e^{-S}. The dominant Wick contractions at leading order in e^{-S} correspond to *non-crossing* pairings. If one orders the $2m$ operators, $A_{i_1 i_2}, \ldots, A_{i_{2m} i_1}$, in a circle and draws lines between them for each contraction, a non-crossing pairing contains no lines that cross [Fig. V.3(b)]. Such a pairing contains 'innermost pairs', where A_{i_{r-1}, i_r} is paired with $A_{i_r, i_{r+1}}$ [for example, $r = 2, 5$ in Fig. V.3(b)]. An innermost pair in a non-crossing pairing contributes a factor

$$\sum_j \mathbb{E}\left[|A_{ij}|^2\right] = \sum_j v_{ij}, \tag{V.103}$$

to the entire Wick contraction. This factor can, in fact, be re-expressed in terms of the connected auto-correlation function [Eq. (V.69)] as follows:

$$\begin{aligned} \sum_j v_{ij} &= \int dE_j \, e^{S(E_j) - S(E_{ij})} |f(E_{ij}, \omega_{ij})|^2 \cdot \frac{\sin(\omega_{ij} t/2)^2}{\omega_{ij}^2} \\ &= \sum_j e^{-S(E_{ij})} |f(E_{ij}, \omega_{ij})|^2 \cdot \int_0^t dt_1 \int_0^t dt_2 \, e^{i\omega_{ij}(t_1 - t_2)} \\ &= t\sqrt{2} \int_{-t}^t ds \left(1 - \frac{|s|}{t}\right) \cdot G^c(s; \beta(E_i)), \end{aligned} \tag{V.104}$$

where the third line follows from re-arranging the integration coordinates and applying the definition of the connected auto-correlation function [Eq. (V.69)]. The sum over v_{ij} is therefore upper bounded by

$$\sum_j v_{ij} \leq \frac{F t}{8} \tag{V.105}$$

from the definition of F [Eq. (V.98)]. The product $Ft/8$ serves as the scaling factor of a comparable Gaussian unitary ensemble [see discussion under Eq. (V.100)].

Using the above lower bound, we can remove the sums over all indices that lie in the center of an innermost pairing from Eq. (V.101). This leaves a non-crossing pairing on the remaining indices [for example, $r = 1, 4$ remain after upper bounding innermost pairs in Fig. V.3(b)]. We can successively iterate the above procedure to bound the total contribution of the original non-crossing pairing by $d(Ft)^m$, where the factor of the Hilbert space dimension, d, arises from the final index summation [339]. Summing over all non-crossing pairings gives the following upper bound on the leading order contribution to the $2m^{\text{th}}$ moment:

$$(\mu'_{2m})_{\text{leading order}} \leq d\, C_m (Ft)^m, \tag{V.106}$$

where d is again the matrix dimension, and C_m is total number of non-crossing pairings of $2m$ elements, also known as the m^{th} Catalan number [339]. The right hand side is also equal to the $2m^{\text{th}}$ moment of the Gaussian unitary ensemble with scaling factor $Ft/8$.

We now explicitly bound the sum over all Wick contractions. Consider the contribution of a pairing π. The independent energy indices that are summed over are determined by π as well as the ordering of the operators in the trace. This results in a sum over $k_\pi = m + 1 - 2g_\pi$ independent energy indices, where the integer g_π is known as the *genus* of the pairing π [339]. To proceed, let us associate the k_π independent energy indices with vertices in a graph G, which are connected by edges corresponding to the matrices $A_{i_n, i_{n+1}}$ [see bottom of Fig. V.3(b,c)]. Note that G is necessarily connected, as G can be viewed as a circular graph with some vertices identified. The contribution of the pairing π can now be written in the form

$$\begin{aligned} c_\pi &= \left(\prod_{r \in V(G)} \sum_{i_r} \right) \prod_{(p,q) \in E(G)} e^{-S(E_{i_p i_q})} |f_{i_p i_q}|^2 \frac{\sin\left(\omega_{i_p i_q} t/2\right)^2}{\omega_{i_p i_q}^2} \\ &= \left(\prod_{r \in V(G)} \int dE_{i_r}\, e^{S(E_{i_r})} \right) \prod_{(p,q) \in E(G)} e^{-S(E_{i_p i_q})} |f_{i_p i_q}|^2 \frac{\sin\left(\omega_{i_p i_q} t/2\right)^2}{\omega_{i_p i_q}^2}, \end{aligned} \tag{V.107}$$

where we denote the vertices and edges of the graph G by $V(G)$ and $E(G)$, respectively. Each vertex in G contributes a sum over its associated energy index, and each edge in G contributes a factor of $\mathbb{E}[|A_{i_p i_q}|^2]$.

To bound Eq. (V.107), we consider a spanning tree T of G, which will contain k_π vertices and $k_\pi - 1$ edges, since each edge in T can be uniquely associated with its child vertex, which leaves only the root vertex without an

associated edge. There are $2g_\pi$ edges that are in G but not in T. Now, for each such edge (p, q), we apply the inequality

$$e^{-S(E_{i_p i_q})}|f_{i_p i_q}|^2 \frac{\sin\left(\omega_{i_p i_q} t/2\right)^2}{\omega_{i_p i_q}^2} \leq \frac{\mathfrak{m} t}{2}, \tag{V.108}$$

which is derived by applying the inequality, $\sin^2(x)/x^2 \leq 1$, and defining the maximum average squared off-diagonal matrix element,

$$\mathfrak{m} = \max_{ij} e^{-S(E_{ij})}|f_{ij}|^2 = \mathcal{O}(\exp(-N)). \tag{V.109}$$

Note that $\mathfrak{m}$ is exponentially small in the system size due to the factor of e^{-S}. This leaves the following expression:

$$c_\pi \leq \left(\tfrac{\mathfrak{m} t}{2}\right)^{2g_\pi} \left(\prod_{r \in V(T)} \int dE_{i_r}\, e^{S(E_{i_r})}\right) \prod_{(p,q)\in E(T)} e^{-S(E_{i_p i_q})}|f_{i_p i_q}|^2 \frac{\sin\left(\omega_{i_p i_q} t/2\right)^2}{\omega_{i_p i_q}^2}, \tag{V.110}$$

where the RHS differs from Eq. (V.107) because the product is now over edges in T and not G. We recall $V(T) = V(G)$ since T is spanning.

The benefit of the above expression is that every edge term is uniquely associated with a vertex integral because T is a tree. Namely, each edge can be associated with its child vertex. Let us denote p as the parent vertex and q the child vertex for each edge $(p, q) \in E(T)$ Now, note that if q is a leaf of T (i.e. it has no children of its own), then the entire dependence of expression Eq. (V.110) on i_q is contained in the integral

$$\int_{-\infty}^{\infty} dE_{i_q}\, e^{S(E_{i_q}) - S(E_{i_p i_q})}|f_{i_p i_q}|^2 \frac{\sin\left(\omega_{i_p i_q} t/2\right)^2}{\omega_{i_p i_q}^2} \leq \frac{Ft}{8}, \tag{V.111}$$

which we have already upper bounded in Eq. (V.104). Applying this upper bound to each leaf of T produces an expression analogous to Eq. (V.104), but where each leaf vertex of T is deleted and replaced a factor of $Ft/8$. We can then repeat this procedure until the tree is reduced to a single root vertex, at which point the only integral that remains, $\int dE_{i_R} e^{S(E_{i_R})}$, evaluates to the Hilbert space dimension d. This gives

$$c_\pi \leq d \left(\frac{Ft}{8}\right)^m \left(\frac{\mathfrak{m} t}{2}\right)^{2g_\pi}, \tag{V.112}$$

which is our final bound on c_π.

We can now bound the sum over all pairings. We have

$$\mu'_{2m} = \sum_{\pi} c_\pi \leq d\,(Ft/8)^m \sum_{g\geq 0} \varepsilon_g(m)\, w^{-2g}, \tag{V.113}$$

where $\varepsilon_g(m)$ is the number of pairings of $2m$ elements with genus g. Here we define

$$w = \left(\frac{\mathfrak{m}t}{2}\right)^{-1}. \tag{V.114}$$

The above sum is familiar in random matrix theory and is upper bounded by [339]

$$\sum_{g\geq 0} \varepsilon_g(m)\, w^{-2g} \leq C_m \exp\!\big(m^3/2w^2\big), \tag{V.115}$$

where C_m are the Catalan numbers. Combining with Eq. (V.113), we have our final bound on μ'_{2m}

$$\mu'_{2m} \leq d\,(Ft/8)^m C_m \exp\!\big(m^3/2w^2\big) < d\,(Ft/2)^m \exp\!\big(m^3/2w^2\big), \tag{V.116}$$

where in the latter inequality we use $C_m < 4^m/(m+1)\sqrt{\pi m} < 4^m$ [340]. Note that this is equal to the contribution of the leading order Wick contractions [Eq. (V.106)] up to an overall factor, $\exp\!\big(m^3/2w^2\big)$, which is near unity.

We can now show that the probability that A has a maximum eigenvalue that scales as t is exponentially suppressed in the effective Hilbert space dimension. First, note that the $2m^{\text{th}}$ moment of a matrix A provides an upper bound on the maximum eigenvalue, $\lambda_{\max}^{2m} \leq \text{Trace}(A^{2m})$. Now consider the probability that $\lambda_{\max}$ takes value greater than $\sqrt{Ft}$ for some constant $\mathfrak{c}$. We have

$$P\left[\lambda_{\max} > \mathfrak{c}\sqrt{Ft/2}\right] \leq P\left[\text{Trace}(A^{2m}) > (\mathfrak{c}t)^{2m}\right] \tag{V.117}$$

$$\leq \frac{\mathbb{E}\left[\text{Trace}(A^{2m})\right]}{(\mathfrak{c}t)^{2m}} \tag{V.118}$$

$$< d\left(\frac{1}{\mathfrak{c}^2}\right)^m \exp\!\big(m^3/2w^2\big), \tag{V.119}$$

where in the second step we use Markov's inequality. Taking $m = w^{2/3} = \mathcal{O}(\exp(N))$, we find that the probability is doubly exponentially suppressed in N whenever $\mathfrak{c} > 1$,

$$P\left[\lambda_{\max} > \mathfrak{c}\sqrt{Ft/2}\right] \leq d\left(\frac{1}{\mathfrak{c}^2}\right)^{w^{2/3}} \sqrt{e}. \tag{V.120}$$

Note that the prefactor d grows singly exponentially in the system size, so the probability is dominated by the doubly exponential suppression. Taking $\mathfrak{c} = \sqrt{2}$ gives Lemma V.8. □

V.D.5. Upper bounding the rank of the diagonal component

The goal of this section is to establish the following Lemma.

Lemma V.9
Consider the setting in Definition V.6, and define the vector of operators $\boldsymbol{A}_d(t)$ as in Eq. (V.91). Then $\boldsymbol{A}_d(t)$ possesses a rank-R approximation with error δ (Definition V.7), with R defined via

$$R(\delta, t) \leq 2B|\beta_2 - \beta_1| \log_2(t\sqrt{N_p}/\delta)^{\gamma+1}. \tag{V.121}$$

Here B, β_1, β_2, and γ are constants defined in Theorem V.9.

Proof of Lemma V.9—We recall that the component operators of $\boldsymbol{A}_d$ take the form:

$$\boldsymbol{A}_{\mathrm{d}}(t) = t \sum_i \langle \boldsymbol{P} \rangle_{\beta(E_i)} |E_i\rangle\langle E_i| . \tag{V.122}$$

The central idea of our proof is to exploit the bounded derivatives of $\langle P_a \rangle_\beta$ with respect to β [i.e. the assumption Eq. (V.88)] to express $\boldsymbol{A}_d$ as a sum of a low number of vectors.

We do so by Taylor expanding. Specifically, we first divide the inverse temperature range $[\beta_1, \beta_2]$ into $|\beta_2 - \beta_1|/\Delta$ windows of width Δ. We then form an approximation, $\boldsymbol{Q}(\beta; n, \Delta)$, that is equal to the order-$(n-1)$ Taylor expansion of $\langle \boldsymbol{P} \rangle_\beta$ within each window

$$\boldsymbol{Q}(\beta; n, \Delta) \equiv \langle \boldsymbol{P} \rangle_{\beta_j} + \ldots + \frac{(\beta - \beta_j)^{n-1}}{(n-1)!} \partial_\beta^{n-1} \langle \boldsymbol{P} \rangle_{\beta_j} , \qquad \beta \in [\beta_j, \beta_{j+1}) \tag{V.123}$$

where $\beta_j = \beta_1 + j\Delta$ for $j \in \mathbb{N}$. The approximation carries two free parameters, Δ and n. According to the remainder theorem, errors in the approximation are suppressed exponentially in n:

$$|\langle P_a \rangle (\beta) - Q_a(\beta; n, \Delta)| \leq \Delta^n \max_\beta \left[\frac{\partial_\beta^n \langle P_a \rangle}{n!} \right] \leq (n^\gamma B \Delta)^n , \tag{V.124}$$

and therefore,

$$\boldsymbol{w}^T \left(\langle \boldsymbol{P} \rangle (\beta) - \boldsymbol{Q}(\beta; n, \Delta) \right) \leq \sqrt{N_p} \, (n^\gamma B \Delta)^n , \tag{V.125}$$

for any normalized vector $\boldsymbol{w}^T\boldsymbol{w} = 1$, where N_p is the number of unknown parameters.

