

Quantum-enhanced estimation of physical properties

February 2026

Kaito Wada

A Thesis for the Degree of Ph.D. in Science

Quantum-enhanced estimation of physical properties

February 2026

Graduate School of Science and Technology
Keio University

Kaito Wada

Abstract

Quantum computers have great potential to uncover new and unexpected physical phenomena. They can simulate classically intractable many-particle quantum systems and probe them in a programmable, computational manner. They can also process experimental quantum data directly, enabling us to probe unseen aspects of physical systems that cannot be reached by any classical means. In both roles, refining quantum algorithms toward their fundamental limits expands and sharpens the reach of quantum computers.

Among quantum algorithms, quantum estimation is of central importance because it elicits the physical properties of interest from quantum systems. Even in an ideal, noiseless setting, however, obtaining precise estimates of such properties is significantly resource-intensive, unless one leverages coherent quantum operations. While many quantum-enhanced (i.e., higher performance than any classical means enabled by quantum effects) estimation algorithms have been proposed under reasonable access assumptions, it is still an ongoing quest to design quantum algorithms achieving ultimate estimation efficiency allowed by quantum mechanics. In addition, rapid experimental progress makes us envision an early fault-tolerant regime with large-scale yet not fully error-corrected quantum computers. Toward this regime, it is also crucial to develop quantum estimation algorithms that elicit as much information of interest as possible from a quantum system subject to noise. In these contexts, this thesis advances quantum-enhanced algorithms for estimating physical properties in two respects: Heisenberg-limited multi-observable estimation and single-observable estimation for the early fault-tolerant quantum computing (FTQC) regime.

First, we provide an adaptive quantum algorithm for estimating expectation values of multiple observables simultaneously in an ideal (noiseless) setting. Given common access to block-encoded observables, a state-preparation unitary, and its inverse, this algorithm achieves a double quantum enhancement in query complexity to the state preparation and its inverse: the mean-squared-error (MSE) scaling at the Heisenberg limit and sublinear scaling in the number of observables. It also essentially matches the query lower bound in the small-additive-error regime (not in MSE), previously derived in quantum complexity theory. In addition, the proposed algorithm significantly reduces the number of ancilla qubits and the classical computation for circuit synthesis, compared to prior art for this task. The double enhancement and the circuit simplification come from a novel algorithm design that integrates key ideas from adaptive phase estimation, quantum amplification, and quantum gradient estimation.

Next, we focus on expectation value estimation for a single observable in the early FTQC regime. In this regime, available coherent resources are mostly consumed for a state preparation unitary, and the dominant noise stems from that costly unitary and its inverse. Under a specific model reflecting these aspects, we derive the quantum-enhanced precision limit, that is, the minimal quantum Cramér-Rao bound over all possible estimation strategies in the model. Then, by designing an adaptive optimization based on Fisher information, we propose a quantum-enhanced estimation algorithm, with rigorous estimation guarantees, that asymptotically achieves the quantum Cramér-Rao bound, when the target system size is sufficiently large. Finally, we numerically demonstrate that the asymptotic behavior holds with only a few hundred measurements per step and a few tens of qubits.

Acknowledgment

This research was conducted while the author was enrolled in the doctoral program of the Graduate School of Science and Technology, Keio University, under the supervision of Professor Naoki Yamamoto. I would like to express my sincere gratitude to Professor Naoki Yamamoto, who served as the chair of my thesis committee and my supervisor, for his guidance throughout my research. Through our daily discussions and his advice during the writing of manuscripts, I gained deep insight into what it means to conduct research with integrity and rigor. Moreover, thanks to him, I was able to write many papers on topics beyond the main focus of this thesis. There were difficult times when my submissions were repeatedly rejected; however, I can now genuinely say that I find research deeply enjoyable. The true fascination of the research world he taught me will remain a guiding principle in my life. I am also deeply grateful to Professor Hiroshi Watanabe, Professor Masahiro Takeoka, and Associate Professor Shunsuke Furukawa at Keio University for serving as committee members and for providing valuable comments and advice from diverse perspectives.

I would like to extend my heartfelt thanks to Associate Professor Nobuyuki Yoshioka at the University of Tokyo for holding in-depth discussions with me on a biweekly basis from the early stages of this work. He provided invaluable insights and guidance at every step, leading to the completion of my papers, including how to identify a meaningful research topic and how to bring a project to completion. I was also given opportunities to discuss further developments stemming from this research, for which I am truly grateful.

I am grateful to Hiroshi C Watanabe, Rudy Raymond, Yuki Sato, Michihiko Sugawara, Yu-ya Ohnishi, and Eriko Kaminishi, who were, at the time, affiliated with the Keio University Quantum Computing Center (KQCC), for their generous support from the very early stages of my research. In particular, it was a valuable experience to have my first paper carefully reviewed and thoroughly corrected at the draft stage. In addition, I would like to offer my special thanks to Kohei Oshio, Jumpei Kato, Kosuke Ito, and Rei Sakuma at KQCC for arranging opportunities for intensive discussions on important themes in quantum algorithm research, which are also mentioned in this thesis. These discussions ultimately led to the writing of several papers, and I am especially grateful for their support. I would also like to thank Yohichi Suzuki and Tomoki Tanaka, then at KQCC, for deepening discussions related to my research topics.

I would like to thank Yasunari Suzuki and Yuuki Tokunaga, who were, at the time, with the NTT Computer and Data Science Laboratories, for the many lessons I learned through the MITOU Target Program run by the Information-technology Promotion Agency, Japan—not only regarding how to advance research, but also concerning presentation techniques and related skills. Furthermore, during my subsequent internship at NTT, I received substantial support from Suguru Endo, through which I gained knowledge of state-of-the-art quantum algorithms and was given an important opportunity to determine the future direction of my research. I am sincerely grateful.

I would like to express my sincere appreciation to Hiroyuki Harada and Kazuma Fukuchi, my colleagues in the Yamamoto laboratory, for the many fruitful discussions we had daily. It was a stimulating experience that ultimately led to the joint writing of a paper.

I am also grateful to a Project Manager at KQCC, Shinichi Niimi, and to the secretaries, Naoko Oue and Shino Aizawa, for their administrative support regarding my activities at KQCC and my participation in academic conferences. I am deeply grateful for their assistance.

Finally, I would like to express my deepest gratitude to Orié Honda for supporting me throughout the years leading up to the completion of my doctoral program. Thanks to her, my research years were fulfilling, and I was able to persevere until the very end. I would also like to acknowledge my family's support, which was indispensable to maintaining a healthy research life.

Contents

1	Introduction	7
2	Preliminary	11
2.1	Quantum state, operation, and measurement	11
2.2	Quantum circuit	13
2.3	Statistical estimation theory	13
2.3.1	Classical Cramér-Rao inequality	13
2.3.2	Maximum likelihood estimation	14
2.3.3	Quantum Cramér-Rao inequality	15
2.3.4	Heisenberg limit (HL) and standard quantum limit (SQL)	16
3	Fundamental quantum algorithms	19
3.1	Quantum estimation algorithms	19
3.1.1	Quantum phase estimation (QPE)	19
3.1.2	Quantum amplitude estimation (QAE)	21
3.1.3	Quantum gradient estimation (QGE)	22
3.2	Quantum state preparation algorithms	26
3.2.1	Virtual methods	26
3.2.2	Variational methods with sequential optimization	30
3.3	Other basic quantum algorithms	36
3.3.1	Block encoding	36
3.3.2	Quantum singular value transformation (QSVT)	37
4	Heisenberg-limited multi-observable estimation	39
4.1	Introduction	39
4.1.1	Background	39
4.1.2	Summary of results	41
4.2	Problem setup	42
4.3	Adaptive gradient estimation for multiple observables	42
4.3.1	Proposed adaptive algorithm	42
4.3.2	Estimation performance and query complexity	44
4.4	State preparation for adaptive gradient estimation	46
4.5	Circuit implementation and theoretical guarantees	53
4.5.1	Probing-state preparation	53
4.5.2	Maximum mean squared error (MSE) and query complexity	62
4.5.3	Circuit implementation	66
4.6	Numerical simulation	66
4.6.1	More details	68
4.7	Concluding remarks and outlook	76
5	Single-observable estimation for early FTQC	77
5.1	Introduction	77
5.2	Maximum-likelihood QAE	78
5.3	Problem setup	80
5.4	Precision limit with noisy amplitude-amplified states	81
5.5	Adaptive observable estimation under white noise	83

5.5.1	The two-type measurement	83
5.5.2	The proposed adaptive estimation algorithm	85
5.5.3	Statistical properties of the estimator	87
5.5.4	Summary of the proposed protocol	89
5.5.5	Detailed derivations	89
5.6	Numerical experiment	99
5.6.1	Asymptotic properties with respect to the number of measurements	100
5.6.2	Efficiency of the estimator	102
5.6.3	Impact of calibration errors	105
5.7	Concluding remarks and outlook	105
6	Conclusion	107
7	Appendix	109
7.1	Theorems for sequential quantum optimizers	109
7.1.1	Proof of Theorem 7	109
7.1.2	Proof of Theorem 8	112

Chapter 1

Introduction

The development of methods for probing unseen regimes of nature has repeatedly led to drastic changes in our understanding of physics. Here, we use “probe” in a broad sense to include experimental and computational methods. Violations of Bell inequalities [53, 7, 171], demonstrated via advances in precise control and detection of photons, clarified non-classical regimes by ruling out local hidden variable theories and laid the foundations of quantum information science. Likewise, the detection of gravitational waves at the Laser Interferometer Gravitational-wave Observatory (LIGO) [2] has deepened the understanding of strong-gravity dynamics in general relativity by developing a refined probe to extremely small variations of the waves. What comes next in extending our probing capability in nature? In this thesis, we focus on quantum computers as a prominent programmable probe.

Quantum computers provide a computational probe capable of directly handling quantum systems. Every quantum system that evolves by local interactions can be simulated by quantum computers efficiently in terms of the system dimension [112]. The dimension tractable in quantum computers reaches far beyond what can be treated efficiently by classical computers. Thus, quantum computers can perform *ab initio* calculations for strongly correlated many-particle quantum systems in condensed matter physics and quantum chemistry, allowing us to explore new and unexpected phenomena in such systems that remain inaccessible even with sophisticated classical methods. More abstractly, programmable quantum simulation can, in principle, generate macroscopic phenomena from microscopic operations or rules, and further, enable us to probe physical properties in such emergence processes. This sharpens our understanding of how hierarchies across different scales of nature emerge, particularly for the higher organizing principle at macroscopic scales showing insensitivity to microscopic details, making concrete aspects of Anderson’s “more is different” [5, 107].

From an (ideal) experimental perspective, quantum computers can play a crucial role as part of the experimental apparatus to probe target physical systems. At a higher level, experimental procedures can be considered as generalizations of quantum algorithms (which usually have classical input)—state preparation, quantum operations between target physical systems and quantum computers, and final measurement, leading to the concept of quantum experiments with quantum input and classical final output, e.g., unknown states/processes and estimates of their physical properties [3]. Under this description, for instance, ideal quantum computers coherently process the information gathered by quantum sensors, and quantum algorithms with this direct quantum data processing are significantly more efficient than any classical method in terms of the number of required experimental queries in several tasks [31, 81]. The expectation value estimation of multiple non-commuting Pauli strings can be performed with an exponentially small number of state copies with respect to the number of observables, when carefully designed joint measurements on multiple states are available [83, 98]. Also, the principal component of an unknown quantum state (i.e., the most dominant eigenvector of a density matrix) can be measured with constant error by coherently consuming a constant number of copies of an unknown target n -qubit state under certain spectral assumptions; by contrast, conventional approaches measuring each of copies separately requires an exponentially large number 2^n of copies [113, 81]. Hereafter, we use the term “quantum algorithm” to denote a sequence of quantum and classical information processing steps, which in particular includes quantum experiments.

Quantum estimation algorithms have exceptional importance as they extract classical information of interest from quantum systems. Quantum computers can simulate a target quantum [112, 115, 36] and classical process [9, 4], but such a simulation itself is not enough because it only generates resulting

states in a quantum form; their properties of interest are not accessible unless we measure them. We actually need to know several (possibly many) properties of the target state, such as energy, population, magnetization, correlation, and so on. However, obtaining precise estimates of such properties for quantitative analysis is resource-intensive. To see this, we here consider the expectation value estimation of M non-commuting observables in standard sampling approaches. Specifically, our primary resource metric is the total number of calls (i.e., queries) to the state-preparation unitary for a target (pure) state, for estimating the expectation values of M observables within a root mean-squared-error (MSE) ε simultaneously. In the standard sampling, one simply measures the state prepared by the unitary with each observable and repeats this procedure to obtain estimates of the expectation values as averaged values of the measurement outcomes. From the central limit theorem in classical statistics, the standard sampling approach needs a large number $\mathcal{O}(M/\varepsilon^2)$ of queries to the state preparation (or copies of a target state), which reaches quickly unfeasible runtime as precision increases. This scaling $\mathcal{O}(1/\varepsilon^2)$ for the root MSE is known as the standard quantum limit (SQL). In contrast, several additional capabilities can change the attainable resource scaling drastically such as randomized measurements, joint measurements across multiple states, and inverse unitaries of the state preparation, motivating quantum-enhanced (i.e., higher performance than any classical means enabled by quantum effects) estimation algorithms discussed next [82, 83, 100, 86].

By exploiting intriguing quantum effects, we obtain several quantum-enhanced estimation algorithms. The quantum phase estimation (QPE) algorithm [99, 128] yields a precise estimate of the energy of a ground or excited eigenstate, given its reasonable state preparation routine. In cases where a target state is not an eigenstate of a Hamiltonian or another observable, we can use the quantum amplitude estimation (QAE) algorithm [21, 100] to precisely estimate the expectation value of a single but highly general observable. Both algorithms (under some modification from the original proposals) utilize quantum coherence or entanglement for achieving the scaling of the so-called Heisenberg limit (HL) $\mathcal{O}(1/\varepsilon)$ for the root MSE in terms of query complexity to proper oracles: the Hamiltonian evolution for QPE and the state preparation unitary and its inverse for QAE. On the other hand, shadow tomography [83, 98] uses another mechanism of quantum enhancement (i.e., joint measurements) yielding the exponential reduction in query complexity to state preparation (or the number of state copies) with respect to the Pauli observable count M , at the cost of a poor precision scaling almost quadratically worse than SQL. These results lead to our first question in this thesis: in expectation value estimation of multiple observables, can one attain the HL scaling in query complexity while also accelerating with respect to the number M of arbitrary observables with a reasonable access model? Chapter 4 provides an affirmative answer to this question by explicitly constructing a new adaptive quantum algorithm that integrates ideas from adaptive QPE [77, 97, 48, 47], quantum amplification [114, 57], and quantum gradient estimation (QGE) algorithm [91, 56, 159] to realize a double quantum enhancement in the root MSE ε and M . Both our algorithm and QAE require access to the inverse of the state-preparation unitary, such as the time-backward evolution e^{+iHt} . This assumption is natural when quantum computers are used as computational probes, and its relevance to quantum experimental scenarios will be discussed in the conclusion.

Besides the ideal setting, there has been remarkable progress in real quantum devices. Indeed, so-called noisy intermediate-scale quantum (NISQ) computers are now available on various platforms: superconducting circuits, trapped ions, neutral Rydberg atoms, and others. This development has spurred research on NISQ-oriented quantum algorithms, particularly for variational quantum algorithms (VQAs) [27, 18]. However, potential obstacles have been identified that such NISQ algorithms, like VQAs, have severe scalability and/or trainability problems, e.g., so-called barren plateaus [121, 28, 168], suggesting that their quantum advantage is doubtful. Meanwhile, there has been rapid progress in experimental demonstrations of small-scale quantum error correction [49, 140, 20, 64, 63, 104], making us envision an early era of fault-tolerant quantum computing (FTQC) with large-scale yet not fully error-corrected quantum computers [148, 93]. These developments motivate a key paradigm in quantum algorithm design: retain an essential quantum mechanism for quantum advantage or speedup, while leveraging classical pre- and post-processing to minimize required quantum resources (e.g., qubit and gate counts).

Along these lines, developing quantum-enhanced algorithms for estimating expectation values that work well in the early FTQC era is highly important, as they have various applications in quantum computing tasks including the amplitude estimation. We naturally expect that available quantum resources are mostly consumed for a state preparation unitary, and the dominant source of noise is that costly unitary and its inverse in early FTQC. Thus, we model this situation such that the state

preparation unitary and its inverse are subjected to white noise (more precisely, a global depolarization noise) and no further demanding quantum operations (e.g., controlled Grover operation) are available. Within this model, we study a quantum-enhanced estimation algorithm for a single expectation value tailored for the early FTQC era in Chapter 5. Then, we encounter the following questions: what is the quantum-enhanced precision limit, that is, the minimal quantum Cramér-Rao bound over all estimation strategies in this model? Also, can we construct a quantum algorithm that is implementable and optimal in the precision limit? Crucially, we see that the standard strategy in noiseless cases is no longer optimal in the presence of noise. Despite the oracle assumption on the state preparation severely restricting available quantum operations, we can give a positive answer to these questions, namely, we develop a new quantum estimation algorithm with a carefully designed adaptive optimization that asymptotically achieves the quantum Cramér-Rao bound we derived, if target systems have a few tens of qubits.

The remainder of this thesis is organized as follows. Chapter 2 provides preliminaries to the results in this thesis. Chapter 3 introduces several fundamental quantum algorithms with a particular focus on quantum estimation, quantum state preparation, and the handling of block-encoded matrices. Chapter 4 is devoted to a Heisenberg-limited adaptive estimation algorithm for multiple observables that simultaneously achieves the HL query scaling and sublinear dependence on the observable count. Chapter 5 focuses on a quantum-enhanced single-observable estimation tailored to the early FTQC era, with an adaptive algorithm that asymptotically achieves the quantum Cramér-Rao bound under realistic noise models. Chapter 6 concludes this thesis.

Chapter 2

Preliminary

In this chapter, we provide the basics of quantum information, quantum computation, and quantum statistical estimation.

2.1 Quantum state, operation, and measurement

Let us consider a d -dimensional Hilbert space H . We write $L(H)$ as the set of all linear operators on H , which is also a finite-dimensional Hilbert space with the Hilbert-Schmidt inner product, i.e., $\langle A, B \rangle = \text{tr}[A^\dagger B]$ for any $A, B \in L(H)$. In this thesis, we often take $d = 2^n$ for a positive number n .

In the postulate of quantum mechanics [128], for any physical system, there is a Hilbert space known as the state space of the system. The system is completely described by a density matrix $\rho \in L(H)$ such that $\text{tr}[\rho] = 1$ and $\rho \geq 0$ holds. If the rank of a density matrix ρ is one (2 or more), the quantum state described by ρ is called pure (mixed). Any pure state $\rho = |\psi\rangle\langle\psi|$ can also be described as a unit vector $|\psi\rangle$ on the state space up to global phase. For a composite system consisting of subsystems A and B whose state spaces are H_A and H_B , respectively, the state space is given by the tensor product of H_A and H_B , i.e., $H_A \otimes H_B$. For instance, a single-qubit pure state is given by

$$\alpha |0\rangle + \beta |1\rangle \quad (2.1)$$

with the amplitude α, β satisfying $|\alpha|^2 + |\beta|^2 = 1$. In the n -qubit case, any pure state $|\psi\rangle$ can always be written as

$$|\psi\rangle = \sum_{j_0, \dots, j_{n-1}=0}^1 c_{j_0 j_1 \dots j_{n-1}} |j_0 j_1 \dots j_{n-1}\rangle \equiv \sum_{j=0}^{d-1} c_j |j\rangle, \quad (2.2)$$

where $|j_0 \dots j_{n-1}\rangle$ is the tensor product of $|j_k\rangle$ ($k = 0, 1, \dots, n-1$) and the complex values $c_{j_0 j_1 \dots j_{n-1}}$ satisfy the normalization

$$\sum_{j_0, j_1, \dots, j_{n-1}=0}^1 |c_{j_0 j_1 \dots j_{n-1}}|^2 = \sum_{j=0}^{d-1} |c_j|^2 = 1 \quad (2.3)$$

for the condition $\text{tr}[\rho] = 1$. We often shorten $|\psi\rangle$ as in the most right-hand side of Eq. (2.2). Each basis $|j\rangle$ (or $|j_0 j_1 \dots j_{n-1}\rangle$) is called the computational basis.

Transformation of quantum states is generally described by a linear map $\mathcal{E}$ between the sets of linear operators. In particular, the evolution of a closed system is described by a unitary operator $U \in L(H)$, or a unitary transform $\mathcal{E}(\bullet) = U \bullet U^\dagger$:

$$|\psi\rangle \mapsto U |\psi\rangle, \quad \text{or} \quad \rho \mapsto \mathcal{E}(\rho) = U \rho U^\dagger. \quad (2.4)$$

Another important transformation is discarding the subsystem of a composite system AB. This transformation is described by the partial trace operation:

$$\rho_{AB} \mapsto \rho_A = \text{tr}_B[\rho_{AB}]. \quad (2.5)$$

Including these transformations, we can define a more general class of transformations based on the principle that any physically relevant transformation should map a quantum state to another valid

quantum state. The most general physical transformation is described by a completely positive and trace-preserving (CPTP) map, which we also call a quantum channel. Here, a linear map $\mathcal{E}$ is said to be completely positive if $(\mathcal{E} \otimes \mathcal{I})(X) \geq 0$ for any positive semidefinite operator $X \geq 0$ on an extended system $\mathcal{H} \otimes \mathbb{C}^k$ for any integer k , where $\mathcal{I}$ denotes the identity map. Any CPTP map $\mathcal{E}$ can be written as in the Kraus form

$$\mathcal{E}(\bullet) = \sum_i K_i \bullet K_i^\dagger, \quad (2.6)$$

where K_i is a linear operator between input-output systems and satisfies $\sum_i K_i^\dagger K_i = \mathbf{1}$ for the identity operator $\mathbf{1}$. Also, any CPTP map can be realized by a unitary operation on an enlarged system followed by a partial trace. This is known as the Stinespring dilation theorem. Any quantum noise acting on the target system can also be described by CPTP maps. For instance, if 1-qubit Pauli operators X, Y, Z act on the target system with a uniform probability $p/3$ ($p \in [0, 1]$), the resulting transformation is the local depolarizing channel

$$\rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z). \quad (2.7)$$

Similarly, if n -qubit Pauli strings $P \in \{I, X, Y, Z\}^{\otimes n}$ occur uniform-randomly with the probability $p/4^n$, then the resulting global depolarizing channel is

$$\rho \mapsto (1-p)\rho + \frac{p}{4^n} \sum_P P\rho P^\dagger = (1-p)\rho + p \frac{I^{\otimes n}}{d}, \quad (2.8)$$

where the quantum state $I^{\otimes n}/d$ is called the maximally mixed state. The equality directly follows from the discussion about random unitary as mentioned later.

Quantum measurements on quantum states are described by a set of specific operators $\{M_m\}$ with the measurement outcome m . The operators act on the state space of the system being measured. When we measure the quantum state ρ , we obtain the measurement outcome m with the probability $p(m) = \text{tr}[M_m^\dagger M_m \rho]$. After the measurement, the quantum state ρ turns into

$$\rho \mapsto \frac{M_m \rho M_m^\dagger}{\text{tr}[M_m^\dagger M_m \rho]} \quad (2.9)$$

with the probability $p(m)$. Thus, the measurement operators $\{M_m\}$ should satisfy the completeness condition

$$\sum_m M_m^\dagger M_m = \mathbf{1}, \quad (2.10)$$

in order to ensure $\sum_m p(m) = 1$. When we do not need the post-measurement state, it is sufficient to determine the probability $p(m)$. In this case, we can use a positive operator-valued measure (POVM) $\{\Pi_m\}$ satisfying $\sum_m \Pi_m = \mathbf{1}$ and $\Pi_m \geq 0$, instead of $\{M_m\}$. Each POVM element Π_m is connected with the measurement operator M_m as $\Pi_m = M_m^\dagger M_m$, but such a decomposition of Π_m is not unique. The most important subclass of POVMs is projective measurements associated with observables, Hermitian operators on the state space being measured. Any observable O has a spectral decomposition

$$O = \sum_m m \Pi_m^O \quad (2.11)$$

for the projector Π_m^O onto the eigenspace of O with the eigenvalue m . These projectors form a projection-valued measure (a special subclass of POVMs), i.e., $\Pi_m^O \Pi_n^O = \delta_{mn} \Pi_m^O$ and $\sum_m \Pi_m^O = \mathbf{1}$. If we measure ρ with the observable O , we obtain m with probability $\text{tr}[\Pi_m^O \rho]$ as a measurement outcome. Thus, by repeating the measurement several times and taking the average of the outcomes, we can estimate the expectation value $\text{tr}[O\rho]$ of the observable O with respect to ρ . We further discuss the statistical aspects of quantum estimation (on parametric models) in Sec. 2.3.3.

Finally, we mention the notion of a random unitary. We first introduce the Haar random unitary. This random unitary is distributed according to the probability measure, called the Haar measure, defined on a unitary group. The Haar measure is invariant under left and right multiplication by any unitary in the group, which also makes the (normalized) Haar measure unique. Although the Haar random unitary is widely used in quantum information science, its exact implementation is extremely

difficult as the implementation cost (e.g., the number of elementary gates) scales polynomially with the dimension of the unitary operator, or exponentially with the number of qubits. This motivates us to define another ensemble of random unitaries approximating the Haar random unitary.

Definition 1 (Unitary t -design). For a probability measure μ on a unitary group and a positive integer t , we call μ a unitary t -design if the random unitary V distributed according to μ (we denote this as $V \sim \mu$) satisfies

$$\mathbb{E}_{V \sim \mu} [V^{\otimes t} A (V^\dagger)^{\otimes t}] = \mathbb{E}_{U \sim \text{Haar}} [U^{\otimes t} A (U^\dagger)^{\otimes t}]. \quad (2.12)$$

for any operator A .

From the definition, the random unitary $V \sim \mu$ recovers the t -th order moments of the Haar random unitary. If we know a random unitary V forms a unitary t -design, the action of the channel $\mathbb{E} [V^{\otimes t} \bullet (V^\dagger)^{\otimes t}]$ can be computed by the integration over the Haar measure. For instance, when $t = 1$, the d -dimensional Haar random unitary satisfies

$$\mathbb{E}_{U \sim \text{Haar}} [U A U^\dagger] = \frac{\text{tr}[A]}{d} \mathbf{1} \quad (2.13)$$

for any operator A . Because it is known that the uniform probability distribution over the n -qubit Pauli strings forms a unitary 1-design, Eq. (2.13) holds when we replace U with the uniformly-random Pauli strings.

2.2 Quantum circuit

A quantum circuit typically consists of (i) an initial state preparation, (ii) a sequence of unitary operations (also called quantum gates in this context), and (iii) the computational basis measurement. Here, the n -qubit computational basis measurement is given by the POVM $\{|j\rangle\langle j|\}_{j=0}^{d-1}$ for $d = 2^n$. In addition, one may use the mid-circuit measurement to perform feed-forward operations depending on the mid-circuit measurement outcomes and to simulate completely positive and trace-non-increasing (CPTN) maps [50]. After the mid-circuit measurement, by using a qubit reset operation, we may reuse the qubit for the subsequent computation.

To construct quantum circuits, one can use a set of elementary gates and their specific combinations. A standard choice is a universal gate set consisting of the Hadamard gate, T gate, and the controlled-NOT (CNOT) gate; any unitary can be approximated within arbitrary error by a sequence of gates in the universal set. Here, the Hadamard gate Had, the T gate, and the CNOT gate are defined as

$$\text{Had} := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad T := \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}, \quad \text{CNOT} := \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & X \end{pmatrix} \quad (2.14)$$

in the computational basis. We here provide a few examples. The reflection gate R_ψ about a pure state $|\psi\rangle$ is defined as

$$R_\psi := 2|\psi\rangle\langle\psi| - \mathbf{1}, \quad (2.15)$$

which is a crucial component of the Grover operator; see Sec. 3.1.2. The quantum Fourier transform QFT_n on n qubits is defined as

$$\text{QFT}_n |j\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2\pi i j k / 2^n} |k\rangle, \quad (2.16)$$

which can be decomposed into $\mathcal{O}(n^2)$ elementary gates and takes a crucial role in quantum estimation algorithms as we will see later.

2.3 Statistical estimation theory

2.3.1 Classical Cramér-Rao inequality

First, we focus on the problem of estimating an unknown parameter in a family of classical probability distributions. Let $\mathbf{X}$ be a multi-dimensional (discrete) random variable following a joint probability

density $\mathcal{L}(\mathbf{x}; \theta)$, where θ is an unknown parameter in a subset $\Theta \subset \mathbb{R}$. Our goal is to estimate the true parameter θ or generally a function of θ , say $g(\theta)$ with $g(\cdot)$ a bijective function, by some processing for samples from $\mathbf{X}$. To describe the processing, we define an estimator for estimating the target parameter θ as a function $\hat{\theta}(\mathbf{X})$ from samples to Θ . In addition, the value $\hat{\theta}(\mathbf{x})$ with a realized value $\mathbf{x}$ of the random variable $\mathbf{X}$ is called an estimate.

A natural measure to assess the performance of an estimator is the bias defined as

$$\mathbb{E} [\hat{\theta}(\mathbf{X})] - \theta.$$

If an estimator $\hat{\theta}(\mathbf{X})$ for $g(\theta)$ has no bias, that is, $\mathbb{E} [\hat{\theta}(\mathbf{X})] = g(\theta)$, the estimator is said to be an unbiased estimator for $g(\theta)$. However, an unbiased estimator may yield an estimate far from the true value, although the difference is canceled out in its expectation. Therefore, it is important to quantify the uncertainty of an estimator around its true value. To this end, the mean-squared-error (MSE) of an estimator is useful, which is defined as

$$\text{MSE} [\hat{\theta}(\mathbf{X})] := \mathbb{E} [(\hat{\theta}(\mathbf{X}) - g(\theta))^2]. \quad (2.17)$$

Note that if an estimator is unbiased, then its MSE is equivalent to its variance. Naturally, an unbiased estimator with a smaller MSE is better. The following inequality provides a lower bound of MSE for any unbiased estimator (in some conditions [184, 39, 144]):

$$\text{MSE} [\hat{\theta}(\mathbf{X})] \geq \left(\frac{\partial g(\theta)}{\partial \theta} \right)^2 \frac{1}{\mathcal{I}_{\text{c,tot}}(\theta)}. \quad (2.18)$$

This is called the classical Cramér-Rao inequality. Also, the quantity $\mathcal{I}_{\text{c,tot}}(\theta)$ is the (total) classical Fisher information defined as

$$\mathcal{I}_{\text{c,tot}}(\theta) := \mathbb{E}_{\mathbf{X}} \left[\left\{ \frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta; \mathbf{X}) \right\}^2 \right]. \quad (2.19)$$

We here rewrite the $\mathcal{L}(\mathbf{X}; \theta)$ as $\mathcal{L}(\theta; \mathbf{X})$ to emphasize that $\mathcal{L}$ is now a function of θ for a given sample $\mathbf{X}$. For simplicity, we write $\partial/\partial\theta$ as ∂_θ in what follows. By the Cramér-Rao inequality, the (classical) Fisher information can be considered as a quantity that determines the limit of estimation precision. Because the Fisher information is additive with respect to additional independent samples, we can also consider this quantity per a single sample as information on the target parameter obtained from the sample.

2.3.2 Maximum likelihood estimation

We introduce estimators based on the maximization of a *likelihood function*, which is the most popular method in statistical estimation. This method is called the maximum likelihood method. Let us consider a sample $\mathbf{x}$ of $\mathbf{X}$ is observed. Then, we imagine that the observation of $\mathbf{x}$ is attributed to the high probability at $\mathbf{x}$ in the underlying distribution. That is, the parameter $\hat{\theta}$ maximizing the probability at $\mathbf{x}$ is assumed to be a plausible estimate of the target value in the maximum likelihood method.

Definition 2 (The maximum likelihood estimator). For a sample $\mathbf{x}$ of $\mathbf{X}$, the probability density $\mathcal{L}(\theta; \mathbf{x})$ considered as a function of θ is called the likelihood function. Then, a maximum likelihood estimator $\hat{\theta}$ of θ is defined as

$$\hat{\theta} := \sup_{\theta \in \Theta} \mathcal{L}(\theta; \mathbf{X}).$$

By definition, the possible candidates for a maximum likelihood estimate are the value of $\theta \in \Theta$ satisfying the (log-)likelihood equation:

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta; \mathbf{X}) = 0 \quad \text{or} \quad \frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta; \mathbf{X}) = 0.$$

As for the estimation of $g(\theta)$, the corresponding maximum likelihood estimator is given by $g(\hat{\theta})$ if $\hat{\theta}$ is the maximum likelihood estimator of θ . This is called the invariance property of maximum likelihood estimators.

The theoretical justification of maximum likelihood estimators stems from their asymptotic properties under some regularity conditions.

Lemma 1 (Ref. [144] etc.). *Let $\mathbf{X} = (X_1, \dots, X_n)$ be a set of independent and identically distributed samples from a probability density $f(x; \theta)$ with a support independent from θ , where θ is in an open set $\Theta \subset \mathbb{R}$. If some regularity conditions in [144] hold, then there exists a sequence $\{\hat{\theta}_n\}_n$ of estimators of θ (based on $\mathbf{X}$) such that*

$$\Pr \left[\frac{\partial}{\partial \theta} \ln \mathcal{L}(\hat{\theta}_n; \mathbf{X}) = 0 \right] \rightarrow 1, \quad \text{and} \quad \hat{\theta}_n \xrightarrow{p} \theta, \quad (2.20)$$

where $\mathcal{L}(\mathbf{x}; \theta)$ is the joint probability density of $\mathbf{X}$. In addition, the sequence $\hat{\theta}_n$ is asymptotically efficient, that is,

$$\sqrt{I_n(\theta)} (\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, 1), \quad (2.21)$$

where $\mathcal{N}(0, 1)$ is the standard normal distribution, and $I_n(\theta)$ is the classical Fisher information associated with $\mathcal{L}(\theta; \mathbf{X})$. The $\xrightarrow{p}$ and $\xrightarrow{d}$ mean the convergence in probability and distribution, respectively.

Here, a sequence of estimators that converges to the true value like $\hat{\theta}_n \xrightarrow{p} \theta$ ($\forall \theta$) in the sense of probability is called consistent. The lemma 1 says that under some regularity conditions, there is a consistent sequence $\{\hat{\theta}_n\}_n$ each of which is a root of the likelihood equation. Strictly speaking, this lemma does not guarantee that the consistent root is the maximum likelihood estimator. However, if there is a unique root of the likelihood equation for every n that maximizes the likelihood, then this root is equivalent to the maximum likelihood estimator. Therefore, we can establish that the maximum likelihood estimator is consistent in such cases. On the other hand, if a sequence $\{\hat{\theta}_n\}$ of estimators for θ satisfies

$$\frac{1}{\sqrt{V_n(\theta)}} (\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, 1), \quad (2.22)$$

then $V_n(\theta)$ is called the asymptotic variance of $\hat{\theta}_n$. From Eq. (2.21), the asymptotic variance of maximum likelihood estimators matches the inverse of the Fisher information $I_n(\theta)$. That is, the asymptotic variance achieves the lower bound of the Cramér-Rao inequality. For this reason, such a sequence of estimators is said to be asymptotically efficient. Also, the sequence $\{\hat{\theta}_n\}_n$ of estimators in Eq. (2.21) follows the Gaussian (or normal) distribution in the asymptotic regime. Such a property of estimators is called asymptotic normality.

2.3.3 Quantum Cramér-Rao inequality

Now, we introduce a quantum extension of the classical estimation theory. Specifically, we focus on the estimation problem for a single parameter embedded in a quantum state $\rho(\theta)$. Note that in quantum estimation theory, various studies such as quantum channel estimation [88, 180] and quantum multi-parameter estimation [151, 43, 110] have been conducted beyond the quantum state estimation with a single parameter. Fixing a POVM $\{M_k\}$ for measuring the quantum state $\rho(\theta)$, we can obtain the following probability function $p(k; \theta)$ parameterized by θ :

$$p(k; \theta) := \text{tr}[M_k \rho(\theta)]. \quad (2.23)$$

Thus, we have a study in Eq. (2.18). However, unlike the classical case, there is freedom for designing a POVM in the quantum case. Utilizing this freedom, the lower bound of the classical Cramér-Rao inequality Eq. (2.18) can in fact be improved; that is, the following quantum Cramér-Rao inequality holds (for simplicity, $g(\theta) = \theta$) [76, 75, 73, 23, 133]:

$$\text{MSE}[\hat{e}(\mathbf{X})] \geq \frac{1}{\mathcal{I}_c(\theta)} \geq \frac{1}{\mathcal{I}_q(\theta)}, \quad (2.24)$$

where $\mathcal{I}_c$ is the classical Fisher information associated with the probability function $p(k; \theta)$ for a POVM $\{M_k\}$ and the target state $\rho(\theta)$. Here, $\mathcal{I}_q$ denotes the quantum Fisher information defined by only the quantum state $\rho(\theta)$ as follows:

$$\mathcal{I}_q(\theta) := \text{tr}[L_{\text{SLD}}^2 \rho(\theta)], \quad (2.25)$$

where L_{SLD} is the symmetric logarithmic derivative (SLD), which is defined as the Hermitian operator satisfying

$$\frac{\partial}{\partial \theta} \rho(\theta) = \frac{1}{2} (L_{\text{SLD}} \rho(\theta) + \rho(\theta) L_{\text{SLD}}). \quad (2.26)$$

Note that other definitions of quantum Fisher information can be used such as the *right symmetric derivative (RLD)* quantum Fisher information [79, 73]. However, it is known that the SLD gives the tightest bound (among the *logarithmic derivatives*) for the quantum state estimation problem with a single parameter [79, 72]. The second inequality Eq. (2.24) holds for any POVM $\{M_k\}$ independent of θ . Therefore, the quantum Fisher information characterizes the fundamental limit of estimation precision that cannot be improved by any measurement.

The calculation of the quantum Fisher information of a given parameterized quantum state is complicated, but some strategies are established [89, 176, 141]. For instance, the SLD for a quantum state in the form of

$$p\sigma(\theta) + \frac{1-p}{d}I_{d \times d}, \quad \sigma(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)|,$$

(where $p \in (0, 1)$ and $I_{d \times d}$ is the d -dimensional identity matrix) is given by

$$L_{\text{SLD}} \left(p\sigma(\theta) + \frac{1-p}{d}I_{d \times d} \right) = 2 \tanh \left(\frac{\alpha}{2} \right) \frac{\partial}{\partial \theta} \sigma(\theta).$$

Here, α satisfies $e^\alpha = (1 + (d-1)p)/(1-p)$. Therefore, the corresponding quantum Fisher information is calculated as

$$\mathcal{I}_q(\theta) = \left(\frac{2p}{\frac{2}{d} + (1 - \frac{2}{d})p} \right)^2 \text{tr} \left[\left(p\sigma(\theta) + \frac{1-p}{d}I_{d \times d} \right) \left\{ \frac{\partial}{\partial \theta} \sigma(\theta) \right\}^2 \right].$$

2.3.4 Heisenberg limit (HL) and standard quantum limit (SQL)

We here describe a quantum-enhanced estimation via a generalized phase estimation problem [59, 60]. Let us consider the estimation of θ encoded in a unitary operator $e^{i\theta H}$ with any known generator H . This essentially captures a broad class of metrological tasks. The total number of applications of $e^{i\theta H}$ is the resource metric we employed here. We focus on two eigenstates $|\lambda_0\rangle$ and $|\lambda_1\rangle$ of H with the eigenvalue λ_0 and λ_1 ($\lambda_0 < \lambda_1$), respectively. If we apply the unitary $e^{i\theta H}$ α times to the equal superposition of the two eigenstates, the resulting state becomes

$$|\psi_\alpha(\theta)\rangle = \frac{e^{i\alpha\theta\lambda_0}|\lambda_0\rangle + e^{i\alpha\theta\lambda_1}|\lambda_1\rangle}{\sqrt{2}} = e^{i\alpha\theta\lambda_0} \frac{|\lambda_0\rangle + e^{i\alpha\theta(\lambda_1-\lambda_0)}|\lambda_1\rangle}{\sqrt{2}}. \quad (2.27)$$

The quantum Fisher information $\mathcal{I}_q[|\psi_\alpha(\theta)\rangle]$ of the state $|\psi_\alpha(\theta)\rangle$ is calculated as

$$\mathcal{I}_q[|\psi_\alpha(\theta)\rangle] = 4 \left(\langle \partial_\theta \psi(\theta) | \partial_\theta \psi(\theta) \rangle - |\langle \psi(\theta) | \partial_\theta \psi(\theta) \rangle|^2 \right) = \alpha^2 (\lambda_1 - \lambda_0)^2. \quad (2.28)$$

By taking the maximal (minimal) eigenvalue as λ_1 (λ_0), the quantum Fisher information is maximized. Thus, the quantum Cramér-Rao inequality says for any unbiased estimator $\hat{\theta}$

$$\sqrt{\text{MSE}[\hat{\theta}]} \geq \frac{1}{\sqrt{k}\alpha(\lambda_1 - \lambda_0)} \quad (2.29)$$

for the total number k of separable preparation of $|\psi_\alpha(\theta)\rangle$ in the experiment. Here, we used the additivity of quantum Fisher information for separable states. The total number of $e^{i\theta H}$ in the whole estimation procedure is given by $N = k\alpha$, as a single experiment has α queries in a coherent and sequential way.

The lower bound implies several estimation limits. First, we consider a fully classical case with $\alpha = 1$. Here, the *classical* means that target quantum systems (probes) are separable and have a single query ($\alpha = 1$) to $e^{i\theta H}$. In this case, Eq. (2.29) indicates that the root MSE decreases as $\mathcal{O}(1/\sqrt{N})$:

$$\sqrt{\text{MSE}[\hat{\theta}]} \geq \frac{1}{\sqrt{N}(\lambda_1 - \lambda_0)}, \quad (2.30)$$

which represents the standard quantum limit (SQL). On the other hand, for fixed k , Eq. (2.29) shows the quantum-enhanced scaling $\mathcal{O}(1/N)$ due to the quantum coherence (more precisely, the sequential applications of the sensing operation $e^{i\theta H}$). This represents the ultimate limit on the estimation precision within standard setting in quantum metrology, under an appropriate definition of the resource metric [59, 183]; this universal limit is called the Heisenberg limit (HL). More precisely, in the current setup, it has been proved that the lower bound of Eq. (2.29) is not tight and we actually need the correction by a constant factor π [66]. We here show an example of HL based on quantum coherence, but there are several examples of HL by using other quantum effects, e.g., entanglement-enhanced sensing with GHZ states and N00N states [41, 22]. Moreover, even when quantum Markovian noise exists, a quantum error correction strategy enables us to attain the HL scaling iff the generator and noises satisfy a specific condition [181].

Chapter 3

Fundamental quantum algorithms

In this chapter, we describe several fundamental quantum algorithms including quantum estimation and quantum state preparation. Also, we provide basic results on an efficient calculation of block-encoded matrices on a quantum computer. These algorithms form the basis for the subsequent chapters.

3.1 Quantum estimation algorithms

3.1.1 Quantum phase estimation (QPE)

Quantum phase estimation (QPE) algorithm [99, 128] is the most famous quantum estimation algorithm, and it can estimate an eigenphase of a unitary operator when we have access to a (controlled) unitary and its eigenstate. Let U be a target unitary operator, and let $|\phi\rangle$ is the eigenstate of U such that $U|\phi\rangle = e^{i\phi}|\phi\rangle$ holds for an unknown parameter $\phi \in [0, 2\pi)$. The textbook style QPE algorithm [128] uses the quantum circuit with $\chi = \text{Had}^{\otimes m}$ (i.e., the tensor product of Hadamard gates) in Fig. 3.1. Here, m is the number of ancilla qubits. The quantum state before the measurement is given by

$$\sum_{k=0}^{2^m-1} a_k e^{ik\phi} \text{QFT}_m^\dagger |k\rangle \otimes |\phi\rangle = \frac{1}{\sqrt{2^m}} \sum_{k=0}^{2^m-1} a_k e^{ik\phi} \sum_{l=0}^{2^m-1} e^{-i2\pi \frac{kl}{2^m}} |l\rangle \otimes |\phi\rangle, \quad (3.1)$$

where $\sum_{k=0}^{2^m-1} a_k |k\rangle$ is the initial ancilla state prepared by the unitary χ . Then, we measure the ancilla qubits by the computational basis and obtain the measurement outcome $l \in \{0, 1, \dots, 2^m - 1\}$ with the following probability

$$\text{Pr}_{m,\phi}(l) = \left| \frac{1}{\sqrt{2^m}} \sum_{k=0}^{2^m-1} a_k e^{2\pi i k (\phi/2\pi - l/2^m)} \right|^2. \quad (3.2)$$

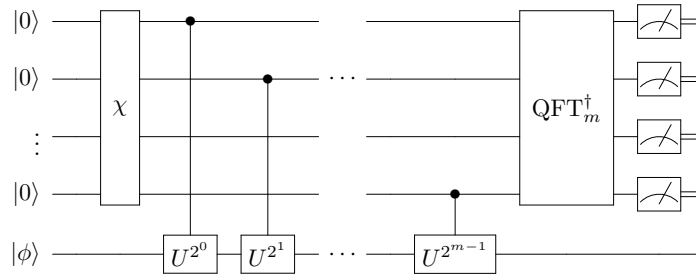


Figure 3.1: Quantum circuit for the QPE algorithm. The unitary χ prepares the initial state of m -qubit ancilla registers.

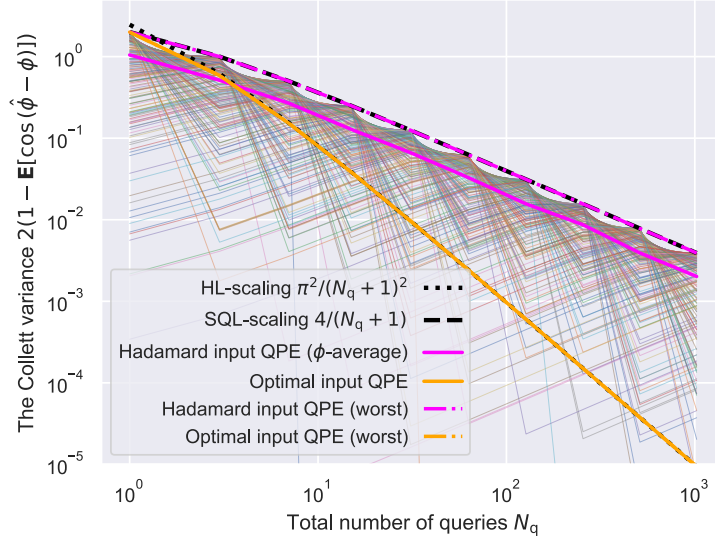


Figure 3.2: Numerical experiment for the QPE algorithm. We assume the preparation of an exact eigenstate.

By defining a phase estimator $\hat{\phi} := 2\pi\hat{l}/2^m$ ($\hat{l}$ is the random variable realizing l with the probability $\text{Pr}_{m,\phi}(l)$), we can estimate the target parameter ϕ from the measurement outcome in Fig. 3.1. The standard analysis of the estimator $\hat{\phi}$ is based on a confidence interval, but hereafter we focus on the MSE of the estimator $\hat{\phi}$. More specifically, taking the periodicity into account, we here consider the following uncertainty metric for phase estimators $\hat{\phi}$, called the Collett phase variance [14],

$$2 \left(1 - \mathbb{E} \left[\cos \left(\hat{\phi} - \phi \right) \right] \right) \approx \mathbb{E} \left[\left(\hat{\phi} - \phi \right)^2 \right] = \text{MSE}_{\phi} \left[\hat{\phi} \right]. \quad (3.3)$$

Here, the approximation holds if the distribution of $\hat{\phi} - \phi$ is sufficiently narrow around 0. The total number of applications of controlled U is the resource metric here.

We expect the textbook-style QPE to achieve the HL in query complexity for U as the circuit structure queries U sequentially. However, surprisingly, it is known that this QPE fails to attain the HL, and moreover, it end up realizing the SQL scaling [78, 160, 92]. The reason for this poor scaling is that the probability distribution $\text{Pr}_{m,\phi}(l)$ has a heavy tail; if we use an appropriate entangled state, this heavy tail is suppressed and the HL can be achieved. Such an entangled state is given by [88]

$$\sqrt{\frac{1}{2^{m-1}}} \sum_{k=0}^{2^m-1} \sin \left(\frac{k\pi}{2^m} \right) |k\rangle. \quad (3.4)$$

These results are summarized as follows. We also numerically confirm the HL and SQL in Fig. 3.2.

Theorem 1 (Heisenberg-limited QPE [88]). *Let U be a unitary operator with an eigenstate $|\phi\rangle$ satisfying*

$$U |\phi\rangle = e^{i\phi} |\phi\rangle \quad (3.5)$$

for an unknown phase $\phi \in [0, 2\pi)$. Suppose we have access to the controlled U and a single copy of $|\phi\rangle$. For any positive integer m , the random output $\hat{l} \in \{0, 1, \dots, 2^m - 1\}$, from the circuit Fig. (3.1) with the unitary χ for Eq. (3.4), satisfies

$$2 \left(1 - \mathbb{E} \left[\cos \left(\frac{2\pi}{2^m} \hat{l} - \phi \right) \right] \right) = 4 \sin^2 \left(\frac{\pi}{2N_q + 2} \right) \approx \left(\frac{\pi}{N_q + 1} \right)^2 \quad (3.6)$$

where $N_q := 2^m - 1$ is the total number of U in the circuit. For the case of the textbook QPE with $\chi = \text{Had}^{\otimes m}$, the random output $\hat{l}$ satisfies

$$2 \left(1 - \mathbb{E} \left[\cos \left(\frac{2\pi}{2^m} \hat{l} - \phi \right) \right] \right) = \frac{2}{N_q + 1} (1 - \cos[(N_q + 1)\phi]) = \mathcal{O}(1/N_q). \quad (3.7)$$

Although the Ref. [88] proves this theorem in a special phase estimation problem on a (control-free) unitary $\text{diag}(1, e^{i\phi})$, this is essentially the same as the problem addressed here as we assume the noiseless eigenstate preparation. Also, a further historical review of the QPE algorithm is provided in Sec. 4.1.1.

3.1.2 Quantum amplitude estimation (QAE)

Quantum amplitude estimation (QAE) algorithm [21] is essential for many applications such as chemistry, finance, and machine learning. The QAE algorithm estimates the amplitude $a \in [0, 1]$ of the following quantum state prepared by a black-box unitary operator $\mathcal{A}$:

$$\mathcal{A}|\mathbf{0}\rangle = \sqrt{a}|\Phi_1\rangle + \sqrt{1-a}|\Phi_0\rangle, \quad (3.8)$$

where $|\Phi_i\rangle$ ($i = 0, 1$) are generally unknown orthonormal quantum states and $|\mathbf{0}\rangle$ is an initial state. By defining $a = \sin^2 \theta$ ($\theta \in [0, \pi/2]$), we can rewrite the target state into

$$\mathcal{A}|\mathbf{0}\rangle = \sin \theta |\Phi_1\rangle + \cos \theta |\Phi_0\rangle. \quad (3.9)$$

Now, we focus on the 1-qubit space spanned by $|\Phi_i\rangle$ ($i = 0, 1$) and define Pauli operators $\bar{I}, \bar{X}, \bar{Y}$, and $\bar{Z}$ in this basis. For simplicity, let Π_G be the projection operator for the 1-qubit space, i.e., $\Pi_G = |\Phi_0\rangle\langle\Phi_0| + |\Phi_1\rangle\langle\Phi_1|$. We observe that

$$\mathcal{A}R_0\mathcal{A}^\dagger\Pi_G = \cos(2\theta)\bar{Z} + \sin(2\theta)\bar{X}, \quad (3.10)$$

where we recall $R_0 := 2|\mathbf{0}\rangle\langle\mathbf{0}| - \mathbf{1}$ in Eq. (2.15). Thus, this allows us to define the rotation gate G , called the Grover operator,

$$G = \mathcal{A}R_0\mathcal{A}^\dagger V_{\bar{Z}}, \quad G\Pi_G = \mathcal{A}R_0\mathcal{A}^\dagger V_{\bar{Z}}\Pi_G = \cos(2\theta)\bar{I} - i\sin(2\theta)\bar{Y} = e^{-i2\theta\bar{Y}}, \quad (3.11)$$

where $V_{\bar{Z}}$ is any known unitary operator such that

$$V_{\bar{Z}}\Pi_G = \bar{Z} \quad (3.12)$$

holds. Here, the eigenstates and the associated eigenvalues of the Grover operator G are given by

$$|\Phi_\pm\rangle = \frac{|\Phi_0\rangle \pm i|\Phi_1\rangle}{\sqrt{2}}, \quad G|\Phi_\pm\rangle = e^{\mp i2\theta}|\Phi_\pm\rangle. \quad (3.13)$$

Before proceeding to the details of QAE, we here describe the amplitude amplification, a building block for efficient quantum algorithm design. The quantum state $\mathcal{A}|\mathbf{0}\rangle$ is the equal-weight superposition of the eigenstates:

$$\mathcal{A}|\mathbf{0}\rangle = \frac{e^{-i\theta}|\Phi_+\rangle + e^{i\theta}|\Phi_-\rangle}{\sqrt{2}}. \quad (3.14)$$

Thus, the t applications of the Grover operator yield

$$G^t\mathcal{A}|\mathbf{0}\rangle = \frac{e^{-i(2t+1)\theta}|\Phi_+\rangle + e^{i(2t+1)\theta}|\Phi_-\rangle}{\sqrt{2}} = \sin[(2t+1)\theta]|\Phi_1\rangle + \cos[(2t+1)\theta]|\Phi_0\rangle. \quad (3.15)$$

Thus, by choosing the number t such that $\sin^2[(2t+1)\theta] \approx 1$, we can obtain the state $|\Phi_1\rangle$ with a high probability. When θ is small, it is sufficient to take $t \approx 1/\theta \approx 1/\sqrt{a}$, leading to $\sin^2[(2t+1)\theta] \approx 1$. By contrast, if we repeatedly measure $\mathcal{A}|\mathbf{0}\rangle$ until we obtain $|\Phi_1\rangle$, the total number of querying $\mathcal{A}$ (i.e., the total number copies of $\mathcal{A}|\mathbf{0}\rangle$ in this case) is approximately given by $1/a$. This means that the total number of querying $\mathcal{A}$ is quadratically improved when we can sequentially use $\mathcal{A}$ and its inverse.

The target amplitude a is encoded in the eigenphases of the Grover operator, and thus we can estimate it by the QPE algorithm if we can prepare the corresponding eigenstate. Since $\mathcal{A}|\mathbf{0}\rangle$ is the

equal superposition of the eigenstates, we can use it for the input state to QPE, while the superposition makes the performance analysis slightly more complicated. The resulting circuit for the standard QAE algorithm is given by the same one for QPE in Fig. 3.1 with the substitution $U \rightarrow G$ and $|\phi\rangle \rightarrow \mathcal{A}|\mathbf{0}\rangle$. However, if we slightly modify the circuit, we can halve the total number of queries to $\mathcal{A}$ and its inverse. The key observation is

$$\begin{aligned} (|0\rangle\langle 0| \otimes G + |1\rangle\langle 1| \otimes G^\dagger) (\mathbf{1} \otimes |\Phi_\pm\rangle\langle \Phi_\pm|) &= \begin{pmatrix} e^{\mp i2\theta} & 0 \\ 0 & e^{\pm i2\theta} \end{pmatrix} \otimes |\Phi_\pm\rangle\langle \Phi_\pm| \\ &= e^{\mp i2\theta} \begin{pmatrix} 1 & 0 \\ 0 & e^{\pm i4\theta} \end{pmatrix} \otimes |\Phi_\pm\rangle\langle \Phi_\pm|. \end{aligned} \quad (3.16)$$

This controlled unitary $|0\rangle\langle 0| \otimes G + |1\rangle\langle 1| \otimes G^\dagger$ is constructed by two uses of $\mathcal{A}$ and its inverse. By repeating this j times, which leads to $2j$ uses of $\mathcal{A}$ and its inverse, the resulting action is essentially the same as

$$(|0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes G^{2j}) (\mathbf{1} \otimes |\Phi_\pm\rangle\langle \Phi_\pm|) = \begin{pmatrix} 1 & 0 \\ 0 & e^{\mp i4j\theta} \end{pmatrix} \otimes |\Phi_\pm\rangle\langle \Phi_\pm|, \quad (3.17)$$

which requires $4j$ uses of $\mathcal{A}$ and its inverse. Thus, we can replace the controlled- G^{2j} in the QPE circuit (now, we set $U = G$) with the j uses of $|0\rangle\langle 0| \otimes G + |1\rangle\langle 1| \otimes G^\dagger$. This leads to the following theorem.

Theorem 2 (Heisenberg-limited QAE [103]). *Let $\mathcal{A}$ be a unitary operator preparing the state*

$$\mathcal{A}|\mathbf{0}\rangle = \sqrt{a}|\Phi_1\rangle + \sqrt{1-a}|\Phi_0\rangle$$

for an unknown amplitude $a \in [0, 1]$, unknown orthonormal quantum states $|\Phi_i\rangle$ ($i = 0, 1$), and an initial state $|\mathbf{0}\rangle$. Suppose we have access to the controlled $V_{\bar{Z}}$ operator defined in Eq. (3.12). Then, the random output $\hat{l} \in [0, 1, \dots, 2^m - 1]$, from the circuit Fig. (3.1) with the unitary χ in Eq. (3.4) and the modification for Eq. (3.17), satisfies

$$\max_{a \in [0, 1]} \mathbb{E} \left[\left(\sin^2 \left(\frac{\hat{l}\pi}{2^m} \right) - a \right)^2 \right] = \left(\frac{\pi}{2^{m+1}} \right)^2 + \mathcal{O} \left(\frac{1}{2^{3m}} \right). \quad (3.18)$$

Also, if we use $\chi = \text{Had}^{\otimes m}$, the random output $\hat{l}$ satisfies

$$\max_{a \in [0, 1]} \mathbb{E} \left[\left(\sin^2 \left(\frac{\hat{l}\pi}{2^m} \right) - a \right)^2 \right] = \frac{1}{2^{m+1}}. \quad (3.19)$$

3.1.3 Quantum gradient estimation (QGE)

For a given blackbox of a real scalar function $f(\mathbf{x})$ on $\mathbf{x} \in \mathbb{R}^M$, the quantum algorithm introduced by Stephen P. Jordan [91] can efficiently estimate the M -dimensional gradient of $\nabla f(\mathbf{0})$, with use of less queries to the blackbox compared to classical case. We here review this algorithm based on its analysis by Ref. [56]. Here, the target point $\mathbf{x} = \mathbf{0}$ can be taken as $\mathbf{x} \neq \mathbf{0}$ by trivially redefining $f(\mathbf{x})$. The quantum algorithm consists of three steps: (i) prepare a superposition state of grid points $\mathbf{x}$ around the target point $\mathbf{x} = \mathbf{0}$, (ii) apply the blackbox of the target function f to the state and evaluate a phase $e^{if(\mathbf{x})}$ at each grid point, and (iii) measure the resulting state by the computational basis labeled by the grid points after the inverse quantum Fourier transformations.

To begin with, we define a set of grid points G_p^M around $\mathbf{x} = \mathbf{0}$ to evaluate f , as follows:

$$G_p^M := \left\{ \frac{\mu}{2^p} - \frac{1}{2} + \frac{1}{2^{p+1}} : \mu \in \{0, 1, \dots, 2^p - 1\} \right\}^M, \quad (3.20)$$

where p denotes a positive integer which specifies the estimation precision later. Note that because there is a bijection map

$$\varphi : \mu \mapsto \varphi(\mu) := \frac{\mu}{2^p} - \frac{1}{2} + \frac{1}{2^{p+1}} \in G_p, \quad (3.21)$$

we always label the p -qubit computational basis $|\mu\rangle$ by the corresponding element $x \equiv \phi(\mu) \in G_p$. Then, applying the Hadamard gates to initialized pM -qubit registers, we have the superposition state of the grid points $\mathbf{x} := (x_1, \dots, x_M) \in G_p^M$:

$$\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} |\mathbf{x}\rangle = \frac{1}{\sqrt{2^{pM}}} \sum_{(x_1, \dots, x_M) \in G_p^M} |x_1\rangle |x_2\rangle \cdots |x_M\rangle. \quad (3.22)$$

Here, each $|x_j\rangle$ contains p -qubit registers. Next, we assume access to the *phase oracle* O_f for $f(\mathbf{x})$ defined as [56]

$$O_f : |\mathbf{x}\rangle \rightarrow e^{if(\mathbf{x})} |\mathbf{x}\rangle \quad \text{for all } \mathbf{x} \in G_p^M. \quad (3.23)$$

Note that Ref. [56] provides a generic analysis of Jordan's algorithm based on the phase (and probability) oracle, while the original paper [91] assumes another powerful oracle (i.e., η -accurate Binary oracle for f that outputs $f(\mathbf{x})$ binarily with accuracy η from an input $\mathbf{x}$).

Now, to clarify the basic idea of Jordan's algorithm, we consider a special case where the target function f is an affine linear function, i.e., for a target gradient vector $\mathbf{g} \in [-1/3, 1/3]^M$ and some constant $f_0 \in \mathbb{R}$, $f(\mathbf{x}) = \mathbf{g} \cdot \mathbf{x} + f_0$. For such linear functions, the application of the (modified) phase oracle $O_f^{2\pi 2^p}$ to the input state Eq. (3.22) yields

$$\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p f(\mathbf{x})} |\mathbf{x}\rangle = e^{2\pi i 2^p f_0} \cdot \bigotimes_{j=1}^M \left[\frac{1}{\sqrt{2^p}} \sum_{x_j \in G_p} e^{2\pi i 2^p x_j g_j} |x_j\rangle \right]. \quad (3.24)$$

Then, we apply a slightly modified version of the inverse quantum Fourier transformation $\text{QFT}_{G_p}^\dagger$ over the p -qubit system: for all $x \in G_p$,

$$\text{QFT}_{G_p} : |x\rangle \mapsto \frac{1}{\sqrt{2^p}} \sum_{k \in G_p} e^{2\pi i 2^p x k} |k\rangle. \quad (3.25)$$

Note that QFT_{G_p} is the same as the usual p -qubit QFT up to conjugation with a tensor product of p single-qubit gates [56]. The statistics of the computational basis measurement to Eq. (3.24) with $\text{QFT}_{G_p}^\dagger$ are similar to that of the output of the standard QPE algorithm [128]. More precisely, let $(k_1, k_2, \dots, k_M) \in G_p^M$ be a result of the computational basis measurement, then the following holds [56]:

$$\Pr \left[|k_j - g_j| > \frac{3}{2^p} \right] \leq \frac{1}{4} \quad \text{for every } i = 1, 2, \dots, M. \quad (3.26)$$

In the above description, we assumed that the target function is affine linear. Moreover, if a target function is very close to some affine linear function and thereby the equality of Eq. (3.24) approximately holds, we can prove similar results to Eq. (3.26), as follows.

Lemma 2 ([56, 159]). *Let $\mathbf{g} \in \mathbb{R}^M$ such that $\|\mathbf{g}\|_\infty \leq 1/3$. Suppose we can prepare the quantum state $|\Psi\rangle$ that is $1/12$ -close in the Euclidean distance to the following state*

$$\left(\text{QFT}_{G_p}^\dagger \right)^{\otimes M} \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \mathbf{g} \cdot \mathbf{x}} |\mathbf{x}\rangle. \quad (3.27)$$

Then, measuring the quantum state $|\Psi\rangle$ in the computational basis, we obtain a coordinate-wise estimate $(k_1, \dots, k_M) \in G_p^M$ satisfying

$$\Pr \left[|k_j - g_j| > \frac{3}{2^p} \right] \leq \frac{1}{3} \quad \text{for every } j = 1, 2, \dots, M.$$

Proof. By the closeness assumption, the trace distance between $|\Psi\rangle$ and Eq. (3.27) is upper bounded by $1/12$. Since the trace distance provides the upper bound of the total variation distance of two probability measures defined by the two quantum states and an arbitrary common POVM (Theorem 9.1 [128]), the probability (3.26) is modified at most $1/12$ by measuring $|\Psi\rangle$ instead of Eq. (3.27). $\square$

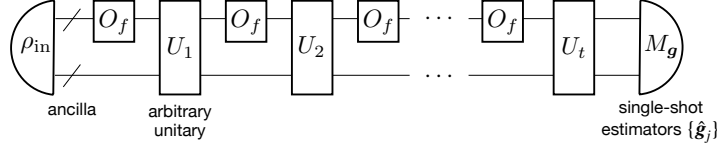


Figure 3.3: A general adaptive protocol for estimating $\mathbf{g}$ in O_f . Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

The above analysis shows that we can obtain an $\mathcal{O}(1/2^p)$ -precise estimate of the gradient g_j for all $j = 1, 2, \dots, M$, with a remarkable property that the implementation is independent from M in terms of query to an oracle for $f(\mathbf{x}) \approx \sum_j g_j x_j$. However, it must be noted that such an advantage is achieved under the condition that the approximation $f(\mathbf{x}) \approx \sum_j g_j x_j$ holds very well. In most cases, the higher order cannot be negligible and additional oracle queries are required to enhance the linearity. For certain smooth functions, the smallest overhead is proven to be $\mathcal{O}(\sqrt{M})$, and this can be achieved, up to poly-logarithmic factors, by the method using higher-degree central-difference formula [56]. For practical applications of this algorithm, it is crucial in terms of total quantum resources to efficiently prepare an oracle for each task. Namely, we want to design a sufficiently linear function $f(\mathbf{x}) \approx \sum_j g_j x_j$ whose gradient $\{g_j\}$ encodes the target quantities for each task and further prepare the corresponding oracle for $|\mathbf{x}\rangle \mapsto e^{if(\mathbf{x})} |\mathbf{x}\rangle$ or a final state close to Eq. (3.27).

