

Resource-efficient algorithm for estimating the trace of quantum state powers

Myeongjin Shin^{*1,2}, Junseo Lee^{*2,3}, Seungwoo Lee^{1,2}, and Kabgyun Jeong^{2,4,5}

¹School of Computing, KAIST, Daejeon 34141, Korea

²Team QST, Seoul National University, Seoul 08826, Korea

³Quantum AI Team, Norma Inc., Seoul 04799, Korea

⁴Research Institute of Mathematics, Seoul National University, Seoul 08826, Korea

⁵School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea

Estimating the trace of quantum state powers, $\text{Tr}(\rho^k)$, for k identical quantum states is a fundamental task with numerous applications in quantum information processing, including nonlinear function estimation of quantum states and entanglement detection. On near-term quantum devices, reducing the required quantum circuit depth, the number of multi-qubit quantum operations, and the copies of the quantum state needed for such computations is crucial. In this work, inspired by the Newton-Girard method, we significantly improve upon existing results by introducing an algorithm that requires only $\mathcal{O}(\tilde{r})$ qubits and $\mathcal{O}(\tilde{r})$ multi-qubit gates, where $\tilde{r} = \min\{\text{rank}(\rho), \lceil \ln(2k/\epsilon) \rceil\}$. This approach is efficient, as it employs the $\tilde{r}$ -entangled copy measurement instead of the conventional k -entangled copy measurement, while asymptotically preserving the known sample complexity upper bound. Furthermore, we prove that estimating $\{\text{Tr}(\rho^i)\}_{i=1}^{\tilde{r}}$ is sufficient to approximate $\text{Tr}(\rho^k)$ even for large integers $k > \tilde{r}$. This leads to a rank-dependent complexity for solving the problem, providing an efficient algorithm for low-rank quantum states while also improving existing methods when the rank is unknown or when the state is not low-rank. Building upon these advantages, we extend our algorithm to the estimation of $\text{Tr}(M\rho^k)$ for arbitrary observables and $\text{Tr}(\rho^k\sigma^l)$ for multiple quantum states.

Myeongjin Shin*: hanwoolmj@kaist.ac.kr, <https://myeongjinshin.github.io/>

Junseo Lee*: harris.junseo@gmail.com, <https://harris-junseo-lee.github.io/>

Seungwoo Lee: smilelee9@kaist.ac.kr

Kabgyun Jeong: kgjeong6@snu.ac.kr,

*The first two authors contributed equally to this work.

Contents

1	Introduction	3
1.1	Trace of quantum state powers	3
1.2	Organization of the paper	4
1.3	Literature review	4
2	Iterative algorithm for estimating the trace of quantum state powers	7
2.1	Intuition: Newton-Girard method	7
2.2	Explicit algorithm construction	9
3	Analysis of the proposed algorithm	11
3.1	Rank is all you need	12
3.2	Effective rank is all you need	14
3.3	Trace of quantum state powers with arbitrary observables	17
4	Numerical simulations	20
4.1	Simulation setup	20
4.2	Simulation result	21
5	Applications in quantum information	25
5.1	Nonlinear function calculations for quantum states	25
5.2	Quantum Gibbs state preparation	26
5.3	Efficient estimation of the trace of products of quantum state powers	26
5.4	Entanglement detection	28
6	Concluding remarks	28
6.1	Summary of findings	28
6.2	Future research directions	29
A	Omitted proofs	35
A.1	Proof of Lemma 3.1	35
A.2	Proof of Theorem 3.1	36
A.3	Proof of Corollary 3.1	37
A.4	Proof of Lemma 3.2	38
A.4.1	Bounding a_t	39
A.5	Proof of Theorem 3.2	40
A.6	Proof of Theorem 3.3	41
A.7	Proof of Corollary 3.3	43
A.8	Proof of Theorem 5.2	44
B	Additional numerical simulations	44

1 Introduction

1.1 Trace of quantum state powers

Estimation task for the trace of the product of identical density matrices, which is represented as

$$\text{Tr}(\rho^k) \quad \text{‘trace of quantum state powers’}$$

given access to copies of the quantum state ρ , is a core subroutine for many algorithms and applications in quantum information theory. We refer to this quantity as the ‘trace of quantum state powers,’ which is used to calculate the value of integer Rényi entropy [1, 2, 3], nonlinear functions of quantum states [4, 5, 6, 7, 8], and deducing the eigenvalues of the quantum state, a process known as entanglement spectroscopy [1, 9].

We focus on estimating $\text{Tr}(\rho^k)$ for large integer k . The main applications are calculating the nonlinear functions of quantum states, which need estimation of the trace of large powers. Precisely, Yirka and Subaşı [9] proved that the trace of ‘well-behaved’ polynomials $g(\rho)$, such as $g(x) = (1 + x)^\alpha$ and $\log(1 + x)$, can be efficiently estimated using the trace of quantum state powers. Moreover, $\text{Tr}(e^{\beta\rho})$ is an example with applications in thermodynamics.

The preparation of quantum Gibbs states [10, 11, 12, 13, 14] is an essential part of quantum computation, used in various applications such as quantum simulation, quantum optimization, and quantum machine learning. The truncated Taylor series

$$S_k(\rho) = \sum_{i=1}^k \text{Tr}((\rho - I)^i \rho) \tag{1.1}$$

is exploited as the cost function for variational quantum Gibbs state preparation [10], which can be calculated by $\{\text{Tr}(\rho^i)\}_{i=1}^{k+1}$.

Several methods for the estimation of the trace of quantum state powers have been proposed, such as the generalized swap test [4], entanglement spectroscopy via Hadamard test [1], two-copy test [15], qubit-efficient entanglement spectroscopy [9], multivariate trace estimation [16], and methods using randomized measurement such as classical shadows [6, 17, 18]. An analysis of these methods is performed in Section 1.3.

Our work is inspired by the Newton-Girard method, as demonstrated in Section 2.1. Specifically, we use quantum devices only to estimate $\{\text{Tr}(\rho^i)\}_{i=1}^{\tilde{r}}$, where $\tilde{r} = \min\{r, \lceil \ln(2k/\epsilon) \rceil\}$. (From this point onward, we consistently use r to denote the rank of the quantum state ρ throughout the paper.) Subsequently, we use a classical computer with a recursive formula to calculate $\{\text{Tr}(\rho^i)\}_{i=1}^k$, with an additive error of less than ϵ for large $k \in \mathbb{N}$. In Section 3.1, we prove that the rank r is sufficient, implying that quantum devices are required only for $\{\text{Tr}(\rho^i)\}_{i=1}^r$. By defining the notion of ‘effective rank’ in Section 3.2, we further prove a more advanced theorem that the effective rank $\tilde{r}$ is sufficient for estimating the trace of quantum state powers. The Newton-Girard method and recursion are used in the proof.

Furthermore, we argue that combining our work with previous ones [1, 4, 9, 15, 16, 19, 20] improves its algorithmic performance. The number of needed qubits (i.e., width of the circuit) and the required multi-qubit gates are reduced. We support our work with numerical simulations. To emphasize the importance of our work, we demonstrate advantages when applying our method to applications such as calculating nonlinear functions of quantum states, preparation of quantum Gibbs states, and entanglement detection.

Quantity	Quantum Resource Needed	Upper bound on t
$\text{Tr}(\rho^k)$ Theorem 3.1, 3.2	$\{\text{Tr}(\rho^i)\}_{i=1}^t$	$\min\{\text{rank}(\rho), \lceil \ln(2k/\epsilon) \rceil\}$
$\text{Tr}(M\rho^k)$ Theorem 3.3	$\{\text{Tr}(\rho^i)\}_{i=1}^t, \{\text{Tr}(M\rho^i)\}_{i=1}^t$	$\min\{\text{rank}(\rho), \lceil \ln(2k\ M\ _\infty/\epsilon) \rceil\}$
$\text{Tr}(\rho^k\sigma^l)$ Theorem 5.2	$\{\text{Tr}(\rho^i)\}_{i=1}^t, \{\text{Tr}(\sigma^i)\}_{i=1}^t, \{\text{Tr}(\rho^i\sigma^j)\}_{(i,j)=(1,1)}^{(t,t)}$	$\min\{\max\{\text{rank}(\rho), \text{rank}(\sigma)\}, \lceil \ln((4k+4l)/\epsilon) \rceil\}$

Table 1: **Summary of quantum resource requirements and effective rank conditions for ϵ -additive estimations.** This table summarizes the key results of the paper. It presents the range of values to be obtained through quantum resources for each of the three physical quantities, and further details can be found in the corresponding theorems.