Putting the above results together, we have the following low-rank approximation for $\boldsymbol{A}_{\mathrm{d}}(t)$:

$$\boldsymbol{A}_{\mathrm{d}}(t) = t \sum_{j=1}^{\frac{|\beta_2-\beta_1|}{\Delta}} \sum_{m=0}^{n-1} \partial_\beta^m \langle \boldsymbol{P} \rangle_{\beta_j} \cdot \left(\sum_{E_i=E(\beta_j)}^{E(\beta_{j+1})} \frac{(\beta(E_i)-\beta_j)^m}{m!} |E_i\rangle\langle E_i| \right) + \boldsymbol{E}_{\mathrm{d}}, \tag{V.126}$$

where $\boldsymbol{E}_{\mathrm{d}}$ is the approximation error (see below). Note that the left term in the summand is a vector, while the right term is an operator. The above expression thus matches Definition V.7 of a low-rank approximation, with vectors $\{\boldsymbol{b}_\alpha\} \to \{\partial_\beta^m \langle \boldsymbol{P} \rangle (\beta_j)\}$ and operators $\{B_\alpha\}$ equal to the term in parentheses in Eq. (V.126). The approximation has rank

$$\dim\left(\mathrm{span}\{\partial_\beta^m \langle \boldsymbol{P} \rangle (\beta_i)\}\right) \leq \frac{n|\beta_2-\beta_1|}{\Delta}. \tag{V.127}$$

The error in our approximation is

$$\boldsymbol{E}_{\mathrm{d}} = t \sum_i \left(\langle \boldsymbol{P} \rangle_{\beta(E_i)} - \boldsymbol{Q}(\beta_i; n, \Delta) \right) |E_i\rangle\langle E_i| \,. \tag{V.128}$$

To quantify the error as in Definition V.7, consider any normalized vector $\boldsymbol{w}$. The operator $\boldsymbol{w}^T\boldsymbol{E}_{\mathrm{d}}$ is diagonal in the energy eigenbasis, with matrix elements bounded in magnitude by $t\sqrt{N_p}(n^\gamma B\Delta)^n$ [see Eq. (V.125)]. We thus have $\|\boldsymbol{w}^T\boldsymbol{A}_d\|_s \leq t\sqrt{N_p}(n^\gamma B\Delta)^n$. The collection of vectors $\{\partial_\beta^m \langle \boldsymbol{P} \rangle_{\beta_i}\}$ therefore forms a rank-$(n|\beta_1-\beta_2|/\Delta)$ approximation of $\boldsymbol{A}_{\mathrm{d}}(t)$ with error $\delta = t\sqrt{N_p}(n^\gamma B\Delta)^n$.

The above bound holds for any choice of parameters n and Δ. To conclude, we make the particular choice $\Delta = (2n^\gamma B)^{-1}$. This provides a low-rank approximation of $\boldsymbol{A}_{\mathrm{d}}(t)$ comprised of

$$R(\delta, t) \leq 2Bn^{\gamma+1}|\beta_2-\beta_1| \tag{V.129}$$

vectors, for $\delta = t\sqrt{M}/2^n$. Solving for $n = \log_2(Mt/\delta)$ and substituting into the above expression, we have our final bound,

$$R(\delta, t) \leq 2B|\beta_2-\beta_1| \log_2(t\sqrt{M}/\delta)^{\gamma+1}, \tag{V.130}$$

which holds for any δ. □

V.D.6. Bound on derivatives of expectation values in local Hamiltonians

In Theorem V.9, we require that the derivatives of thermal expectation values with respect to the inverse temperature are bounded above by constants independent of the system size. This behavior is widely observed in physical many-body systems and quantum field theories. In this section, we prove that this bound is obeyed by any local Hamiltonian in D-dimensions that exhibits an exponential decay of two-point correlation functions.

We begin by defining the exponential decay of connected two-point correlation functions as in [319]:

Definition V.9
We say that the two-point correlation functions of a set of operators $\mathcal{O}$ decay exponentially if

$$|\langle AB\rangle_\beta - \langle A\rangle_\beta\langle B\rangle_\beta| < a_\beta \|A\|_s \|B\|_s \exp(-d(A,B)/\xi_\beta) \tag{V.131}$$

for all $A, B \in \mathcal{O}$. Here $d(A,B)$ is the Euclidean distance between the support of A and the support of B, and a_β, ξ_β are constants that depend only on β (as in [319]).

Exponential decay of correlations has been proven for all operators in one-dimensional systems with finite-range interactions at any non-zero temperature [320], and for operators with bounded spatial support in d-dimensional systems with finite-range interactions above a critical temperature [319].

The aim of this section is to establish the following Lemma.

Lemma V.10
Consider a Hamiltonian on a D-dimensional cubic lattice with range r and degree κ. The exponential decay of two-point correlation functions as in Definition V.9 implies that the n^{th} derivative of thermal expectation value is bounded,

$$\left|\partial_\beta^n \langle A\rangle_\beta\right| \leq c_\beta \|A\|_s \left(n^{D+2}\left(\frac{2D}{e^{1/2\xi_\beta}-1}\right)^D \frac{\kappa J}{\log(2)}\right)^n, \tag{V.132}$$

for any observable A with range r. Here we define $c_\beta = 4a_\beta \mathfrak{b}\, e^{\lceil r/2\rceil/\xi_\beta}$, with $a_\beta, \xi_\beta, \mathfrak{b}$ as in Definition V.9. Note that the bound is independent of the system size.

In the above, we define the range, r, of a Hamiltonian to be the maximum Euclidean distance between two qubits acted on by the same Hamiltonian

term. We also define the degree, κ, of a Hamiltonian to be the maximum number of terms acting on a single qubit.

Before proceeding to the proof of Lemma V.10, let us first establish some basic connections between correlation functions and the derivatives of thermal expectation values. The first derivative of the thermal expectation value gives

$$\partial_\beta \langle A \rangle_\beta = \partial_\beta \left[\frac{\mathrm{Trace}(Ae^{-\beta H})}{\mathrm{Trace}(e^{-\beta H})} \right] = -\langle AH \rangle_\beta + \langle A \rangle_\beta \langle H \rangle_\beta, \tag{V.133}$$

which is precisely equal to the connected two-point correlation function in Definition V.9. The second derivative gives

$$\partial_\beta^2 \langle A \rangle_\beta = \langle AH^2 \rangle_\beta - 2\langle AH \rangle_\beta \langle H \rangle_\beta - \langle A \rangle_\beta \langle H^2 \rangle_\beta + 2\langle A \rangle_\beta \langle H \rangle_\beta \langle H \rangle_\beta, \tag{V.134}$$

which is some particular connected three-point correlation function between A and two copies of H. More systematically, for the n^{th} derivative, we can calculate

$$\partial_\beta^n \langle A \rangle_\beta = C_\beta^{n+1}(A, H, \ldots, H) \tag{V.135}$$

where we define the *connected m-point correlation function* $C_\beta^m(O_1, \ldots, O_m)$,

$$C_\beta^m(O_1, \ldots, O_m) = -\sum_{P \in \mathcal{P}_m} (-1)^{m+|P|} \cdot (|P|-1)! \cdot \prod_{\sigma=1}^{|P|} \langle \prod_{i \in P_\sigma} O_i \rangle_\beta \tag{V.136}$$

Here the sum is over partitions, P, of the first m positive integers, $\{1, \ldots, m\}$. We denote the set of such partitions as $\mathcal{P}_m$. A partition, $P = \{P_1, \ldots, P_{|P|}\}$, is specified by $|P|$ disjoint blocks, i.e. sets P_σ for $\sigma = 1, \ldots, |P|$, such that the union of the P_σ comprises the full set, i.e. $\cup_{\sigma=1}^{|P|} P_\sigma = \{1, \ldots, m\}$. For example, $P = \{\{0, 3\}, \{2, 4, 5\}, \{1\}\}$ is a partition of $\{0, 1, 2, 3, 4, 5\}$ with $|P| = 3$ blocks. The number of partitions of m elements into k blocks is given by the Stirling numbers of the second kind, denoted $\mathfrak{s}(m, k)$. For our purposes, the ordering of operators within the expectation values in Eq. (V.136) does not matter, since the thermal density matrix ρ and all copies of H commute.

With this formula in hand, we now establish a simple Lemma, which shows that the exponential decay of two-point correlation functions implies exponential decay of n-point correlation functions.

Lemma V.11
The exponential decay of two-point correlation functions as in Definition V.9,

implies exponential decay of n-point correlation functions

$$|C_\beta^{n+1}(O_1,\dots,O_{n+1})| \leq a_\beta \mathfrak{b} \frac{n!}{\log(2)^n} \Big(\prod_{i=1}^{n+1} \|O_i\|_s \Big) e^{-d(O_1,\dots,O_{n+1})/\xi_\beta} \quad \text{(V.137)}$$

for all $O_1,\dots,O_{n+1} \in \mathcal{O}$. Here,

$$d(O_1,\dots,O_n) = \max_i \min_j d(O_i,O_j), \quad \text{(V.138)}$$

$\mathfrak{b}$ is constant, and a_β, ξ_β are constants from Definition V.9 which depend only on β.

In local Hamiltonians, we do not expect the above bound to be tight. In particular, we conjecture that the prefactor of $n!$ can be eliminated, and that the distance in the exponential decay can be strengthened (previous works have speculated that the relevant distance is equal to the length of the minimal Euclidean Steiner tree between $O_1,\dots,O_{n+1}$ [341]). Nevertheless, the above bound is sufficient for our purposes, and is convenient since the exponential decay of two-point correlation functions has been rigorously proven in several cases [319, 320]. Looking forward, we expect that tighter bounds above would improve the exponent γ in Theorem V.9.

We can now proceed to the proof of Lemma V.11.

Proof of Lemma V.11— We first re-write the $n+1$-point correlation function as a sum of connected two-point correlation functions

$$\begin{aligned} C_\beta^{n+1}(O_1,\dots,O_{n+1}) = \\ = \sum_{P\in\mathcal{P}_n} (-1)^{n+|P|}(|P|-1)! \sum_{\tau=1}^{|P|} \left(\prod_{\sigma\neq\tau} \langle \prod_{i\in P_\sigma} O_i \rangle_\beta \right) \cdot \\ \cdot \left(\langle O_{n+1} \prod_{i\in P_\tau} O_i \rangle_\beta - \langle O_{n+1} \rangle_\beta \langle \prod_{i\in P_\tau} O_i \rangle_\beta \right) \end{aligned} \quad \text{(V.139)}$$

Note that the first sum is over partitions of the first n positive integers (not $n+1$), and that the final factor is equal to a connected two-point correlation function between O_{n+1} and $\prod_{i\in P_\tau} O_i$. Also, while we have written the above expression in terms of O_{n+1} for convenience, the same expression can be applied to any O_m since the n-point correlation functions are symmetric under index permutations.

To bound the n-point correlation function we first isolate the most distant operator, corresponding to $i^* = \text{argmax}_i \min_j d(O_i,O_j)$. This operator lies at a distance of at least $d(O_1,\dots,O_{n+1})$ [see Eq. (V.138)] from all other O_j.

Without loss of generality we permute indices such that $i^* = n+1$. We have

$$\begin{aligned}\left|C_\beta^{n+1}(O_1,\ldots,O_{n+1})\right| \leq & \sum_{P\in\mathcal{P}_n}(|P|-1)!\cdot \\ & \cdot\sum_{\tau=1}^{|P|}\left|\prod\nolimits_{\sigma\neq\tau}\langle\prod\nolimits_{i\in P_\sigma} O_i\rangle_\beta\right|\cdot\left|\langle O_{n+1}\prod\nolimits_{i\in P_\tau} O_i\rangle_\beta-\langle O_{n+1}\rangle_\beta\langle\prod\nolimits_{i\in P_\tau} O_i\rangle_\beta\right| \\ & \leq\sum\nolimits_{P\in\mathcal{P}_n}(|P|-1)!\sum\nolimits_{\tau=1}^{|P|}\left(\prod\nolimits_{i=1}^{n+1}\|O_i\|_s\right)\left(a_\beta\, e^{-d(O_{n+1},\prod_{i\in P_\tau} O_i)/\xi_\beta}\right) \\ & \leq a_\beta\Big(\prod_{i=1}^{n+1}\|O_i\|_s\Big)e^{-d(O_1,\ldots,O_{n+1})/\xi_\beta}\left(\sum\nolimits_{|P|=1}^{n}\mathfrak{s}(n,|P|)\cdot|P|!\right).\end{aligned} \tag{V.140}$$

The latter sum is known in mathematics as an ordered Bell number, and satisfies [342]

$$\sum_{k=1}^{n}\mathfrak{s}(n,k)\cdot k! = \frac{n!}{2\log(2)}\left(\frac{1}{\log(2)}\right)^n+\varepsilon(n),\qquad |\varepsilon(n)|<\frac{\pi}{12}\frac{n!}{(2\pi)^n}. \tag{V.141}$$

The error bound can be absorbed into the prefactor of the first term to give the following bound

$$\sum_{k=1}^{n}\mathfrak{s}(n,k)\cdot k!\ \leq\ \mathfrak{b}\, n!\left(\frac{1}{\log(2)}\right)^n, \tag{V.142}$$

with $\mathfrak{b}=\frac{1}{2\log(2)}+\frac{\log(2)}{24}$. Inserting the above into Eq. (V.140) produces our final bound, Eq. (V.137). □

Proof of Lemma V.10—Expanding the Hamiltonian, $H=\sum_i f_a P_a$, we have

$$\partial_\beta^n\langle A\rangle_\beta=\sum_{a_1,\ldots,a_n}\left(\prod_{i=1}^{n}f_{a_i}\right)C_\beta^{n+1}(A,P_{a_1},\ldots,P_{a_n}), \tag{V.143}$$

and hence

$$\begin{aligned}\left|\partial_\beta^n \langle A\rangle_\beta\right| &\leq J^n \sum_{a_1,\ldots,a_n} \left|C_\beta^{n+1}(A, P_{a_1}, \ldots, P_{a_n})\right| \\ &\leq a_\beta \mathfrak{b} \|A\|_s n! \left(\frac{J}{\log(2)}\right)^n \sum_{a_1,\ldots,a_n} e^{-d(A,P_{a_1},\ldots,P_{a_n})/\xi_\beta} \\ &= a_\beta \mathfrak{b} \|A\|_s n! \left(\frac{J}{\log(2)}\right)^n \sum_d \gamma_A(d) e^{-d/\xi_\beta}\end{aligned} \tag{V.144}$$

where $\gamma_A(d)$ is the number of choices of $\{a_1, \ldots, a_n\}$ with $d(A, P_{a_1}, \ldots, P_{a_n}) = d$.