Heisenberg limit

The phase oracle O_f for a smooth real function $f(\mathbf{x})$ on $\mathbf{x} \in \mathbb{R}^M$ with gradient $\mathbf{g} := \nabla f(\mathbf{0})$ corresponds to the time evolution operator (with unit time) generated by the following Hamiltonian

$$H(\mathbf{g}) + V, \quad H(\mathbf{g}) := \sum_{\mathbf{x} \in G_p^M} (\mathbf{g} \cdot \mathbf{x}) |\mathbf{x}\rangle \langle \mathbf{x}|, \quad (3.28)$$

where V is a $\mathbf{g}$ -independent Hermitian operator that commutes with $H(\mathbf{g})$.¹ Thus, the estimation of $\mathbf{g}$ from O_f with an appropriate choice of an input state and a final measurement can be considered as one of the (multi-parameter) unitary estimation problems. Such unitary estimation problems are widely studied in the field of quantum metrology. Here, we derive the HL in quantum gradient estimation.

To derive the fundamental bound, we here consider a general adaptive estimation protocol [59], which is illustrated in Fig. 3.3. In the general protocol, we first prepare an arbitrary input state ρ_{in} between the system on which O_f acts and an ancilla system with an arbitrary number of qubits. Then, we apply the following sequence of quantum operations to the input state:

$$U_t(O_f \otimes \mathbf{1}) \cdots U_2(O_f \otimes \mathbf{1}) U_1(O_f \otimes \mathbf{1}), \quad (3.29)$$

where $\mathbf{1}$ denotes the identity and U_i ($i = 1, 2, \dots, t$) are $\mathbf{g}$ -independent arbitrary unitary operators acting on the whole system. This sequence contains t uses of the phase oracle in total. Finally, the output state is measured by an arbitrary POVM $\{M_{\mathbf{g}}\}$, and we write the corresponding single-shot estimator for g_j as $\hat{g}_j$. Because the interaction U_i to the ancilla systems can extract the information during the estimation process, this protocol also includes adaptive techniques. Obviously, the original method for gradient estimation is captured in this general protocol, by setting ρ_{in} to the uniform superposition state without ancilla qubits, $U_i = \mathbf{1}$ for all i , and $M_{\mathbf{g}}$ to the computational basis measurement with the inverse quantum Fourier transformation. Note that the $\mathbf{g}$ -independent part of O_f can be included in each U_j , and therefore in the following, we consider $O_f = e^{iH(\mathbf{g})}$ without loss of generality.

We here remark that $e^{iH(\mathbf{g})}$ can be written as a tensor product of time evolutions

$$\bigotimes_{j=1}^M \left(\sum_{x_j \in G_p} e^{ig_j x_j} |x_j\rangle \langle x_j| \right) \quad (3.30)$$

¹The results in this part are from the author's work [166].

by the generators $H_j := \sum_{x_j} x_j |x_j\rangle \langle x_j|$ with the bounded spectral norm $\|H_j\| \leq 1/2$. Therefore, the estimation of $\mathbf{g} = (g_1, \dots, g_M)$ is essentially equal to the estimation of evolving time g_j in each $e^{ig_j H_j}$. Now, we are ready to prove the HL in quantum gradient estimation.

Theorem 3 (Heisenberg limit in quantum gradient estimation [166]). *Suppose we have access to the phase oracle O_f for a smooth real function f on $\mathbb{R}^M$ with gradient $\mathbf{g} := \nabla f(\mathbf{0})$ and g_j belongs to some finite interval $\Theta \subset \mathbb{R}$ for all j . Then, the single-shot estimators $\hat{g}_j$, obtained from the general adaptive estimation protocol in Fig. 3.3 with t uses of the phase oracle O_f , satisfy the following inequality:*

$$\min_{j=1,2,\dots,M} \widehat{\text{MSE}}[\hat{g}_j] \geq \frac{\pi^2}{t^2} \quad \text{as } t \rightarrow \infty, \quad (3.31)$$

where $\widehat{\text{MSE}}[\hat{g}_j]$ denotes the supremum of the MSE for $\hat{g}_j$ over the known interval Θ , i.e., the maximum value of MSE over all possible target values $g_j \in \Theta$.

Proof. We prove this theorem by contradiction. Suppose there is a protocol such that for an index $j \in \{1, \dots, M\}$, $\widehat{\text{MSE}}[\hat{g}_j] < \pi^2/(t^2 \Delta G_p^2)$ holds as t goes to ∞ , where ΔG_p denotes the difference between extreme eigenvalues of $H_j = \sum_{x_j \in G_p} x_j |x_j\rangle \langle x_j|$ i.e., $\max G_p - \min G_p \leq 1$. Now, we recall that the phase oracle O_f can be considered as the product of $e^{ig_j H_j}$ without loss of generality. Then, focusing on the estimation of g_j in the single-parameter unitary $e^{ig_j H_j}$, it can be shown that there is no protocol using the total t uses of $e^{ig_j H_j}$ such that $\widehat{\text{MSE}}[\hat{g}_j]$ can be less than $\pi^2/t^2 \Delta G_p^2$ as t increases [66, 65], which contradicts the above assumption. Therefore, we conclude that

$$\widehat{\text{MSE}}[\hat{g}_j] \geq \frac{\pi^2}{t^2 \Delta G_p^2} \geq \frac{\pi^2}{t^2} \quad \text{as } t \rightarrow \infty$$

holds for all j , which completes the proof of Theorem 3. $\square$

As for the achievability of the lower bound, the same methodology as the phase estimation with the minimum phase uncertainty [118, 160, 92, 10] can be applied to design an optimal protocol in a special case. Hereafter, we describe an optimal protocol of gradient estimation that is applicable to an affine linear function $f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{g}$ without any adaptive operations. Let us consider M tensor products of an ansatz state $\sum_{x \in G_p} a_x |x\rangle$, where a_x is a real amplitude, as an initial input state instead of the equal superposition state Eq. (3.22). Then, applying the phase oracle O_f 2^{p+1} times to the input state and performing the computational basis measurement after $\text{QFT}_{G_p}^\dagger$, we obtain the measurement result k_j with the probability

$$\Pr[k_j] = \left| \sum_{x_j=0}^{2^p-1} \frac{a_{x_j}}{\sqrt{2^p}} e^{2\pi i 2^p x_j (g'_j - k_j)} \right|^2, \quad (3.32)$$

where $g'_j := g_j/\pi$. For simplicity, we drop the subscript j in the following. Taking the measurement outcome k as an estimator of g' , we can approximately evaluate the MSE of k as follows:

$$\mathbb{E}[(k - g')^2] \simeq \frac{1}{2\pi^2} (1 - \mathbb{E}[\cos(2\pi(k - g'))]), \quad (3.33)$$

where we can check the right hand side is close to the MSE when $(k - g')$ is small by the Taylor expansion. As shown in Ref. [88], the expectation of cosine can be analytically calculated:

$$\mathbb{E}[\cos(2\pi(k - g'))] = \sum_{k=1}^{2^p-1} a_{k-1} a_k + a_0 a_{2^p-1} \cos \left[2\pi 2^p \left(g' + \frac{1}{2} - \frac{1}{2^{p+1}} \right) \right]. \quad (3.34)$$

Here, a_k is equivalent to $a_{\varphi(k)}$ for the bijection map defined in Eq. (3.21). Therefore, if we take $a_0 = 0$, then the right hand side of Eq. (3.33) can be written as a quadratic form $(1/2\pi^2) \sum_{k,l=1}^{2^p-1} A_{kl} a_k a_l$ for the symmetric matrix A :

$$A = \begin{pmatrix} 1 & -1/2 & & \\ -1/2 & 1 & -1/2 & \\ & -1/2 & 1 & \ddots \\ & & \ddots & \ddots \end{pmatrix}. \quad (3.35)$$

The minimum eigenvalue and the corresponding eigenvector of A is known as

$$2 \sin^2 \frac{\pi}{2^{p+1}} \quad (3.36)$$

and

$$a_k = \sqrt{\frac{2}{2^p}} \sin \frac{k\pi}{2^p}, \quad (3.37)$$

respectively. Consequently, combining Eqs. (3.33) and (3.36), we can confirm that this protocol achieves the lower bound of Theorem 3 when $t = 2^{p+1}$ is sufficiently large, as

$$\mathbb{E}[(\pi k - g)^2] \simeq \left(\frac{\pi}{2^{p+1}}\right)^2. \quad (3.38)$$

Finally, we remark that the original implementation of the gradient estimation cannot achieve the quadratic speedup regarding total queries t , i.e., $\varepsilon^2 = \mathcal{O}(1/t^2)$ for a MSE ε^2 , even if the target function is an affine linear. This can be checked from Eq. (3.33); the uniform superposition state $a_x = 1/\sqrt{2^p}$ for all x gives

$$\min_{g'} \mathbb{E}[\cos(2\pi(k - g'))] = 1 - \frac{2}{2^p}, \quad (3.39)$$

and thus, $\max_g \mathbb{E}[(\pi k - g)^2] = \mathcal{O}(1/t)$, which is the same scaling as the SQL. Note that this result is consistent with that the QPE algorithm fails to achieve the HL when the input ancilla qubits are initialized as the uniform superposition state, unlike the optimal entangled state as in Eq. (3.37) [78, 77, 14]. We remark that the proposed adaptive method in Chapter 4 performs gradient estimation at the same scaling as the HL $\mathcal{O}(1/\varepsilon)$ even when the target function is *approximately* affine linear.

3.2 Quantum state preparation algorithms

Another important class of quantum algorithms is quantum state preparation, which prepares a target quantum state, e.g., the ground state of Hamiltonian and a time-evolved state on quantum devices. State preparation also serves as the starting point of various industrial applications of quantum computing such as functional material designs.

3.2.1 Virtual methods

In usual, the goal of quantum state preparation is to directly prepare a target quantum state on a quantum device. However, in most cases, we aim for calculating properties (e.g., expectation values of observables) of a target state, and the direct state preparation is not always necessary. For this goal, we can effectively use a wide class of transformations of quantum states that expands the class of CPTP maps, while we cannot directly realize such a transformation on quantum devices; this is the reason why we call *virtual* methods [85, 135, 179]. In this subsection, we focus on the property estimation of a target state and mention several efficient methods with significant quantum resource reduction. We remark that such a virtual method appears only in this subsection of this thesis.

Sample-based Hamiltonian simulation

Simulating time-independent Hamiltonian evolution e^{-iHt} is one of the most important tasks in quantum computing. Among various access models of Hamiltonian, we here focus on the case where the target Hamiltonian H is encoded into an *unknown* quantum state ρ as

$$\rho = \frac{H + c\mathbf{1}}{\text{tr}[H + c\mathbf{1}]}, \quad (3.40)$$

where c is a real value chosen so that ρ is non-negative. This access model is studied in Refs. [96, 62]. Now, our goal is to estimate the expectation value of a given observable O

$$\text{tr}[O(\mathcal{U}[e^{-i\rho T}] \otimes \mathcal{I}_R)(\sigma)] \quad (3.41)$$

for a given initial state σ , by consuming the copies of ρ . Here, the ρ acts on the subsystem of the observable O and initial state σ , and $\mathcal{I}_R$ is the identity channel on the other system. We recall that due to the no-cloning theorem, we cannot duplicate the unknown ρ .

While a naive solution to this problem is to measure the multiple copies of ρ and then synthesize the unitary gate $e^{-i\rho T}$ from the measurement outcomes, this requires the $\text{poly}(d, 1/\varepsilon)$ number of state copies of ρ for the system dimension d and the accuracy ε [113, 96]. Then, Ref. [113] provides an elegant method, called the density matrix exponentiation (DME), to remove the d -dependence. Surprisingly, we show that this copy complexity is further exponentially improved in terms of $1/\varepsilon$ for the goal to estimate Eq. (3.41) (and more general properties). This result is summarized as follows.²

Theorem 4 (Virtual DME (for general states) [163]). *Let ρ be an unknown but accessible d -dimensional quantum state, and let $\mathcal{U}[e^{-i\rho T}]$ be the unitary channel of $e^{-i\rho T}$ for $T \in \mathbb{R} \setminus \{0\}$. For any integer $r \geq \max\{1, |T|\}$, we can construct a random modified-quantum map $\hat{\Phi}_r$ using at most N copies of ρ and a single-ancilla measurement such that*

$$\|C^{2r} \mathbb{E}[\hat{\Phi}_r] - \mathcal{U}[e^{-i\rho T}]\|_{\diamond} \leq \varepsilon \quad (3.42)$$

holds for a given $\varepsilon \in (0, 1/2)$, if N satisfies

$$N \geq 2r + 4r \left\lceil \frac{1}{2} \frac{\ln(3r/\varepsilon)}{W_0((1/e) \ln(3r/\varepsilon))} - \frac{1}{2} \right\rceil. \quad (3.43)$$

Here, W_0 is the principal branch of the Lambert- W function (that is, $W_0(c)$ is defined as the unique solution of $xe^x = c$ for $x \geq -1$ and $c \geq -1/e$), and C is a positive value that satisfies

$$1 < C^{2r} \leq e^{2T^2/r}. \quad (3.44)$$

Here, the random modified-quantum map $\hat{\Phi}_r$ always have the following structure

$$\hat{\Phi}_r(\bullet) = \text{tr}_{\text{anc}} \left[(X_{\text{anc}} \otimes \mathbf{1}) \tilde{\Phi}_{\mathbf{l}_r}(|+\rangle \langle +|_{\text{anc}} \otimes \bullet) \right], \quad (3.45)$$

where $\tilde{\Phi}_{\mathbf{l}_r}$ is a map from the length- r index $\mathbf{l}_r$ to a CPTP map with multiple copies of ρ . The index $\mathbf{l}_r$ is randomly chosen according to the probability distribution:

$$p_{\mathbf{l}}(T/r) \propto \frac{(T/r)^{2l}}{(2l)!} \sqrt{1 + \left(\frac{T/r}{2l+1} \right)^2}. \quad (3.46)$$

The Theorem 4 indicates that the target expectation value can be written as

$$\text{tr}[O(\mathcal{U}[e^{-i\rho T}] \otimes \mathcal{I}_R)(\sigma)] \approx^{\varepsilon} C^{2r} \text{tr}[O\sigma_{\text{eff}}], \quad (3.47)$$

where

$$\sigma_{\text{eff}} = \mathbb{E} \left[\left(\text{tr}_{\text{anc}}[(X_{\text{anc}} \otimes \mathbf{1}) \tilde{\Phi}_{\mathbf{l}_r}(|+\rangle \langle +|_{\text{anc}} \otimes \bullet)] \otimes \mathcal{I}_R \right) (\sigma) \right]. \quad (3.48)$$

Here, the value $\text{tr}[O\sigma_{\text{eff}}]$ from the effective state σ_{eff} (not a valid quantum state) can be measured by preparing the quantum state

$$\left(\tilde{\Phi}_{\mathbf{l}_r} \otimes \mathcal{I}_R \right) (|+\rangle \langle +|_{\text{anc}} \otimes \sigma) \quad (3.49)$$

with randomly chosen $\tilde{\Phi}_{\mathbf{l}_r}$ and then measuring the observable $X_{\text{anc}} \otimes O$.

Finally, we remark that the prefactor C^{2r} amplifies the variance of estimating $\text{tr}[O\sigma_{\text{eff}}]$ by the above procedure. The variance of estimating $\text{tr}[O\sigma_{\text{eff}}]$ itself is at most $\|O\|^2/N$ for N shots. Since we can take the parameter $r \sim 2T^2$ so that $C^{2r} \leq e$ holds from the Theorem 4, the amplification of variance is a small constant e or below. As a result, the number of measurements is $\mathcal{O}(1/\varepsilon^2)$ and the number of state copies per circuit is $\mathcal{O}(T^2 \log(T/\varepsilon)/\log \log(T/\varepsilon))$ in order to estimate the target expectation value Eq. (3.41) within the estimation error ε . We here focus on the case where a target circuit contains a single use of $\mathcal{U}[e^{-i\rho T}]$, but this can be generalized to multiple cases [163]. When we use $\mathcal{U}[e^{-i\rho T}]$ (with a fixed T without loss of generality) M times in a circuit, the amplification of the variance becomes C^{2rM} . We can make this factor C^{2rM} a small constant ($\approx e$ or below) by adjusting the parameter r .

²The result is from the author's work [163].

Lindblad simulation

Next, we move to the simulation of open quantum systems. Specifically, we focus on the Lindblad equation or Gorini–Kossakowski–Sudarshan–Lindblad equation [109, 68]:

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) := -i[H, \rho] + \sum_{k=1}^K \left(L_k \rho L_k^\dagger - \frac{1}{2} \{L_k^\dagger L_k, \rho\} \right), \quad (3.50)$$

where H and $\{L_k\}_{k=1}^K$ are the Hamiltonian and jump operators of a target system, respectively. Here, we employ a common access model to H and $\{L_k\}$, that is, we know that they are specified by n -qubit Pauli strings $\{P_{kj}\}$ as

$$H = \sum_{j=1}^m \alpha_{0j} P_{0j}, \quad L_k = \sum_{j=1}^M \alpha_{kj} P_{kj}, \quad (3.51)$$

for $\alpha_{0j} \in \mathbb{R}, \alpha_{kj} \in \mathbb{C}$. Similar to the previous Hamiltonian simulation, our goal is to estimate the expectation value of a given observable O

$$\text{tr}[Oe^{t\mathcal{L}}(\sigma)] \quad (3.52)$$

for an initial state σ and a simulation time t . A canonical FTQC algorithm [36] to prepare the evolved state $e^{t\mathcal{L}}(\sigma)$ requires the gate complexity $\tilde{O}(\tau \log^2(\tau/\varepsilon) K^2 \text{poly}(m, M))$ and the number of ancilla qubits $\mathcal{O}(\log(\tau/\varepsilon) \log(\tau K(m+M)/\varepsilon))$, where τ is a rescaled simulation time and ε is a simulation accuracy.

For the dynamical map $e^{t\mathcal{L}}$, we can derive a decomposition similar to Eq. (3.45) of virtual DME as follows.³

Theorem 5 (Virtual Lindblad simulation [94]). *Let $\mathcal{L}$ be an n -qubit Lindbladian with a Hamiltonian H and jump operators $\{L_k\}_{k=1}^K$ that are specified by Eq. (3.51), and let $\|\mathcal{L}\|_{\text{pauli}} := 2(\alpha_0 + \sum_{k=1}^K \alpha_k^2)$ for $\alpha_k := \sum_j |\alpha_{kj}|$. Then, for any $t > 0$, $\Delta \in (0, 1/e)$, and any positive integer $r \geq \|\mathcal{L}\|_{\text{pauli}} t$, there exists an approximate decomposition of $e^{t\mathcal{L}}$ such that*

$$\left\| e^{t\mathcal{L}}(\bullet) - \sum_{v \in S_\Delta} c_v \text{tr}_{\text{anc}}[(X_{\text{anc}} \otimes \mathbf{1}) \widetilde{\mathcal{W}}_v(|+\rangle \langle +|_{\text{anc}} \otimes \bullet)] \right\|_{1 \rightarrow 1} \leq \Delta \quad (3.53)$$

holds for some finite index set S_Δ , $(n+1)$ -qubit CPTN maps $\{\widetilde{\mathcal{W}}_v\}$, and real values $c_v > 0$ satisfying

$$\sum_{v \in S_\Delta} c_v \leq e^{2\|\mathcal{L}\|_{\text{pauli}}^2 t^2 / r}. \quad (3.54)$$

Furthermore, for any $v \in S_\Delta$, the $(n+1)$ -qubit CPTN map $\widetilde{\mathcal{W}}_v$ can be effectively simulated by using an $n+4 + \lceil \log_2 M \rceil$ qubits quantum circuit, including mid-circuit measurement and qubit reset, with at most

$$\mathcal{O}\left(r \left(\frac{\log(r/\Delta)}{\log \log(r/\Delta)} + M \log M \right)\right) \quad (3.55)$$

one- or two-qubit gates. The corresponding quantum circuits for the CPTN maps $\widetilde{\mathcal{W}}_v$ can be efficiently sampled according to the probability distribution proportional to $c_v > 0$.

The Theorem 5 indicates that

$$\text{tr}[Oe^{t\mathcal{L}}(\sigma)] \approx^\Delta \left(\sum_{v \in S_\Delta} c_v \right) \text{tr}[O\sigma_{\text{eff, lindblad}}(t)], \quad (3.56)$$

where

$$\sigma_{\text{eff, lindblad}}(t) = \text{tr}_{\text{anc}} \left[(X_{\text{anc}} \otimes \mathbf{1}) \sum_{v \in S_\Delta} \left(\frac{c_v}{\sum_{v \in S_\Delta} c_v} \widetilde{\mathcal{W}}_v \right) (|+\rangle \langle +|_{\text{anc}} \otimes \sigma) \right]. \quad (3.57)$$

³This result is from Ref. [94]. The first author of Ref. [94] and I contributed equally to this paper.

Thus, the value $\text{tr}[O\sigma_{\text{eff},\text{Lindblad}}(t)]$ from the effective state (not valid quantum state) can be measured by applying a unitary circuit for the CPTN map $\widetilde{\mathcal{W}}_v$ to $|+\rangle\langle +|_{\text{anc}} \otimes \sigma$ and then measuring it by the observable $X_{\text{anc}} \otimes O$ with a two-outcome POVM to extract the CPTN part.

As far as we know, the proposed Lindblad simulation algorithm is the first one satisfying the following three features simultaneously. First, the gate complexity Eq. (3.55) is independent from the following two important parameters: the number m of terms in Hamiltonian and the number K of jump operators. Next, this algorithm achieves the logarithmic scaling of the gate complexity in terms of accuracy Δ . Finally, we need $4 + \lceil \log_2 M \rceil$ ancilla qubits, which only depends on the parameter M , the number of Pauli strings in a single jump operator. A thorough comparison with previous works is provided in Ref. [94].

Optimal parallel wire-cutting for large system simulation

The power of virtual methods is not limited to the dynamics simulation. Indeed, several circuit cutting methods [135, 122, 138, 24] allow for simulating quantum states with a large number of qubits that are beyond the reach of an actual quantum device. We here introduce one example of such circuit cutting methods based on the cutting of quantum wires (i.e., multi-qubit identity channels).

In the wire-cutting framework, we decompose the multi-qubit identity channel $\mathcal{I}^{\otimes n}$ into a linear combination of measure-and-prepare (m-p) channels [71]. Here, a single-qubit example of m-p channels is

$$\sum_{j=0,1} \text{tr}[|j\rangle\langle j| \bullet] X |j\rangle\langle j| X. \quad (3.58)$$

This m-p channel measures the entering single-qubit state by the computational basis and then prepare single-qubit state $X |j\rangle\langle j|$ according to the measurement outcome. This measure-and-prepare operation clearly separates the input and output via classical communication; thus, the cutting of a single-qubit wire is partially achieved. However, a single measurement basis is not sufficient to recover the full information of the unknown entering state. Thus, the following question naturally arises: how many types of m-p channels are required to cut the n -qubit wires? On the other hand, if one decomposes the ideal identity channel into a linear combination of m-p channels, the absolute sum of the coefficients of the linear combination becomes the sampling overhead, as in the previous cases (this overhead corresponds to C^{2r} for virtual DME and $\sum_v c_v$ for virtual Lindblad simulation). Indeed, the previous work [24] proves that the sampling overhead is lower bounded by $2^{n+1} - 1$. Thus, in addition to the above question about the number of types of m-p channels, we wonder whether we can obtain an optimal decomposition with the minimal sampling overhead.

Here, we introduce a parallel wire-cutting method that achieves the minimal values in both the sampling overhead and the number of m-p channels without any use of ancilla qubits.⁴

Theorem 6 (Ancilla-free doubly optimal parallel wire-cutting [71]). *The n -qubit identity channel $\mathcal{I}^{\otimes n}(\bullet)$ can be decomposed as*

$$\mathcal{I}^{\otimes n}(\bullet) = \sum_{i=1}^{2^n} \sum_{j \in \{0,1\}^n} \text{tr} \left[U_i |j\rangle\langle j| U_i^\dagger \bullet \right] U_i |j\rangle\langle j| U_i^\dagger - (2^n - 1) \sum_{j \in \{0,1\}^n} \text{tr} [|j\rangle\langle j| \bullet] \rho_j, \quad (3.59)$$

where $|j\rangle$ denotes the n -qubit computational basis. The new input state ρ_j , which depends on the measurement result j , is defined as

$$\rho_j = \sum_{\mathbf{k} \in \{0,1\}^n} \frac{1}{2^n - 1} (1 - \delta_{j,\mathbf{k}}) |\mathbf{k}\rangle\langle \mathbf{k}|, \quad (3.60)$$

where $\delta_{j,\mathbf{k}}$ is the Kronecker delta. Also, $\{U_i\}_{i=1}^{2^n} \cup \{I^{\otimes n}\}$ denotes a set of unitary operators that transform the computational base into the $2^n + 1$ mutually unbiased bases. In particular, each unitary U_i can be implemented by a Clifford circuit with the maximal depth of $n + 2$ on a device with fully-connected qubits. Moreover, the number $2^n + 1$ of m-p channels in the decomposition is provably minimal under an ancilla-free assumption.

⁴This method is provided in Ref. [71]. The first author of Ref. [71] and I contributed equally to this paper.

Two orthonormal bases $\{|\phi_i\rangle\}$ and $\{|\psi_j\rangle\}$ in the n -qubit system are said to be mutually unbiased basis sets if and only if $|\langle\phi_i|\psi_j\rangle|^2 = \frac{1}{2^n}$ holds for all $i, j = 0, 1, \dots, 2^n - 1$. In the n -qubit system, there exist $2^n + 1$ distinct orthonormal bases that are mutually unbiased [174]. Since the sampling overhead is the absolute sum of the coefficients, we can calculate it of Eq. (3.59) as $2^{n+1} - 1$, matching the lower bound of Ref. [24].

3.2.2 Variational methods with sequential optimization

Here, we introduce a quantum-classical hybrid quantum state preparation method that fully utilizes the degree of freedom of a single-qubit rotation gate during the circuit optimization process [165, 164].⁵

Free Quaternion Selection for Variational Quantum Algorithm

We first show that cost functions for variational quantum algorithms (VQAs) [27, 154, 170] are formulated as a quadratic form on a single-qubit gate. To this end, we use a quaternion representation of general single-qubit gates:

$$R_{\mathbf{n}}(\psi) := e^{-i\frac{\psi}{2}\mathbf{n}\cdot\vec{\sigma}} = \mathbf{q} \cdot \vec{\zeta}. \quad (3.61)$$

The rotational axis $\mathbf{n}$ and angle ψ in the single-qubit gate are encoded together in a four-dimensional parameter $\mathbf{q}$ with the unit norm, that is, a unit quaternion. The four-component operator $\vec{\zeta} := (I, -iX, -iY, -iZ)$ is an extension of the Pauli matrices $\vec{\sigma} = (X, Y, Z)$ with the 1-qubit identity I . Hereafter, we write $R_{\mathbf{n}}(\psi)$ as $R(\mathbf{q})$ for simplicity.

To prepare a target state, we solve an optimization task with the following objective function

$$\sum_{k=1}^K \text{tr} \left[\rho_k U(\{\mathbf{q}_d\}_{d=1}^D)^\dagger H_k U(\{\mathbf{q}_d\}_{d=1}^D) \right], \quad (3.62)$$

where ρ_k is an n -qubit initial state from a training set, and H_k is some observable. Here, U is an n -qubit parameterized quantum circuit (PQC) comprised of D parameterized single-qubit gates $\{R(\mathbf{q}_d)\}_{d=1}^D$ and parameter-free gates such as CNOT gate. In the following, without loss of generality, we focus on a single expectation value in the objective function (i.e., $K = 1$, which can simply be considered as the minimization of energy for the Hamiltonian $H := H_1$ and the input state $\rho_{\text{in}} := \rho_1$). The extension of our algorithm to the whole objective function is trivial due to the linearity. We remark that a similar idea below can be applied to other objective functions for simulating real- and imaginary-time evolutions [164].

For the energy minimization, we focus on a sequential optimization regarding $R(\mathbf{q}_d)$ where all parameters are fixed except for the d th single-qubit gate as shown in Fig. 3.4. The energy expectation is written as

$$\langle H \rangle(\mathbf{q}_d) := \text{tr} \left[\rho_{\text{in}} U_1^\dagger R(\mathbf{q}_d)^\dagger U_2^\dagger H U_2 R(\mathbf{q}_d) U_1 \right] = \text{tr} \left[\rho'_{\text{in}} R(\mathbf{q}_d)^\dagger H' R(\mathbf{q}_d) \right], \quad (3.63)$$

where U_1 and U_2 are the quantum circuits before and after $R(\mathbf{q}_d)$, respectively. H' and ρ'_{in} are defined as

$$H' := U_2^\dagger H U_2, \quad \rho'_{\text{in}} := U_1 \rho_{\text{in}} U_1^\dagger. \quad (3.64)$$

Here we omit the subscript d for simplicity. Substituting Eq. (3.61) into Eq. (3.63), we can obtain the following quadratic form

$$\langle H \rangle(\mathbf{q}) = \mathbf{q}^\top S \mathbf{q}, \quad (3.65)$$

where superscript $\top$ denotes a transpose operation. Denoting each element of the four-component operator $\vec{\zeta}$ as ς_μ ($\mu = 0, 1, 2, 3$), $S = (S_{\mu\nu})$ is a 4×4 real-symmetric matrix whose elements are defined as

$$S_{\mu\nu} := \frac{1}{2} \text{tr} \left[\rho'_{\text{in}} (\varsigma_\mu^\dagger H' \varsigma_\nu + \varsigma_\nu^\dagger H' \varsigma_\mu) \right], \quad (3.66)$$

⁵This subsection is based on the author's works [165, 164].

and more explicitly

$$S = \begin{pmatrix} \text{tr}(H' \rho'_{\text{in}}) & \frac{i}{2} \text{tr}(H' [\rho'_{\text{in}}, X]) & \frac{i}{2} \text{tr}(H' [\rho'_{\text{in}}, Y]) & \frac{i}{2} \text{tr}(H' [\rho'_{\text{in}}, Z]) \\ \cdot & \text{tr}(H' X \rho'_{\text{in}} X) & \frac{1}{2} [\text{tr}(H' X \rho'_{\text{in}} Y + H' Y \rho'_{\text{in}} X)] & \frac{1}{2} [\text{tr}(H' X \rho'_{\text{in}} Z + H' Z \rho'_{\text{in}} X)] \\ \cdot & \cdot & \text{tr}(H' Y \rho'_{\text{in}} Y) & \frac{1}{2} [\text{tr}(H' Y \rho'_{\text{in}} Z + H' Z \rho'_{\text{in}} Y)] \\ \cdot & \cdot & \cdot & \text{tr}(H' Z \rho'_{\text{in}} Z) \end{pmatrix}, \quad (3.67)$$

where $[\cdot, \cdot]$ denotes the commutation relation. The derivation of the quadratic form is as follows. Substituting the quaternion representation of single-qubit gates $R(\mathbf{q}) = \mathbf{q} \cdot \vec{\zeta}$ into the energy (3.63), we obtain

$$\begin{aligned} \langle H \rangle(\mathbf{q}) &= \text{tr} [\rho'_{\text{in}} R(\mathbf{q})^\dagger H' R(\mathbf{q})] = \sum_{\mu, \nu=0}^3 q_\mu q_\nu \text{tr} [\rho'_{\text{in}} \zeta_\mu^\dagger H' \zeta_\nu] \\ &= \sum_{\mu, \nu=0}^3 A_{\mu\nu} q_\mu q_\nu, \quad A_{\mu\nu} := \text{tr} [\rho'_{\text{in}} \zeta_\mu^\dagger H' \zeta_\nu] \\ &= \frac{1}{2} \left(\sum_{\mu, \nu=0}^3 A_{\mu\nu} q_\mu q_\nu + \sum_{\nu, \mu=0}^3 A_{\nu\mu} q_\nu q_\mu \right) = \frac{1}{2} \left(\sum_{\mu, \nu=0}^3 A_{\mu\nu} q_\mu q_\nu + \sum_{\mu, \nu=0}^3 A_{\nu\mu} q_\mu q_\nu \right) \\ &= \sum_{\mu, \nu=0}^3 \frac{A_{\mu\nu} + A_{\nu\mu}}{2} q_\mu q_\nu = \mathbf{q}^\top S \mathbf{q}. \end{aligned} \quad (3.68)$$

Here, the S is obviously a real symmetric matrix since $\zeta_\mu^\dagger H' \zeta_\nu + \zeta_\nu^\dagger H' \zeta_\mu$ is a Hermitian operator.

The matrix S can be obtained by running and measuring ten quantum circuits, each corresponding to the element in the upper diagonal of S , as detailed below. Note that the minimization of the quadratic form is exactly achieved by calculating the eigenvector corresponding to the lowest eigenvalue of S . All the elements of S in Eq. (3.67) are calculated from ten expectation values classified into three types as below.

Type-A

$$\text{tr} [\rho'_{\text{in}} H']$$

Type-B for $k = 1, 2, 3$,

$$\text{tr} \left[\rho'_{\text{in}} \left(\frac{I \pm i\sigma_k}{\sqrt{2}} \right)^\dagger H' \left(\frac{I \pm i\sigma_k}{\sqrt{2}} \right) \right]$$

Type-C for $(k, m) = (1, 2), (1, 3), (2, 3)$

$$\text{tr} \left[\rho'_{\text{in}} \left(\frac{\sigma_k + \sigma_m}{\sqrt{2}} \right)^\dagger H' \left(\frac{\sigma_k + \sigma_m}{\sqrt{2}} \right) \right].$$

Here, Type-A is the $(0, 0)$ -the element of S and corresponds to the expectation value of H on the PQC when the single-qubit gate of interest, as in Fig. 3.4, is replaced with identity. The other diagonal elements are produced by the type-A and type-B values with the following identity:

$$\begin{aligned} &\text{tr}(H' \sigma_k \rho'_{\text{in}} \sigma_k) + \text{tr}(H' \rho'_{\text{in}}) \\ &= \text{tr} \left[\rho'_{\text{in}} \left(\frac{I - i\sigma_k}{\sqrt{2}} \right)^\dagger H' \left(\frac{I - i\sigma_k}{\sqrt{2}} \right) \right] + \text{tr} \left[\rho'_{\text{in}} \left(\frac{I + i\sigma_k}{\sqrt{2}} \right)^\dagger H' \left(\frac{I + i\sigma_k}{\sqrt{2}} \right) \right]. \end{aligned} \quad (3.69)$$

Note that the Type-B corresponds to the expectation value of H on the PQC when the single-qubit gate of interest, as in Fig. 3.4, is replaced with, respectively, $R_x(-\pi/2)$, $R_x(\pi/2)$, $R_y(-\pi/2)$, $R_y(\pi/2)$, $R_z(-\pi/2)$, and $R_z(\pi/2)$. In contrast, subtracting the type-B values with different signs yields the

other elements in the first row directly. The remaining off-diagonal elements are produced by the type-C expectation values and the already obtained diagonal elements with the following identity:

$$\begin{aligned} & \frac{1}{2} [\text{tr}(H' \sigma_i \rho'_{\text{in}} \sigma_j) + \text{tr}(H' \sigma_j \rho'_{\text{in}} \sigma_i)] \\ &= \text{tr} \left[\rho'_{\text{in}} \left(\frac{\sigma_i + \sigma_j}{\sqrt{2}} \right)^\dagger H' \left(\frac{\sigma_i + \sigma_j}{\sqrt{2}} \right) \right] - \frac{1}{2} [\text{tr}(H' \sigma_i \rho'_{\text{in}} \sigma_i) + \text{tr}(H' \sigma_j \rho'_{\text{in}} \sigma_j)]. \end{aligned} \quad (3.70)$$

Note that the Type-C values correspond to the expectation values of H on the PQC when the single-qubit gate to be optimized, as in Fig. 3.4, is replaced with, respectively, $R_{\mathbf{n}}(\pi)$ for $\mathbf{n} \propto (1, 1, 0), (1, 0, 1), (0, 1, 1)$. All expectation values of Type-A, B, and C can be evaluated with direct measurements without any control operation such as the Hadamard test. Since the degree of freedom for a 4×4 real-symmetric matrix is ten, the number of required direct measurements should be optimal.

The optimization over the whole $SU(2)$ is a generalization of other sequential optimizers [127, 132, 169], which optimize only a part of $SU(2)$. Also, we call the sequential algorithm repeating the local update of single-qubit gates based on the quadratic form *Free Quaternion Selection (FQS)*. Since FQS can select the optimal gate from the whole $SU(2)$ for minimizing the energy expectation in Eq. (3.63), it can incorporate the multi-parameter correlation in $SU(2)$ and thus achieves better performance for optimizing PQCs than other sequential optimizers. Rotoselect, Rotosolve, and their variants [127, 132, 161, 134, 101, 173] can be used to optimize a general single-qubit gate by first decomposing the multi-parameter gate using ZYZ-decomposition [128] to obtain three single-parameter gates. In Rotosolve/select and their variants, each parameter of the decomposed gate is optimized locally in contrast to FQS and Fraxis, where a general single-qubit gate is decomposed into two Fraxis gates defined as $R_{\mathbf{n}}(\pi)$. Fraxis [169] updates these gates individually but simultaneously optimizes two parameters within the gate.

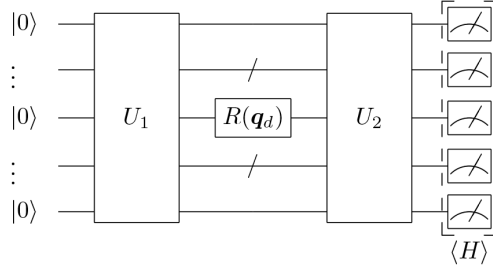


Figure 3.4: Quantum circuit to optimize the single-qubit gate $R(\mathbf{q}_d)$. The variational parameters in U_1 and U_2 are fixed during the optimization of the gate of interest. Our algorithm sequentially replaces $R(\mathbf{q}_d)$ in the circuit by the optimal one, which is computed by the diagonalization of the corresponding matrix S whose elements can be evaluated from the circuit by replacing $R(\mathbf{q}_d)$ with ten different gates. Reproduced from K. Wada *et al.*, *Sequential optimal selections of single-qubit gates in parameterized quantum circuits*, Quantum Sci. Technol. 9 035030 (2024), DOI: 10.1088/2058-9565/ad4583. Licensed under CC BY 4.0.

Unification of sequential quantum optimizers

FQS generalizes all known sequential optimizers for single-qubit gates of PQCs, such as Rotosolve [132] (NFT [127]), Rotoselect [132], Fraxis [169]; those methods can be regarded as special cases of FQS. For NFT, a single-qubit gate is restricted to a fixed axis $\mathbf{m}$ such as $U_{\text{NFT}} := R_{\mathbf{m}}(\psi)$. Then the corresponding objective function in quadratic form is

$$\text{tr} [\rho'_{\text{in}} U_{\text{NFT}}^\dagger H' U_{\text{NFT}}] = \mathbf{c}^\top \begin{pmatrix} S_{00} & \vec{S}_0 \cdot \mathbf{m} \\ \vec{S}_0 \cdot \mathbf{m} & \mathbf{m}^\top \tilde{S} \mathbf{m} \end{pmatrix} \mathbf{c}, \quad (3.71)$$

where $\mathbf{c} := (\cos \psi/2, \sin \psi/2)^\top$ and $\tilde{S}_0 := (S_{01}, S_{02}, S_{03})$. $\tilde{S}$ denotes the lower right 3×3 part of the S matrix. The real symmetric matrix in Eq. (3.71) can be regarded as a contraction of the FQS matrix S with respect to the rotational axis $\mathbf{m}$, reducing its degree of freedom to three. In Rotoselect, three contracted matrices in Eq. (3.71) are constructed for $\mathbf{m} \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and the lowest eigenvalue is selected after three separate diagonalization procedures. Since the three matrices share S_{00} , the total number of circuit evaluations can be reduced to seven.

As for Fraxis, the target gate $U_{\text{Fraxis}} := R_{\mathbf{n}}(\pi)$ is simply expressed by the quaternion $\mathbf{q} = (0, \mathbf{n})$. Thus, substituting $\mathbf{q} = (0, \mathbf{n})$ into the objective function, we can reproduce the previous results derived in [169] as follows

$$\text{tr} \left[\rho'_{\text{in}} U_{\text{Fraxis}}^\dagger H' U_{\text{Fraxis}} \right] = \mathbf{n}^\top \tilde{S} \mathbf{n}. \quad (3.72)$$

Note that the θ -Fraxis of [169], in which the rotation angle is fixed to arbitrary values θ instead of π , can be regarded as minimizing Eq. (3.65) for $\mathbf{q}' := (q_1, q_2, q_3) \in \mathbb{R}^3$ under the constraint $|\mathbf{q}'|^2 = 1 - q_0^2$, which results in solving simultaneous equations rather than diagonalization.

It is worth noting that the required number of circuit evaluations for each sequential quantum optimizer coincides with the degrees of freedom for the real-symmetric matrix of respective methods, i.e., $3 = 1 + 2$ circuit evaluations for NFT/Rotosolve, $6 = 1 + 2 + 3$ circuit evaluations for Fraxis, and $10 = 1 + 2 + 3 + 4$ circuit evaluations for FQS.

Barren plateaus in sequential quantum optimizers

We have shown that the energy landscape of all sequential optimizers can be derived from the eigenvalues of the S matrices whose elements are computed from circuit evaluations of slightly modified PQC. Sequential quantum optimizers can effectively obtain a better single-qubit gate minimizing the cost function as long as the eigenvalues of S are *significantly far* from degenerate. Hence, it is useful to quantify degeneracy of S for evaluation of performance of sequential quantum optimizers. To this end, we introduce a centered matrix $S_c^{(p)}$ defined as,

$$S_c^{(p)} := S^{(p)} - \frac{\text{tr} [S^{(p)}]}{p} I_{p \times p}, \quad (3.73)$$

where $I_{p \times p}$ denotes a $p \times p$ identity matrix. Here, $S^{(p)}$ is a $p \times p$ real-symmetric matrix for each sequential quantum optimizer, and the degrees of freedom of target single-qubit gate are $p - 1$ e.g., $p = 2$ (Rotosolve/NFT), $p = 3$ (Fraxis), and $p = 4$ (FQS). Henceforth, we omit the superscript p for simplicity because it is clear from the context. Since the mean of the eigenvalues of S_c is zero, the spectral radius of S_c indicates the spread of the spectrum of S from the mean of the eigenvalues of S . More precisely, defining $\hat{\lambda}(S)$ as the eigenvalue of S with the largest distance from $\text{tr}[S]/p$, which is the mean of eigenvalues of S , the spectral radius r of S_c is equivalent to

$$r(S_c) \equiv \left| \hat{\lambda}(S) - \frac{\text{tr}[S]}{p} \right|. \quad (3.74)$$

In the following, we provide two theorems on the behavior of the spectrum of each real-symmetric matrix in Eqs. (3.65), (3.71), and (3.72) on the condition that the parameters in U_1 and U_2 , as in Fig. 3.4, are randomly initialized, which is standard for studying barren plateaus in gradient-based methods [121, 28]. The proof of the theorems is proved in Appendix 7.1.

An upper bound of Eq. (3.74) that decays exponentially with the number of qubits hints at the existence of a flat energy landscape (barren plateaus). Indeed, we reveal that the flat landscape can happen under similar conditions assumed in [121] i.e., if the circuits become sufficiently deep to achieve a unitary 2-design, the second moment of the spectral radius of the centered matrix shows exponential shrink regardless of the global or local cost.

Theorem 7. *Suppose that the quantum circuits U_1 and U_2 are randomly and independently generated. If either U_1 or U_2 forms a unitary t -design with $t \geq 2$, the second moment of the spectral radius r of*

the centered matrix Eq. (3.73) is upper bounded as,

$$\begin{aligned} & \mathbb{E}_{U_1, U_2} \left[r(S_c)^2 \right] \\ & \leq \begin{cases} \frac{p^2 \text{tr}[H^2] \Delta \rho_{\text{in},2}}{2(2^{2n}-1)} + \frac{\Delta \rho_{\text{in},2}}{4p(2^{2n}-1)} \sum_{\mu, \nu=0}^{p-1} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} [\mathbb{E}_{U_2} [H' \zeta_\mu \zeta_\nu^\dagger H' \zeta_\nu \zeta_\mu^\dagger + \text{h.c.}]] \\ \frac{p^2 \text{tr}[\rho_{\text{in}}^2] \Delta H_2}{2(2^{2n}-1)} + \frac{\Delta H_2}{4p(2^{2n}-1)} \sum_{\mu, \nu=0}^{p-1} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} [\mathbb{E}_{U_1} [\rho'_{\text{in}} \zeta_\mu^\dagger \zeta_\nu \rho'_{\text{in}} \zeta_\nu^\dagger \zeta_\mu + \text{h.c.}]] \end{cases}, \end{aligned} \quad (3.75)$$

where the first case corresponds to U_1 being a t -design, and the second case corresponds to U_2 being a t -design. Here, $\mathbb{E}_{U_1, U_2}[\cdot]$ is defined as the expectation over the random quantum circuits U_1 and U_2 , $\delta_{\mu\nu}$ denotes the Kronecker delta, h.c. means the Hermite conjugate of the preceding term, and $\Delta \rho_{\text{in},2}$, ΔH_2 are defined as

$$\Delta \rho_{\text{in},2} := \text{tr}[\rho_{\text{in}}^2] - \frac{1}{2^n}, \quad \Delta H_2 := \text{tr}[H^2] - \frac{\text{tr}^2[H]}{2^n}. \quad (3.76)$$

In particular, in the case of $p = 2$ (Rotosolve/NFT), the right hand side of Eq. (3.75) is equivalent to the variance of gradient of the objective function with respect to the rotation angle in a target gate, which is calculated in Ref. [121]. Therefore, we can rewritten Eq. (3.75) with $p = 2$ as follows:

$$\mathbb{E}_{U_1, U_2} \left[r(S_c^{(2)})^2 \right] \leq \text{Var} \left[\frac{\partial}{\partial \theta} \langle H \rangle \right],$$

where $\langle H \rangle$ denotes the energy expectation for the quantum state in Fig. (3.4) with $R = e^{-i\theta \mathbf{n} \cdot \vec{\sigma}}$. This theorem means that if either U_1 or U_2 has sufficient expressibility i.e., unitary 2-design over n -qubit, the spectrum of S matrix concentrates on a single value and its deviation is exponentially small with respect to the number of qubits. As a result, this spectrum concentration implies that the energy of the output quantum state becomes insensitive exponentially on the selection of single-qubit gates.

Although Theorem 7 assumes the relatively deep circuit forming unitary t -design with $t \geq 2$, we are interested in shallower circuits due to the existence of noise-induced barren plateau [168]. Here, we show that, in contrast to Theorem 7, there are shallow circuits avoiding the exponential shrink of the spectral radius when the target Hamiltonian is local.

To this end, let us consider an n -qubit alternating layered ansatz U [28, 126] with m -qubit parametrized unitary blocks, as depicted in Fig. 3.5(a). Each block consists of parametrized single-qubit gates and parameter-free gates. Suppose the ansatz consists of L layers, and each layer contains ξ blocks (i.e., $n = \xi m$). We define S_k ($k = 1, 2, \dots, \xi$) as the m -qubit subsystem on which the k th block from the top in the final layer acts. In the following, we focus on a block W in the l th layer and a parameterized single-qubit gate R in the block as in Fig. 3.5(b). Moreover, we define the forward light-cone $\mathcal{L}$ of the block W as a series of gates that has at least one input qubit causally connected to the output qubits of W . As well, the backward light-cone $\mathcal{L}_B$ of W is defined in the reverse direction of $\mathcal{L}$.

Theorem 8. Suppose that the whole quantum circuit U is an n -qubit alternating layered ansatz with m -qubit blocks as described in Fig. 3.5. Here, we focus on a block W in l th layer and a parameterized single-qubit gate R in the block. We assume that the quantum circuits $W_A, W_B \subset W$, which are located after and before the target gate respectively, and the other blocks form a local 2-design independently. In addition, the Hamiltonian H is assumed to be m -local such as

$$H = c_0 I^{\otimes n} + \sum_i c_i h_i, \quad c_0, c_i \in \mathbb{R}, \quad (\exists i, c_i \neq 0), \quad (3.77)$$

where h_i is a tensor product of Pauli matrices that acts non-trivially on at most m -qubit. (The same assumption was used in [28].) Then, the second moment of the spectral radius r of the centered matrix Eq. (3.73) satisfies

$$\mathbb{E}_{U_1, U_2} \left[r(S_c)^2 \right] \geq \frac{(p+2)(p-1)2^{m(l+1)-1}}{p(2^{2m}-1)^2(2^m+1)^{L+l}} \sum_{i \in \mathcal{L}} \sum_{\substack{(k, k') \in \mathcal{L}_B \\ k' \geq k}} c_i^2 \epsilon(\rho_{k, k'}) \epsilon(h_i), \quad (3.78)$$

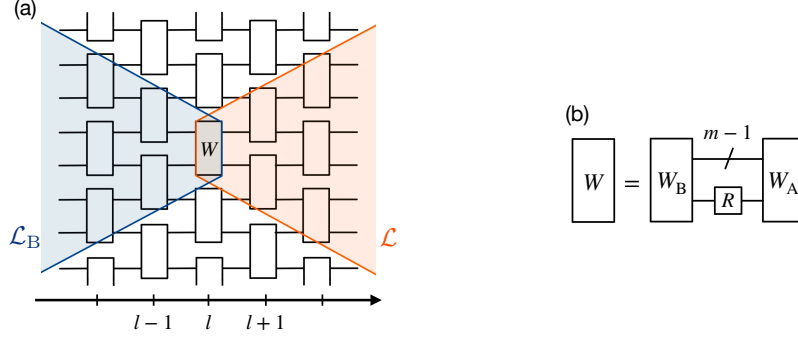


Figure 3.5: Schematic diagram of the alternating layered ansatz with m -qubit unitary blocks. In Theorem 8, we focus on a single-qubit gate R in a block W of the l th layer as depicted in (b), where W_A and W_B are a set of unitary gates located after and before the target single-qubit gate. $\mathcal{L}$ ($\mathcal{L}_B$) denotes the forward (backward) light-cone with respect to W . Here, U_1 in Fig. 3.4 includes the gates in $\mathcal{L}$ and W_A , and U_2 includes the other gates. Reproduced from K. Wada *et al.*, *Sequential optimal selections of single-qubit gates in parameterized quantum circuits*, Quantum Sci. Technol. 9 035030 (2024), DOI: 10.1088/2058-9565/ad4583. Licensed under CC BY 4.0.

where L denotes the total number of layers. Here, $i_{\mathcal{L}}$ is the set of i indices whose associated operators h_i act on qubits in the forward light-cone $\mathcal{L}$ of W , and $k_{\mathcal{L}_B}$ is the set of k indices whose associated subsystems S_k are in the backward light-cone $\mathcal{L}_B$ of W . The quantum state $\rho_{k,k'}$ is the reduced density matrix of the input state ρ_{in} on $S_k S_{k+1} \cdots S_{k'}$, and the function $\epsilon(M)$ for a matrix M is defined as $\epsilon(M) = D_{\text{HS}}(M, \text{tr}(M)\mathbf{1}/d_M)$ where D_{HS} is the Hilbert-Schmidt distance and d_M is the dimension of the matrix M .

We prove this theorem by firstly deriving the relation between the lower bound of the spectral radius and the variance of the gradient regarding θ in the case $R = e^{-i\theta \mathbf{n} \cdot \vec{\sigma}}$, under the condition of Theorem 8. More precisely, we prove the following relation:

$$\mathbb{E}_{U_1, U_2} [r(S_c)^2] \geq \frac{(p+2)(p-1)}{4p} \times \text{Var} \left[\frac{\partial}{\partial \theta} \langle H \rangle \right].$$

Then, combining the previous result in Ref. [28], we complete the proof of Theorem 8; see Appendix 7.1. This theorem means the lower bound of the second moment of the spectral radius does not vanish exponentially fast when we employ shallow L , which leads to similar properties as the gradient-based counterparts [28]. More precisely, if at least one term $c_i^2 \epsilon(\rho_{k,k'}) \epsilon(h_i)$ vanishes no faster than $\Omega(1/\text{poly}(n))$, and if the number of layers L is $\mathcal{O}(\log(n))$, then

$$\mathbb{E}_{U_1, U_2} [r(S_c)^2] = \Omega \left(\frac{1}{\text{poly}(n)} \right) \quad (3.79)$$

holds. This guarantees trainability of sequential quantum optimizers in the beginning of optimization. On the other hand, if at least one term $c_i^2 \epsilon(\rho_{k,k'}) \epsilon(h_i)$ vanishes no faster than $\Omega(1/2^{\text{poly}(\log(n))})$, and if the number of layer L is $\mathcal{O}(\text{poly}(\log(n)))$ then,

$$\mathbb{E}_{U_1, U_2} [r(S_c)^2] = \Omega \left(\frac{1}{2^{\text{poly}(\log(n))}} \right) \quad (3.80)$$

holds. This suggests a transition region of trainability where the spread of the spectrum decays faster than polynomial but slower than exponential. Note that Cerezo et al. [28] showed a detailed analysis of variance of gradient with respect to angles of fixed-axis rotation gates. In contrast, for a global cost function, the spectral radius is expected to shrink exponentially even on a constant-depth alternating layered ansatz in line with the gradient-based methods as [28].

There is a new finding in Eq. (3.78) which is not available from the previous study on the gradient; the lower bound depends on the parameter p . We can see this from the simplest case when $p = 2$ for Rotosolve (NFT). In this case, the lower bound of the spectral radius appears to be twofold of that of

the gradient-based method in [28]. This is probably because the spectral radius measures all possible energy variations that can occur by replacing the target gate, while the gradient-based analysis focuses on only a local response along a slight move in the parameter space.

Also, the lower bound implies the larger the value of p the better the chance to escape from barren plateaus. Although the values of p are limited to 2 (NFT), 3 (Fraxis), and 4 (FQS) for single-qubit gates in this study, our spectral radius analysis can be extended to a multi-qubit gate with $p > 4$ that seems to grow at least polynomially in the number of qubits. Such an extension of our method is actually an important open problem, but it would be achieved by imposing an appropriate structure of parameterized gate to the method in the previous paper [145] or a more detailed analysis of the unified expressions of sequential optimizers.

3.3 Other basic quantum algorithms

We review some important results in efficiently calculating block-encoded matrices on a quantum computer. These results provide the basis of quantum algorithms in the subsequent chapters, as well as for a broad class of modern quantum algorithm designs [119].

3.3.1 Block encoding

The basic tool to represent matrices by unitary operators of dilated quantum systems is *block encoding* which is formally defined as follows.

Definition 3 (Block encoding [57]). For positive values α, ε and a non-negative integer a , we say that an $(n + a)$ -qubit unitary U is an (α, a, ε) -block-encoding of an n -qubit operator A , if

$$\|A - \alpha(|0\rangle^{\otimes a} \otimes \mathbf{1})U(|0\rangle^{\otimes a} \otimes \mathbf{1})\| \leq \varepsilon,$$

where $\|\bullet\|$ denotes the spectral norm. For simplicity, we shorten the perfect (i.e., $\alpha = 1$ and $\varepsilon = 0$) block encoding of A as a -block-encoding of A .

For instance, any unitary operator (e.g., a Pauli operator $X \otimes Z \otimes \dots$) is trivially a 0-block-encoding of itself. In general, there are various ways to construct block encodings [116, 57], and the quest for developing more efficient encoding is still ongoing. In the following, we review two basic constructions: (i) block-encodings obtained via the method called the linear combination of unitaries (LCU) [33, 13], and (ii) block-encodings of sparse matrices.

We first review the LCU-based construction. Let $A := \sum_{i=1}^m c_i U_i$ be a linear combination of unitary operators $\{U_i\}_{i=1}^m$ with real coefficients $c_i \in \mathbb{R}$. Without loss of generality, we assume $c_i > 0$ because -1 can be absorbed into U_i . In order to encode A , we use the following two unitary operations. The first unitary, called PREPARE, encodes the positive coefficients $\{c_i\}_{i=1}^m$ as

$$V : |0\rangle \mapsto \sum_{i=1}^m \sqrt{\frac{c_i}{\|c\|_1}} |i\rangle,$$

where $\|\bullet\|_1$ denotes L^1 -norm, and $|0\rangle$ and $|i\rangle$ denote an initial state and the computational basis in a $\lceil \log_2 m \rceil$ -qubit ancilla system, respectively. The second unitary, called SELECT, encodes the unitary operators U_i conditioned by the $\lceil \log_2 m \rceil$ -qubit ancilla system:

$$W = \sum_{i=1}^m |i\rangle\langle i| \otimes U_i.$$

Using the two operations V and W , it can be shown that the unitary operator $(V^\dagger \otimes \mathbf{1})W(V \otimes \mathbf{1})$ is a $(\|c\|_1, \lceil \log_2 m \rceil, 0)$ -block-encoding of A , as in Fig. 3.6. Note that if the coefficients in LCU are controlled by other qubit registers, it may be useful to modify the PREPARE operator instead of including the phase of c_i to the SELECT operator, in order to save the number of controlled operations.

We next recall a standard block-encoding for sparse matrices [116], which is particularly useful for the case where we have matrix representations of target observables.

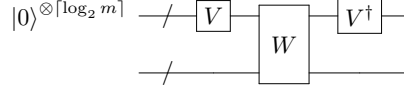


Figure 3.6: Quantum circuit for the linear combination of unitaries (LCU) method. The circuit is a $\lceil \log_2 m \rceil$ -block-encoding of $\|c\|_1^{-1} \sum_{i=1}^m c_i U_i$ ($c_i > 0$). PREPARE $V : |0\rangle^{\otimes \lceil \log_2 m \rceil} \mapsto \sum_{i=1}^m \sqrt{\frac{c_i}{\|c\|_1}} |i\rangle$ and SELECT $W = \sum_{i=1}^m |i\rangle\langle i| \otimes U_i$. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

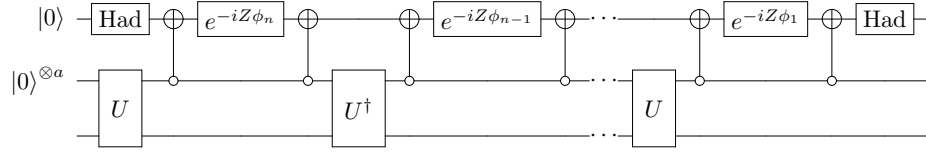


Figure 3.7: Quantum circuit for quantum singular value (eigenvalue) transformation for real polynomials P of odd degree n . U denotes an a -block-encoding of a Hermitian operator A . The NOT gate is controlled by $|0\rangle^{\otimes a}$ of the a qubits, which is represented by the white circle $\circ$. For a given real polynomial P of degree n , the n circuit parameters $\{\phi_i\}_{i=1}^n$ are calculated in $\mathcal{O}(\text{poly}(n))$ classical computation time. With such parameters, this circuit results in an $(a+1)$ -block-encoding of $P(A)$. Adapted from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

Lemma 3 (Lemma 6 in Ref. [116], rephrased). Let O_H, O_F be the standard oracles specifying a d -sparse Hamiltonian H with max-norm $\|H\|_{\max}$:

$$O_H |j\rangle |k\rangle |z\rangle = |j\rangle |k\rangle |z \oplus \hat{H}_{jk}\rangle, \quad O_F |j\rangle |l\rangle = |j\rangle |f(j, l)\rangle, \quad (3.81)$$

where j, k denote the index of row and column, respectively, $f(j, l)$ is the column index of the l th non-zero element of H in row j , and $\hat{H}_{jk}$ represents the (j, k) element of H in some binary format. Then, we can implement a block encoding of $H/d\|H\|_{\max}$ using $\mathcal{O}(1)$ queries to O_F and O_H .

In many applications, observables naturally appear either as linear combinations of Pauli strings or as sparse matrices. For example, the Hamiltonian of molecular systems can be written as a linear combination of Pauli strings via qubit-mapping techniques such as Jordan-Wigner or Bravyi-Kitaev transformations. Likewise, Ising or Heisenberg Hamiltonians are naturally expressed as sums of Pauli strings. Typical examples of physical relevant properties include the (average) magnetization $N^{-1} \sum_i Z_i$ and the two-point correlation $N^{-1} \sum_i Z_i Z_{i+r}$ on N sites.

3.3.2 Quantum singular value transformation (QSVT)

Once we have a block encoding of a target operator, we can systematically transform the block encoding to perform various tasks. We first show several concrete examples of such transformations in the form of lemmas, and then explain the underlying structure that unifies them. These lemmas will be used in the following chapter.

Lemma 4 (Uniform singular value amplification [114, 57]). Let $\gamma > 1$ and let $\delta, \varepsilon \in (0, 1/2)$. Suppose we have an a -block-encoding U of A , ($\|A\| \leq (1-\delta)/\gamma$). Then, we can implement a $(1, a+1, \varepsilon)$ -block-encoding of γA with $m = \mathcal{O}(\gamma \delta^{-1} \log(\gamma/\varepsilon))$ queries to U and $U^\dagger$, $2m$ uses of NOT gates controlled by a-qubit, $\mathcal{O}(m)$ single-qubit gates, and $\mathcal{O}(\text{poly}(m))$ classical computation to find quantum circuit parameters.

Lemma 5 (Quantum eigenvalue transformation by Chebyshev polynomials [57]). *Let m be a positive integer, and let U be an a -block-encoding of a Hamiltonian H . Then, for the m -th Chebyshev polynomial of the first kind $T_m(x)$, we can implement a $(1, a, 0)$ -block-encoding of $T_m(H)$, with m uses of U and $U^\dagger$ and m uses of reflection on $|0\rangle^{\otimes a}$.*

Lemma 6 (Optimal block-Hamiltonian simulation [116]). *Let $t \in \mathbb{R} \setminus \{0\}$, $\varepsilon'' \in (0, 1)$, and let U be a $(1, a, 0)$ -block-encoding of a Hamiltonian H . Then, we can implement a $(1, a + 2, \varepsilon'')$ -block-encoding of e^{itH} , with $4Q$ queries to controlled U and its inverse, $2Q$ uses of NOT gates controlled by $(a+1)$ -qubit, $\mathcal{O}(Q)$ uses of single-qubit or two-qubit gates, and $\mathcal{O}(\text{poly}(Q))$ classical computation to find quantum circuit parameters, where $Q = \mathcal{O}(|t| + \log(1/\varepsilon''))$.*

Lemma 4 and Lemma 5 can be implemented with a quantum circuit in Fig. 3.7. Also, we can implement the optimal Hamiltonian simulation Lemma 6 with a circuit similar to Fig. 3.7; the explicit circuit constructions are provided in Refs. [115, 116, 57].

Importantly, the quantum circuit in Fig. 3.7 reflects the underlying structure that are common in the above lemmas; this quantum circuit implements a general method, called the quantum singular value transformation (QSVT), to transform singular values (eigenvalues) of a block-encoded matrix based on a large class of polynomials [57]. The QSVT uses the idea of quantum signal processing (QSP) [117, 116] that characterizes achievable 1-qubit unitary transformations comprised of an alternating 1-qubit gate sequence of the *signal* rotation with an unknown angle and the *processing* rotation with a controllable angle. To bridge the gap between QSP and QSVT, *Qubitization* [116] is a crucial technique that splits (a part of) ancilla-target systems into some qubits labeled by the eigenvalue (singular values [57]) and constructs parallel signal rotations over the qubits (e.g., U and $U^\dagger$ in Fig. 3.7). Here, the rotation angle depends on the corresponding singular value. Then, using additional processing rotations with controllable parameters $\{\phi_i\}$ (likewise the controlled rotation between U and $U^\dagger$ in Fig. 3.7), we can transform the singular values in parallel by a polynomial that depends on the controllable parameters $\{\phi_i\}$; the achievable polynomials in QSVT are characterized by QSP. See the review [119] for details of the theoretical perspective of QSVT.