1.2 Organization of the paper

Our paper is structured as follows. In Section 1.3, we review existing results on attempts to estimate the trace of quantum state powers. This includes results derived from variations of the swap test, several other approaches, and key related studies. In Section 2.1, we introduce the Newton-Girard method, which serves as the fundamental principle of our algorithm. Then, in Section 2.2, we describe how we specifically design our algorithm using this method. Subsequently, we analyze our algorithm in detail. In Section 3.1, we explain how the quantum resources required for our algorithm are related to the rank of the quantum state. In Section 3.2, we strengthen our algorithm by introducing the concept of effective rank, allowing it to be applied even when the exact rank of the quantum state is unknown. In Section 3.3, we extend the problem to the case of arbitrary observables, which is a more generalized version based on the quantum resources required for estimating the trace of quantum state powers, as determined in previous sections. Section 4 discusses the results of numerical simulations demonstrating the operation of our algorithm, and Section 5 explores how our algorithm can be applied to other quantum information tasks. Finally, in Section 6, we summarize our study, discuss its limitations, and outline potential directions for future research. The proofs of all the theorems, corollaries, and lemmas presented in the paper are provided in Appendix A. The table summarizing our main results is presented in Table 1. The $\tilde{O}(\cdot)$ asymptotic notation used in our paper hides polylogarithmic factors in certain variables. To ensure clarity in each context, we explicitly specify which variables are involved whenever necessary, and repeat the explanation when appropriate. Throughout this paper, unless otherwise noted, n denotes the number of qubits and d denotes the dimension of the quantum state, with $d = 2^n$.

1.3 Literature review

The swap test (ST) [19, 20, 21, 22, 23] estimates $\text{Tr}(\rho\sigma)$, the trace of the product of two matrices ρ and σ :

$$\text{Tr}(S(\rho \otimes \sigma)) = \text{Tr}(\rho\sigma), \quad (1.2)$$

where S denotes the swap operator. The ST can be performed using 1 ancilla qubit with 1 controlled-SWAP (CSWAP) operation and 2 Hadamard gates as shown in Fig. 1. The ST can be thought of as performing the observable S on Eq. (1.2). The observation that quantities like $\text{Tr}(\rho\sigma)$ can be estimated without the need for full-state tomography was a significant development.

Following this line of thinking, Ekert *et al.* [4] proposed a cyclic shift permutation operator W^π for a generalized ST:

$$\text{Tr}(W^\pi(\rho_1 \otimes \dots \otimes \rho_k)) = \text{Tr}(\rho_1 \dots \rho_k). \quad (1.3)$$

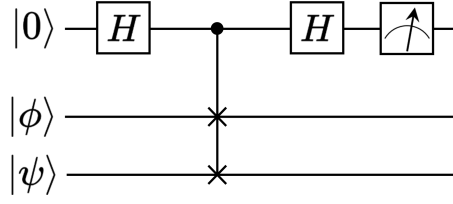


Figure 1: **Circuit implementing the swap test between two states.** The simplest case of a quantum circuit for calculating the trace of the product of two density matrices using the swap test is illustrated. It shows that 2 single-qubit gates, 1 three-qubit gate, and 1 ancilla qubit are required.

By using Eq. (1.3) above, the trace of quantum state powers $\text{Tr}(\rho^k)$ can be easily calculated. Note that regardless of the dimension $d = \dim(\rho)$ and the number of quantum states k , the generalized ST needs only $\mathcal{O}(1/\epsilon^2)$ runs on a quantum device for ϵ additive error prediction. Thus, $\mathcal{O}(k/\epsilon^2)$ copies are needed for the estimation of $\text{Tr}(\rho^k)$. This method requires $\mathcal{O}(k)$ qubits, a quantum circuit of $\mathcal{O}(k)$ depth, and $\mathcal{O}(k)$ multi-qubit gates.

Various methods have been proposed for better estimation [1, 9, 15, 16] of the trace of quantum state powers. A comparison of these methods is shown in Table 2.

The entanglement spectroscopy via hadamard test (HT) [1] is a generalized algorithm that estimates the expectation value of an arbitrary unitary operator or observable M . Specifically, ST can be thought of as a special case of HT when $M = S$. The HT has linear depth $\mathcal{O}(k)$ and uses $\mathcal{O}(k)$ copies of the state. A more improved algorithm, the entanglement spectroscopy via two-copy test (TCT) [15], achieves constant depth and uses $\mathcal{O}(k)$ copies of the state. Thus, both use $\mathcal{O}(k)$ qubits in the estimation circuit. That is, both HT and TCT are improved algorithms but need the original entangled pure state $|\psi\rangle_{AB}$ for the estimation of $\text{Tr}(\rho_A^k)$, where $\rho_A = \text{Tr}_B(|\psi\rangle\langle\psi|_{AB})$.

Qubit-efficient entanglement spectroscopy [9] employs qubit-reset strategies to reduce the number of qubits in the quantum circuit. This method requires only n qubits, constant in terms of power k . When combined with TCT, it requires a linear circuit depth $\mathcal{O}(k)$. Also, Yirka and Subař [9] defines the notion of ‘effective depth,’ and TCT with qubit-reset strategy requires only a constant effective circuit depth $\mathcal{O}(1)$. However, this qubit-reset strategy still demands $\mathcal{O}(k)$ copies of the original entangled pure state $|\psi\rangle_{AB}$, and qubit-reset could lead to more vulnerability to noise.

Without the need for the entangled pure state $|\psi\rangle_{AB}$, multivariate trace estimation [24, 25, 26] $\text{Tr}(\rho_1\rho_2\ldots\rho_k)$, a general case of the trace of quantum state powers, has been proposed with constant quantum depth [16]. Inspired by the method of Shor error correction [27], this approach requires only constant quantum circuit depth, utilizing $\mathcal{O}(k)$ multi-qubit gates and $\mathcal{O}(k)$ qubits, and establishes numerous applications for multivariate trace and trace-of-powers estimation. By combining our work with these advancements, we provide an advantageous solution for estimating $\text{Tr}(\rho^k)$ with large k . Specifically, leveraging multivariate trace estimation [16], we can reduce the number of required qubits from $\mathcal{O}(k)$ to $\mathcal{O}(\tilde{r})$ and multi-qubit gates from $\mathcal{O}(k)$ to $\mathcal{O}(\tilde{r})$ for $\text{Tr}(\rho^k)$ estimation, where $\tilde{r} = \min\{r, \lceil \ln(2k/\epsilon) \rceil\}$.

There are alternative methods that use classical shadows [17, 18] to estimate $\text{Tr}(\rho^k)$. Using

$$\text{Tr}(W^\pi(\rho \otimes \ldots \otimes \rho)) = \text{Tr}(\rho^k), \quad (1.4)$$

and linearly combining the classical snapshots of ρ , we can obtain a classical random variable whose expectation is $\text{Tr}(\rho^k)$. The advantage of these alternative methods is that they allow for measurements to be taken sequentially and do not rely on the assumption that the samples of ρ used by the algorithm are identical and independent [16]. However,

due to the exponential scaling of $\text{Tr}((W^\pi)^2)$, the sample and computational complexity are exponential in terms of qubits. So, this method requires the number of copies as the dimension d of the states. Recently, Pelecanos, Tan, Tang, and Wright [28] proposed a nonlinear extension of classical shadow estimation to estimate $\text{Tr}(\rho^k)$ using a natural unbiased estimator motivated by U-statistics. Suppose we are given N copies of a quantum state ρ and a fixed positive integer $k \leq N$. For each copy, a uniform POVM is performed, and the outcome $|u_i\rangle$ on the i -th copy is used to define the associated observable $\hat{\sigma}_i := (d+1)|u_i\rangle\langle u_i| - I$. An unbiased estimator for $\text{Tr}(\rho^k)$ is given by

$$Z_k := \frac{1}{N^{\underline{k}}} \sum_{\substack{i_1, \dots, i_k \in [n] \\ \text{distinct}}} \text{Tr}(\hat{\sigma}_{i_1} \cdots \hat{\sigma}_{i_k}), \quad (1.5)$$

where $N^{\underline{k}} := N(N-1) \cdots (N-k+1)$ denotes the falling factorial. For any quantum state ρ of dimension d , and any fixed $k \geq 2$, it was shown that with probability at least 0.99, the estimator Z_k approximates $\text{Tr}(\rho^k)$ up to a multiplicative error ϵ , using $N = \mathcal{O}(\max\{d^{2-2/k}/\epsilon^2, d^{3-2/k}/\epsilon^{2/k}\})$ copies of ρ . Naturally, this results in exponential scaling with respect to the number of qubits. Note that their estimator provides a multiplicative-error approximation for $\text{Tr}(\rho^k)$, which immediately yields an additive-error estimator for the quantum Rényi entropy. In contrast, our work focuses on estimating $\text{Tr}(\rho^k)$ up to an additive error, which in turn implies an additive-error estimator for the quantum Tsallis entropy.