We define $K(d)$ to be the number of Hamiltonian terms that lie within a distance d of any given Hamiltonian term. We have the upper bound $K(d) < \kappa(2d + r)^D$. We can use $K(d)$ to upper bound the distance $\gamma_A(d)$. Specifically, we have

$$\sum_{d'=0}^{d} \gamma_A(d') \leq n!\, K(d)^n = n!\, \kappa^n\, (2d + r)^{nD}. \tag{V.145}$$

The LHS is equal to the total number of choices of $\{a_1, \ldots, a_n\}$ with $d(A, P_{a_1}, \ldots, P_{a_n}) \leq d$. To derive the RHS, note that first operator P_{a_1} must lie within distance d of A, the second operator must lie within distance d of either A or P_{a_1}, and the m^{th} operator must lie within distance d of at least one of A, $P_{a_1}, \ldots P_{a_{m-1}}$. Since e^{-d/ξ_β} is strictly decreasing in d, we have

$$\begin{aligned}\sum_d \gamma_A(d) e^{-d/\xi_\beta} &\leq n! \kappa^n \sum_{d=0}^{\infty} \left[(2d + r)^{nD} - (2(d-1) + r)^{nD}\right] e^{-d/\xi_\beta} \\ &\leq n! \kappa^n \left(e^{1/\xi_\beta} - 1\right) \sum_{d=0}^{\infty} (2d - 2 + r)^{nD} e^{-d/\xi_\beta} \\ &\leq n! (2^D \kappa)^n e^{\lceil r/2 \rceil/\xi_\beta} \left(1 - e^{-1/\xi_\beta}\right) \sum_{x=\lceil r/2 \rceil - 1}^{\infty} x^{nD} e^{-x/2\xi_\beta}\end{aligned} \tag{V.146}$$

The sum can be expressed in terms of Eulerian numbers and upper bounded,

$$\sum_{x=0}^{\infty} x^{nD} e^{-x/2\xi_\beta} = \frac{e^{1/2\xi_\beta}}{(e^{1/2\xi_\beta} - 1)^{nD+1}} \sum_{i=1}^{nD} A(nD, i) e^{-(i-1)/2\xi_\beta}$$

$$\leq (nD)! \frac{e^{1/2\xi_\beta}}{(e^{1/2\xi_\beta}-1)^{nD+1}}, \quad \text{(V.147)}$$

which gives

$$\sum_d \gamma_A(d) e^{-d/\xi_\beta} \leq 2\,(nD)!\,(n)! \left(\frac{2^D \kappa}{(e^{1/2\xi_\beta}-1)^D}\right)^n e^{\lceil r/2 \rceil/\xi_\beta} \frac{1-e^{-1/\xi_\beta}}{1-e^{-1/2\xi_\beta}}. \quad \text{(V.148)}$$

Inserting the above bound into Eq. (V.144) and using inequalities, $(1 - e^{-1/\xi_\beta})/(1 - e^{-1/2\xi_\beta}) \leq 2$ and $x! \leq x^x$, gives Eq. (V.132). □

V.D.7. Learning thermalizing Hamiltonians with $\mathcal{O}(1)$ unknown parameters

In many Hamiltonian learning scenarios, the number of unknown parameters, N_p, scales with some power of the system size, N. In the prior sections, we showed that learning at the Heisenberg limit is not possible in this scenario, if the Hamiltonian thermalizes and one does not have sufficient quantum control. In this section, we turn instead to learning in thermalizing Hamiltonians with a constant number of unknown parameters, $N_p = \mathcal{O}(1)$. That is, the Hamiltonians we consider take the form

$$H(u) = uP + \sum_{a=1}^{N_k} v_a Q_a, \quad \text{(V.149)}$$

where u is the single unknown parameter, v_a are the known parameters (of total number N_k), and P, Q_a are k-local Pauli operators. Few-parameter learning problems are potentially easier than many-parameter learning problems, since, in the latter, a lack of knowledge about a given parameter can inhibit learning of other parameters. We will show that for thermalizing Hamiltonians this is indeed the case, and that a constant number of unknown parameters can, in principle, be learned at the Heisenberg limit even in the absence of quantum control. Nonetheless, we find that the measurements required to perform such learning may be difficult in practice.

For brevity, we focus on the case of a single unknown parameter, u, in the no quantum control model. Our main results are summarized by the following theorem:

Theorem V.10
Consider the learning problem (Definition V.5) for an N-qubit Hamiltonian as in Eq. (V.149) in the no quantum control model (Definition V.1). Suppose that the Hamiltonian $H(u)$ obeys the eigenstate thermalization hypothesis

(Definition V.6) within an energy range $[E_1, E_2]$*. Then:*

1. *Suppose that there exist energies* $E_3, E_4 \in [E_1, E_2]$ *such that* $|\langle P\rangle_{\beta(E_3)} - \langle P\rangle_{\beta(E_4)}| \geq c$ *for a constant* $c = \Omega(1)$ *independent of the system size. Then there exists a state for which the quantum Fisher information of* u *scales as* $\mathcal{I}^{(Q)} = \Theta(t^2)$*.*

2. *Suppose that the Hamiltonian has bounded degree* κ *and interaction strengths* $|u_a|, |v_a| \leq J = \mathcal{O}(1)$*, and has bounded derivatives,*

$$\left|\partial_E \langle P\rangle_{\beta(E)}\right| \leq \frac{\tilde{B}}{N}. \tag{V.150}$$

Further suppose that the initial state of each experiment is related to a product state by a finite-depth unitary quantum circuit and has energies within $[E_1, E_2]$*. Then learning the parameter* u *at the Heisenberg limit incurs an overhead proportional to the square root of the system size,* $T = \Omega(\sqrt{N}/\epsilon)$*.*

Proof— We prove both statements 1 and 2 using the quantum Fisher information. For a single unknown parameter, the quantum Fisher information matrix is simply a positive number. For pure initial states ρ_x and no quantum control, it takes the form

$$\mathcal{I}^{(Q)} = 4\sum_x \left(\mathrm{Trace}[\rho_x A_x^2] - \mathrm{Trace}[\rho_x A_x]^2\right), \qquad A_x = A_H(t_x) = \int_0^{t_x} ds\, P(s). \tag{V.151}$$

We refer to Appendix V.A for full details. As in the multi-parameter case, we decompose $A_H(t) = A_{\mathrm{d}}(t) + A_{\mathrm{od}}(t) + \delta A(t)$ as in Eq. (V.90). Again, the latter two terms cannot contribute to the Heisenberg limit owing to Lemma V.8 and Definition V.6, respectively. The first term is diagonal in the energy basis, with matrix elements $\langle E_i| A_d(t) |E_i\rangle = t \langle P\rangle_{\beta(E_i)}$. Writing the diagonal matrix elements of ρ_x as $\langle E_i| \rho |E_i\rangle = p_{x,i}$, we thus have

$$\mathcal{I}^{(Q)} = 4\sum_x t_x^2 \left(\left[\sum_i p_{x,i} \langle P\rangle_{\beta(E_i)}^2\right] - \left[\sum_i p_{x,i} \langle P\rangle_{\beta(E_i)}\right]^2\right) + \mathcal{O}(t^{3/2}). \tag{V.152}$$

The term in parentheses is the variance of $\langle P\rangle_{\beta(E_i)}$ over the probability distribution p_i. The sub-leading contribution, $\mathcal{O}(t^{3/2})$, arises from covariances between the diagonal and off-diagonal components of $A_H(t)$.

The quantum Fisher information clearly scales as t^2 whenever the above variance is non-zero. This is possible whenever the thermal expectation value

of P varies a non-zero amount with the energy. This is precisely the condition assumed in statement 1 of the theorem. From the assumption in statement 1, we can construct a simple state

$$\rho = |\psi_{34}\rangle\langle\psi_{34}|, \qquad |\psi_{34}\rangle = \frac{|E_3\rangle + |E_4\rangle}{\sqrt{2}}, \tag{V.153}$$

which features a quantum Fisher information

$$\mathcal{I}_x^{(Q)} = t_x^2 \left(\langle P \rangle_{\beta(E_4)} - \langle P \rangle_{\beta(E_3)} \right)^2 + \mathcal{O}(t_x^{3/2}). \tag{V.154}$$

This proves statement 1. We mention more realistic states that can achieve a quadratic Fisher information in the discussion following this proof.

To prove statement 2, we would like to upper bound the quantum Fisher information in the specific case that ρ is a related to a product state by a finite-depth unitary quantum circuit,

$$\rho = V |0\rangle\langle 0| V^\dagger, \tag{V.155}$$

where $|0\rangle$ is the all zeroes state and V is a finite-depth unitary. We will achieve this by upper bounding the energy variance of the state ρ, and using condition Eq. (V.150) to then upper bound the variance of $\langle P \rangle_{\beta(E_i)}$.

To bound the energy variance, we first consider the rotated Hamiltonian, $H^V = uP^V + \sum_a v_a Q_a^V$, where we denote $O^V = VOV^\dagger$ for an operator O. Note that if V has depth d and H is k-local with degree κ, then H^V is $\tilde{k}$-local with degree $\tilde{\kappa}$, with $\tilde{k} = 2^d k, \tilde{\kappa} = 2^d \kappa$. Now consider the energy variance of H in the state $V|0\rangle$

$$\delta E^2 \equiv \langle 0| VH^2V^\dagger |0\rangle - \langle 0| VHV^\dagger |0\rangle^2, \tag{V.156}$$

which is equal to the variance of H^V in the state $|0\rangle$. We have

$$\delta E^2 = \sum_{a,b=0}^{N_k} v_a v_b \Big(\langle 0| Q_a^V Q_b^V |0\rangle - \langle 0| Q_a^V |0\rangle \langle 0| Q_b^V |0\rangle \Big), \tag{V.157}$$

abbreviating $u \equiv v_0, P \equiv Q_0$. The summand above vanishes unless Q_a and Q_b share support. Each Q_a has support on at most $\tilde{k}$ sites, and thus overlaps with at most $\tilde{k}\tilde{\kappa}$ operators Q_b. We therefore have

$$\delta E^2 \leq J^2 4^d k\kappa N_k \leq J^2 4^d k\kappa^2 N. \tag{V.158}$$

where in the latter inequality, we use that the bounded degree κ upper bounds

$N_k \leq \kappa N$. We thus have our desired result, that the energy variance is $\mathcal{O}(N)$.

We now return to the upper bounding the quantum Fisher information. Since $\langle P \rangle_{\beta(E)}$ is a continuous and differentiable function of the energy E, we have

$$\begin{aligned}\sum_i p_{x,i} \langle P \rangle^2_{\beta(E_i)} - \left(\sum_i p_{x,i} \langle P \rangle_{\beta(E_i)} \right)^2 \\ \leq \left(\partial_E \langle P \rangle_{\beta(E)} \right)^2 \left[\sum_i p_{x,i} E_i^2 - \left(\sum_i p_{x,i} E_i \right)^2 \right] \\ = \left(\partial_E \langle P \rangle_{\beta(E)} \right)^2 \delta E^2. \quad \text{(V.159)}\end{aligned}$$

Our upper bound on the energy variance, combined with the assumption in statement 2 of the theorem, give the following bound on the quantum Fisher information

$$\mathcal{I}_x^{(Q)} \leq 4 \sum_x t_x^2 \left(\frac{\tilde{B}}{N} \right)^2 J^2 4^d k \kappa^2 N + \mathcal{O}(t_x^{3/2}) = \mathcal{O}(T^2/N). \quad \text{(V.160)}$$

The Cramer-Rao bound now gives our final result, that learning u at the Heisenberg limit requires total evolution time $T = \sum_x t_x = \Omega(\sqrt{N}/\epsilon)$. □

In the above proof, we presented a conceptually simple state [Eq. (V.153)] that can achieve a quadratically-growing quantum Fisher information. The state is a superposition of two energy eigenstates of the Hamiltonian, which is unrealistic in practice since eigenstates are hard to prepare. We now argue that a much more general class of initial states and measurements can achieve a similar Fisher information. The initial states will be 'GHZ-like', in that they are a superposition of states with extensively different energies. The final measurements will involve backwards time-evolution under a best estimate of the Hamiltonian, as first proposed in Refs. [270, 271].

To see this, suppose that the all zero product state $|0 \ldots 0\rangle$, differs in energy by an extensive amount from the all one product state, $|1 \ldots 1\rangle$. This implies that the two states will thermalize to different inverse temperatures, β_0 and β_1, after time-evolution. Generically, the thermal expectation value of P will differ by an order one amount between β_0 and β_1. If we take the initial state of the experiment to be a GHZ state,

$$\rho = |\text{GHZ}\rangle\langle\text{GHZ}|, \qquad |\text{GHZ}\rangle = \frac{|0 \ldots 0\rangle + |1 \ldots 1\rangle}{\sqrt{2}}, \quad \text{(V.161)}$$

this leads to a Fisher information that grows quadratically in time, $\mathcal{I}^{(Q)} \approx t_x^2(\langle P\rangle_{\beta_0} - \langle P\rangle_{\beta_1})^2$. We note that the temperatures can be adjusted by choosing different initial product states.

It remains to specify what measurement basis is needed to achieve this Fisher information. As in quantum sensing with GHZ states, the measurement basis will need to be sensitive to the coherence between the all zero and all one state after time-evolution. That is, we would like to measure the operator

$$e^{-iH(u)t}\,|0\ldots0\rangle\langle 1\ldots1|\,e^{iH(u)t} + h.c. \tag{V.162}$$

If the Hamiltonian were known, this could be achieved by first performing backwards time-evolution, i.e. applying $e^{iH(u)t}$, and then measuring the operator $|0\ldots0\rangle\langle 1\ldots1|$.

In practice, we expect that backwards time-evolution under the unknown Hamiltonian can be replaced with backwards time-evolution under a best estimate of the Hamiltonian, $H(\tilde{u})$. If the estimate $\tilde{u}$ has error ϵ, then the backwards time-evolution under $\tilde{u}$ will approximate that under u up to times $t \approx \epsilon$. This should be sufficient to improve the estimate $\tilde{u}$ by a constant factor. The experiments can then be updated adaptively as learning proceeds. This idea was proposed under the moniker 'quantum-assisted' Hamiltonian learning in Refs. [270, 271], where 'quantum-assisted' refers to the ability to backwards time-evolve under $H(\tilde{u})$. Our analysis thus provides a more thorough justification for the ability of the algorithms in these works to achieve the Heisenberg limit, in the case that the Hamiltonian has a single unknown parameter. On the other hand, our Theorem V.9 establishes that these schemes *cannot* always learn at the Heisenberg limit when the number of unknown parameters is extensive.