In practice, a typical flow of QSVT consists of two steps: (i) finding the circuit parameter $\{\phi_i\}_{i=1}^n$ (called the phase sequence) for a given degree- n real polynomial P on classical computers, (ii) running the $\mathcal{O}(n)$ -depth quantum circuit in Fig. 3.7 on a quantum computer, using the classically tuned parameters. Note that in the process (ii), we may need a post-selection on ancilla qubits. For a given degree- n polynomial P that has definite parity and $P(x) \in [-1, 1]$ for $x \in [-1, 1]$, the n circuit parameters $\{\phi_i\}_{i=1}^n$ in Fig. 3.7 can be found by $\mathcal{O}(\text{poly}(n, \log(1/\delta)))$ classical computation for some error δ [57]. Then, using the parameters this circuit results in an $(a + 1)$ -block-encoding of $P(A)$. In Lemma 4, we can take an odd real polynomial $P(x)$ with degree $m = \mathcal{O}(\gamma\delta^{-1} \log(\gamma/\varepsilon))$ such that $P(x) \approx \gamma x$ holds [114, 57]. In particular, the phase sequence $\{\phi_i\}$ for the Chebyshev polynomial of the first kind T_n is analytically calculated and has a unique structure (Lemma 9 in Ref. [57]); as a result, we can eliminate the additional ancilla qubit and replace the $2n$ controlled NOT gates with n reflections on $|0\rangle^{\otimes a}$ in the circuit of Fig. 3.7.

Chapter 4

Heisenberg-limited multi-observable estimation

4.1 Introduction

4.1.1 Background

Attaining quantum enhancement in unknown parameter estimation lies as one of the most fundamental tasks in quantum technology.¹ It has been noticed in the quantum metrological community from the late 20th century that there exists a gap between the SQL, the statistical scaling of the measurement count $\mathcal{O}(1/\varepsilon^2)$ based on the central limit theorem, and the HL, the scaling $\mathcal{O}(1/\varepsilon)$ due to quantum uncertainty relations that are quantified by the root MSE ε . Theoretically, the HL can be achieved using entanglement or coherence (i.e., sequential applications of a sensing channel) in quantum probes under the absence of noise [58, 60], while its experimental verification has not been realized for a long time until the phase estimation procedure by Higgins *et al.* that uses a novel adaptive measurement [78]. Aside from Ref. [78], intensive quest for experimental realization of the scaling beyond the SQL has provoked numerous interesting ideas such as the use of quantum error correction [95, 6, 46, 181], exploiting the non-Markovianity of the environment [120, 34].

The preceding pursuit for the fundamental limitation by quantum mechanics naturally pertains to the quantum-enhanced measurement of multiple observables. This long-standing question has been posed mainly in the community of quantum computing, since an overwhelming number of quantum algorithms estimate the expectation values of local and/or global observables; this task is typically summarized as $M = \mathcal{O}(N^k)$ (possibly non-commuting) observables estimation on an N -qubit target system, for some k . Not to mention the scientific application of quantum computers for quantum simulation of many-body systems in natural science [112, 8, 32, 57, 17, 178], such a task is also ubiquitous in industrial uses such as quantum machine learning [19, 111, 84] and quantum finance [146, 29]. Note that, while one may employ the modified QAE algorithm [149, 167, 162] individually for M observables to simply obtain the HL scaling of $\mathcal{O}(M/\varepsilon)$ in queries to a target state preparation with respect to the root MSE ε , here we seek for simultaneous improvement on the target precision ε and observable count M . Simply put, our goal is to construct an algorithm that achieves both the HL scaling $\mathcal{O}(1/\varepsilon)$ and the sublinear scaling with respect to M in estimating all of M observables.

Indeed, some previous works based on the QGE algorithm [91, 56] have reported nearly HL scaling, although with multiplicative logarithmic correction $\log(1/\varepsilon)$ [159, 86]; there is no work that achieves the ultimate precision scaling following the HL in light of its definition. Furthermore, existing works argue the estimation precision in terms of confidence intervals, which crucially fails to bound the worst-case behavior of a single run, unlike the MSE. As is well known in the field of quantum metrology [16, 77, 60, 12], the MSE successfully bounds other measures of uncertainty, while the converse does not hold in general. Considering that many quantum algorithms that estimate observables make queries to the state preparation oracle of a target quantum state whose complexity usually scales with the system size, it is crucial to design an estimation algorithm that achieves the HL scaling in terms of the MSE.

¹This chapter is based on the author's work [166].

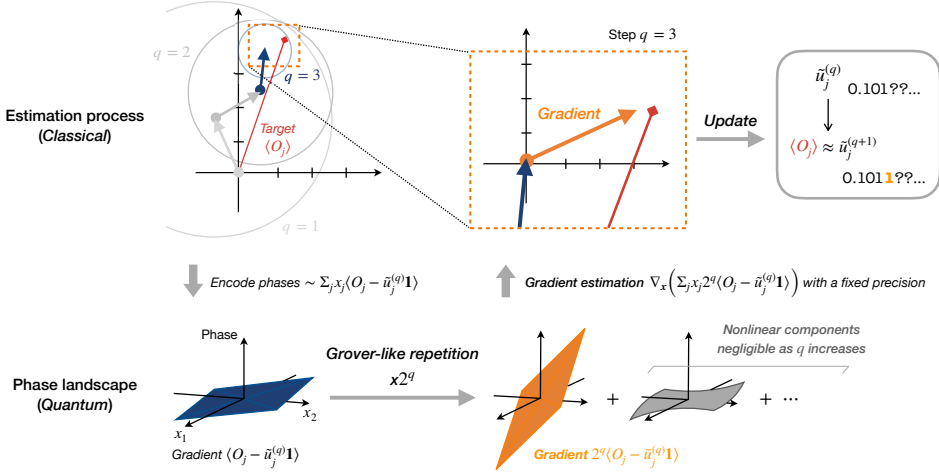


Figure 4.1: Graphical summary of the proposed method. Our algorithm estimates the expectation values of general observables $\{\langle O_j \rangle\}_{j=1}^M$ regarding a quantum state $|\psi\rangle$ prepared by a unitary U_ψ , with an adaptive procedure. In the q th iteration step, we first encode the values of estimation errors $\{\langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle\}$ for a unitary gate, where $\tilde{u}_j^{(q)} \in [-1, 1]$ denotes a temporal estimated value from the previous step, for the target value $\langle O_j \rangle$. In the upper left illustration, each circle denotes a closed ball with $1/2^q$ radius centered at the q th estimate $\{\tilde{u}_j^{(q)}\}$. Then, we coherently amplify the estimation errors by the factor of 2^q via $\mathcal{O}(2^q \sqrt{M})$ repetitions of Grover-like operation (or Hamiltonian simulation) on the encoding unitary and prepare the (approximate) probing state $|\Upsilon(q)\rangle$ (defined as in Eq. (4.1)) whose phase at each point $|\mathbf{x}\rangle = |x_1\rangle \cdots |x_M\rangle$ is depicted in the lower right illustration. The nonlinear components come from the Grover-like repetition, as detailed in Sec. 4.4. From the probing state $|\Upsilon(q)\rangle$, we read out *zoomed-in* values $\{2^q \langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle\}$ simultaneously by quantum gradient estimation protocol with a fixed measurement precision. Using the measurement results, we update the temporal estimates $\tilde{u}_j^{(q)}$ to $\tilde{u}_j^{(q+1)}$. By repeating this procedure for $q = 0, 1, \dots, \lceil \log_2(1/\varepsilon) \rceil$, this algorithm estimates the M observables simultaneously within at most root MSE ε . The algorithm uses totally $\mathcal{O}(\varepsilon^{-1} \sqrt{M} \log M)$ queries to the state preparation unitary U_ψ that indicates the scaling of Heisenberg limit, that is, the inverse of root MSE ε , together with the nearly squared root dependence on the number of observables M , which comes from the uniform singular value amplification process in the encoding of $\{\langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle\}$. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

A key insight for enhancement can be borrowed from the history of phase estimation algorithm [99]. It has been well-recognized that the textbook style of the QPE algorithm [128] encounters two major bottlenecks that prevents the algorithm from practical benefits; the large ancilla consumption of $\mathcal{O}(\log(1/\varepsilon))$ and the poor query complexity scaling with the target root MSE ε , i.e., only obeying the SQL $\mathcal{O}(1/\varepsilon^2)$ [78, 160, 92]. While the former can be addressed by the iterative phase estimation [99], it was not until the work by Higgins *et al.* [78] that both issues were overcome by utilizing an adaptive measurement scheme to achieve the HL with a constant number of ancilla. Although a follow-up work has shown that HL scaling can be achieved even without relying on the adaptive scheme (but using an iterative procedure), there is in practice, an increase of the estimation variance by a constant factor [77]. On the other hand, if the large ancilla consumption is acceptable, it is known that the QPE algorithm with an entangled ancillary state, instead of a uniform superposition state, can achieve the HL [118, 160, 92, 10]. From the above observations, we argue that adaptive strategy is crucial for resource-efficient implementation of quantum estimation that saturates the ultimate scaling limited by purely fundamental principles of quantum physics.

4.1.2 Summary of results

Driven by such a situation, we make a significant contribution to multiple observables estimation of number M with root MSE of ε . Concretely, we explicitly construct an adaptive estimation algorithm as summarized as Algorithm 1, or graphically as Fig. 4.1 which satisfies the following features (each corresponds to Theorem 10, Theorem 11, Lemmas 8, and 9, respectively):

- *Heisenberg-limited scaling $\mathcal{O}(1/\varepsilon)$ and sublinear scaling regarding M .* The proposed observable estimation achieves the pure HL scaling with the root MSE ε as

$$\mathcal{O}(\varepsilon^{-1}\sqrt{M}\log M)$$

in query to the state preparation and its inverse. Also, this scaling indicates a nearly quadratic improvement regarding M , compared to using the (modified) QAE algorithm [21, 100, 139, 149, 167, 162] M times. For a particular set of observables e.g., projectors onto each basis state, we can also achieve the query complexity

$$\mathcal{O}(\varepsilon^{-1}\log M)$$

i.e., the logarithmic dependence of M combined with the HL scaling.

- *Constant space overhead.* The quantum circuits in the proposed scheme require at most additional $\mathcal{O}(M)$ qubits, which is independent from the target root MSE ε unlike non-iterative methods [86, 159] that use additional $\mathcal{O}(M\log(1/\varepsilon))$ qubits.
- *Robustness in high-precision regime.* When highly precise estimates (i.e., $\varepsilon \ll 1$) are required, the quantum circuits in our method have at most $\mathcal{O}(\log(1/\varepsilon))$ parameterized gates for QSVT [57], which can be tuned by $\mathcal{O}(\text{polylog}(1/\varepsilon))$ classical computation, while the previous methods [86, 159] require to tune $\mathcal{O}(1/\varepsilon)$ gates with $\mathcal{O}(\text{poly}(1/\varepsilon))$ classical computation. This significant reduction in classical computation allows us to avoid the numerical instability problem [70, 30, 45, 177, 124, 175] that spoils the quantum enhancement.

The first feature is a mathematical guarantee of estimation performance, and the other features highlight ease of practical implementation of our method. We also mention as another important feature that the proposed algorithm is the first adaptive extension of the QGE algorithm [91, 56], to the best of our knowledge. Therefore, our method would be applicable to various tasks in which the QGE algorithm offers quantum speedup, such as quantum machine learning [56, 38, 37], financial risk calculation [147], and molecular force calculation [129].

Let us briefly describe the abstract structure of the algorithm and where the features stem from. In our algorithm, we sample from an $\mathcal{O}(M + \log_2 d)$ -qubit circuit with an alternating sequence of a global interaction to encode the expectation values of target observables and a controlled rotation over the target $\log_2 d$ -qubit system (and some ancillary system). The global interaction for observables and the total circuit length are adjusted in an adaptive manner, so that the expectation values can be read out in high resolution with the space overhead kept to be independent from the target precision ε . Then, we classically process the samples and use the processing results to construct the quantum circuit at the next step, in a similar way as the HL QPE algorithms [77, 97, 12, 48]. The sublinear scaling in the number M of observables comes from the uniform singular value amplification process [114, 57] embedded in the global interaction. This reason for the speedup on M is the same as the previous method [159].

We remark that the classical computation mentioned in the third point is required for the circuit construction, especially for QSP [117, 116]. In the framework of QSP (also QSVT), we need to tune the parameterized gates classically for a desired polynomial. Although finding this parameter for a degree- n polynomial can be achieved in $\mathcal{O}(\text{poly}(n))$ classical computation time, it exhibits numerical instability for large n ; this instability leads to undesired algorithmic errors in the resulting quantum circuit, posing a central challenge in the practical application of QSVT [70]. To resolve this, various optimization techniques have been investigated [70, 30, 45, 177, 124, 175], and currently, they require 10^2 – 10^4 seconds to tune $n \sim 10^4$ parameters. In the previous method for multiple observables estimation [159], the number of circuit parameters is given by $n = \tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$, leading to a large runtime in classical computation. For instance, we can naively extrapolate that classical runtime for $M = 10^4$ and $\varepsilon = 10^{-4}$ already requires long 10^4 – 10^6 seconds. In contrast, our method requires to

tune only $\mathcal{O}(\sqrt{M} \log(M/\varepsilon))$ parameters under a certain condition, by partially using a special polynomial whose parameters can be analytically determined. This exponential improvement in classical computation time, regarding estimation precision $1/\varepsilon$, significantly lowers the barrier of the practical implementation of our method.

Finally, we provide numerical simulations to compare the number of total queries and ancilla qubits in our adaptive method and the previous non-iterative method [86]. The results clearly show the quantitative advantages of our method in terms of query and space complexity. In particular, the previous method indicates the HL scaling $\mathcal{O}(1/\varepsilon)$ with logarithmic factors on $1/\varepsilon$ regarding the root MSE ε , while our method has no ε dependence except for $\mathcal{O}(1/\varepsilon)$ in the total queries.

4.2 Problem setup

We consider estimating quantum expectation values of given d -dimensional M Hermitian operators $\{O_j\}_{j=1}^M$ with the spectral norm $\|O_j\| \leq 1$, regarding a quantum state $|\psi\rangle$. Here, d is a power of 2 and N denotes the number of target system qubits i.e., $N \equiv \log_2 d$. The target state $|\psi\rangle$ is assumed to be prepared by a state preparation oracle $U_\psi : |\mathbf{0}\rangle \mapsto |\psi\rangle$ and an initial state $|\mathbf{0}\rangle := |\mathbf{0}\rangle^{\otimes \log_2 d}$, and we assume oracular access to U_ψ and $U_\psi^\dagger$. The observables are assumed to be accessed by some block-encoded unitaries over the d -dimensional system and some ancilla system. That is, for each observable O_j , we assume access to a unitary gate B_j whose top-left block matrix is O_j ; see Sec. 3.3.1 for the definition of this block-encoding. We remark that we only assume access to B_j and additional information on properties of the encoded O_j e.g., spectrum and eigenstates is not available. In this setup, our goal is to *efficiently* obtain samples from estimators for the target values $\langle O_j \rangle := \langle \psi | O_j | \psi \rangle$ within the root MSE ε . In particular, we aim to simultaneously achieve the HL scaling regarding ε and the sublinear scaling regarding M , for estimating all $\langle O_j \rangle$.

The performance of quantum algorithms for this task is usually quantified by the total number of queries to the state preparation U_ψ and $U_\psi^\dagger$. This is because U_ψ has complexity that usually scales with the system size, and as a result, the state preparation U_ψ is the most dominant factor in the total execution time for various settings. Under the natural assumption that $U_\psi^\dagger$ has the same cost as U_ψ , it is crucial to design an estimation algorithm that minimizes the statistical uncertainty using a limited number of queries to U_ψ and $U_\psi^\dagger$.

4.3 Adaptive gradient estimation for multiple observables

To address the problem stated in Sec. 4.2, we propose an adaptive extension of the gradient estimation algorithm (see Sec. 3.1.3). In the following, we first describe the adaptive gradient estimation algorithm for multiple observables estimation in Sec. 4.3.1. Then, the performance guarantee of the algorithm is provided in Sec. 4.3.2.

4.3.1 Proposed adaptive algorithm

Now we introduce our algorithm for estimating the expectation values of multiple observables that uses the gradient estimation as a subroutine. The outline of the algorithm is provided in Algorithm 1, and also its high-level graphical overview is shown in Fig. 4.1. In the following, we first introduce the foundation of Algorithm 1, with a particular focus on the quantum part (i.e., the quantum state $|\Upsilon(q)\rangle$ in Step 3). Then, we prove our main theorem (Theorem 10) to guarantee that Algorithm 1 achieves both the HL scaling $\mathcal{O}(1/\varepsilon)$ and the sublinear scaling on M for the total query complexity regarding the use of U_ψ and $U_\psi^\dagger$. Sec. 4.5 gives the complete proof of theorems together with showing concrete implementation method for the algorithm.

The key idea of the proposed adaptive algorithm is to prepare the following *probing* state $|\Upsilon(q)\rangle$ at each iteration step q , from which the approximated expectation values can be extracted via the gradient estimation algorithm described in Sec. 3.1.3:

$$|\Upsilon(q)\rangle := \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^q \pi^{-1} \langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle} |\mathbf{x}\rangle, \quad (4.1)$$

Algorithm 1 Adaptive gradient estimation for multiple observables

Input: $\log_2 d$ -qubit state preparation unitary U_ψ and $U_\psi^\dagger$; observables $\{O_j\}_{j=1}^M$ with the spectral norm $\|O_j\| \leq 1$ such that $M > \mathcal{O}(\log d)$ holds; confidence parameter $c \in (0, 3/8(1 + \pi)^2]$; target root mean squared error (MSE) $\varepsilon \in (0, 1)$.

Output: A sample $(\tilde{u}_1, \dots, \tilde{u}_M)$ from an estimator $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_M)$ whose j -th element estimates $\langle \psi | O_j | \psi \rangle$ within the MSE ε^2 as

$$\max_{j=1,2,\dots,M} \mathbb{E} \left[(\hat{u}_j - \langle \psi | O_j | \psi \rangle)^2 \right] \leq \varepsilon^2$$

- 1: Set a fixed precision parameter $p := 3$ and temporal estimates $\tilde{u}_j^{(0)} := 0$ for all j .
- 2: **for** $q = 0, 1, \dots, q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$ **do**
- 3: Measure $\mathcal{O}(\log M / \delta^{(q)})$ approximate copies of the probing state

$$|\Upsilon(q)\rangle := \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^q \pi^{-1} \langle O_j - \tilde{u}_j^{(q)} | \mathbf{1} \rangle} |\mathbf{x}\rangle$$

after p -qubit inverse quantum Fourier transformations. Here, $\delta^{(q)} := c/8^{q_{\max}-q}$ is a failure probability.

- 4: Set the coordinate-wise median of the measurement outputs as $g_j^{(q)}$.
- 5: Update $\tilde{u}_j^{(q+1)} := \tilde{u}_j^{(q)} + \pi 2^{-q} g_j^{(q)}$
- 6: Truncate $\tilde{u}_j^{(q+1)}$ in $[-1, 1]$
- 7: **end for**
- 8: **return** final estimates $\tilde{u}_j := \tilde{u}_j^{(q_{\max}+1)}$

where $\mathbf{1}$ denotes the identity operator. Also, p denotes a fixed precision parameter and G_p^M is defined as in Eq. (3.20). The quantities $\tilde{u}_j^{(q)} \in [-1, 1]$ in the probing state Eq. (4.1) are temporal estimated values for $\langle O_j \rangle$ at the iteration step q . Later we will show that $p = 3$ is sufficient for our algorithm to successfully work. We remark that the probing state $|\Upsilon(q)\rangle$ is equivalent to the quantum state Eq. (3.24) with gradient $g_j = 2^q \pi^{-1} \langle O_j - \tilde{u}_j^{(q)} | \mathbf{1} \rangle$.

The probing state $|\Upsilon(q)\rangle$ can be prepared by first initializing the pM -qubit probe system, the $\log_2 d$ -qubit target system, and some ancilla systems to encode the observables. Then, applying a q -dependent interaction unitary on the whole system, we approximately obtain the probing state. The query complexity regarding U_ψ and $U_\psi^\dagger$ of this state preparation process is given as follows.

Proposition 9 (Probing-state preparation). *Suppose that we have access to block-encoded observables $\{O_j\}_{j=1}^M$ in d dimension via some unitaries, a $\log_2 d$ -qubit state preparation U_ψ , and its inverse $U_\psi^\dagger$, such that $M > \mathcal{O}(\log d)$ holds. Then, we can prepare the probing state $|\Upsilon(q)\rangle$, up to $1/12$ Euclidean distance error, for (any) integer $q \geq 0$ and $\tilde{u}_j^{(q)} \in [-1, 1]$, using $\mathcal{O}(2^q \sqrt{M \log d})$ queries to U_ψ and $U_\psi^\dagger$ in total.*

In the next section, we describe this probing-state preparation in detail. In particular, this proposition directly follows from Lemmas 8 and 9 shown in the next section, both of which give methods to prepare the probing state $|\Upsilon(q)\rangle$: one is based on the Hamiltonian simulation protocol (Lemma 8), and the other is based on the Grover-like repetition (Lemma 9).

Suppose we have the probing state $|\Upsilon(q)\rangle$ for Step 3 at the q th iteration, with a temporal estimate $\tilde{u}_j^{(q)}$ for $\langle O_j \rangle$. Then, on the phase of $|\mathbf{x}\rangle$ in $|\Upsilon(q)\rangle$, the estimation error $\langle O_j - \tilde{u}_j^{(q)} | \mathbf{1} \rangle$ is amplified by the factor of 2^q ; this means that when $\tilde{u}_j^{(q)}$ matches $\langle O_j \rangle$ in the first q binary (fraction) digits as

$$\left| \langle O_j \rangle - \tilde{u}_j^{(q)} \right| \leq \frac{1}{2^q}, \quad (4.2)$$

the 2^q -fold amplification *zooms in* on the significant digits of $\langle O_j \rangle - \tilde{u}_j^{(q)}$ to shift the values toward upper digits. As proved in Ref. [56], the gradient estimation for $|\Upsilon(q)\rangle$ can simultaneously extract these zoomed-in values $2^q \pi^{-1} \langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle$ or equivalently the gradient of the linear function on the phase (with $1/\pi$):

$$f(\mathbf{x}) := \sum_j x_j 2^q \langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle, \quad (4.3)$$

with an additive error specified by the precision parameter p and a certain success probability (see Lemma 2). Note that, in our case, the gradient estimation algorithm is simply to perform the computational basis measurement (with rewriting $|\mu\rangle$ as $|x\rangle$) on the probing state after applying a slightly modified version of p -qubit inverse quantum Fourier transformation, which is defined in Eq. (3.25).

In Step 4 of Algorithm 1, we use the measurement outputs $\mathbf{k}^{(1)}, \mathbf{k}^{(2)}, \dots, \mathbf{k}^{(\mathcal{O}(\log M/\delta^{(q)}))}$ obtained in Step 3 to construct the coordinate-wise median $g_j^{(q)}$; that is, $g_j^{(q)}$ is defined as the middle value separating the greater and lesser halves of $\{k_j^{(1)}, k_j^{(2)}, \dots\}$. Then, under the condition of Eq. (4.2), we can prove that the median $g_j^{(q)}$ satisfies

$$\left| g_j^{(q)} - \frac{2^q \langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle}{\pi} \right| \leq \frac{1}{2\pi}, \quad (4.4)$$

for all $j = 1, 2, \dots, M$ with success probability bigger than $1 - \delta^{(q)}$; later we will carefully choose the probability $\delta^{(q)}$ in order that a final estimator has the minimal statistical uncertainty. Using the successfully obtained $g_j^{(q)}$ satisfying Eq. (4.4), we update the temporal estimate in Step 5 as

$$\tilde{u}_j^{(q+1)} := \tilde{u}_j^{(q)} + \pi 2^{-q} g_j^{(q)}.$$

Then, it is straightforward to prove

$$\left| \langle O_j \rangle - \tilde{u}_j^{(q+1)} \right| \leq \frac{1}{2^{q+1}},$$

for all j . Thus, using the outcomes from the gradient estimation algorithm at each step, we can iteratively determine one binary fraction digit of the target value $\langle O_j \rangle$ by updating $\tilde{u}_j^{(q)}$ to $\tilde{u}_j^{(q+1)}$. Repeating this iteration step for $q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$ times, we obtain estimates for $\{\langle O_j \rangle\}_j$ within the root MSE ε .

4.3.2 Estimation performance and query complexity

Now, we provide a performance guarantee, namely the HL scaling of the multiple observables estimation precision in terms of MSE, for the proposed Algorithm 1. While a more detailed proof is given in Sec. 4.5.2, here we present the Theorem with a sketch of its proof in the following:

Theorem 10 (Heisenberg-limited multiple observables estimation). *Let $\varepsilon \in (0, 1)$ be a target precision. For given M observables $\{O_j\}_{j=1}^M$ ($\|O_j\| \leq 1 \forall j$) and a state preparation oracle U_ψ , there exists a quantum algorithm that outputs a sample from estimators $\{\hat{u}_j\}_{j=1}^M$ for $\{\langle O_j \rangle\}$ satisfying*

$$\max_{j=1,2,\dots,M} \text{MSE}[\hat{u}_j] \leq \varepsilon^2, \quad (4.5)$$

using $\mathcal{O}(\varepsilon^{-1} \sqrt{M} \log M)$ queries to the state preparation oracles U_ψ and $U_\psi^\dagger$ in total. Here, the mean squared error of an estimator $\hat{u}_j$ is defined as

$$\text{MSE}[\hat{u}_j] := \mathbb{E}[(\hat{u}_j - \langle O_j \rangle)^2].$$

Sketch of the proof. While the outline of the proof is inspired from the Heisenberg-limited phase estimation or its application [77, 97, 48, 47], it is nontrivial for the case of adaptive gradient estimation to show that the condition given in Eq. (4.2) is satisfied for all j . Note that, if this condition holds and the

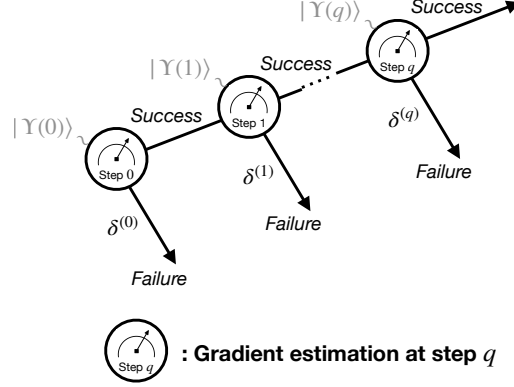


Figure 4.2: Probability tree diagram in Algorithm 1. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

precision parameter p is taken as $p = 3$, then a single shot measurement result $\mathbf{k} := (k_1, \dots, k_M) \in G_p^M$ in Step 3 of Algorithm 1 follows

$$\Pr \left[\left| k_j - \frac{2^q (\langle O_j \rangle - \tilde{u}_j^{(q)})}{\pi} \right| > \frac{1}{2\pi} \right] < \frac{1}{3}, \quad (4.6)$$

for every $j = 1, 2, \dots, M$. The derivation of this inequality is based on our numerical finding; see Sec. 4.5.2. We remark that the failure probability represented by the left hand side of Eq. (4.6) can be exponentially suppressed to $\delta^{(q)}/M$ by using the coordinate-wise median $g_j^{(q)}$ of the measurement results over $\mathcal{O}(\log M/\delta^{(q)})$ copies of the (approximate) probing state, instead of the single shot result k_j . This repetition is the origin of the $\log M$ term in the total query complexity.

Let us first discuss how to assure that the condition (4.2) is satisfied. In the first iteration step $q = 0$, the condition Eq. (4.2) is trivially satisfied, and the gradient estimation can yield $1/2\pi$ -close estimates of the target quantities for all j . This probability is at least $1 - \delta^{(0)}$ due to the union bound. As for the iteration step $q \geq 1$, if all of the previous steps succeed in the gradient estimation, we can show that Eq. (4.2) holds for all j at the iteration step q , as discussed in the previous subsection. On the other hand, if the gradient estimation fails at the iteration step q (while all processes have been successfully executed up to the $(q - 1)$ th step), the condition Eq. (4.2) is not satisfied; as a result, outputs of the gradient estimation may not improve the temporal estimate in the subsequent processes. However, in this case, it can be shown that the additive error $|\tilde{u}_j - \langle O_j \rangle|$ of the final estimate $\tilde{u}_j := \tilde{u}_j^{(q_{\max}+1)}$ is at most $(1 + \pi)/2^q$ because the outputs of gradient estimation are always in $[-1/2, 1/2]$ from the definition of G_p^M . From the above analysis, we can bound the additive error of the final estimate for all branches in Fig. 4.2.

The failure probability $\delta^{(q)}$ should be very small at the beginning of iteration steps q because the error of the top fraction bits has a significant impact on the MSE of the final estimator. From the analysis in Ref. [47], the choice of $\delta^{(q)} := c/8^{q_{\max}-q}$ ($c \in (0, 3/8(1 + \pi)^2)$) is sufficient, and then the MSE of the final estimator $\hat{u}_j$ (its realization is $\tilde{u}_j$) is calculated as at most ε^2 for all $j = 1, 2, \dots, M$, from the tree diagram Fig. 4.2 and the above error evaluation in each branch.

To derive the query complexity for the state preparation U_ψ , we use the result in Proposition 9. In the iteration step q , preparing $\mathcal{O}(\log M/\delta^{(q)})$ copies of the probing state Eq. (4.1) uses $\mathcal{O}(2^q \sqrt{M \log d \log(M/\delta^{(q)})})$ queries to the state preparation unitary U_ψ and its inverse. Thus, the summation of the queries over $q = 0, 1, \dots, q_{\max}$ ($q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$) is directly calculated as $\mathcal{O}(\varepsilon^{-1} \sqrt{M \log d \log M})$, which completes the proof of Theorem 10. $\square$

In Theorem 10, the scaling $1/\varepsilon$ of queries to U_ψ with respect to the root MSE ε achieves the same scaling to the Heisenberg limit (HL) in the gradient estimation, which is derived in Sec. 3.1.3. We remark that, to the best of our knowledge, this is the first derivation of the HL in gradient estimation

in terms of MSE, while the previous work [56] shows a similar lower bound (additionally, including $\sqrt{M}$ dependency) on queries to an oracle for a target function $f(\mathbf{x})$ on $\mathbf{x} \in G_p^M$, in another metric of statistical uncertainty (i.e., a confidence interval). Also, in the stringent measure of statistical uncertainty, the root MSE ε [16], all the existing works [86, 159] in multiple observables estimation only prove the nearly HL scaling that includes logarithmic terms on ε such as $\log(M/\varepsilon)$.

In addition to the Heisenberg-limited scaling regarding ε , our protocol has the nearly squared root dependence regarding the total number of observables M . This is a nearly quadratic improvement regarding M compared to the standard method (i.e., quantum amplitude estimation) of estimating the expectation values of observables.

Furthermore, when we focus on another metric of statistical uncertainty — a confidence interval comprised of estimators with an additive error $\pm\epsilon_{\text{add}}$ for some confidence level, which is considered in Refs. [86, 159], the query complexity of our method is essentially optimal in terms of M and ϵ_{add} as well as these previous methods. Here, the query lower bound in multiple observables estimation in terms of a confidence interval is proved in Ref. [86] as follows.

Lemma 7 (Corollary 3 in Ref. [86]). *Let M be a positive integer power of 2 and let $\epsilon_{\text{add}} \in (0, 1/3\sqrt{M})$. Let $\mathcal{A}$ be any algorithm that takes as an input an arbitrary set of M observables $\{O_j\}_{j=1}^M$. Suppose that, for every quantum state $|\psi\rangle$, accessed via a state preparation oracle U_ψ , $\mathcal{A}$ outputs estimates of each $\langle\psi|O_j|\psi\rangle$ to within additive error ϵ_{add} with high probability at least $2/3$. Then, there exists a set of observables $\{O_j\}_{j=1}^M$ such that $\mathcal{A}$ applied to $\{O_j\}_{j=1}^M$ must use $\Omega(\epsilon_{\text{add}}^{-1}\sqrt{M})$ queries to U_ψ .*

For a given ε , Algorithm 1 also satisfies

$$\Pr[\max_j |\hat{u}_j - \langle O_j \rangle| \leq \varepsilon/2] \geq 1 - 8c/7, \quad (4.7)$$

by using $\mathcal{O}(\varepsilon^{-1}\sqrt{M}\log(M/c))$ queries to U_ψ and $U_\psi^\dagger$. This can be confirmed by calculating the probability of the all-success branch in Fig. 4.2. Consequently, we establish from the Lemma 7 that our protocol achieves the worst-case query optimality (up to the $\log M$ correction) in the high-precision regime $\epsilon_{\text{add}} \in (0, 1/3\sqrt{M})$.

For a particular set of observables, Ref. [159] shows an improvement on the dependence of M in the query complexity. A similar improvement can be attained while retaining the HL scaling $\mathcal{O}(1/\varepsilon)$ in MSE.

Theorem 11 (Improved M -dependence). *In Theorem 10, we can improve the total number of queries to U_ψ and $U_\psi^\dagger$ in terms of M , while keeping the HL scaling $\mathcal{O}(\varepsilon^{-1})$ for root MSE ε , to*

$$\mathcal{O}\left(\varepsilon^{-1}\sqrt{\mathcal{B}_M}\log M\right) \quad (4.8)$$

for any known $\mathcal{B}_M$ satisfying

$$\left\| \sum_{j=1}^M O_j^2 \right\| \leq \mathcal{B}_M \leq M. \quad (4.9)$$

The proof of this theorem is essentially the same as that of Theorem 10 except for the state preparation (more precisely, we use Algorithm 1 with a modified Step 3); see Sec. 4.5.2. This theorem recovers the query complexity in Theorem 10 when we only know $\|O_j\| \leq 1$ on observables, leading to $\mathcal{B}_M = M$. For a particular set of observables, $\mathcal{B}_M$ can be taken as e.g., $\mathcal{B}_M = \mathcal{O}(1)$ for $O_j = |j\rangle\langle j|$ in multidimensional amplitude estimation [158, 159], so the logarithmic dependence of M together with the HL scaling $\mathcal{O}(1/\varepsilon)$ is attained in such cases.

Finally, we remark that the preparation and measurement for multiple copies of the probing state in Step 3 of Algorithm 1 can be performed in parallel if we are allowed to use multiple quantum computers, which may be practically important to reduce the total execution time.

4.4 State preparation for adaptive gradient estimation

In the performance guarantee (Theorem 10) provided in Sec. 4.3.1, we have exploited Proposition 9 that states the query complexity to prepare the probing state defined in Eq. (4.1). In this section, we aim to show Proposition 9 by providing two quantum algorithms to prepare the probing state Eq. (4.1),

using the state preparation U_ψ , $U_\psi^\dagger$, and the unitary gates $\{B_j\}_{j=1}^M$ that encode the target observables $\{O_j\}_{j=1}^M$ with the help of some ancilla system. The two methods are summarized in Lemmas 8 and 9; Lemma 8 provides a method based on Hamiltonian simulation protocol, and Lemma 9 provides a method based on Grover-like repetition. Both Lemmas 8 and 9 prove that the total space complexity to prepare the probing state $|\Upsilon(q)\rangle$ is $\mathcal{O}(M + \log_2 d)$. In addition, Lemma 9 shows an exponential improvement in the classical computation time for constructing explicit quantum circuits under a certain condition, compared to Lemma 8.

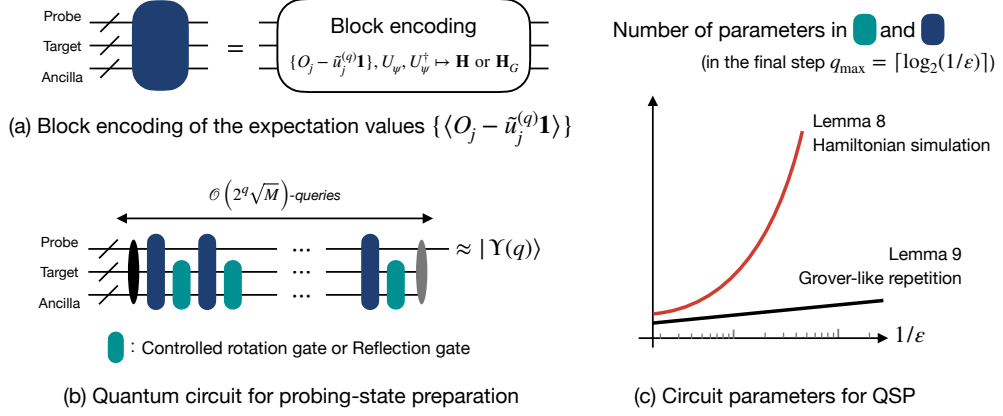


Figure 4.3: Preparation of the probing state $|\Upsilon(q)\rangle$ at the q -th iteration. The blue unitary gate in (a) and (b) denotes a block-encoding of the Hamiltonian $\mathbf{H}$ or $\mathbf{H}_G$ in Eq. (4.10) or (4.21), respectively, and the green unitary gate in (b) is a controlled rotation or reflection gate. The blue and green unitary gates contain parameterized gates for QSP, and the total number of circuit parameters scales as in (c) with respect to the target root MSE ϵ . As the graph in (c) indicates, the method in Lemma 9 has a much smaller number of circuit parameters for QSP compared to that of Lemma 8. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

Before proceeding to the proof of Lemmas 8 and 9, we first provide an overview of our method to prepare $|\Upsilon(q)\rangle$ in Fig. 4.3. As seen in the figure, the proposed two methods to prepare $|\Upsilon(q)\rangle$ have a similar structure: Fig. 4.3(a) shows the block encoding of a Hamiltonian that encodes the expectation values of observables $\langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle$ and Fig. 4.3(b) shows alternating applications of the block encoding and a processing operation with tuned parameters, which is based on Lemma 5 or Lemma 6. In the proof of Lemmas 8 and 9, we depict the circuit for (a) and (b), respectively; then we evaluate the approximation error between the final state of the circuit and the probing state $|\Upsilon(q)\rangle$. More detailed proofs and explicit quantum circuit diagrams for these lemmas are provided in Sec. 4.5.

Hamiltonian simulation

Let us first present an informal lemma (its formal version is Lemma 11 in Sec. 4.5.1) regarding the complexity of probing-state preparation when we utilize the Hamiltonian simulation based on QSP:

Lemma 8 (Informal). *Suppose that we have access to block-encoded observables $\{O_j\}_{j=1}^M$ in d dimension, a $\log_2 d$ -qubit state preparation U_ψ , and its inverse $U_\psi^\dagger$, such that $M > \mathcal{O}(\log d)$. Then, we can prepare the probing state $|\Upsilon(q)\rangle$ for any integer $q \geq 0$ and $\tilde{u}_j^{(q)} \in [-1, 1]$ up to $1/12$ Euclidean distance error, using an*

$$\mathcal{O}(M + \log_2 d)\text{-qubit}$$

circuit regardless of q . Furthermore, each quantum circuit with q requires

$$\mathcal{O}(\text{poly}(2^q \sqrt{M \log d}) + \text{poly}(\sqrt{M}(q + \log M)))$$

classical computation for finding circuit parameters, and it consists of

$$\mathcal{O}(2^q \sqrt{M \log d}) \text{ uses of } U_\psi \text{ and } U_\psi^\dagger,$$

$\mathcal{O}(2^q M(q + \log M))$ uses of unitary gates for block-encoded observables, and $\mathcal{O}(2^q M(q + \log M) \log dM)$ uses of single-qubit and two-qubit gates.

Sketch of the proof. To encode the ideal phase $f(\mathbf{x})$ in Eq. (4.3), we construct the Hamiltonian $\mathbf{H}$ with the observables $\{O_j\}$ as

$$\mathbf{H} := \sum_{\mathbf{x} \in G_p^M} \tilde{f}(\mathbf{x}) |\mathbf{x}\rangle \langle \mathbf{x}|, \quad (4.10)$$

where the target observables are approximately encoded in the eigenvalues $\tilde{f}(\mathbf{x})$ as

$$\tilde{f}(\mathbf{x}) \approx \frac{1}{\sigma} \sum_{j=1}^M x_j \langle \tilde{O}_j^{(q)} \rangle, \quad \sigma = \mathcal{O}(\sqrt{M \log d}), \quad (4.11)$$

Here, σ denotes the rescaling factor of the Hamiltonian $\mathbf{H}$ such that $\mathbf{H}$ can be encoded in a unitary operator. Also, we defined $\tilde{O}_j^{(q)} := (O_j - \tilde{u}_j^{(q)} \mathbf{1})/2$ for the identity $\mathbf{1}$. It is crucial for the total query complexity of U_ψ to construct the block encoding of $\mathbf{H}$ with the $\mathcal{O}(\sqrt{M})$ rescaling factor, while naive block encoding for $\mathbf{H}$ has $\mathcal{O}(M)$ rescaling factor, as mentioned below.

Setting aside the details for now, we suppose that we have a (perfect) block encoding of the Hamiltonian $\mathbf{H}$ that acts on the pM -qubit probe system, $\log_2 d$ -qubit target system, and ancilla systems (specified below), as illustrated in Fig. 4.3. Then, the optimal Hamiltonian simulation protocol Lemma 6 yields a quantum circuit W for an ϵ' -precise block encoding of time evolution operator

$$e^{i\mathbf{H}t} = \sum_{\mathbf{x} \in G_p^M} e^{i\tilde{f}(\mathbf{x})t} |\mathbf{x}\rangle \langle \mathbf{x}| \quad (4.12)$$

with time $t > 0$, using $\mathcal{O}(t + \log(1/\epsilon'))$ queries to the block encoding of $\mathbf{H}$. Note that in the case $t = 2\sigma t'$ for some positive integer t' , the resulting time evolution operator approximates t' times applications of the *phase oracle* for an affine linear function $f(\mathbf{x}) = \sum_j x_j \langle O_j \rangle$ in the theory of gradient estimation. Then, applying W for

$$t := 2^{p+q+2}\sigma$$

to the uniform superposition state $|+\rangle^{\otimes pM}$ in the probe system and the initial state $|\mathbf{0}\rangle$ in the ancilla-target systems, we can approximately prepare the probing state:

$$W |+\rangle^{\otimes pM} |\mathbf{0}\rangle \approx |\Upsilon(q)\rangle |\mathbf{0}\rangle. \quad (4.13)$$

From the triangle inequality, this approximation error is given by the sum of ϵ' (more precisely, $\epsilon' + \sqrt{2\epsilon'}$ in Euclidean distance) from the transformation $\mathbf{H} \mapsto e^{i\mathbf{H}t}$ and the error from the observable encoding in Eq. (4.11), that is,

$$\epsilon' + \sqrt{2\epsilon'} + \left\| \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{i\tilde{f}(\mathbf{x})t} |\mathbf{x}\rangle - |\Upsilon(q)\rangle \right\|. \quad (4.14)$$

To quantify the error of the third term in Eq. (4.14) (or the error of Eq. (4.11)), we here construct the block encoding of the Hamiltonian $\mathbf{H}$ from the state preparation U_ψ ($U_\psi^\dagger$) and unitary gates $\{B_j\}$ that are a -block-encodings of observables $\{O_j\}$ for some $a \in \mathbb{N}$. The block encoding of $\mathbf{H}$ consists of (i) the LCU method, (ii) adding the pM -qubit control to the LCU operations, (iii) the singular value amplification, and (iv) multiplying the conjugation of U_ψ . Let us describe these procedures step by step in the following.

First, we utilize the LCU method with the controlled version of each B_j to construct a quantum circuit $U^{(\mathbf{x})}$ for a block encoding of

$$\frac{1}{M} \sum_{j=1}^M x_j \tilde{O}_j^{(q)}$$

(with the rescaling factor M , instead of σ) for a given $\mathbf{x} \in G_p^M$. The LCU requires additional $\mathcal{O}(\log M)$ qubits to encode the coefficients $\{x_j/M\}$. In step (ii), using controlled versions of $U^{(\mathbf{x})}$, we can obtain a quantum circuit for

$$U'_{\text{SEL}} := \sum_{\mathbf{x} \in G_p^M} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U^{(\mathbf{x})}, \quad (4.15)$$

which can be implemented with at most $\mathcal{O}(pM \log(1/\delta))$ depth regarding the implementation error δ of the coefficients $\{x_j/M\}$; see Remark 1 in Sec. 4.5.

To understand why we must have step (iii), it is informative to observe the action realized in step (iv). Namely, using U'_{SEL} , U_ψ , and its conjugation $U_\psi^\dagger$, we obtain

$$\left(\mathbf{1} \otimes U_\psi^\dagger\right) \cdot U'_{\text{SEL}} \cdot \left(\mathbf{1} \otimes U_\psi\right), \quad (4.16)$$

which is a block encoding of the following Hamiltonian:

$$\mathbf{H}' := \sum_{\mathbf{x}} \left(\sum_j x_j \cdot \frac{\langle \tilde{O}_j^{(q)} \rangle}{M} \right) |\mathbf{x}\rangle \langle \mathbf{x}|.$$

This Hamiltonian $\mathbf{H}'$ has the normalization factor M that is quadratically larger than $\sigma = \mathcal{O}(\sqrt{M})$ in $\mathbf{H}$. In estimating the expectation value $\langle \tilde{O}_j^{(q)} \rangle$ via gradient estimation, we need to amplify $\mathbf{H}'$ by the factor M to prepare Eq. (4.1). This can be achieved by using a time evolution $e^{i\mathbf{H}'t}$ of $t = 2^{p+q+2}M$, but the linear dependence on M in the evolution time t results in no speedup for the total queries to U_ψ regarding the number M of observables.

To obtain the quadratic speedup regarding M , the method in Ref. [159] uses the singular value amplification Lemma 4. Importantly, this amplification can be performed with no use of U_ψ . From the random matrix series inequality, we can show that for a large part of G_p^M (more precisely, for a subset $F \subset G_p^M$ such that $|F| \geq (1 - \delta')|G_p^M|$ for any $\delta' > 0$), the condition of the amplification is satisfied for $\gamma = M/\sigma = \mathcal{O}(\sqrt{M})$, as well as the analysis in Ref. [159]. Therefore, we can amplify the block encoding $U^{(\mathbf{x})}$ by γ for such $\mathbf{x} \in F$; as a result, we have a quantum circuit for

$$U_{\text{obs}} := \sum_{\mathbf{x} \in G_p^M} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U_{\text{obs}}^{(\mathbf{x})}, \quad (4.17)$$

where $U_{\text{obs}}^{(\mathbf{x})}$ is an ϵ'' -precise block encoding of the Hamiltonian $\sigma^{-1} \sum_j x_j \tilde{O}_j^{(q)}$ if $\mathbf{x} \in F$. By multiplying U_ψ and $U_\psi^\dagger$ into U_{obs} as well as Eq. (4.16), we finally arrive at the block encoding of the Hamiltonian $\mathbf{H}$. We note that the amplification is valid when $\sigma = \mathcal{O}(\sqrt{M \log_2 d})$ is smaller than M , and this is satisfied by the condition $M > \mathcal{O}(\log d)$.

The main contribution to the space complexity comes from the steps (i) the LCU method and (ii) adding the pM -qubit control to $U^{(\mathbf{x})}$. The other processes (iii), (iv), and the Hamiltonian simulation protocol $\mathbf{H} \mapsto e^{i\mathbf{H}t}$ introduce constant or no ancilla qubits. The LCU method (i) introduces $\mathcal{O}(\log M)$ -qubit registers, and thus the number of qubits of U'_{SEL} scales as $pM + \log M + \log d$. Recalling that the precision parameter p is fixed to a constant (i.e., $p = 3$) in the probing state $|\Upsilon(q)\rangle$, we conclude that the total space complexity is $\mathcal{O}(M + \log_2 d)$, which is independent from the root MSE ε .

As for the gate complexity, we here focus on the number of queries to U_ψ (or $U_\psi^\dagger$) and U'_{SEL} . A comprehensive analysis on the total gate complexity is provided in Lemma 11 in Sec. 4.5.1. Here, we need to carefully choose the parameters ε' , ε'' , and δ' so that the entire approximation error Eq. (4.14) is smaller than $1/12$; in particular, this is achieved by taking ε' and δ' as some constants, and $\varepsilon'' = \mathcal{O}(1/(2^q \sigma))$. Since the block encoding of $\mathbf{H}$ consists of two uses of U_ψ and $U_\psi^\dagger$ and $m := \mathcal{O}(M\sigma^{-1} \log(M/(\sigma\varepsilon'')))$ uses of control- U'_{SEL} , the total queries can be evaluated by multiplying $\mathcal{O}(t)$ for the Hamiltonian simulation protocol, which yields the gate complexity as stated in Lemma 8. In addition, the number of circuit parameters for the singular value amplification in the step (iii) and the Hamiltonian simulation protocol are given by m and t , respectively. Thus, we need to tune the parameters via classical computation with time complexity of $\mathcal{O}(\text{poly}(m))$ and $\mathcal{O}(\text{poly}(t))$ time, respectively. $\square$

Before proceeding to the Grover-like repetition, we remark that our scheme has a significant improvement in space complexity compared to the previous non-iterative counterparts [86, 159]. While the non-iterative methods determine all of the $\mathcal{O}(\log 1/\varepsilon)$ binary fraction bits of $\langle O_j \rangle$ by a quantum circuit with additional $\mathcal{O}(M \log(1/\varepsilon))$ (or $\mathcal{O}(M \log(M/\varepsilon))$) qubits to read out the observables, our scheme determines only 1 binary fraction bit of $\langle O_j \rangle$ at each iteration step. Specifically, the previous non-iterative method [159] uses the Hamiltonian simulation with $t = 2^p$ in Eq. (4.12), where the

number of read-out qubits is $p = \mathcal{O}(\log(\sqrt{M \log d}/\varepsilon))$ (i.e., $q = 0$). In contrast, by the adaptive nature of our scheme, space overhead of quantum circuits for the probing state $|\Upsilon(q)\rangle$ is $\mathcal{O}(M)$ ($\mathcal{O}(M + \log_2 d)$ in total), and this is independent of the estimation precision ε . Here, a similar improvement of space overhead can be found in the previous works on the adaptive (iterative) versions of quantum phase estimation [99, 78, 77, 97, 48]. Also, we remark that the significant reduction of space overhead directly leads to the reduction of the number of controlled operations, which is also crucial in practical implementation.

Grover-like repetition

The quantum circuit employed in Lemma 8 requires classical tuning of $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$ circuit parameters (more precisely, $\tilde{\mathcal{O}}(2^q \sqrt{M})$ parameters in step q) for QSP. The classical computation for finding $\tilde{\mathcal{O}}(\sqrt{M}/\varepsilon)$ parameters is also required in the existing work by van Apeldoorn *et al.* [159]. As discussed in Sec. 4.1.2, this is in practice a significant overhead in classical computation time, for instance, 10^4 – 10^6 seconds for the case of $M = 10^4$ and $\varepsilon = 10^{-4}$. Although the classical computation of n parameters can be done with time complexity of $\mathcal{O}(\text{poly}(n))$ [57], it is extremely challenging if we desire to compute for $n > 10^6$ which may be required in, e.g., quantum chemistry applications. Even if we choose to compensate for the computation time with low accuracy in parameter finding, this may eventually negate the quantum enhancement in the gradient estimation algorithm.

To avoid the numerical instability, we here provide an alternative way to prepare the probing state (4.1) using the *Grover-like repetition*, which is a special case of QSP such that the corresponding quantum circuit parameters are analytically derived.

Lemma 9 (Informal). *Suppose we have access to block-encoded observables $\{O_j\}_{j=1}^M$ in d dimension, a $\log_2 d$ -qubit state preparation U_ψ , and its inverse $U_\psi^\dagger$, such that $M > \mathcal{O}(\log d)$. If the iteration step $q \geq 0$ satisfies the following condition ($\delta' := 2^{-14}$)*

$$q \geq \log_4 \left[\frac{2^3 \cdot 33^3}{625 \ln(2d/\delta')} \frac{\left\lceil \sqrt{2(M+1) \ln(2d/\delta')} \right\rceil}{\sqrt{\ln(2d/\delta')}} \right], \quad (4.18)$$

and for given $\tilde{u}_j^{(q)} \in [-1, 1]$,

$$\left| \langle \psi | \left(O_j - \tilde{u}_j^{(q)} \mathbf{1} \right) | \psi \rangle \right| \leq 2^{-q} \quad \text{for all } j = 1, 2, \dots, M, \quad (4.19)$$

holds, then we can successfully prepare the probing state $|\Upsilon(q)\rangle$ up to $1/12$ Euclidean distance error, with probability at least 0.462 with ancilla qubits measurement result indicating success, using an

$$\mathcal{O}(M + \log_2 d)\text{-qubit}$$

circuit regardless of q . Furthermore, each quantum circuit with q requires

$$\mathcal{O}(\text{poly}(\sqrt{M}(q + \log M)))$$

classical computation, and it has the same gate complexity as that of Lemma 8.

The formal version of this lemma is Lemma 12 in Sec. 4.5.1. Here, we remark that the condition Eq. (4.19) is trivially satisfied at the iteration step q in Algorithm 1 when all the previous gradient estimations are successfully performed, as discussed in Sec. 4.3.

Sketch of the proof. This method uses eigenvalue transformation of (a slightly modified version of) the Hamiltonian $\mathbf{H}$ in Eq. (4.10) based on the Chebyshev polynomial of the first kind $T_t(x)$, that is, Lemma 5. Here, the degree t is given by $t = 2^{p+q+2}\sigma$. We note that since the resulting operator $T_t[\mathbf{H}]$ is non-unitary, post-selection of ancilla qubits in the block encoding is required to implement $T_t[\mathbf{H}]$, meaning that the state preparation by this method is stochastic. As mentioned in Sec. 3.3.2, the eigenvalue transformation based on the Chebyshev polynomials $T_t(x)$ can be considered as a special case in QSVT; that is, it has an analytical solution of circuit parameters (or phase sequence). Therefore, in this alternative method, the circuit parameter finding is required only for the construction of the Hamiltonian $\mathbf{H}$; the total runtime for the classical computation is at most $\mathcal{O}(\text{polylog}(1/\varepsilon))$ (more precisely, $\mathcal{O}(\text{poly}(\sqrt{M}(q + \log M)))$ at step $q \leq \lceil \log_2(1/\varepsilon) \rceil$).

Now, we consider the action of the resulting operator $T_t[\mathbf{H}]$. Applying $T_t[\mathbf{H}]$ to the uniform superposition state $|+\rangle^{\otimes pM}$, we obtain an unnormalized state proportional to

$$\sum_{s=\pm 1} \sum_{\mathbf{x} \in G_p^M} e^{ist \arccos[\tilde{f}(\mathbf{x})]} |\mathbf{x}\rangle. \quad (4.20)$$

Therefore, we need to correct the phase of $|\mathbf{x}\rangle$ from $e^{\pm it \arccos[\tilde{f}(\mathbf{x})]}$ to $e^{it\tilde{f}(\mathbf{x})}$ in order to match the probing state $|\Upsilon(q)\rangle$.

As for the sign $s = \pm 1$ of the phase, we can correct it by slightly modifying the Hamiltonian $\mathbf{H}$ with the help of an additional 1-qubit ancilla system. Let $\mathbf{H}_G$ be a Hamiltonian defined as

$$\mathbf{H}_G := \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} \tilde{f}'(\mathbf{x}, y) |\mathbf{x}, y\rangle \langle \mathbf{x}, y|, \quad (4.21)$$

where $|y\rangle$ denotes a computational basis on the additional ancilla system and

$$\tilde{f}'(\mathbf{x}, y) \approx \frac{1}{\sigma'} \left(y \langle O_{M+1} \rangle + \sum_{j=1}^M x_j \langle \tilde{O}_j^{(q)} \rangle \right). \quad (4.22)$$

Here, O_{M+1} denotes a 1-qubit observable proportional to the identity, and $\sigma' = \mathcal{O}(\sqrt{M \log d})$ is the rescaling factor for block encoding. The block encoding of $\mathbf{H}_G$ can be constructed in a similar way to that of the Hamiltonian $\mathbf{H}$; the corresponding circuit has the same complexity as that of $\mathbf{H}$. Then, applying $T_t[\mathbf{H}_G]$ to the uniform superposition state $|+\rangle^{\otimes pM+1}$, we can show that the following holds: (for simplicity, we write O_{M+1} as $\tilde{O}_{M+1}^{(q)}$)

$$\begin{aligned} & \frac{1}{\mathcal{N}_t \sqrt{2^{pM}}} \sum_{\mathbf{x}, y} T_t \left(\tilde{f}'(\mathbf{x}, y) \right) |\mathbf{x}, y\rangle \\ & \approx \frac{1}{2\sqrt{2^{pM}}} \sum_{s=\pm 1} \sum_{(\mathbf{x}, x_{M+1}) \in G_p^M \times G_1} e^{ist \left(\pi/2 - (\sigma')^{-1} \sum_{j=1}^{M+1} x_j \langle \tilde{O}_j^{(q)} \rangle \right)} |\mathbf{x}\rangle |x_{M+1}\rangle \\ & = \frac{1}{\sqrt{2}} \sum_{s=\pm 1} \left(\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M (sx_j) \pi^{-1} 2^{q+1} \langle \tilde{O}_j^{(q)} \rangle} |\mathbf{x}\rangle \otimes \frac{1}{\sqrt{2}} \sum_{k \in G_1} e^{2\pi i 2k(s/4)} |k\rangle \right), \end{aligned} \quad (4.23)$$

where $\mathcal{N}_t$ is the normalization factor. In the second equality, we chose the constant factor of O_{M+1} as

$$O_{M+1} := \frac{\pi(1/4 + 4l)}{2^{p+q}} I,$$

where l is an arbitrary integer satisfying $\|2^{q+1} O_{M+1}\| \leq 1$ and I denotes the 1-qubit identity. In the next paragraph, we explain the approximation of the first line in Eq. (4.23) in detail. Suppose that the approximation holds, then we can correct the sign $s = \pm 1$ in Eq. (4.23) and as a result obtain the probing state $|\Upsilon(q)\rangle$. That is, by applying the inverse quantum Fourier transformation to the 1-qubit additional ancilla system in the state Eq. (4.23), followed by controlled $X^{\otimes pM}$ gate (which flips the sign as $X^{\otimes pM} |\mathbf{x}\rangle = |-\mathbf{x}\rangle$), we obtain the probing state Eq. (4.1) in the pM -qubit registers.

Now, we argue that the approximation in the first line of Eq. (4.23) follows from the non-linearity of the phase function $\arccos x$ in Eq. (4.20). Here, we recall that $\tilde{f}'(\mathbf{x}, y)$ in Eq. (4.22) is approximately given by the linear combination of $\langle O_j - \tilde{u}_j^{(q)} \mathbf{1} \rangle / \sigma'$. Using this fact, we can show that $|\tilde{f}'(\mathbf{x}, y)|$ scales as $\mathcal{O}(\sqrt{M}/(2^q \sigma'))$ for a large part of $G_p^M \times G_1$, from the assumption of $|\langle \tilde{O}_j^{(q)} \rangle| \leq 1/2^{q+1}$ and the Hoeffding's inequality for $(\mathbf{x}, y)$ which is uniformly distributed. Moreover, the error between the function $\arccos(x)$ and a linear function $\pi/2 - x$ can be upper bounded by $\mathcal{O}(|x|^3)$. As a result, the non-linearity of $\arccos x$ in Eq. (4.20) can be suppressed for a large part of $G_p^M \times G_1$ as

$$\left| \arccos[\tilde{f}'(\mathbf{x}, y)] - \frac{\pi}{2} + \tilde{f}'(\mathbf{x}, y) \right| = \mathcal{O} \left(\left(\frac{\sqrt{M}}{2^q \sigma'} \right)^3 \right).$$

In the approximation error of Eq. (4.23), the above error is amplified by the factor of $t = \mathcal{O}(2^q \sigma')$ and the amplified error decreases

$$\mathcal{O}\left(\frac{\sqrt{M}}{4^q} \frac{M}{(\sigma')^2}\right) \quad (4.24)$$

for $\sigma' = \mathcal{O}(\sqrt{M \log d})$. Then, we prove the approximation error in Eq. (4.23) is at most $1/12$ in Euclidean distance error under the condition that the amplified error in Eq. (4.24) is sufficiently small, which leads to the assumption in Eq. (4.18) of this lemma. At the same time, we can evaluate the normalization factor as $|\mathcal{N}_t - 1| \leq 2\sqrt{6}\delta'$, thereby proving the post-selection probability $\mathcal{N}_t^2/2 > 0.462$.

Finally, we mention the space and gate complexity of this Lemma. As well as in the proof of Lemma 8, we can count the number of qubits in the block encoding of $\mathbf{H}_G$, and it scales as $pM + \log M + \log d$ in total. Thus, the method has the space complexity $\mathcal{O}(M + \log_2 d)$ that is also independent of the target precision ε . As for the gate complexity, the main contribution comes from the eigenvalue transformation $\mathbf{H}_G \mapsto T_t[\mathbf{H}_G]$, which uses $t = \mathcal{O}(2^q \sigma')$ queries of the block encoding of $\mathbf{H}_G$. This indicates that the gate complexity of the method is the same as that of the method in Lemma 8. $\square$

In this method, we require finding $\tilde{\mathcal{O}}(q\sqrt{M})$ circuit parameters, which is due to the singular value amplification to construct the Hamiltonian Eq. (4.21), while the Grover-like repetition itself does not require any phase angle computation. Thus, the alternative method shows an exponential improvement of runtime in classical computation with respect to the iteration q (or ε), compared to the method in Lemma 8 and the previous method [159], if certain conditions on the iteration step q are satisfied. Recalling that the iteration step is upper bounded by $q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$ for the target root MSE ε , the alternative method can be used if there exists the iteration steps q satisfying $q \leq q_{\max}$ and the inequality Eq. (4.18). Later, we numerically investigate the range of iteration steps q for the use of Lemma 9, especially for the case of $M = \mathcal{O}((\log_2 d)^k)$ for some k .

Finally, we mention the success probability of Lemma 9. While we cannot prepare the probing state $|\Upsilon(q)\rangle$ deterministically in the alternative method, the success probability can be easily boosted by a constant number of repetitions of the circuit runs. Therefore, the total query complexity remains unchanged when we use Lemma 9 instead of Lemma 8 to prepare the probing state of Eq. (4.1).

Applicability condition for the probing-state preparation by Grover-like repetition

Here, we investigate the threshold of iteration step q in Lemma 9. We rewrite the threshold Eq. (4.18) as $q^* := \log_2(1/\varepsilon^*)$ by defining ε^* as

$$1/\varepsilon^* := \sqrt{C \ln^{-3/2}(2d/\delta')} \left\lceil \sqrt{2(M+1) \ln(2d/\delta')} \right\rceil, \quad (4.25)$$

where $C := 2^3 \cdot 33^3/625$. Since the iteration step q runs from 0 to $q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$, we need that the threshold $1/\varepsilon^*$ is smaller than the inverse of the target root MSE $1/\varepsilon$ in order to use the state preparation by the Grover-like repetition.

To clarify this point, we here plot the threshold $1/\varepsilon^*$ for several cases $M = N^2, N^3, N^4$, and 2^N under target system with the number of qubits $N \equiv \log_2 d$. In Fig. 4.4, we confirm that the threshold $\log_2(1/\varepsilon^*)$ converges to some constant (~ 5) in the case of $M = N^2$, which is consistent with Eq. (4.25) because $(1/\varepsilon^*)^2 = \mathcal{O}(\sqrt{M}/N)$ holds. Thus, particularly in this case $M = \mathcal{O}(N^2)$, the method in Lemma 9 is available in a wide range of target precision ε regardless of N . On the other hand, if the number of observables scales as 2^N , the threshold $\log_2(1/\varepsilon^*)$ increases linearly regarding N (note that the horizontal axis in Fig. 4.4 is logarithmic). This means that in the case of $M = \mathcal{O}(2^N)$, the Grover-based method is applicable only for small-size systems, otherwise the target precision ε is exponentially small with respect to N (the base of the exponential is $2^{-1/4} \approx 0.841$).

Now, we focus on the case that we require more precise estimates as the number of observables M increases. From the definition of ε^* , if the desired precision ε satisfy $\varepsilon^2 \ll N/(C\sqrt{M})$, then there exists iteration steps q for the Grover-based method. Specifically, considering the desired precision is given by $\varepsilon = c_{\text{mse}}/\sqrt{M} \in (0, 1/\sqrt{M})$ for some $c_{\text{mse}} \in (0, 1)$, we can evaluate the difference between the upper bound $q_{\max}$ of iteration steps and the threshold q^* as follows:

$$q_{\max} - q^* \geq \Omega\left(\log\left(N\sqrt{M}\right)\right). \quad (4.26)$$

Thus, the range of iteration step q such that the Grover-based method is available enlarges in this case, as N or M increases.