Several studies have explored the relationship between the trace of quantum state powers, quantum entanglement, and separability testing. Among them, Bradshaw *et al.* [29] investigates quantum separability tests from the perspective of combinatorial group theory, uncovering a fundamental link between the acceptance probabilities of these tests and the cycle index polynomials of finite groups. The cycle index polynomial of a permutation group $\mathcal{G}$ is defined as

$$Z(\mathcal{G})(x_1, \dots, x_n) := \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \prod_{i=1}^n x_i^{c_i(g)}, \quad (1.6)$$

where $c_j(g)$ represents the number of cycles of length j in the disjoint cycle decomposition of g . Notably, in the generalization of the bipartite pure-state separability algorithm, the acceptance probability associated with a group $\mathcal{G}$, denoted as $p_{\mathcal{G}}$, takes the form

$$p_{\mathcal{G}} = Z(\mathcal{G})(1, \dots, \text{Tr}(\rho^k)). \quad (1.7)$$

This implies that $p_{\mathcal{G}}$ is determined by evaluating the cycle index polynomial of $\mathcal{G}$ at $x_j = \text{Tr}(\rho^j)$ for $j \in \{1, \dots, k\}$. The study first derives an exact analytical expression for the probability of a mixedness test accepting as the number of state copies increases, showing that this probability is governed by the cycle index polynomial of the symmetric group. Building on this insight, the authors extend the framework to develop a family of separability tests corresponding to arbitrary finite groups, proving that the acceptance probability aligns with the cycle index polynomial of the respective group. Furthermore, they propose explicit quantum circuit implementations for these tests, leveraging CSWAP gates in a resource-efficient manner—scaling as $\mathcal{O}(k^2)$ for the symmetric group and $\mathcal{O}(k \ln(k))$ for the cyclic group, where k denotes the number of state copies used in the test. The study of partial transpose moments and entanglement detection was discussed in the work by Neven *et al.* [30], and Section 5.4 provides a more detailed discussion on how our work can be applied to that research. Additionally, Wagner *et al.* [31] proposed simple quantum circuits for measuring weak values, Kirkwood–Dirac (KD) quasiprobability distributions,

and the spectra of quantum states without post-selection, particularly by interpreting the trace of quantum state powers from the perspective of measuring unitary-invariant and relational properties of quantum states using Bargmann invariants.

Finally, the work by Liu and Wang [32], published several months after the first version of our paper appeared on arXiv, provides a detailed computational complexity analysis of estimating the trace of powers of an n -qubit mixed quantum state ρ , given a state-preparation circuit of size $\text{poly}(n)$. Leveraging efficiently computable uniform approximations of positive power functions within the framework of quantum singular value transformation, the authors achieved an exponential improvement over previously known methods. Their study focused particularly on estimating the quantum Tsallis entropy,

$$S_k(\rho) = \frac{1 - \text{Tr}(\rho^k)}{k - 1}, \quad (1.8)$$

and precisely identified the thresholds at which the computational complexity of the problem undergoes qualitative changes. Specifically, they showed that for $k = 1$, the problem is **NIQSZK-complete**; for $1 < k \leq 1 + (n - 1)^{-1}$, it is **NIQSZK-hard**; for $1 + \Omega(1) \leq k \leq 2$, it becomes **BQP-complete**; and for $k > 2$, it remains within **BQP**. They also established rigorous bounds on the query and sample complexity across different regimes of k , with particular attention to rank-dependent behavior. In their publication, the authors characterized the method proposed in our initial preprint as a rank-dependent estimator for the quantum Tsallis entropy in the regime where k exceeds the rank of the quantum state. Our current results further strengthen this interpretation, as our method now provides an ϵ -additive estimator whenever k exceeds the effective rank $\tilde{r} = \min \{r, \lceil \ln(2k/\epsilon) \rceil\}$.

2 Iterative algorithm for estimating the trace of quantum state powers

In this section, we explain the Newton-Girard method and discuss how it is utilized in the design of our algorithm. In proving the main theorems, the key idea underlying our improvement is the use of the Newton-Girard identities, which establish an explicit relationship between power sums (e.g., $x_1^k + x_2^k$) and elementary symmetric polynomials (e.g., $x_1 + x_2$, x_1x_2). By leveraging these identities alongside a careful and systematic analysis, we can efficiently express higher-order moments in terms of lower-order symmetric functions, thereby reducing the overall estimation complexity. This connection provides a clear algebraic intuition behind our approach and highlights why fewer moment estimations suffice.

2.1 Intuition: Newton-Girard method

The main idea of entanglement spectroscopy demonstrates that the trace of quantum state powers can be used to estimate the largest eigenvalues [1, 9, 33]. The k largest eigenvalues can be estimated using $\{\text{Tr}(\rho^i)\}_{i=1}^k$. The Newton-Girard method [34, 35] provides the mathematical foundation of entanglement spectroscopy and serves as an important component in our method. Therefore, we describe the details of the Newton-Girard method and explain the inspiration that leads to the notion that ‘*rank is sufficient*’ for estimating the trace of quantum state powers.

Let $r = \text{rank}(\rho)$, and the eigenvalues of ρ are $\{p_i\}_{i=1}^r$, sorted in descending order. We utilize the Newton-Girard method to leverage the following well-known result from linear algebra and provide an intuition for it: knowing the trace of quantum state powers

$\{\text{Tr}(\rho^i)\}_{i=1}^r$ is equivalent to knowing $\{p_i\}_{i=1}^r$. Consider the equation having these eigenvalues as root in the form of

$$\prod_{m=1}^r (x - p_m) = 0. \quad (2.1)$$

The values of $\text{Tr}(\rho^i)$ are now the i -th power sum of the roots. Denote the power sum as

$$P_i := \sum_{m=1}^r p_m^i = \text{Tr}(\rho^i). \quad (2.2)$$

Here, Simply expanding the terms of the Eq. (2.1) above as follows:

$$\prod_{m=1}^r (x - p_m) = \sum_{k=0}^r (-1)^k a_k x^{r-k}, \quad (2.3)$$

where a_k is the elementary symmetric polynomial, defined as the sum of all distinct products of k distinct variables, such as:

$$\begin{aligned} a_0 &= 1, \\ a_1 &= p_1 + p_2 + \dots + p_r = \sum_{1 \leq \alpha \leq r} p_\alpha, \\ a_2 &= p_1 p_2 + p_1 p_3 + \dots + p_{r-1} p_r = \sum_{1 \leq \alpha < \beta \leq r} p_\alpha p_\beta, \\ a_3 &= \sum_{1 \leq \alpha < \beta < \gamma \leq r} p_\alpha p_\beta p_\gamma, \\ &\vdots \\ a_r &= p_1 p_2 \dots p_r. \end{aligned}$$

The Newton-Girard method states the relationship of the elementary symmetric polynomials and the power sums recursively as follows. For all $r \geq k \geq 1$,

$$a_k = \frac{1}{k} \sum_{i=1}^k (-1)^{i-1} a_{k-i} P_i. \quad (2.4)$$

Given P_i for $1 \leq i \leq r$, we can uniquely determine the values of a_k on the right-hand side of Eq. (2.3). Moreover, the set of eigenvalues $\{p_i\}_{i=1}^r$ is also uniquely determined as the roots of Eq. (2.1).

Unfortunately in real-world situations, we cannot exactly calculate the trace of quantum state powers; instead, we can obtain the estimation with errors using previous strategies. Then, it is natural to ask the following question:

“If the error of estimated power sums is small, are the roots obtained by the Newton-Girard method close to the eigenvalues of ρ ?”

No, the statement is not always true. A counterexample is Wilkinson’s polynomial [36], which shows that the location of the roots can be very sensitive to perturbations in the coefficients of a polynomial. Generally, to obtain the eigenvalues, the estimation error of the trace of quantum state powers should be exponential, causing the copy and time complexity to be exponential [1]. Therefore, estimating the eigenvalues with the estimated values of $\{\text{Tr}(\rho^i)\}_{i=1}^r$ is unfeasible.

However, we get an intuition from the Newton-Girard method that estimating $\{\text{Tr}(\rho^i)\}_{i=1}^r$ contains valuable information about the quantum states. In [Section 3.1](#), we prove that estimating the trace of quantum state powers $\{\text{Tr}(\rho^i)\}_{i=1}^r$ is sufficient for estimating the trace of larger powers $\text{Tr}(\rho^i)$ for $i > r$. The error of each eigenvalue obtained by the Newton-Girard method is large, but as the power of the eigenvalues is summed up, the error diminishes to a smaller extent.