Bibliography

[1] R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute et al. Quantum error correction below the surface code threshold. *Nature*, **638**, 920—-926 (2025).

[2] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute et al. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, **614**, 676–681 (2023).

[3] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, **626**, 58–65 (2024).

[4] A. Paetznick, M. da Silva, C. Ryan-Anderson, J. Bello-Rivas, J. Campora III, A. Chernoguzov et al. Demonstration of logical qubits and repeated error correction with better-than-physical error rates, arXiv:2404.02280 (2024).

[5] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin & X. Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, **92**, 015003 (2020).

[6] B. Bauer, S. Bravyi, M. Motta & G. K.-L. Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chemical Reviews*, **120**, 12685–12717 (2020).

[7] C. W. Bauer, Z. Davoudi, A. B. Balantekin, T. Bhattacharya, M. Carena, W. A. De Jong et al. Quantum simulation for high-energy physics. *PRX Quantum*, **4**, 027001 (2023).

[8] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, **1**, 195 (1964).

[9] D. Zhu, G. D. Kahanamoku-Meyer, L. Lewis, C. Noel, O. Katz, B. Harraz et al. Interactive cryptographic proofs of quantumness using mid-circuit measurements. *Nature Physics*, **19**, 1725–1731 (2023).

[10] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, **526**, 682–686 (2015).

[11] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, **41**, 303–332 (1999).

[12] C. Gidney & M. Ekerå. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits (2021).

[13] C. Gidney. How to factor 2048 bit RSA integers with less than a million noisy qubits, arXiv:2505.15917 (2025).

[14] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends et al. Quantum supremacy using a programmable superconducting processor. *Nature*, **574**, 505–510 (2019).

[15] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen et al. Strong quantum computational advantage using a superconducting quantum processor. *Physical Review Letters*, **127**, 180501 (2021).

[16] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo et al. Quantum computational advantage using photons. *Science*, **370**, 1460–1463 (2020).

[17] B. M. Terhal & D. P. DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games, arXiv:quant-ph/0205133 (2002).

[18] D. Hangleiter & J. Eisert. Computational advantage of quantum random sampling. *Reviews of Modern Physics*, **95**, 035001 (2023).

[19] D. A. Abanin, R. Acharya, L. Aghababaie-Beni, G. Aigeldinger, A. Ajoy, R. Alcaraz et al. Constructive interference at the edge of quantum ergodic dynamics, arXiv:2506.10191 (2025).

[20] A. D. King, A. Nocera, M. M. Rams, J. Dziarmaga, R. Wiersema, W. Bernoudy et al. Beyond-classical computation in quantum simulation. *Science*, **388**, 199–204 (2025).

[21] R. Haghshenas, E. Chertkov, M. Mills, W. Kadow, S.-H. Lin, Y.-H. Chen et al. Digital quantum magnetism at the frontier of classical simulations, arXiv:2503.20870 (2025).

[22] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, **618**, 500–505 (2023).

[23] D. Aharonov, X. Gao, Z. Landau, Y. Liu & U. Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 945–957 (2023).

[24] J. Tindall, M. Fishman, E. M. Stoudenmire & D. Sels. Efficient tensor network simulation of IBM's eagle kicked ising experiment. *PRX quantum*, **5**, 010308 (2024).

[25] J. Tindall, A. Mello, M. Fishman, M. Stoudenmire & D. Sels. Dynamics of disordered quantum systems with two-and three-dimensional tensor networks, arXiv:2503.05693 (2025).

[26] L. Mauron & G. Carleo. Challenging the quantum advantage frontier with large-scale classical simulations of annealing dynamics, arXiv:2503.08247 (2025).

[27] T. Begušić, J. Gray & G. K.-L. Chan. Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance. *Science Advances*, **10**, eadk4321 (2024).

[28] F. Pan, K. Chen & P. Zhang. Solving the sampling problem of the Sycamore quantum circuits. *Physical Review Letters*, **129**, 090502 (2022).

[29] M. A. Nielsen & I. L. Chuang. *Quantum Computation and Quantum Information*, volume 54. Cambridge University Press, 2 edition (2001).

[30] D. Aharonov, M. Ben-Or, R. Impagliazzo & N. Nisan. Limitations of noisy reversible computation, arXiv:quant-ph/9611028 (1996).

[31] I. L. Chuang, R. Laflamme, P. W. Shor & W. H. Zurek. Quantum computers, factoring, and decoherence. *Science*, **270**, 1633–1635 (1995).

[32] W. K. Wootters & W. H. Zurek. A single quantum cannot be cloned. *Nature*, **299**, 802–803 (1982).

[33] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall & T. J. Yoder. High-threshold and low-overhead fault-tolerant quantum memory. *Nature*, **627**, 778–782 (2024).

[34] P. W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, **52**, R2493 (1995).

[35] A. Y. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of physics*, **303**, 2–30 (2003).

[36] A. G. Fowler, M. Mariantoni, J. M. Martinis & A. N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, **86**, 032324 (2012).

[37] H. Bombin & M. A. Martin-Delgado. Topological quantum distillation. *Physical review letters*, **97**, 180501 (2006).

[38] P. Panteleev & G. Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In *Proceedings of the 54th annual ACM SIGACT symposium on theory of computing*, pages 375–388 (2022).

[39] E. Campbell. A series of fast-paced advances in quantum error correction. *Nature Reviews Physics*, **6**, 160–161 (2024).

[40] B. Eastin & E. Knill. Restrictions on transversal encoded quantum gate sets. *Physical Review Letters*, **102**, 110502 (2009).

[41] D. Gottesman. The Heisenberg representation of quantum computers, arXiv:quant-ph/9807006 (1998).

[42] E. T. Campbell, B. M. Terhal & C. Vuillot. Roads towards fault-tolerant universal quantum computation. *Nature*, **549**, 172–179 (2017).

[43] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. Mcclean, N. Wiebe & R. Babbush. Even more efficient quantum computations of chemistry through tensor hypercontraction. *Physical Review X* (2021).

[44] C. Gidney, N. Shutty & C. Jones. Magic state cultivation: Growing T states as cheap as CNOT gates, arXiv:2409.17595 (2024).

[45] E. Knill. Quantum computing with realistically noisy devices. *Nature*, **434**, 39–44 (2005).

[46] J. J. Goings, A. White, J. Lee, C. S. Tautermann, M. Degroote, C. Gidney, T. Shiozaki, R. Babbush & N. C. Rubin. Reliably assessing the electronic structure of cytochrome p450 on today's classical computers and tomorrow's quantum computers. *Proceedings of the National Academy of Sciences*, **119**, e2203533119 (2022).

[47] D. Castelvecchi. IBM releases first-ever 1,000-qubit quantum chip. *Nature*, **624**, 238–238 (2023).

[48] J. Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, **2**, 79 (2018).

[49] K. Temme, S. Bravyi & J. M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, **119**, 180509 (2017).

[50] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean & T. E. O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, **95**, 045005 (2023).

[51] S. Endo, S. C. Benjamin & Y. Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, **8**, 031027 (2018).

[52] J. Preskill. Beyond NISQ: The megaquop machine (2025).

[53] N. S. Blunt, G. P. Gehér & A. E. Moylett. Compilation of a simple chemistry application to quantum error correction primitives. *Physical Review Research*, **6**, 013325 (2024).

[54] Y. Akahoshi, K. Maruyama, H. Oshima, S. Sato & K. Fujii. Partially fault-tolerant quantum computing architecture with error-corrected Clifford gates and space-time efficient analog rotations. *PRX Quantum*, **5**, 010337 (2024).

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

[56] G. Wang, D. S. França, G. Rendon & P. D. Johnson. Efficient ground-state-energy estimation and certification on early fault-tolerant quantum computers. *Physical Review A*, **111**, 012426 (2025).

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

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

[59] K. Wan, M. Berta & E. T. Campbell. Randomized quantum algorithm for statistical phase estimation. *Physical Review Letters*, **129**, 030503 (2022).

[60] Y. Dong, L. Lin & Y. Tong. Ground state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices. *PRX Quantum*, **3**, 040305 (2022).

[61] Z. Ding & L. Lin. Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation. *PRX Quantum*, **4**, 020331 (2023).

[62] Z. Ding & L. Lin. Simultaneous estimation of multiple eigenvalues with short-depth quantum circuit on early fault-tolerant quantum computers. *Quantum*, **7**, 1136 (2023).

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

[64] J. S. Nelson & A. D. Baczewski. An assessment of quantum phase estimation protocols for early fault-tolerant quantum computers, arXiv:2403.00077 (2024).

[65] I. Y. Akhalwaya, S. Ubaru, K. L. Clarkson, M. S. Squillante, V. Jejjala, Y.-H. He, K. Naidoo, V. Kalantzis & L. Horesh. Topological data analysis on noisy quantum computers, arXiv:2209.09371 (2024).

[66] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross & Y. Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, **115**, 9456–9461 (2018).

[67] E. T. Campbell. Early fault-tolerant simulations of the Hubbard model. *Quantum Science and Technology*, **7**, 015007 (2022).

[68] Y. Suzuki, S. Endo, K. Fujii & Y. Tokunaga. Quantum error mitigation as a universal error-minimization technique: Applications from NISQ to FTQC eras. *PRX Quantum*, **3**, 010345 (2022).

[69] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta & K. Temme. Error mitigation for universal gates on encoded qubits. *Physical Review Letters*, **127**, 200505 (2021).

[70] M. Lostaglio & A. Ciani. Error mitigation and quantum-assisted simulation in the error corrected regime. *Physical Review Letters*, **127**, 200506 (2021).

[71] R. Toshio, Y. Akahoshi, J. Fujisaki, H. Oshima, S. Sato & K. Fujii. Practical quantum advantage on partially fault-tolerant quantum computer (2024).

[72] A. Katabarwa, K. Gratsea, A. Caesura & P. D. Johnson. Early fault-tolerant quantum computing, arXiv:2311.14814 (2023).

[73] Z. Ding, H. Li, L. Lin, H. Ni, L. Ying & R. Zhang. Quantum multiple eigenvalue gaussian filtered search: an efficient and versatile quantum phase estimation method. *Quantum*, **8**, 1487 (2024).

[74] S. Kimmel, G. H. Low & T. J. Yoder. Robust calibration of a universal single-qubit gate set via robust phase estimation. *Physical Review A*, **92**, 062315 (2015).

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

[76] A. Dutkiewicz, S. Polla, M. Scheurer, C. Gogolin, W. J. Huggins & T. E. O'Brien. Error mitigation and circuit division for early fault-tolerant quantum phase estimation. *PRX Quantum*, **6**, 040318 (2025).

[77] K. Liu & Z. Cai. Quantum error mitigation for sampling algorithms, arXiv:2502.11285 (2025).

[78] Y. Akahoshi, R. Toshio, J. Fujisaki, H. Oshima, S. Sato & K. Fujii. Compilation of Trotter-Based Time Evolution for Partially Fault-Tolerant Quantum Computing Architecture, arXiv:2408.14929 (2024).

[79] Y. Manin. Computable and uncomputable. *Sovetskoye Radio, Moscow*, **128**, 15 (1980).

[80] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics* (1982).

[81] A. Di Meglio, K. Jansen, I. Tavernelli, C. Alexandrou, S. Arunachalam, C. W. Bauer et al. Quantum computing for high-energy physics: State of the art and challenges. *PRX Quantum*, **5**, 037001 (2024).

[82] T. E. O'Brien, M. Streif, N. C. Rubin, R. Santagati, Y. Su, W. J. Huggins et al. Efficient quantum computation of molecular forces and other energy gradients. *Physical Review Research*, **4**, 043210 (2022).

[83] H. Perrin, T. Scoquart, A. I. Pavlov & N. V. Gnezdilov. Dynamic thermalization on noisy quantum hardware. *Communications Physics*, **8**, 95 (2025).

[84] A. M. Childs, J. Leng, T. Li, J.-P. Liu & C. Zhang. Quantum simulation of real-space dynamics. *Quantum*, **6**, 860 (2022).

[85] K. Lee, F. Turro & X. Yao. Quantum computing for energy correlators. *Physical Review D*, **111**, 054514 (2025).

[86] E. Knill, G. Ortiz & R. D. Somma. Optimal quantum measurements of expectation values of observables. *Physical Review A—Atomic, Molecular, and Optical Physics*, **75**, 012328 (2007).

[87] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová et al. Quantum chemistry in the age of quantum computing. *Chemical reviews*, **119**, 10856–10915 (2019).

[88] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love & M. Head-Gordon. Simulated quantum computation of molecular energies. *Science*, **309**, 1704–1707 (2005).

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

[90] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker & M. Troyer. Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences*, **114**, 7555–7560 (2017).

[91] N. C. Jones, J. D. Whitfield, P. L. McMahon, M.-H. Yung, R. Van Meter, A. Aspuru-Guzik & Y. Yamamoto. Faster quantum chemistry simulation on fault-tolerant quantum computers. *New Journal of Physics*, **14**, 115023 (2012).

[92] G. Wang, S. Sim & P. D. Johnson. State preparation boosters for early fault-tolerant quantum computation. *Quantum*, **6**, 829 (2022).

[93] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush & G. K.-L. Chan. Low rank representations for quantum simulation of electronic structure. *npj Quantum Information*, **7**, 83 (2021).

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

[95] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung et al. Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms. *Physical Review X*, **5**, 041041 (2015).

[96] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven & G. K.-L. Chan. Low-depth quantum simulation of materials. *Physical Review X*, **8**, 011044 (2018).

[97] Y. Su, D. W. Berry, N. Wiebe, N. Rubin & R. Babbush. Fault-tolerant quantum simulations of chemistry in first quantization. *PRX Quantum*, **2**, 040332 (2021).

[98] T. Eckstein, R. Mansuroglu, P. Czarnik, J.-X. Zhu, M. J. Hartmann, L. Cincio, A. T. Sornborger & Z. Holmes. Large-scale simulations of Floquet physics on near-term quantum computers. *npj Quantum Information*, **10**, 84 (2024).

[99] Z. Ding, Y. Dong, Y. Tong & L. Lin. Robust ground-state energy estimation under depolarizing noise, arXiv:2307.11257 (2023).