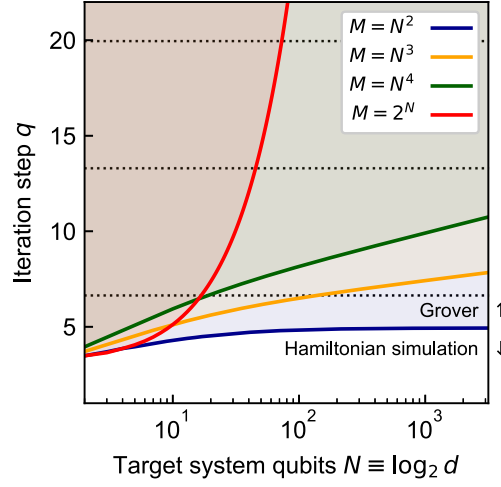


Figure 4.4: The condition for the Grover-based state preparation (Lemma 9). The solid lines represent the threshold $q^* := \log_2(1/\varepsilon^*)$ of iteration steps in Eq. (4.25) (or Eq. (4.18)). When the iteration step q exceeds the solid lines, we can use the Grover-like repetition to prepare the probing state Eq. (4.1), instead of using Hamiltonian simulation. The horizontal dotted lines represent $q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$ for $\varepsilon = 10^{-2}$ (bottom), 10^{-4} (middle), and 10^{-6} (top), respectively. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

4.5 Circuit implementation and theoretical guarantees

In Sec. 4.4, we have provided two methods to prepare the probing state $|\Upsilon(q)\rangle$ that was defined in Eq. (4.1) as

$$|\Upsilon(q)\rangle := \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^q \pi^{-1} \langle O_j - \bar{u}_j^{(q)} \mathbf{1} \rangle} |\mathbf{x}\rangle.$$

In the following, we give more details regarding the state preparation in Sec. 4.5.1, and then analyze the algorithm of the adaptive observables estimation using those state preparations in Sec. 4.5.2. The quantum circuits are provided in Sec. 4.5.3. In this section, we often refer to the detailed version Algorithm 1*, instead of Algorithm 1 in Sec. 4.3.

4.5.1 Probing-state preparation

Here, we provide two quantum algorithms to prepare the probing state $|\Upsilon(q)\rangle$ in Eq. (4.1), using block encodings of observables $\{O_j\}$ and the state preparation oracle U_ψ . First, we describe a method to construct the block encoding of a linear combination of observables.

Amplified block encoding for observables

Let B_j ($j = 1, \dots, M$) be an a -block-encoding of a $d \times d$ observable O_j , and let $\mathbf{p} := (p_1, \dots, p_M)$ be a sequence of positive integers. Here, we define a set $G_{\mathbf{p}}$ of grid points as

$$G_{\mathbf{p}} := G_{p_1} \times \dots \times G_{p_M},$$

where each G_p is defined as in Eq. (3.20). For any element $\mathbf{x} = (x_1, \dots, x_M) \in G_{\mathbf{p}}$, we can construct an $(a + \lceil \log_2 M \rceil + 1)$ -block-encoding of

$$\tilde{H}^{(\mathbf{x})} := M^{-1} \sum_{j=1}^M x_j O_j$$

Algorithm 1* Adaptive gradient estimation for multiple observables (A detailed version.)

Input: $\log_2 d$ -qubit state preparation unitary U_ψ and $U_\psi^\dagger$; observables $\{O_j\}_{j=1}^M$ with the spectral norm $\|O_j\| \leq 1$ such that

$$M > 2 \ln d + 24 = \mathcal{O}(\log d) \quad (4.27)$$

holds; confidence parameter $c \in (0, 3/8(1 + \pi)^2]$; target precision parameter $\varepsilon \in (0, 1)$.

Output: A sample $(\tilde{u}_1, \dots, \tilde{u}_M)$ from an estimator $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_M)$ whose j -th element estimates $\langle O_j \rangle := \langle \psi | O_j | \psi \rangle$ within MSE ε^2 as $\max_{j=1,2,\dots,M} \mathbb{E} \left[(\tilde{u}_j - \langle O_j \rangle)^2 \right] \leq \varepsilon^2$

1: Set $p := 3$ and $\tilde{u}_j^{(0)} := 0$, ($j = 1, 2, \dots, M$)

2: **for** $q = 0, 1, \dots, q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$ **do**

3: Set

$$\tilde{O}_j^{(q)} := \frac{O_j - \tilde{u}_j^{(q)} \mathbf{1}}{2} \quad \text{and} \quad \delta^{(q)} := \frac{c}{8^{q_{\max} - q}}$$

4: Prepare $\mathcal{O}(\log(M/\delta^{(q)}))$ copies of a quantum state that is $1/12$ -close (in Euclidean distance) to

$$\left(\text{QFT}_{G_p}^\dagger \right)^{\otimes M} |\Upsilon(q)\rangle, \text{ where } |\Upsilon(q)\rangle := \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^{q+1} \pi^{-1} \langle \tilde{O}_j^{(q)} \rangle} |\mathbf{x}\rangle,$$

using Lemma 11 or Lemma 12 and perform the computational basis measurement on each of them. Note that each measurement outputs a result in the form of $(x_1, \dots, x_M) \in G_p^M$.

5: Set coordinate-wise medians of the measurement results as $g_j^{(q)}$

6: Set $\tilde{u}_j^{(q+1)} := \tilde{u}_j^{(q)} + \pi 2^{-q} g_j^{(q)}$

7: **if** there are some j such that $\tilde{u}_j^{(q+1)} \geq 1$ (or ≤ -1) **then**

8: Set $\tilde{u}_j^{(q+1)} = 1$ (or -1) for such j

9: **end if**

10: **end for**

11: **return** $\tilde{u}_j := \tilde{u}_j^{(q_{\max}+1)}$

by using the LCU method. The corresponding SELECT operation denoted by U_{SEL} is given by

$$U_{\text{SEL}} := \sum_{j=1}^M |j\rangle \langle j| \otimes B_j.$$

As for the PREPARE operation, we consider implementation with and without reflecting the signs of coefficients as $P_{\text{R}}^{(\mathbf{x})}$ and $P_{\text{L}}^{(\mathbf{x})}$, defined as

$$P_{\text{R}}^{(\mathbf{x})} : |0\rangle^{\otimes \lceil \log_2 M \rceil} \mapsto \sum_{j=1}^M \text{sgn}(x_j) \sqrt{\frac{|x_j|}{\|\mathbf{x}\|_1}} |j\rangle. \quad (4.28)$$

$$P_{\text{L}}^{(\mathbf{x})} : |0\rangle^{\otimes \lceil \log_2 M \rceil} \mapsto \sum_{j=1}^M \sqrt{\frac{|x_j|}{\|\mathbf{x}\|_1}} |j\rangle \quad (4.29)$$

Then, we can confirm that the unitary $(P_{\text{L}}^{(\mathbf{x})} \otimes \mathbf{1})^\dagger U_{\text{SEL}} (P_{\text{R}}^{(\mathbf{x})} \otimes \mathbf{1})$ is a block encoding of $\propto \tilde{H}^{(\mathbf{x})}$ because for any $|\psi\rangle$,

$$\begin{aligned} & (P_{\text{L}}^{(\mathbf{x})} \otimes \mathbf{1})^\dagger U_{\text{SEL}} (P_{\text{R}}^{(\mathbf{x})} \otimes \mathbf{1}) |0\rangle^{\otimes \lceil \log_2 M \rceil} |\psi\rangle \\ &= \sum_{j=1}^M \text{sgn}(x_j) \sqrt{\frac{|x_j|}{\|\mathbf{x}\|_1}} (P_{\text{L}}^{(\mathbf{x})})^\dagger |j\rangle \otimes B_j |\psi\rangle = |0\rangle^{\otimes \lceil \log_2 M \rceil} \otimes \sum_{j=1}^M \frac{x_j}{\|\mathbf{x}\|_1} B_j |\psi\rangle + |\tilde{\psi}^\perp\rangle, \end{aligned} \quad (4.30)$$

where $\langle 0 |^{\otimes \lceil \log_2 M \rceil} |\tilde{\psi}^\perp\rangle = 0$. While the block encoding has the normalization factor $\|\mathbf{x}\|_1$ which depends on $\mathbf{x}$, it can be modified to M by introducing an additional single ancilla qubit and a single rotation gate $R_y(\theta) = e^{i\theta Y}$. As a result, the following unitary is $(a + \lceil \log_2 M \rceil + 1)$ -block-encoding of $\tilde{H}^{(\mathbf{x})}$:

$$(I \otimes P_L^{(\mathbf{x})} \otimes \mathbf{1})^\dagger (I \otimes U_{\text{SEL}}) (R_y(\theta_{\mathbf{x}}) \otimes P_R^{(\mathbf{x})} \otimes \mathbf{1}), \quad (4.31)$$

where I denotes the 1-qubit identity, and $R_y(\theta_{\mathbf{x}}) : |0\rangle \mapsto \frac{\|\mathbf{x}\|_1}{M} |0\rangle + \sqrt{1 - \left(\frac{\|\mathbf{x}\|_1}{M}\right)^2} |1\rangle$.

In this block encoding of $\tilde{H}^{(\mathbf{x})}$, only PREPARE operation depends on the grid point $\mathbf{x} \in G_{\mathbf{p}}$. Thus, by adding a control to $P_L^{(\mathbf{x})}$, $P_R^{(\mathbf{x})}$, and R_y , and then sequentially applying them to each $\mathbf{x}$, we can implement the following unitary

$$U'_{\text{SEL}} := \sum_{\mathbf{x} \in G_{\mathbf{p}}} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U^{(\mathbf{x})}, \quad (4.32)$$

where $U^{(\mathbf{x})}$ denotes the block encoding of $\tilde{H}^{(\mathbf{x})}$. Because the detailed circuit description of U'_{SEL} is not required in the following, we deal with the unitary U'_{SEL} as an oracle for observables $\{O_j\}_{j=1}^M$, instead of U_{SEL} , for a simple expression of gate complexity.

Remark 1. The above straightforward construction of U'_{SEL} results in a very large circuit with $\mathcal{O}(2^{\|\mathbf{p}\|_1})$ -depth. Fortunately, we can construct a more efficient circuit by further modifying the PREPARE operation, using a similar way to Ref. [106]. As shown in the Sec. 3 of Ref. [56], we can implement $\sum_{x \in G_p} |x\rangle \langle x| \otimes e^{-ixZ}$ with $\mathcal{O}(p)$ -depth circuit using $p+1$ (controlled) single-qubit rotation gates. By sequentially applying the $|j\rangle$ -controlled version of this gate (see Fig. 4.10) over x_j -register, the additional single ancilla qubit, and the $\lceil \log_2 M \rceil$ qubits, we obtain a $\mathbf{x}$ -controlled block encoding of $e^{i\mathcal{H}^{(\mathbf{x})}}$, where $\mathcal{H}^{(\mathbf{x})} := \sum_{j=1}^M x_j |j\rangle \langle j|$. The corresponding circuit diagram is shown in Fig. 4.11. Then, we can obtain an ϵ -precise block encoding of $(2/\pi)\mathcal{H}^{(\mathbf{x})}$, via the eigenvalue transformation for the logarithm of unitaries: $e^{i\mathcal{H}^{(\mathbf{x})}} \mapsto \mathcal{H}^{(\mathbf{x})}$ (Corollary 71 in [57]). This implementation consists of $\mathcal{O}(\log(1/\epsilon))$ uses of block-encoding of $e^{i\mathcal{H}^{(\mathbf{x})}}$ or its inverse; see Fig. 4.12. Now, introducing the SELECT U_{SEL} and Hadamard gates, we have an ϵ -precise block-encoding of $(2M/(\pi 2^{\lceil \log_2 M \rceil}))\tilde{H}^{(\mathbf{x})}$ controlled by $\mathbf{x}$, using at most $\mathcal{O}(\|\mathbf{p}\|_1 \log(1/\epsilon))$ -depth circuit over $\mathcal{O}(\|\mathbf{p}\|_1 + a + \lceil \log_2 M \rceil + \log_2 d)$ qubits in total:

$$\langle \mathbf{0} | \text{Had}^{\otimes \lceil \log_2 M \rceil} \otimes \mathbf{1} \cdot U_{\text{SEL}} \cdot \frac{2}{\pi} \mathcal{H}^{(\mathbf{x})} \text{Had}^{\otimes \lceil \log_2 M \rceil} | \mathbf{0} \rangle \otimes \mathbf{1} = \frac{2M}{\pi 2^{\lceil \log_2 M \rceil}} \frac{1}{M} \sum_{j=1}^M x_j B_j. \quad (4.33)$$

Note that this block encoding has a single use of $U_{\text{SEL}} = \sum_j |j\rangle \langle j| \otimes B_j$ regardless of ϵ ; see Fig. 4.9.

Using U'_{SEL} , U_ψ , and its conjugation $U_\psi^\dagger$, we obtain a block encoding of the following Hamiltonian:

$$\sum_{\mathbf{x}} \langle \psi | \tilde{H}^{(\mathbf{x})} | \psi \rangle |\mathbf{x}\rangle \langle \mathbf{x}| = \sum_{\mathbf{x}} \left(\sum_j x_j \cdot \frac{\langle O_j \rangle}{M} \right) |\mathbf{x}\rangle \langle \mathbf{x}|.$$

This allows us to simulate the phase oracle for the affine linear function $\sum_j x_j \langle O_j \rangle / M$ by using the block-Hamiltonian simulation (Lemma 6). However, in estimating the expectation value $\langle O_j \rangle$, we need to amplify the function by the factor M due to the normalization factor $1/M$ in $\tilde{H}^{(\mathbf{x})}$, which may result in no speedup for the total queries to U_ψ regarding the number M of observables. To obtain the quadratic speedup regarding M , Ref. [159] (Lemma 36) uses a linear amplification of $\tilde{H}^{(\mathbf{x})}$ by Lemma 4, and this amplification can be performed with no use of U_ψ . Here, we provide a slightly modified version of the Lemma 36 in Ref. [159] as follows.

Lemma 10. *Let $\delta' > 0$, $\epsilon' \in (0, 1/2)$, and let $\mathbf{p} = (p_1, p_2, \dots, p_M)$ be any sequence of positive integers. Suppose that for observables $\{O_j\}_{j=1}^M$ in d dimension (d is a power of 2) with $\|O_j\| \leq 1$, we have access to the oracle*

$$U_{\text{SEL}} = \sum_{\mathbf{x} \in G_{\mathbf{p}}} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U^{(\mathbf{x})},$$

where $U^{(\mathbf{x})}$ is an $(a + \lceil \log_2 M \rceil)$ -block encoding of observable $\tilde{H}^{(\mathbf{x})} := M^{-1} \sum_{j=1}^M x_j O_j$ for some $a \in \mathbb{N}$. If the following inequality holds

$$\sigma := \left\lceil \sqrt{2M \ln(2d/\delta')} \right\rceil < M, \quad (4.34)$$

then there exists a subset F of $G_{\mathbf{p}}$ with the cardinality $|F|$ satisfying $|F| \geq (1 - \delta')|G_{\mathbf{p}}|$, and we can implement a unitary (with $\|\mathbf{p}\|_1 + \lceil \log_2 M \rceil + \log_2 d + a + 1$ qubits in total)

$$U_{\text{obs}} := \sum_{\mathbf{x} \in G_{\mathbf{p}}} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U_{\text{obs}}^{(\mathbf{x})} \quad (4.35)$$

such that $U_{\text{obs}}^{(\mathbf{x})}$ is a $(1, a + \lceil \log_2 M \rceil + 1, \varepsilon')$ -block-encoding of the Hamiltonian

$$H^{(\mathbf{x})} := \frac{1}{\sigma} \sum_{j=1}^M x_j O_j \quad \text{if } \mathbf{x} \in F \subset G_{\mathbf{p}}. \quad (4.36)$$

For $m := \mathcal{O}(M\sigma^{-1} \log(M\sigma^{-1}/\varepsilon'))$, this implementation of U_{obs} uses m queries to U'_{SEL} or its inverse, $2m$ NOT gates controlled by $(a + \lceil \log_2 M \rceil)$ -qubit, and $\mathcal{O}(m)$ single-qubit gates, with $\mathcal{O}(\text{poly}(m))$ classical precomputation for finding quantum circuit parameters. The circuit diagram for U_{obs} is shown in Fig. 4.13.

Proof of Lemma 10. Let $\gamma := M/\sigma > 1$. To perform the amplification in Lemma 4, it is required that the spectral norm $\|\tilde{H}^{(\mathbf{x})}\|$ is upper bounded by $(1 - \delta)/\gamma$ for some $\delta \in (0, 1/2)$. To this end, we define a subset $F \subset G_{\mathbf{p}}$ such that

$$F := \left\{ \mathbf{x} \in G_{\mathbf{p}} : \|\tilde{H}^{(\mathbf{x})}\| < \frac{1}{2\gamma} \right\}.$$

If $\mathbf{x} \in F$, we can perform the linear amplification by γ and obtain $(1, a + \lceil \log_2 M \rceil + 1, \varepsilon')$ -block-encoding of $\gamma \tilde{H}^{(\mathbf{x})}$ from Lemma 4. Next, we show the cardinality of F is equal or more than $(1 - \delta') \prod_{j=1}^M 2^{p_j}$. Let us consider independent random variables X_j ($j = 1, \dots, M$) that are uniformly distributed on each G_{p_j} . Because the random variables are also symmetrically distributed on each G_{p_j} , the following matrix series inequality holds (Theorem 35 in [159]):

$$\Pr_{X_1, \dots, X_M} \left[\left\| \sum_{j=1}^M (2X_j) O_j \right\| \geq t \right] \leq 2d \cdot \exp \left[- \frac{t^2}{2 \left\| \sum_{j=1}^M O_j^2 \right\|} \right]. \quad (4.37)$$

Thus, inserting M/γ into t , we obtain

$$\Pr_{X_1, \dots, X_M} \left[\left\| \tilde{H}^{(\mathbf{X})} \right\| \geq 1/2\gamma \right] \leq \delta'. \quad (4.38)$$

Since the probability of an event $X_j = x_j$ ($j = 1, \dots, M$) is equal to $1/\prod_{j=1}^M 2^{p_j}$ for all x_j , we obtain $|F| \geq (1 - \delta') \prod_{j=1}^M 2^{p_j}$. The gate complexity follows from that of Lemma 4. $\square$

Here, we make some remarks for Lemma 10. In the proof of this lemma, we used the loose inequality: $\left\| \sum_{j=1}^M O_j^2 \right\| \leq M$ (we assumed $\|O_j\| \leq 1$) to derive Eq. (4.38) from Eq. (4.37), because explicitly calculating the factor $\left\| \sum_{j=1}^M O_j^2 \right\|$ may be involved. If we have some evaluation of the factor $\left\| \sum_{j=1}^M O_j^2 \right\|$ in a particular setup, we can further improve the block encoding Eq. (4.36); see Ref. [159].

Also, the condition Eq. (4.34) for the amplification, $M > \mathcal{O}(\log(d/\delta'))$, is crucial in our algorithm, and for a constant δ' , this condition is satisfied when the number of observables is bigger than the number of qubits such as Eq. (4.27).

When we use the circuit construction of U'_{SEL} in Remark 1, $U^{(\mathbf{x})}$ in the assumption of Lemma 10 should be replaced as the ϵ -precise block encoding of $(2M/(\pi 2^{\lceil \log_2 M \rceil})) \tilde{H}^{(\mathbf{x})}$. In this case, the error ε' in the block encoding $U_{\text{obs}}^{(\mathbf{x})}$ of $H^{(\mathbf{x})}$ becomes $\varepsilon' + \mathcal{O}(m\sqrt{\epsilon})$, by the error propagation associated with the uniform amplification Lemma 4 (see; Lemma 22 in Ref. [57]). Thus, replacing ε' and ϵ with $\varepsilon'/2$ and $\mathcal{O}((\varepsilon'/m)^2)$, respectively, we obtain the ε' -precise block encoding $U_{\text{obs}}^{(\mathbf{x})}$.

Approximated state preparation

Now, we are ready to show two quantum algorithms to prepare the probing state $|\Upsilon(q)\rangle$. Note that in the following lemmas, we explicitly construct a quantum circuit in its proof.

Lemma 11 (State preparation by Hamiltonian simulation). *Let O_j ($j = 1, 2, \dots, M$) be observables in d dimension (d is a power of 2) with $\|O_j\| \leq 1$, and let p be a positive integer. Also, let U_ψ be a $\log_2 d$ -qubit state preparation oracle, and $\langle O_j \rangle := \langle \psi | O_j | \psi \rangle$. Suppose we have access to*

$$U'_{\text{SEL}} := \sum_{\mathbf{x} \in G_p^M} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U^{(\mathbf{x})},$$

where $U^{(\mathbf{x})}$ is an $(a + \lceil \log_2 M \rceil)$ -block-encoding of $M^{-1} \sum_{j=1}^M x_j O_j$ for some $a \in \mathbb{N}$. We assume that

$$\sigma := \left\lceil \sqrt{2M \ln(2d/\delta')} \right\rceil < M$$

holds for $\delta' = 2^{-10}$. Then, for any non-negative integer q , we can prepare the pM -qubit quantum state

$$\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^{q+1} \pi^{-1} \langle O_j \rangle} |\mathbf{x}\rangle$$

up to $1/12$ Euclidean distance error, with $2Q$ uses of U_ψ or $U_\psi^\dagger$, $4mQ$ uses of controlled U'_{SEL} or its inverse, $\mathcal{O}(mQ)$ uses of NOT gates controlled by at most $\mathcal{O}(a + \lceil \log_2 M \rceil + \log_2 d)$ qubits, $\mathcal{O}(mQ + pM)$ uses of single-qubit or two-qubit gates, additional $(\lceil \log_2 M \rceil + \log_2 d + a + 3)$ ancilla qubits, and $\mathcal{O}(\text{poly}(Q) + \text{poly}(m))$ classical precomputation for finding quantum circuit parameters, where

$$Q = \mathcal{O}(2^{p+q+2}\sigma) \quad \text{and} \quad m = \mathcal{O}\left(\frac{M}{\sigma}(p + q + \log M)\right).$$

Note that if we have a block encoding B_j of O_j , then a block encoding of $(O_j - \tilde{u}_j^{(q)} \mathbf{1})/2$ for any $\tilde{u}_j^{(q)} \in [-1, 1]$ is easily constructed by the LCU method; see Fig. 4.14. Therefore, Lemma 11 provides a method to approximately prepare the probing state $|\Upsilon(q)\rangle$.

Proof of Lemma 11. Let $\varepsilon'' = 2^{-14} \in (0, 1)$ and $\varepsilon' = \sqrt{\delta'}/(2^{p+q+2}\sigma) \in (0, 1/2)$. From the assumption and Lemma 10, we have a unitary

$$U_{\text{obs}} := \sum_{\mathbf{x} \in G_p^M} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U_{\text{obs}}^{(\mathbf{x})}, \quad (4.39)$$

where $U_{\text{obs}}^{(\mathbf{x})}$ is a $(1, a'' = a + \lceil \log_2 M \rceil + 1, \varepsilon')$ -block-encoding of the Hamiltonian $H^{(\mathbf{x})} := \sigma^{-1} \sum_{j=1}^M x_j O_j$ if $\mathbf{x} \in F$ with $|F| \geq (1 - \delta')|G_p^M|$. Thus, using U_ψ , $U_\psi^\dagger$, and U_{obs} , we can implement a $(1, a'' + \log_2 d, 0)$ -block-encoding of the following Hamiltonian

$$\sum_{\mathbf{x} \in G_p^M} \tilde{f}(\mathbf{x}) |\mathbf{x}\rangle \langle \mathbf{x}|, \quad (4.40)$$

where $\tilde{f}(\mathbf{x}) := \langle 0 |^{\otimes a'' + \log_2 d} U_\psi^\dagger \cdot U_{\text{obs}}^{(\mathbf{x})} \cdot U_\psi | 0 \rangle^{\otimes a'' + \log_2 d}$. The block-Hamiltonian simulation (Lemma 6) yields the quantum circuit W for a $(1, a'' + \log_2 d + 2, \varepsilon'')$ -block-encoding of time evolution operator

$$\sum_{\mathbf{x} \in G_p^M} e^{i\tilde{f}(\mathbf{x})t} |\mathbf{x}\rangle \langle \mathbf{x}| \quad \text{and} \quad t := 2^{p+2+q}\sigma. \quad (4.41)$$

An example of this quantum circuit is shown in Fig. 4.15. Note that a block encoding W' of ε'' -precise time evolution operator e^{iHt} for some Hamiltonian H yields $\mathcal{O}(\sqrt{\varepsilon''})$ -precise time evolved states, that is, for any state $|\psi\rangle$,

$$\|W' |\mathbf{0}\rangle |\psi\rangle - |\mathbf{0}\rangle e^{iHt} |\psi\rangle\| \leq \varepsilon'' + \sqrt{2\varepsilon''},$$

where $|\mathbf{0}\rangle$ is the signal state for W' . Therefore, for $|\mathbf{0}\rangle = |0\rangle^{\otimes a'' + \log_2 d + 2}$ and the $+1$ eigenstate $|+\rangle$ of Pauli X , we obtain

$$\begin{aligned} & \left\| W |\mathbf{0}\rangle |+\rangle^{\otimes pM} - \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j \pi^{-1} 2^{q+1} \langle O_j \rangle} |\mathbf{0}\rangle |\mathbf{x}\rangle \right\| \\ & \leq \varepsilon'' + \sqrt{2\varepsilon''} + \frac{1}{\sqrt{2^{pM}}} \left\| \sum_{\mathbf{x} \in G_p^M} e^{i\tilde{f}(\mathbf{x})t} |\mathbf{x}\rangle - \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j \pi^{-1} 2^{q+1} \langle O_j \rangle} |\mathbf{x}\rangle \right\| \\ & \leq \varepsilon'' + \sqrt{2\varepsilon''} + \sqrt{5\delta'} < \frac{1}{12}. \end{aligned} \quad (4.42)$$

Here, in the second inequality we used the following evaluation.

$$\begin{aligned} & \left\| \sum_{\mathbf{x} \in G_p^M} e^{i\tilde{f}(\mathbf{x})t} |\mathbf{x}\rangle - \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j \pi^{-1} 2^{q+1} \langle O_j \rangle} |\mathbf{x}\rangle \right\|^2 \\ & \leq \sum_{\mathbf{x} \in G_p^M} \left| e^{i\tilde{f}(\mathbf{x})t} - e^{2\pi i 2^p \sum_{j=1}^M x_j \pi^{-1} 2^{q+1} \langle O_j \rangle} \right|^2 \\ & \leq 4 \times 2^{pM} \delta' + \sum_{\mathbf{x} \in F} \left| e^{i\tilde{f}(\mathbf{x})t} - e^{2\pi i 2^p \sum_{j=1}^M x_j \pi^{-1} 2^{q+1} \langle O_j \rangle} \right|^2 \quad (\because |F| \geq (1 - \delta') |G_p^M|) \\ & \leq 4 \times 2^{pM} \delta' + \sum_{\mathbf{x} \in F} \left| \tilde{f}(\mathbf{x})t - 2\pi 2^p \sum_{j=1}^M x_j \pi^{-1} 2^{q+1} \langle O_j \rangle \right|^2 \quad (\because |e^{ia} - e^{ib}| \leq |a - b|) \\ & \leq 4 \times 2^{pM} \delta' + \sum_{\mathbf{x} \in F} (t\varepsilon')^2 \leq 2^{pM} \times 5\delta' \quad (\because \varepsilon' := \sqrt{\delta'}/t). \end{aligned} \quad (4.43)$$

The gate complexity follows from that of Lemma 10 and Lemma 6. $\square$

Lemma 12 (State preparation by Grover-like repetition). *Let O_j ($j = 1, 2, \dots, M$) be observables in d dimension (d is a power of 2) with $\|O_j\| \leq 1$, and let p be a positive integer. We assume the following two conditions holds for $\delta' := 2^{-14} \in (0, 1)$:*

$$(i) \quad \sigma := \left\lceil \sqrt{2(M+1) \ln(2d/\delta')} \right\rceil < M + 1.$$

(ii) *For a given $(\log_2 d)$ -qubit quantum state $|\psi\rangle$ prepared by U_ψ , $|\langle O_j \rangle| \leq 2^{-q-1}$ holds for all $j = 1, 2, \dots, M$, where q denotes a non-negative integer satisfying*

$$q \geq \log_4 \left[\frac{2^p \cdot 33^3}{625 \ln(2d/\delta')} \frac{\left\lceil \sqrt{2(M+1) \ln(2d/\delta')} \right\rceil}{\sqrt{\ln(2d/\delta')}} \right] \quad (4.44)$$

Suppose we have access to

$$U'_{\text{SEL}} := \sum_{\mathbf{x} \in G_p^M \times G_1} |\mathbf{x}\rangle \langle \mathbf{x}| \otimes U^{(\mathbf{x})},$$

where $U^{(\mathbf{x})}$ is an $(a + \lceil \log_2(M+1) \rceil)$ -block-encoding of

$$\frac{1}{M+1} \left(\sum_{j=1}^{M+1} x_j O_j \right)$$

for some $a \in \mathbb{N}$. Here, O_{M+1} denotes a 2×2 observable defined as

$$O_{M+1} := \frac{\pi(1/4 + 4l)}{2^{p+q}} I,$$

where l is an arbitrary integer satisfying $\|2^{q+1}O_{M+1}\| \leq 1$ and I denotes the 1-qubit identity. Then, we can successfully prepare the following pM -qubit quantum state with a measurement result of ancilla qubits indicating the success:

$$\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^{q+1} \pi^{-1} \langle O_j \rangle} |\mathbf{x}\rangle$$

up to $1/12$ Euclidean distance error with the success probability ≥ 0.462 . This implementation consists of $2t$ uses of U_ψ or $U_\psi^\dagger$, mt uses of $U_{\text{SEL}}^\dagger$ or its inverse, $(2m+1)t$ uses of NOT gates controlled by at most $(a + \lceil \log_2(M+1) \rceil + \log_2 d)$ qubits, $\mathcal{O}(mt + pM)$ uses of single-qubit or two-qubit gates, additional $(\lceil \log_2(M+1) \rceil + \log_2 d + a + 2)$ ancilla qubits, and $\mathcal{O}(\text{poly}(m))$ classical precomputation for finding quantum circuit parameters, where

$$t = 2^{p+q+2}\sigma \quad \text{and} \quad m = \mathcal{O}\left(\frac{M}{\sigma}(p + q + \log M)\right).$$

Proof of Lemma 12. Let $\varepsilon' = \sqrt{\delta'}/(2^{p+q+2}\sigma) \in (0, 1/2)$. From the assumption and Lemma 10, we have a unitary

$$U_{\text{obs}} := \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} |\mathbf{x}, y\rangle \langle \mathbf{x}, y| \otimes U_{\text{obs}}^{(\mathbf{x}, y)}, \quad (4.45)$$

where $U_{\text{obs}}^{(\mathbf{x}, y)}$ is a $(1, a'' = a + \lceil \log_2(M+1) \rceil + 1, \varepsilon')$ -block-encoding of the Hamiltonian

$$H^{(\mathbf{x}, y)} := \frac{1}{\sigma} \left(y O_{M+1} + \sum_{j=1}^M x_j O_j \right)$$

if $(\mathbf{x}, y) \in F$ with $|F| \geq (1 - \delta')2|G_p^M|$. Thus, using U_ψ , $U_\psi^\dagger$, and U_{obs} , we can implement a $(1, a'' + \log_2 d, 0)$ -block-encoding of the following Hamiltonian

$$\sum_{(\mathbf{x}, y) \in G_p^M \times G_1} \tilde{f}(\mathbf{x}, y) |\mathbf{x}, y\rangle \langle \mathbf{x}, y|, \quad \text{where} \quad \tilde{f}(\mathbf{x}, y) := \langle 0|^{\otimes a'' + \log_2 d} U_\psi^\dagger \cdot U_{\text{obs}}^{(\mathbf{x}, y)} \cdot U_\psi |0\rangle^{\otimes a'' + \log_2 d}. \quad (4.46)$$

Since $U_{\text{obs}}^{(\mathbf{x}, y)}$ is the ε' -precise block encoding of $H^{(\mathbf{x}, y)}$, $\tilde{f}(\mathbf{x}, y)$ is close to an affine linear function as

$$\left| \tilde{f}(\mathbf{x}, y) - \frac{1}{\sigma} \left(y \langle O_{M+1} \rangle + \sum_{j=1}^M x_j \langle O_j \rangle \right) \right| \leq \varepsilon' \quad \text{for all } (\mathbf{x}, y) \in F. \quad (4.47)$$

Thus, Grover-like repetition on Eq. (4.46), that is, Lemma 5 yields the quantum circuit W for an $(a'' + \log_2 d)$ -block-encoding of the following operator

$$\sum_{(\mathbf{x}, y) \in G_p^M \times G_1} T_t \left(\tilde{f}(\mathbf{x}, y) \right) |\mathbf{x}, y\rangle \langle \mathbf{x}, y| \quad \text{and} \quad t := 2^{p+q+2}\sigma.$$

Applying W to the state $|+\rangle^{\otimes pM+1} |\mathbf{0}\rangle$ with $|\mathbf{0}\rangle = |0\rangle^{\otimes a'' + \log_2 d}$, we obtain

$$W |+\rangle^{\otimes pM+1} |\mathbf{0}\rangle = \frac{\mathcal{N}_t}{\sqrt{2}} \left[\frac{1}{\mathcal{N}_t} \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x}, y} T_t \left(\tilde{f}(\mathbf{x}, y) \right) |\mathbf{x}, y\rangle \right] |\mathbf{0}\rangle + \sqrt{1 - \left(\frac{\mathcal{N}_t}{\sqrt{2}} \right)^2} |\Phi^\perp\rangle, \quad (4.48)$$

where $|\Phi^\perp\rangle$ denotes a normalized quantum state satisfying $(\mathbf{1} \otimes |\mathbf{0}\rangle \langle \mathbf{0}|) |\Phi^\perp\rangle = \mathbf{0}$ and $\mathcal{N}_t$ is defined as:

$$\mathcal{N}_t := \left\| \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x}, y} T_t \left(\tilde{f}(\mathbf{x}, y) \right) |\mathbf{x}, y\rangle \right\|. \quad (4.49)$$

Measuring the $(a'' + \log_2 d)$ -qubit ancilla registers in (4.48), we obtain the outcome $|\mathbf{0}\rangle$ and the following state with the probability $\mathcal{N}_t^2/2$:

$$\frac{1}{\mathcal{N}_t} \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x}, y} T_t \left(\tilde{f}(\mathbf{x}, y) \right) |\mathbf{x}, y\rangle. \quad (4.50)$$

In the following, we show this quantum state is close to the following state

$$\begin{aligned} & \frac{1}{\sqrt{2^{pM}}} \sum_{(\mathbf{x}, x_{M+1}) \in G_p^M \times G_1} \cos \left[t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right] |\mathbf{x}\rangle |x_{M+1}\rangle \\ &= \frac{1}{\sqrt{2}} \sum_{s=\pm 1} \sum_{\mathbf{x} \in G_p^M} \frac{1}{\sqrt{2^{pM}}} e^{2\pi i 2^p \sum_{j=1}^M (sx_j) \pi^{-1} 2^{q+1} \langle O_j \rangle} |\mathbf{x}\rangle \otimes \text{QFT}_{G_1} |s/4\rangle, \end{aligned} \quad (4.51)$$

where we recall that $|-1/4\rangle \equiv |0\rangle$ and $|1/4\rangle \equiv |1\rangle$ hold by the labelling in Eq. (3.21). Note that the equality in Eq. (4.51) can be shown as follows.

$$\begin{aligned} & \sum_{(\mathbf{x}, x_{M+1}) \in G_p^M \times G_1} \cos \left[t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right] |\mathbf{x}\rangle |x_{M+1}\rangle \\ &= \frac{1}{2} \sum_{\mathbf{x}, y} \sum_{s=\pm 1} e^{-is \frac{t}{\sigma} (y \langle O_{M+1} \rangle + \sum_{j=1}^M x_j \langle O_j \rangle)} |\mathbf{x}\rangle |y\rangle \quad (\because e^{ist\pi/2} = e^{2\pi i s 2^{p+q} \sigma} = 1) \\ &= \frac{1}{\sqrt{2}} \sum_{s=\pm 1} \sum_{\mathbf{x} \in G_p^M} e^{-ist \frac{1}{\sigma} \sum_{j=1}^M x_j \langle O_j \rangle} |\mathbf{x}\rangle \otimes \frac{1}{\sqrt{2}} \sum_{y \in G_1} e^{-ist \frac{1}{\sigma} y \langle O_{M+1} \rangle} |y\rangle \\ &= \frac{1}{\sqrt{2}} \sum_{s=\pm 1} \sum_{\mathbf{x} \in G_p^M} e^{-ist \frac{1}{\sigma} \sum_{j=1}^M x_j \langle O_j \rangle} |\mathbf{x}\rangle \otimes \text{QFT}_{G_1} |-s/4\rangle \quad (\because \langle O_{M+1} \rangle = \frac{\pi(1/4 + 4l)}{2^{p+q}}) \\ &= \frac{1}{\sqrt{2}} \sum_{s=\pm 1} \sum_{\mathbf{x} \in G_p^M} e^{-is 2\pi 2^p \sum_{j=1}^M x_j \pi^{-1} (2^{q+1}) \langle O_j \rangle} |\mathbf{x}\rangle \otimes \text{QFT}_{G_1} |-s/4\rangle. \end{aligned} \quad (4.52)$$

Thus, by applying $\text{QFT}_{G_1}^\dagger$ and controlled $X^{\otimes pM}$ gate to the state (4.48), we can prepare the target state:

$$\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^{q+1} \pi^{-1} \langle O_j \rangle} |\mathbf{x}\rangle \otimes |+\rangle \quad (4.53)$$

up to some Euclidean distance error with the probability $\mathcal{N}_t^2/2$, because $X^{\otimes pM} |\mathbf{x}\rangle = |-\mathbf{x}\rangle$ holds. Note that $\text{QFT}_{G_1}^\dagger$ and controlled $X^{\otimes pM}$ does not change the probability to obtain the success flag $|0\rangle$. Here, we show the corresponding quantum circuit to prepare Eq. (4.53) in Fig. 4.16.

For the subsequent proof, we here show that for any $\delta' \in (0, 1)$ and any real weight w_j ($j = 1, 2, \dots, M+1$), a subset F' of $G_p^M \times G_1$ defined as

$$F' := \left\{ (\mathbf{x}, y) \in G_p^M \times G_1 : \left| \sum_{j=1}^{M+1} w_j x_j \right| < \sqrt{\frac{\ln(2/\delta')}{2} \sum_{j=1}^{M+1} |w_j|^2} \right\} \quad (4.54)$$

has the cardinality $|F'| \geq (1 - \delta') 2^{pM+1}$. This fact can be proved from the Hoeffding's inequality as follows. Let us consider independent random variables X_j ($j = 1, 2, \dots, M$) and Y . Here, X_j are identically and uniformly distributed in G_p and Y is uniformly distributed in G_1 . Note that the random variables are upper bounded by $1/2$ from the definition of G_p . Therefore, from the Hoeffding's inequality, we obtain

$$\Pr_{X_1, \dots, X_M, Y} \left[\left| \sum_{j=1}^{M+1} w_j X_j \right| \geq c \right] \leq 2 \exp \left[-\frac{2c^2}{\sum_j |w_j|^2} \right]. \quad (4.55)$$

Thus, taking c as $c = \sqrt{\frac{\ln(2/\delta')}{2} \sum_{j=1}^{M+1} |w_j|^2}$, we obtain

$$\Pr_{X_1, \dots, X_M, Y} \left[\left| \sum_{j=1}^{M+1} w_j X_j \right| \geq \sqrt{\frac{\ln(2/\delta')}{2} \sum_{j=1}^{M+1} |w_j|^2} \right] \leq \delta'. \quad (4.56)$$

Since the independent random variables $X_1, \dots, X_M, Y$ has uniform distributions, F' has the cardinality $\geq 2^{pM+1}(1 - \delta')$.

Now, we are ready to show the closeness between Eq. (4.50) and Eq. (4.51). Defining F' for $\delta' = 2^{-14}$ and $w_j = \langle O_j \rangle$ as

$$F' := \left\{ (\mathbf{x}, y) \in G_p^M \times G_1 : \left| \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right| < \sqrt{\frac{\ln(2/\delta')}{2} \sum_{j=1}^{M+1} |\langle O_j \rangle|^2} \right\} \quad (4.57)$$

and recalling that $|F'|, |F| \geq (1 - \delta')2^{pM+1}$, we proceed as follows.

$$\begin{aligned} & \left\| \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} T_t(\tilde{f}(\mathbf{x}, y)) |\mathbf{x}\rangle |y\rangle - \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} \cos \left[t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right] |\mathbf{x}\rangle |y\rangle \right\|^2 \\ &= \sum_{(\mathbf{x}, x_{M+1}) \in G_p^M \times G_1} \left| T_t(\tilde{f}(\mathbf{x}, x_{M+1})) - \cos \left[t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right] \right|^2 \\ &\leq 4 \times 2 \times 2^{pM+1} \delta' + \sum_{(\mathbf{x}, x_{M+1}) \in F \cap F'} \left| t \cos^{-1} [\tilde{f}(\mathbf{x}, x_{M+1})] - t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right|^2, \quad (4.58) \end{aligned}$$

where we used $|\cos(a) - \cos(b)| \leq |a - b|$ and $G_p^M \times G_1 = (F \cap F') \cup (F^c \cup (F')^c)$ in the third line. Now, we focus on the case $(\mathbf{x}, x_{M+1}) \in F \cap F'$.

$$\begin{aligned} |\tilde{f}(\mathbf{x}, x_{M+1})| &\leq \varepsilon' + \left| \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right| \quad (\because \text{Eq. (4.47)}) \\ &\leq \varepsilon' + \sqrt{\frac{\sum_j |\langle O_j \rangle|^2}{4(M+1)}} \sqrt{\frac{\ln(2/\delta')}{\ln(2d/\delta')}} \quad (\because \text{Eq. (4.57)}) \\ &\leq \frac{\sqrt{\delta'}}{2^{p+q+2}\sigma} + \frac{1}{2^{q+2}} \sqrt{\frac{\ln(2/\delta')}{\ln(2d/\delta')}} < \frac{33/10}{2^{q+2}\sqrt{\ln(2d/\delta')}} < 1/4. \quad (4.59) \end{aligned}$$

In the case of $|x| \leq 1/4$, $|\cos^{-1}(x) - \pi/2 + x| \leq \frac{1}{5}|x|^3$ holds, and therefore we obtain

$$\begin{aligned} & \left| t \cos^{-1} [\tilde{f}(\mathbf{x}, x_{M+1})] - \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right| \\ &= \left| t \cos^{-1} [\tilde{f}(\mathbf{x}, x_{M+1})] - \frac{\pi}{2} + \tilde{f}(\mathbf{x}, x_{M+1}) + \left(\frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle - \tilde{f}(\mathbf{x}, x_{M+1}) \right) \right| \\ &\leq t\varepsilon' + t \left| \cos^{-1} [\tilde{f}(\mathbf{x}, x_{M+1})] - \frac{\pi}{2} + \tilde{f}(\mathbf{x}, x_{M+1}) \right| \\ &\leq t\varepsilon' + \frac{t}{5} \left| \frac{33/10}{2^{q+2}\sqrt{\ln(2d/\delta')}} \right|^3. \quad (4.60) \end{aligned}$$

From the assumption, the non-negative integer q satisfies $\frac{t}{5} \left| \frac{33/10}{2^{q+2}\sqrt{\ln(2d/\delta')}} \right|^3 \leq \sqrt{\delta'}$, then we conclude

that

$$\begin{aligned}
& \left\| \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} T_t(\tilde{f}(\mathbf{x}, y)) |\mathbf{x}\rangle |y\rangle - \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} \cos \left[t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right] |\mathbf{x}\rangle |y\rangle \right\| \\
& \leq \sqrt{4 \times 2 \times 2^{pM+1} \delta' + \sum_{(\mathbf{x}, x_{M+1}) \in F \cap F'} (2\sqrt{\delta'})^2} \\
& \leq 2\sqrt{6\delta'} \times \sqrt{2^{pM}}.
\end{aligned} \tag{4.61}$$

This leads to

$$1 - 2\sqrt{6\delta'} \leq \mathcal{N}_t = \left\| \frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x}, y} T_t(\tilde{f}(\mathbf{x}, y)) |\mathbf{x}, y\rangle \right\| \leq 1 + 2\sqrt{6\delta'} \tag{4.62}$$

and

$$\begin{aligned}
& \left\| \frac{1}{\mathcal{N}_t} \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} T_t(\tilde{f}(\mathbf{x}, y)) |\mathbf{x}\rangle |y\rangle - \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} \cos \left[t \left(\frac{\pi}{2} - \frac{1}{\sigma} \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right) \right] |\mathbf{x}\rangle |y\rangle \right\| \\
& \leq \sqrt{2^{pM}} \times \frac{4\sqrt{6\delta'}}{1 - 2\sqrt{6\delta'}} < \sqrt{2^{pM}} \times \frac{1}{12}.
\end{aligned} \tag{4.63}$$

Therefore, we can prepare the target state up to $1/12$ Euclidean distance error with the probability $\mathcal{N}_t^2/2 > 0.462$.

Finally, the gate complexity follows from that of Lemma 10 and Lemma 5.

□

Note that we can coherently amplify the success probability in Lemma 12 by quantum amplitude amplification [21], while this requires a quantum circuit with 3-fold depth compared to the quantum circuit (before measurement) in Fig. 4.16.

4.5.2 Maximum mean squared error (MSE) and query complexity

The proof of Theorem 10

Here, we evaluate the relation between the root mean squared error ε and the total queries to the state preparation in Algorithm 1*.

Proof of Theorem 10. The core idea of this proof is similar to the previous methods [97, 48, 47], but we need carefully to deal with the condition of gradient estimation. In the following, we prove this theorem based on the state preparation by Lemma 11; see remarks after this proof for the case of Lemma 12.

We start by clarifying the statistical property of $g_j^{(q)}$ in the step 5 of Algorithm 1*. The truncation in Step 7 guarantees $\tilde{u}_j^{(q)} \in [-1, 1]$ for all q ; we can construct a block-encoding U'_{SEL} for $\{\tilde{O}_j^{(q)}\}$, which is assumed to be accessible in Lemma 11, from the block encodings of $\{O_j\}$; see Fig. 4.14. Therefore, using the Lemma 11, we can prepare the quantum state in Step 4.

Now, we consider the case that the following condition for gradient estimation holds at the beginning of an iteration q : for all j ,

$$\left| \langle \tilde{O}_j^{(q)} \rangle \right| = \left| \frac{\langle O_j \rangle - \tilde{u}_j^{(q)}}{2} \right| \leq 2^{-q-1}. \tag{4.64}$$

Then, a single shot measurement result $\mathbf{k} := (k_1, \dots, k_M) \in G_p^M$ in Step 4 follows

$$\Pr \left[\left| k_j - \frac{2^q (\langle O_j \rangle - \tilde{u}_j^{(q)})}{\pi} \right| > \frac{3}{2^p} \right] \leq \frac{1}{3},$$

for every $j = 1, 2, \dots, M$, from the analysis of gradient estimation in Lemma 2. If we take $p = 5$, then the additive error $3/2^p$ is smaller than $1/2\pi$, and therefore, the temporal estimate $\tilde{u}_j^{(q)} + \frac{\pi}{2^q} k_j$ becomes

a 1-bit more precise estimate of $\langle O_j \rangle$ at least $2/3$ probability. However, the choice of $p = 5$ is sufficient but not tight. Here, we employ the following tighter bound that we numerically found²: if we take $p = 3$,

$$\Pr \left[\left| k_j - \frac{2^q(\langle O_j \rangle - \tilde{u}_j^{(q)})}{\pi} \right| > \frac{1}{2\pi} \right] < 0.18 + \frac{1}{12}, \quad (4.65)$$

holds for every j , where the term $1/12$ arises from the Euclidean distance error as well as Lemma 2. Since the $g_j^{(q)}$ is defined as the coordinate-wise median of independent samples $\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(\#)}$, the Hoeffding's inequality for the independent and bounded random variables $\{\chi[|k_j^{(i)} - 2^q(\langle O_j \rangle - \tilde{u}_j^{(q)})/\pi| > 1/(2\pi)]\}_i$ yields

$$\Pr \left[\left| g_j^{(q)} - \frac{2^q(\langle O_j \rangle - \tilde{u}_j^{(q)})}{\pi} \right| > \frac{1}{2\pi} \right] \leq e^{-\frac{\#}{9}}, \quad (4.66)$$

for every j . Here, $\chi[\bullet]$ denotes the indicator function. Therefore, taking $\# := 9 \ln(M/\delta^{(q)})$ and using the union bound, we can bound the probability of the event

$$\mathbf{A}^{(q)} : \max_j \left| g_j^{(q)} - 2^q(\langle O_j \rangle - \tilde{u}_j^{(q)})/\pi \right| \leq \frac{1}{2\pi}$$

as $\Pr[\mathbf{A}^{(q)}] \geq 1 - \delta^{(q)}$. In the event $\mathbf{A}^{(q)}$, $|\langle \tilde{O}_j^{(q+1)} \rangle| \leq 2^{-q-2}$ holds even if we truncate $\tilde{u}_j^{(q+1)}$ in Step 7. On the other hand, if the condition Eq. (4.64) is false, the gradient estimation does not work, and we only say that the measurement result $g_j^{(q)}$ is bounded as $g_j^{(q)} \in [-1/2, 1/2]$ because of the definition of G_p^M .

In the case of $q = 0$, the condition Eq. (4.64) holds, and thus, the event $\mathbf{A}^{(0)}$ occurs with the probability $1 - \delta^{(0)}$. By repeating this, in branches such that all of $\{\mathbf{A}^{(q)}\}_{q=0}^{q'-1}$ occur, the temporal estimate $\tilde{u}_j^{(q')}$ satisfy $|\tilde{u}_j^{(q')} - \langle O_j \rangle| \leq 1/2^{q'}$ for all j . Moreover, considering branches such that all of $\{\mathbf{A}^{(q)}\}_{q=0}^{q'-1}$ occur but the complement of $\mathbf{A}^{(q')}$ occurs at the iteration q' , we bound the additive error of the final estimate $\tilde{u}_j$ in such branches as follows:

$$|\tilde{u}_j - \langle O_j \rangle| \leq |\tilde{u}_j - \tilde{u}_j^{(q')}| + |\tilde{u}_j^{(q')} - \langle O_j \rangle| \leq \pi \left| \sum_{q \geq q'} \frac{g_j^{(q)}}{2^q} \right| + \frac{1}{2^{q'}} \leq \frac{1 + \pi}{2^{q'}}. \quad (4.67)$$

In the third line, we use the fact $|g_j^{(q)}| \leq 1/2$. Thus, we can calculate the mean squared error of $\hat{u}_j$ as

$$\mathbb{E} \left[(\hat{u}_j - \langle O_j \rangle)^2 \right] \leq \frac{1}{2^{2(q_{\max}+1)}} + (1 + \pi)^2 \sum_{q=0}^{q_{\max}} \frac{\delta^{(q)}}{4^q} \leq \frac{1}{2^{2(q_{\max}+1)}} + (1 + \pi)^2 c 2^{-2q_{\max}+1} \leq \varepsilon^2, \quad (4.68)$$

where we defined $q_{\max} := \lceil \log_2(1/\varepsilon) \rceil$. In the final line, we used the fact $c \in (0, 3/(8(1 + \pi)^2)]$. The inequality holds for all j , which completes the proof of Eq. (4.5).

Next, we count the total queries to the state preparation U_ψ . At each iteration q , we prepare $\# := 9 \ln(M/\delta^{(q)})$ copies of a quantum state that approximates the probing state $|\Upsilon(q)\rangle$. If we prepare a single copy of this state using the method in Lemma 11, then the $\mathcal{O}(2^{p+q+2}\sigma)$ queries to U_ψ and its inverse are required, and therefore the total queries are calculated as

$$\begin{aligned} & \sum_{q=0}^{q_{\max}} 2^{p+q+2}\sigma \times 9 \ln(M/\delta^{(q)}) = 9 \cdot 2^{p+2}\sigma \times \sum_{q=0}^{q_{\max}} 2^q \ln(M/\delta^{(q)}) \\ & = 9 \cdot 2^{p+2}\sigma \left[2^{q_{\max}+1} \ln \frac{8M}{c} - (q_{\max} + 2) \ln 8 - \ln \frac{M}{c} \right] \\ & = \mathcal{O}(\varepsilon^{-1} \sqrt{M} \log M). \end{aligned}$$

□

²More precisely, in the setup of Lemma 2 without any Euclidean distance error, we numerically evaluate the probability of an event $|k_j - g_j| > 1/2\pi$ for $g_j \in [-1/\pi, 1/\pi]$ when $p = 3$.

Here, we discuss the case that we employ Lemma 12 to the state preparation in Step 4 of Algorithm 1*. This alternative method can prepare the probing state $|\Upsilon(q)\rangle$ under the following conditions: (i) $|\langle \tilde{O}_j^{(q)} \rangle| \leq 2^{-q-1}$ and (ii) the iteration step q satisfies Eq. (4.44). In the case that only the condition (i) (or equivalently, Eq. (4.64)) is violated, the quantum circuit in Fig. 4.16 for Lemma 12 yields a pM -qubit quantum state that may be far from the target probing state. However, this does not affect the proof of Theorem 10 because it is only required that the measurement results at the end of Fig. 4.16 are in the range of $[-1/2, 1/2]$. On the other hand, the condition (ii) restricts the usage of the alternative method, as discussed in Sec. 4.4.

The proof of Theorem 11

Hereafter, we prove Theorem 11 by explicitly constructing a quantum algorithm. The quantum algorithm is the same as Algorithm 1* except for the Step 4. For Theorem 11, we use an improved version of Lemma 11 or Lemma 12, with a *known* parameter

$$\mathcal{B}_M \in \left[\left\| \sum_{j=1}^M O_j^2 \right\|, M \right] \quad (4.69)$$

for the state preparation in Step 4. At the level of quantum circuit, we modify the following in Fig. 4.15 for the improved Lemma 11:

- Replace $U_{\text{obs}}^{(\mathbf{x})}$ with a quantum circuit $V_{\text{obs}}^{(\mathbf{x})}$ for $(\sqrt{\delta'}/2^{p+q+1}\bar{\sigma})$ -precise block encoding of the Hamiltonian $(\bar{\sigma})^{-1} \sum_{j=1}^M x_j O_j$, where $\bar{\sigma}$ is defined as

$$\bar{\sigma} := \sqrt{2\mathcal{B}_M \ln(2d/\delta')} = \mathcal{O}(\sqrt{\mathcal{B}_M \log d}). \quad (4.70)$$

- Set the evolution time $t = 2^{p+q+1}\bar{\sigma}$
- Apply the following phase oracle for given $\tilde{u}_j^{(q)} \in [-1, 1]$ ($j = 1, 2, \dots, M$) at the end of circuit

$$|\mathbf{x}\rangle \mapsto e^{-2\pi i 2^p \sum_{j=1}^M x_j 2^q \pi^{-1} \tilde{u}_j^{(q)}} |\mathbf{x}\rangle = \bigotimes_{j=1}^M \left(e^{-i(2^{p+q+1}\tilde{u}_j^{(q)})x_j} |x_j\rangle \right), \quad (4.71)$$

which can be easily constructed due to the separable structure.

The total number of qubits is given by Eq. (4.82), where the additive constant 9 can be reduced to 7 because the block encoding of $\tilde{O}_j^{(q)}$ is unnecessary.

If the iteration step q in Algorithm 1* exceeds a new threshold specified below, we can also use an improved version of Lemma 12. In this case, the modification for the circuit in Fig. 4.16 is as follows.

- Replace $U_{\text{obs}}^{(\mathbf{x},y)}$ with a quantum circuit $V_{\text{obs}}^{(\mathbf{x},y)}$ for $(\sqrt{\delta'}/2^{p+q+2}\bar{\sigma}')$ -precise block encoding of the Hamiltonian $(\bar{\sigma}')^{-1} \sum_{j=1}^{M+1} x_j \tilde{O}_j^{(q)}$, where $\tilde{O}_{M+1}^{(q)} := O_{M+1}$ and $\bar{\sigma}'$ is defined as

$$\begin{aligned} \bar{\sigma}' &:= \left\lceil \sqrt{2(\mathcal{B}_M + 4^{-q-1}) \ln(2d/\delta')} + \frac{1}{2^{q+1}} \sqrt{2(M+1) \ln(2/\delta')} \right\rceil \\ &= \mathcal{O}(\sqrt{\mathcal{B}_M \log d}), \end{aligned} \quad (4.72)$$

where the final evaluation holds because of the threshold Eq. (4.74).

- Set $t = 2^{p+q+2}\bar{\sigma}'$.

Here, the total number of qubits is given by Eq. (4.83). This state preparation is available when the iteration step q satisfies (we recall $p = 3$ and $\delta' = 2^{-14}$)

$$\frac{t}{5} \left(\frac{1}{2^{q+2}\bar{\sigma}'} \left(\frac{\sqrt{\delta'}}{2^p} + \sqrt{2(M+1) \ln(2/\delta')} \right) \right)^3 \leq \sqrt{\delta'} \quad (4.73)$$

for ensuring the approximation error in the final state and

$$2^{q+1} \sqrt{2(\mathcal{B}_M + 4^{-q-1}) \ln(2d/\delta')} > \sqrt{2(M+1) \ln(2/\delta')} \quad (4.74)$$

for $\bar{\sigma}' = \mathcal{O}(\sqrt{\mathcal{B}_M \log d})$.

Because the final state in Figs. 11 or 4.16 after these modifications is $1/12$ -close to the target state Eq. (4.1) in the Euclidean distance as proved below, we complete the Step 4 by measuring the final $3M$ -qubit state by Fourier basis (see Sec. 4.5.3). Thus, the same analysis on MSE in the proof of Theorem 10 holds, and as for the query complexity, we simply replace σ with $\bar{\sigma}$ or $\bar{\sigma}'$. Consequently, the total number of queries to U_ψ and $U_\psi^\dagger$ is given by $\mathcal{O}(\varepsilon^{-1} \sqrt{\mathcal{B}_M \log d} \log M)$.

In the following, we show the correctness of these modifications. First of all, Lemma 10 holds when we replace σ with $\bar{\sigma}$ from the discussion after the proof of Lemma 10. The same replacement is available in Lemma 11, so we can prepare the pM -qubit quantum state

$$\frac{1}{\sqrt{2^{pM}}} \sum_{\mathbf{x} \in G_p^M} e^{2\pi i 2^p \sum_{j=1}^M x_j 2^q \pi^{-1} \langle O_j \rangle} |\mathbf{x}\rangle \quad (4.75)$$

up to $1/12$ Euclidean distance error, along with the parameter $Q = \mathcal{O}(2^{p+q+1} \bar{\sigma})$ in the statement of Lemma 11. Thus, applying the phase oracle Eq. (4.71), we obtain the target state Eq. (4.1). Note that the Euclidean distance error remains unchanged under the unitary transformation.

Under the same assumption in Lemma 12, a subset $F' \subset G_p^M \times G_1$ defined as Eq. (4.57) with $\tilde{O}_j^{(q)}$ instead of O_j satisfies $|F'| \geq (1 - \delta') |G_p^M \times G_1|$. Also, we have another subset $F \subset G_p^M \times G_1$ with $|F| \geq (1 - \delta') |G_p^M \times G_1|$, and in this subset, we obtain

$$\left\| \sum_{j=1}^{M+1} x_j O_j \right\| \leq \frac{1}{2} \sqrt{2(\mathcal{B}_M + 4^{-q-1}) \ln(2d/\delta')}. \quad (4.76)$$

This follows from the Lemma 10 and the discussion after this lemma. By combining the above observations, we can say that if $(\mathbf{x}, y) \in F \cap F'$, then

$$\begin{aligned} & \left\| \sum_{j=1}^M x_j \tilde{u}_j^{(q)} \mathbf{1} - y O_{M+1} \right\| \leq \left| \sum_{j=1}^M x_j \tilde{u}_j^{(q)} - y \frac{\pi(1/4 + 4l)}{2^{p+q}} \right| \\ &= \left| \sum_{j=1}^{M+1} x_j \langle O_j \rangle - 2 \sum_{j=1}^{M+1} x_j \langle \tilde{O}_j^{(q)} \rangle \right| \leq \left| \sum_{j=1}^{M+1} x_j \langle O_j \rangle \right| + \sqrt{2 \ln(2/\delta') \sum_{j=1}^{M+1} \langle \tilde{O}_j^{(q)} \rangle^2} \\ &\leq \left\| \sum_{j=1}^{M+1} x_j O_j \right\| + \frac{1}{2} \frac{1}{2^q} \sqrt{2(M+1) \ln(2/\delta')} \end{aligned} \quad (4.77)$$

and thus,

$$\begin{aligned} & \left\| \sum_{j=1}^{M+1} x_j \tilde{O}_j^{(q)} \right\| \leq \frac{1}{2} \left\| \sum_{j=1}^{M+1} x_j O_j \right\| + \frac{1}{2} \left\| \sum_{j=1}^M x_j \tilde{u}_j^{(q)} \mathbf{1} - y O_{M+1} \right\| \\ &\leq \left\| \sum_{j=1}^{M+1} x_j O_j \right\| + \frac{1}{2} \frac{1}{2^{q+1}} \sqrt{2(M+1) \ln(2/\delta')} \leq \frac{\bar{\sigma}'}{2}. \end{aligned} \quad (4.78)$$

Therefore, by using Lemma 4 for a quantum circuit for the block encoding of $\propto \sum_j x_j \tilde{O}_j^{(q)}$, we obtain

$$V_{\text{obs}} = \sum_{(\mathbf{x}, y) \in G_p^M \times G_1} |\mathbf{x}, y\rangle \langle \mathbf{x}, y| \otimes V_{\text{obs}}^{(\mathbf{x}, y)}, \quad (4.79)$$

where $V_{\text{obs}}^{(\mathbf{x}, y)}$ is $(\sqrt{\delta'}/2^{p+q+2} \bar{\sigma}')$ -precise block encoding of

$$\frac{1}{\bar{\sigma}'} \left(y \tilde{O}_{M+1}^{(q)} + \sum_{j=1}^M x_j \tilde{O}_j^{(q)} \right) \equiv \frac{1}{\bar{\sigma}'} \sum_{j=1}^{M+1} x_j \tilde{O}_j^{(q)}, \quad (4.80)$$

if $(\mathbf{x}, y) \in F \cap F'$. Then, replacing the U_{obs} and $t = 2^{p+q+2}\sigma$ with V_{obs} and $t = 2^{p+q+2}\sigma'$, respectively, in the circuit Fig. 4.16, we can prepare the target state $|\Upsilon(q)\rangle|+\rangle$ *approximately* after we obtain all 0 outcomes in the other registers of the circuit Fig. 4.16.

To ensure that the approximation error is at most $1/12$ in the Euclidean distance, the non-negative integer q needs to satisfy Eq. (4.73). We obtain this condition in a similar way to prove Eqs. (4.59) and (4.60). Also, this condition is roughly evaluated as

$$2^{q+1}\sqrt{2(\mathcal{B}_M + 4^{-q-1})\ln(2d/\delta')} \geq \mathcal{O}\left(M^{\frac{3}{4}}\right) \quad (4.81)$$

which is asymptotically tighter than the other condition Eq. (4.74).

4.5.3 Circuit implementation

We first show quantum circuits for Lemma 10. For clarification, input oracle unitary gates are provided in Fig. 4.8. Final output circuit of Lemma 10 is given by Fig. 4.13, which corresponds to U_{obs} in Eq. (4.35). In Figs. 4.9–4.12, we describe an efficient circuit construction of (a slightly modified version of) U'_{SEL} with a single use of U_{SEL} ; see Remark 1.

We give quantum circuits for preparing $|\Upsilon(q)\rangle$ via Hamiltonian simulation protocol Lemma 11. The full circuit is shown in Fig. 4.15. In particular, the gate $U_{\text{obs}}^{(\mathbf{x})}$ and its inverse contained in this figure are decomposed as in Fig. 4.13, where $\mathbf{p} = (3, \dots, 3)$. When we use the circuits in Figs. 4.9–4.12 for U'_{SEL} , the total number of qubits in the full circuit Fig. 4.15 is given by

$$3M + \lceil \log_2 M \rceil + \log_2 d + a + 9. \quad (4.82)$$

Note that each B_j in Fig. 4.9 is replaced by the circuit in Fig. 4.14.