2.2 Explicit algorithm construction

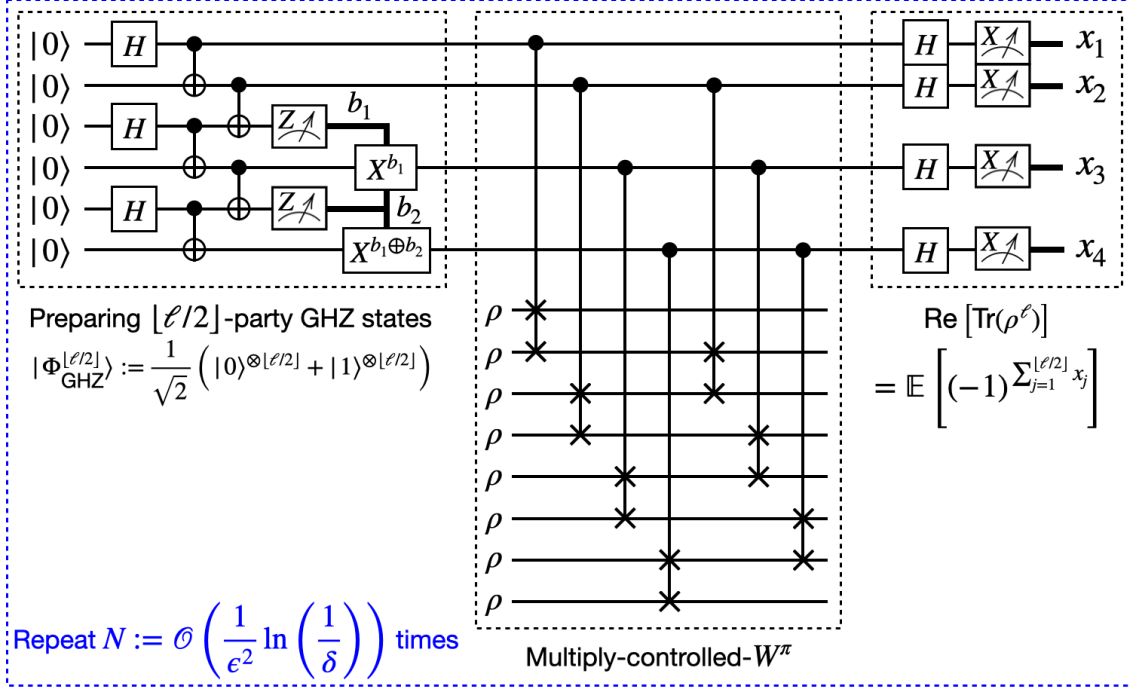


Figure 2: **Quantum circuit for Step 1 of the [Algorithm 1]: A detailed example for $\ell = 8$.** This quantum circuit is used to estimate $\text{Tr}(\rho^\ell)$. The first part of the circuit corresponds to the GHZ state preparation described in Step 1(a). As mentioned in the main text, this step can be implemented differently if necessary. Following this, a multiply-controlled cyclic shift operation is applied, with slight structural variations depending on whether $\ell \bmod 2$ is 0 or 1. The type of gate applied before measurement and the measurement basis used depend on whether $\text{Re}[\text{Tr}(\rho^\ell)]$ or $\text{Im}[\text{Tr}(\rho^\ell)]$ is being estimated. By calculating the expectation of the measured outcomes, the desired physical quantity can be estimated. To ensure a good estimate with an additive error of at most ϵ with high probability $1 - \delta$, as guaranteed by [Eq. \(2.9\)](#), $\mathcal{O}(\ln(1/\delta)/\epsilon^2)$ repetitions of the steps within the blue box in the figure are required.

Based on the insights gained in [Section 2.1](#), the specific algorithm for calculating the trace of quantum state powers is as follows. (In this section, the index is denoted by ℓ to avoid confusion with $i = \sqrt{-1}$, which represents the imaginary unit.)

[Algorithm 1] Estimation of $\text{Tr}(\rho^k)$

1. Based on the circuit presented in [Fig. 2](#), the values of $\text{Tr}(\rho^\ell)$ for $\ell = 1, 2, \dots, t$ are obtained. This process is based on research on multivariate trace estimation using constant quantum depth [\[16\]](#), and the detailed procedure is as follows: (An example for $\ell = 8$ is illustrated in [Fig. 2](#) for reference.)

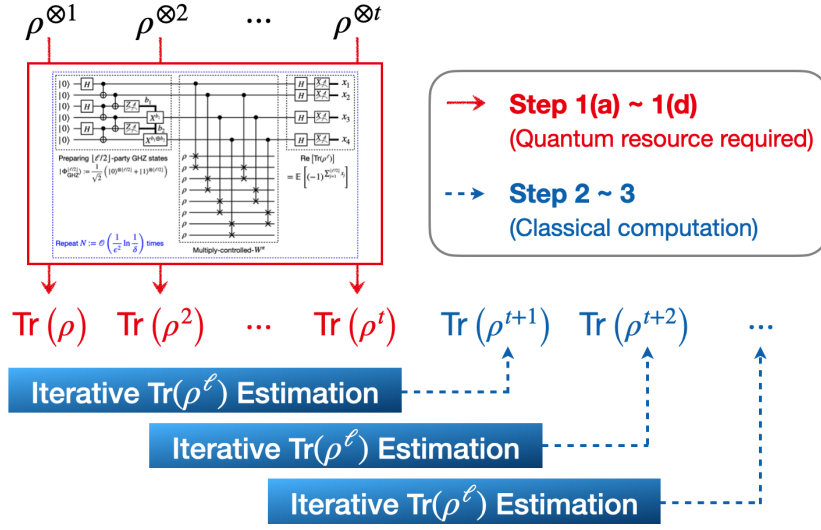


Figure 3: **Diagram of the complete process of the [Algorithm 1].** The red box represents the process shown in Fig. 2, corresponding to Step 1 of the algorithm described in Section 2.2. Quantum resources are required only for this step, during which the values from $\text{Tr}(\rho)$ to $\text{Tr}(\rho^t)$ are obtained. The subsequent blue dashed lines indicate computations performed using a simple recurrence relation without requiring quantum resources, following the processes outlined in Steps 2 and 3.

- (a) Generate an $\lfloor \ell/2 \rfloor$ -party GHZ state

$$|\Phi_{\text{GHZ}}^{\lfloor \ell/2 \rfloor}\rangle := \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes \lfloor \ell/2 \rfloor} + |1\rangle^{\otimes \lfloor \ell/2 \rfloor} \right). \quad (2.5)$$

This process utilizes mid-circuit measurement and requires a constant quantum-depth circuit along with classical feedback, while a logarithmic-depth classical circuit is needed for parity computation. Besides the method illustrated in Fig. 2, it is also possible to employ other methods for generating a GHZ state.

- (b) Next, a multiply-controlled cyclic shift operation is performed. Depending on whether ℓ is odd or even, slight structural modifications to the circuit may be necessary. The specifics are discussed in detail in of [16, Section 3.2], and this process can be achieved with constant quantum depth.
- (c) To estimate $\text{Re}[\text{Tr}(\rho^\ell)]$, apply an H gate to all $\lfloor \ell/2 \rfloor$ qubits and measure in the X -basis.

Note: While the multivariate trace estimation problem in the original study requires the estimation of both the real and imaginary parts, our problem focuses solely on estimating the trace of quantum powers, making the estimation of the real part sufficient. However, for the sake of completeness, we also describe the process for estimating the imaginary part: replace the H gate with an $HS^\dagger$ gate and measure in the Y -basis to estimate $\text{Im}[\text{Tr}(\rho^\ell)]$.

- (d) Repeat the process from Step 1(a) to 1(c)

$$N := \mathcal{O} \left(\frac{\ln(1/\delta)}{\epsilon^2} \right) \quad (2.6)$$

times, and let the measurement outcomes (0 or 1) obtained in Step 1(c) for the m -th iteration be denoted as $x_1^m, \dots, x_{\lfloor \ell/2 \rfloor}^m, y_1^m, \dots, y_{\lfloor \ell/2 \rfloor}^m$. Then, the quantity

we aim to estimate, $\text{Tr}(\rho^\ell)$, is expressed as:

$$Q_{\ell(\leq t)} := \hat{\mathcal{R}} + i\hat{\mathcal{J}} \approx \text{Tr}(\rho^\ell), \quad (2.7)$$

$$\text{where } \hat{\mathcal{R}} = \frac{\sum_{m=1}^N \sum_{j=1}^{\lfloor \ell/2 \rfloor} (-1)^{x_j^m}}{N}, \quad \text{and } \hat{\mathcal{J}} = \frac{\sum_{m=1}^N \sum_{j=1}^{\lfloor \ell/2 \rfloor} (-1)^{y_j^m}}{N}. \quad (2.8)$$

Then this estimate satisfies the inequality below for $\ell = 1, 2, \dots, t$.

$$\Pr \left(|Q_\ell - \text{Tr}(\rho^\ell)| \leq \epsilon \right) \geq 1 - \delta. \quad (2.9)$$

Note: As mentioned in Step 1(c), $\text{Tr}(\rho^\ell) \in \mathbb{R}$, so it does not matter if we set $Q_{\ell(\leq t)} = \hat{\mathcal{R}}$. This is because, through this algorithm, we have

$$\sqrt{(\hat{\mathcal{R}} - \text{Tr}(\rho^\ell))^2 + \hat{\mathcal{J}}^2} \leq \epsilon, \quad (2.10)$$

which also ensures that $|\hat{\mathcal{R}} - \text{Tr}(\rho^\ell)| \leq \epsilon$.