[100] O. Kiss, U. Azad, B. Requena, A. Roggero, D. Wakeham & J. M. Arrazola. Early fault-tolerant quantum algorithms in practice: Application to ground-state energy estimation. *Quantum*, **9**, 1682 (2025).

[101] C. Zhang, R. G. Cortiñas, A. H. Karamlou, J. P. N. Noll, J. Bausch, S. Shirobokov et al. Quantum computation of molecular geometry via many-body nuclear spin echoes, arXiv:2510.19550 (2025).

[102] N. Drummond & R. Needs. Phase diagram of the low-density two-dimensional homogeneous electron gas. *Physical review letters*, **102**, 126402 (2009).

[103] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar et al. Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry. *Nature Communications*, **14**, 1952 (2023).

[104] M. Stroeks, D. Lenterman, B. Terhal & Y. Herasymenko. Solving free fermion problems on a quantum computer, arXiv:2409.04550 (2024).

[105] R. D. Somma, R. King, R. Kothari, T. E. O'Brien & R. Babbush. Shadow Hamiltonian simulation. *Nature Communications*, **16**, 2690 (2025).

[106] W. J. Huggins, O. Leimkuhler, T. F. Stetina & K. B. Whaley. Efficient state preparation for the quantum simulation of molecules in first quantization. *PRX Quantum*, **6**, 020319 (2025).

[107] G. H. Low, R. King, D. W. Berry, Q. Han, A. E. DePrince III, A. White, R. Babbush, R. D. Somma & N. C. Rubin. Fast quantum simulation of electronic structure by spectrum amplification, arXiv:2502.15882 (2025).

[108] Z. Li, J. Li, N. S. Dattani, C. Umrigar & G. K. Chan. The electronic complexity of the ground-state of the femo cofactor of nitrogenase as relevant to quantum simulations. *The Journal of chemical physics*, **150** (2019).

[109] D. W. Berry, Y. Tong, T. Khattar, A. White, T. I. Kim, S. Boixo et al. Rapid initial state preparation for the quantum simulation of strongly correlated molecules, arXiv:2409.11748 (2024).

[110] M. Mörchen, G. H. Low, T. Weymuth, H. Liu, M. Troyer & M. Reiher. Classification of electronic structures and state preparation for quantum computation of reaction chemistry, arXiv:2409.08910 (2024).

[111] J. Bausch. Fast black-box quantum state preparation. *Quantum*, **6**, 773 (2022).

[112] L. Lin & Y. Tong. Near-optimal ground state preparation. *Quantum*, **4**, 372 (2020).

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

[114] G. H. Low & I. L. Chuang. Hamiltonian simulation by qubitization. *Quantum*, **3**, 163 (2019).

[115] J. D. Whitfield, J. Biamonte & A. Aspuru-Guzik. Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics*, **109**, 735–750 (2011).

[116] I. Moflic & A. Paler. On the constant depth implementation of Pauli exponentials, arXiv:2408.08265 (2024).

[117] E. Campbell. Random compiler for fast Hamiltonian simulation. *Physical Review Letters*, **123**, 070503 (2019).

[118] J. Günther, F. Witteveen, A. Schmidhuber, M. Miller, M. Christandl & A. Harrow. Phase estimation with partially randomized time evolution, arXiv:2503.05647 (2025).

[119] V. Giovannetti, S. Lloyd & L. Maccone. Advances in quantum metrology. *Nature photonics*, **5**, 222–229 (2011).

[120] C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, **1**, 231–252 (1969).

[121] R. Demkowicz-Dobrzański. Quantum measurement and estimation theory.

[122] S. Zhou & L. Jiang. Asymptotic theory of quantum channel estimation. *PRX Quantum*, **2**, 010343 (2021).

[123] H. Yuan & C.-H. F. Fung. Optimal feedback scheme and universal time scaling for Hamiltonian parameter estimation. *Physical Review Letters*, **115**, 110401 (2015).

[124] P. Sekatski, M. Skotiniotis, J. Kołodyński & W. Dür. Quantum metrology with full and fast quantum control. *Quantum*, **1**, 27 (2017).

[125] J. Liu, H. Yuan, X.-M. Lu & X. Wang. Quantum fisher information matrix and multiparameter estimation. *Journal of Physics A: Mathematical and Theoretical*, **53**, 023001 (2020).

[126] S. Zhou, M. Zhang, J. Preskill & L. Jiang. Achieving the Heisenberg limit in quantum metrology using quantum error correction. *Nature communications*, **9**, 78 (2018).

[127] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman & D. W. Berry. π-corrected Heisenberg limit. *Physical Review Letters*, **124**, 030501 (2020).

[128] R. Demkowicz-Dobrzański, W. Górecki & M. Guţă. Multi-parameter estimation beyond quantum fisher information. *Journal of Physics A: Mathematical and Theoretical*, **53**, 363001 (2020).

[129] A. Luis & J. Peřina. Optimum phase-shift estimation and the quantum description of the phase difference. *Physical Review A*, **54**, 4564–4570 (1996).

[130] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman & G. J. Pryde. Entanglement-free Heisenberg-limited phase estimation. *Nature*, **450**, 393–396 (2007).

[131] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman & G. J. Pryde. Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements. *New Journal of Physics*, **11**, 073023 (2009).

[132] P. Najafi, P. C. S. Costa & D. W. Berry. Optimum phase estimation with two control qubits, arXiv:2303.12503 (2023).

[133] A. W. Harrow, A. Hassidim & S. Lloyd. Quantum algorithm for solving linear systems of equations. *Physical Review Letters*, **15**, 150502 (2009).

[134] G. Rendon & P. D. Johnson. Low-depth gaussian state energy estimation, arXiv:2309.16790 (2023).

[135] Y. Ge, J. Tura & J. I. Cirac. Faster ground state preparation and high-precision ground energy estimation with fewer qubits. *Journal of Mathematical Physics*, **60**, 022202 (2019).

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

[137] P. Wocjan & S. Zhang. Several natural BQP-complete problems, arXiv:quant-ph/0606179 (2006).

[138] R. D. Somma. Quantum eigenvalue estimation via time series analysis. *New Journal of Physics*, **21**, 123025 (2019).

[139] W. van Dam, G. M. D'Ariano, A. Ekert, C. Macchiavello & M. Mosca. Optimal quantum circuits for general phase estimation. *Physical Review Letters*, **98**, 090501 (2007).

[140] G. Rendon, T. Izubuchi & Y. Kikuchi. Effects of cosine tapering window on quantum phase estimation. *Physical Review D*, **106**, 034503 (2022).

[141] G. Rendon, J. Watkins & N. Wiebe. Improved accuracy for trotter simulations using chebyshev interpolation. *Quantum*, **8**, 1266 (2024).

[142] D. W. Berry, Y. Su, C. Gyurik, R. King, J. Basso, A. D. T. Barba, A. Rajput, N. Wiebe, V. Dunjko & R. Babbush. Analyzing prospects for quantum advantage in topological data analysis. *PRX Quantum*, **5**, 010319 (2024).

[143] A. Y. Kitaev. Quantum measurements and the Abelian Stabilizer Problem, arXiv:quant-ph/9511026 (1995).

[144] F. Belliardo & V. Giovannetti. Achieving Heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error. *Physical Review A*, **102**, 042613 (2020).

[145] J. M. Martyn, Z. M. Rossi, A. K. Tan & I. L. Chuang. Grand unification of quantum algorithms. *PRX quantum*, **2**, 040203 (2021).

[146] R. Cleve, A. Ekert, C. Macchiavello & M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **454**, 339–354 (1998).

[147] V. Giovannetti, S. Lloyd & L. Maccone. Quantum metrology. *Physical Review Letters*, **96**, 010401 (2006).

[148] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde & H. M. Wiseman. How to perform the most accurate possible phase measurements. *Physical Review A*, **80**, 052114 (2009).

[149] R. B. Griffiths & C.-S. Niu. Semiclassical Fourier transform for quantum computation. *Physical Review Letters*, **76**, 3228–3231 (1996).

[150] A. Y. Kitaev. Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026 (1995).

[151] D. W. Berry, G. Ahokas, R. Cleve & B. C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Communications in Mathemathical Physics*, **270** (2007).

[152] N. Wiebe & C. Granade. Efficient Bayesian phase estimation. *Physical Review Letters*, **117**, 010503 (2016).

[153] K. M. Svore, M. B. Hastings & M. Freedman. Faster phase estimation, arXiv:1304.0741 (2013).

[154] E. van den Berg. Efficient Bayesian phase estimation using mixed priors, arXiv:2007.11629 (2020).

[155] D. C. Rife & R. R. Boorstyn. Single-tone parameter estimation from discrete-time observations. *IEEE Transactions on information theory*, **20**, 591–598 (1974).

[156] S. Lu, M. C. Bañuls & J. I. Cirac. Algorithms for quantum simulation at finite energies. *PRX Quantum*, **2**, 020321 (2020).

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

[158] A. Roggero. Spectral-density estimation with the gaussian integral transform. *Physical Review A*, **102**, 022409 (2020).

[159] A. Gilyén, Y. Su, G. H. Low & N. Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st annual ACM SIGACT symposium on theory of computing*, pages 193–204 (2019).

[160] O. Regev. A subexponential time algorithm for the dihedral hidden subgroup problem with polynomial space, arXiv:quant-ph/0406151 (2004).

[161] V. Gebhart, A. Smerzi & L. Pezzè. Bayesian quantum multiphase estimation algorithm. *Physical Review Applied*, **16**, 014035 (2021).

[162] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press (1946). ISBN 0691080046.

[163] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, **37**, 81–89 (1945).

[164] Y. Hua & T. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustic Speech and Signal Processing*, **38** (1990).

[165] A. Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 821–830. Association for Computing Machinery, New York, NY, USA (2015). ISBN 978-1-4503-3536-2.

[166] M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, A. Chernoguzov, J. M. Dreiling, C. Foltz et al. Demonstration of logical qubits and repeated error correction with better-than-physical error rates, arXiv:2404.02280 (2024).

[167] A. G. Fowler & C. Gidney. Low overhead quantum computation using lattice surgery, arXiv:1808.06709 (2019).

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

[169] C. Gidney & A. G. Fowler. Efficient magic state factories with a catalyzed $|CCZ\rangle$ to $2|T\rangle$ transformation. *Quantum*, **3**, 135 (2019).

[170] R. Kshirsagar, A. Katabarwa & P. D. Johnson. On proving the robustness of algorithms for early fault-tolerant quantum computers. *Quantum*, **8**, 1531 (2024).

[171] D. Bultrini, S. Wang, P. Czarnik, M. H. Gordon, M. Cerezo, P. J. Coles & L. Cincio. The battle of clean and dirty qubits in the era of partial error correction. *Quantum*, **7**, 1060 (2023).

[172] Q. Liang, Y. Zhou, A. Dalal & P. Johnson. Modeling the performance of early fault-tolerant quantum algorithms. *Physical Review Research*, **6**, 023118 (2024).

[173] D. Wang, O. Higgott & S. Brierley. Accelerated Variational Quantum Eigensolver. *Physical Review Letters*, **122**, 140504 (2019).

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

[175] Y. Quek, D. S. França, S. Khatri, J. J. Meyer & J. Eisert. Exponentially tighter bounds on limitations of quantum error mitigation, arXiv:2210.11505 (2023).

[176] R. Takagi, H. Tajima & M. Gu. Universal sampling lower bounds for quantum error mitigation. *Physical Review Letters*, **131**, 210602 (2023).

[177] K. Tsubouchi, T. Sagawa & N. Yoshioka. Universal cost bound of quantum error mitigation based on quantum estimation theory. *Physical Review Letters*, **131**, 210601 (2023).

[178] Y. Li & S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, **7**, 021050 (2017).

[179] S. McArdle, X. Yuan & S. Benjamin. Error-mitigated digital quantum simulation. *Physical Review Letters*, **122**, 180501 (2019).

[180] X. Bonet-Monroig, R. Sagastizabal, M. Singh & T. E. O'Brien. Low-cost error mitigation by symmetry verification. *Physical Review A*, **98**, 62339 (2018).

[181] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush & J. R. McClean. Virtual distillation for quantum error mitigation. *Physical Review X*, **11**, 041036 (2021).

[182] A. Montanaro & S. Stanisic. Error mitigation by training with fermionic linear optics, arXiv:2102.02120 (2021).

[183] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow & J. M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, **567**, 491–495 (2019).

[184] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu et al. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. *Physical Review A*, **100**, 010302 (2019).

[185] E. van den Berg, Z. K. Minev, A. Kandala & K. Temme. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nature Physics*, **19**, 1116–1121 (2023).

[186] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby & A. Montanaro. Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer. *Nature Communications*, **13**, 5743 (2022).

[187] T. E. O'Brien, G. Anselmetti, F. Gkritsis, V. E. Elfving, S. Polla, W. J. Huggins et al. Purification-based quantum error mitigation of pair-correlated electron simulations. *Nature Physics*, pages 1–6 (2023).

[188] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends et al. Hartree-Fock on a superconducting qubit quantum computer. *Science*, **369**, 1084–1089 (2020).

[189] S. Lu, M. C. Bañuls & J. I. Cirac. Algorithms for quantum simulation at finite energies. *PRX Quantum*, **2**, 020321 (2021).

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

[191] Y. Yang, A. Christianen, M. C. Bañuls, D. S. Wild & J. I. Cirac. Phase-sensitive quantum measurement without controlled operations. *Physical Review Letters*, **132**, 220601 (2024).

[192] D. Castaldo, S. Jahangiri, A. Migliore, J. M. Arrazola & S. Corni. A differentiable quantum phase estimation algorithm, arXiv:2406.14113 (2024).

[193] V. Dunjko, Y. Ge & J. I. Cirac. Computational speedups using small quantum devices. *Physical Review Letters*, **121**, 250501 (2018).

[194] Y. Ge & V. Dunjko. A hybrid algorithm framework for small quantum computers with application to finding Hamiltonian cycles. *Journal of Mathematical Physics*, **61**, 012201 (2020).

[195] A. M. Childs, R. Kothari, M. Kovacs-Deak, A. Sundaram & D. Wang. Quantum divide and conquer, arXiv:2210.06419 (2022).