Then, we show quantum circuits for preparing $|\Upsilon(q)\rangle$ via Grover-like repetition Lemma 12. In this case, the full quantum circuit is shown in Fig. 4.16, which is simpler than that of Hamiltonian-simulation based protocol. When we construct $U_{\text{obs}}^{(\mathbf{x}, y)}$ by the circuits in Figs. 4.9–4.14, the total number of qubits in the full circuit Fig. 4.16 is calculated as

$$3M + \lceil \log_2(M + 1) \rceil + \log_2 d + a + 8. \quad (4.83)$$

Finally, we make a remark on Step 3 in Algorithm 1 (or Step 4 in Algorithm 1*). To complete this step, we simply measure only the $3M$ qubits of Fig. 4.15 in the computational basis after inverse QFTs. This step does not require any post-selection because the final output state without post-selection is ensured to be $1/12$ -close to the separable state $|\mathbf{0}\rangle|\Upsilon(q)\rangle$. Similarly, we can remove the post-selection in Lemma 12 by using a single-step amplitude amplification and adjusting δ' to ensure the closeness condition on a final output state.

4.6 Numerical simulation

To evaluate the performance of observables estimation methods based on quantum gradient estimation, we here consider the following simple problem. Let U_ψ be a 1-qubit unitary that prepares a target state $|\psi\rangle$. As for target observables $\{O_j\}_{j=1}^M$ with the spectral norm $\|O_j\| \leq 1$, we put the following assumptions for the sake of efficient classical simulation: $[O_i, O_j] = 0$ holds for any pair of i, j , and the target state $|\psi\rangle$ is a common eigenstate of the observables $\{O_j\}_{j=1}^M$, that is,

$$O_j |\psi\rangle = g_j |\psi\rangle \quad (4.84)$$

holds for $j = 1, 2, \dots, M$ and eigenvalues $g_j \in [-1, 1]$. The goal is to estimate the expectation values $\langle O_j \rangle := \langle \psi | O_j | \psi \rangle = g_j$ with a root MSE ε for all $j = 1, 2, \dots, M$. In the following, we show the number of ancilla qubits and total queries to U_ψ and its inverse in our method and the previous method [86], for solving the above problem with $M = 30$. For more details on the simulation setup, please see Sec. 4.6.1.

Figure 4.5 shows the number of additional ancilla qubits (except for target system qubits) in our method and the previous method [86]. We numerically calculate the number of ancilla qubits for the previous method, which is well approximated by

$$\lceil \log_2(24/\varepsilon_{\text{add}}) \rceil \times M. \quad (4.85)$$

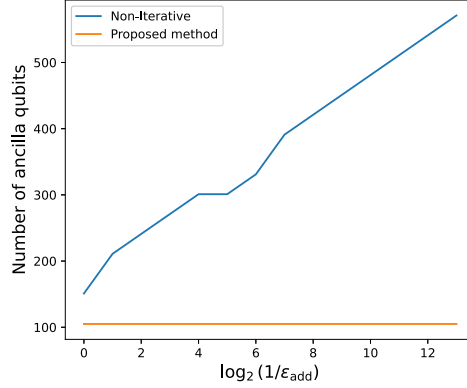


Figure 4.5: The number of additional ancilla qubits for $M = 30$. The ancilla count in the proposed adaptive method does not depend on target precision ϵ_{add} (more precisely, target root MSE ϵ), unlike the previous one [86]. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

As for our method, the total number of qubits is explicitly calculated as

$$3M + \lceil \log_2 M \rceil + \log_2 d + a + 9 \quad (4.86)$$

when we use the method based on Lemma 8; the multiplicative constant hidden in $\mathcal{O}(M)$ is exactly given by 3. Note that in the figure, we ignore the factor a for a block encoding of a single observable in our method and ancilla qubits that might be required for implementing Hamiltonian simulation for given observables in the previous method. This figure clearly shows that our adaptive method has a constant ancilla overhead with respect to the target precision, unlike the non-iterative method.

Next, we show the relationship between the total queries to state preparation U_ψ (and its inverse) and the resulting root MSE in Fig. 4.6. The number of total queries in our method is based on an upper bound of actually required queries, which is derived in the proof of Theorem 10. As for the previous method which takes an additive error ϵ_{add} and a failure probability δ as input, we evaluate the resulting total queries and root MSEs by changing input pairs $(\epsilon_{\text{add}}, \delta)$. More precisely, we plot the dots with the same colors for a fixed $\epsilon_{\text{add}} = 2^{-a}$ ($a = -3, -2, \dots, 13$) and various δ , by estimating the root MSE of a single observable $\langle O_1 \rangle$ and calculating the corresponding total queries $\tilde{T}_{\text{NonIter}}$.

We remark that $\tilde{T}_{\text{NonIter}}$ in the previous method represents the actual number of queries divided by a factor (> 1) that rapidly decreases and converges to a constant as $1/\epsilon_{\text{add}}$ increases, in order to highlight the asymptotic behavior. When including the asymptotically negligible factor (> 1), all the colored dots shift above our horizontal line that represents a theoretical upper bound; see Fig. 4.7 in Sec. 4.6.1 for the worst-case and average-case results without any rescaling.

Our method achieves the Heisenberg-limited scaling in MSE; the line for our method is horizontal i.e., $\epsilon T = \mathcal{O}(1)$ in Fig. 4.6. On the other hand, in the previous method, the sequence of minimal values of the same-colored dots shows the (asymptotic) scaling of $\Omega(\log^2 T \log \log T)$ for the total queries T . This indicates that the Heisenberg-limited scaling with a logarithmic overhead is achieved by carefully choosing δ or N_{med} for a final median calculation. When we take a single shot outcome from a circuit as a final estimate ($N_{\text{med}} = 1$), the previous method has the scaling of SQL.

We remark that the previous method has $\mathcal{O}(\log(M/\delta))$ factor in the total query because it takes the median of independent $N_{\text{med}} = \mathcal{O}(\log(M/\delta))$ samples as a final estimate. Naively, in order to achieve MSE ϵ^2 , we can take $\delta = \mathcal{O}(\epsilon^2)$ [108]. Such a $N_{\text{med}} = \mathcal{O}(\log(1/\epsilon))$ (for a fixed M) factor might not be tight. However, Fig. 4.6 also indicates that the logarithmic correction $N_{\text{med}} = \mathcal{O}(\log(1/\epsilon))$ is required, as follows. Figure 4.6 indicates the following lower bound for the (rescaled) total number of queries: $\epsilon \tilde{T}_{\text{NonIter}} = \Omega(\log^2(\tilde{T}_{\text{NonIter}}) \log \log(\tilde{T}_{\text{NonIter}}))$, or

$$\tilde{T}_{\text{NonIter}} = \Omega(\epsilon^{-1} \log^2(1/\epsilon) \log \log(1/\epsilon)).$$

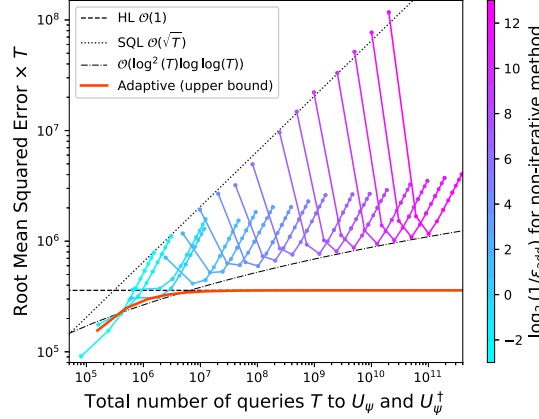


Figure 4.6: The comparison between the total number of queries and the root MSE. The orange line represents a theoretical upper bound in our method. The colored dots show the worst-case results of the previous non-iterative method [86] among randomly-generated 26 sets of $\{g_j\}$, with various input pairs $(\varepsilon_{\text{add}}, \delta)$. For each color, we fix ε_{add} as in the color bar and vary the failure probability $\delta = 2^{-j}$ ($j = 0, 1, \dots$), which gives the number of samples $N_{\text{med}} = 2\lceil \log(1/\delta) \rceil + 1$ for a final median calculation. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

Therefore, by comparing an upper bound $\tilde{T}_{\text{NonIter}} = N_{\text{med}} \times \mathcal{O}(\varepsilon_{\text{add}}^{-1} \log(1/\varepsilon_{\text{add}}) \log \log(1/\varepsilon_{\text{add}}))$ [56] with the above empirical lower bound, we consider $N_{\text{med}} = \mathcal{O}(\log(1/\varepsilon))$ to be tight for a given root MSE ε .

Finally, these two algorithms cannot work in the query regime $< 10^5$ under the current problem setup. (Actually, the query limit is more severe in the previous method; see Fig. 4.7.) This is attributed to a large constant factor in the total number of queries; even if we want to obtain estimates within ~ 1 error, these algorithms require a much larger queries to run.

4.6.1 More details

To solve the problem in Sec. 4.6, our adaptive Algorithm 1 with Lemma 8 uses

$$T_{\text{adapt}} := \sum_{q=0}^{q_{\text{max}}} 2Q(q) \times 9 \ln \left(\frac{M}{\delta^{(q)}} \right) = \mathcal{O}(\varepsilon^{-1} \sqrt{M} \log M) \quad (4.87)$$

queries of U_ψ and its inverse in total, where $q_{\text{max}} := \lceil \log_2(1/\varepsilon) \rceil$ for a target root MSE ε ; see the proof of Theorem 10 in Sec. 4.5.2. Here, $Q(q)$ is the minimal integer Q satisfying $4t^Q/(2^Q Q!) \leq 2^{-17}$ for the total evolving time $t = 2^{5+q} \sqrt{2M \ln(2^{11}d)}$, and we use an explicit upper bound $Q(q) < 1.5t + 126$ for the numerical simulation. It should be noted that T_{adapt} itself is an upper bound of total queries in Algorithm 1 to theoretically ensure root MSE $\leq \varepsilon$ for every target observables, and thus the required queries might be smaller than the upper bound in practice. In addition, from the entire circuit diagrams in Sec. 4.5.3, the total number of qubits required for quantum circuits in our algorithm is calculated as Eqs. (4.82) and (4.83).

For comparison, we briefly review the previous method in Ref. [86]. The method essentially uses the gradient estimation method [56] that can be summarized as the following theorem.

Theorem 12 (Theorem 25 in Ref. [56]). *Let $\mathbf{x} \in \mathbb{R}^M$, $\varepsilon_{\text{add}} \leq c \in \mathbb{R}_+$ be fixed constants and suppose $f : \mathbb{R}^M \rightarrow \mathbb{R}$ is analytic and satisfies the following: for every $k \in \mathbb{N}$ and $\alpha \in [M]^k$, k -th order partial derivatives of f at $\mathbf{x}$ (denoted by $\partial_\alpha f(\mathbf{x})$) are upper bounded as $|\partial_\alpha f(\mathbf{x})| \leq c^k 2^{k/2}$. Then, there is a quantum algorithm that works for all such functions, and outputs an ε_{add} -approximate gradient $\tilde{\mathbf{g}} \in \mathbb{R}^M$ such that $\|\nabla f(\mathbf{x}) - \tilde{\mathbf{g}}\|_\infty \leq \varepsilon_{\text{add}}$ with probability at least $1 - \delta$, using $\mathcal{O}(c\sqrt{M} \log(M/\delta)/\varepsilon_{\text{add}})$ queries to the phase oracle O_f .*

Then, taking the function f in this theorem as

$$f(\mathbf{x}) := \frac{1}{2} - \frac{1}{2} \text{Im} \left[\langle \psi | \prod_{j=1}^M e^{-2ix_j O_j} | \psi \rangle \right], \quad (4.88)$$

we can estimate the expectation values of target observables because $\nabla f(\mathbf{0})$ is the collection of target expectation values. Note that this function for observables with $\|O_j\| \leq 1$ satisfies the condition in Theorem 12 by taking $c = 2$.

For circuit implementation, Ref. [86] first provides an explicit quantum circuit for the *probability* oracle to encode the function Eq. (4.88). Here, the probability oracle U_f for a function f is defined as

$$U_f : |\mathbf{x}\rangle |0\rangle \mapsto |\mathbf{x}\rangle \left(\sqrt{f(\mathbf{x})} |1\rangle |\phi_1(\mathbf{x})\rangle + \sqrt{1-f(\mathbf{x})} |0\rangle |\phi_0(\mathbf{x})\rangle \right) \quad (4.89)$$

for all $\mathbf{x} \in G_n^M$ (n is a positive integer), where $|\phi_0(\mathbf{x})\rangle$ and $|\phi_1(\mathbf{x})\rangle$ are arbitrary (normalized) quantum states. The quantum circuit for the probability oracle has a single use of the state preparation oracle U_ψ , controlled time evolutions for target observables, and additional 1 ancilla qubit except for the target system qubits. Then, we obtain the phase oracle for the function Eq. (4.88) from the probability oracle.

Theorem 13 (Theorem 14 in Ref. [56]). *Let U_f be a probability oracle for a function $f(\mathbf{x})$. For any $\epsilon \in (0, 1/3)$, we can implement an ϵ -approximate phase oracle $\tilde{O}_f$ such that*

$$\left\| \tilde{O}_f |\Psi\rangle |0\rangle - O_f |\Psi\rangle |0\rangle \right\| \leq \epsilon \quad (4.90)$$

for all input states $|\Psi\rangle$, where O_f denotes an exact phase oracle for f . This implementation uses $\mathcal{O}(\log(1/\epsilon))$ invocations of the probability oracle U_f or its inverse and $\mathcal{O}(\log \log(1/\epsilon))$ additional ancilla qubits beyond those required by U_f .

In the numerical simulation, we assume the oracle conversion (from the probability oracle to a fractional phase oracle) is perfect $\epsilon = 0$ and ignore the overhead in space complexity, for simplicity. As for the query complexity, we only consider the overhead 10 (for the 2-round oblivious amplitude amplification), while the actual overhead is larger.

Given the phase oracle (and its fractional version) for the function Eq. (4.88), the estimation method in Theorem 12 effectively constructs a phase oracle for the following function

$$h(\mathbf{x}) := \frac{1}{2^{\lceil \log_2(3cr) \rceil}} \sum_{l=-m}^m a_l^{(2m)} f(l\mathbf{x}) \simeq \sum_{j=1}^M \frac{r \langle O_j \rangle}{2^{\lceil \log_2(3cr) \rceil}} x_j + \dots, \quad (4.91)$$

where $a_l^{(2m)} = 0$ ($l = 0$) and for $l \neq 0$,

$$a_l^{(2m)} := \frac{(-1)^{l-1}}{l} \frac{\binom{m}{|l|}}{\binom{m+|l|}{|l|}}. \quad (4.92)$$

Here, $m = \lceil \log(c\sqrt{M}/\varepsilon_{\text{add}}) \rceil$, and r is a rescaled parameter defined as

$$r^{-1} := 9cm\sqrt{M} \left(81 \cdot 8 \cdot 42\pi m \frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right)^{1/2m}. \quad (4.93)$$

This function h is closer to a linear function in the grid G_n^M than the original function f , due to the higher-order central-difference formula. Using the (effective) phase oracle O_h on G_n^M , we can estimate the gradient of h at $\mathbf{x} = 0$, i.e., the rescaled expectation values $r \langle O_j \rangle / 2^{\lceil \log_2(3cr) \rceil}$. More precisely, measuring the following state by the Fourier basis, we obtain estimates $\mathbf{k} = (k_1, \dots, k_M) \in G_n^M$ for the gradient:

$$\frac{1}{\sqrt{2^{nM}}} \sum_{\mathbf{x} \in G_n^M} e^{2\pi i 2^n h(\mathbf{x})} |\mathbf{x}\rangle. \quad (4.94)$$

Then, we can estimate $\langle O_j \rangle$ by $(2^{\lceil \log_2(3cr) \rceil} / r) k_j$. According to Ref. [56] (see Theorem 21 of Ref. [56]), we take the number of qubits n for the 1 dimensional grid G_n as

$$n = \lceil \log_2(4/\varepsilon_{\text{add}} r) \rceil + \lceil \log_2(3cr) \rceil \quad (4.95)$$

to successfully bound the estimation error. Thus, the total number of qubits required for quantum circuits in this method is given by

$$(\lceil \log_2(4/\varepsilon_{\text{add}} r) \rceil + \lceil \log_2(3cr) \rceil) M + 1 + \log_2 d. \quad (4.96)$$

To enhance the success probability to $1 - \delta$, we take the coordinate-wise median of independent $N_{\text{med}} = \mathcal{O}(\log(M/\delta))$ samples $\mathbf{k}^{(1)}, \mathbf{k}^{(2)} \dots$ from the gradient estimation procedure as a final estimate. As a result, the entire procedure for the observable estimation based on the Theorem 12 uses

$$\begin{aligned} T_{\text{NonIter}} &:= N_{\text{med}} \times \sum_{l=-m}^m \left[2\pi \cdot 2^{\lceil \log_2(4/\varepsilon_{\text{add}} r) \rceil} \cdot |a_l^{(2m)}| \right] \\ &= N_{\text{med}} \times \mathcal{O} \left(\frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \log \left(\frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right) \log \log \left(\frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right) \right) \end{aligned} \quad (4.97)$$

queries to the state preparation U_ψ and its inverse (up to the oracle conversion overhead in Theorem 13); the derivation of this query count is provided in the proof of Theorem 25 in Ref. [56].

To evaluate MSE, we need full information regarding the output probability distribution for the Fourier-basis measurement on the state Eq. (4.94). However, direct calculation of the distribution is difficult because of many qubits $nM = \mathcal{O}(M \log(1/\varepsilon_{\text{add}}))$. To overcome this, we perform the following numerical simulation. Focusing on the top registers related to the estimation for the 1st gradient in Eq. (4.94), the reduced state can be written as

$$\sum_{(x_2, \dots, x_M) \in G_n^{M-1}} \frac{1}{2^{n(M-1)}} |\Psi(x_2, \dots, x_M)\rangle \langle \Psi(x_2, \dots, x_M)|, \quad (4.98)$$

where

$$|\Psi(x_2, \dots, x_M)\rangle = \frac{1}{\sqrt{2^n}} \sum_{x_1 \in G_n} e^{2\pi i 2^n h(x_1, x_2, \dots, x_M)} |x_1\rangle. \quad (4.99)$$

Therefore, we can estimate the full probability distribution $p(k_1)$ for the Fourier-basis measurement on the x_1 registers by

$$p(k_1) \simeq \frac{1}{\#} \sum_{i=1}^{\#} \left| \langle k_1 | \text{QFT}_{G_n}^\dagger |\Psi(x_2^{(i)}, \dots, x_M^{(i)})\rangle \right|^2, \quad (4.100)$$

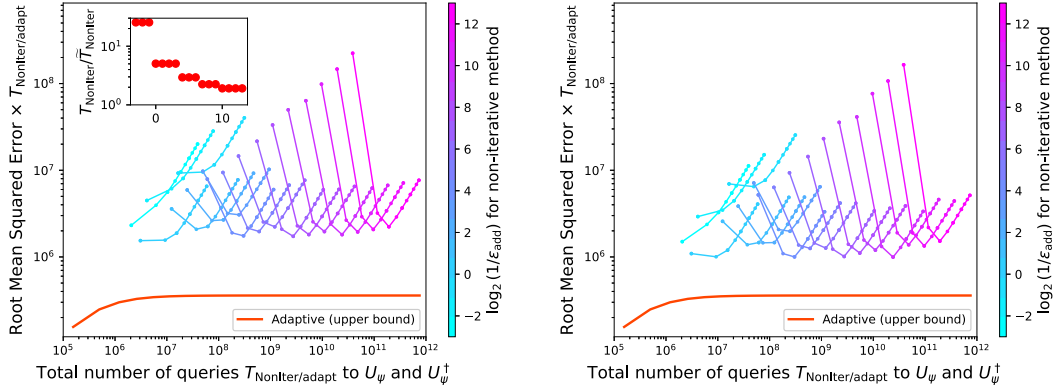
where $i = 1, 2, \dots, \#$ denotes the index of trials and $(x_2^{(i)}, \dots, x_M^{(i)})$ is independently and uniformly randomly sampled over G_n^{M-1} . Here, we take $\# = 10000$ for the numerical simulation in this work.

Given the probability distribution of a single shot $p(k_1)$, we can calculate the probability distribution of median k^{med} calculated by (an odd number) N_{med} independent samples of k_1 , from the theory of order statistics [144]. The explicit form of the median distribution is given by

$$\Pr(k^{\text{med}} = x_i) = \sum_{l=(N_{\text{med}}+1)/2}^{N_{\text{med}}} \binom{N_{\text{med}}}{l} \{P_i^l (1 - P_i)^{N_{\text{med}}-l} - P_{i-1}^l (1 - P_{i-1})^{N_{\text{med}}-l}\}, \quad (4.101)$$

where $G_n = \{x_i\}_{i=1}^{2^n}$ for $x_i = (2i-1)/2^{n+1} - 1/2$, and P_i is the cumulative distribution $P_i := \Pr(k_1 \leq x_i)$, $P_0 = 0$. Using the median distribution, we can calculate MSE of estimates in the observable estimation method. The corresponding number of total queries is given by Eq. (4.97).

To confirm the tightness of $N_{\text{med}} = \mathcal{O}(\log(1/\varepsilon))$ (M is fixed) for achieving MSE ε^2 , we have focused on a modified factor $\tilde{T}_{\text{NonIter}}$ in Fig. 4.6. In the following, we describe how the modification on T_{NonIter}



(a) Worst-case for the non-iterative method

(b) Average-case for the non-iterative method

Figure 4.7: The comparison between the total number of queries (given by T_{adapt} and T_{NonIter} including the oracle-conversion overhead) and the root MSE. The colored dots show the (a) worst-case or (b) average-case results of the previous non-iterative method [86] among randomly-generated 26 sets of $\{g_j\}$, with various input pairs $(\varepsilon_{\text{add}}, \delta)$. For each color, we fix ε_{add} as in the color bar and vary the failure probability $\delta = 2^{-j}$ ($j = 0, 1, \dots$), which gives the number of samples $N_{\text{med}} = 2\lceil \log(1/\delta) \rceil + 1$ for a final median calculation (see Eq. (4.97)). In the upper left subplot of (a), we show the value of rescaling in Eq. (4.104), where the horizontal axis is $\log_2(1/\varepsilon_{\text{add}})$. The orange line for our adaptive method represents an upper bound. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

was performed. From Eqs. (4.93) and (4.97), we can obtain

$$T_{\text{NonIter}} = N_{\text{med}} \times \sum_{l=-m}^m \left[2\pi \cdot 2^{\lceil \log_2(4/\varepsilon_{\text{add}} r) \rceil} \cdot |a_l^{(2m)}| \right] \sim N_{\text{med}} \times \frac{cm\sqrt{M}}{\varepsilon_{\text{add}}} \sum_{l=-m}^m |a_l^{(2m)}| \times \left(81 \cdot 8 \cdot 42\pi m \frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right)^{1/2m} \quad (4.102)$$

$$\sim N_{\text{med}} \times \frac{c\sqrt{M}}{\varepsilon_{\text{add}}} m \log(m) \times \left(81 \cdot 8 \cdot 42\pi m \frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right)^{1/2m} \quad (4.103)$$

where we keep only dominant factors in the second and third lines. As m increases, the factor $\left(81 \cdot 8 \cdot 42\pi m \frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right)^{1/2m}$ decreases and converges to a constant, while the other terms increase. Thus, to mitigate the effect of the contrary and asymptotically negligible factor, we define $\tilde{T}_{\text{NonIter}}$ as

$$\tilde{T}_{\text{NonIter}} := \frac{1}{(81 \cdot 8)^{1/2m}} T_{\text{NonIter}} = N_{\text{med}} \times \mathcal{O} \left(\frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \log \left(\frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right) \log \log \left(\frac{c\sqrt{M}}{\varepsilon_{\text{add}}} \right) \right). \quad (4.104)$$

In Fig. 4.7, we show a similar plot as Fig. 4.6 using the total number of queries T_{NonIter} with no rescaling, instead of $\tilde{T}_{\text{NonIter}}$.

While we have discussed the comparison with Ref. [86], we note that another previous method in Ref. [159] follows a similar estimation procedure with measurements on a uniform superposition state evolved under a phase oracle. In particular, it also uses the median trick essentially to obtain final estimates for expectation values. Similar to the numerical simulation suggesting $N_{\text{med}} = \Omega(\log(1/\varepsilon))$ in order to achieve $\text{MSE} \leq \varepsilon^2$, such a logarithmic correction regarding ε in the total queries might be *necessary* in the method of Ref. [159].

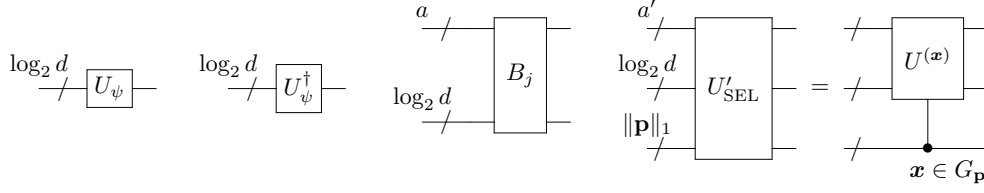


Figure 4.8: Input oracles. (Left) U_ψ prepares a target state $|\psi\rangle$ from an initial state $|0\rangle$. (Middle) For any $j \in \{1, 2, \dots, M\}$, B_j is an a -block-encoding of observable O_j for a nonnegative integer a . (Right) $U^{(\mathbf{x})}$ is a block encoding of $\tilde{H}^{(\mathbf{x})} := M^{-1} \sum_{j=1}^M x_j O_j$ with additional $a' := \mathcal{O}(\log M + a)$ qubits. For simplicity, we often assume U'_{SEL} to be an oracle for target observables $\{O_j\}$. Note that we can construct (a slightly modified) U'_{SEL} from B_j with arbitrary precision; see Remark 1 and Figs. 4.9–4.12. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

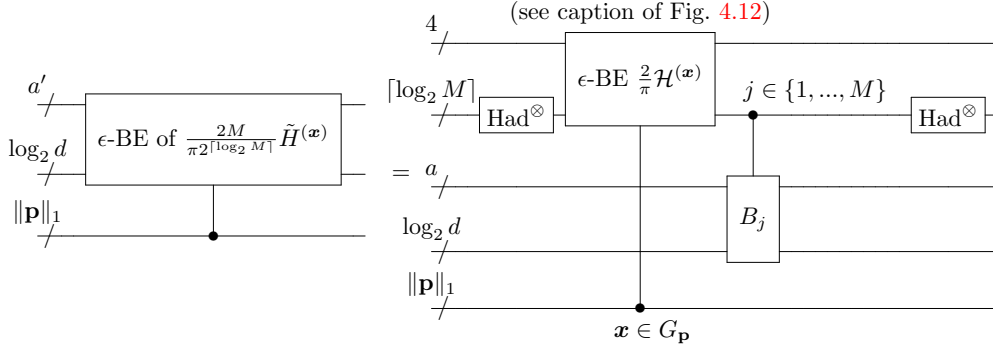


Figure 4.9: Quantum circuit for a slightly modified version of U'_{SEL} , where $a' := 4 + \lceil \log_2 M \rceil + a$. For any $\mathbf{x} \in G_{\mathbf{p}}$, the circuit (except for the bottom line) is a $(1, 4 + \lceil \log_2 M \rceil + a, \epsilon < 1/2)$ -block-encoding of $(2M/\pi 2^{\lceil \log_2 M \rceil}) \tilde{H}^{(\mathbf{x})}$. Had $^{\otimes}$ means Had $^{\otimes \lceil \log_2 M \rceil}$. Adapted from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

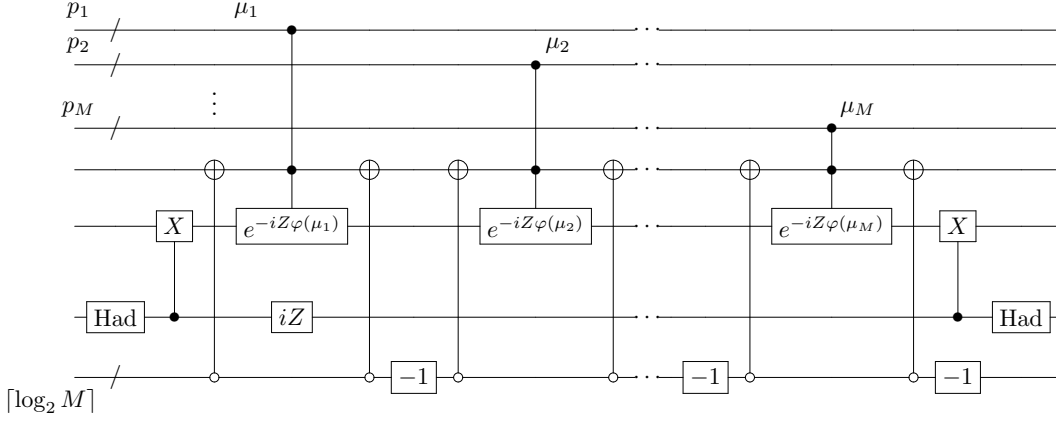


Figure 4.12: Quantum circuit for 3-block-encoding of $\sin \mathcal{H}^{(x)}$. Substituting this circuit and its inverse into U and $U^\dagger$, respectively, in the circuit of Fig. 3.7 (a is taken as 3) and setting $\mathcal{O}(\log(1/\epsilon))$ circuit parameters $\{\phi_i\}$ for $P(x) \approx (2/\pi) \arcsin(x)$, we obtain an $(1, 4, \epsilon)$ -block-encoding of $\sum_x |x\rangle \langle x| \otimes (2/\pi) \mathcal{H}^{(x)}$ for any $\epsilon \in (0, 1/2)$. Adapted from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

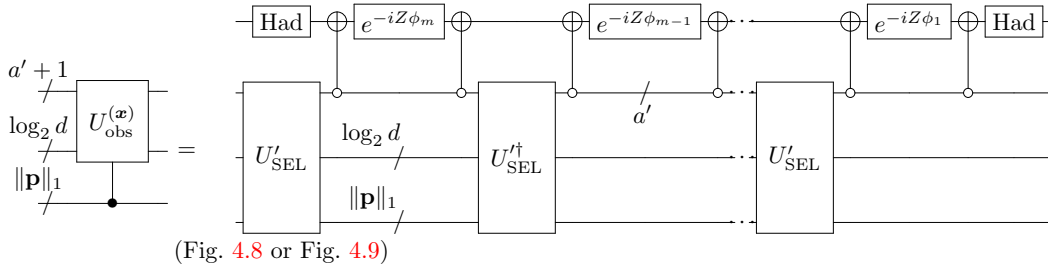


Figure 4.13: Quantum circuit for U_{obs} in Eq. (4.35). Here, $a' = \mathcal{O}(\log M + a)$. The circuit parameters $\{\phi_j\}_{j=1}^m$ are adjusted for real QSP based on the degree- m odd polynomial $P(x) \approx \gamma x$, $\gamma = \mathcal{O}(M/\sigma)$. Also, U'_{SEL} can be replaced with the circuit in Fig. 4.9 (where a' is taken as $\lceil \log_2 M \rceil + a + 4$), and in this case, the circuit equality holds by adjusting ϵ in Fig. 4.9 and m . Adapted from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

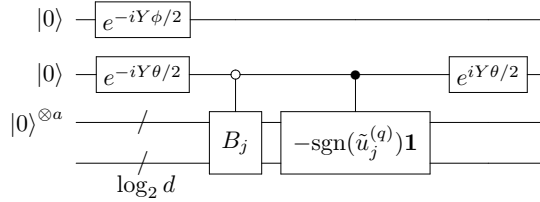


Figure 4.14: Quantum circuit for an $(a + 2)$ -block-encoding of $\tilde{O}_j^{(q)} := (O_j - \tilde{u}_j^{(q)} \mathbf{1})/2$, where $\tilde{u}_j^{(q)} \in [-1, 1]$. B_j is an a -block-encoding of O_j . The angles are defined as $\theta := 2 \tan^{-1} \sqrt{|\tilde{u}_j^{(q)}|}$ and $\phi := 2 \tan^{-1} \sqrt{4 - (1 + |\tilde{u}_j^{(q)}|)^2 / (1 + |\tilde{u}_j^{(q)}|)}$. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

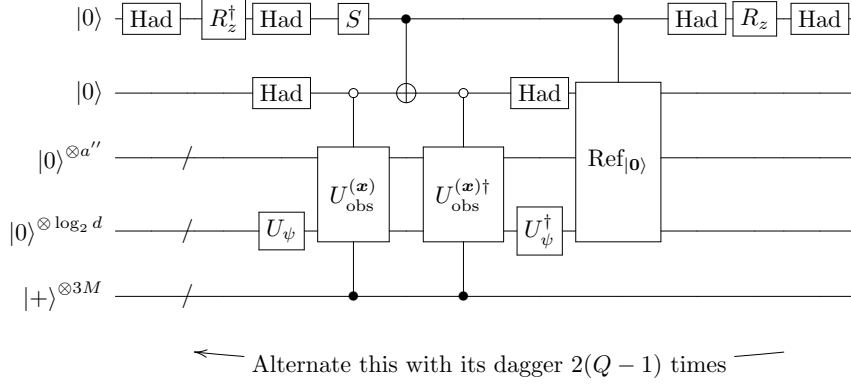


Figure 4.15: Quantum circuit for the approximate probing-state preparation $|\Upsilon(q)\rangle$ by Hamiltonian simulation. When we use the circuits Figs. 4.9 and 4.14 for encoding of target observables $\{\tilde{O}_j^{(q)}\}$, a'' of $U_{\text{obs}}^{(x)}$ is explicitly given by $a'' = a + \lceil \log_2 M \rceil + 7$, while Lemma 11 itself assumes the oracles in Fig. 4.8 (thereby, $a'' = a + \lceil \log_2 M \rceil + 1$ in Lemma 11). $Q = \mathcal{O}(2^q \sigma + \log(1/\varepsilon''))$ and $\text{Ref}_{|0\rangle}$ is the reflection on $|0\rangle$, i.e., $\text{Ref}_{|0\rangle} := 2|0\rangle\langle 0| - \mathbf{1}$. The angles of rotations R_z in the top line are adjusted by QSP to realize ε'' -precise Hamiltonian simulation [115, 116]. Adapted from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

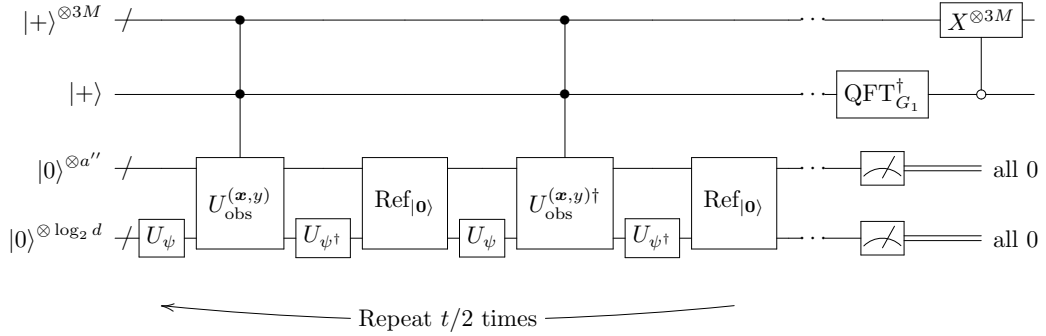


Figure 4.16: Quantum circuit for the approximate probing-state preparation $|\Upsilon(q)\rangle$ by Grover-like repetition. When we use the circuits Figs. 4.9 and 4.14 for encoding of target observables $\{\tilde{O}_j^{(q)}\}$, a'' of $U_{\text{obs}}^{(x,y)}$ is explicitly given by $a'' = a + \lceil \log_2(M+1) \rceil + 7$, while Lemma 12 itself assumes the oracles in Fig. 4.8 (thereby, $a'' = a + \lceil \log_2(M+1) \rceil + 1$ in Lemma 12). $|0\rangle := |0\rangle^{\otimes a'' + \log_2 d}$ and $\text{Ref}_{|0\rangle}$ is the reflection on $|0\rangle$, i.e., $\text{Ref}_{|0\rangle} := 2|0\rangle\langle 0| - \mathbf{1}$. The measurement outcome $|0\rangle$ is obtained with the probability $\mathcal{N}_t^2/2 \geq 0.462$ (see Lemma 12). This post-selection can be removed by using a single-step amplitude amplification and adjusting the constant δ' for our algorithm. Reproduced from K. Wada *et al.*, *Heisenberg-Limited Adaptive Gradient Estimation for Multiple Observables*, PRX Quantum 6, 020308 (2025), DOI: 10.1103/PRXQuantum.6.020308. Licensed under CC BY 4.0.

4.7 Concluding remarks and outlook

In this chapter, we proved that the expectation values of multiple observables can be simultaneously estimated with the HL scaling in the total number of queries to a state preparation unitary and its inverse, by explicitly constructing an adaptive quantum algorithm. At the same time, the query complexity shows a nearly quadratic improvement with respect to the number of general observables, compared to the conventional method for this task i.e., the (modified) QAE algorithm [21, 100, 139, 149, 167, 162]. The key idea of the proposed algorithm is to prepare a quantum state with phases that encode the expectation values of observables, followed by the measurement for the QGE algorithm [91, 56]. Then, our method determines a single binary digit of the target observables from the measurement outcomes and updates the quantum state so that the next measurement has sufficient resolution to determine the next fraction bit. Importantly, our method can be considered as an extension of the iterative or adaptive QPE algorithms [99, 78, 77, 97, 48] to the QGE algorithm. In addition, the proposed method significantly reduces the requirement of actual implementation (e.g., the number of ancilla qubits and classical computation for circuit synthesis), compared to the state-of-the-art algorithms for multiple observables estimation [86, 159].

Finally, we list several open questions to be further investigated.

- *Minimizing the hidden prefactor in the query complexity.* Although we focus on the scaling of query complexity, it is practically important to minimize the hidden prefactor in the scaling. If the prefactor is large, the advantage doesn't exist in a practical parameter regime, compared to QAE and moreover, standard sampling approaches.
- *Further reduction of ancilla qubits.* Our algorithm requires $\mathcal{O}(M)$ ancilla qubits, a linear scaling on M , but is this M -dependence optimal? Or can we achieve better scaling?
- *Improving query complexity for a specific set of states and non-commuting observables.* In standard sampling, the query complexity can scale as the order of $\sum_{i=1}^M \sigma_i^2 / \varepsilon^2$ for the quantum variance σ_i^2 determined by a target state and observable. Is it possible to achieve the quadratic speedup to this scaling, that is, the order of $\sqrt{\sum_{i=1}^M \sigma_i^2} / \varepsilon$? Because $\sum_i \sigma_i^2 \leq \|\sum_i O_i^2\|$ holds, there is a gap from Theorem 11.
- *Application to multi-parameter quantum metrology.* It is another interesting direction to further explore the quantum speedup regarding the number of target parameters, observed in our algorithm, in the field of quantum metrology.

Chapter 5

Single-observable estimation for early FTQC

5.1 Introduction

In quantum estimation, the statistical estimation theory [39, 144] and their quantum extension [79, 73] are of central importance.¹ In particular, the seminal Cramér-Rao inequality and Fisher information, which identify the limit of estimation precision, have been extended to quantum versions [76, 75, 73, 23, 133]; that is, the quantum Cramér-Rao (QCR) inequality and quantum Fisher information. As in the classical case, they can be used to determine the ultimate performance of several quantum information processing systems such as quantum sensors [60, 155, 41, 137, 143, 80]. Moreover, the QCR acts as a fundamental limitation on information extractable from quantum mechanical systems, e.g., [51]. Therefore, extensive investigations have been conducted to characterize estimators that (almost) achieve the QCR lower bound [142, 15, 11, 72, 54, 55, 131, 67, 157]. Note that whether such optimal estimators are implementable depends on available quantum operations in the target system. Devising an implementable and optimal (in the sense of QCR) estimator is thus one of the most important subjects in quantum sensing and metrology.

In this way, the quantum statistical estimation theory should be appropriately applied to develop efficient estimators for various quantities in quantum computing, such as expectation values of quantum observables. At the same time, toward the early success of quantum computing, there is an increasing need for efficient estimators that can be realized with as few demanding quantum operations as possible. For instance, some less-demanding implementation methods for the QPE algorithm [78, 77, 150, 172, 130, 35] and QAE algorithm [182, 149, 1, 69, 125, 25, 152, 156, 61, 153, 26] have been developed, and their performance has been investigated using the statistical estimation theories. Nevertheless, in practical noisy settings before the fault-tolerant era, devising an estimator that is both easy to implement and capable of achieving the ultimate QCR bound remains an ongoing subject of research. Focusing on the expectation value estimation problem in a noisy environment, Ref. [167] formulated this problem within the framework of amplitude estimation and proposed a Bayesian inference method for this problem; however, there was no discussion on how to achieve the QCR bound. To our best knowledge, only Ref. [156] gave a detailed analysis on the QCR bound in the noisy amplitude estimation problem; in particular, under the assumption that the system is subjected to depolarization noise, the estimator given in [156] achieves the precision that is yet strictly lower than the ultimate limit.

In this chapter, we focus on the problem of constructing a less-demanding quantum expectation-value estimator in a global depolarizing noisy environment. In light of quantum estimation theory and quantum metrology, the following questions are important. (i) *What is the quantum-enhanced precision limit (or, the optimal QCR bound) over less-demanding protocols for estimating an expectation value?* (ii) *Can we construct an implementable estimator such that the estimation precision achieves the QCR bound?* (iii) *Also, is there any mathematical guarantee of the estimation performance?* The main contribution in this chapter is to give answers to all the above questions.

A summary of those contributions is as follows. First, the answer to the question (i) is given in

¹This chapter is based on the author's work [162].

Theorem 14; we derive the precision limit for estimating an expectation value, in the setup where we are allowed to use noisy quantum states but no demanding controlled amplifications. Based on the findings obtained from the precision limit, we construct a modified maximum likelihood estimator that adaptively optimizes a specific measurement and the number of amplitude amplification operations, by maximizing the magnitude of quantum enhancement. The optimality of this estimator is supported by Theorem 15, stating that there exists a POVM that makes the ratio of the quantum and classical Fisher information close enough to 1 (leading that the QCR bound is almost achieved) in the limit of large number of qubits. Note that the implementation of such an optimal estimator is in general nontrivial, but our estimator can be efficiently implemented without any knowledge of the target state, which is usually prepared by a black-box unitary in the expectation value estimation problem. Thus, this is the answer to the question (ii). As for the question (iii), we prove that the estimator satisfies the following statistical properties.

- Consistency (Theorem 16): As the number of measurements increases, the estimation accuracy becomes better. More precisely, the probability that the estimated value is arbitrarily close to the true value approaches 1 in the infinite limit of the number of measurements.
- Asymptotic normality (Theorem 17): The estimator regularized by its Fisher information asymptotically follows the standard normal distribution, as the number of measurements increases. As a result, the estimation error asymptotically achieves the classical Cramér-Rao lower bound.

Therefore, the proposed estimator has mathematical guarantees in its performance.

In addition to the above theoretical contributions, Sec. 5.6 is devoted to thorough numerical simulations. In particular, we will demonstrate that the above asymptotic properties hold even with a modest number of each measurement (~ 500) and a modest number of qubits (~ 20). Also, we focus on the previously-reported problem that the precision of the expectation value estimation under the depolarizing noise may significantly deteriorate depending on the target value to be estimated; but we will numerically demonstrate that the proposed method can also get around this problem due to the optimization. Note here that the above-mentioned Bayesian inference method [167] optimizes a quantum circuit containing multi-parameter in order to alleviate the target value dependency of estimation efficiency, while our method optimizes just one discrete parameter, which is clearly computationally advantageous.

5.2 Maximum-likelihood QAE

The amplitude estimation algorithm studied in [149, 152, 156, 153], which does not use the hard-to-implement phase estimation algorithm [21], consists of two major steps. The first step is to perform amplitude amplification on the quantum state whose amplitude is to be estimated, and then make a measurement; this procedure is repeated with different operation times. The second step is to estimate the amplitude by classical post-processing for the combined measurement result, i.e., the maximum likelihood estimation.

The problem is to estimate the amplitude $\sin^2(\phi^*)$, with unknown $\phi^* \in (0, \pi/2)$, of the following $(n+1)$ -qubit quantum state $\mathcal{A}|0\rangle_{n+1}$:

$$\mathcal{A}|0\rangle_{n+1} := \cos(\phi^*)|\psi_0\rangle_n|0\rangle_1 + \sin(\phi^*)|\psi_1\rangle_n|1\rangle_1, \quad (5.1)$$

where $|0\rangle_{n+1}$ is the computational basis of $(n+1)$ -qubit, $|\psi_0\rangle_n$ and $|\psi_1\rangle_n$ are normalized n -qubit quantum states. Here, $\mathcal{A}$ is a black-box (i.e., unknown yet implementable) operation on a device, which is often called an oracle in the Grover search algorithm. Note that a single application of $\mathcal{A}$ or $\mathcal{A}^\dagger$ is counted as one query. The first step is to amplify the amplitude via the operator $\mathcal{Q}$ defined as

$$\mathcal{Q} := \mathcal{A}(2|0\rangle_{n+1}\langle 0|_{n+1} - I_{n+1})\mathcal{A}^\dagger(I_n \otimes Z), \quad (5.2)$$

where I_n is the identity operator on the n -qubit system, and Z denotes the 2×2 Pauli Z matrix. From the assumption of implementability of $\mathcal{A}$ and $\mathcal{A}^\dagger$ as oracle operators, the amplification operator $\mathcal{Q}$ is also implementable. As shown originally in [21], the amplification operator corresponds to the following rotation gate in the subspace $\mathcal{S}$ spanned by $|\psi_0\rangle_n|0\rangle_1$ and $|\psi_1\rangle_n|1\rangle_1$:

$$\mathcal{Q}|_{\mathcal{S}} = \begin{bmatrix} \cos(2\phi^*) & -\sin(2\phi^*) \\ \sin(2\phi^*) & \cos(2\phi^*) \end{bmatrix} = e^{-i2\phi^* \tilde{Y}}, \quad (5.3)$$

where $\tilde{Y}$ denotes the Pauli Y matrix with respect to the basis of $\tilde{\mathcal{S}}$. Thus, m applications of $\mathcal{Q}$ to the state (5.1) yield

$$\mathcal{Q}^m \mathcal{A} |0\rangle_{n+1} = \cos[(2m+1)\phi^*] |\psi_0\rangle_n |0\rangle_1 + \sin[(2m+1)\phi^*] |\psi_1\rangle_n |1\rangle_1. \quad (5.4)$$

We now measure the last qubit of Eq. (5.4) by the computational basis $|0\rangle_1$ and $|1\rangle_1$. Then the probability of obtaining "1" is given by

$$\tilde{\mathbf{P}}(m; \phi^*) := \frac{1}{2} - \frac{1}{2} \cos[2(2m+1)\phi^*]. \quad (5.5)$$

Note that this is equivalent to the Bernoulli trial with success probability $\tilde{\mathbf{P}}(m; \phi^*)$. By the applications of $\mathcal{Q}$, the resolution on ϕ^* in the probability (5.5) is enhanced by the factor $2m+1$; on the other hand, the ambiguity emerges in the sense that the probability $\tilde{\mathbf{P}}(m; \phi)$ becomes periodic with respect to ϕ .

In the second step, we estimate the amplitude based on the measurement results obtained in the first step by the maximum likelihood estimation method. Note that in the first step, we prepare the quantum states with different resolutions in order to eliminate the ambiguities, as similar to the methods for phase measurement [123, 77, 14, 12, 48, 35]. Specifically, for each of predefined odd numbers $2m_k + 1$ ($k = 1, 2, \dots, M$), we prepare the state (5.4) and measure it N times with the computational basis independently. Here, M denotes the total number of types of quantum states to be prepared. We write $x^{(k)} \in \{0, 1, \dots, N\}$ as the number of hitting "1" for the state $\mathcal{Q}^{m_k} \mathcal{A} |0\rangle_{n+1}$. The measurement results for M different states are put together as $\mathbf{x}_M := (x^{(1)}, x^{(2)}, \dots, x^{(M)})$. Then the likelihood function to have $\mathbf{x}_M$ is given by

$$\tilde{\mathcal{L}}_M(\phi; \mathbf{x}_M) := \prod_{k=1}^M \tilde{F}_k(x^{(k)}; m_k, \phi), \quad (5.6)$$

where $\tilde{F}_k(x^{(k)}; m_k, \phi)$ is the probability of obtaining $x^{(k)}$. Given the measurements $\mathbf{x}_M$, the maximum likelihood estimation assumes that the value of ϕ that maximizes $\tilde{\mathcal{L}}_M(\phi; \mathbf{x}_M)$ is a plausible estimate of the true value ϕ^* . In other words, the maximum likelihood estimate $\hat{\phi}_M$ for ϕ^* from the M series of measurements is defined as

$$\hat{\phi}_M := \operatorname{argmax}_{\phi \in [0, \pi/2]} \tilde{\mathcal{L}}_M(\phi; \mathbf{x}_M). \quad (5.7)$$

Note that the maximum likelihood estimator for $\sin^2(\phi^*)$ corresponds to $\sin^2(\hat{\phi}_M)$ due to the invariance property of maximum likelihood estimators. Thanks to the quantum-enhanced resolution via amplitude amplification and the elimination of its ambiguities by the product of likelihoods, the estimation precision of ϕ^* and $\sin^2(\phi^*)$ can be improved quadratically with respect to the number of total queries of $\mathcal{A}$ and $\mathcal{A}^\dagger$, by using a specific sequence $m_k = 2^{k-1}$ as demonstrated in [149].

We now consider the n -qubit depolarizing channel:

$$\mathcal{D}[\rho] := p\rho + \frac{1-p}{d} I_n, \quad d := 2^n, \quad (5.8)$$

where ρ is an arbitrary density operator and $1-p$ is the probability that depolarization occurs. As in [167, 156, 152], we here consider a noise model in which depolarization occurs with probability $1-p_q$ for a single application $\mathcal{A}$ or $\mathcal{A}^\dagger$. This assumption may be justified when the actual implementation of oracles $\mathcal{A}$ and $\mathcal{A}^\dagger$ requires complicated circuits compared to other operations in Eq. (5.2). Then the probability of obtaining "1" for the final state under this noise model is given by

$$\tilde{\mathbf{P}}_{\mathcal{D}}(m; \phi^*) := \frac{1}{2} - \frac{p_q^{2m+1}}{2} \cos[2(2m+1)\phi^*]. \quad (5.9)$$

The subscript $\mathcal{D}$ means that the quantum state to be measured has passed through the above-defined depolarizing channels. Then the classical Fisher information associated with the probability (5.9) for ϕ^* is calculated as

$$\tilde{\mathcal{I}}_c(m; \phi^*) := \frac{(2m+1)^2 p_q^{2(2m+1)} \sin^2[2(2m+1)\phi^*]}{\tilde{\mathbf{P}}_{\mathcal{D}}(m; \phi^*) (1 - \tilde{\mathbf{P}}_{\mathcal{D}}(m; \phi^*))}. \quad (5.10)$$

Equation (5.10) indicates that the estimation may become ineffective if m satisfies $\sin[2(2m+1)\phi^*] \simeq 0$. More precisely, if m and ϕ^* satisfy this condition, the classical Cramér-Rao lower bound of the estimation error for ϕ^* significantly deteriorates [167, 152]. In addition, even if we use the above-described maximum likelihood estimation based on the M series of measurements for m_k ($k = 1, 2, \dots, M$), which are predefined independently to ϕ^* , the same deterioration may occur [152]. Importantly, this phenomena are also seen in the quantum expectation value estimation problem discussed in the next section.

5.3 Problem setup

Our goal is to estimate the expectation value of a known Hermitian operator $\mathcal{O}$ with eigenvalues ± 1 , i.e., $\langle \mathcal{O} \rangle := \langle A | \mathcal{O} | A \rangle$, where $|A\rangle := A |0\rangle_n$ is an n -qubit quantum state with a state preparation unitary A . Here, we assume that A and $A^\dagger$ are black-box operations yet implementable on a device, similar to $\mathcal{A}$ in the previous section. Note that $\mathcal{O}$ is a unitary operator, which is assumed to be implementable; also, the projective measurement $\{(I_n + \mathcal{O})/2, (I_n - \mathcal{O})/2\}$ is implementable, which is often assumed in quantum algorithms. A typical $\mathcal{O}$ is a single Pauli string such as $\mathcal{O} = X \otimes Z \otimes \dots$. This setup includes the amplitude estimation problem because by setting $A = \mathcal{A}$ and $\mathcal{O} = I_n \otimes Z$ in the notation of Sec. 5.2, the expectation value $\langle \mathcal{O} \rangle$ matches $1 - 2\sin^2(\phi^*)$.

We remark that our algorithm below can also be used to estimate the expectation value of *any* block-encoded observable $\mathcal{O}'$ with an arbitrary spectrum in $[-1, 1]$. Let A' be a target state preparation $|A'\rangle = A' |0\rangle_n$. Suppose that B is a unitary operator for an a -block-encoding of $\mathcal{O}'$. Now, we take

$$A = \text{Had} \otimes I^{\otimes a} \otimes A', \quad \mathcal{O} = (X \otimes I^{\otimes a} \otimes \mathbf{1}) (|0\rangle \langle 0| \otimes B + |1\rangle \langle 1| \otimes B^\dagger). \quad (5.11)$$

Here, $\mathcal{O}$ is still an implementable unitary Hermitian operator, and moreover the corresponding two-outcome projective measurement can be performed via the Hadamard test. Then, we can show that the expectation value $\langle A | \mathcal{O} | A \rangle$ is equivalent to the target one $\langle A' | \mathcal{O}' | A' \rangle$ because of

$$(\langle + | \langle 0 |_a \otimes \mathbf{1} \rangle \mathcal{O} (| + \rangle | 0 \rangle_a \otimes \mathbf{1}) = \mathcal{O}'. \quad (5.12)$$

In the estimation problem, we aim for maximizing the estimation precision of the target value $\langle \mathcal{O} \rangle$ in given total queries to A and $A^\dagger$ [100, 139, 167, 86]. In particular, we address this problem without the quantum phase estimation algorithm which requires many controlled operations [100]. The amplitude estimation method (with no controlled- $\mathcal{Q}$ operations) in the previous section can be applied to this problem; hence, ideally, we need quadratically less queries of A and $A^\dagger$ compared to the standard algorithm without amplitude amplification, to estimate the expectation value with specified precision.

Here, we describe the expectation value estimation problem in a similar way to the previous section. Suppose $|A\rangle$ is not an eigenstate of $\mathcal{O}$ (thus $|\langle \mathcal{O} \rangle| \neq 1$). Then the two quantum states $|A\rangle$ and $\mathcal{O}|A\rangle$ are linearly independent and form the following subspace [167, 102]:

$$\mathcal{S} := \text{Span}\{|A\rangle, \mathcal{O}|A\rangle\} = \text{Span}\{|A\rangle, |A^\perp\rangle\}, \quad (5.13)$$

where $|A^\perp\rangle$ is an (unknown) normalized state orthogonal to $|A\rangle$ obtained by Gram-Schmidt procedure to $|A\rangle$ and $\mathcal{O}|A\rangle$. We denote $|A\rangle$ and $|A^\perp\rangle$ as $|\bar{0}\rangle$ and $|\bar{1}\rangle$, respectively, and identify $\mathcal{S}$ as the 1-qubit Hilbert space likewise the idea of qubitization [116, 119]. We then define an n -qubit operator Q , which corresponds to the amplitude amplification operator in Eq. (5.2), as

$$Q := A(2|0\rangle_n \langle 0|_n - I_n) A^\dagger \mathcal{O}. \quad (5.14)$$

Recall that the oracles A and $A^\dagger$ are implementable, meaning that the amplification operator Q is also implementable. This operator keeps the subspace $\mathcal{S}$ invariant because $\mathcal{O}|A\rangle = \langle \mathcal{O} \rangle |A\rangle + \sqrt{1 - \langle \mathcal{O} \rangle^2} |A^\perp\rangle$ and $\mathcal{O}|A^\perp\rangle = \sqrt{1 - \langle \mathcal{O} \rangle^2} |A\rangle - \langle \mathcal{O} \rangle |A^\perp\rangle$ hold from the definition of $|A^\perp\rangle$ and the fact that $\mathcal{O}^2 = I_n$. Then, similar to Eq. (5.3), the representation of Q on $\mathcal{S}$ is expressed as a unitary operator: $Q|_{\mathcal{S}} = \bar{Z}(\cos(\theta^*)\bar{Z} + \sin(\theta^*)\bar{X}) = \exp(i\theta^*\bar{Y})$, where we define $\bar{X}, \bar{Y}, \bar{Z}$ as the Pauli X, Y, Z matrices in the basis $|\bar{0}\rangle$ and $|\bar{1}\rangle$. The rotational angle θ^* is related to the target value as $\theta^* := \arccos(\langle \mathcal{O} \rangle)$; in what follows θ^* is also referred to as the target value. Note that, for the target

expectation value $\langle \mathcal{O} \rangle = \cos \theta^*$, the domain of θ^* is given by $\theta^* \in (0, \pi)$, which is twice as that of ϕ^* in the previous section.

Using several queries of the rotation Q on $\mathcal{S}$, we can prepare and measure the quantum state with quantum-enhanced resolution by $\alpha \in \mathbb{N}$, at the cost of α queries to A and $A^\dagger$. Applying m times of the rotation Q to $|\bar{0}\rangle$ followed by $\mathcal{O}$ and going back to the computational basis by $A^\dagger$, we obtain the following quantum state

$$\cos\left(\frac{(2m+2)\theta^*}{2}\right)|\bar{0}\rangle_n + \sin\left(\frac{(2m+2)\theta^*}{2}\right)A^\dagger|\bar{1}\rangle, \quad (5.15)$$

where this state uses $2(m+1)$ queries to A and $A^\dagger$ in total. Here, the basis $A^\dagger|\bar{1}\rangle$ is actually unknown by the black-box assumption of $A^\dagger$; thus, the class of implementable measurements on the state (5.15) is restricted, as detailed in the following sections. On the other hand, the projective measurement of $\mathcal{O}$ on $Q^m|\bar{0}\rangle$ should be considered in $\mathcal{S}$ because of the implicit θ^* -dependence of $\mathcal{O}$. That is, this process is equivalent to the projective measurement of $\bar{Z} \oplus \mathcal{O}|_{\mathcal{S}^\perp}$ on the following state with $2m+1$ queries to A or $A^\dagger$:

$$\cos\left(\frac{(2m+1)\theta^*}{2}\right)|\bar{0}\rangle + \sin\left(\frac{(2m+1)\theta^*}{2}\right)|\bar{1}\rangle. \quad (5.16)$$

This means that we can measure the quantum state with $(2m+1)$ -fold enhanced resolution of θ^* , although available POVM is restricted. Therefore, for a given $\alpha \in \mathbb{N}$, we can prepare and measure the quantum state Eq. (5.15) ($\alpha = 2m+2$) or Eq. (5.16) ($\alpha = 2m+1$), according to the parity of α ; we write these states in a unified manner as $\rho(\alpha; \theta^*)$. In the following, we call α the *amplified level*. Note that $\rho(\alpha; \theta^*)$ requires α queries to A and $A^\dagger$ in total.

Here, we assume the state $\rho(\alpha; \theta^*)$ is subjected to the (global) depolarization noise with probability $1 - p_q (> 0)$ for each use of A ($A^\dagger$), as in the previous case. Then, the output state with α queries to A and $A^\dagger$ is given by

$$\rho_{\mathcal{D}}(\alpha; \theta^*) := p_q^\alpha \rho(\alpha; \theta^*) + \frac{1 - p_q^\alpha}{d} I_n. \quad (5.17)$$

We remark that A and $A^\dagger$ (e.g., the time evolution of a global Hamiltonian) are usually comprised of a large number of gates, which would sufficiently scramble local noise into global depolarization noise. Such assumptions are verified in terms of theoretical [52, 40] and (small-size) experimental [152] aspects, while we leave further verification in a large-size experiment as future work.

5.4 Precision limit with noisy amplitude-amplified states

In the following, we derive the precision limit in estimating θ^* with the use of the noisy amplitude-amplified states $\rho_{\mathcal{D}}(\alpha; \theta^*)$, based on the quantum Fisher information of these states. To this end, we suppose that any POVM including the joint measurement on multiple n -qubit systems can be performed, for the time being.

For a given number N_q of queries to A and $A^\dagger$, let us prepare a set of $\rho_{\mathcal{D}}(\alpha; \theta^*)$ by splitting the total queries. More specifically, we consider any partition $\{\alpha'_k\}$ of given total queries N_q such that $N_q = \sum_{k=1}^{M'} \alpha'_k$ for some positive integer M' . Then, we define the quantum state before measurement as

$$\rho_{\mathcal{D}}^{(\text{bm})}(\theta^*) := \rho_{\mathcal{D}}(\alpha'_1; \theta^*) \otimes \cdots \otimes \rho_{\mathcal{D}}(\alpha'_{M'}; \theta^*). \quad (5.18)$$

Note that each $\rho_{\mathcal{D}}(\alpha'_k; \theta^*)$ uses α'_k queries to A and $A^\dagger$, and thus the quantum state $\rho_{\mathcal{D}}^{(\text{bm})}$ uses N_q queries in total. Here, for all partitions, we provide the upper bound of the total quantum Fisher information $\mathcal{I}_{q, \text{tot}}[\rho_{\mathcal{D}}^{(\text{bm})}]$ of the state $\rho_{\mathcal{D}}^{(\text{bm})}$ in estimating θ^* .

Theorem 14 (Precision limit in the noisy expectation value estimation). *For a given number N_q of queries to n -qubit state preparation oracles A and $A^\dagger$, we consider all partitions $\{\alpha'_k\}$ ($\alpha'_k \in \mathbb{N}$) of N_q such that $N_q = \sum_{k=1}^{M'} \alpha'_k$ for some positive integer M' . Here, each use of A and $A^\dagger$ induces the n -qubit depolarization noise with probability $1 - p_q$. Then, the total quantum Fisher information $\mathcal{I}_{q, \text{tot}}$ regarding θ^* of the quantum state $\rho_{\mathcal{D}}^{(\text{bm})}(\theta^*)$ in Eq. (5.18) satisfies the following inequalities:*

$$\mathcal{I}_{q, \text{tot}}[\rho_{\mathcal{D}}^{(\text{bm})}] \leq \frac{N_q^2 p_q^{2N_q}}{2^{1-n} + (1 - 2^{1-n}) p_q^{N_q}}, \quad N_q \leq \alpha_B,$$

and

$$\mathcal{I}_{q,\text{tot}}[\rho_{\mathcal{D}}^{(\text{bm})}] \leq \frac{N_q \alpha_B p_q^{2\alpha_B}}{2^{1-n} + (1 - 2^{1-n}) p_q^{\alpha_B}}, \quad N_q > \alpha_B,$$

where α_B is defined by

$$\alpha_B := \operatorname{argmax}_{\alpha \in \mathbb{N}} \frac{\alpha p_q^{2\alpha}}{2^{1-n} + (1 - 2^{1-n}) p_q^{\alpha}}. \quad (5.19)$$

Furthermore, if $N_q \leq \alpha_B$ or $N_q = r\alpha_B$ for some $r \in \mathbb{N}$, there exists a partition $\{\alpha'_k\}$ satisfying the above equality.

Proof. First, for any partition $\{\alpha'_k\}$, $\mathcal{I}_{q,\text{tot}}[\rho_{\mathcal{D}}^{(\text{bm})}]$ is given by the sum of each quantum Fisher information for $\rho_{\mathcal{D}}(\alpha'_k; \theta^*)$ with respect to θ^* :

$$\mathcal{I}_{q,\text{tot}}[\rho_{\mathcal{D}}^{(\text{bm})}] = \sum_{k=1}^{M'} \frac{(\alpha'_k)^2 p_q^{2\alpha'_k}}{\frac{2}{2^n} + (1 - \frac{2}{2^n}) p_q^{\alpha'_k}} \equiv \sum_{k=1}^{M'} \mathcal{I}_q(\alpha'_k). \quad (5.20)$$

Considering $\mathcal{I}_q(\alpha) = \alpha^2 p_q^{2\alpha} / (2^{1-n} + (1 - 2^{1-n}) p_q^{\alpha})$, we can directly confirm that $\mathcal{I}_q(\alpha)/\alpha$ is an increasing function in the regime $0 < \alpha \leq \alpha_B$, from the definition of α_B . Note that if the argmax in the definition of α_B returns multiple values, we take the minimum as α_B . In the case of $N_q \leq \alpha_B$, for any partition $\{\alpha'_k\}$, $\alpha'_k \leq N_q \leq \alpha_B$ holds, and therefore we have

$$\sum_{k=1}^{M'} \mathcal{I}_q(\alpha'_k) = \sum_{k=1}^{M'} \alpha'_k \frac{\mathcal{I}_q(\alpha'_k)}{\alpha'_k} \leq \sum_{k=1}^{M'} \alpha'_k \frac{\mathcal{I}_q(N_q)}{N_q} = \mathcal{I}_q(N_q), \quad (5.21)$$

where we recall $N_q = \sum_{k=1}^{M'} \alpha'_k$. The second equality holds for the following partition: $M' = 1$ and $\alpha'_1 = N_q$.