2. Calculate the elementary symmetric polynomial b_k ($1 \leq k \leq t$) defined as:

$$b_k = \frac{1}{k} \sum_{\ell=1}^k (-1)^{\ell-1} b_{k-\ell} Q_\ell, \quad b_0 = 1. \quad (2.11)$$

3. Using $Q_1, \dots, Q_t$ obtained from Step 1 and $b_1, \dots, b_t$ obtained from Step 2, the value of $\text{Tr}(\rho^\ell)$ ($\ell > t$) can be estimated through the following recurrence relation:

$$Q_{\ell(>t)} := \sum_{k=1}^t (-1)^{k-1} b_k Q_{\ell-k} \approx \text{Tr}(\rho^\ell). \quad (2.12)$$

Through Step 3, we can obtain values for $Q_{t+1}, Q_{t+2}, \dots$, and in [Section 3.1](#) and [3.2](#), we analyze in detail the conditions on t required to ensure that the estimated values obtained in this process are within an additive error of at most ϵ . See [Fig. 3](#) for the overall process of the algorithm we propose. Note that quantum devices are only used to estimate $\{\text{Tr}(\rho^i)\}_{i=1}^t$. At most $\mathcal{O}(t)$ qubits and $\mathcal{O}(t)$ multi-qubit gates are required (used only in Step 1).

3 Analysis of the proposed algorithm

We analyze our proposed algorithm in two phases.

- (1) In [Section 3.1](#), we show that $t \geq r$ is sufficient, identifying the rank dependence.
- (2) Then, in [Section 3.2](#), we prove that $t \geq \lceil \ln(2k/\epsilon) \rceil$ is sufficient for estimating trace of quantum state powers within an additive error of ϵ .

This introduces the new concept of the effective rank, leading to stronger results and enabling the algorithm to be applicable even when the exact rank is unknown. In [Section 3.3](#), we discuss the case that includes arbitrary observables, which is a more generalized version of the problem of estimating the trace of quantum state powers.

For clarity, we summarize the notations used in this section. Let P_i represent the *exact values* of the trace of quantum state powers:

$$P_i := \text{Tr}(\rho^i) = \sum_{j=1}^r p_j^i. \quad (3.1)$$

Similarly, let Q_i denote the *estimated values* of the trace of quantum state powers. For $Q_{i(\leq t)}$, the values are estimated by the quantum device, while for $Q_{i(>t)}$, they are defined by Eq. (2.12). The estimation (additive) error is denoted as

$$\epsilon_i := Q_i - P_i. \quad (3.2)$$

Next, let a_k and b_k represent the elementary symmetric polynomials corresponding to P_i and Q_i , respectively (see Eq. (2.4) and Eq. (2.11)). In Lemma 3.1, we analyze the bound on the difference between these two elementary symmetric polynomials.

In cases where quantum resources are so limited that even utilizing resources commensurate with the rank is infeasible, we may not be able to estimate all elements of $\{\text{Tr}(\rho^i)\}_{i=1}^r$ but only up to $\{\text{Tr}(\rho^i)\}_{i=1}^{t(<r)}$. To account for this limitation, we introduce a new quantity, denoted as $\tilde{P}$, which is defined as follows for $i \leq t$:

$$\tilde{P}_{i(\leq t)} := \text{Tr}(\rho^i) = \sum_{j=1}^r p_j^i. \quad (3.3)$$

For $i > t$, $\tilde{P}_i$ is recursively defined based on the Newton-Girard recurrence relations, where the elementary symmetric polynomials a_k are identical to P_i : (Since $\tilde{P} = P$ when $t = r$, the recurrence relation can also be applied to P in this case.)

$$\tilde{P}_{i(>t)} := \sum_{k=1}^t (-1)^{k-1} a_k \tilde{P}_{i-k}. \quad (3.4)$$

While the introduction of $\tilde{P}$ may appear indistinguishable from the original definition of P in Eq. (3.1), a fundamental distinction arises when $t < r$, as $\tilde{P}_{i(>t)} \neq P_{i(>t)}$. The concept of $\tilde{P}$ is particularly useful for quantifying the impact of information loss on error when only partial spectral information is available, and in Lemma 3.2, we provide a rigorous quantitative analysis of the discrepancy between $\tilde{P}$ and P .

For the problem of the trace of quantum state powers with arbitrary observables, given a quantum state

$$\rho = \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i| \quad (3.5)$$

and an arbitrary observable M , we can define $P_{i,M}$ similarly to P_i as follows:

$$P_{k,M} := \text{Tr}(M\rho^k) = \sum_{i=1}^r \langle\psi_i|M|\psi_i\rangle p_i^k. \quad (3.6)$$

Likewise, $Q_{i,M}$ represents its estimated value.

3.1 Rank is all you need

In this section, we prove that $t \geq r$ is sufficient to proceed with the algorithm while maintaining low error, and we derive the required number of quantum circuit runs. Although

the method is simple, we argue that it offers advantages in terms of the number of required qubits and multi-qubit gates. To the best of our knowledge, our work is the first to prove that the traces of rank-at-most- r powers, $\{\text{Tr}(\rho^i)\}_{i=1}^r$, are sufficient for estimating $\text{Tr}(\rho^k)$ when k is large. Furthermore, it provides an efficient algorithm, particularly for low-rank quantum states. Our goal is to first establish a quantitative bound on the difference between the elementary symmetric polynomials derived from the true values P_i and those obtained from the estimated values Q_i .

Lemma 3.1. *Let $d_k := b_k - a_k$, then the following holds:*

$$|d_k| \leq \sum_{j=1}^k \frac{|\epsilon_j|}{j}. \quad (3.7)$$

Proof. See the details in [Appendix A.1](#). $\square$

Now, we prove our first theorem, which demonstrates that $t = r$ is sufficient to execute our algorithm with low error.

Theorem 3.1. *Suppose that,*

$$\varepsilon_i := |\epsilon_i| = |Q_i - P_i| < \frac{\epsilon}{kt \ln t} \quad (3.8)$$

holds for $i = 1, 2, \dots, t$. Setting $t = r$ and proceeding with [\[Algorithm 1\]](#) based on the recurrence relation [Eq. \(2.12\)](#), the following relation always holds:

$$|\epsilon_i| = |Q_i - P_i| < \epsilon \quad (3.9)$$

for $i = t + 1, \dots, k$.

Proof. See the details in [Appendix A.2](#). $\square$

Based on [Theorem 3.1](#), the quantum resources required to solve the problem of estimating the trace of quantum state powers are derived in [Corollary 3.1](#).

Corollary 3.1. *To estimate $\text{Tr}(\rho^i)$ for all $i \leq k$ within an additive error of ϵ and with a success probability of at least $1 - \delta$, where $\delta \in (0, 1)$, it suffices to estimate each $\text{Tr}(\rho^j)$ for $j \leq r$ within an additive error of ϵ_j , as defined in [Theorem 3.1](#). This can be achieved by using*

$$\mathcal{O}\left(\frac{k^2 r^2 \ln^2 r \ln(1/\delta)}{\epsilon^2}\right) \quad (3.10)$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(j)$ qubits and $\mathcal{O}(j)$ CSWAP operations.

Proof. See the details in [Appendix A.3](#). $\square$

Using a quantum device, j copies of ρ are required for each run of the quantum circuit to estimate $\text{Tr}(\rho^j)$. The number of runs is the same for every $\text{Tr}(\rho^j)$, as specified in [Eq. \(3.10\)](#). Since $j \leq r$, the total number of copies needed to estimate $\{\text{Tr}(\rho^i)\}_{i=1}^k$ within an additive error of ϵ is

$$\mathcal{O}\left(\sum_{j=1}^r \frac{jk^2 r^2}{\epsilon^2} \ln^2 r\right) = \mathcal{O}\left(\frac{k^2 r^4}{\epsilon^2} \ln^2 r\right). \quad (3.11)$$

To highlight the significance of our work and aid understanding, we present the following proposition, which provides a simplified version of [Theorem 3.1](#) and [Corollary 3.1](#).