[196] K. Fujii, K. Mizuta, H. Ueda, K. Mitarai, W. Mizukami & Y. O. Nakagawa. Deep Variational Quantum Eigensolver: A Divide-And-Conquer method for solving a larger problem with smaller size quantum computers. *PRX Quantum*, **3**, 010346 (2022).

[197] A. Pérez-Salinas, R. Draš škić ć, J. Tura & V. Dunjko. Shallow quantum circuits for deeper problems. *Physical Review A: Atomic, Molecular, and Optical Physics*, **108**, 062423 (2023).

[198] GitHub repository (2024).

[199] D. Horsman, A. G. Fowler, S. Devitt & R. V. Meter. Surface code quantum computing by lattice surgery. *New Journal of Physics*, **14**, 123011 (2012).

[200] R. Raussendorf & J. Harrington. Fault-tolerant quantum computation with high threshold in two dimensions. *Physical Review Letters*, **98**, 190504 (2007).

[201] A. G. Fowler, A. M. Stephens & P. Groszkowski. High-threshold universal quantum computation on the surface code. *Physical Review A*, **80**, 052312 (2009).

[202] D. Rocca, C. L. Cortes, J. F. Gonthier, P. J. Ollitrault, R. M. Parrish, G.-L. Anselmetti, M. Degroote, N. Moll, R. Santagati & M. Streif. Reducing the runtime of fault-tolerant quantum simulations in chemistry through symmetry-compressed double factorization. *Journal of Chemical Theory and Computation*, **20**, 4639–4653 (2024).

[203] M. P. Harrigan, T. Khattar, C. Yuan, A. Peduri, N. Yosri, F. D. Malone, R. Babbush & N. C. Rubin. Expressing and analyzing quantum algorithms with Qualtran, arXiv:2409.04643 (2024).

[204] J. Kołodyński & R. Demkowicz-Dobrzański. Efficient tools for quantum metrology with uncorrelated noise. *New Journal of Physics*, **15**, 073043 (2013).

[205] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan & M. Steffen. Self-consistent quantum process tomography. *Physical Review A—Atomic, Molecular, and Optical Physics*, **87**, 062119 (2013).

[206] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier & P. Maunz. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nature communications*, **8**, 14485 (2017).

[207] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young & R. Blume-Kohout. Gate Set Tomography. *Quantum*, **5**, 557 (2021).

[208] N. M. Tubman, C. Mejuto-Zaera, J. M. Epstein, D. Hait, D. S. Levine, W. Huggins et al. Postponing the orthogonality catastrophe: Efficient state preparation for electronic structure simulations on quantum devices, arXiv:1809.05523 (2018).

[209] S. Fomichev, K. Hejazi, M. S. Zini, M. Kiser, J. F. Morales, P. A. M. Casares et al. Initial state preparation for quantum chemistry on quantum computers, arXiv:2310.18410 (2024).

[210] M. McEwen, L. Faoro, K. Arya, A. Dunsworth, T. Huang, S. Kim et al. Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits. *Nature Physics*, **18**, 107–111 (2022).

[211] X. Pan, Y. Zhou, H. Yuan, L. Nie, W. Wei, L. Zhang et al. Engineering superconducting qubits to reduce quasiparticles and charge noise. *Nature Communications*, **13**, 7196 (2022).

[212] M. McEwen, K. C. Miao, J. Atalaya, A. Bilmes, A. Crook, J. Bovaird et al. Resisting high-energy impact events through gap engineering in superconducting qubit arrays, arXiv:2402.15644 (2024).

[213] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush & J. Lee. Unbiasing fermionic quantum Monte Carlo with a quantum computer. *Nature*, **603**, 416–420 (2022).

[214] J. Kempe, A. Kitaev & O. Regev. The complexity of the local Hamiltonian problem. *SIAM Journal on Computing*, **35**, 1070–1097 (2006).

[215] S. Gharibian & F. L. Gall. Dequantizing the quantum singular value transformation: Hardness and applications to quantum chemistry and the quantum PCP conjecture. *SIAM Journal on Computing*, **52**, 1009–1038 (2023).

[216] R. Cleve, A. Ekert, C. Macchiavello & M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **454**, 339–354 (1998).

[217] D. W. Berry & H. M. Wiseman. Optimal states and almost optimal adaptive measurements for quantum interferometry. *Physical Review Letters*, **85**, 5098 (2000).

[218] R. B. Griffiths & C.-S. Niu. Semiclassical Fourier transform for quantum computation. *Physical Review Letters*, **76**, 3228–3231 (1996).

[219] P. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134 (1994).

[220] M. Mosca & A. Ekert. The Hidden Subgroup Problem and eigenvalue estimation on a quantum computer. arXiv:quant-ph/9903071. arXiv (1999).

[221] C. Zalka. Fast versions of Shor's quantum factoring algorithm, arXiv:quant-ph/9806084 (1998).

[222] S. Parker & M. B. Plenio. Efficient factorization with a single pure qubit and $\log N$ mixed qubits. *Physical Review Letters*, **85**, 3049–3052 (2000).

[223] S. Beauregard. Circuit for Shor's algorithm using 2n+3 qubits, arXiv:quant-ph/0205095 (2003).

[224] N. Wiebe & C. Granade. Efficient Bayesian phase estimation. *Physical Review Letters*, **117**, 010503 (2016).

[225] A. E. Russo, K. M. Rudinger, B. C. Morrison & A. D. Baczewski. Evaluating energy differences on a quantum computer with robust phase estimation. *Physical Review Letters*, **126**, 210501 (2021).

[226] E. van den Berg. Efficient Bayesian phase estimation using mixed priors. *Quantum*, **5**, 469 (2021).

[227] Harald Cramer. *Mathematical Methods Of Statistics*. Princeton University Press (1946).

[228] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. In S. Kotz & N. L. Johnson, editors, *Breakthroughs in Statistics*, pages 235–247. Springer New York, New York, NY (1992). ISBN 978-0-387-94037-3 978-1-4612-0919-5.

[229] G. Casella & R. L. Berger. *Statistical Inference*. Thomson Learning, Australia ; Pacific Grove, CA, 2nd ed edition (2002). ISBN 978-0-534-24312-8.

[230] V. Giovannetti, S. Lloyd & L. Maccone. Quantum-enhanced measurements: Beating the standard quantum limit. *Science (New York, N.Y.)*, **306**, 1330–1336 (2004).

[231] C. King. The capacity of the quantum depolarizing channel, arXiv:quant-ph/0204172 (2002).

[232] J. Vovrosh, K. E. Khosla, S. Greenaway, C. Self, M. Kim & J. Knolle. Simple mitigation of global depolarizing errors in quantum simulations. *Physical Review E*, **104**, 035309 (2021).

[233] C. Gidney. Halving the cost of quantum addition. *Quantum*, **2**, 74 (2018).

[234] Y. R. Sanders, D. W. Berry, P. C. S. Costa, L. W. Tessler, N. Wiebe, C. Gidney, H. Neven & R. Babbush. Compilation of fault-tolerant quantum heuristics for combinatorial optimization. *PRX Quantum*, **1**, 020312 (2020).

[235] L. Hales & S. Hallgren. An improved quantum Fourier transform algorithm and applications. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 515–525. IEEE (2000).

[236] M. Mosca & C. Zalka. Exact quantum Fourier transforms and discrete logarithm algorithms. *International Journal of Quantum Information*, **2**, 91–100 (2004).

[237] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill et al. Information scrambling in quantum circuits. *Science*, **374**, 1479–1483 (2021).

[238] S. Filippov, M. Leahy, M. A. C. Rossi & G. García-Pérez. Scalable tensor-network error mitigation for near-term quantum computing, arXiv:2307.11740 (2023).

[239] S. Bravyi, M. Englbrecht, R. Koenig & N. Peard. Correcting coherent errors with surface codes. *npj Quantum Information*, **4**, 55 (2018).

[240] N. Linden & R. de Wolf. Average-case verification of the quantum Fourier transform enables worst-case phase estimation. *Quantum*, **6**, 872 (2022).

[241] J. van Apeldoorn, A. Cornelissen, A. Gilyén & G. Nannicini. Quantum tomography using state-preparation unitaries, arXiv:2207.08800 (2022).

[242] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt et al. Recent developments in the PySCF program package. *The Journal of Chemical Physics*, **153**, 024109 (2020).

[243] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan & G. Knizia. Automated construction of molecular active spaces from atomic valence orbitals. *Journal of Chemical Theory and Computation*, **13**, 4063–4078 (2017).

[244] GitHub repository (2024).

[245] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao et al. OpenFermion: The electronic structure package for quantum computers. *Quantum Science and Technology*, **5**, 034014 (2020).

[246] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo & L. Visscher. Calculating energy derivatives for quantum chemistry on a quantum computer. *npj Quantum Information*, **5**, 113 (2019).

[247] L. Clinton, T. S. Cubitt, R. Garcia-Patron, A. Montanaro, S. Stanisic & M. Stroeks. Quantum phase estimation without controlled unitaries, arXiv:2410.21517 (2024).

[248] D. W. Berry, M. Kieferová, A. Scherer, Y. R. Sanders, G. H. Low, N. Wiebe, C. Gidney & R. Babbush. Improved techniques for preparing eigenstates of fermionic Hamiltonians. *npj Quantum Information*, **4**, 22 (2018).

[249] S.-i. Amari & H. Nagaoka. *Methods of Information Geometry*. American Mathematical Soc. (2000). ISBN 978-0-8218-4302-4.

[250] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning Series. MIT Press, Cambridge, MA (2012). ISBN 978-0-262-01802-9.

[251] F. Nielsen. WHAT IS...an Information Projection? *Notices of the American Mathematical Society*, **65**, 1 (2018).

[252] D. Patel, S. J. S. Tan, Y. Subasi & A. T. Sornborger. Optimal coherent quantum phase estimation via tapering, arXiv:2403.18927 (2024).

[253] A. Cornelissen & Y. Hamoudi. A sublinear-time quantum algorithm for approximating partition functions. In *Proceedings of the 2023 annual ACM-Siam symposium on discrete algorithms (SODA)*, pages 1245–1264. SIAM (2023).

[254] J. van Apeldoorn, A. Cornelissen, A. Gilyén & G. Nannicini. Quantum tomography using state-preparation unitaries. In *Proceedings of the 2023 annual ACM-SIAM symposium on discrete algorithms (SODA)*, pages 1265–1318. SIAM (2023).

[255] G. Wang, D. S. França, G. Rendon & P. Johnson. Faster ground state energy estimation on early fault-tolerant quantum computers via rejection sampling, arXiv:2304.09827v1 (2024).

[256] G. H. Low & I. L. Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Physical Review Letters*, **118**, 010501 (2017).

[257] Z. Cai. A practical framework for quantum error mitigation, arXiv:2110.05389 (2021).

[258] D. Aharonov, O. Alberton, I. Arad, Y. Atia, E. Bairey, M. B. Dov et al. Reliable high-accuracy error mitigation for utility-scale quantum circuits, arXiv:2508.10997 (2025).

[259] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**, 1–26 (1979).

[260] T. Baumgratz & A. Datta. Quantum enhanced estimation of a multidimensional field. *Physical Review Letters*, **116**, 030801 (2016).

[261] A. Dutt, E. Pednault, C. W. Wu, S. Sheldon, J. Smolin, L. Bishop & I. L. Chuang. Active learning of quantum system Hamiltonians yields query advantage. *Physical Review Research*, **5**, 033060 (2023).

[262] C. Ferrie, C. E. Granade & D. G. Cory. How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies. *Quantum Information Processing*, **12**, 611–623 (2013).

[263] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett & H. M. Wiseman. Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis. *Physical Review A*, **84**, 052315 (2011).

[264] S. Pang & T. A. Brun. Quantum metrology for a general Hamiltonian parameter. *Physical Review A*, **90**, 022117 (2014).

[265] J. M. E. Fraïsse & D. Braun. Enhancing sensitivity in quantum metrology by Hamiltonian extensions. *Physical Review A*, **95**, 062342 (2017).

[266] Z. Hou, R.-J. Wang, J.-F. Tang, H. Yuan, G.-Y. Xiang, C.-F. Li & G.-C. Guo. Control-enhanced sequential scheme for general quantum parameter estimation at the Heisenberg limit. *Physical Review Letters*, **123**, 040501 (2019).

[267] K. C. Young, M. Sarovar, R. Kosut & K. B. Whaley. Optimal quantum multiparameter estimation and application to dipole-and exchange-coupled qubits. *Physical Review A*, **79**, 062301 (2009).

[268] F. M. Sbahi, A. J. Martinez, S. Patel, D. Saberi, J. H. Yoo, G. Roeder & G. Verdon. Provably efficient variational generative modeling of quantum many-body systems via quantum-probabilistic information geometry, arXiv:2206.04663 (2022).

[269] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe et al. Experimental quantum Hamiltonian learning. *Nature Physics*, **13**, 551–555 (2017).

[270] N. Wiebe, C. Granade, C. Ferrie & D. G. Cory. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, **112**, 190501 (2014).

[271] N. Wiebe, C. Granade & D. G. Cory. Quantum bootstrapping via compressed quantum Hamiltonian learning. *New Journal of Physics*, **17**, 022005 (2015).

[272] A. Valenti, E. van Nieuwenburg, S. Huber & E. Greplova. Hamiltonian learning for quantum error correction. *Physical Review Research*, **1**, 033092 (2019).

[273] D. Hangleiter, I. Roth, J. Fuksa, J. Eisert & P. Roushan. Robustly learning the Hamiltonian dynamics of a superconducting quantum processor. *Nature Communications*, **15**, 9595 (2024).

[274] M. P. da Silva, O. Landon-Cardinal & D. Poulin. Practical characterization of quantum devices without tomography. *Physical Review Letters*, **107**, 210404 (2011).

[275] R. D. Somma & S. Boixo. Parameter estimation with mixed-state quantum computation. *Physical Review A*, **77**, 052320 (2008).

[276] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut & H. Rabitz. Estimation of many-body quantum Hamiltonians via compressive sensing. *Physical Review A*, **84**, 012107 (2011).

[277] J. Zhang & M. Sarovar. Quantum Hamiltonian identification from measurement time traces. *Physical Review Letters*, **113**, 080401 (2014).

[278] Y. Wang, D. Dong, B. Qi, J. Zhang, I. R. Petersen & H. Yonezawa. A quantum Hamiltonian identification algorithm: Computational complexity and error analysis. *IEEE Transactions on Automatic Control*, **63**, 1388–1403 (2017).