Next, we consider the second case $N_q > \alpha_B$. By definition, $\mathcal{I}_q(\alpha)/\alpha \leq \mathcal{I}_q(\alpha_B)/\alpha_B$ holds for any $\alpha \in \mathbb{N}$, and this leads to

$$\sum_{k=1}^{M'} \mathcal{I}_q(\alpha'_k) \leq \sum_{k=1}^{M'} \frac{\alpha'_k}{\alpha_B} \mathcal{I}_q(\alpha_B) = \frac{\mathcal{I}_q(\alpha_B)}{\alpha_B} N_q. \quad (5.22)$$

If there exists a natural number r such that $N_q = r\alpha_B$, then the first equality is saturable for the partition: $M' = r$, $\alpha'_k = \alpha_B$ for all k . This completes the proof of Theorem 14. $\square$

A similar evaluation (i.e., the loss of quadratic term on N_q) of quantum Fisher information can also be found in the previous works on noisy quantum metrology [44, 42, 105]. Note that, if the argmax returns multiple values, we take the minimum as α_B . A particularly important fact is that the equality in the case $N_q = r\alpha_B > \alpha_B$ holds for the partition $M' = r$ and $\alpha'_k = \alpha_B \forall k$; that is,

$$\rho_{\mathcal{D}}^{(\text{bm})}(\theta^*) = \rho_{\mathcal{D}}(\alpha_B; \theta^*) \otimes \cdots \otimes \rho_{\mathcal{D}}(\alpha_B; \theta^*) \quad (5.23)$$

enables us to achieve the upper bound of $\mathcal{I}_{q,\text{tot}}$. Here, α_B can be expressed as

$$\alpha_B = \operatorname{argmax}_{\alpha \in \mathbb{N}} \frac{\mathcal{I}_q(\alpha)}{\alpha}, \quad (5.24)$$

where $\mathcal{I}_q(\alpha)$ denotes the quantum Fisher information of $\rho_{\mathcal{D}}(\alpha; \theta^*)$ regarding θ^* :

$$\mathcal{I}_q(\alpha) := \frac{\alpha^2 p_q^{2\alpha}}{2^{1-n} + (1 - 2^{1-n}) p_q^{\alpha}}. \quad (5.25)$$

We can derive Eq. (5.25) using the standard recipe found in e.g., [89, 176, 87]. That is, fixing α to α_B provides the best usage of A and $A^\dagger$ in estimating θ^* under the noisy condition. In the following, we particularly focus on the regime $\alpha_B \gtrsim 100$ (i.e., $\lesssim 1\%$ depolarization noise on each A) for the early stage of fault-tolerant quantum computing [148]. As for the case $N_q \leq \alpha_B$, if we coherently use all queries (i.e., $M' = 1$ and $\alpha'_1 = N_q$), the quantum Fisher information of $\rho_{\mathcal{D}}^{(\text{bm})}$ achieves the upper bound, which contains the quadratic term with respect to the total queries N_q .

From the quantum Cramér-Rao (QCR) inequality [76, 75, 73, 23, 133], the upper bounds in Theorem 14 set the precision limit in estimating θ^* with given queries N_q , via the quantum states $\rho_{\mathcal{D}}(\alpha'_k; \theta^*)$. That is, for any (unbiased) estimator $\hat{\theta}$ of θ^* based on the measurement result from the quantum states, the mean squared error (MSE) of $\hat{\theta}$ is lower bounded as

$$\text{MSE}(\hat{\theta}) \geq \frac{1}{N_q^2} \frac{2^{1-n} + (1 - 2^{1-n}) p_q^{N_q}}{p_q^{2N_q}}, \quad N_q \leq \alpha_B \quad (5.26)$$

and

$$\text{MSE}(\hat{\theta}) \geq \frac{1}{N_q \alpha_B} \frac{2^{1-n} + (1 - 2^{1-n}) p_q^{\alpha_B}}{p_q^{2\alpha_B}}, \quad N_q > \alpha_B. \quad (5.27)$$

Note that we can obtain the lower bound of MSE regarding $\langle \mathcal{O} \rangle = \cos \theta^*$ by simply multiplying this inequality by $1 - \langle \mathcal{O} \rangle^2$. These inequalities hold for any POVM independent of θ^* ; thus they characterize the limit of estimation precision that cannot be improved by any measurement. In the following, we mainly focus on the precision limit given by Eq. (5.27) because the other bound is asymptotically negligible in a large number of queries N_q .

5.5 Adaptive observable estimation under white noise

5.5.1 The two-type measurement

To most efficiently read out the expectation value from $\rho_{\mathcal{D}}$, we need a POVM whose classical Fisher information achieves the quantum Fisher information defined for this state. In fact, we found the following 3-valued POVM that satisfies the above requirement for specific values of θ^* and $\rho_{\mathcal{D}}(\alpha; \theta^*)$ with even $\alpha = 2m + 2$, as shown in Sec. 5.5.5:

$$M_0 := |0\rangle_n \langle 0|, \quad M_1 := A^\dagger |\bar{1}\rangle \langle \bar{1}| A, \quad M_2 := I_n - M_0 - M_1, \quad (5.28)$$

where we recall that $\langle 0|_n A^\dagger |\bar{1}\rangle = \langle \bar{0}|\bar{1}\rangle = 0$. Note that this POVM can be obtained from an optimal POVM that exactly achieves the quantum Fisher information at the cost of θ^* -dependency in the POVM elements; also see Sec. 5.5.5.

Suppose we can perform such an optimal measurement on each n -qubit system of $\rho_{\mathcal{D}}^{(\text{bm})}$ in a separable way. Then, the corresponding (total) classical Fisher information $\mathcal{I}_{\text{c,tot}}$, which depends on $\rho_{\mathcal{D}}^{(\text{bm})}$ and the selection of POVM, is equal to $\mathcal{I}_{\text{q,tot}}[\rho_{\mathcal{D}}^{(\text{bm})}]$ and furthermore achieves the upper bound in Theorem 14 by setting $\alpha'_k = \alpha_B \forall k$. However, both the optimal measurement and the POVM (5.28) contain the unknown state $A^\dagger |\bar{1}\rangle$, and thus it is highly nontrivial to perform the measurements on standard quantum computing devices under the black-box assumption of A and $A^\dagger$. Hence, we approximate the ideal 3-valued POVM by marginalizing it, to have the following 2-valued POVM:

$$M_1^{(\text{even})} := I_n - M_0^{(\text{even})}, \quad M_0^{(\text{even})} := |0\rangle_n \langle 0|, \quad (5.29)$$

where the superscript *even* means that we can perform this measurement on the quantum state $\rho_{\mathcal{D}}$ with *even* α (i.e., $\alpha = 2m + 2$). This POVM can be implemented without any knowledge of the black-box operations A and $A^\dagger$, by the computational basis measurement followed by post-processing for classifying the obtained measurements into "0" or "1".

The measurement (5.29) has powerful estimation capabilities in the sense of its Fisher information, as follows. Measuring the state (5.17) with $\alpha = 2m + 2$ by this POVM, we obtain the measurement result corresponding to $M_1^{(\text{even})}$ with probability

$$\mathbf{P}_{\mathcal{D}}^{(\text{even})}(m; \theta^*) := \frac{d-1}{d} + p_q^{2(m+1)} \left[\sin^2[(m+1)\theta^*] - \frac{d-1}{d} \right]. \quad (5.30)$$

Then, the classical Fisher information with Eq. (5.30) is calculated as

$$\mathcal{I}_{\text{c}}^{(\text{even})}(m; \theta^*) := \frac{(2m+2)^2 p_q^{4(m+1)} \sin^2[(2m+2)\theta^*]}{4\mathbf{P}_{\mathcal{D}}^{(\text{even})}(m; \theta^*) \left(1 - \mathbf{P}_{\mathcal{D}}^{(\text{even})}(m; \theta^*)\right)}. \quad (5.31)$$

Also, we rewrite the corresponding quantum Fisher information (5.25) as $\mathcal{I}_{\text{q}}^{(\text{even})}(m) := \mathcal{I}_{\text{q}}(2m+2)$ for simplicity.

Theorem 15. (Optimality of the marginalized POVM)

$$\mathcal{I}_c^{(\text{even})}(m; \theta^*) = \kappa \mathcal{I}_q^{(\text{even})}(m) \times \left[1 + \frac{1}{2^n} \frac{2(1 - p_q^{2m+2})}{p_q^{2m+2}} \right] \quad (5.32)$$

holds, where κ is defined as

$$\kappa := \frac{\sin^2[(m+1)\theta^*]}{1 - p_q^{2m+2} \cos^2[(m+1)\theta^*] + \varepsilon_n(m; \theta^*)}. \quad (5.33)$$

Here, $\varepsilon_n(m; \theta^*) = \mathcal{O}(2^{-n})$ if $\cos[(m+1)\theta^*] \neq 0$. When $\theta^*/2\pi$ is an irrational number, there exists an m (i.e., the number of querying Q) such that $\kappa_\infty := \lim_{n \rightarrow \infty} \kappa$ is arbitrarily close to 1.

This theorem means that, to estimate a *particular* θ^* , the POVM (5.29) can become optimal in the sense of Fisher information, as the number of qubits increases; the proof of Theorem 15 and additional analysis are provided in Sec. 5.5.5. Note that a more detailed analysis shows that the factor $\varepsilon_n(m; \theta^*)$ diverges at certain points (m, θ^*) such that $\cos[(m+1)\theta^*] = 0$, while in other points, it converges to zero exponentially fast with respect to the number of qubits n . In the limit of a large number of qubits n , $\mathcal{I}_c^{(\text{even})}$ still depends on θ^* by the factor κ_∞ in the domains $\sin[2(m+1)\theta^*] \neq 0$, which differs from the classical Fisher information of the original 3-valued POVM; see Sec. 5.5.5.

Based on Theorem 15, in the next subsection we introduce an algorithm that optimizes m (or equivalently α) in a certain finite range. Although this theorem does not guarantee $\mathcal{I}_c^{(\text{even})} \approx \mathcal{I}_q^{(\text{even})}$ for the optimized α , we numerically verify this approximation holds well when using modest number of qubits. Furthermore, this optimization is often valid even when $\theta^*/2\pi$ is a rational number as demonstrated later. As a consequence, this algorithm enables us to have nearly optimal POVMs for *almost all* θ^* in the large system dimension, in the sense of total Fisher information i.e., $\mathcal{I}_{c,\text{tot}} \approx \mathcal{I}_{q,\text{tot}}$.

Although the even POVM (5.29) can become optimal in the sense of Fisher information, this does not necessarily guarantee that the actual estimator (particularly the maximum likelihood estimator based on the POVM) produces the best estimate that achieves the QCR bound. Actually, the probability (5.30) is an even function with respect to the value $\theta^* - \pi/2$, and the maximum likelihood estimation associated with this even POVM cannot distinguish the sign of expectation values. To construct an estimator that may correctly estimate the true target value, we therefore introduce the following second POVM that breaks the above-mentioned symmetry:

$$M_1^{(\text{odd})} := I_n - M_0^{(\text{odd})}, \quad M_0^{(\text{odd})} := \frac{I_n + \mathcal{O}}{2}. \quad (5.34)$$

From the assumption, the projective measurement of $\mathcal{O}$ is implementable. Here, we recall that the process to measure $Q^m |\bar{0}\rangle$ by this POVM is equivalent to the projective measurement of $\bar{Z} \oplus \mathcal{O}|_{\mathcal{S}^\perp}$ on the state $\rho(\alpha; \theta^*)$ with odd $\alpha = 2m+1$, as mentioned in Sec. 5.3. For this reason, we have added the superscript *odd* for the POVM (5.34). Then, the probability of obtaining "1" is given as

$$\mathbf{P}_{\mathcal{D}}^{(\text{odd})}(m; \theta^*) = \frac{1}{2} - \frac{p_q^{2m+1}}{2} \cos[(2m+1)\theta^*]. \quad (5.35)$$

The coefficient of θ^* differs by the factor 2 compared to that of ϕ^* in Eq. (5.9), because the presence of sign for expectation values doubles the domain of the target value. Since the probability (5.35) is an odd function unlike Eq. (5.30), the measurement can distinguish the sign of the target value. Note that, in contrast to the even POVM, the classical Fisher information corresponding to the odd POVM, $\mathcal{I}_c^{(\text{odd})}(m; \theta^*)$, always deviates from the quantum Fisher information even if the number of qubits increases; see Sec. 5.5.5.

Hence, our estimator is composed of the even and odd POVMs, where the probability of obtaining the measurement result "1" is summarized as follows:

$$\mathbf{P}_{\mathcal{D}}(\alpha; \theta^*) := \begin{cases} \frac{1}{2} - \frac{p_q^\alpha}{2} \cos(\alpha\theta^*), & \alpha \text{ is odd} \\ \frac{d-1}{d} + p_q^\alpha \left[\sin^2\left(\frac{\alpha}{2}\theta^*\right) - \frac{d-1}{d} \right], & \alpha \text{ is even} \end{cases}. \quad (5.36)$$

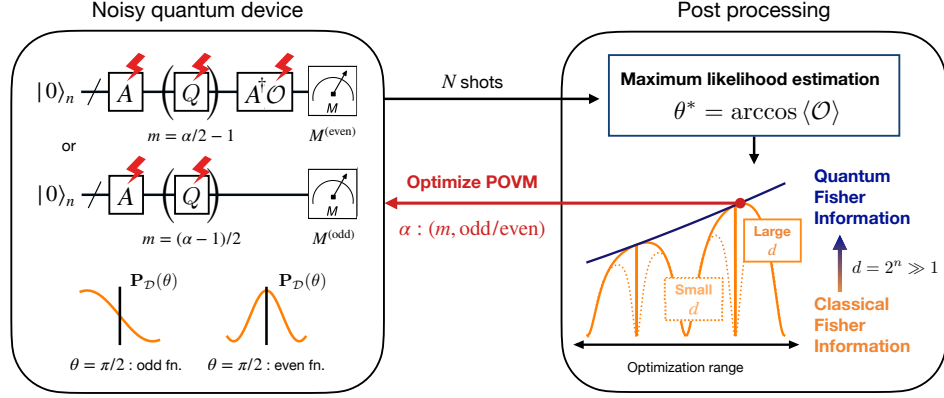


Figure 5.1: Overview of the proposed algorithm for estimating $\langle \mathcal{O} \rangle = \langle A | \mathcal{O} | A \rangle$. Here, $|A\rangle$ is a target state prepared by operating an n -qubit quantum circuit A on the computational basis $|0\rangle_n$, and $\mathcal{O}$ is a target Hermitian operator with eigenvalues ± 1 . We use the modified maximum likelihood estimation method based on the specific quantum circuits with the amplitude amplification operator Q and the two POVMs $M^{(\text{even})}, M^{(\text{odd})}$. The measurement based on the two types of POVM yields probability distributions with different symmetries, and as a result the algorithm can estimate the expectation value $\langle A | \mathcal{O} | A \rangle$ correctly including its sign. We assume that the depolarization noise is induced when implementing the state preparation A and $A^\dagger$. Even under this noisy condition, the algorithm can estimate the target value efficiently in the sense of the Fisher information by optimizing the number of querying Q (i.e., m) and the type of POVM (i.e., even or odd), which are encoded together into a 1-dimensional discrete parameter α . That is, the optimization improves the classical Fisher information so that it gets sufficiently close to the quantum Fisher information, when the system dimension $d = 2^n$ is enough large. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

Here $\alpha \in \mathbb{N}$ represents the number of queries to the operator A and $A^\dagger$. We perform the measurement via Eq. (5.34) when α is odd ($\alpha = 2m + 1$) and via Eq. (5.29) when α is even ($\alpha = 2m + 2$). Also, the classical Fisher information corresponding to each measurement is represented as

$$\mathcal{I}_c(\alpha; \theta^*) := \frac{\alpha^2 p_q^{2\alpha} \sin^2(\alpha\theta^*)}{4P_D(\alpha; \theta^*)(1 - P_D(\alpha; \theta^*))}. \quad (5.37)$$

5.5.2 The proposed adaptive estimation algorithm

Based on the discussion so far, the algorithm should satisfy the following conditions:

- When $\sin(\alpha\theta^*) \simeq 0$, the classical Fisher information (5.37) becomes nearly zero, and thus the estimation becomes inefficient by the Cramér-Rao inequality. Hence, such α should be avoided, in both POVMs (5.29) and (5.34).
- We want to employ measurements whose classical Fisher information approaches the quantum Fisher information for enhancing the estimation power.
- We want the total quantum Fisher information $\mathcal{I}_{q,\text{tot}}$ to achieve the upper bound, shown in Theorem 14; thus α should be chosen to approach $\alpha_B = \arg\max_{\alpha \in \mathbb{N}} \mathcal{I}_q(\alpha)/\alpha$, generating the state (5.23).

Here we describe our expectation value estimation method that adaptively adjusts the amplified level α and thereby satisfies the above three requirements. Importantly, this method inherits several good properties of the maximum likelihood estimation, as proven in Sec. 5.5.3. Note that, in the context of the amplitude estimation algorithms described in Sec. 5.2, Ref. [156] discusses the similar POVM in Eq. (5.29). However, unlike our method, the algorithm of [156] does not employ an adaptive optimization algorithm for estimation; thus it cannot achieve the quantum Fisher information for almost all θ^* and also suffers from the problem of vanishing classical Fisher information.

Our algorithm is based on the maximum likelihood estimation method for the random variables subjected to the probability distribution (5.36). An overview is shown in Fig. 5.1. The algorithm is composed of the following five procedures.

- (i) Set the first amplified level $\alpha_1 = 1$, and set the number of measurements N and the number of steps M as some natural numbers.
- (ii) If the amplified level α_k is even, then we measure N copies of the quantum state $\rho_{\mathcal{D}}(\alpha_k; \theta^*)$ of Eq. (5.17) with the POVM (5.29). If α_k is odd, we perform the POVM (5.34) to each of N copies of $\rho_{\mathcal{D}}(\alpha_k; \theta^*)$. We write the total number of hitting "1" as $x^{(k)} \in \{0, 1, \dots, N\}$. Here, $x^{(k)}$ is a realization of the Binomial random variable

$$X^{(k)} \sim \text{Bin}(N, \mathbf{P}_{\mathcal{D}}(\alpha_k; \theta^*)). \quad (5.38)$$

- (iii) Calculate the maximum likelihood estimate $\hat{\theta}_k$ based on the total k measurement results $\mathbf{x}_k := (x^{(1)}, x^{(2)}, \dots, x^{(k)})$. That is, $\hat{\theta}_k$ maximizes the likelihood function

$$\mathcal{L}_k(\theta; \mathbf{x}_k) := \prod_{l=1}^k F_l(x^{(l)}; \alpha_l(\mathbf{x}_{l-1}), \theta), \quad (5.39)$$

where $F_l(x^{(l)}; \alpha_l(\mathbf{x}_{l-1}), \theta^*)$ is the probability distribution of the random variable $X^{(l)}$.

- (iv) Determine α_{k+1} for the next measurement based on the current estimate $\hat{\theta}_k$ by numerically solving the following optimization problem:

$$\alpha_{k+1} := \underset{\alpha \in D_k}{\operatorname{argmax}} \frac{\mathcal{I}_c(\alpha; \hat{\theta}_k)}{\alpha} \frac{\sin^2(\alpha \hat{\theta}_k)}{1 - \delta \cos^2(\alpha \hat{\theta}_k)}, \quad (5.40)$$

where $D_k \subset \mathbb{N}$ is a (finite) optimization range and $\delta \in (0, 1]$ is a parameter for regularization.

- (v) Repeating the above procedure (ii)–(iv) for $k = 1, 2, \dots, M$, we obtain a final estimate $\hat{\theta}_M$. (We omit the optimization in (iv) when $k = M$.)

We make some remarks on the proposed algorithm. First of all, we assume that the probability of depolarization $1 - p_q$ is identified by some experiments on the quantum device to be used in advance and the calibration on p_q is perfect. The impact of calibration errors of p_q on this algorithm is analyzed numerically in Sec. 5.6.3. Unlike Bayesian estimation, our algorithm does not require any prior information on the target value.

Then, we note that the measurement $\mathbf{x}_k$ in (iii) is the collection of all the results via the measurements with $\alpha_l \forall l = 1, 2, \dots, k$. As well as the method in Sec. 5.2, the ambiguity in estimating θ^* , which arises from the periodic nature of $\mathbf{P}_{\mathcal{D}}(\alpha_k; \theta^*)$, can be eliminated by the product of likelihoods with different resolutions. Because the amplified level α_k depends on the previous results $\mathbf{x}_{k-1}$ (except for the case $k = 1$), the likelihood function $\mathcal{L}_k(\theta; \mathbf{x}_k)$ has a hierarchical structure, unlike the case of Eq. (5.6).

The optimization problem (5.40) comes from the third requirement described in the beginning of this subsection; we will provide a detailed explanation in the next two paragraphs. Since the elements of D_k are integers, the optimization problem can be efficiently solved using a simple brute force method. Also, it is obvious that the optimization does not yield an amplified level such that $\sin(\alpha \hat{\theta}_k) \simeq 0$, and thus our method can avoid vanishing classical Fisher information regardless of the target value. It should be noted that this algorithm finally outputs an estimate $\hat{\theta}_M$ together with the realized amplified levels $\{\alpha_k\}_{k=1}^M$ in a single run which uses $N_q = N \sum_{k=1}^M \alpha_k$ queries to A and $A^\dagger$ in total.

Here, we explain the meaning of the term $\mathcal{I}_c/\alpha$ in the optimization problem (5.40), which is the objective function when $\delta = 1$ (i.e., no regularization). By Theorem 15, the objective function can be written as $\kappa_\infty \mathcal{I}_q(\alpha)/\alpha$ for even α when $d = 2^n$ is sufficiently large, because we can assume $\cos(\alpha \hat{\theta}_k/2) \neq 0$ without loss of generality. Now, $\kappa_\infty \in [0, 1)$ is a periodic function while $\mathcal{I}_q(\alpha)/\alpha$ is a unimodal function, with respect to α ; the former function oscillates much faster than the change

of the latter unless the noise level is too big, resulting that the maximum point of the objective function $\kappa_\infty \mathcal{I}_q(\alpha)/\alpha$ is close to a point that maximizes $\mathcal{I}_q(\alpha)/\alpha$ within D_k . Furthermore, the numerical simulation shown later suggests that, at around the maximum point of $\kappa_\infty \mathcal{I}_q(\alpha)/\alpha$, we observe $\kappa_\infty \approx 1$. This means that, from Theorem 15, our POVM becomes nearly optimal i.e., $\mathcal{I}_c \approx \mathcal{I}_q$. Note that the odd classical Fisher information is comparable to the quantum Fisher information when α_{k+1} takes a relatively small value, and therefore the odd α (i.e., $\alpha = 2m + 1$) also takes an important role in such a regime.

Importantly, if we set D_k such that $D_k \subset D_{k+1}$ and $\alpha_B \in D_{k'}$ for some k' , then the output of this optimization problem converges to $\alpha_{B,c}(\theta^*)$ as the step k increases, under the condition $\hat{\theta}_k \rightarrow \theta^*$ which is actually valid as proven in the next subsection. Here, $\alpha_{B,c}(\theta^*)$ is defined as a solution of Eq. (5.40) with $D_k = \mathbb{N}$ and $\hat{\theta}_k = \theta^*$. The above discussion says that the factor $\alpha_{B,c}(\theta^*)$ would be close to $\alpha_B = \operatorname{argmax}_{\alpha \in \mathbb{N}} \mathcal{I}_q(\alpha)/\alpha$, which yields the state (5.23) and thereby achieves the upper bound of the total quantum Fisher information. Thus, our algorithm uses the maximal quantum enhancement specified by α_B in Theorem 14 as the step k increases. In the following section, we numerically verify these desirable properties of our algorithm.

The optimization presented above has the following issue that forced us to introduce a regularization function. That is, if κ_∞ is too close to one, the probability $1 - \mathbf{P}_D(\alpha; \theta^*)$ becomes small for such an α when $d \gg 1$, meaning that we need a large number of shots N to have a good estimate. To circumvent this issue, we introduce a regularization function $\operatorname{Reg}(\alpha; \hat{\theta}_k)$ such that $\operatorname{Reg} \simeq 0$ at $\kappa_\infty \simeq 1$ and take the objective function $\kappa_\infty \mathcal{I}_q/\alpha \times \operatorname{Reg}$ rather than $\kappa_\infty \mathcal{I}_q/\alpha$. Then, $\kappa_\infty \mathcal{I}_q/\alpha \times \operatorname{Reg}$ is not maximized at an α such that κ_∞ is too close to 1. But we want κ_∞ to be close enough to 1, hence the regularization term Reg should satisfy $\operatorname{Reg} \simeq 0$ only in the narrow region of α satisfying $\kappa_\infty \simeq 1$. We employ the following regularization function satisfying those requirements:

$$\operatorname{Reg}(\alpha; \hat{\theta}_k) := \frac{\sin^2(\alpha \hat{\theta}_k)}{1 - \delta \cos^2(\alpha \hat{\theta}_k)}. \quad (5.41)$$

When we employ the parameter δ closer to 1, the regularization function becomes sharper around α such that $\kappa_\infty \simeq 1$. Hence, we arrive at the objective function given in Eq. (5.40). In the numerical simulation shown later, we choose $\delta = 0.95$.

Also, the optimization ranges of $\{D_k\}$ control the following trade-off. That is, a large α is preferred to enhance the Fisher information per shot, but if α dramatically (e.g., super-exponentially) increases with respect to k , then the estimation will be failed for a practical N because the likelihood function $\mathcal{L}_k(\theta; \mathbf{x}_k)$ has many peaks around the true target value. A more detailed analysis related to this trade-off can be found in Ref. [74].

5.5.3 Statistical properties of the estimator

Although the random variables $X^{(k)}$ describing the measurement results depend on each other and have the hierarchical structure as shown in Eq. (5.39), we can prove the following desirable statistical property of the estimator; i.e., the consistency.

Theorem 16. (Consistency; informal version.) *There exists a unique maximum likelihood estimator $\hat{\theta}_M$ of $\mathcal{L}_M(\theta; \mathbf{X}_M)$ such that*

$$\cos \hat{\theta}_M \rightarrow \langle \mathcal{O} \rangle, \quad (5.42)$$

where $\rightarrow$ means the convergence in probability as $N \rightarrow \infty$. In addition, the random variable $\alpha_k(\mathbf{X}_{k-1})$ converges to a constant α_k^* in the sense of probability as follows

$$\lim_{N \rightarrow \infty} P[\alpha_k(\mathbf{X}_{k-1}) = \alpha_k^*] = 1, \quad k = 2, \dots, M, \quad (5.43)$$

where $\alpha_k(\mathbf{x}_{k-1})$ is the optimized amplified level based on the measurement results $\mathbf{x}_{k-1}$; that is, $\{\alpha_k\}$ become independent with each other, as $N \rightarrow \infty$.

The proof is given in Sec. 5.5.5, where α_k^* is given in Eq. (5.69). Theorem 16 means that the proposed estimator is valid in the sense that more data leads to more accurate estimation. That is, unlike the previous method [167] which uses Bayesian inference based on the measurement with the odd POVM (5.34) (not the even POVM), the estimate approaches the true value with high probability

as the number of measurements N increases, as implied by Eq. (5.42). Note that the Bayesian inference may return estimates far from the true value as discussed in [90].

Furthermore, the asymptotic variance of the estimator can achieve the Cramér-Rao lower bound when N is large. To show this result, we focus on the following total classical/quantum Fisher information in our method:

$$\mathcal{I}_{c,\text{tot}}(\theta^*) = N \sum_{k=1}^M \mathbb{E}_{\mathbf{X}_{k-1}} [\mathcal{I}_c(\alpha_k(\mathbf{X}_{k-1}); \theta^*)], \quad (5.44)$$

$$\mathcal{I}_{q,\text{tot}}(\theta^*) = N \sum_{k=1}^M \mathbb{E}_{\mathbf{X}_{k-1}} [\mathcal{I}_q(\alpha_k(\mathbf{X}_{k-1}))]. \quad (5.45)$$

The derivation is provided in Sec. 5.5.5. Since Theorem 16 states that $\alpha_k(\mathbf{X}_{k-1})$ converges to a constant α_k^* with high probability as N increases, the total Fisher information also converges as follows;

$$\frac{\mathcal{I}_{c/q,\text{tot}}(\theta^*)}{N} \rightarrow \frac{\mathcal{I}_{c/q,\text{tot}}^*(\theta^*)}{N}, \quad (5.46)$$

where $\mathcal{I}_{c/q,\text{tot}}^*(\theta^*)$ is the asymptotic value of the total classical/quantum Fisher information defined as

$$\mathcal{I}_{c,\text{tot}}^*(\theta^*) := N \sum_{k=1}^M \mathcal{I}_c(\alpha_k^*; \theta^*), \quad (5.47)$$

$$\mathcal{I}_{q,\text{tot}}^*(\theta^*) := N \sum_{k=1}^M \mathcal{I}_q(\alpha_k^*). \quad (5.48)$$

Note that α_k^* depends on the target value θ^* implicitly as in Eq. (5.69). Here, we provide the convergence theorem for the distribution of our estimator.

Theorem 17. (Asymptotic normality; informal version.) *If $\hat{\theta}_M$ is a maximum likelihood estimator of $\mathcal{L}_M(\theta; \mathbf{X}_M)$, then the following convergence holds;*

$$\sqrt{\mathcal{I}_{c,\text{tot}}^*(\arccos \langle \mathcal{O} \rangle)} \left(\cos \hat{\theta}_M - \langle \mathcal{O} \rangle \right) \rightarrow \mathcal{N}(0, 1 - \langle \mathcal{O} \rangle^2), \quad (5.49)$$

where $\mathcal{N}(0, 1 - \langle \mathcal{O} \rangle^2)$ denotes a centered normal distribution with variance $1 - \langle \mathcal{O} \rangle^2$, and $\rightarrow$ means the convergence in distribution as N increases.

This theorem shows that the distribution of the estimator $\cos \hat{\theta}_M$ gets arbitrarily close to the normal distribution with mean $\langle \mathcal{O} \rangle$ in large N . In addition, the asymptotic variance of the estimator is proportional to the inverse of the asymptotic value of the total classical Fisher information. Thus, our estimator can asymptotically achieve the classical Cramér-Rao lower bound. In the next section, we numerically demonstrate how fast the above two asymptotic properties behave with respect to N .

As mentioned in Sec. 5.5.2 and demonstrated later, the adaptive optimization (5.40) yields the nearly optimal measurement (i.e., $\mathcal{I}_c \approx \mathcal{I}_q$) at each step in a large number of qubits. Furthermore, the output α_k of (5.40) converges to $\alpha_{B,c}(\theta^*) \approx \alpha_B$ as the step k increases for an appropriate choice of $\{D_k\}$. As a result, the properties of our algorithm lead to the following relations regarding the total Fisher information:

$$\mathcal{I}_{c,\text{tot}}(\theta^*) \approx \mathcal{I}_{q,\text{tot}}(\theta^*) \underset{M \gg 1}{\approx} N_q^* \frac{\mathcal{I}_q(\alpha_B)}{\alpha_B}, \quad (5.50)$$

where N_q^* denotes the total number of queries for the asymptotic sequence $\{\alpha_k^*\}$:

$$N_q^* := N \sum_{k=1}^M \alpha_k^*. \quad (5.51)$$

Note that this can be rewritten as $N_q^* = \sum_{l=1}^{M'} \alpha_l'$ for $M' = NM$ and $\alpha_l' = \alpha_k^*$ if $(k-1)N + 1 \leq l \leq kN$ in the notation of Theorem 14. Importantly, the most right hand side in Eq. (5.50) matches the upper bound in Theorem 14, where we use $N_q^* (\geq \alpha_B)$ queries to A and $A^\dagger$ in total. In the next section, we demonstrate that the approximation errors are sufficiently small for *almost all* θ^* , via an appropriate choice of $\{D_k\}_{k=1}^{M-1}$; that is, together with the asymptotic normality, our estimator nearly achieves the ultimate precision given by the inverse of total quantum Fisher information (5.27).

5.5.4 Summary of the proposed protocol

Here we summarize the notable feature of the proposed adaptive method, with remarks on which part is theoretically guaranteed and which part is left for numerical verification.

The goal is to ultimately estimate $\langle \mathcal{O} \rangle = \cos \theta^*$, for the given n -qubit implementable operators A and $\mathcal{O}$; note that, likewise the Grover operator, the implementability of these operators does not mean that all the operator components are known and accordingly $\langle \mathcal{O} \rangle$ is directly computable. The noise parameter p_q are assumed to be known. Another important assumption is that $d = 2^n \gg 1$. We do not need other prerequisites for the target system.

The estimation procedure is illustrated in Fig. 5.1, where the amplification level α_k , which determines the number of querying Q and the type of POVM, is updated via Eq. (5.40) at each step k (each step contains N shots). This optimization (5.40) is the maximization of classical Fisher information per queries, $\mathcal{I}_c/\alpha$, reflecting the goal to achieve quantum Fisher information $\mathcal{I}_q$ at each step and $\alpha_k \rightarrow \alpha_B := \operatorname{argmax}_{\alpha \in \mathbb{N}} \mathcal{I}_q(\alpha)/\alpha$. Here, α_B provides the best coherent use of noisy A and $A^\dagger$ in estimating θ^* , which is proved in Theorem 14. From the analysis of our POVM (5.29) in Theorem 15, the objective function $\mathcal{I}_c/\alpha$ can be written as $\kappa_\infty \mathcal{I}_q/\alpha$, where κ_∞ is a bounded periodic function regarding α . Then, to achieve the above goal, the following properties should be satisfied: (i) $\max \kappa_\infty \mathcal{I}_q/\alpha$ in a given optimization range $D_k (\ni \alpha)$ is close to $\max \mathcal{I}_q/\alpha$ in D_k and (ii) $\kappa_\infty \approx 1$ (i.e., $\mathcal{I}_c \approx \mathcal{I}_q$) at the maximum point. Note that we gradually expand the range of D_k as k increases to eliminate the estimation ambiguity; see Sec. 5.2. In Sec. 5.5.2, we provide an intuitive explanation (not a rigorous proof) that the properties (i) and (ii) hold, and furthermore, these properties will be numerically verified in the next section.

On the other hand, Theorems 16 and 17 guarantee the statistical properties of the estimator. More precisely, Theorem 16 guarantees the consistency, stating that the estimate approaches the true value with high probability as the number of shots N increases. Also, Theorem 17 guarantees the asymptotic normality, stating that the estimator asymptotically achieves the classical Cramér-Rao lower bound $1/\mathcal{I}_{c,\text{tot}}^*(\theta^*)$. Here, the (asymptotic) total classical/quantum Fisher information $\mathcal{I}_{c,\text{tot}}^*(\theta^*)$ and $\mathcal{I}_{q,\text{tot}}^*(\theta^*)$ are defined in Eqs. (5.47) and (5.48), respectively. Therefore, if the nearly optimal sequence α_k is properly chosen, the estimator achieves the estimation precision $1/\mathcal{I}_{c,\text{tot}}^*(\theta^*) \approx 1/\mathcal{I}_{q,\text{tot}}^*(\theta^*)$, and furthermore, it also achieves the precision $\alpha_B/(N_q^* \mathcal{I}_q(\alpha_B))$ for the (asymptotic) total queries N_q^* in Eq. (5.51), when the iteration k is sufficiently large. This matches the precision limit in Theorem 14 (more precisely, Eq. (5.27)). We numerically find that this happens in a moderate number of shots N , as will be demonstrated in Figs. 5.4, 5.5 and 5.8.

5.5.5 Detailed derivations

Further analysis of the POVMs

Our problem is to estimate the target parameter $\theta \in (0, \pi)$ embedded into the noisy quantum state (5.17). Here, let us consider the following POVM:

$$M_0 := |0\rangle_n \langle 0|, \quad M_1 := A^\dagger |\bar{1}\rangle \langle \bar{1}| A, \quad M_2 := I_n - M_0 - M_1. \quad (5.52)$$

The POVM has the following meanings: M_0 and M_1 correspond to the measurement in the subspace basis $|0\rangle_n$ and $A^\dagger |\bar{1}\rangle$, respectively, and M_2 corresponds to the event such that the measured state is not in the subspace $\text{Span} \{|0\rangle_n, A^\dagger |\bar{1}\rangle\}$. Now, the classical Fisher information $\mathcal{I}'_c(\theta)$ for the probability distribution

$$\{\operatorname{tr}[M_k \rho_D(\alpha; \theta)]\}_{k=0}^2, \quad \alpha = 2m + 2 \quad (5.53)$$

is calculated as follows:

$$\begin{aligned} \mathcal{I}'_c(\theta) &= 4(\eta')^2 (m+1)^2 \sin^2 [2(m+1)\theta] \left\{ \eta' + \frac{2(1-\eta')}{d} - \frac{d(\eta')^2 \cos^2 [2(m+1)\theta]}{d\eta' + 2(1-\eta')} \right\}^{-1}, \quad d = 2^n \\ &= 4(\eta')^2 (m+1)^2 \frac{d\eta' + 2(1-\eta')}{d(\eta')^2} \sin^2 [2(m+1)\theta] \left[\left(1 + \frac{2(1-\eta')}{d\eta'} \right)^2 - 1 + \sin^2 [2(m+1)\theta] \right]^{-1} \\ &\leq 4(\eta')^2 (m+1)^2 \frac{d\eta' + 2(1-\eta')}{d(\eta')^2} \left(1 + \frac{2(1-\eta')}{d\eta'} \right)^{-2} = \frac{d(\eta')^2 (2m+2)^2}{d\eta' + 2(1-\eta')}, \end{aligned} \quad (5.54)$$

where $\eta' := p_q^\alpha = p_q^{2m+2}$. Since a function $x^2/(K+x^2)$, $x \in [-1, 1]$ with a positive constant K is maximized at $x = \pm 1$, we obtain the third line, and therefore the condition for the equality is given by $\sin[2(m+1)\theta] = \pm 1$. The most right hand side in the final line corresponds to the quantum Fisher information $\mathcal{I}_q(\alpha)$ for $\alpha = 2m+2$. Thus, for θ satisfying the equality condition, the POVM (5.52) is optimal in the sense that the corresponding classical Fisher information gives the quantum Fisher information. Moreover, in the limit of $d = 2^n \rightarrow \infty$, the classical Fisher information is equal to the quantum Fisher information except for (m, θ) satisfying $\sin[2(m+1)\theta] = 0$. Note that, in general, the eigenstates of SLD operator can be used to construct the optimal measurement that exactly achieves the quantum Fisher information [23]. In our case, the POVM of optimal measurement consists of quantum states in the form of superposition of $|0\rangle_n$ and $A^\dagger |\bar{1}\rangle$, where the coefficients of the states depend on the unknown target value; the POVM (5.52) is obtained by removing this target-dependency in the coefficients.

Although the POVM (5.52) is an optimal measurement for certain θ satisfying $\sin[2(m+1)\theta] = \pm 1$, its implementation is nontrivial especially for M_1 because $A^\dagger |\bar{1}\rangle$ is unknown. Note that the optimal POVM obtained from the SLD operator has the same difficulty in implementation. Therefore, in this paper, we focus on the 2-valued POVM (5.29), which can be obtained from the POVM (5.52) as follows:

$$M_0^{(\text{even})} = M_0, \quad M_1^{(\text{even})} = M_1 + M_2. \quad (5.55)$$

This POVM removes the element M_1 from the original 3-valued POVM (5.52), meaning that the quantum Fisher information is not achieved in general.

The classical Fisher information associated with the 2-valued POVM is calculated as, assuming $\cos[(m+1)\theta] \neq 0$,

$$\begin{aligned} \mathcal{I}_c^{(\text{even})}(m; \theta) &= \frac{(2m+2)^2 \eta' \sin^2[(m+1)\theta]}{1 - \eta' \cos^2[(m+1)\theta] + \frac{1-\eta'}{d\eta' \cos^2[(m+1)\theta]} \left(1 - \frac{1-\eta'}{d} - 2\eta' \cos^2[(m+1)\theta]\right)} \\ &= \mathcal{I}_q^{(\text{even})}(m) \frac{\sin^2[(m+1)\theta] \left(1 + 2\frac{1-\eta'}{d\eta'}\right)}{1 - \eta' \cos^2[(m+1)\theta] + \frac{1-\eta'}{d\eta' \cos^2[(m+1)\theta]} \left(1 - \frac{1-\eta'}{d} - 2\eta' \cos^2[(m+1)\theta]\right)} \\ &\equiv \mathcal{I}_q^{(\text{even})}(m) \frac{\sin^2[(m+1)\theta]}{1 - \eta' \cos^2[(m+1)\theta] + \varepsilon_n(m; \theta)} \left(1 + 2\frac{1-\eta'}{d\eta'}\right), \end{aligned} \quad (5.56)$$

where we recall $\eta' := p_q^\alpha = p_q^{2m+2}$ and in the final line, we defined ε_n as

$$\varepsilon_n(m; \theta) := \frac{1-\eta'}{d\eta' \cos^2[(m+1)\theta]} \left(1 - \frac{1-\eta'}{d} - 2\eta' \cos^2[(m+1)\theta]\right). \quad (5.57)$$

This factor ε_n vanishes exponentially fast with respect to the number of qubits n if $\cos[(m+1)\theta] \neq 0$, as follows:

$$|\varepsilon_n(m; \theta)| \leq \frac{1-\eta'}{\eta' \cos^2[(m+1)\theta]} \frac{1}{2^n} = O\left(\frac{1}{2^n}\right), \quad (5.58)$$

where we recall $d = 2^n$. Here, we define the coefficient of $\mathcal{I}_q^{(\text{even})}(m)$ as κ :

$$\kappa := \frac{\sin^2[(m+1)\theta]}{1 - \eta' \cos^2[(m+1)\theta] + \varepsilon_n(m; \theta)} \quad (5.59)$$

Then, Eq. (5.56) is expressed as

$$\mathcal{I}_c^{(\text{even})}(m; \theta) = \kappa \mathcal{I}_q^{(\text{even})}(m) \left[1 + 2\frac{1-\eta'}{2^n \eta'}\right]. \quad (5.60)$$

Now, when $\theta/2\pi$ is an irrational number, the set of complex numbers $\{e^{i(m+1)\theta}\}_{m=0,1,\dots}$ with sufficiently large m fill the unit circle in the complex plane. Thus, there exists an m such that $\cos[(m+1)\theta]$ is arbitrarily close to 0 and accordingly

$$\kappa_\infty := \lim_{d \rightarrow \infty} \kappa = \frac{\sin^2[(m+1)\theta]}{1 - \eta' \cos^2[(m+1)\theta]} \in [0, 1) \quad (5.61)$$

is arbitrarily close to 1. Therefore, from Eq. (5.60), the classical Fisher information sufficiently approaches the quantum Fisher information, under certain condition on (m, θ) ; moreover, taking an appropriate m , we can make the convergence exponentially fast with respect to the number of qubits (recall $d = 2^n$ with n the number of qubits). This establishes Theorem 15.

Here we point out that there is discontinuity in the classical Fisher information, which implies the instability of estimation. For this purpose, let us see the Fisher information in the limit $d \rightarrow \infty$:

$$\lim_{d \rightarrow \infty} \mathcal{I}_c^{(\text{even})}(m; \theta) = \begin{cases} 0, & \cos[(m+1)\theta] = 0 \\ \frac{\mathcal{I}_q^{(\text{even})}(m) \sin^2[(m+1)\theta]}{1 - \eta' \cos^2[(m+1)\theta]}, & \cos[(m+1)\theta] \neq 0 \end{cases}. \quad (5.62)$$

That is, for $\cos[(m+1)\theta] \neq 0$, we obtain

$$\lim_{d \rightarrow \infty} \mathcal{I}_c^{(\text{even})}(m; \theta) \leq 4\eta'(m+1)^2 = \lim_{d \rightarrow \infty} \mathcal{I}_q^{(\text{even})}(m). \quad (5.63)$$

The first equality holds for $\cos[(m+1)\theta] = 0$ even though $\mathcal{I}_c^{(\text{even})}(m; \theta)$ vanishes at such points as shown in Eq. (5.62). Therefore, the classical Fisher information is not continuous at $\cos[(m+1)\theta] = 0$ in the limit $d = 2^n \rightarrow \infty$.

Finally, as for the other 2-valued POVM (5.34), comparing the Fisher information $\mathcal{I}_{q/c}(\alpha; \theta)$ for $\alpha = 2m+1$, we can obtain the following inequality for $d > 2$:

$$\mathcal{I}_c^{(\text{odd})}(m; \theta) := \frac{(2m+1)^2 (p_q^{2m+1})^2 \sin^2[(2m+1)\theta]}{4\mathbf{P}_D^{(\text{odd})}(m; \theta) (1 - \mathbf{P}_D^{(\text{odd})}(m; \theta))} \leq (2m+1)^2 p_q^{2(2m+1)} < \mathcal{I}_q(2m+1). \quad (5.64)$$

Hence, unlike the even classical Fisher information, the odd classical Fisher information is always smaller than the quantum Fisher information even when the number of qubits increases.

Statistical structure of the algorithm

In our model, the distribution of random variables $\mathbf{X}_M$ has a hierarchical model as follows;

$$\mathcal{L}_M(\mathbf{x}_M; \theta) = \prod_{m=1}^M F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta), \quad (5.65)$$

where F_m is the probability function of Binomial distribution with success probability $\mathbf{P}_D(\alpha_m(\mathbf{x}_{m-1}); \theta)$. Substituting Eq. (5.65) into Eq. (2.19), we obtain the total classical Fisher information as

$$\begin{aligned} \mathcal{I}_{c, \text{tot}}(\theta) &= - \sum_{m=1}^M \mathbb{E}_{\mathbf{X}_m} \left[\partial_\theta^2 \ln F_m(x^{(m)}; \alpha_m(\mathbf{X}_{m-1}), \theta) \right] \\ &= - \sum_{m=1}^M \sum_{\mathbf{x}_m = (x^{(1)}, \dots, x^{(m)})} \partial_\theta^2 \ln F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta) \mathcal{L}_m(\mathbf{x}_m; \theta) \\ &= - \sum_{m=1}^M \sum_{x^{(m)}} \sum_{\mathbf{x}_{m-1} = (x^{(1)}, \dots, x^{(m-1)})} \partial_\theta^2 \ln F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta) F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta) \mathcal{L}_{m-1}(\mathbf{x}_{m-1}; \theta) \\ &= - \sum_{m=1}^M \sum_{\mathbf{x}_{m-1}} \mathcal{L}_{m-1}(\mathbf{x}_{m-1}; \theta) \sum_{x^{(m)}} \partial_\theta^2 \ln F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta) F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta) \\ &= - \sum_{m=1}^M \sum_{\mathbf{x}_{m-1}} \mathcal{L}_{m-1}(\mathbf{x}_{m-1}; \theta) \mathbb{E}_{Y^{(m)}} \left[\partial_\theta^2 \ln F_m(Y^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta) \right], \end{aligned} \quad (5.66)$$

where we used the conditional probability (5.65) in the third line. $Y^{(m)}$ denotes a random variable following $F_m(y^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta)$, with $y^{(m)} \in \{0, 1, \dots, N\}$. Note here that $\mathbf{x}_{m-1}$ is not a random variable, but a realized value of $\mathbf{X}_{m-1}$. The expectation in the final line corresponds to the classical

Fisher information in Eq. (5.37) multiplied by N . Thus, we can write the total classical Fisher information as

$$\mathcal{I}_{c,\text{tot}}(\theta) = N \sum_{m=1}^M \mathbb{E}_{\mathbf{X}_{m-1}} [\mathcal{I}_c(\alpha_m(\mathbf{X}_{m-1}); \theta)]. \quad (5.67)$$

Similarly, we define the corresponding total quantum Fisher information as

$$\mathcal{I}_{q,\text{tot}}(\theta) = N \sum_{m=1}^M \mathbb{E}_{\mathbf{X}_{m-1}} [\mathcal{I}_q(\alpha_m(\mathbf{X}_{m-1}))]. \quad (5.68)$$

As proved in the following subsection, $\alpha_m(\mathbf{X}_{m-1}) = \alpha_m^*$ holds with high probability for all m as N increases, where α_m^* is a solution of the following optimization problem for the target value θ (not an estimate $\hat{\theta}$):

$$\alpha_m^* := \operatorname{argmax}_{\alpha \in D_{m-1}} \frac{\mathcal{I}_c(\alpha; \theta)}{\alpha} \frac{\sin^2(\alpha\theta)}{1 - \delta \cos^2(\alpha\theta)}, \quad (5.69)$$

where we define $\alpha_1^* = 1$. Using Eq. (5.69), the asymptotic values of the total classical/quantum Fisher information are given by

$$\mathcal{I}_{c/q,\text{tot}}^*(\theta) := N \sum_{m=1}^M \mathcal{I}_{c/q}(\alpha_m^*; \theta). \quad (5.70)$$

Then we have

$$\lim_{N \rightarrow \infty} \frac{\mathcal{I}_{c/q,\text{tot}}(\theta)}{N} = \sum_{m=1}^M \mathcal{I}_{c/q}(\alpha_m^*; \theta) = \frac{\mathcal{I}_{c/q,\text{tot}}^*(\theta)}{N}, \quad (5.71)$$

which characterizes the variance of our estimator in the asymptotic regime.

Finally, we show the impact of δ on the asymptotic total classical/quantum Fisher information in Fig. 5.2. Here, the experimental settings are equal to those in Sec. 5.6. As expected from the form of the objective function (5.69), we confirm that the ratio of $\mathcal{I}_{c/q,\text{tot}}^*(\theta)$ gets closer to 1 when $\delta \rightarrow 1$. Also, in the case of $\delta = 0.99$, the peaks at the rational points π/j , ($j = 2, 3, 4, 6$) become more distinct than the case of $\delta = 0.95$.

Proof for statistical properties

Let $\theta^* \in (0, \pi)$ satisfying $\cos \theta^* = \langle \mathcal{O} \rangle$ be a target value to be estimated. Since the objective function in Eq. (5.69) vanishes at $\alpha\theta^* = l\pi$ ($l \in \mathbb{Z}$), then the definition of α_m^* leads to $\alpha_m^*\theta^* \neq l\pi$ for $m = 2, 3, \dots, M$. Thus, we can take an integer z_m satisfying

$$\theta^* \in \Theta^{(m)} := \left(\frac{z_m}{\alpha_m^*} \pi, \frac{z_m + 1}{\alpha_m^*} \pi \right), \quad \forall m, \quad (5.72)$$

and define $\Theta_M := \cap_{m=1}^M \Theta^{(m)}$.

Theorem 18. *There exists a unique maximum likelihood estimator $\hat{\theta}_M$ of $\mathcal{L}_M(\theta; \mathbf{X}_M)$ such that $\cos \hat{\theta}_M \rightarrow \cos \theta^*$ (convergence in probability) with the probability approaching 1 as $N \rightarrow \infty$, and*

$$\lim_{N \rightarrow \infty} P[\alpha_k(\mathbf{X}_{k-1}) = \alpha_k^*] = 1 \quad (5.73)$$

holds for all $k = 2, \dots, M$.

As a demonstration of Theorem 16, we provide an example of the realized amplified levels in Fig. 5.3. Here, the experimental settings are equivalent to those in Sec. 5.6.1. Figure 5.3 shows that the amplified levels converge rapidly to the asymptotic values, and therefore the total Fisher information and the total number of queries also show rapid convergence.

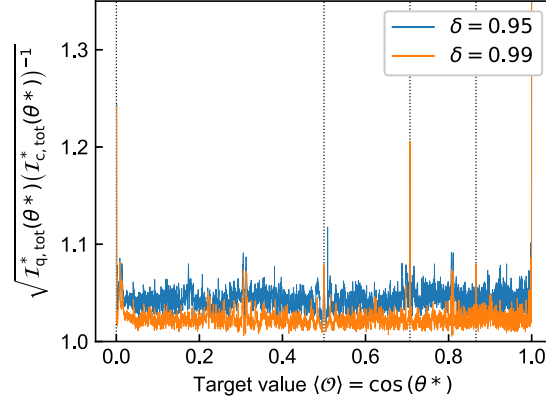


Figure 5.2: Comparison of the asymptotic total classical and quantum Fisher information of our method in the case of 20-qubit system and $M = 8$. The blue (top) and orange (bottom) lines represent the results for the case of the regularization parameter $\delta = 0.95$ and $\delta = 0.99$, respectively. The vertical dotted lines represent the target values corresponding to $\theta^* = \pi/j$, $j = 2, 3, 4, 6$ from left to right. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

Proof. Let $E(m)$ be the following event : $\alpha_k(\mathbf{X}_{k-1}) = \alpha_k^*$ holds for all $k = 1, 2, \dots, m$, and there exists a unique maximum likelihood estimator $\hat{\theta}_m$ of $\mathcal{L}_m(\theta; \mathbf{X}_m)$ which converges to θ^* . Note that if we establish $\hat{\theta}_m \rightarrow \theta^*$, then this immediately leads to $\cos \hat{\theta}_m \rightarrow \cos \theta^*$ due to the continuous mapping theorem.

In the case of $m = 1$, the likelihood function $\mathcal{L}_1(\theta; \mathbf{X}_1)$ is equivalent to $F_1(X^{(1)}; \alpha_1^*, \theta)$. Because the success probability $\mathbf{P}_{\mathcal{D}}(\alpha_1^*; \cdot) : (0, \pi) \rightarrow (0, 1)$ is injective and the likelihood equation

$$\partial_{\theta} \ln F_1(X^{(1)}; \alpha_1^*, \theta) = 0 \quad (5.74)$$

has a unique solution when a solution exists, it follows that

$$P[E(1)] \rightarrow 1 \text{ as } N \rightarrow \infty, \quad (5.75)$$

from the theory of maximum likelihood estimation [144].

Here, we assume that $P[E(M-1)] \rightarrow 1$ for $M \geq 2$. For simplicity, we identify the classical Fisher information $\mathcal{I}_c(\alpha; \theta)$ with the objective function in Eq. (5.40) in this proof. Fixing α in the objective function $\mathcal{I}_c(\alpha; \theta)$, we can confirm that $\mathcal{I}_c(\alpha; \hat{\theta}_{M-1})$ converges smoothly to $\mathcal{I}_c(\alpha; \theta^*)$ when $\hat{\theta}_{M-1} \rightarrow \theta^*$. If there is a unique point maximizing $\mathcal{I}_c(\alpha; \theta^*)$ in the range of D_{M-1} , then there exists some small number $\epsilon' > 0$, and the following implication holds

$$\left\{ \left| \hat{\theta}_{M-1} - \theta^* \right| < \epsilon' \right\} \subseteq \left\{ \alpha_M(\mathbf{X}_{M-1}) = \alpha_M^* \right\}, \quad (5.76)$$

where $\{\cdot\}$ denotes a set of events such as

$$\left\{ \mathbf{x}_{M-1} : \left| \hat{\theta}_{M-1}(\mathbf{x}_{M-1}) - \theta^* \right| < \epsilon' \right\}. \quad (5.77)$$

Note that even if the maximum point of $\mathcal{I}_c(\alpha; \theta^*)$ in D_{M-1} is not unique, Eq. (5.76) also holds when we adopt the minimum value in a subset $\bar{D}_{M-1}$ of D_{M-1} as the realization of $\alpha_M(\mathbf{X}_{M-1})$. Here, the subset $\bar{D}_{M-1}$ consists of elements in D_{M-1} satisfying

$$\left| \mathcal{I}_c(\alpha; \hat{\theta}_{M-1}) - \mathcal{I}_{c, M-1}^{(\max)}(\hat{\theta}_{M-1}) \right| < \epsilon_{\text{threshold}}, \quad (5.78)$$

where $\epsilon_{\text{threshold}}$ is a small positive number, and

$$\mathcal{I}_{c, M-1}^{(\max)}(\hat{\theta}_{M-1}) := \max_{\alpha \in D_{M-1}} \mathcal{I}_c(\alpha; \hat{\theta}_{M-1}). \quad (5.79)$$

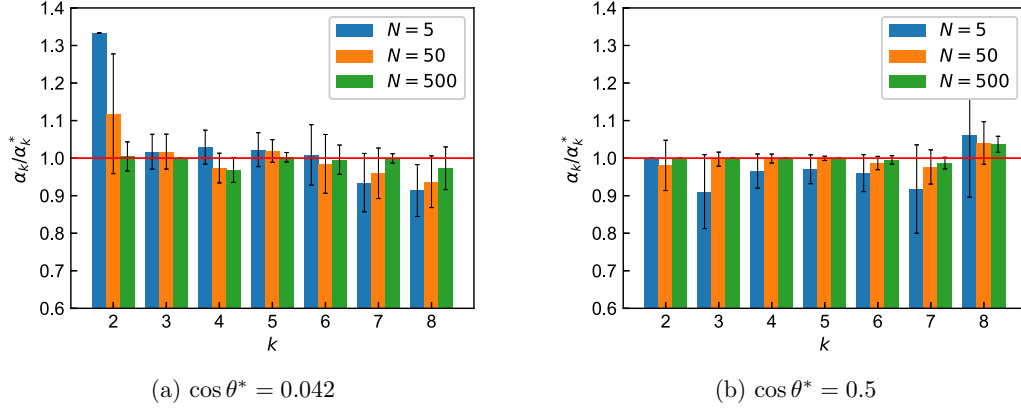


Figure 5.3: Convergence of the amplified levels α_k with respect to N . The color bars represent the average of the realized amplified levels normalized by its asymptotic values α_k^* over 3000 trials, and the error bars denote the standard deviation. With a fixed value of k , the three color bars show the results corresponding to $N = 5$ (left, blue), $N = 50$ (center, orange), and $N = 500$ (right, green), respectively. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

Note that α_M^* is defined as the minimum value of $\bar{D}_{M-1}$ for θ^* . In the limit of $\hat{\theta}_{M-1} \rightarrow \theta^*$, $\bar{D}_{M-1}$ for $\hat{\theta}_{M-1}$ has the same elements in $\bar{D}_{M-1}$ for θ^* , and therefore the implication in Eq. (5.76) holds. From the assumption $P[E(M-1)] \rightarrow 1$, there exists $N_1^* \in \mathbb{N}$ for all $\eta > 0$ such that

$$P \left[E(M-1) \cap \left\{ \left| \hat{\theta}_{M-1} - \theta^* \right| < \epsilon' \right\} \right] > 1 - \eta \quad (5.80)$$

holds for all $N > N_1^*$. The implication in Eq. (5.76) yields the following inequality

$$\begin{aligned} 1 - \eta &< P \left[E(M-1) \cap \left\{ \left| \hat{\theta}_{M-1} - \theta^* \right| < \epsilon' \right\} \right] \\ &\leq P \left[E(M-1) \cap \left\{ \alpha_M(\mathbf{X}_{M-1}) = \alpha_M^* \right\} \right] \leq P \left[\left\{ \alpha_k(\mathbf{X}_{k-1}) = \alpha_k^* \forall k \right\} \right]. \end{aligned} \quad (5.81)$$

Note that if $\alpha_k(\mathbf{x}_{k-1}) = \alpha_k^*$ holds for all $k = 1, 2, \dots, M$, then the likelihood function $\mathcal{L}_M(\theta; \mathbf{x}_M)$ can be written as

$$\mathcal{L}_M(\theta; \mathbf{x}_M) = \prod_{m=1}^M F_m \left(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta \right) = \prod_{m=1}^M F_m \left(x^{(m)}; \alpha_m^*, \theta \right). \quad (5.82)$$

On the other hand, Lemma 2 guarantees the existence of some natural number N_2^* such that

$$P[\mathbf{A}] > 1 - \eta, \quad (5.83)$$

for all $N > N_2^*$, where an event $\mathbf{A}$ is defined as

$$\mathbf{A}: \prod_{m=1}^M F_m \left(X^{(m)}; \alpha_m^*, \theta \right) \text{ has a unique maximum point } \hat{\theta}_M \text{ such that } \hat{\theta}_M \rightarrow \theta^*.$$

From Eqs. (5.81) and (5.83), if we choose $N > \max\{N_1^*, N_2^*\}$, then

$$\begin{aligned} P[E(M)] &= P \left[\left\{ \mathcal{L}_M(\theta; \mathbf{X}_M) \text{ has a unique maximum point } \hat{\theta}_M \text{ s.t. } \hat{\theta}_M \rightarrow \theta^* \right\} \cap \left\{ \alpha_k(\mathbf{X}_{k-1}) = \alpha_k^* \forall k = 1, 2, \dots, M \right\} \right] \\ &= P \left[\left\{ \prod_{m=1}^M F_m \left(X^{(m)}; \alpha_m^*, \theta \right) \text{ has a unique maximum point } \hat{\theta}_M \text{ s.t. } \hat{\theta}_M \rightarrow \theta^* \right\} \cap \left\{ \alpha_k(\mathbf{X}_{k-1}) = \alpha_k^* \forall k \right\} \right] \\ &= P \left[\mathbf{A} \cap \left\{ \alpha_k(\mathbf{X}_{k-1}) = \alpha_k^* \forall k \right\} \right] > 1 - 2\eta. \end{aligned} \quad (5.84)$$

In conclusion, we have proved that $P[E(M)] \rightarrow 1$ also holds, and the mathematical induction establishes the theorem. $\square$

To prove Lemma 2 or Eq. (5.83), we first establish the existence of a solution for the likelihood equation on F_m .

Lemma 13. *A natural number $2 \leq m \leq M$ is fixed arbitrarily. Suppose $P[\alpha_m(\mathbf{X}_{m-1}) = \alpha_m^*] \rightarrow 1$ as the number of measurements N increases, then $F_m(X^{(m)}; \alpha_m^*, \theta)$ has a unique maximum point $\hat{\theta}^{(m)}$ in Θ_m such that $\hat{\theta}^{(m)} \rightarrow \theta^*$ with the probability approaching 1.*

Proof. We first show

$$P[F_m(X^{(m)}; \alpha_m^*, \theta^*) > F_m(X^{(m)}; \alpha_m^*, \theta)] \rightarrow 1 \quad (5.85)$$

holds for any $\theta \in \Theta_m$, $\theta \neq \theta^*$. The l.h.s can be transformed as

$$\begin{aligned} & P[F_m(X^{(m)}; \alpha_m^*, \theta^*) > F_m(X^{(m)}; \alpha_m^*, \theta)] \\ &= \sum_{x^{(m)}} \sum_{\mathbf{x}_{m-1}} \mathcal{L}_{m-1}(\mathbf{x}_{m-1}; \theta^*) F_m(x^{(m)}; \alpha_m(\mathbf{x}_{m-1}), \theta^*) \mathbf{1}\{F_m(X^{(m)}; \alpha_m^*, \theta^*) > F_m(X^{(m)}; \alpha_m^*, \theta)\}(x^{(m)}) \\ &= \sum_{\alpha'_m \in D_{m-1}} \sum_{\substack{\mathbf{x}_{m-1}: \\ \alpha_m(\mathbf{x}_{m-1}) = \alpha'_m}} \mathcal{L}_{m-1} \sum_{x^{(m)}} F_m(x^{(m)}; \alpha'_m, \theta^*) \mathbf{1}\{F_m(X^{(m)}; \alpha_m^*, \theta^*) > F_m(X^{(m)}; \alpha_m^*, \theta)\}(x^{(m)}) \\ &= \sum_{\alpha'_m \in D_{m-1}} P[\alpha_m(\mathbf{X}_{m-1}) = \alpha'_m] \sum_{x^{(m)}} F_m(x^{(m)}; \alpha'_m, \theta^*) \mathbf{1}\{F_m(X^{(m)}; \alpha_m^*, \theta^*) > F_m(X^{(m)}; \alpha_m^*, \theta)\}(x^{(m)}), \end{aligned} \quad (5.86)$$

where $\mathbf{1}\{\cdot\}(x)$ denotes an indicator function. By assumption, the summation for α'_m of the last line vanishes except for the term with $\alpha'_m = \alpha_m^*$ as N increases. Thus, the remaining term can be written as

$$P[F_m(Y^{(m)}; \alpha_m^*, \theta^*) > F_m(Y^{(m)}; \alpha_m^*, \theta)], \quad (5.87)$$

where $Y^{(m)}$ denotes a random variable following

$$\text{Bin}(N, \mathbf{P}_{\mathcal{D}}(\alpha_m^*; \theta^*)). \quad (5.88)$$

Since the probability $\mathbf{P}_{\mathcal{D}}(\alpha_m^*; \theta)$ is an injective function with respect to $\theta \in \Theta_m$; and its support is independent to θ , these conditions yield

$$P[F_m(Y^{(m)}; \alpha_m^*, \theta^*) > F_m(Y^{(m)}; \alpha_m^*, \theta)] \rightarrow 1, \quad (5.89)$$

as $N \rightarrow \infty$. We can prove this from the law of large numbers for independent Bernoulli trials on $Y^{(m)}$ and the non-negativity of the Kullback-Leibler divergence. Combining Eqs. (5.89), (5.86), we complete the proof of Eq. (5.85).