Proposition 3.1 (Informal, see [Theorem 3.1](#) and [Corollary 3.1](#)). *In [\[Algorithm 1\]](#), setting $t \geq r$ is sufficient to efficiently estimate the trace of quantum state powers, even for large powers. (In other words, the problem of estimating the trace of quantum state powers requires quantum resources proportional to the rank of the given quantum state, rather than its power k .)*

3.2 Effective rank is all you need

Method	# Depth	# Qubits	# CSWAP	# Copies	Original $ \psi\rangle$
Generalized swap test [4]	$\mathcal{O}(k)$	$\mathcal{O}(k)$	$\mathcal{O}(k)$	$\mathcal{O}(k^2/\epsilon^2)$	NOT required
Hadamard test [1]	$\mathcal{O}(k)$	$\mathcal{O}(k)$	$\mathcal{O}(k)$	$\mathcal{O}(k^2/\epsilon^2)$	Required
Two-copy test [15]	$\mathcal{O}(1)$	$\mathcal{O}(k)$	$\mathcal{O}(k)$	$\mathcal{O}(k^2/\epsilon^2)$	Required
Two-copy test & Qubit-reset [9]	$\mathcal{O}(k)$	$\mathcal{O}(1)$	$\mathcal{O}(k)$	$\mathcal{O}(k^2/\epsilon^2)$	Required
Multivariate trace estimation [16]	$\mathcal{O}(1)$	$\mathcal{O}(k)$	$\mathcal{O}(k)$	$\mathcal{O}(k^2/\epsilon^2)$	NOT required
Ours (this work)	$\mathcal{O}(1)$	$\mathcal{O}(\tilde{r})$	$\mathcal{O}(\tilde{r})$	$\tilde{\mathcal{O}}(k^2/\epsilon^2)$	NOT required

Table 2: **Summary of resources required by different algorithms to estimate the values of $\{\text{Tr}(\rho^i)\}_{i=1}^k$ within an error margin of ϵ .** The comparison includes a total of six algorithms, including ours. The algorithms are categorized based on quantum circuit depth, the number of required qubits, the number of required CSWAP operations, the number of required quantum states ρ , and whether the original state $|\psi\rangle$ is needed for the algorithm to operate. Here, the notation $\tilde{\mathcal{O}}(\cdot)$ hides polylogarithmic factors in k , and $\tilde{r} = \min\{r, \lceil \ln(2k/\epsilon) \rceil\}$ is the effective rank defined in [Eq. \(3.17\)](#).

In [Section 3.1](#), we identified for the first time how the complexity of estimating $\text{Tr}(\rho^k)$ can improve when the quantum state ρ has low rank. However, this advantage critically relies on the exact knowledge of the rank. In practice, such knowledge is rarely available, and even when the rank is known, the improvement may be marginal unless the state is *very* low-rank compared to its dimension. This motivates a broader question: under what conditions can we still benefit from our algorithm when the rank is unknown, approximately known, or when quantum resources are limited?

To address this, we move away from analyses that require precise knowledge of the rank of ρ , and instead seek a more nuanced understanding of the complexity in terms of the parameters k and ϵ , which govern the exponent in $\text{Tr}(\rho^k)$ and the target additive precision. Although the rank can have a noticeable impact when it is very small, its effect diminishes rapidly as the state becomes more full-rank. In such cases, the value of $\text{Tr}(\rho^k)$ often becomes so small that it cannot be meaningfully distinguished from zero within realistic precision bounds. This motivates us to treat k and ϵ as the central parameters driving the complexity, rather than relying on exact rank information.

This observation naturally leads to the notion of an effective rank, which captures how many eigenvalues of ρ make a meaningful contribution to $\text{Tr}(\rho^k)$ within the desired precision. Rather than focusing solely on the full rank of ρ , the effective rank offers a more nuanced and practical understanding of when the algorithm remains useful in realistic settings. It emphasizes that the true complexity is more fundamentally governed by the interplay between k and the target accuracy ϵ , rather than the sheer dimension of the system.

First, let us examine the quantitative difference between $\tilde{P}$ and P .

Lemma 3.2. *Suppose that $\tilde{P}_i$ is defined by Eq. (3.3), Eq. (3.4). Then the following holds:*

$$|\tilde{P}_k - P_k| \leq \frac{k}{t!} \left(1 - \frac{t}{r}\right). \quad (3.12)$$

Proof. See the details in Appendix A.4. $\square$

Now, we prove our second theorem, which demonstrates that $t = \lceil \ln(2k/\epsilon) \rceil$ is sufficient to execute our algorithm with low error.

Theorem 3.2. *Suppose that,*

$$\epsilon_i := |\epsilon_i| = |Q_i - P_i| < \frac{\epsilon}{2kt \ln t} \quad (3.13)$$

holds for $i = 1, 2, \dots, t$. Setting $t = \lceil \ln(2k/\epsilon) \rceil$ and proceeding with [Algorithm 1] based on the recurrence relation Eq. (2.12), the following relation always holds:

$$|\epsilon_i| = |Q_i - P_i| < \epsilon \quad (3.14)$$

for $i = t + 1, \dots, k$.

Proof. See the details in Appendix A.5. $\square$

Based on Theorem 3.2, the quantum resources required to solve the problem of estimating the trace of quantum state powers are derived in Corollary 3.2.

Corollary 3.2. *To estimate $\text{Tr}(\rho^i)$ for all $i \leq k$ within an additive error of ϵ and with a success probability of at least $1 - \delta$, where $\delta \in (0, 1)$, it suffices to estimate each $\text{Tr}(\rho^j)$ for $j \leq \lceil \ln(2k/\epsilon) \rceil$ within an additive error of ϵ_j , as defined in Theorem 3.2. This can be achieved by using*

$$\tilde{\mathcal{O}}\left(\frac{k^2 \ln(1/\delta)}{\epsilon^2}\right) \quad (3.15)$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(j)$ qubits and $\mathcal{O}(j)$ CSWAP operations. Here, the notation $\tilde{\mathcal{O}}(\cdot)$ hides polylogarithmic factors in k .

Proof. It follows the same logic as the proof of Corollary 3.1. Please refer to Appendix A.3. $\square$

Using a quantum device, j copies of ρ are required for each run of the quantum circuit to estimate $\text{Tr}(\rho^j)$. The number of runs is the same for every $\text{Tr}(\rho^j)$, as specified in Eq. (3.15). Since $j \leq \lceil \ln(2k/\epsilon) \rceil$, the total number of copies needed to estimate $\{\text{Tr}(\rho^i)\}_{i=1}^k$ within an additive error of ϵ is

$$\tilde{\mathcal{O}}\left(\frac{k^2}{\epsilon^2}\right), \quad (3.16)$$

where $\tilde{\mathcal{O}}(\cdot)$ hides polylogarithmic factors in k .

Again, to highlight the significance of our work and aid understanding, we present the following proposition, which provides a simplified version of Theorem 3.2 and Corollary 3.2.

Proposition 3.2 (Informal, see Theorem 3.2 and Corollary 3.2). *In [Algorithm 1], setting $t \geq \lceil \ln(2k/\epsilon) \rceil$ is sufficient to efficiently estimate the trace of quantum state powers $\{\text{Tr}(\rho^i)\}_{i=1}^k$ with an additive error of at most ϵ . (In other words, the problem of estimating the trace of quantum state powers requires quantum resources proportional to the logarithm of the number of powers, rather than the power k .)*

To conclude [Section 3.1](#) and [3.2](#), we summarize our findings in the following theorem:

Proposition 3.3 (Informal description of the main results). *For the problem of estimating the trace of quantum state powers, given a large integer k , it is possible to approximate $\{\text{Tr}(\rho^i)\}_{i=1}^k$ within an additive error of ϵ using quantum resources only up to $\{\text{Tr}(\rho^i)\}_{i=1}^t$, where t is given by*

$$t \geq \tilde{r} = \min \left\{ r, \left\lceil \ln \left(\frac{2k}{\epsilon} \right) \right\rceil \right\}. \quad (3.17)$$

We define $\tilde{r}$ as the effective rank.

In this way, we present a strengthened result from [Section 3.1](#), incorporating the concept of effective rank to achieve a more refined analysis.

As mentioned, our work provides an advantage in terms of the number of needed qubits and multi-qubit gates. Since we only need to estimate $\{\text{Tr}(\rho^i)\}_{i=1}^{\tilde{r}}$, $n\tilde{r}$ qubits and $\mathcal{O}(\tilde{r})$ CSWAP operations are sufficient for the estimation. We emphasize that reducing the number of qubits and CSWAP operations used in the quantum circuit is an important improvement because it is less sensitive to noise, and having fewer qubits is advantageous for implementation on near-term quantum devices [\[37, 38\]](#). The comparison of the quantum resources required by existing methods and our algorithm is summarized in [Table 2](#).

For estimating $\text{Tr}(\rho^k)$ to within additive error ϵ , our approach leverages the algorithm from [\[16\]](#) as a subroutine. When k exceeds $\tilde{r} = \min\{r, \lceil \ln(k/\epsilon) \rceil\}$, our method reduces the circuit size of each iteration from $\mathcal{O}(k)$ to $\mathcal{O}(\tilde{r})$. As a result, it becomes possible to estimate $\text{Tr}(\rho), \text{Tr}(\rho^2), \dots, \text{Tr}(\rho^k)$ for sufficiently large k with a total copy complexity of $\mathcal{O}(k^2/\epsilon^2)$, matching that of prior works [\[19, 4, 16\]](#). However, when estimating $\text{Tr}(\rho^k)$ for a single value of k , our method retains a copy complexity of $\mathcal{O}(k^2/\epsilon^2)$, whereas previous approaches require only $\mathcal{O}(k/\epsilon^2)$ in this case.