[279] S. Krastanov, S. Zhou, S. T. Flammia & L. Jiang. Stochastic estimation of dynamical variables. *Quantum Science and Technology*, **4**, 035003 (2019).

[280] T. J. Evans, R. Harper & S. T. Flammia. Scalable bayesian Hamiltonian learning, arXiv:1912.07636 (2019).

[281] Z. Li, L. Zou & T. H. Hsieh. Hamiltonian tomography via quantum quench. *Physical Review Letters*, **124**, 160502 (2020).

[282] C. Kokail, B. Sundar, T. V. Zache, A. Elben, B. Vermersch, M. Dalmonte, R. van Bijnen & P. Zoller. Quantum variational learning of the entanglement Hamiltonian. *Physical Review Letters*, **127**, 170501 (2021).

[283] A. Zubida, E. Yitzhaki, N. H. Lindner & E. Bairey. Optimal short-time measurements for Hamiltonian learning, arXiv:2108.08824 (2021).

[284] D. Rattacaso, G. Passarelli & P. Lucignano. High-accuracy Hamiltonian learning via delocalized quantum state evolutions. *Quantum*, **7**, 905 (2023).

[285] N. Kura & M. Ueda. Finite-error metrological bounds on multiparameter Hamiltonian estimation. *Physical Review A*, **97**, 012101 (2018).

[286] H.-Y. Huang, Y. Tong, D. Fang & Y. Su. Learning many-body Hamiltonians with heisenberg-limited scaling. *Physical Review Letters*, **130**, 200403 (2023).

[287] W. Yu, J. Sun, Z. Han & X. Yuan. Robust and efficient Hamiltonian learning. *Quantum*, **7**, 1045 (2023).

[288] A. Anshu, S. Arunachalam, T. Kuwahara & M. Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, **17**, 931–935 (2021).

[289] J. Haah, R. Kothari & E. Tang. Learning quantum Hamiltonians from high-temperature gibbs states and real-time evolutions. *Nature Physics*, **20**, 1027–1031 (2024).

[290] S.-T. Wang, D.-L. Deng & L.-M. Duan. Hamiltonian tomography for quantum many-body systems with arbitrary couplings. *New Journal of Physics*, **17**, 093017 (2015).

[291] A. Gu, L. Cincio & P. J. Coles. Practical Hamiltonian learning with unitary dynamics and Gibbs states. *Nature Communications*, **15**, 312 (2024).

[292] C. Di Franco, M. Paternostro & M. Kim. Hamiltonian tomography in an access-limited setting without state initialization. *Physical Review Letters*, **102**, 187203 (2009).

[293] D. Burgarth, K. Maruyama & F. Nori. Indirect quantum tomography of quadratic Hamiltonians. *New Journal of Physics*, **13**, 013019 (2011).

[294] E. Bairey, I. Arad & N. H. Lindner. Learning a local Hamiltonian from local measurements. *Physical Review Letters*, **122**, 020504 (2019).

[295] J. R. Garrison & T. Grover. Does a single eigenstate encode the full Hamiltonian? *Physical Review X*, **8**, 021026 (2018).

[296] X.-L. Qi & D. Ranard. Determining a local Hamiltonian from a single eigenstate. *Quantum*, **3**, 159 (2019).

[297] J. Liu & H. Yuan. Quantum parameter estimation with optimal control. *Physical Review A*, **96**, 012117 (2017).

[298] J. Kiukas, K. Yuasa & D. Burgarth. Remote parameter estimation in a quantum spin chain enhanced by local control. *Physical Review A*, **95**, 052132 (2017).

[299] M. P. Stenberg, Y. R. Sanders & F. K. Wilhelm. Efficient estimation of resonant coupling between quantum systems. *Physical Review Letters*, **113**, 210404 (2014).

[300] T. E. O'Brien, L. B. Ioffe, Y. Su, D. Fushman, H. Neven, R. Babbush & V. Smelyanskiy. Quantum computation of molecular structure using data from challenging-to-classically-simulate nuclear magnetic resonance experiments. *PRX Quantum*, **3**, 030345 (2022).

[301] M. Holzäpfel, T. Baumgratz, M. Cramer & M. B. Plenio. Scalable reconstruction of unitary processes and Hamiltonians. *Physical Review A*, **91**, 042129 (2015).

[302] D. Sels, H. Dashti, S. Mora, O. Demler & E. Demler. Quantum approximate bayesian computation for nmr model inference. *Nature Machine Intelligence*, **2**, 396–402 (2020).

[303] K. Seetharam, D. Biswas, C. Noel, A. Risinger, D. Zhu, O. Katz et al. Digital quantum simulation of nmr experiments. *Science Advances*, **9**, eadh2594 (2023).

[304] J. Carrasco, A. Elben, C. Kokail, B. Kraus & P. Zoller. Theoretical and experimental perspectives of quantum verification. *PRX Quantum*, **2**, 010102 (2021).

[305] W. Wootters. Statistical distance and Hilbert space. *Physical Review D*, **23**, 357 (1981).

[306] S. L. Braunstein & C. M. Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, **72**, 3439 (1994).

[307] S. L. Braunstein, C. M. Caves & G. Milburn. Generalized uncertainty relations: Theory, examples, and lorentz invariance. *Annals of Physics*, **247**, 135–173 (1996).

[308] V. Giovannetti, S. Lloyd & L. Maccone. Quantum metrology. *Physical Review Letters*, **96**, 010401 (2006).

[309] N. F. Ramsey. A molecular beam resonance method with separated oscillating fields. *Physical Review*, **78** (1950).

[310] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young & R. Blume-Kohout. Gate set tomography. *Quantum*, **5**, 557 (2021).

[311] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends et al. Observation of separated dynamics of charge and spin in the fermi-hubbard model, arXiv:2010.07965 (2020).

[312] K. Wan & R. Lasenby. Bounds on adaptive quantum metrology under markovian noise. *Physical Review Research*, **4**, 033092 (2022).

[313] J. Haah, R. Kothari, R. O'Donnell & E. Tang. Query-optimal estimation of unitary channels in diamond distance. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 363–390. IEEE (2023).

[314] M. Srednicki. Chaos and quantum thermalization. *Physical Review E*, **50**, 888 (1994).

[315] M. Rigol, V. Dunjko & M. Olshanii. Thermalization and its mechanism for generic isolated quantum systems. *Nature*, **452**, 854–858 (2008).

[316] L. D'Alessio, Y. Kafri, A. Polkovnikov & M. Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, **65**, 239–362 (2016).

[317] J. M. Deutsch. Eigenstate thermalization hypothesis. *Reports on Progress in Physics*, **81**, 082001 (2018).

[318] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak & M. Troyer. Solving strongly correlated electron models on a quantum computer. *Physical Review A*, **92**, 062318 (2015).

[319] M. Kliesch, C. Gogolin, M. Kastoryano, A. Riera & J. Eisert. Locality of temperature. *Physical Review X*, **4**, 031019 (2014).

[320] A. Bluhm, Á. Capel & A. Pérez-Hernández. Exponential decay of mutual information for gibbs states of local Hamiltonians. *Quantum*, **6**, 650 (2022).

[321] D. Stilck França, L. A. Markovich, V. Dobrovitski, A. H. Werner & J. Borregaard. Efficient and robust estimation of many-qubit Hamiltonians. *Nature Communications*, **15**, 311 (2024).

[322] J. Liu & H. Yuan. Control-enhanced multiparameter quantum estimation. *Physical Review A*, **96**, 042114 (2017).

[323] S. Pang & A. N. Jordan. Optimal adaptive control for quantum metrology with time-dependent Hamiltonians. *Nature communications*, **8**, 14695 (2017).

[324] D. Abanin, W. De Roeck, W. W. Ho & F. Huveneers. A rigorous theory of many-body prethermalization for periodically driven and closed quantum systems. *Communications in Mathematical Physics*, **354**, 809–827 (2017).

[325] R. M. Wilcox. Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, **8**, 962–982 (1967).

[326] C.-F. Chen & F. G. Brandão. Fast thermalization from the eigenstate thermalization hypothesis, arXiv:2112.07646 (2021).

[327] A. Dymarsky. Bound on eigenstate thermalization from transport. *Physical Review Letters*, **128**, 190601 (2022).

[328] L. Foini & J. Kurchan. Eigenstate thermalization hypothesis and out of time order correlators. *Physical Review E*, **99**, 042139 (2019).

[329] A. Chan, A. De Luca & J. Chalker. Eigenstate correlations, thermalization, and the butterfly effect. *Physical Review Letters*, **122**, 220601 (2019).

[330] C. Murthy & M. Srednicki. Bounds on chaos from the eigenstate thermalization hypothesis. *Physical Review Letters*, **123**, 230606 (2019).

[331] J. Wang, M. H. Lamann, J. Richter, R. Steinigeweg, A. Dymarsky & J. Gemmer. Eigenstate thermalization hypothesis and its deviations from random-matrix theory beyond the thermalization time. *Physical Review Letters*, **128**, 180601 (2022).

[332] D. L. Jafferis, D. K. Kolchmeyer, B. Mukhametzhanov & J. Sonner. Matrix models for eigenstate thermalization. *Physical Review X*, **13**, 031033 (2023).

[333] J. Cotler, N. Hunter-Jones, J. Liu & B. Yoshida. Chaos, complexity, and random matrices. *Journal of High Energy Physics*, **2017**, 1–60 (2017).

[334] M. Schiulaz, E. J. Torres-Herrera & L. F. Santos. Thouless and relaxation time scales in many-body quantum systems. *Physical Review B*, **99**, 174313 (2019).

[335] J. Cotler & N. Hunter-Jones. Spectral decoupling in many-body quantum chaos. *Journal of High Energy Physics*, **2020**, 1–62 (2020).

[336] J. Cotler, N. Hunter-Jones & D. Ranard. Fluctuations of subsystem entropies at late times. *Physical Review A*, **105**, 022416 (2022).

[337] C. Murthy, A. Babakhani, F. Iniguez, M. Srednicki & N. Yunger Halpern. Non-abelian eigenstate thermalization hypothesis. *Physical Review Letters*, **130**, 140402 (2023).

[338] A. Altland & B. D. Simons. *Condensed matter field theory*. Cambridge university press (2010).

[339] J. A. Mingo & R. Speicher. *Free probability and random matrices*, volume 35. Springer (2017).

[340] R. D. Dutton & R. C. Brigham. Computationally efficient bounds for the catalan numbers. *European Journal of Combinatorics*, **7**, 211–213 (1986).

[341] A. Avdoshkin, L. Astrakhantsev, A. Dymarsky & M. Smolkin. Rate of cluster decomposition via Fermat-Steiner point. *Journal of High Energy Physics*, **2019**, 1–15 (2019).

[342] R. W. Bailey. The number of weak orderings of a finite set. *Social Choice and Welfare*, pages 559–562 (1998).

[343] A. Dutkiewicz, T. E. O'Brien & S. Polla. Accurate ground state energy estimation with noise and imperfect state preparation, arXiv:2603.21873 (2026).

[344] A. Dutkiewicz, T. E. O'Brien & T. Schuster. The advantage of quantum control in many-body Hamiltonian learning. *Quantum*, **8**, 1537 (2024).

Acknowledgments

First, I would like to thank my promotors for your guidance and for creating an environment in which I felt both supported and free to grow. Tom, when we first met, I was at the very beginning of my academic journey, and under your mentorship I grew into the researcher I am today. You taught me to aim for ambitious projects and to carry them through with care and rigor. Even though most of our collaboration was online, you consistently made time for discussions and made me feel supported throughout. I am grateful for the many opportunities you introduced me to, and for your confidence in me that encouraged me to pursue them. Carlo, thank you for fostering an inspiring and supportive scientific environment, and for giving me the opportunity to be part of it. I am grateful for your ongoing support, for the independence you offered, and for the trust you extended to me.

The road to this PhD began long before the PhD itself, and I am thankful to those who set me on this path. I thank my family for their encouragement and for always supporting my curiosity. I am grateful to Nina for recognizing my enthusiasm and encouraging me to explore physics further. Rafał, thank you for sparking my interest in quantum information, and for teaching me the tools I would go on to use throughout my PhD. Jakub, you were the first to explain the basics of quantum computing to me, and I am grateful for your mentorship through my first steps in research and for bringing me to Leiden.

Research is a collective endeavour, and I am grateful to all those with whom I have had the pleasure to collaborate. Stefano, thank you for everything I have learned with you and from you. Ake, Armando, Barbara, Jonas, Kshiti, Maarten, Nick, Tommy, thank you for the many discussions, ideas, and projects we shared, including those that never became publications but nonetheless shaped my thinking and made doing research enjoyable. I am also grateful to the many people I met at conferences, for the fresh perspectives, shared experiences, and for making me feel part of a wider scientific community.

I have been fortunate to be part of a vibrant and welcoming community that shaped my daily life throughout this PhD. Ania, Evert, Jordi, Vedran, thank you for creating a stimulating research environment, and Fran and Nina for keeping everything running smoothly behind the scenes. I am thankful to all my colleagues in aQa and the Lorentz Institute for the warm and lively atmosphere you created, making each visit to Leiden a highlight of my work

week. Special thanks to my paranymphs, Mahtab and Ilse, for all the fun moments we shared and for your support during this final stage. I am also thankful to the friends I have made through science – Aravindh, Emiel, Jelena, Kira, Marten, Masha and Norman – for being there for me throughout this journey.

Finally, I want to thank the people outside of academia who were part of my life during these years, with special thanks to those who left their mark on this thesis. Gina, thank you for bringing your creativity to my thesis and for helping me stay connected to that side of myself. Thomas, for your endless enthusiasm and curiosity, which consistently reignite my excitement for science. Peter and Weronika, for the company and good times during much-needed breaks from research. And Kamil, for being there with me, and for me, every step of the way.