By definition of Θ_m , we can choose a small positive number δ such that $(\theta^* - \delta, \theta^* + \delta) \subset \Theta_m$. Let $\mathbf{B}_{\pm}$ be the following events

$$\mathbf{B}_{\pm}: F_m(X^{(m)}; \alpha_m^*, \theta^*) > F_m(X^{(m)}; \alpha_m^*, \theta^* \pm \delta) \text{ holds.}$$

Considering $x^{(m)} \in \mathbf{B}_+ \cap \mathbf{B}_-$, the equation

$$\frac{\partial}{\partial \theta} F_m(x^{(m)}; \alpha_m^*, \theta) = 0, \quad (5.90)$$

has solutions, and we write $\hat{\theta}^{(m)}$ as the solution closest to θ^* among them. Since the difference between $\hat{\theta}^{(m)}$ and θ^* is less than δ ; and for all $\epsilon > 0$ there exists $N^* \in \mathbb{N}$ such that $P[\mathbf{B}_{\pm}] > 1 - \epsilon$ for all $N > N^*$ by Eq. (5.85), we obtain

$$P[|\hat{\theta}^{(m)} - \theta^*| < \delta] \geq P[\mathbf{B}_+ \cap \mathbf{B}_-] > 1 - 2\epsilon. \quad (5.91)$$

In addition, the likelihood equation Eq. (5.90) has a unique solution maximizing $F_m(x^{(m)}; \alpha_m^*, \theta)$ in Θ_m if a solution of Eq. (5.90) exists in Θ_m . Therefore, $\hat{\theta}^{(m)}$ is a unique maximum point of $F_m(x^{(m)}; \alpha_m^*, \theta)$ in Θ_m . $\square$

Applying Lemma 1 to all the functions $\{F_m\}$, we establish the existence of a unique maximum point of the product function $\Pi_m F_m$.

Lemma 14. *Suppose $P[\alpha_k(\mathbf{X}_{k-1}) = \alpha_k^*] \rightarrow 1$ for all $k = 2, \dots, M$ and $\hat{\theta}^{(1)} \rightarrow \theta^*$ (convergence in probability). Then, there exists a unique maximum point $\hat{\theta}_M$ of the product function $\Pi_{m=1}^M F_m(X^{(m)}; \alpha_m^*, \theta)$ in $(0, \pi)$ such that $\hat{\theta}_M \rightarrow \theta^*$ with the probability approaching 1 as $N \rightarrow \infty$.*

Proof. From the assumption and Lemma 1, we can choose $N^* \in \mathbb{N}$ for all $\epsilon > 0$ such that the probability of occurrence of the following event **C** is bigger than $1 - \epsilon$ for all $N > N^*$;

$$\mathbf{C}: \text{For all } m, \hat{\theta}^{(m)} \text{ exists in } \Theta_M.$$

Here, $\hat{\theta}^{(m)}$ denotes a (unique) maximum point of F_m in Θ_M . Actually, F_m is globally maximized at the point $\hat{\theta}^{(m)}$ because of the periodicity of F_m :

$$\max_{\theta \in (0, \pi)} F_m(x^{(m)}; \alpha_m^*, \theta) = F_m(x^{(m)}; \alpha_m^*, \hat{\theta}^{(m)}). \quad (5.92)$$

Let $\hat{\theta}_{\max}$ ($\hat{\theta}_{\min}$) be the maximum (minimum) value of a set $\{\hat{\theta}^{(m)}\}$. Since the product function $\Pi_{m=1}^M F_m$ is a continuous function with respect to θ , it has a maximum point $\hat{\theta}_M$ in $[\hat{\theta}_{\min}, \hat{\theta}_{\max}]$. $\hat{\theta}_{\max}$ and $\hat{\theta}_{\min}$ themselves cannot be a maximum point, because the gradient at these points does not vanish as follows. Since $\hat{\theta}_{\min} \leq \hat{\theta}^{(m)}$ holds, we obtain

$$\begin{aligned} \left. \frac{\partial}{\partial \theta} \prod_{m=1}^M F_m(x^{(m)}; \alpha_m^*, \theta) \right|_{\hat{\theta}_{\min}} &= \sum_{k=1}^M F_1 \cdots F_{k-1} \left. \frac{\partial F_k}{\partial \theta} \right|_{\hat{\theta}_{\min}} F_{k+1} \cdots F_M \\ &= \sum_{k: \hat{\theta}^{(k)} \neq \hat{\theta}_{\min}} F_1 \cdots F_{k-1} \left. \frac{\partial F_k}{\partial \theta} \right|_{\hat{\theta}_{\min}} F_{k+1} \cdots F_M > 0, \end{aligned} \quad (5.93)$$

where $F_m > 0$ and $\partial_\theta F_m > 0$ for $\theta < \hat{\theta}^{(m)}$ lead to the final inequality. In the same way, we can prove that the gradient at $\hat{\theta}_{\max}$ is negative. Therefore, we obtain

$$\begin{aligned} P \left[\text{there exists a maximum point } \hat{\theta}_M \text{ of } \prod_{m=1}^M F_m(X^{(m)}; \alpha_m^*, \theta) \text{ in } \Theta_M \text{ such that } \hat{\theta}_M \xrightarrow{p} \theta^* \right] \\ \geq P[\mathbf{C}] > 1 - \epsilon, \end{aligned} \quad (5.94)$$

where we use the fact that $\hat{\theta}_{\min}, \hat{\theta}_{\max} \rightarrow \theta^*$ yields $\hat{\theta}_M \rightarrow \theta^*$.

Finally, we show that $\Pi_m F_m$ is globally maximized at $\hat{\theta}_M$. For simplicity, we write $F_m(x^{(m)}; \alpha_m^*, \theta)$ as $F_m(\theta)$. By Taylor's theorem, there is a scalar c_m between $\hat{\theta}^{(m)}$ and $\hat{\theta}_M$ such that

$$\ln F_m(\hat{\theta}^{(m)}) - \ln F_m(\hat{\theta}_M) = -\frac{\partial_\theta^2 \ln F_m(c_m)}{2} (\hat{\theta}^{(m)} - \hat{\theta}_M)^2. \quad (5.95)$$

Here, the second derivative in Eq. (5.95) can be upper bounded by N and a positive constant K (independent to N) as

$$-\frac{\partial_\theta^2 \ln F_m(c_m)}{2} (\hat{\theta}^{(m)} - \hat{\theta}_M)^2 < \frac{NK}{2} (\hat{\theta}^{(m)} - \hat{\theta}_M)^2. \quad (5.96)$$

Recalling that $F_1(\theta)$ has a single peak in $(0, \pi)$, for an arbitrary $\lambda \notin \Theta_M$, we obtain

$$\frac{\ln F_1(\hat{\theta}_M) - \ln F_1(\lambda)}{|\hat{\theta}_M - \lambda|} = \frac{N}{2} \left| \frac{p_q \sin(c_1)(x^{(1)}/N - \mathbf{P}_{\mathcal{D}}(1; c_1))}{\mathbf{P}_{\mathcal{D}}(1; c_1)[1 - \mathbf{P}_{\mathcal{D}}(1; c_1)]} \right| > \frac{NL}{2}, \quad (5.97)$$

where c_1 is a scalar between λ and $\hat{\theta}_M$, and $L > 0$ is a positive constant. Combining Eqs. (5.92) and (5.95)–(5.97), the following inequalities hold:

$$\sum_{m=1}^M \ln \frac{F_m(\hat{\theta}_M)}{F_m(\lambda)} \geq \ln \frac{F_1(\hat{\theta}_M)}{F_1(\lambda)} + \sum_{m=2}^M \ln \frac{F_m(\hat{\theta}_M)}{F_m(\hat{\theta}^{(m)})} > \frac{N}{2} \left[L |\lambda - \hat{\theta}_M| - K \sum_{m=2}^M (\hat{\theta}^{(m)} - \hat{\theta}_M)^2 \right]. \quad (5.98)$$

For sufficiently large N , the second term in the square bracket becomes small compared to the first term, and therefore we conclude that

$$\prod_m F_m(\hat{\theta}_M) > \prod_m F_m(\lambda). \quad (5.99)$$

Consequently, the maximum of $\prod_{m=1}^M F_m$ in Θ_M becomes the global maximum in $(0, \pi)$. $\square$

Finally, we prove that, when N is large, the asymptotic variance of our estimator can achieve the classical Cramér-Rao lower bound.

Theorem 19. *If $\hat{\theta}_M$ is a maximum likelihood estimator on $\mathcal{L}_M(\theta; \mathbf{X}_M)$ such that $\hat{\theta}_M \rightarrow \theta^*$ (convergence in probability), then the following convergence holds;*

$$\sqrt{\mathcal{I}_{c,\text{tot}}^*(\theta^*)} (\cos \hat{\theta}_M - \cos \theta^*) \rightarrow \mathcal{N}(0, \sin^2 \theta^*), \quad (5.100)$$

where $\mathcal{N}$ denotes a centered normal distribution with variance $\sin^2 \theta^*$ and $\rightarrow$ means the convergence in distribution as N increases.

Proof. Using the Taylor's theorem, we can expand $\partial_\theta \ln \mathcal{L}_M(\theta; \mathbf{X}_M)$ around $\theta = \theta^*$ as

$$\begin{aligned} \partial_\theta \ln \mathcal{L}_M(\theta; \mathbf{X}_M) &= \partial_\theta \ln \mathcal{L}_M(\theta^*; \mathbf{X}_M) \\ &+ \partial_\theta^2 \ln \mathcal{L}_M(\theta^*; \mathbf{X}_M)(\theta - \theta^*) + \frac{1}{2} \partial_\theta^3 \ln \mathcal{L}_M(\theta'; \mathbf{X}_M)(\theta - \theta^*)^2, \end{aligned} \quad (5.101)$$

where θ' denotes some real number between θ and θ^* . Since $\hat{\theta}_M$ is a root of the likelihood equation, we obtain

$$\sqrt{\mathcal{I}_{c,\text{tot}}^*(\theta^*)} (\hat{\theta}_M - \theta^*) = \frac{A}{B - C}, \quad (5.102)$$

where for simplicity we introduced the random variables A, B , and C defined as

$$A := (\mathcal{I}_{c,\text{tot}}^*(\theta^*))^{-1/2} \partial_\theta \ln \mathcal{L}_M(\theta^*; \mathbf{X}_M), \quad (\mathcal{I}_{c,\text{tot}}^*(\theta^*))^{-1} \partial_\theta^2 \ln \mathcal{L}_M(\theta^*; \mathbf{X}_M),$$

$$C := (2\mathcal{I}_{c,\text{tot}}^*(\theta^*))^{-1} \partial_\theta^3 \ln \mathcal{L}_M(\theta'; \mathbf{X}_M)(\hat{\theta}_M - \theta^*).$$

Here, we show that A converges to the standard normal distribution. To apply Lemma 3, we first calculate the characteristic function of A replacing $\mathbf{X}_M, \alpha_m \rightarrow \mathbf{Y}_M, \alpha_m^*$, where $\mathbf{Y}_M = (Y^{(1)}, \dots, Y^{(M)})$ is a random vector whose element follows Binomial distribution $\text{Bin}(N, \mathbf{P}_D(\alpha_m^*; \theta^*))$ independently. That is, for $t \in \mathbb{R}$ we have

$$\begin{aligned} \mathbb{E} \left[\exp(it A |_{\mathbf{X}_M, \alpha_m \rightarrow \mathbf{Y}_M, \alpha_m^*}) \right] &= \mathbb{E} \left[\exp \left(\frac{it \sum_{m=1}^M \partial_\theta \ln F_m(Y^{(m)}; \alpha_m^*, \theta^*)}{\sqrt{\mathcal{I}_{c,\text{tot}}^*(\theta^*)}} \right) \right] \\ &= \prod_{m=1}^M \left(\mathbb{E} \left[\exp \left(\frac{it \partial_\theta \ln f_m(Y_1^{(m)}; \alpha_m^*, \theta^*)}{\sqrt{\mathcal{I}_{c,\text{tot}}^*(\theta^*)}} \right) \right] \right)^N = \prod_{m=1}^M \left(1 - \frac{t^2}{2N} \frac{\mathcal{I}_c(\alpha_m^*; \theta^*)}{\sum_{m=1}^M \mathcal{I}_c(\alpha_m^*; \theta^*)} + o\left(\frac{1}{N}\right) \right)^N \\ &\rightarrow \exp \left(-\frac{t^2}{2} \right), \end{aligned} \quad (5.103)$$

where $Y_1^{(m)}$ denotes a random variable following Bernoulli distribution with the success probability $\mathbf{P}_D(\alpha_m^*; \theta^*)$, and f_m is the probability function of $Y_1^{(m)}$. Thus, Lemma 3 yields

$$\lim_{N \rightarrow \infty} \mathbb{E} [\exp(itA)] = \exp \left(-\frac{t^2}{2} \right). \quad (5.104)$$

Therefore, $A \rightarrow \mathcal{N}(0, 1)$ holds.

Next, we consider the convergence of B . According to the Chebyshev inequality, for all $c > 0$ we obtain

$$P[|B - 1| > c] \leq c^{-2} \mathbb{E}[(B - 1)^2]. \quad (5.105)$$

Here, $(B - 1)^2$ is upper bounded, and the expectation of $(B - 1)^2$ replacing $\mathbf{X}_M, \alpha_m \rightarrow \mathbf{Y}_M, \alpha_m^*$ also converges to 0 as follows. For simplicity, we define $W_m(Y^{(m)})$ and $w_m(Y^{(m)})$ as

$$W_m(Y^{(m)}) := \partial_\theta^2 \ln F_m(Y^{(m)}; \alpha_m^*, \theta^*) - \mathbb{E}[\partial_\theta^2 \ln F_m(Y^{(m)}; \alpha_m^*, \theta^*)],$$

$$w_m(Y_1^{(m)}) := \partial_\theta^2 \ln f_m(Y_1^{(m)}; \alpha_m^*, \theta^*) - \mathbb{E}[\partial_\theta^2 \ln f_m(Y_1^{(m)}; \alpha_m^*, \theta^*)],$$

respectively.

$$\begin{aligned} \mathbb{E}[(B - 1)^2 | \mathbf{X}_M, \alpha_m \rightarrow \mathbf{Y}_M, \alpha_m^*] &= \frac{\text{Var}[\partial_\theta^2 \ln \prod_{m=1}^M F_m(Y^{(m)}; \alpha_m^*, \theta^*)]}{[\mathcal{I}_{c, \text{tot}}^*(\theta^*)]^2} \\ &= \frac{\sum_{n,m=1}^M \mathbb{E}[W_m(Y^{(m)})W_n(Y^{(n)})]}{[\mathcal{I}_{c, \text{tot}}^*(\theta^*)]^2} = \frac{N}{[\mathcal{I}_{c, \text{tot}}^*(\theta^*)]^2} \sum_{m=1}^M \mathbb{E}[w_m^2(Y_1^{(m)})] \\ &= \frac{1}{N} \frac{\sum_{m=1}^M \mathbb{E}[w_m^2(Y_1^{(m)})]}{[\sum_{m=1}^M \mathcal{I}_c(\alpha_m^*; \theta^*)]^2} \rightarrow 0, \end{aligned} \quad (5.106)$$

where we used $\mathcal{I}_{c, \text{tot}}^*(\theta^*) = -\mathbb{E}[\partial_\theta^2 \ln \prod_{m=1}^M F_m]$ for the first equality. Thus, Lemma 3 establishes that $B \rightarrow 1$ holds.

Since $N^{-1} \partial_\theta^3 \ln \mathcal{L}_M(\theta'; \mathbf{x}_M)$ is also upper bounded, $\hat{\theta}_M - \theta^* \rightarrow 0$ yields $C \rightarrow 0$. To sum up the above asymptotic properties, we conclude that the following convergence holds;

$$\sqrt{\mathcal{I}_{c, \text{tot}}^*(\theta^*)}(\hat{\theta}_M - \theta^*) \rightarrow \mathcal{N}(0, 1). \quad (5.107)$$

In addition, the delta method [144] yields a more suitable expression

$$\sqrt{\mathcal{I}_{c, \text{tot}}^*(\theta^*)}(\cos \hat{\theta}_M - \cos \theta^*) \rightarrow \mathcal{N}(0, \sin^2 \theta^*). \quad (5.108)$$

□

To complete the proof of Theorem 17, we finally show the following lemma.

Lemma 15. *Let $h(\mathbf{x}_M, \alpha_2, \dots, \alpha_M)$ be an upper bounded function on $\mathbf{x}_M \in \{0, 1, \dots, N\}^M$ and $Y^{(m)}$ be an independent random variable following Binomial distribution $\text{Bin}(N, \mathbf{P}_{\mathcal{D}}(\alpha_m^*; \theta^*))$ ($m = 1, 2, \dots, M$). Suppose that $\mathbb{E}[h(\mathbf{Y}_M, \alpha_2^*, \dots, \alpha_M^*)]$ converges to some constant β as $N \rightarrow \infty$, then the following convergence holds*

$$\lim_{N \rightarrow \infty} \mathbb{E}[h(\mathbf{X}_M, \alpha_2(\mathbf{X}_1), \dots, \alpha_M(\mathbf{X}_{M-1}))] = \beta. \quad (5.109)$$

Proof. Expanding the left hand side in Eq. (5.109), we obtain

$$\begin{aligned}
\mathbb{E}[h(\mathbf{X}_M, \alpha_2(\mathbf{X}_1), \dots, \alpha_M(\mathbf{X}_{M-1}))] &= \sum_{\mathbf{x}_M} \mathcal{L}_M(\mathbf{x}_M; \theta^*) h(\mathbf{x}_M, \alpha_2(\mathbf{x}_1), \dots, \alpha_M(\mathbf{x}_{M-1})) \\
&= \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) = \alpha_k^* \forall k}} \mathcal{L}_M(\mathbf{x}_M; \theta^*) h(\mathbf{x}_M, \alpha_2(\mathbf{x}_1), \dots, \alpha_M(\mathbf{x}_{M-1})) \\
&\quad + \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \mathcal{L}_M(\mathbf{x}_M; \theta^*) h(\mathbf{x}_M, \alpha_2(\mathbf{x}_1), \dots, \alpha_M(\mathbf{x}_{M-1})) \\
&= \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) = \alpha_k^* \forall k}} \prod_{m=1}^M F_m(x^{(m)}; \alpha_m^*, \theta^*) h(\mathbf{x}_M, \alpha_2^*, \dots, \alpha_M^*) \\
&\quad + \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \mathcal{L}_M(\mathbf{x}_M; \theta^*) h(\mathbf{x}_M, \alpha_2(\mathbf{x}_1), \dots, \alpha_M(\mathbf{x}_{M-1})) \\
&= \mathbb{E}[h(\mathbf{Y}_M, \alpha_2^*, \dots, \alpha_M^*)] \\
&\quad + \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \mathcal{L}_M(\mathbf{x}_M; \theta^*) h(\mathbf{x}_M, \alpha_2(\mathbf{x}_1), \dots, \alpha_M(\mathbf{x}_{M-1})) \\
&\quad - \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \prod_{m=1}^M F_m(x^{(m)}; \alpha_m^*, \theta^*) h(\mathbf{x}_M, \alpha_2^*, \dots, \alpha_M^*). \tag{5.110}
\end{aligned}$$

The above expression yields the following inequality

$$\begin{aligned}
&|\mathbb{E}[h(\mathbf{X}_M, \alpha_2(\mathbf{X}_1), \dots, \alpha_M(\mathbf{X}_{M-1}))] - \beta| \\
&\leq |\mathbb{E}[h(\mathbf{Y}_M, \alpha_2^*, \dots, \alpha_M^*)] - \beta| \\
&\quad + \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \mathcal{L}_M(\mathbf{x}_M; \theta^*) |h(\mathbf{x}_M, \alpha_2(\mathbf{x}_1), \dots, \alpha_M(\mathbf{x}_{M-1}))| \\
&\quad + \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \prod_{m=1}^M F_m(x^{(m)}; \alpha_m^*, \theta^*) |h(\mathbf{x}_M, \alpha_2^*, \dots, \alpha_M^*)| \\
&\leq |\mathbb{E}[h(\mathbf{Y}_M, \alpha_2^*, \dots, \alpha_M^*)] - \beta| + \gamma \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) \neq \alpha_k^* \exists k}} \left(\mathcal{L}_M(\mathbf{x}_M; \theta^*) + \prod_{m=1}^M F_m(x^{(m)}; \alpha_m^*, \theta^*) \right), \tag{5.111}
\end{aligned}$$

where γ is the upper bound of h . The convergence of amplified levels implies that for all $\epsilon > 0$ there exists $N^* \in \mathbb{N}$ such that

$$P[\alpha_k(\mathbf{X}_{k-1}) = \alpha_k^* \forall k] = \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) = \alpha_k^* \forall k}} \mathcal{L}_M(\mathbf{x}_M; \theta^*) = \sum_{\substack{\mathbf{x}_M \\ \alpha_k(\mathbf{x}_{k-1}) = \alpha_k^* \forall k}} \prod_{m=1}^M F_m(x^{(m)}; \alpha_m^*, \theta^*) > 1 - \frac{\epsilon}{2\gamma}, \tag{5.112}$$

for all $N > N^*$. Combining Eqs. (5.111), (5.112) and the assumption for the convergence of $\mathbb{E}[h]$, we obtain Eq. (5.109). $\square$

5.6 Numerical experiment

In this section, we numerically verify the performance of the proposed algorithm. First, in the practical (i.e., a few hundred of) measurement number N , we demonstrate that the asymptotic properties given

by Theorems 16 and 17 hold. Next, by evaluating the asymptotic value of the total Fisher information, we confirm that the proposed algorithm retains a large classical Fisher information regardless of the target value θ^* . Also, we show that the desirable relations given by Eq. (5.50) hold for a system of dozens of qubits, that is, the total classical Fisher information becomes sufficiently close to the total quantum Fisher information for such systems. Finally, we also show the impact of calibration errors.

5.6.1 Asymptotic properties with respect to the number of measurements

To demonstrate the properties of our algorithm, we numerically evaluate the root mean squared error (RMSE) of $\cos \hat{\theta}$ defined as

$$\text{RMSE} [\cos \hat{\theta}] := \sqrt{\mathbb{E}_{\hat{\theta}} \left[\left(\cos \hat{\theta} - \cos \theta^* \right)^2 \right]} \simeq \sqrt{\frac{1}{\#} \sum_{i=1}^{\#} \left(\cos \hat{\theta}[i] - \cos \theta^* \right)^2}, \quad (5.113)$$

where $\#$ is the total number of trials, and $\hat{\theta}[i]$ denotes the estimate of i th trial. In the following, we use $\# = 300$ samples to evaluate the RMSE. We assume the depolarization noise with the parameter $p_q = 0.995$. The number of qubits is set to 20, which corresponds to the system dimension $d = 2^{20}$. We also fix the number of measurements as $N = 500$ for each circuit. To obtain the maximum likelihood estimates, we used a modified brute force method, in which the search domain becomes narrowed as the measurement process proceeds. The amplified level α_k of the k th measurement process ($k = 2, \dots, M$) is determined by solving the optimization problem (5.40), where the maximum likelihood estimate $\hat{\theta}_{k-1}$ has been obtained at the $(k-1)$ th step. Also, in this work, the optimization range is chosen as

$$D_{k-1} := \{2, 2+1, \dots, 2^k\}. \quad (5.114)$$

The exponential increase of the number of elements in D_{k-1} is inspired by the fact that, when there is no noise, the exponential increment sequence $m_k = 2^{k-1}$ achieves the Heisenberg-limited scaling [149]; also see the discussion below Eq. (5.7).

Figure 5.4 shows the relationship between the RMSE Eq. (5.113) and the total number of queries N_q calculated as

$$N_q = N \sum_{k=1}^M \alpha_k. \quad (5.115)$$

Here, we remark that N_q can be rewritten as $N_q = \sum_{l=1}^{M'} \alpha'_l$ for $M' = NM$ and $\alpha'_l = \alpha_k$ if $(k-1)N+1 \leq l \leq kN$ in the notation of Theorem 14. The red dots indicate the RMSE with $M = 3, 4, \dots, 12$ from left to right in each panel. Since N_q is a random variable due to the randomness of α_k , we plot the RMSE as a function of the maximum value of N_q in $\# = 300$ trials, and thus the RMSE dots are overestimated with respect to the number of queries. In addition, the classical Cramér-Rao (CCR) lower bound and the quantum Cramér-Rao (QCR) lower bound are depicted with the solid blue and dashed green lines, respectively; these are calculated by substituting the true value θ^* and the asymptotic sequence $\{\alpha_k^*\}_{k=1}^M$ into $\mathcal{I}_{c/q, \text{tot}}^*(\theta^*)$ in Eqs. (5.47) and (5.48). Note that the difference between the bounds obtained from the $\# = 300$ average of the total Fisher information (not shown in the figure), $\mathcal{I}_{c/q, \text{tot}}(\theta^*)$, and the asymptotic CCR/QCR bounds from $\mathcal{I}_{c/q, \text{tot}}^*(\theta^*)$ can be negligible, which clearly supports Eq. (5.46). The orange dash-dotted and gray dotted lines represent the Heisenberg-limited (HL) scaling $\text{RMSE} = \mathcal{O}(1/N_q)$ and the precision limit given by Eq. (5.27) with $1 - \langle \mathcal{O} \rangle^2$, respectively.

Several important features are observed. First, the CCR and QCR bounds are close to each other; the first part of Eq. (5.50) holds. Note that this good approximation holds even when the target value $\theta^*/2\pi$ is a rational number as shown in the panel (d); actually, we will see in the next subsection that this approximation (i.e., the first part of (5.50)) holds for almost all θ^* in the numerical simulation. We then find that the RMSE almost achieves the CCR and accordingly QCR bounds in all the cases of six target values. That is, $N = 500$ is sufficient to obtain the asymptotic normality (5.49) in the chosen noise condition. We here remark that by taking the number of shots N (originally, we fix $N = 500$ for all iteration steps) as a linear function N_k of the iteration step k , as was done in Ref. [77, 14], the RMSE of our method can nearly achieve the QCR bound even for a regime with smaller total queries ($N_q \sim \mathcal{O}(10^2)$). Moreover, the empirical probability density of our estimator

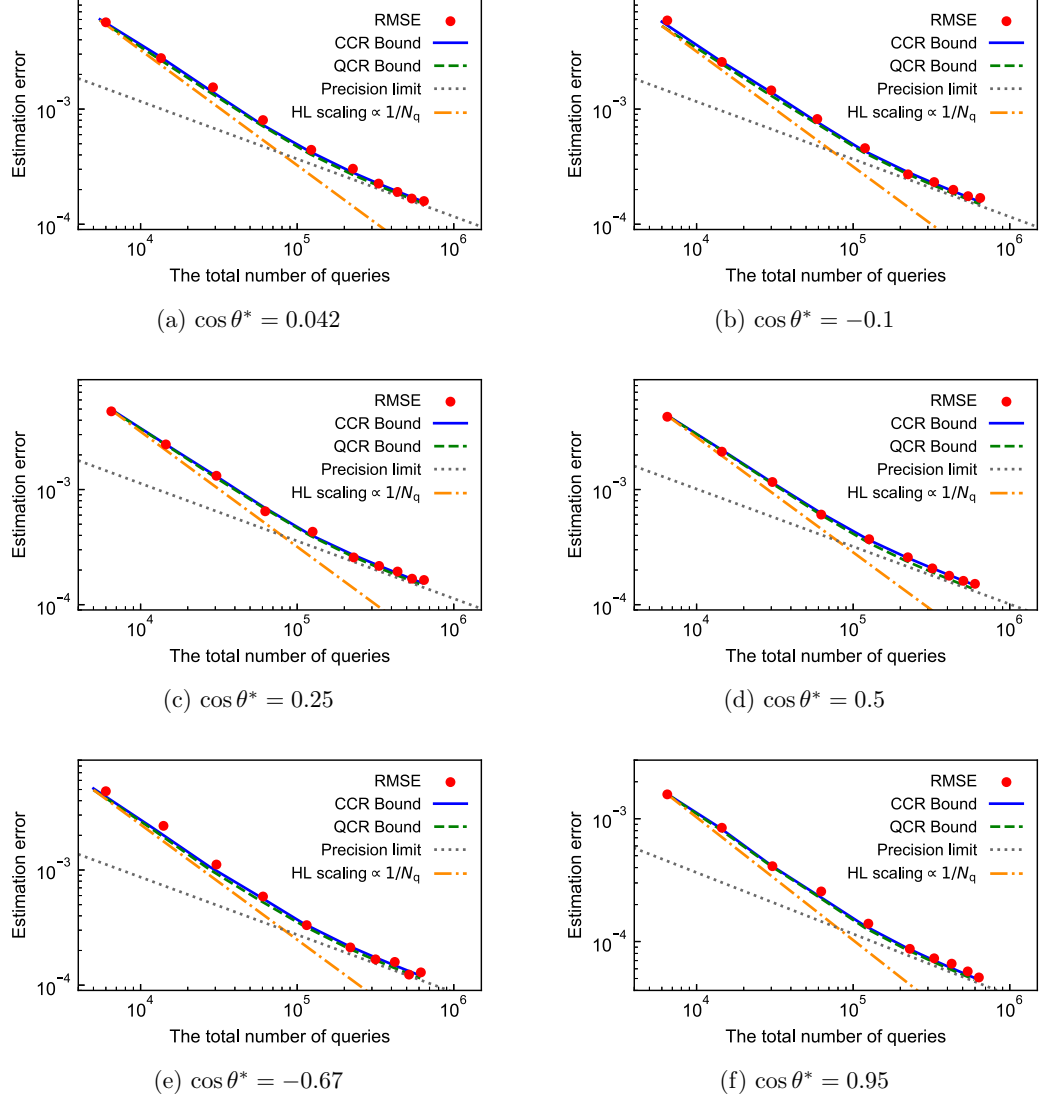


Figure 5.4: The estimation error of $\langle \mathcal{O} \rangle = \cos \theta^*$ and the total number of queries for several target values (a)–(f). The x-axis shows the total number of queries to the state preparation A and $A^\dagger$ in a single trial of our method. The root mean squared error (RMSE) of the estimator $\cos \hat{\theta}$ is estimated by the averaged value over 300 trials, as in Eq. (5.113). Since the number of total queries in a single trial of our method varies stochastically, the corresponding x-axis value of the RMSE dot is taken as the maximum value of total queries in the 300 trials. The blue solid and green dashed lines represent the asymptotic values of CCR/QCR bounds obtained from the corresponding classical/quantum Fisher information $\mathcal{I}_{c/q, \text{tot}}^*(\theta^*)$, respectively. The orange dash-dotted and the gray dotted lines represent the Heisenberg-limited scaling and the precision limit derived in Theorem 14 (in the case of $N_q > \alpha_B = 199$), respectively. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

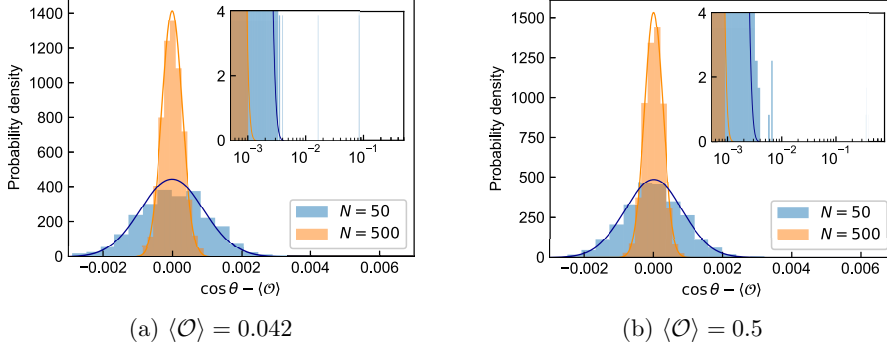


Figure 5.5: The probability density of the maximum likelihood estimator $\cos \hat{\theta}_M$ in the case of $M = 8$. The blue and orange plots correspond to the results with $N = 50$ and $N = 500$, respectively. The histograms represent the empirical probability density over 3000 trials, and the solid lines denote the probability density of the centered normal distribution with the variance corresponding to the asymptotic CCR bound. The upper right panels show the same plots (but the x-axis represents the absolute error) with much wider range in the horizontal axis to show the existence of outliers. The larger N becomes, the sharper the density gets around the target value, which demonstrates the consistency in Theorem 16. Also, the outliers vanish in large N ; thus the empirical density is nearly identical to the normal distribution, which is exactly the asymptotic normality in Theorem 17. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

shown in Fig. 5.5 converges to the normal distribution whose variance corresponds to the asymptotic CCR bound as N increases, which directly demonstrates the asymptotic normality (5.49). Recall that this key property holds because the amplified levels are almost identical to $\{\alpha_k^*\}_{k=1}^M$ due to Theorem 16, which is demonstrated in Fig. 5.3 in Sec. 5.5.5.

Moreover, as the step M increases, we confirm that all of the RMSE, the CCR bound, and the QCR bound nearly achieve the precision limit; the second part of Eq. (5.50) also holds, in addition to the first part of the relations and Theorem 17. Therefore, our algorithm estimates the expectation value $\langle \mathcal{O} \rangle = \cos \theta^*$ with precision close enough to the ultimate limit given by Theorem 14 (or more precisely (5.27)) in the current setup. Note that the RMSE decreases nearly according to the Heisenberg-limited scaling when N_q is small. Actually, this quadratic improvement is more distinct in the case where the depolarization noise is smaller, which certainly recovers the previous result [149] in the ideal amplitude estimation (Fig. 5.6).

Finally, we also provide the case of $p_q = 0.99$ in Fig. 5.6; that is, 1% depolarization noise is added through each operation A or $A^\dagger$, while Fig. 5.4 studies the case of only 0.5% noise. The panels (a) and (b) in Fig. 5.6 show that, as expected, the RMSE loses the quadratic speedup with respect to N_q immediately and approaches the precision limit (i.e., the classical scaling regarding N_q).

5.6.2 Efficiency of the estimator

The classical Fisher information (5.37) vanishes at certain points of the amplified level α . Here we show that, in the same numerical experiment as before, our algorithm can avoid those points thanks to the optimization of α and as a result Eq. (5.50) holds for almost all θ^* . For this purpose, we calculate the asymptotic total Fisher information $\mathcal{I}_{c/q,tot}^*(\theta^*)$ and the asymptotic queries N_q^* with equally distributed 10^5 points of $\cos \theta^*$ in $(0, 1)$, based on Eq. (5.69).

Figure 5.7 shows the target-value dependency of the Cramér-Rao lower bound in terms of the total (asymptotic) classical Fisher information $\mathcal{I}_{c,tot}^*(\theta^*)$, of our method and the following two estimation methods. That is, the classical Fisher information for the standard (i.e., no amplitude amplification) sampling method, which is usually employed for VQE [136] calculations, is defined by Eq. (5.47) with $\alpha_k^* = 1 \forall k$. Also, the classical Fisher information of the method [149, 152, 156], which uses amplitude amplification yet in a non-adaptive way, is defined by Eq. (5.47) with $\alpha_1^* = 1$, $\alpha_k^* = 2^k \forall k \geq 2$, where the corresponding measurement is the even POVM (5.29). Here, the number of measurement processes is chosen as $M = 8$; in this case D_{M-1} ($M \geq 8$) contains α_B under the chosen noise level.

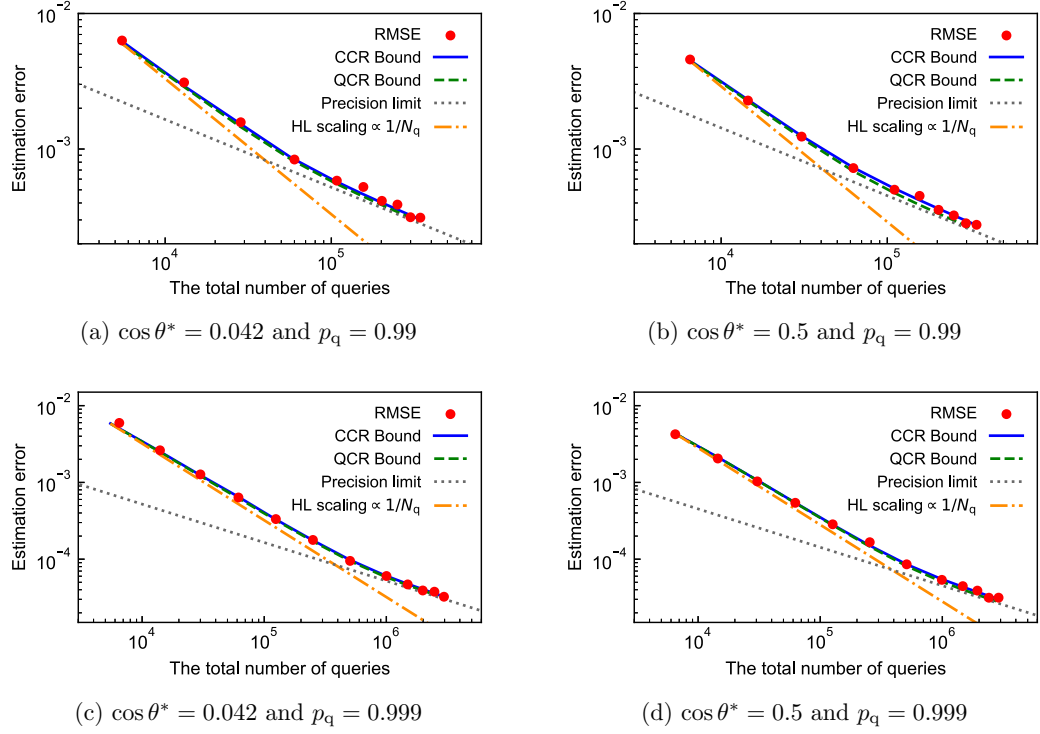


Figure 5.6: The estimation error of $\langle \mathcal{O} \rangle = \cos \theta^*$ and the total number of queries under several depolarization noise parameters. The root mean squared error (RMSE) of the estimator $\cos \hat{\theta}$ is evaluated by 300 trials, and the red dots correspond to $M = 3, 4, \dots, 12$ (or 14) from left to right. Since N_q in Eq. (5.115) is a random variable in our method, the RMSE is plotted as a function of the maximum value of N_q in 300 trials. The blue solid and green dashed lines represent the asymptotic values of CCR/QCR bounds obtained from the corresponding classical/quantum Fisher information $\mathcal{I}_{c/q, \text{tot}}^*(\theta^*)$, respectively. The orange dash-dotted and the gray dotted lines represent the Heisenberg-limited scaling $\text{RMSE} = \mathcal{O}(1/N_q)$ and the precision limit derived in Theorem 14, respectively. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

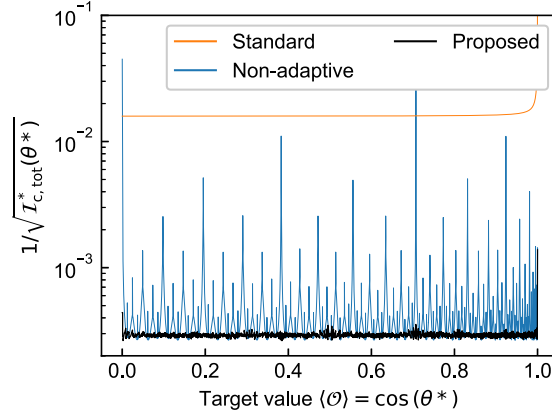


Figure 5.7: Comparison of the Cramér-Rao lower bound in terms of the total (asymptotic) classical Fisher information $\mathcal{I}_{c,\text{tot}}^*(\theta^*)$, of the three estimation methods for several values of θ^* . The black (bottom), orange (top), and blue (middle) lines correspond to our method, the standard classical sampling method with $\alpha_k^* = 1$, and the non-adaptive method with the predefined sequence $\alpha_1^* = 1$, $\alpha_k^* = 2^k$ ($k \geq 2$), respectively. Note that the non-adaptive method uses more queries than our method because $\alpha_k^* = 2^k$ is the maximum number in the optimization range D_{k-1} ; the blue line is lower than the black line at some points of θ^* . Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

First, the black (bottom) line reflects that our classical Fisher information does not almost depend on the target value, meaning the robustness of the estimator. In addition, compared to the standard method, we observe an improvement of about two orders of magnitude for all target values in our Fisher information. The improvement depends on the noise level, and therefore the estimation efficiency is further accelerated if the noise becomes smaller. In contrast to our total classical Fisher information, the performance of the non-adaptive method depicted with the blue (middle) line heavily depends on the target value. Note that, since the total number of queries in the non-adaptive method is bigger than that in our method, there exist some target values $\cos \theta^*$ such that the total Fisher information of ours is less than that of the non-adaptive method. On the other hand, there are some target values for which the estimation error bound deviates by about two orders of magnitude from ours. Therefore, when the amplified level $\{\alpha_k\}_{k=1}^M$ is chosen non-adaptively, even if the asymptotic properties of maximum likelihood estimators hold, the estimation efficiency significantly decreases.

Finally, Fig. 5.8 shows the ratio of the asymptotic total classical and quantum Fisher information in our method. For almost all target values, the ratio of $\mathcal{I}_{q,\text{tot}}^*$ and $\mathcal{I}_{c,\text{tot}}^*$ is close to 1, which is in line with the results discussed in the previous subsection. Moreover, as expected from the objective function in Eq. (5.40), relaxing the regularization as $\delta \rightarrow 1$, we confirm that the ratio gets closer to 1; see Fig. 5.2, meaning that the first relation of Eq. (5.50) holds. Recall now that, in general, the QCR bound (i.e., the ultimate limit of the estimation precision) can only be achieved when using the optimal measurement tailored for a given quantum state. Therefore, the result obtained here means that our estimation method selects the almost optimal measurements (or POVM) in the sense of Fisher information. In addition, the ratio of $N_q^* \mathcal{I}_q(\alpha_B)/\alpha_B$ and $\mathcal{I}_{q,\text{tot}}^*$ gets closer to 1 as M increases, meaning that the second relation in (5.50) holds for almost all target values. Since the right hand side of the second relation in Eq. (5.50) is equivalent to the precision limit by Theorem 14, our algorithm therefore provides the almost best usage of A and $A^\dagger$ in estimating the expectation value from noisy quantum devices.

For the target values $\theta^* = \pi/2, \pi/4$, the difference of $\mathcal{I}_{q,\text{tot}}^*$ and $\mathcal{I}_{c,\text{tot}}^*$ is larger than that for other target values. Note that such a difference is also distinct at $\theta^* = \pi/3, \pi/6$ when $\delta \rightarrow 1$; see Fig. 5.2. This is because the values of $(m+1)\theta^* \pmod{2\pi}$ obtained from the amplitude amplification are limited when $\theta^*/2\pi$ is a rational number. In this case, $(m+1)\theta^* \pmod{2\pi}$ cannot arbitrarily get close to $\pi/2$ or $3\pi/2$, meaning that the classical Fisher information cannot arbitrarily get close to the quantum Fisher information; see Theorem 15 or Sec. 5.5.5. The peaks at these rational points in Fig. 5.8 are

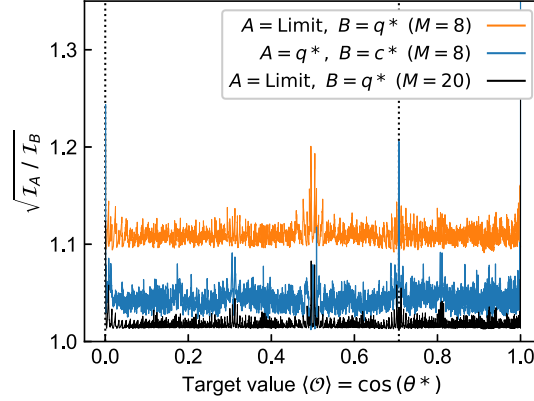


Figure 5.8: Comparison of the asymptotic total classical and quantum Fisher information of our method in the case of 20-qubit system. The middle blue line represents the ratio between the quantum and classical Fisher information: $\mathcal{I}_{q,\text{tot}}^*(\theta^*)$ over $\mathcal{I}_{c,\text{tot}}^*(\theta^*)$ (indicated by $A = q^*$ and $B = c^*$). On the other hand, the orange (top) and black (bottom) lines represent $N_q^* \mathcal{I}_q(\alpha_B) / \alpha_B$ over $\mathcal{I}_{q,\text{tot}}^*(\theta^*)$ (indicated by $A = \text{Limit}$ and $B = q^*$) in the case of $M = 8$ and $M = 20$, respectively. The vertical dotted lines represent the target values corresponding to $\theta^* = \pi/j$, $j = 2, 4$ from left to right. Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

very sharp, because the values of $(m+1)\theta^* \pmod{2\pi}$ fill the domain $[0, 2\pi)$ exponentially fast in the optimization of amplified levels with the chosen $\{D_k\}_{k=1}^{M-1}$, when the target values shift slightly from these points. Thus, these anomalous cases can be ignored in practice.

5.6.3 Impact of calibration errors

Here, we provide the numerical simulations to clarify the impact of calibration errors of p_q on the performance of our method. If an estimate of p_q contains some calibration error, the resulting estimator can be biased even in the asymptotic case. The experimental settings are the same as those in Sec. 5.6.1. Figure 5.9 shows that in the current setting, the $\pm 0.1\%$ calibration errors can be ignored up to $\sim 10^5$ queries. In the case of the $\pm 0.5\%$ calibration errors, the estimation performance deteriorates even for small queries $\sim 6 \times 10^3$. The results indicate that our method requires a relatively high-precision estimate of p_q that have the precision comparable to the target precision of $\cos \theta^*$. Note that, in the quantum amplitude estimation problem, the previous paper [152] provides the (non-adaptive) multiparameter estimation method with respect to both a target parameter and a noise parameter. In addition, they also show the robustness of target parameter estimation against the estimation error of the noise parameter in terms of the Fisher information matrix. Although a similar extension of our adaptive method is possible, it falls outside the scope of this study.

5.7 Concluding remarks and outlook

In this chapter, we proposed a quantum-enhanced single expectation value estimation algorithm under a global depolarization noise that asymptotically achieves the quantum Cramér-Rao bound (and the precision limit), when the target quantum state consists of a modest number of qubits. Here, we derive the precision limit by evaluating the quantum Fisher information for the noisy quantum states with quantum-enhanced resolution regarding the target expectation value, in the setup without demanding controlled amplifications. This algorithm employs a modified maximum likelihood estimation consisting of the amplitude amplification and the adaptive measurement; the latter is derived from the ideal POVM achieving the quantum Fisher information, and notably, it can be implemented without any knowledge of the state preparation oracle for the target state. The measurement and the number of queries per circuit of the amplification operators are adaptively optimized for enhancing the

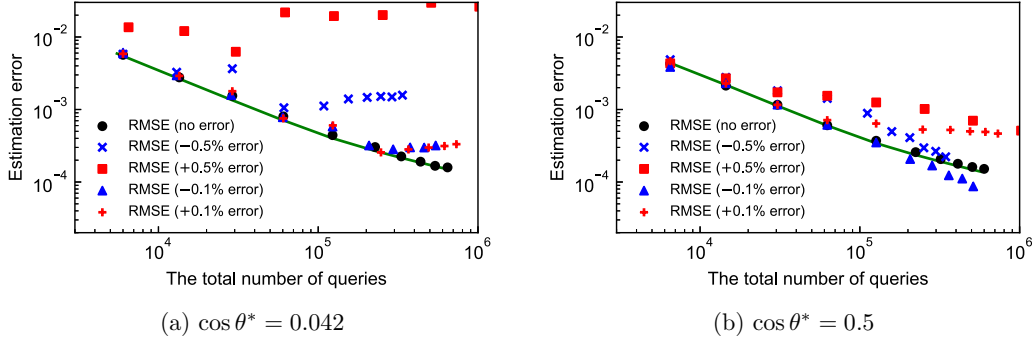


Figure 5.9: The impact of calibration error of p_q on the estimation performance. The plots indicate the root mean squared error (RMSE) of the method, evaluated with 200 trials, using the estimated noise parameter p_q including $\pm 0.1\%$ or $\pm 0.5\%$ calibration errors. The true value of p_q is taken as 0.995. The green solid line represents the (asymptotic) QCR bound under the perfect calibration. We note that the RMSE is plotted as a function of the maximum value of N_q in the trials because of the randomness of N_q . Reproduced from K. Wada *et al.*, Quantum 8, 1463 (2024), DOI: 10.22331/q-2024-09-09-1463. Licensed under CC BY 4.0.

classical Fisher information toward achieving the quantum Fisher information. Importantly, thanks to the maximum likelihood formulation and the two types of measurements with different symmetry in the probability distribution, our estimator enjoys some provable statistical properties such as consistency and asymptotic normality. To show the effectiveness of the proposed estimator, we executed several numerical experiments. The asymptotic statistical properties are evaluated in terms of the root MSE for several target values; the result was that, in all cases, the root MSE saturates the asymptotic classical and quantum Cramér-Rao bound, with a modest number of measurements. In particular, we confirmed that the classical Fisher information almost saturates the ultimate quantum Fisher information in a large system dimension $d = 2^{20}$. In addition, we studied how the total Fisher information depends on the target value by evaluating the asymptotic classical/quantum Cramér-Rao lower bounds. Although the previous researches [167, 152] imply that the classical Fisher information significantly deteriorates for certain target values under depolarization noise, the proposed estimator retains a large Fisher information regardless of the target value, due to the adaptive optimization.

Finally, we list several open questions to be further investigated.

- *Developing methods to cast unknown quantum noises to a global depolarization noise and to precisely calibrate it.* We assume a global depolarization noise and crucially use its mathematical structure, but practical noise itself is not in the form without any modification. In addition, the numerical experiments indicate that our algorithm needs a precise calibration of the noise parameter.
- *Rapid convergence to the precision limit.* Although our algorithm asymptotically achieves the precision limit, the number of queries required for this seems relatively large. How can we improve this? For instance, how about using appropriate query allocation?

Chapter 6

Conclusion

In this thesis, we investigate how efficiently we can probe the physical properties of quantum systems using quantum computers, in light of fundamental limits derived from the law of quantum physics. Specifically, we address an estimation task for expectation values of general observables on a target quantum state prepared by an unknown (black-box) unitary, which is one of the fundamental tasks in probing unknown quantum systems of interest with quantum computers. We mainly presented two estimation algorithms for this task that show quantum-enhanced scalings in the required query complexity to a state-preparation unitary and its inverse. We compared these query scalings with fundamental limits in relevant situations, and furthermore, specified other important quantum resources (i.e., elementary gate and ancilla qubit counts) required in those algorithms beyond the total queries. In our assumption, the state-preparation unitaries and their inverses can be queried at any time, but they are most costly to implement.

The first adaptive quantum algorithm is for multiple observables in an ideal (noiseless) setting. Given common access to block-encoded observables, a state-preparation unitary, and its inverse, this algorithm achieves the mean-squared-error (MSE) scaling at the Heisenberg limit together with the sublinear scaling in the number of observables, in terms of query complexity to the state preparation and its inverse. It also essentially matches the query lower bound of our task in the small-additive-error regime (not in MSE), previously derived in quantum complexity theory. In addition, the proposed algorithm significantly reduces the number of ancilla qubits and the classical computation for circuit synthesis due to its adaptive structure, compared to prior art for this task.

The second adaptive quantum algorithm is for a single observable in the early FTQC regime. In this regime, available coherent resources are mostly consumed for a state-preparation unitary, and the dominant noise stems from that costly unitary and its inverse. Under a specific model reflecting these noisy aspects, we derive the quantum-enhanced precision limit, that is, the minimal quantum Cramér-Rao bound over all possible estimation strategies in the model. Then, by designing an adaptive optimization based on Fisher information, we propose an implementable, quantum-enhanced estimation algorithm, with rigorous estimation guarantees, that asymptotically achieves the quantum Cramér-Rao bound, when the target system size is sufficiently large. This means that quantum circuits in the proposed algorithm are adjusted asymptotically so that they demonstrate the maximum quantum enhancement under the noisy settings. The asymptotic behavior is numerically verified with only a few hundred measurements per step and a few tens of qubits.

Access to the inverse unitary of the state preparation, such as the time-backward evolution e^{+iHt} , is crucial in the quantum algorithms developed in this thesis. From the perspective of quantum computers as fully computational probes, this assumption is natural. If the original state-preparation unitary is given as a known sequence of elementary local unitaries, the inverse can be implemented straightforwardly; we remark that this situation does not imply we know a classical description of the resulting entire unitary in general. In contrast, in quantum experimental scenarios where the target (essentially) closed quantum system is completely unknown, the inverse unitary cannot be efficiently accessible in general. Nevertheless, certain physically motivated systems with partial prior knowledge, such as symmetry, may effectively admit inverse unitary access. For instance, if a target system Hamiltonian H has chiral symmetry, meaning that there exists a known unitary S such that the anticommutator relation $\{H, S\} = 0$ (equivalently, $SHS^\dagger = -H$) holds, then we can apply the time-backward evolution e^{+iHt} by interleaving the symmetry operation S with the forward time evolution.

For such quantum systems, our estimation algorithm (particularly for the first one) provides one of the most query-efficient strategies among all protocols for our task with general multiple observables in ideal coherent quantum experiments. Finding nontrivial physical quantum systems that admit reasonable access to inverse unitaries is an important open problem for realizing our quantum-enhanced experiments; we leave this direction for future research.

We have focused on physical property estimation with quantum computers as a fundamental approach to potentially discovering unknown physics, and developed a constructive understanding of its ultimate limits within the framework of quantum algorithms. From an algorithmic complexity perspective, we sharpen the fundamental limit within reasonable and well-defined access models for our task. Moreover, by explicitly constructing algorithms that approach the corresponding fundamental limits, we expand the region that is actually attainable within those access models, and thus, the reach of quantum computers with finite resources. Also, our algorithms can take arbitrary state-preparation unitaries and their inverses as quantum input, thereby providing a highly general and efficient strategy for estimating general physical properties, which can be naturally combined with a wide range of quantum routines that coherently prepare target states. Together, the results in this thesis clarify and extend the reach of programmable quantum computers as computational and experimental probes for exploring previously inaccessible regions of nature.

Chapter 7

Appendix

7.1 Theorems for sequential quantum optimizers

As derived in Sec. 3.2.2, each sequential quantum optimizer has a $p \times p$ real-symmetric matrix. Here, $p-1$ is the degree of freedom of the target single-qubit gate (-1 means the constraint of normalization). Note that FQS gate or general single-qubit gate $R(\mathbf{q})$ ($p = 4$), Fraxis gate $R_{\mathbf{n}}(\pi)$ ($p = 3$), and NFT gate with $\mathbf{m}$ axis $R_{\mathbf{m}}(\psi)$ ($p = 2$) are written as the following unified form

$$R^{(p)} := \sum_{\mu=0}^{p-1} q_{\mu}^{(p)} \zeta_{\mu}^{(p)}, \quad (7.1)$$

where $q_{\mu}^{(p)}$ is the μ th element of unit vector in $\mathbb{R}^p$. Here, $\zeta_{\mu}^{(p)}$ is an extension of the Pauli matrices such as $(\vec{\zeta})^{(3)} = (-iX, -iY, -iZ)$ for Fraxis gate and $(\vec{\zeta})^{(2)} = (I, -i\mathbf{m} \cdot \vec{\sigma})$ for NFT gate. Accordingly, each $p \times p$ real-symmetric matrix is also written as

$$S_{\mu\nu}^{(p)} := \frac{1}{2} \text{tr} \left[\rho'_{\text{in}} \left(\left(\zeta_{\mu}^{(p)} \right)^{\dagger} H' \zeta_{\nu}^{(p)} + \left(\zeta_{\nu}^{(p)} \right)^{\dagger} H' \zeta_{\mu}^{(p)} \right) \right]. \quad (7.2)$$

In the below, we omit the superscript p for simplicity. Moreover, it is convenient to write the elements of S as

$$\begin{aligned} S_{\mu\nu} &= \frac{1}{2} \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \text{tr} \left[H U_2 \zeta_{\alpha} U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta}^{\dagger} U_2^{\dagger} \right], \quad \mu, \nu = 0, 1, \dots, p-1, \\ &= \frac{1}{2} \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \text{tr} \left[H^{(2)} \zeta_{\alpha} \rho_{\text{in}}^{(1)} \zeta_{\beta}^{\dagger} \right], \quad H^{(2)} := U_2^{\dagger} H U_2, \quad \rho_{\text{in}}^{(1)} := U_1 \rho_{\text{in}} U_1^{\dagger}. \end{aligned} \quad (7.3)$$

7.1.1 Proof of Theorem 7

In the following, d denotes the dimension of n -qubit system i.e., $d = 2^n$. Here, we provide a proof of Theorem 7. For convenience, we recall it.

Theorem. Suppose that the quantum circuits U_1 and U_2 are randomly and independently generated. If either U_1 or U_2 forms a unitary t -design with $t \geq 2$, the second moment of the spectral radius, r , for the centered matrix $S_c = S - \text{tr}[S] I_{p \times p} / p$ ($I_{p \times p}$ denotes the $p \times p$ identity) is upper bounded as follows.

$$\begin{aligned} &\mathbb{E}_{U_1, U_2} \left[r(S_c)^2 \right] \\ &\leq \begin{cases} \frac{p^2 \text{tr}[H^2] \Delta \rho_{\text{in},2}}{2(d^2-1)} + \frac{\Delta \rho_{\text{in},2}}{4p(d^2-1)} \sum_{\mu,\nu=0}^{p-1} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \zeta_{\mu} \zeta_{\nu}^{\dagger} H^{(2)} \zeta_{\nu} \zeta_{\mu}^{\dagger} + \text{h.c.} \right] \right] \\ \frac{p^2 \text{tr}[\rho_{\text{in}}^2] \Delta H_2}{2(d^2-1)} + \frac{\Delta H_2}{4p(d^2-1)} \sum_{\mu,\nu=0}^{p-1} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \zeta_{\mu}^{\dagger} \zeta_{\nu} \rho_{\text{in}}^{(1)} \zeta_{\nu}^{\dagger} \zeta_{\mu} + \text{h.c.} \right] \right] \end{cases}, \end{aligned} \quad (7.4)$$

where the first case corresponds to U_1 being a t -design, and the second case corresponds to U_2 being a t -design. Here, $\mathbb{E}_{U_1, U_2}[\cdot]$ is defined as the expectation over the random quantum circuits U_1 and U_2 , $\delta_{\mu\nu}$ denotes the Kronecker delta, h.c. means the Hermite conjugate of the preceding term, and $\Delta\rho_{\text{in},2}, \Delta H_2$ are defined as

$$\Delta\rho_{\text{in},2} := \text{tr}[\rho_{\text{in}}^2] - \frac{1}{d}, \quad \Delta H_2 := \text{tr}[H^2] - \frac{\text{tr}^2[H]}{d}. \quad (7.5)$$

Proof. Since the centered matrix S_c is a real-symmetric matrix regardless of quantum circuits U_1 and U_2 , the following inequality holds, which comes from the maximum is at most the square root of the sum of the squares.

$$r\left(S - \frac{\text{tr}[S]}{p}I_{p \times p}\right) \leq \left\|S - \frac{\text{tr}[S]}{p}I_{p \times p}\right\|_{\text{F}}, \quad (7.6)$$

where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm. Thus, the second moment of the spectral radius is evaluated as

$$\begin{aligned} \mathbb{E}_{U_1, U_2} \left[r\left(S - \frac{\text{tr}[S]}{p}I_{p \times p}\right)^2 \right] &\leq \mathbb{E}_{U_1, U_2} \left[\left\|S - \frac{\text{tr}[S]}{p}I_{p \times p}\right\|_{\text{F}}^2 \right] \\ &= \sum_{\mu\nu} \mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \sum_{\mu\nu} \mathbb{E}_{U_1, U_2} [S_{\mu\mu}S_{\nu\nu}]. \end{aligned} \quad (7.7)$$

First, we evaluate the r.h.s of Eq. (7.7) on the condition that the random quantum circuit U_1 forms a unitary t -design with $t \geq 2$. Using lemma 16, which is proved later, the following identity holds,

$$\begin{aligned} \mathbb{E}_{U_1, U_2} [S_{\mu\nu}S_{\mu'\nu'}] &= \frac{\text{tr}^2[H]}{d^2 - 1} \Delta\rho_{\text{in},1} \delta_{\mu\nu} \delta_{\mu'\nu'} + \frac{1}{4(d^2 - 1)} \Delta\rho_{\text{in},2} \sum_{(\alpha, \beta)} \sum_{(\alpha', \beta')} \mathbb{E}_{U_2} \left[\text{tr} \left[\zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \right] \right]. \end{aligned} \quad (7.8)$$

where we defined

$$\Delta\rho_{\text{in},1} := 1 - \frac{\text{tr}[\rho_{\text{in}}^2]}{d}. \quad (7.9)$$

Hence, we can calculate the expectations in the final line of Eq. (7.7) as follows.

$$\begin{aligned} \mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] &= \frac{\text{tr}^2[H]}{d^2 - 1} \Delta\rho_{\text{in},1} \delta_{\mu\nu} \\ &\quad + \frac{1}{4(d^2 - 1)} \Delta\rho_{\text{in},2} \sum_{(\alpha, \beta)=(\mu, \nu), (\nu, \mu)} \sum_{(\alpha', \beta')=(\mu, \nu), (\nu, \mu)} \mathbb{E}_{U_2} \left[\text{tr} \left[H^{(2)} \zeta_{\alpha} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \zeta_{\beta}^{\dagger} \right] \right] \\ &= \frac{\text{tr}^2[H]}{d^2 - 1} \Delta\rho_{\text{in},1} \delta_{\mu\nu} + \frac{\text{tr}[H^2]}{2(d^2 - 1)} \Delta\rho_{\text{in},2} \\ &\quad + \frac{1}{4(d^2 - 1)} \Delta\rho_{\text{in},2} \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \zeta_{\mu} \zeta_{\nu}^{\dagger} H^{(2)} \zeta_{\nu} \zeta_{\mu}^{\dagger} + \text{h.c.} \right] \right], \end{aligned} \quad (7.10)$$

$$\begin{aligned} \mathbb{E}_{U_1, U_2} [S_{\mu\mu}S_{\nu\nu}] &= \frac{\text{tr}^2[H]}{d^2 - 1} \Delta\rho_{\text{in},1} \\ &\quad + \frac{1}{4(d^2 - 1)} \Delta\rho_{\text{in},2} \sum_{(\alpha, \beta)=(\mu, \mu), (\mu, \mu)} \sum_{(\alpha', \beta')=(\nu, \nu), (\nu, \nu)} \mathbb{E}_{U_2} \left[\text{tr} \left[H^{(2)} \zeta_{\alpha} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \zeta_{\beta}^{\dagger} \right] \right] \\ &= \frac{\text{tr}^2[H]}{d^2 - 1} \Delta\rho_{\text{in},1} + \frac{1}{d^2 - 1} \Delta\rho_{\text{in},2} \mathbb{E}_{U_2} \left[\text{tr} \left[H^{(2)} \zeta_{\mu} \zeta_{\nu}^{\dagger} H^{(2)} \zeta_{\nu} \zeta_{\mu}^{\dagger} \right] \right] \\ &= \frac{\text{tr}^2[H]}{d^2 - 1} \Delta\rho_{\text{in},1} + \frac{1}{2(d^2 - 1)} \Delta\rho_{\text{in},2} \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \zeta_{\mu} \zeta_{\nu}^{\dagger} H^{(2)} \zeta_{\nu} \zeta_{\mu}^{\dagger} + \text{h.c.} \right] \right], \end{aligned} \quad (7.11)$$

Accordingly, we can evaluate the second moment of the spectral radius as

$$\begin{aligned}
& \mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \\
&= \frac{\text{tr}^2 [H]}{p(d^2 - 1)} \Delta \rho_{\text{in}, 1} (p\delta_{\mu\nu} - 1) + \frac{\text{tr} [H^2]}{2(d^2 - 1)} \Delta \rho_{\text{in}, 2} \\
&+ \frac{\Delta \rho_{\text{in}, 2}}{4p(d^2 - 1)} \left(p \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \varsigma_\mu \varsigma_\nu^\dagger H^{(2)} \varsigma_\mu \varsigma_\nu^\dagger + \text{h.c.} \right] \right] - 2 \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \varsigma_\mu \varsigma_\nu^\dagger H^{(2)} \varsigma_\nu \varsigma_\mu^\dagger + \text{h.c.} \right] \right] \right) \\
&= \frac{\text{tr}^2 [H]}{p(d^2 - 1)} \Delta \rho_{\text{in}, 1} (p\delta_{\mu\nu} - 1) + \frac{\text{tr} [H^2]}{2(d^2 - 1)} \Delta \rho_{\text{in}, 2} \\
&+ \frac{\Delta \rho_{\text{in}, 2}}{4p(d^2 - 1)} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \varsigma_\mu \varsigma_\nu^\dagger H^{(2)} \varsigma_\nu \varsigma_\mu^\dagger + \text{h.c.} \right] \right], \tag{7.12}
\end{aligned}$$

where we used $\varsigma_\mu \varsigma_\nu^\dagger = (-1)^{1-\delta_{\mu\nu}} \varsigma_\nu \varsigma_\mu^\dagger$ in the last equality, and

$$\begin{aligned}
& \mathbb{E}_{U_1, U_2} \left[r \left(S - \frac{\text{tr} [S]}{p} I_{p \times p} \right)^2 \right] \leq \sum_{\mu\nu} \left(\mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \right) \\
&= \frac{p^2 \text{tr} [H^2]}{2(d^2 - 1)} \Delta \rho_{\text{in}, 2} + \frac{\text{tr}^2 [H]}{p(d^2 - 1)} \sum_{\mu\nu} (p\delta_{\mu\nu} - 1) \\
&+ \frac{\Delta \rho_{\text{in}, 2}}{4p(d^2 - 1)} \sum_{\mu\nu} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \varsigma_\mu \varsigma_\nu^\dagger H^{(2)} \varsigma_\nu \varsigma_\mu^\dagger + \text{h.c.} \right] \right] \\
&= \frac{p^2 \text{tr} [H^2]}{2(d^2 - 1)} \Delta \rho_{\text{in}, 2} + \frac{\Delta \rho_{\text{in}, 2}}{4p(d^2 - 1)} \sum_{\mu\nu} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_2} \left[H^{(2)} \varsigma_\mu \varsigma_\nu^\dagger H^{(2)} \varsigma_\nu \varsigma_\mu^\dagger + \text{h.c.} \right] \right]. \tag{7.13}
\end{aligned}$$

Next, we evaluate the r.h.s of Eq. (7.7) on the condition that the random quantum circuit U_2 forms a unitary t -design with $t \geq 2$. Using lemma 17, which is also proved later, we calculate the expectations in the final line of Eq. (7.7) as follows.

$$\begin{aligned}
\mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] &= \frac{\text{tr}^2 [\rho_{\text{in}}]}{d^2 - 1} \Delta H_1 \delta_{\mu\nu} \\
&+ \frac{1}{4(d^2 - 1)} \Delta H_2 \sum_{(\alpha, \beta) = (\mu, \nu), (\nu, \mu)} \sum_{(\alpha', \beta') = (\mu, \nu), (\nu, \mu)} \mathbb{E}_{U_1} \left[\text{tr} \left[\rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger \varsigma_\alpha \right] \right] \\
&= \frac{1}{d^2 - 1} \Delta H_1 \delta_{\mu\nu} + \frac{\text{tr} [\rho_{\text{in}}^2]}{2(d^2 - 1)} \Delta H_2 \\
&+ \frac{1}{4(d^2 - 1)} \Delta H_2 \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu + \text{h.c.} \right] \right], \tag{7.14}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] &= \frac{\text{tr}^2 [\rho_{\text{in}}]}{d^2 - 1} \Delta H_1 \\
&+ \frac{1}{4(d^2 - 1)} \Delta H_2 \sum_{(\alpha, \beta) = (\mu, \mu), (\nu, \nu)} \sum_{(\alpha', \beta') = (\nu, \nu), (\mu, \mu)} \mathbb{E}_{U_1} \left[\text{tr} \left[\rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger \varsigma_\alpha \right] \right] \\
&= \frac{1}{d^2 - 1} \Delta H_1 + \frac{1}{d^2 - 1} \Delta H_2 \mathbb{E}_{U_1} \left[\text{tr} \left[\rho_{\text{in}}^{(1)} \varsigma_\mu^\dagger \varsigma_\nu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu \right] \right] \\
&= \frac{1}{d^2 - 1} \Delta H_1 + \frac{1}{2(d^2 - 1)} \Delta H_2 \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \varsigma_\mu^\dagger \varsigma_\nu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu + \text{h.c.} \right] \right], \tag{7.15}
\end{aligned}$$

where we defined

$$\Delta H_1 := \text{tr}^2 [H] - \frac{\text{tr} [H^2]}{d}. \tag{7.16}$$

Accordingly, we can evaluate the second moment of the spectral radius as

$$\begin{aligned}
& \mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \\
&= \frac{1}{p(d^2 - 1)} \Delta H_1 (p\delta_{\mu\nu} - 1) + \frac{\text{tr} [\rho_{\text{in}}^2]}{2(d^2 - 1)} \Delta H_2 \\
&+ \frac{\Delta H_2}{4p(d^2 - 1)} \left(p \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu + \text{h.c.} \right] \right] - 2 \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \varsigma_\mu^\dagger \varsigma_\nu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu + \text{h.c.} \right] \right] \right) \\
&= \frac{1}{p(d^2 - 1)} \Delta H_1 (p\delta_{\mu\nu} - 1) + \frac{\text{tr} [\rho_{\text{in}}^2]}{2(d^2 - 1)} \Delta H_2 \\
&+ \frac{\Delta H_2}{4p(d^2 - 1)} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \varsigma_\mu^\dagger \varsigma_\nu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu + \text{h.c.} \right] \right], \tag{7.17}
\end{aligned}$$

where we used $\varsigma_\nu^\dagger \varsigma_\mu = (-1)^{1-\delta_{\mu\nu}} \varsigma_\mu^\dagger \varsigma_\nu$ in the last equality, and

$$\begin{aligned}
& \mathbb{E}_{U_1, U_2} \left[r \left(S - \frac{\text{tr} [S]}{p} I_{p \times p} \right)^2 \right] \leq \sum_{\mu\nu} \left(\mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \right) \\
&= \frac{p^2 \text{tr} [\rho_{\text{in}}^2] \Delta H_2}{2(d^2 - 1)} + \frac{\Delta H_2}{4p(d^2 - 1)} \sum_{\mu\nu} (p(-1)^{1-\delta_{\mu\nu}} - 2) \text{tr} \left[\mathbb{E}_{U_1} \left[\rho_{\text{in}}^{(1)} \varsigma_\mu^\dagger \varsigma_\nu \rho_{\text{in}}^{(1)} \varsigma_\nu^\dagger \varsigma_\mu + \text{h.c.} \right] \right]. \tag{7.18}
\end{aligned}$$

□

7.1.2 Proof of Theorem 8

Here, we provide a proof of Theorem 8. For convenience, we recall it.