The significance of our contribution lies in scenarios where one needs to estimate all moments $\{\text{Tr}(\rho^i)\}_{i=1}^k$ simultaneously. In such cases, while we maintain the same copy complexity (up to polylogarithmic factors) as existing approaches, our method substantially reduces the quantum circuit resources required for implementation. This leads to a more resource-efficient and scalable procedure, particularly for large k , where circuit depth, qubit count, and the number of multi-qubit gates pose practical bottlenecks.

We now present an illustrative example that highlights the utility of the effective rank. Consider a d -dimensional quantum state defined by

$$\rho = \text{diag} \left(1 - \frac{1}{d}, \frac{1}{d(d-1)}, \dots, \frac{1}{d(d-1)} \right), \quad (3.18)$$

where ρ has rank d and $d = 2^n$. The trace of the k -th power of ρ is given by:

$$\text{Tr}(\rho^k) = \left(1 - \frac{1}{d} \right)^k + (d-1) \cdot \left(\frac{1}{d(d-1)} \right)^k \quad (3.19)$$

$$= \left(1 - \frac{1}{d} \right)^k + \frac{1}{d^k(d-1)^{k-1}}. \quad (3.20)$$

Now consider the regime of large k , which is the primary focus of our work. For the example state above, setting $k = d$ yields:

$$\text{Tr}(\rho^k) = \left(1 - \frac{1}{d} \right)^d + \frac{1}{d^d(d-1)^{d-1}} \quad (3.21)$$

$$\approx \frac{1}{e} + \exp(-\Theta(d \log d)) \approx \frac{1}{e}. \quad (3.22)$$

Although one might expect $\text{Tr}(\rho^k)$ to become negligibly small as k grows large, this example shows that the trace can still retain a significant value, approximately $1/e$, thanks to the contribution of the dominant eigenvalue.

This demonstrates the advantage of the effective rank perspective. Traditional approaches that rely on worst-case rank assumptions would treat this state as full-rank and thus require $\Omega(d) = \Omega(2^n)$ quantum resources to estimate $\text{Tr}(\rho^k)$ accurately. Here, “quantum resources” refer to the number of qubits, the number of multi-qubit gates, and the circuit depth required to implement the estimation algorithm. In contrast, our method based on the effective rank recognizes that only a small subset of eigenvalues contribute meaningfully to the trace, thereby reducing the quantum resource cost to $\mathcal{O}(\log(d/\epsilon)) = \mathcal{O}(n + \log(1/\epsilon))$. Crucially, this gain in circuit efficiency is achieved without significantly increasing the number of samples required: the overall sample complexity remains essentially unchanged, up to polylogarithmic factors. We believe that the notion of effective rank can offer similar benefits in many other realistic settings.

3.3 Trace of quantum state powers with arbitrary observables

The algorithm we developed for computing the trace of quantum state powers can be extended to address a more generalized problem: estimating $\text{Tr}(M\rho^k)$, where M represents an arbitrary observable. Successfully estimating this quantity would enable applications in calculating values used as subroutines in virtual distillation [39, 40], a quantum error mitigation technique.

In this problem, we consider a Pauli decomposition of the observable

$$M = \sum_{\alpha=1}^{N_M} a_{\alpha} P_{\alpha}, \quad (3.23)$$

where $a_{\alpha} \in \mathbb{R}$ and

$$P_{\alpha} = \sigma_{\alpha_1} \otimes \dots \otimes \sigma_{\alpha_n} \quad (3.24)$$

are tensor products of Pauli operators

$$\sigma_{\alpha_1}, \dots, \sigma_{\alpha_n} \in \{\sigma_x, \sigma_y, \sigma_z, I\}. \quad (3.25)$$

We assume that the bounded condition

$$\sum_{\alpha=1}^{N_M} |a_{\alpha}| = \mathcal{O}(c) \quad (3.26)$$

holds for some constant c .

[Algorithm 2] Estimation of $\text{Tr}(M\rho^k)$

1. Following Steps 1 and 2 of [Algorithm 1] in Section 2.2, we obtain the values of the elementary symmetric polynomials $b_1, \dots, b_t$.
2. Estimate $\text{Tr}(M\rho^{\ell})$ for $\ell = 1, 2, \dots, t$ using the method outlined in [39].

Note (1): The quantum circuits required for this step are designed following in [39, Propositions 1 and 2]. As highlighted in their work, the circuit structure depends on the trade-off between qubit-depth and parallelization. In this paper, we focus on describing the high-level procedure without delving into specific implementation details.

Note (2): Other methods, such as classical shadows, can be employed to estimate $\text{Tr}(M\rho^\ell)$. We emphasize that any method capable of estimating $\text{Tr}(M\rho^\ell)$ for $\ell = 1, 2, \dots, t$ can be used as a substitute for this step.

(a) For each $\alpha = 1, \dots, N_M$, the following steps are repeated

$$N := \mathcal{O} \left(\frac{\left(\sum_{\alpha=1}^{N_M} |a_\alpha| \right)^2 \ln(1/\delta)}{\epsilon^2} \right) \quad (3.27)$$

times:

- i. Prepare a GHZ state and apply a sequence of **CSWAP** gates.
- ii. Apply a controlled- P_α gate to an arbitrary register storing ρ .
- iii. Repeat the above process and measure the ancillary qubits in the X -basis and Y -basis, where the X -basis measurement is used for the real part estimation and the Y -basis measurement is used for the imaginary part estimation. The measurements obtained are then used to estimate $\text{Tr}(P_\alpha \rho^\ell)$ using the similar logic as in Step 1(d) of **[Algorithm 1]**. This estimate, denoted as $\hat{W}_\alpha$, satisfies the following inequality. The value of $\hat{W}_\alpha$ is expressed as the expectation obtained from N repetitions of the measurement process.

$$\Pr \left(\left| \hat{W}_\alpha - \text{Tr}(P_\alpha \rho^\ell) \right| \leq \frac{\epsilon}{\sum_{\alpha=1}^{N_M} |a_\alpha|} \right) \geq 1 - \delta.$$

(b) Finally, the overall expectation value

$$Q_{\ell(\leq t), M} = \frac{1}{N_M} \sum_{\alpha=1}^{N_M} a_\alpha \hat{W}_\alpha \quad (3.28)$$

serves as an estimate for $\text{Tr}(M\rho^\ell)$. Then this estimate satisfies the inequality below for $\ell = 1, 2, \dots, t$.

$$\Pr \left(\left| Q_{\ell, M} - \text{Tr}(M\rho^\ell) \right| \leq \epsilon \right) \geq 1 - \delta. \quad (3.29)$$

For reference, the sample complexity required in Step 2 is given by:

$$\mathcal{O}(N_M \cdot N) = \mathcal{O} \left(\frac{N_M \left(\sum_{\alpha=1}^{N_M} |a_\alpha| \right)^2 \ln(1/\delta)}{\epsilon^2} \right) \quad (3.30)$$

$$= \mathcal{O} \left(\frac{c^2 N_M \ln(1/\delta)}{\epsilon^2} \right). \quad (3.31)$$

3. Using $b_1, \dots, b_t$ obtained from Step 1 and $Q_{1, M}, \dots, Q_{t, M}$ obtained from Step 2, the value of $\text{Tr}(M\rho^\ell)$ for $\ell > t$ can be estimated using the following recurrence relation:

$$Q_{\ell(> t), M} := \sum_{k=1}^t (-1)^{k-1} b_k Q_{\ell-k, M} \approx \text{Tr}(M\rho^\ell). \quad (3.32)$$

Through Step 3, values for $Q_{t+1, M}, Q_{t+2, M}, \dots$ can be obtained. In this section, we analyze in detail the conditions on t required to ensure that the estimated values derived through

this process are within an additive error of at most ϵ . As mentioned, any method capable of estimating $\text{Tr}(M\rho^\ell)$ for $\ell = 1, 2, \dots, t$ can be employed in Step 2. The most suitable method should be chosen based on the specific application. For entanglement detection, classical shadows should be used in Step 2, as discussed in [Section 5.4](#). When applying our algorithm for the efficient estimation of $\text{Tr}(\rho^k \sigma^l)$, multivariate trace estimation [\[16\]](#) is utilized in Step 2, as detailed in [Section 5.3](#).