Summary

After decades of existing only on paper, quantum computers are now becoming real. The last decade has seen the construction of the first working prototypes, and significant engineering effort is underway to scale these devices up in the years ahead. The most promising near-term application for this emerging technology is quantum simulation – using one quantum system to study another, such as modelling the electronic structure of molecules or the behaviour of exotic materials. Yet these early devices are far from the powerful machines theorists have long envisioned. Quantum systems are extraordinarily fragile: even tiny, unavoidable interactions with the surrounding environment introduce errors that corrupt computations, and the number of quantum bits (*qubits*) that can be reliably operated remains limited. This thesis addresses the question of how to extract accurate and efficient scientific results from quantum simulations given these constraints.

Chapter I introduces the theoretical background required for the remainder of the thesis. It begins with the basics of quantum computation, with an emphasis on different approaches to managing noise. We then review quantum simulation as a target application, and explain how parameter estimation fits within this task, before introducing the tools of quantum estimation theory used throughout this thesis. The chapter concludes with a discussion of quantum phase estimation (QPE), an algorithmic primitive central to this thesis, which refers to inferring energy levels of a quantum system from measurements. Chapters II–IV study quantum phase estimation in increasingly realistic settings, progressively relaxing the assumptions of ideal state preparation and noiseless computation.

Chapter II studies a class of QPE algorithms referred to as *single-control*, which use small, low-depth quantum circuits together with classical post-processing to extract energy levels, making them naturally suited to noisy devices with a limited number of qubits. Standard QPE assumes access to an eigenstate – a state with a single, definite energy. In practice this is itself a hard problem, and realistic states will overlap with several energy levels at once. This chapter provides a single-control phase estimation algorithm that, given such a state, simultaneously estimates all the corresponding energy levels. The algorithm achieves the optimal scaling of computational time required with the target precision, known as the *Heisenberg limit*.

Chapter III turns to *QFT-based* QPE algorithms, which use larger quantum circuits to sample energy levels more directly, but require increasing circuit depth to achieve better precision. Because larger circuits accumulate more errors, there is an inherent tension between circuit size and result quality. This chapter analyses this trade-off to identify the optimal circuit depth for a given noise level, and introduces a scheme that uses maximum likelihood estimation to achieve arbitrarily high precision by increasing the number of circuit repetitions rather than their depth. This is enabled by a novel error mitigation technique, *explicit unbiasing*, which corrects for the effect of noise and allows accurate estimation on imperfect devices.

Chapter IV combines the challenges of the two preceding ones. It introduces a *moment-projection estimator* compatible with both QFT-based and single-control circuits. By generalizing the explicit unbiasing technique of Chapter III to the setting where the initial state is not a single eigenstate, we obtain an algorithm that accurately recovers ground state energies even in the presence of both noise and spectral overlap.

Chapter V turns to a different estimation problem: rather than inferring energy levels of a known system, can we learn the governing equations – the Hamiltonian – of a quantum system by observing its dynamics? The answer turns out to depend critically on the level of control available during the experiment. Passive observation is shown to be fundamentally insufficient to achieve optimal scaling for large classes of physically relevant systems. In contrast, with the ability to apply additional operations or modify the Hamiltonian during evolution, the chapter provides an algorithm that learns the Hamiltonian at the Heisenberg limit. This establishes a rigorous quadratic advantage for quantum control in Hamiltonian learning.

Samenvatting

Na decennialang alleen op papier te hebben bestaan, worden quantumcomputers nu werkelijkheid. Het afgelopen decennium zijn de eerste werkende prototypes gebouwd en er wordt momenteel hard gewerkt aan de opschaling van deze apparaten in de komende jaren. De meest veelbelovende toepassing van deze opkomende technologie op korte termijn is quantumsimulatie – het gebruik van een quantumsysteem om een ander quantumsysteem te bestuderen, zoals het modelleren van de elektronische structuur van moleculen of het gedrag van exotische materialen. Deze eerste apparaten zijn echter nog lang niet de krachtige machines die theoretici al lang voor ogen hebben. Quantumsystemen zijn buitengewoon fragiel: zelfs minuscule, onvermijdelijke interacties met de omgeving leiden tot fouten die berekeningen verstoren, en het aantal quantumbits (*qubits*) dat betrouwbaar kan worden gebruikt, blijft beperkt. Deze dissertatie behandelt de vraag hoe nauwkeurige en efficiënte wetenschappelijke resultaten uit quantumsimulaties kunnen worden verkregen, gegeven deze beperkingen.

Hoofdstuk I introduceert de theoretische achtergrond die nodig is voor de rest van de dissertatie. Het begint met de basisprincipes van quantumcomputing, met de nadruk op verschillende benaderingen voor het beheersen van ruis. Vervolgens bespreken we quantumsimulatie als beoogde toepassing en leggen we de rol van parameterestimatie uit, alvorens de instrumenten van de quantumestimatietheorie te introduceren die in dit proefschrift worden gebruikt. Het hoofdstuk sluit af met een bespreking van quantumfase-estimatie (QPE), een groep van algoritmen die centraal staat in dit proefschrift en die verwijst naar het afleiden van energieniveaus van een quantumsysteem uit specifieke metingen. Hoofdstukken II–IV bestuderen quantumfase-estimatie met steeds realistischere restricties, waarbij de aannames van ideale toestandsvoorbereiding en ruisvrije computatie geleidelijk worden versoepeld.

Hoofdstuk II bestudeert een klasse van QPE-algoritmen die *single-control* worden genoemd. Deze algoritmen gebruiken kleine quantumcircuits met een kleine diepte in combinatie met klassieke nabewerking van de meetresultaten om energieniveaus af te leiden, waardoor ze van nature geschikt zijn voor ruisgevoelige apparaten met een beperkt aantal qubits. Standaard QPE veronderstelt toegang tot een eigenstaat – een toestand met één enkele, goed gedefinieerde energie. In de praktijk is het verkrijgen van deze staat op

zichzelf een moeilijk probleem en zullen realistische toestanden overlappen met meerdere energieniveaus tegelijk. Dit hoofdstuk beschrijft een algoritme voor single-control quantumfase-estimatie dat, gegeven een dergelijke toestand, gelijktijdig alle corresponderende energieniveaus schat. De benodigde rekentijd van het algoritme voldoet aan de optimale schaling met de gewenste precisie, bekend als de *Heisenberg-limiet*.

Hoofdstuk III behandelt *QFT-based* QPE-algoritmen, die grotere quantumcircuits gebruiken om energieniveaus directer te meten, maar een toenemende circuitdiepte vereisen om een betere precisie te bereiken. Omdat grotere circuits meer fouten accumuleren, bestaat er een inherente tweestrijd tussen circuitgrootte en resultaatkwaliteit. Dit hoofdstuk analyseert deze afweging om de optimale circuitdiepte voor een gegeven ruisniveau te bepalen en introduceert een methode die willekeurig hoge precisie kan bereiken door het aantal circuitherhalingen te verhogen in plaats van de dieptes te vergroten. Dit wordt mogelijk gemaakt door een nieuwe foutreductietechniek, *explicit unbiasing*, die het effect van ruis corrigeert en nauwkeurige schattingen mogelijk maakt op imperfecte apparaten.

Hoofdstuk IV combineert de uitdagingen van de twee voorgaande hoofdstukken. Het introduceert een *moment-projectie* schatter die compatibel is met zowel op QFT gebaseerde circuits als op single-control circuits. Door de explicit unbiasing techniek van hoofdstuk III te generaliseren naar de situatie waarin de begintoestand geen perfecte eigenstaat is, verkrijgen we een algoritme dat de grondtoestandenergieën nauwkeurig bepaalt, zelfs in aanwezigheid van zowel ruis als spectrale overlap.

Hoofdstuk V behandelt een ander estimatieprobleem: kunnen we, in plaats van energieniveaus van een bekend systeem af te leiden, de drijvende kracht – de Hamiltoniaan – van een quantumsysteem leren door de dynamische ontwikkeling ervan te observeren? Het antwoord blijkt sterk af te hangen van de mate van controle die tijdens het experiment beschikbaar is. Passieve observatie blijkt fundamenteel ontoereikend te zijn om optimale schaling te bereiken voor grote klassen van fysiek relevante systemen. Daarentegen, gegeven de mogelijkheid om extra operaties uit te voeren of de Hamiltoniaan tijdens de evolutie te wijzigen, biedt het hoofdstuk een algoritme dat de Hamiltoniaan leert in de Heisenberg-limiet. Dit toont aan dat quantumcontrole een rigoreus kwadratisch voordeel biedt bij het leren van de Hamiltoniaan.

Curriculum Vitæ

I was born on the 2nd of June 1997 in Kraków, Poland. From an early age, I was drawn to science, a curiosity encouraged by my family and developed further during my education at XXV Liceum Ogólnokształcące im. Tadeusza Czackiego in Warsaw, where I followed an extended curriculum in physics and mathematics. Under the guidance of my physics teacher, Nina Tomaszewska, I developed a lasting interest in the subject.

In 2016, I began my Bachelor's studies in Physics at the University of Warsaw, enrolling in the selective *Individual Track* programme. During my studies, I sought out research experience beyond the curriculum, including an internship in medical physics at the Jagiellonian University in Kraków. A turning point came when I attended a course on quantum estimation and metrology taught by Rafał Demkowicz-Dobrzański, which led me toward quantum information science. I subsequently carried out research under the supervision of Jakub Tworzydło and graduated *summa cum laude* in 2019. In my final undergraduate year, I visited Leiden as an Erasmus intern in the group of Carlo Beenakker at the Lorentz Institute, working with Thomas O'Brien—an experience that shaped my future academic path.

Motivated by this experience, I continued to Leiden for my Master's studies, supported by a LION scholarship. I joined the Applied Quantum Algorithms Leiden group and contributed to its early development, while also gaining experience in experimental quantum optics in the group of Wolfgang Löffler. I completed my Master's thesis under the supervision of Thomas O'Brien and Barbara Terhal, graduating *summa cum laude* in 2021.

Following my Master's, I was awarded a Google PhD Fellowship to pursue research on quantum estimation methods for early fault-tolerant quantum simulation, forming the basis of this thesis. I remained in Leiden for my doctoral studies under the supervision of Thomas O'Brien and Carlo Beenakker. Shortly before the start of my PhD, Thomas O'Brien joined Google Quantum AI, which shaped the structure of my research: I conducted my work largely independently in Leiden while maintaining close collaboration with him and his colleagues. I further strengthened this connection through two research internships at Google Quantum AI, including one with Nicholas Rubin. I also collaborated with researchers across the Netherlands, including at QuSoft in Amsterdam and QuTech in Delft, benefiting from the country's highly

interconnected quantum research community. I presented my work at several international conferences and schools, including a contributed talk at the Quantum Information Processing (QIP) conference in Taipei in 2024 and an invited talk at a Quantum Sensing workshop in Chicago in the same year. Alongside my research, I was actively involved in teaching, serving as a teaching assistant, delivering guest lectures, and supervising a Master's student. I also contributed to community initiatives through my involvement with Women in Quantum Development.

Upon completing my doctorate, I was awarded the Ada Lovelace Fellowship and joined the group of Jonas Helsen at QuSoft (CWI, Amsterdam) as a postdoctoral researcher. There, I continue my work on early fault-tolerant quantum algorithms, building on the foundations developed in this thesis.

List of publications

[343] A. Dutkiewicz, S. Polla & T. E. O'Brien.

Accurate ground state energy estimation with noise and imperfect state preparation. arXiv:2603.21873.

[Chapter IV is based on this preprint.]

[76] A. Dutkiewicz, S. Polla, M. Scheurer, C. Gogolin, W. J. Huggins & T. E. O'Brien.

Error mitigation and circuit division for early fault-tolerant quantum phase estimation. *PRX Quantum*, **6**, 040318 (2025).

[Chapter III is based on this publication.]

[344] A. Dutkiewicz, T. E. O'Brien & T. Schuster.

The advantage of quantum control in many-body Hamiltonian learning. *Quantum*, **8**, 1537 (2024).

[Chapter V is based on this publication.]

[75] A. Dutkiewicz, B. M. Terhal & T. E. O'Brien.

Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits. *Quantum*, **6**, 830 (2022).

[Chapter II is based on this publication.]

[246] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, & L. Visscher.

Calculating energy derivatives for quantum chemistry on a quantum computer. *npj Quantum Information*, **5 (1)**, 113 (2019).

Stellingen

Behorende bij het proefschrift
"Quantum Parameter Estimation
for Early Fault-Tolerant Quantum Simulation"

1. Heisenberg-limited estimation of multiple eigenphases can be achieved without increasing the size of the control register.
   [Chapter 2; Z. Ding et al, Quantum 8, 1487 (2024).]

2. Quantum error mitigation can be applied to quantum Fourier transform–based phase estimation. [Chapter 3, 4].

3. Phase estimation in the early fault-tolerant regime requires sophisticated classical post-processing beyond simple estimators such as single-shot outcomes or majority voting. [Chapter 3, 4]

4. Passive observation of a quantum system is insufficient for Heisenberg-limited learning of many-body Hamiltonians. Achieving the Heisenberg limit in many-body Hamiltonian learning requires coherent quantum control. [Chapter 5]

5. Quantum control allows many-body Hamiltonians to be learned with polynomial classical resources.
   [Chapter 5; H.Y Huang, et al. PRL 130.20 (2023): 200403.]

6. Ideas developed in quantum sensing and quantum metrology provide a useful source of inspiration for quantum algorithm design.

7. Combining quantum error correction with quantum error mitigation will be essential for extracting useful computations from early fault-tolerant quantum computers. Achieving this in practice will require efficient methods for characterizing large quantum systems.

8. The most impactful real-world applications of quantum computing have likely not yet been identified. The search for useful quantum algorithms should explore both problems that are classically intractable and ways to perform existing computations more efficiently, for example faster or with lower energy consumption.

9. Research in quantum computing is already producing valuable results by inspiring improved classical algorithms and deepening understanding of fundamental disciplines such as quantum mechanics, information theory, and computational complexity. These benefits arise independently of whether large-scale quantum computers are ultimately realized.

10. Academic institutions should offer a wider diversity of career paths and value the distinct skills required for each. Different forms of contribution such as teaching, mentorship, and management, should be rewarded on par with research output.

11. Efforts to improve diversity in academia should address a broad range of marginalized groups rather than focusing exclusively on women.

12. Diversity policies such as committee quotas are counterproductive, as they overburden academics from underrepresented groups with invisible labor without addressing structural barriers to hiring and retention.

Alicja Dutkiewicz
Leiden, 22/05/2026