Theorem. Suppose that the whole quantum circuit U is an n -qubit alternating layered ansatz with m -qubit blocks as described in Fig. 3.5. Here, we focus on a block W in l th layer and a parameterized single-qubit gate R in the block. We assume that the quantum circuits $W_A, W_B \subset W$, which are located after and before the target gate respectively, and the other blocks form a local 2-design independently. In addition, the Hamiltonian H is assumed to be m -local such as

$$H = c_0 I^{\otimes n} + \sum_i c_i h_i, \quad c_0, c_i \in \mathbb{R}, \quad (\exists i, c_i \neq 0), \tag{7.19}$$

where a tensor product of Pauli matrices h_i acts non-trivially on at most m -qubit. (This is the same assumption in [28].) Then, the second moment of the spectral radius is lower bounded as

$$\mathbb{E}_{U_1, U_2} \left[r \left(S - \frac{\text{tr} [S]}{p} I_{p \times p} \right)^2 \right] \geq \frac{(p+2)(p-1)2^{m(l+1)-1}}{p(2^{2m}-1)^2(2^m+1)^{L+l}} \sum_{i \in i_{\mathcal{L}}} \sum_{\substack{(k, k') \in k_{\mathcal{L}_B} \\ k' \geq k}} c_i^2 \epsilon(\rho_{k, k'}) \epsilon(h_i), \tag{7.20}$$

where L denotes the total number of layers. Here, $i_{\mathcal{L}}$ is the set of i indices whose associated operators h_i act on qubits in the forward light-cone $\mathcal{L}$ of W , and $k_{\mathcal{L}_B}$ is the set of k indices whose associated subsystems S_k are in the backward light-cone $\mathcal{L}_B$ of W . The quantum state $\rho_{k, k'}$ is the reduced density matrix of the input state ρ_{in} on $S_k S_{k+1} \cdots S_{k'}$, and the function $\epsilon(M)$ for a matrix M is defined as $\epsilon(M) = D_{\text{HS}}(M, \text{tr}(M) \mathbf{1}/d_M)$ where D_{HS} is the Hilbert-Schmidt distance and d_M is the dimension of the matrix M .

Proof. To establish the lower bound of the second moment, we begin with the following inequality, which comes from the maximum is at least the square root of the average of the sum of the squares.

$$r \left(S - \frac{\text{tr} [S]}{p} I_{p \times p} \right) \geq \frac{1}{\sqrt{p}} \left\| S - \frac{\text{tr} [S]}{p} I_{p \times p} \right\|_{\text{F}}, \tag{7.21}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Since the inequality holds for any quantum circuits U_1 and U_2 in Fig. 3.4,

$$\begin{aligned} \mathbb{E}_{U_1, U_2} \left[r \left(S - \frac{\text{tr}[S]}{p} I_{p \times p} \right)^2 \right] &\geq \frac{1}{p} \mathbb{E}_{U_1, U_2} \left[\left\| S - \frac{\text{tr}[S]}{p} I_{p \times p} \right\|_F^2 \right] \\ &= \frac{1}{p} \sum_{\mu\nu} \left(\mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \right). \end{aligned} \quad (7.22)$$

We first evaluate the expectation of $S_{\mu\nu} S_{\mu'\nu'}$ over the block of interest. From the setting, the block W containing a single-qubit gate R is decomposed as

$$W = W_A (\mathbf{1}_{m-1} \otimes R) W_B, \quad (7.23)$$

where $\mathbf{1}_{m-1}$ is the identity on $m-1$ qubits system, and $W_A, W_B \subset W$ are the quantum circuits after and before R , respectively. As shown in Fig. 3.4, the quantum circuits U_1 and U_2 for an alternating layered ansatz can be written as

$$U_1 = (\mathbf{1}_{\bar{w}} \otimes W_B) V_1, \quad U_2 = V_2 (\mathbf{1}_{\bar{w}} \otimes W_A), \quad (7.24)$$

where $\mathbf{1}_{\bar{w}}$ denotes the identity over the qubits on which the block W acts trivially. Here, V_2 contains the gates in the forward light-cone $\mathcal{L}$ of W , i.e., all gates with at least one input qubit causally connected to the qubits of W as in Fig. 3.5, and V_1 contains other gates. Accordingly, we can write the elements of S matrix as

$$S_{\mu\nu} = \frac{1}{2} \sum_{(\alpha, \beta) = (\mu, \nu), (\nu, \mu)} \text{tr} \left[H V_2 (\mathbf{1}_{\bar{w}} \otimes W_A (\mathbf{1}_{m-1} \otimes \varsigma_\alpha) W_B) V_1 \rho_{\text{in}} V_1^\dagger \left(\mathbf{1}_{\bar{w}} \otimes W_B^\dagger (\mathbf{1}_{m-1} \otimes \varsigma_\beta^\dagger) W_A^\dagger \right) V_2^\dagger \right], \quad (7.25)$$

and we obtain

$$\begin{aligned} S_{\mu\nu} S_{\mu'\nu'} &= \frac{1}{4} \sum_{\substack{(\alpha, \beta) \\ (\alpha', \beta')}} \text{tr} \left[H V_2 (\mathbf{1}_{\bar{w}} \otimes W_A (\mathbf{1}_{m-1} \otimes \varsigma_\alpha) W_B) V_1 \rho_{\text{in}} V_1^\dagger \left(\mathbf{1}_{\bar{w}} \otimes W_B^\dagger (\mathbf{1}_{m-1} \otimes \varsigma_\beta^\dagger) W_A^\dagger \right) V_2^\dagger \right] \\ &\quad \times \text{tr} \left[H V_2 (\mathbf{1}_{\bar{w}} \otimes W_A (\mathbf{1}_{m-1} \otimes \varsigma_{\alpha'}) W_B) V_1 \rho_{\text{in}} V_1^\dagger \left(\mathbf{1}_{\bar{w}} \otimes W_B^\dagger (\mathbf{1}_{m-1} \otimes \varsigma_{\beta'}^\dagger) W_A^\dagger \right) V_2^\dagger \right] \\ &= \frac{1}{4} \sum_{\substack{(\alpha, \beta) \\ (\alpha', \beta')}} \sum_{\substack{\mathbf{i}, \mathbf{j} \\ \mathbf{i}', \mathbf{j}'}} \text{tr} \left[W_A (\mathbf{1}_{m-1} \otimes \varsigma_\alpha) W_B \rho_{\text{in}, \mathbf{i}\mathbf{j}}^{(1)} W_B^\dagger (\mathbf{1}_{m-1} \otimes \varsigma_\beta^\dagger) W_A^\dagger H_{\mathbf{j}\mathbf{i}}^{(2)} \right] \\ &\quad \times \text{tr} \left[W_A (\mathbf{1}_{m-1} \otimes \varsigma_{\alpha'}) W_B \rho_{\text{in}, \mathbf{i}'\mathbf{j}'}^{(1)} W_B^\dagger (\mathbf{1}_{m-1} \otimes \varsigma_{\beta'}^\dagger) W_A^\dagger H_{\mathbf{j}'\mathbf{i}'}^{(2)} \right]. \end{aligned} \quad (7.26)$$

Here, we defined

$$\rho_{\text{in}, \mathbf{i}\mathbf{j}}^{(1)} := \text{tr}_{\bar{w}} \left[(|\mathbf{j}\rangle \langle \mathbf{i}| \otimes \mathbf{1}_w) V_1 \rho_{\text{in}} V_1^\dagger \right], \quad H_{\mathbf{j}\mathbf{i}}^{(2)} := \text{tr}_{\bar{w}} \left[(|\mathbf{i}\rangle \langle \mathbf{j}| \otimes \mathbf{1}_w) V_2^\dagger H V_2 \right], \quad (7.27)$$

where $\text{tr}_{\bar{w}}[\cdot]$ means the partial trace over the qubits which are not in W , and $\{|\mathbf{i}\rangle\}$ denotes the computational basis on $(n-m)$ -qubit system.

Since W_A forms a local 2-design, we first compute the expectation of Eq. (7.26) over W_A as

$$\begin{aligned}
& \mathbb{E}_{W_A} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{1}{4(2^{2m}-1)} \sum_{\substack{(\alpha,\beta) \\ (\alpha',\beta')}} \sum_{\substack{i,j \\ i',j'}} \left(\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}]}{2^m} \right) \\
&\quad \times \text{tr} \left[(\mathbf{1}_{m-1} \otimes \varsigma_{\beta}^{\dagger} \varsigma_{\alpha}) W_B \rho_{\text{in},ij}^{(1)} W_B^{\dagger} \right] \text{tr} \left[(\mathbf{1}_{m-1} \otimes \varsigma_{\beta'}^{\dagger} \varsigma_{\alpha'}) W_B \rho_{\text{in},i'j'}^{(1)} W_B^{\dagger} \right] \\
&+ \frac{1}{4(2^{2m}-1)} \sum_{\substack{(\alpha,\beta) \\ (\alpha',\beta')}} \sum_{\substack{i,j \\ i',j'}} \left(\text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}]}{2^m} \right) \\
&\quad \times \text{tr} \left[(\mathbf{1}_{m-1} \otimes \varsigma_{\beta}^{\dagger} \varsigma_{\alpha}) W_B \rho_{\text{in},ij}^{(1)} W_B^{\dagger} (\mathbf{1}_{m-1} \otimes \varsigma_{\beta'}^{\dagger} \varsigma_{\alpha'}) W_B \rho_{\text{in},i'j'}^{(1)} W_B^{\dagger} \right]. \quad (7.28)
\end{aligned}$$

Here, we employed the following integration formula from the Weingarten calculus [28]

$$\begin{aligned}
\mathbb{E}_W \text{tr} [W A W^{\dagger} B W C W^{\dagger} D] &= \frac{1}{2^{2m}-1} (\text{tr} [A] \text{tr} [C] \text{tr} [BD] + \text{tr} [AC] \text{tr} [B] \text{tr} [D]) \\
&\quad - \frac{1}{2^m(2^{2m}-1)} (\text{tr} [AC] \text{tr} [BD] + \text{tr} [A] \text{tr} [B] \text{tr} [C] \text{tr} [D]), \quad (7.29)
\end{aligned}$$

where W is Haar-distributed on the unitary group of degree 2^m , and A, B, C and D are arbitrary linear operators on an m -qubit system. Noting that

$$\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \text{tr} \left[(\mathbf{1}_{m-1} \otimes \varsigma_{\beta}^{\dagger} \varsigma_{\alpha}) W_B \rho_{\text{in},ij}^{(1)} W_B^{\dagger} \right] = 2\delta_{\mu\nu} \text{tr} \left[\rho_{\text{in},ij}^{(1)} \right] \quad (7.30)$$

holds, then the first term of Eq. (7.28) is independent of W_B . Since W_B also forms a local 2-design, a part of the second term of Eq. (7.28) is evaluated as

$$\begin{aligned}
& \sum_{\substack{(\alpha,\beta) \\ (\alpha',\beta')}} \mathbb{E}_{W_B} \text{tr} \left[(\mathbf{1}_{m-1} \otimes \varsigma_{\beta}^{\dagger} \varsigma_{\alpha}) W_B \rho_{\text{in},ij}^{(1)} W_B^{\dagger} (\mathbf{1}_{m-1} \otimes \varsigma_{\beta'}^{\dagger} \varsigma_{\alpha'}) W_B \rho_{\text{in},i'j'}^{(1)} W_B^{\dagger} \right] \\
&= \frac{1}{2^{2m}-1} \sum_{\substack{(\alpha,\beta) \\ (\alpha',\beta')}} \text{tr} \left[\mathbf{1}_{m-1} \otimes \varsigma_{\beta}^{\dagger} \varsigma_{\alpha'} \varsigma_{\beta'}^{\dagger} \varsigma_{\alpha} \right] \left(\text{tr} \left[\rho_{\text{in},ij}^{(1)} \right] \text{tr} \left[\rho_{\text{in},i'j'}^{(1)} \right] - \frac{\text{tr} \left[\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)} \right]}{2^m} \right) \\
&+ \frac{1}{2^{2m}-1} \sum_{\substack{(\alpha,\beta) \\ (\alpha',\beta')}} \text{tr} \left[\mathbf{1}_{m-1} \otimes \varsigma_{\beta}^{\dagger} \varsigma_{\alpha'} \right] \text{tr} \left[\mathbf{1}_{m-1} \otimes \varsigma_{\beta'}^{\dagger} \varsigma_{\alpha} \right] \left(\text{tr} \left[\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)} \right] - \frac{\text{tr} \left[\rho_{\text{in},ij}^{(1)} \right] \text{tr} \left[\rho_{\text{in},i'j'}^{(1)} \right]}{2^m} \right) \\
&= \frac{4 \cdot 2^m}{2^{2m}-1} \delta_{\mu\nu} \delta_{\mu'\nu'} \left(\text{tr} \left[\rho_{\text{in},ij}^{(1)} \right] \text{tr} \left[\rho_{\text{in},i'j'}^{(1)} \right] - \frac{\text{tr} \left[\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)} \right]}{2^m} \right) \\
&+ \frac{8 \cdot 2^{2(m-1)}}{2^{2m}-1} (\delta_{\mu'\mu} \delta_{\nu'\nu} + \delta_{\mu\nu} \delta_{\mu'\nu'}) \left(\text{tr} \left[\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)} \right] - \frac{\text{tr} \left[\rho_{\text{in},ij}^{(1)} \right] \text{tr} \left[\rho_{\text{in},i'j'}^{(1)} \right]}{2^m} \right), \quad (7.31)
\end{aligned}$$

where we used the formula Eq. (7.29) in the first equality. In the second equality, we used the following relation, which can be derived from direct calculation, as

$$\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')} \text{tr} \left[\varsigma_{\beta}^{\dagger} \varsigma_{\alpha'} \right] \text{tr} \left[\varsigma_{\beta'}^{\dagger} \varsigma_{\alpha} \right] = 8(\delta_{\mu'\mu} \delta_{\nu'\nu} + \delta_{\mu\nu} \delta_{\mu'\nu'}). \quad (7.32)$$

Taking the expectation over W_B and substituting Eqs. (7.30), (7.31) back into Eq. (7.28), we obtain

$$\begin{aligned}
& \mathbb{E}_{W_A, W_B} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{\delta_{\mu\nu} \delta_{\mu'\nu'}}{2^{2m} - 1} \sum_{\substack{i,j \\ i',j'}} \text{tr} [\rho_{\text{in},ij}^{(1)}] \text{tr} [\rho_{\text{in},i'j'}^{(1)}] \left(\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}]}{2^m} \right) \\
&+ \frac{2^m \delta_{\mu\nu} \delta_{\mu'\nu'}}{(2^{2m} - 1)^2} \sum_{\substack{i,j \\ i',j'}} \left(\text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}]}{2^m} \right) \left(\text{tr} [\rho_{\text{in},ij}^{(1)}] \text{tr} [\rho_{\text{in},i'j'}^{(1)}] - \frac{\text{tr} [\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)}]}{2^m} \right) \\
&+ \frac{2 \cdot 2^{2m}}{4(2^{2m} - 1)^2} (\delta_{\mu'\mu} \delta_{\nu'\nu} + \delta_{\mu\nu} \delta_{\mu'\nu'}) \\
&\quad \times \sum_{\substack{i,j \\ i',j'}} \left(\text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}]}{2^m} \right) \left(\text{tr} [\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)}] - \frac{\text{tr} [\rho_{\text{in},ij}^{(1)}] \text{tr} [\rho_{\text{in},i'j'}^{(1)}]}{2^m} \right) \\
&= \delta_{\mu\nu} \delta_{\mu'\nu'} T^{(1,2)} + \frac{2 \cdot 2^{2m}}{4(2^{2m} - 1)^2} (\delta_{\mu'\mu} \delta_{\nu'\nu} + \delta_{\mu\nu} \delta_{\mu'\nu'}) \sum_{\substack{i,j \\ i',j'}} \Delta H_{ij}^{i'j'} \Delta \rho_{ij}^{i'j'}. \tag{7.33}
\end{aligned}$$

Here, we defined

$$\Delta H_{ij}^{i'j'} := \text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}]}{2^m}, \tag{7.34}$$

$$\Delta \rho_{ij}^{i'j'} := \text{tr} [\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)}] - \frac{\text{tr} [\rho_{\text{in},ij}^{(1)}] \text{tr} [\rho_{\text{in},i'j'}^{(1)}]}{2^m}, \tag{7.35}$$

$$\begin{aligned}
T^{(1,2)} &:= \frac{1}{2^{2m} - 1} \sum_{\substack{i,j \\ i',j'}} \text{tr} [\rho_{\text{in},ij}^{(1)}] \text{tr} [\rho_{\text{in},i'j'}^{(1)}] \left(\text{tr} [H_{ji}^{(2)}] \text{tr} [H_{j'i'}^{(2)}] - \frac{\text{tr} [H_{ji}^{(2)} H_{j'i'}^{(2)}]}{2^m} \right) \\
&+ \frac{2^m}{(2^{2m} - 1)^2} \sum_{\substack{i,j \\ i',j'}} \Delta H_{ij}^{i'j'} \left(\text{tr} [\rho_{\text{in},ij}^{(1)}] \text{tr} [\rho_{\text{in},i'j'}^{(1)}] - \frac{\text{tr} [\rho_{\text{in},ij}^{(1)} \rho_{\text{in},i'j'}^{(1)}]}{2^m} \right). \tag{7.36}
\end{aligned}$$

Before proceeding to evaluate the expectation over V_1 and V_2 , we calculate the summation in Eq. (7.22) as follows.

$$\sum_{\mu\nu} \mathbb{E}_{W_A, W_B} [S_{\mu\nu}^2] = p T^{(1,2)} + \frac{2 \cdot 2^{2m}}{4(2^{2m} - 1)^2} (p^2 + p) \sum_{\substack{i,j \\ i',j'}} \Delta H_{ij}^{i'j'} \Delta \rho_{ij}^{i'j'}, \tag{7.37}$$

$$\sum_{\mu\nu} \mathbb{E}_{W_A, W_B} [S_{\mu\mu} S_{\nu\nu}] = p^2 T^{(1,2)} + \frac{2^{2m}}{(2^{2m} - 1)^2} p \sum_{\substack{i,j \\ i',j'}} \Delta H_{ij}^{i'j'} \Delta \rho_{ij}^{i'j'}. \tag{7.38}$$

Accordingly,

$$\begin{aligned}
\sum_{\mu\nu} \left(\mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \right) &= \mathbb{E}_{V_1, V_2} \left[\sum_{\mu\nu} \mathbb{E}_{W_A, W_B} [S_{\mu\nu}^2] - \frac{1}{p} \sum_{\mu\nu} \mathbb{E}_{W_A, W_B} [S_{\mu\mu} S_{\nu\nu}] \right] \\
&= \frac{2 \cdot 2^{2m} (p+2)(p-1)}{4(2^{2m} - 1)^2} \sum_{\substack{i,j \\ i',j'}} \mathbb{E}_{V_1} [\Delta \rho_{ij}^{i'j'}] \mathbb{E}_{V_2} [\Delta H_{ij}^{i'j'}]. \tag{7.39}
\end{aligned}$$

Finally, we evaluate the expectations over V_1, V_2 in Eq. (7.39), which can be calculated basically with a series of integration for the m -qubit blocks in V_1, V_2 . In the same assumption of ours, the previous study [28] has showed the following inequality holds:

$$\sum_{\substack{i,j \\ i',j'}} \mathbb{E}_{V_1} [\Delta \rho_{ij}^{i'j'}] \mathbb{E}_{V_2} [\Delta H_{ij}^{i'j'}] \geq \frac{2^{m(l-1)}}{(2^m+1)^{L+l}} \sum_{i \in i_{\mathcal{L}}} \sum_{\substack{(k,k') \in k_{\mathcal{L}_{\mathbb{B}}} \\ k' \geq k}} c_i^2 \epsilon(\rho_{k,k'}) \epsilon(h_i), \quad (7.40)$$

where we recall that L is the total number of layers, and the block W of interest is in the l th layer. Combining this inequality with Eq. (7.39), we establish the lower bound for the second moment of spectral radius as

$$\begin{aligned} \mathbb{E}_{U_1, U_2} \left[r \left(S - \frac{\text{tr}[S]}{p} I_{p \times p} \right)^2 \right] &\geq \frac{1}{p} \sum_{\mu\nu} \left(\mathbb{E}_{U_1, U_2} [S_{\mu\nu}^2] - \frac{1}{p} \mathbb{E}_{U_1, U_2} [S_{\mu\mu} S_{\nu\nu}] \right) \\ &= \frac{1}{p} \frac{2 \cdot 2^{2m} (p+2)(p-1)}{4(2^{2m}-1)^2} \sum_{\substack{i,j \\ i',j'}} \mathbb{E}_{V_1} [\Delta \rho_{ij}^{i'j'}] \mathbb{E}_{V_2} [\Delta H_{ij}^{i'j'}] \\ &\geq \frac{(p+2)(p-1)2^{m(l+1)-1}}{p(2^{2m}-1)^2(2^m+1)^{L+l}} \sum_{i \in i_{\mathcal{L}}} \sum_{\substack{(k,k') \in k_{\mathcal{L}_{\mathbb{B}}} \\ k' \geq k}} c_i^2 \epsilon(\rho_{k,k'}) \epsilon(h_i). \end{aligned} \quad (7.41)$$

□

Proof of some useful lemmas

Hereafter, we provide some lemmas for the proof of Theorem 7.

Lemma 16. *Suppose that the random quantum circuits U_1 and U_2 are independent. If U_1 forms a unitary t -design with $t \geq 2$, the expectation of the product of two matrix elements for S can be evaluated as*

$$\mathbb{E}_{U_1, U_2} [S_{\mu\nu} S_{\mu'\nu'}] = \frac{\text{tr}^2[H]}{d^2-1} \Delta \rho_{\text{in},1} \delta_{\mu\nu} \delta_{\mu'\nu'} + \frac{1}{4(d^2-1)} \Delta \rho_{\text{in},2} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \mathbb{E}_{U_2} \left[\text{tr} \left[\zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \right] \right]. \quad (7.42)$$

Here, the summation runs over the set of μ, ν as follows

$$\sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} = \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')}. \quad (7.43)$$

Proof. Substituting Eq. (7.3) into the l.h.s of Eq. (7.42) and omitting the expectation over U_2 , we obtain

$$\begin{aligned} &\mathbb{E}_{U_1} [S_{\mu\nu} S_{\mu'\nu'}] \\ &= \frac{1}{4} \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')} \mathbb{E}_{U_1} \text{tr} \left[H^{(2)} \zeta_{\alpha} U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta}^{\dagger} \right] \text{tr} \left[H^{(2)} \zeta_{\alpha'} U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta'}^{\dagger} \right] \\ &= \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \sum_{i,j} \mathbb{E}_{U_1} \text{tr} \left[U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} |i\rangle \langle j| U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} |j\rangle \langle i| \right], \end{aligned} \quad (7.44)$$

where $\{|i\rangle\}$ denotes the computational basis on n -qubit system. If U_1 forms a unitary t -design with $t \geq 2$, we can evaluate

$$\begin{aligned} &\sum_{i,j} \mathbb{E}_{U_1} \text{tr} \left[U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} |i\rangle \langle j| U_1 \rho_{\text{in}} U_1^{\dagger} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} |j\rangle \langle i| \right] \\ &= \frac{1}{d^2-1} \left(\text{tr} \left[\zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} \right] \text{tr} \left[\zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \right] + \text{tr} \left[\rho_{\text{in}}^2 \right] \text{tr} \left[\zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \right] \right) \\ &\quad - \frac{1}{d(d^2-1)} \left(\text{tr} \left[\rho_{\text{in}}^2 \right] \text{tr} \left[\zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} \right] \text{tr} \left[\zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \right] + \text{tr} \left[\zeta_{\beta}^{\dagger} H^{(2)} \zeta_{\alpha} \zeta_{\beta'}^{\dagger} H^{(2)} \zeta_{\alpha'} \right] \right), \end{aligned} \quad (7.45)$$

where we used the formula in Eq. (7.29). Substituting Eq. (7.45) back into Eq. (7.44), we obtain

$$\begin{aligned}
& \mathbb{E}_{U_1} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \frac{1}{d^2-1} \left(\text{tr} [\zeta_\beta^\dagger H^{(2)} \zeta_\alpha] \text{tr} [\zeta_{\beta'}^\dagger H^{(2)} \zeta_{\alpha'}] \left(1 - \frac{\text{tr} [\rho_{\text{in}}^2]}{d} \right) \right) \\
&+ \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \frac{1}{d^2-1} \left(\text{tr} [\zeta_\beta^\dagger H^{(2)} \zeta_\alpha \zeta_{\beta'}^\dagger H^{(2)} \zeta_{\alpha'}] \left(\text{tr} [\rho_{\text{in}}^2] - \frac{1}{d} \right) \right) \\
&= \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \frac{1}{d^2-1} \left(\Delta \rho_{\text{in},1} \text{tr} [\zeta_\beta^\dagger H^{(2)} \zeta_\alpha] \text{tr} [\zeta_{\beta'}^\dagger H^{(2)} \zeta_{\alpha'}] + \Delta \rho_{\text{in},2} \text{tr} [\zeta_\beta^\dagger H^{(2)} \zeta_\alpha \zeta_{\beta'}^\dagger H^{(2)} \zeta_{\alpha'}] \right).
\end{aligned} \tag{7.46}$$

In addition, the first term in Eq. (7.46) can be further simplified as

$$\begin{aligned}
\sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \text{tr} [\zeta_\beta^\dagger H^{(2)} \zeta_\alpha] \text{tr} [\zeta_{\beta'}^\dagger H^{(2)} \zeta_{\alpha'}] &= \left(\sum_{(\alpha,\beta)} \text{tr} [H^{(2)} \zeta_\alpha \zeta_\beta^\dagger] \right) \left(\sum_{(\alpha',\beta')} \text{tr} [H^{(2)} \zeta_{\alpha'} \zeta_{\beta'}^\dagger] \right) \\
&= 4 \delta_{\mu\nu} \text{tr} [H^{(2)}] \delta_{\mu'\nu'} \text{tr} [H^{(2)}] \\
&= 4 \delta_{\mu\nu} \delta_{\mu'\nu'} \text{tr}^2 [H].
\end{aligned} \tag{7.47}$$

Finally, we obtain

$$\begin{aligned}
& \mathbb{E}_{U_1, U_2} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{\text{tr}^2 [H]}{d^2-1} \Delta \rho_{\text{in},1} \delta_{\mu\nu} \delta_{\mu'\nu'} + \frac{1}{4(d^2-1)} \Delta \rho_{\text{in},2} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \mathbb{E}_{U_2} \left[\text{tr} [\zeta_\beta^\dagger H^{(2)} \zeta_\alpha \zeta_{\beta'}^\dagger H^{(2)} \zeta_{\alpha'}] \right].
\end{aligned} \tag{7.48}$$

□

Lemma 17. Suppose that the random quantum circuits U_1 and U_2 are independent. If U_2 forms a unitary t -design with $t \geq 2$, the expectation of the product of two matrix elements for S can be evaluated as

$$\mathbb{E}_{U_1, U_2} [S_{\mu\nu} S_{\mu'\nu'}] = \frac{\text{tr}^2 [\rho_{\text{in}}]}{d^2-1} \Delta H_1 \delta_{\mu\nu} \delta_{\mu'\nu'} + \frac{1}{4(d^2-1)} \Delta H_2 \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \mathbb{E}_{U_1} \left[\text{tr} [\zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger \zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger] \right]. \tag{7.49}$$

Here, the summation runs in the same way as lemma 16.

Proof. This proof follows the same flow in lemma 16.

$$\begin{aligned}
& \mathbb{E}_{U_2} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{1}{4} \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')} \mathbb{E}_{U_2} \left[\text{tr} [U_2^\dagger H U_2 \zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger] \text{tr} [U_2^\dagger H U_2 \zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger] \right] \\
&= \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \sum_{\mathbf{i}, \mathbf{j}} \mathbb{E}_{U_2} \left[\text{tr} [U_2 \zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger U_2^\dagger |\mathbf{i}\rangle \langle \mathbf{j}| H U_2 \zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger U_2^\dagger |\mathbf{j}\rangle \langle \mathbf{i}| H] \right].
\end{aligned} \tag{7.50}$$

If U_2 forms a unitary t -design with $t \geq 2$, we obtain

$$\begin{aligned}
& \sum_{\mathbf{i}, \mathbf{j}} \mathbb{E}_{U_2} \left[\text{tr} [U_2 \zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger U_2^\dagger |\mathbf{i}\rangle \langle \mathbf{j}| H U_2 \zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger U_2^\dagger |\mathbf{j}\rangle \langle \mathbf{i}| H] \right] \\
&= \frac{1}{d^2-1} \left(\text{tr} [\zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger] \text{tr} [\zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger] \text{tr}^2 [H] + \text{tr} [\zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger \zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger] \text{tr} [H^2] \right) \\
&- \frac{1}{d(d^2-1)} \left(\text{tr} [\zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger \zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger] \text{tr}^2 [H] + \text{tr} [\zeta_\alpha \rho_{\text{in}}^{(1)} \zeta_\beta^\dagger] \text{tr} [\zeta_{\alpha'} \rho_{\text{in}}^{(1)} \zeta_{\beta'}^\dagger] \text{tr} [H^2] \right),
\end{aligned} \tag{7.51}$$

where we used the formula in Eq. (7.29). Thus, this leads to

$$\begin{aligned}
& \mathbb{E}_{U_2} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \frac{1}{d^2-1} \left(\text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger] \text{tr} [\varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] \left(\text{tr}^2 [H] - \frac{\text{tr} [H^2]}{d} \right) \right) \\
&+ \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \frac{1}{d^2-1} \left(\text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] \left(\text{tr} [H^2] - \frac{\text{tr}^2 [H]}{d} \right) \right) \\
&= \frac{1}{4} \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \frac{1}{d^2-1} \left(\Delta H_1 \text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger] \text{tr} [\varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] + \Delta H_2 \text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] \right). \quad (7.52)
\end{aligned}$$

In addition, the first term in Eq. (7.52) can be further simplified as

$$\begin{aligned}
\sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger] \text{tr} [\varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] &= \left(\sum_{(\alpha,\beta)} \text{tr} [\rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_\alpha] \right) \left(\sum_{(\alpha',\beta')} \text{tr} [\rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger \varsigma_{\alpha'}] \right) \\
&= 4\delta_{\mu\nu} \text{tr} [\rho_{\text{in}}^{(1)}] \delta_{\mu'\nu'} \text{tr} [\rho_{\text{in}}^{(1)}] \\
&= 4\delta_{\mu\nu} \delta_{\mu'\nu'}. \quad (7.53)
\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}
& \mathbb{E}_{U_1, U_2} [S_{\mu\nu} S_{\mu'\nu'}] \\
&= \frac{1}{4(d^2-1)} \mathbb{E}_{U_1} \left[\Delta H_1 \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger] \text{tr} [\varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] \right] \\
&+ \frac{1}{4(d^2-1)} \mathbb{E}_{U_1} \left[\Delta H_2 \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] \right] \\
&= \frac{1}{d^2-1} \Delta H_1 \delta_{\mu\nu} \delta_{\mu'\nu'} + \frac{1}{4(d^2-1)} \Delta H_2 \sum_{(\alpha,\beta)} \sum_{(\alpha',\beta')} \mathbb{E}_{U_1} \left[\text{tr} [\varsigma_\alpha \rho_{\text{in}}^{(1)} \varsigma_\beta^\dagger \varsigma_{\alpha'} \rho_{\text{in}}^{(1)} \varsigma_{\beta'}^\dagger] \right]. \quad (7.54)
\end{aligned}$$

□

Bibliography

- [1] S. Aaronson and P. Rall. Quantum approximate counting, simplified. In *Symposium on Simplicity in Algorithms*, pages 24–32. SIAM, 2020.
- [2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116(6):061102, 2016.
- [3] D. Aharonov, J. Cotler, and X.-L. Qi. Quantum algorithmic measurement. *Nature communications*, 13(1):887, 2022.
- [4] D. An, J.-P. Liu, and L. Lin. Linear combination of hamiltonian simulation for nonunitary dynamics with optimal state preparation cost. *Phys. Rev. Lett.*, 131:150603, 2023.
- [5] P. W. Anderson. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, 1972.
- [6] G. Arrad, Y. Vinkler, D. Aharonov, and A. Retzker. Increasing sensing resolution with error correction. *Phys. Rev. Lett.*, 112:150801, 2014.
- [7] A. Aspect, J. Dalibard, and G. Roger. Experimental test of bell’s inequalities using time-varying analyzers. *Phys. Rev. Lett.*, 49:1804–1807, 1982.
- [8] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.
- [9] R. Babbush, D. W. Berry, R. Kothari, R. D. Somma, and N. Wiebe. Exponential quantum speedup in simulating coupled classical oscillators. *Phys. Rev. X*, 13:041041, 2023.
- [10] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven. Encoding electronic spectra in quantum circuits with linear t complexity. *Phys. Rev. X*, 8:041015, 2018.
- [11] O. E. Barndorff-Nielsen and R. D. Gill. Fisher information in quantum statistics. *Journal of Physics A: Mathematical and General*, 33(24):4481, 2000.
- [12] F. Belliard and V. Giovannetti. Achieving heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error. *Phys. Rev. A*, 102:042613, 2020.
- [13] D. W. Berry, A. M. Childs, and R. Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, FOCS ’15, page 792–809, USA, 2015.
- [14] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman. How to perform the most accurate possible phase measurements. *Phys. Rev. A*, 80:052114, 2009.
- [15] D. W. Berry and H. M. Wiseman. Optimal states and almost optimal adaptive measurements for quantum interferometry. *Phys. Rev. Lett.*, 85:5098–5101, 2000.
- [16] D. W. Berry, H. M. Wiseman, and J. K. Breslin. Optimal input states and feedback for interferometric phase estimation. *Phys. Rev. A*, 63:053804, 2001.

- [17] M. E. Beverland, P. Murali, M. Troyer, K. M. Svore, T. Hoeffler, V. Kliuchnikov, G. H. Low, M. Soeken, A. Sundaram, and A. Vashchilo. Assessing requirements to scale to practical quantum advantage. *arXiv preprint arXiv:2211.07629*, 2022.
- [18] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, 2022.
- [19] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- [20] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024.
- [21] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.
- [22] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola. Quantum-enhanced measurements without entanglement. *Rev. Mod. Phys.*, 90:035006, 2018.
- [23] S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, 1994.
- [24] L. Brenner, C. Piveteau, and D. Sutter. Optimal wire cutting with classical communication. *IEEE Transactions on Information Theory*, 2025.
- [25] E. G. Brown, O. Goktas, and W. Tham. Quantum amplitude estimation in the presence of noise. *arXiv preprint arXiv:2006.14145*, 2020.
- [26] A. Callison and D. Browne. Improved maximum-likelihood quantum amplitude estimation. *arXiv preprint arXiv:2209.03321*, 2022.
- [27] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.
- [28] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12, 2021.
- [29] S. Chakrabarti, R. Krishnakumar, G. Mazzola, N. Stamatopoulos, S. Woerner, and W. J. Zeng. A threshold for quantum advantage in derivative pricing. *Quantum*, 5:463, 2021.
- [30] R. Chao, D. Ding, A. Gilyen, C. Huang, and M. Szegedy. Finding angles for quantum signal processing with machine precision. *arXiv preprint arXiv:2003.02831*, 2020.
- [31] S. Chen, J. Cotler, H.-Y. Huang, and J. Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585, 2022.
- [32] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, 115(38):9456–9461, 2018.
- [33] A. M. Childs and N. Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *arXiv preprint arXiv:1202.5822*, 2012.
- [34] A. W. Chin, S. F. Huelga, and M. B. Plenio. Quantum metrology in non-markovian environments. *Phys. Rev. Lett.*, 109:233601, 2012.
- [35] V. Cimini, E. Polino, F. Belliardo, F. Hoch, B. Piccirillo, N. Spagnolo, V. Giovannetti, and F. Sciarrino. Experimental metrology beyond the standard quantum limit for a wide resources range. *npj Quantum Information*, 9(1):20, 2023.

- [36] R. Cleve and C. Wang. Efficient quantum algorithms for simulating lindblad evolution. *arXiv preprint arXiv:1612.09512*, 2016.
- [37] A. Cornelissen, Y. Hamoudi, and S. Jerbi. Near-optimal quantum algorithms for multivariate mean estimation. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 33–43, New York, NY, USA, 2022.
- [38] A. Cornelissen and S. Jerbi. Quantum algorithms for multivariate monte carlo estimation. *arXiv preprint arXiv:2107.03410*, 2021.
- [39] H. Cramér. *Mathematical methods of statistics*. Princeton University Press, 1946.
- [40] A. M. Dalzell, N. Hunter-Jones, and F. G. Brandão. Random quantum circuits transform local noise into global white noise. *Communications in Mathematical Physics*, 405(3):78, 2024.
- [41] C. L. Degen, F. Reinhard, and P. Cappellaro. Quantum sensing. *Rev. Mod. Phys.*, 89:035002, 2017.
- [42] R. Demkowicz-Dobrzański and L. Maccone. Using entanglement against noise in quantum metrology. *Phys. Rev. Lett.*, 113:250801, 2014.
- [43] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă. Multi-parameter estimation beyond quantum fisher information. *Journal of Physics A: Mathematical and Theoretical*, 53(36):363001, 2020.
- [44] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă. The elusive heisenberg limit in quantum-enhanced metrology. *Nature communications*, 3(1):1063, 2012.
- [45] Y. Dong, X. Meng, K. B. Whaley, and L. Lin. Efficient phase-factor evaluation in quantum signal processing. *Phys. Rev. A*, 103:042419, 2021.
- [46] W. Dür, M. Skotiniotis, F. Fröwis, and B. Kraus. Improved quantum metrology using quantum error correction. *Phys. Rev. Lett.*, 112:080801, 2014.
- [47] A. Dutkiewicz, T. E. O’Brien, and T. Schuster. The advantage of quantum control in many-body Hamiltonian learning. *Quantum*, 8:1537, 2024.
- [48] A. Dutkiewicz, B. M. Terhal, and T. E. O’Brien. Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits. *Quantum*, 6:830, 2022.
- [49] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, et al. Fault-tolerant control of an error-corrected qubit. *Nature*, 598(7880):281–286, 2021.
- [50] S. Endo, S. C. Benjamin, and Y. Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, 2018.
- [51] P. Faist, M. P. Woods, V. V. Albert, J. M. Renes, J. Eisert, and J. Preskill. Time-energy uncertainty relation for noisy quantum metrology. In *Quantum Information and Measurement VI 2021*, page W2A.3. Optica Publishing Group, 2021.
- [52] J. Foldager and B. Koczor. Can shallow quantum circuits scramble local noise into global white noise? *Journal of Physics A: Mathematical and Theoretical*, 57(1):015306, 2023.
- [53] S. J. Freedman and J. F. Clauser. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941, 1972.
- [54] A. Fujiwara. Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *Journal of Physics A: Mathematical and General*, 39(40):12489, 2006.
- [55] A. Fujiwara. Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *Journal of Physics A: Mathematical and Theoretical*, 44(7):079501, 2011.

- [56] A. Gilyén, S. Arunachalam, and N. Wiebe. Optimizing quantum optimization algorithms via faster quantum gradient computation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, page 1425–1444, USA, 2019.
- [57] A. Gilyén, Y. Su, G. H. Low, and 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*, STOC 2019, page 193–204, New York, NY, USA, 2019.
- [58] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum-enhanced measurements: beating the standard quantum limit. *Science*, 306(5700):1330–1336, 2004.
- [59] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, 2006.
- [60] V. Giovannetti, S. Lloyd, and L. Maccone. Advances in quantum metrology. *Nature Photonics*, 5(4):222–229, 2011.
- [61] T. Giurgica-Tiron, I. Kerenidis, F. Labib, A. Prakash, and W. Zeng. Low depth algorithms for quantum amplitude estimation. *Quantum*, 6:745, 2022.
- [62] B. Go, H. Kwon, S. Park, D. Patel, and M. M. Wilde. Sample-based hamiltonian and lindbladian simulation: non-asymptotic analysis of sample complexity. *Quantum Science and Technology*, 10(4):045058, 2025.
- [63] Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614(7949):676–681, 2023.
- [64] Google Quantum AI and Collaborators. Quantum error correction below the surface code threshold. *Nature*, 638(8052):920–926, 2025.
- [65] W. Górecki. Heisenberg limit beyond quantum fisher information. *arXiv preprint arXiv:2304.14370*, 2023.
- [66] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry. π -corrected heisenberg limit. *Phys. Rev. Lett.*, 124:030501, 2020.
- [67] W. Górecki, A. Riccardi, and L. Maccone. Quantum metrology of noisy spreading channels. *Phys. Rev. Lett.*, 129:240503, 2022.
- [68] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17(5):821–825, 1976.
- [69] D. Grinko, J. Gacon, C. Zoufal, and S. Woerner. Iterative quantum amplitude estimation. *npj Quantum Information*, 7(1):1–6, 2021.
- [70] J. Haah. Product Decomposition of Periodic Functions in Quantum Signal Processing. *Quantum*, 3:190, 2019.
- [71] H. Harada, K. Wada, and N. Yamamoto. Doubly optimal parallel wire cutting without ancilla qubits. *PRX Quantum*, 5:040308, 2024.
- [72] M. Hayashi. *Asymptotic Theory of Quantum Statistical Inference*, volume 125. WORLD SCIENTIFIC, 2005.
- [73] M. Hayashi. *Quantum information*. Springer, 2006.
- [74] M. Hayashi, S. Vinjanampathy, and L. Kwek. Resolving unattainable cramer-rao bounds for quantum sensors. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 52(1):015503, 2018.
- [75] C. Helstrom. The minimum variance of estimates in quantum signal detection. *IEEE Transactions on Information Theory*, 14(2):234–242, 1968.

- [76] C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, 1969.
- [77] B. Higgins, D. Berry, S. Bartlett, M. Mitchell, H. Wiseman, and G. Pryde. Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements. *New Journal of Physics*, 11(7):073023, 2009.
- [78] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde. Entanglement-free heisenberg-limited phase estimation. *Nature*, 450(7168):393–396, 2007.
- [79] A. S. Holevo. *Probabilistic and statistical aspects of quantum theory*, volume 1. Springer Science & Business Media, 2011.
- [80] S. Hong, J. Ur Rehman, Y.-S. Kim, Y.-W. Cho, S.-W. Lee, H. Jung, S. Moon, S.-W. Han, and H.-T. Lim. Quantum enhanced multiple-phase estimation with multi-mode n 00 n states. *Nature communications*, 12(1):5211, 2021.
- [81] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, et al. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022.
- [82] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.
- [83] H.-Y. Huang, R. Kueng, and J. Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Phys. Rev. Lett.*, 126:190505, 2021.
- [84] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022.
- [85] W. J. Huggins, S. McArdle, T. E. O’Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean. Virtual distillation for quantum error mitigation. *Phys. Rev. X*, 11:041036, 2021.
- [86] W. J. Huggins, K. Wan, J. McClean, T. E. O’Brien, N. Wiebe, and R. Babbush. Nearly optimal quantum algorithm for estimating multiple expectation values. *Phys. Rev. Lett.*, 129:240501, 2022.
- [87] D. Šafránek. Simple expression for the quantum fisher information matrix. *Phys. Rev. A*, 97:042322, 2018.
- [88] Z. Ji, G. Wang, R. Duan, Y. Feng, and M. Ying. Parameter estimation of quantum channels. *IEEE Transactions on Information Theory*, 54(11):5172–5185, 2008.
- [89] Z. Jiang. Quantum fisher information for states in exponential form. *Phys. Rev. A*, 89:032128, 2014.
- [90] P. D. Johnson, A. A. Kunitsa, J. F. Gonthier, M. D. Radin, C. Buda, E. J. Daskocil, C. M. Abuan, and J. Romero. Reducing the cost of energy estimation in the variational quantum eigensolver algorithm with robust amplitude estimation. *arXiv preprint arXiv:2203.07275*, 2022.
- [91] S. P. Jordan. Fast quantum algorithm for numerical gradient estimation. *Phys. Rev. Lett.*, 95:050501, 2005.
- [92] T. Kaftal and R. Demkowicz-Dobrzański. Usefulness of an enhanced kitaev phase-estimation algorithm in quantum metrology and computation. *Phys. Rev. A*, 90:062313, 2014.
- [93] A. Katabarwa, K. Gratsea, A. Caesura, and P. D. Johnson. Early fault-tolerant quantum computing. *PRX Quantum*, 5:020101, 2024.
- [94] J. Kato, K. Wada, K. Ito, and N. Yamamoto. Exponentially accurate open quantum simulation via randomized dissipation with minimal ancilla. *arXiv preprint arXiv:2412.19453*, 2024.

- [95] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin. Quantum error correction for metrology. *Phys. Rev. Lett.*, 112:150802, 2014.
- [96] S. Kimmel, C. Y.-Y. Lin, G. H. Low, M. Ozols, and T. J. Yoder. Hamiltonian simulation with optimal sample complexity. *npj Quantum Information*, 3(1):13, 2017.
- [97] S. Kimmel, G. H. Low, and T. J. Yoder. Robust calibration of a universal single-qubit gate set via robust phase estimation. *Phys. Rev. A*, 92:062315, 2015.
- [98] R. King, D. Gosset, R. Kothari, and R. Babbush. Triply efficient shadow tomography. *PRX Quantum*, 6:010336, 2025.
- [99] A. Y. Kitaev. Quantum measurements and the abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.
- [100] E. Knill, G. Ortiz, and R. D. Somma. Optimal quantum measurements of expectation values of observables. *Phys. Rev. A*, 75(1):012328, 2007.
- [101] B. Koczor and S. C. Benjamin. Quantum analytic descent. *Phys. Rev. Res.*, 4(2):023017, 2022.
- [102] D. E. Koh, G. Wang, P. D. Johnson, and Y. Cao. A framework for engineering quantum likelihood functions for expectation estimation. *arXiv preprint arXiv:2006.09349*, 2020.
- [103] Y. Koizumi, K. Wada, W. Mizukami, and N. Yoshioka. Comprehensive study on heisenberg-limited quantum algorithms for multiple observables estimation. *arXiv preprint arXiv:2505.00698*, 2025.
- [104] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, et al. Realizing repeated quantum error correction in a distance-three surface code. *Nature*, 605(7911):669–674, 2022.
- [105] S. Kurdziałek, W. Górecki, F. Albarelli, and R. Demkowicz-Dobrzański. Using adaptiveness and causal superpositions against noise in quantum metrology. *Phys. Rev. Lett.*, 131:090801, 2023.
- [106] L. Laneve. Robust oracle quantum-state preparation via quantum signal processing. *arXiv preprint arXiv:2305.04705*, 2023.
- [107] R. B. Laughlin and D. Pines. The theory of everything. *Proceedings of the National Academy of Sciences*, 97(1):28–31, 2000.
- [108] L. Lin and Y. Tong. Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers. *PRX Quantum*, 3(1):010318, 2022.
- [109] G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, 1976.
- [110] J. Liu, H. Yuan, X.-M. Lu, and X. Wang. Quantum fisher information matrix and multiparameter estimation. *Journal of Physics A: Mathematical and Theoretical*, 53(2):023001, 2020.
- [111] Y. Liu, S. Arunachalam, and K. Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9):1013–1017, 2021.
- [112] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.
- [113] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014.
- [114] G. H. Low and I. L. Chuang. Hamiltonian simulation by uniform spectral amplification. *arXiv preprint arXiv:1707.05391*, 2017.
- [115] G. H. Low and I. L. Chuang. Optimal hamiltonian simulation by quantum signal processing. *Phys. Rev. Lett.*, 118:010501, 2017.
- [116] G. H. Low and I. L. Chuang. Hamiltonian Simulation by Qubitization. *Quantum*, 3:163, 2019.

- [117] G. H. Low, T. J. Yoder, and I. L. Chuang. Methodology of resonant equiangular composite quantum gates. *Phys. Rev. X*, 6:041067, 2016.
- [118] A. Luis and J. Peřina. Optimum phase-shift estimation and the quantum description of the phase difference. *Phys. Rev. A*, 54:4564–4570, 1996.
- [119] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang. Grand unification of quantum algorithms. *PRX Quantum*, 2:040203, 2021.
- [120] Y. Matsuzaki, S. C. Benjamin, and J. Fitzsimons. Magnetic field sensing beyond the standard quantum limit under the effect of decoherence. *Phys. Rev. A*, 84:012103, 2011.
- [121] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9, 2018.
- [122] K. Mitarai and K. Fujii. Constructing a virtual two-qubit gate by sampling single-qubit operations. *New Journal of Physics*, 23(2):023021, 2021.
- [123] M. Mitchell. Metrology with entangled states. In *Quantum Communications and Quantum Imaging III*, volume 5893, pages 263–272. SPIE, 2005.
- [124] K. Mizuta and K. Fujii. Recursive quantum eigenvalue and singular-value transformation: Analytic construction of matrix sign function by newton iteration. *Phys. Rev. Res.*, 6:L012007, 2024.
- [125] K. Nakaji. Faster amplitude estimation. *Quantum Inf. Comput.*, 2020.
- [126] K. Nakaji and N. Yamamoto. Expressibility of the alternating layered ansatz for quantum computation. *Quantum*, 5:434, 2021.
- [127] K. M. Nakanishi, K. Fujii, and S. Todo. Sequential minimal optimization for quantum-classical hybrid algorithms. *Phys. Rev. Res.*, 2(4):043158, 2020.
- [128] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [129] T. E. O’Brien, M. Streif, N. C. Rubin, R. Santagati, Y. Su, W. J. Huggins, J. J. Goings, N. Moll, E. Kyoseva, M. Degroote, C. S. Tautermann, J. Lee, D. W. Berry, N. Wiebe, and R. Babbush. Efficient quantum computation of molecular forces and other energy gradients. *Phys. Rev. Res.*, 4:043210, 2022.
- [130] T. E. O’Brien, B. Tarasinski, and B. M. Terhal. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New Journal of Physics*, 21(2):023022, 2019.
- [131] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi. Experimental demonstration of adaptive quantum state estimation. *Phys. Rev. Lett.*, 109:130404, 2012.
- [132] M. Ostaszewski, E. Grant, and M. Benedetti. Structure optimization for parameterized quantum circuits. *Quantum*, 5:391, 2021.
- [133] M. G. Paris. Quantum estimation for quantum technology. *International Journal of Quantum Information*, 7(supp01):125–137, 2009.
- [134] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon. A jacobi diagonalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization. *arXiv preprint arXiv:1904.03206*, 2019.
- [135] T. Peng, A. W. Harrow, M. Ozols, and X. Wu. Simulating large quantum circuits on a small quantum computer. *Phys. Rev. Lett.*, 125:150504, 2020.
- [136] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):1–7, 2014.

- [137] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd. Advances in photonic quantum sensing. *Nature Photonics*, 12(12):724–733, 2018.
- [138] C. Piveteau and D. Sutter. Circuit knitting with classical communication. *IEEE Transactions on Information Theory*, 70(4):2734–2745, 2023.
- [139] P. Rall. Quantum algorithms for estimating physical quantities using block encodings. *Phys. Rev. A*, 102:022408, 2020.
- [140] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, et al. Realization of real-time fault-tolerant quantum error correction. *Phys. Rev. X*, 11(4):041058, 2021.
- [141] D. Šafránek. Simple expression for the quantum fisher information matrix. *Phys. Rev. A*, 97(4):042322, 2018.
- [142] B. C. Sanders and G. J. Milburn. Optimal quantum measurements for phase estimation. *Phys. Rev. Lett.*, 75:2944–2947, 1995.
- [143] K. P. Seshadreesan, S. Kim, J. P. Dowling, and H. Lee. Phase estimation at the quantum cramer-rao bound via parity detection. *Phys. Rev. A*, 87:043833, 2013.
- [144] J. Shao. *Mathematical statistics*. Springer Science & Business Media, 2003.
- [145] L. Slattery, B. Villalonga, and B. K. Clark. Unitary block optimization for variational quantum algorithms. *Phys. Rev. Res.*, 4(2):023072, 2022.
- [146] N. Stamatopoulos, D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner. Option Pricing using Quantum Computers. *Quantum*, 4:291, 2020.
- [147] N. Stamatopoulos, G. Mazzola, S. Woerner, and W. J. Zeng. Towards Quantum Advantage in Financial Market Risk using Quantum Gradient Algorithms. *Quantum*, 6:770, 2022.
- [148] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga. Quantum error mitigation as a universal error reduction technique: Applications from the nisq to the fault-tolerant quantum computing eras. *PRX Quantum*, 3:010345, 2022.
- [149] Y. Suzuki, S. Uno, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto. Amplitude estimation without phase estimation. *Quantum Information Processing*, 19(2):1–17, 2020.
- [150] K. M. Svore, M. B. Hastings, and M. Freedman. Faster phase estimation. *Quantum Inf. Comput.*, 14:306–328, 2014.
- [151] M. Szczykulska, T. Baumgratz, and A. Datta. Multi-parameter quantum metrology. *Advances in Physics: X*, 1(4):621–639, 2016.
- [152] T. Tanaka, Y. Suzuki, S. Uno, R. Raymond, T. Onodera, and N. Yamamoto. Amplitude estimation via maximum likelihood on noisy quantum computer. *Quantum Information Processing*, 20(9):1–29, 2021.
- [153] T. Tanaka, S. Uno, T. Onodera, N. Yamamoto, and Y. Suzuki. Noisy quantum amplitude estimation without noise estimation. *Phys. Rev. A*, 105:012411, 2022.
- [154] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, et al. The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986:1–128, 2022.
- [155] G. Tóth and I. Apellaniz. Quantum metrology from a quantum information science perspective. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424006, 2014.
- [156] S. Uno, Y. Suzuki, K. Hisanaga, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto. Modified grover operator for quantum amplitude estimation. *New Journal of Physics*, 23(8):083031, 2021.

- [157] M. Valeri, V. Cimini, S. Piacentini, F. Ceccarelli, E. Polino, F. Hoch, G. Bizzarri, G. Corrielli, N. Spagnolo, R. Osellame, and F. Sciarrino. Experimental multiparameter quantum metrology in adaptive regime. *Phys. Rev. Res.*, 5:013138, 2023.
- [158] J. van Apeldoorn. Quantum Probability Oracles & Multidimensional Amplitude Estimation. In M.-H. Hsieh, editor, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)*, volume 197 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:11, Dagstuhl, Germany, 2021.
- [159] J. van Apeldoorn, A. Cornelissen, A. Gilyén, and 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.
- [160] W. van Dam, G. M. D’Ariano, A. Ekert, C. Macchiavello, and M. Mosca. Optimal quantum circuits for general phase estimation. *Phys. Rev. Lett.*, 98:090501, 2007.
- [161] J. G. Vidal and D. O. Theis. Calculus on parameterized quantum circuits. *arXiv preprint arXiv:1812.06323*, 2018.
- [162] K. Wada, K. Fukuchi, and N. Yamamoto. Quantum-enhanced mean value estimation via adaptive measurement. *Quantum*, 8:1463, 2024.
- [163] K. Wada, J. Kato, H. Harada, and N. Yamamoto. State-to-hamiltonian conversion with a few copies. *arXiv preprint arXiv:2509.14791*, 2025.
- [164] K. Wada, R. Raymond, Y.-y. Ohnishi, E. Kaminishi, M. Sugawara, N. Yamamoto, and H. C. Watanabe. Simulating time evolution with fully optimized single-qubit gates on parametrized quantum circuits. *Phys. Rev. A*, 105(6):062421, 2022.
- [165] K. Wada, R. Raymond, Y. Sato, and H. C. Watanabe. Sequential optimal selections of single-qubit gates in parameterized quantum circuits. *Quantum Science and Technology*, 9(3):035030, 2024.
- [166] K. Wada, N. Yamamoto, and N. Yoshioka. Heisenberg-limited adaptive gradient estimation for multiple observables. *PRX Quantum*, 6:020308, 2025.
- [167] G. Wang, D. E. Koh, P. D. Johnson, and Y. Cao. Minimizing estimation runtime on noisy quantum computers. *PRX Quantum*, 2:010346, 2021.
- [168] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature communications*, 12(1):1–11, 2021.
- [169] H. C. Watanabe, R. Raymond, Y.-Y. Ohnishi, E. Kaminishi, and M. Sugawara. Optimizing parameterized quantum circuits with free-axis single-qubit gates. *IEEE Transactions on Quantum Engineering*, 4:1–16, 2023.
- [170] D. Wecker, M. B. Hastings, and M. Troyer. Progress towards practical quantum variational algorithms. *Phys. Rev. A*, 92(4):042303, 2015.
- [171] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger. Violation of bell’s inequality under strict einstein locality conditions. *Phys. Rev. Lett.*, 81:5039–5043, 1998.
- [172] N. Wiebe and C. Granade. Efficient bayesian phase estimation. *Phys. Rev. Lett.*, 117:010503, 2016.
- [173] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, 2022.
- [174] W. K. Wootters and B. D. Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, 1989.
- [175] S. Yamamoto and N. Yoshioka. Robust angle finding for generalized quantum signal processing. *arXiv preprint arXiv:2402.03016*, 2024.

- [176] Y. Yao, L. Ge, X. Xiao, X. Wang, and C. P. Sun. Multiple phase estimation for arbitrary pure states under white noise. *Phys. Rev. A*, 90:062113, 2014.
- [177] L. Ying. Stable factorization for phase factors of quantum signal processing. *Quantum*, 6:842, 2022.
- [178] N. Yoshioka, T. Okubo, Y. Suzuki, Y. Koizumi, and W. Mizukami. Hunting for quantum-classical crossover in condensed matter problems. *npj Quantum Information*, 10(1):45, 2024.
- [179] X. Yuan, B. Regula, R. Takagi, and M. Gu. Virtual quantum resource distillation. *Phys. Rev. Lett.*, 132:050203, 2024.
- [180] S. Zhou and L. Jiang. Asymptotic theory of quantum channel estimation. *PRX Quantum*, 2:010343, 2021.
- [181] S. Zhou, M. Zhang, J. Preskill, and L. Jiang. Achieving the heisenberg limit in quantum metrology using quantum error correction. *Nature communications*, 9(1):78, 2018.
- [182] I. Zintchenko and N. Wiebe. Randomized gap and amplitude estimation. *Phys. Rev. A*, 93:062306, 2016.
- [183] M. Zwiernik, C. A. Pérez-Delgado, and P. Kok. Ultimate limits to quantum metrology and the meaning of the heisenberg limit. *Phys. Rev. A*, 85:042112, 2012.
- [184] 久保川達也. **現代数理統計学の基礎**. 共立出版, 2017.