Theorem 3.3. *Suppose that*

$$\varepsilon_{i,M} := |\epsilon_{i,M}| = |P_{i,M} - Q_{i,M}| < \frac{\epsilon}{4}, \quad (3.33)$$

and

$$\varepsilon_i := |\epsilon_i| = |P_i - Q_i| < \frac{\epsilon}{2 \|M\|_\infty k t \ln t}, \quad (3.34)$$

holds for $i = 1, 2, \dots, t$, where the operator norm $\|M\|_\infty$ is defined corresponding to the ∞ -norm for vectors $\|x\|$, as

$$\|M\|_\infty = \sup_{x \neq 0} \frac{\|Mx\|_\infty}{\|x\|_\infty}. \quad (3.35)$$

Setting $t = \tilde{r}_M$ and proceeding with [\[Algorithm 2\]](#) based on the recurrence relation [Eq. \(3.32\)](#), the following relation always holds:

$$|\epsilon_{i,M}| = |P_{i,M} - Q_{i,M}| \leq \epsilon \quad (3.36)$$

for $i = t + 1, \dots, k$. Where $\tilde{r}_M$ is the effective rank for the observable M defined as:

$$\tilde{r}_M = \min \left\{ r, \left\lceil \ln \left(\frac{2k \|M\|_\infty}{\epsilon} \right) \right\rceil \right\}. \quad (3.37)$$

Proof. See the details in [Appendix A.6](#). □

Based on [Theorem 3.3](#), the quantum resources required to estimate the trace of quantum state powers with arbitrary observables are derived in [Corollary 3.3](#).

Corollary 3.3. *To estimate $\text{Tr}(M\rho^i)$ for all $i \leq k$ within an additive error of ϵ and with a success probability of at least $1 - \delta$, where $\delta \in (0, 1)$, it is necessary to estimate each $\text{Tr}(M\rho^j)$ for $j \leq \tilde{r}_M$ within an additive error of $\varepsilon_{j,M}$ as defined in [Theorem 3.3](#). This can be achieved by using*

$$\mathcal{O} \left(\frac{c^2 N_M \ln(1/\delta)}{\epsilon^2} \right) \quad (3.38)$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(j)$ qubits and $\mathcal{O}(j)$ CSWAP operations, and estimating each $\text{Tr}(\rho^{j'})$ for $j' \leq \tilde{r}_M$ within an additive error of $\varepsilon_{j'}$ as defined in [Theorem 3.3](#), by using

$$\tilde{\mathcal{O}} \left(\frac{k^2 \|M\|_\infty^2 \ln(1/\delta)}{\epsilon^2} \right) \quad (3.39)$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(j')$ qubits and $\mathcal{O}(j')$ CSWAP operations. Here, the notation $\tilde{\mathcal{O}}(\cdot)$ hides polylogarithmic factors in k .

Proof. See the details in [Appendix A.7](#). □

To conclude [Section 3.3](#), we summarize our findings in the following proposition:

Proposition 3.4 (Informal, see [Theorem 3.3](#) and [Corollary 3.3](#)). *For the problem of estimating the trace of quantum state powers with arbitrary observables, given a large integer k , it is possible to approximate $\{\text{Tr}(M\rho^i)\}_{i=1}^k$ within an additive error of ϵ using quantum resources only up to $\{\text{Tr}(\rho^i)\}_{i=1}^t$ and $\{\text{Tr}(M\rho^i)\}_{i=1}^t$, where t is given by*

$$t \geq \tilde{r}_M = \min \left\{ r, \left\lceil \ln \left(\frac{2k \|M\|_\infty}{\epsilon} \right) \right\rceil \right\}. \quad (3.40)$$

The estimation of the trace of quantum state powers with arbitrary observables also applies to the efficient estimation of $\text{Tr}(\rho^k \sigma^l)$ and is discussed in [Section 5.3](#).

4 Numerical simulations

4.1 Simulation setup

To validate the findings obtained in [Section 3](#), we conduct numerical simulations to examine the performance of our algorithm. The problem setup to be estimated, including the eigenvalue pattern, is defined as follows and the legend to be used in the graph is shown in [Fig. 4](#).

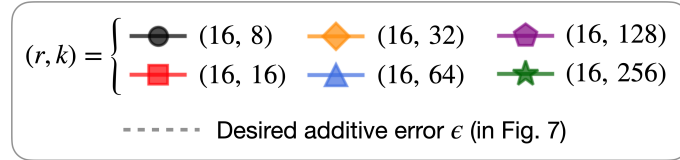


Figure 4: **Legends used in the graph.** There are six based on the (r, k) combinations, and in [Fig. 7](#), gray dashed lines are used to further represent the guarantee of estimation within additive error.

- Types of eigenvalue distributions:
 - (1) **Geometrically decaying eigenvalues**, $p_{\max}/p_{\min} = 2^{15}$.
 - (2) **Arithmetically decaying eigenvalues**, $p_{\max} - p_{\min} = 0.124$.
 - (3) **One dominant eigenvalue** $p_{\max} \approx 1$, while the remaining eigenvalues are randomly chosen small values.
 - (4) **Identical eigenvalues**, $p_i = 1/r$ for all i .
- Rank of the quantum state: $r = 16$.
- Target power k for estimating $\text{Tr}(\rho^k)$: $k \in \{8, 16, 32, 64, 128, 256\}$.
- Additive error bound ϵ for estimation: $\epsilon \in \{10^{-1}, 10^{-2}, \dots, 10^{-7}\}$.

In the case of the eigenvalue distribution, the settings for geometrically decaying and arithmetically decaying distributions are mathematically inspired problem setups. For the case of one dominant eigenvalue, the model was formulated under the assumption of an experimental situation where, due to hardware noise or other factors, it is impossible to create a perfect pure state. This situation can be generalized as a scenario where $\tilde{r} \ll r$ during the operation of our algorithm.

The simulation will be conducted for two different scenarios.

(k, ϵ)	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
8	6	8	10	12	15	16	16
16	6	9	11	13	15	16	16
32	6	9	11	13	15	16	16
64	8	10	12	15	16	16	16
128	8	11	13	15	16	16	16
256	9	11	14	16	16	16	16

Table 3: **The value of $\tilde{r}$ as a function of (k, ϵ) .** The value of t used in Scenario 1 is $\tilde{r} = \min\{r, \lceil \ln(2k/\epsilon) \rceil\}$. Note that $r = 16$.

- (1) **Scenario 1 (simulation of [Algorithm 1]):** We evaluate the actual additive error that arises when following the procedure outlined in [Algorithm 1] under a given (r, k, ϵ) setting, using $t = \tilde{r}$ for different eigenvalue distributions. (The values of $\tilde{r}$ for different (k, ϵ) are listed in Table 3.) Although Step 1 of [Algorithm 1] originally requires a quantum circuit simulation, in our case, we do not employ quantum circuits. Instead, the true value corresponding to $\{\text{Tr}(\rho^i)\}_{i=1}^t$ is numerically computed, and a simulation is performed using a sampling-based approximation. Specifically, sampling is conducted from a binomial distribution

$$B\left(n = \left\lceil \frac{k^2}{\epsilon^2} \right\rceil, p = \text{Tr}(\rho^i)\right). \quad (4.1)$$

To approximate the true value, n independent random variables are drawn from this binomial distribution, and their empirical mean is used as the estimate. Since n is chosen to satisfy Corollary 3.2, the estimation error can be maintained below ϵ .

- (2) **Scenario 2 (simulation of Lemma 3.2):** We investigate how the error evolves as the value of t is varied. In particular, we examine the error trend when $t < r$ or even when $t < \tilde{r}$. The objective is to determine the minimum value of t required to ensure that the estimation remains within a sufficiently small additive error across various distributions. In this scenario, k is fixed at 32.

4.2 Simulation result

The simulation results for geometrically decaying, arithmetically decaying, one dominant, and identical eigenvalues are shown in Fig. 5, Fig. 6, Fig. 7, and Fig. 8, respectively. Each figure consists of three subfigures: (a) the distribution of the eigenvalues, (b) Scenario 1—simulation of [Algorithm 1], and (c) Scenario 2—simulation of Lemma 3.2.

For every eigenvalue distribution, the experimental error in Scenario 1 is smaller than the target additive error ϵ , which strengthens the credibility of [Algorithm 1]. In the cases of geometrically decaying, arithmetically decaying, and identical eigenvalues, the discrepancy between the target error and the experimental error is quite large. The case of one dominant eigenvalue gives the tightest result.

For every eigenvalue distribution, the experimental error in Scenario 2 is also smaller than the target additive error ϵ , further enhancing the credibility of Lemma 3.2. Only $\{\text{Tr}(\rho^i)\}_{i=1}^t$ is obtained from quantum resources, while $\text{Tr}(\rho^{t+1})$ to $\text{Tr}(\rho^k)$ are computed using the recurrence relation described in the algorithm. As mentioned earlier, our simulation uses a sampling-based approximation instead of quantum resources. The graph presents both

$$\max_{j \in \{t+1, \dots, k\}} |P_j - \tilde{P}_j| \quad (4.2)$$

and the theoretical bound we derived, $t/k!$. In the cases of geometrically decaying, arithmetically decaying eigenvalues, and one dominant eigenvalue, the discrepancy between the theoretical bound $k/t!$ and the experimental error is quite large. The case of identical eigenvalues gives the tightest result. For every distribution we simulated, $t = 8$ is sufficient to keep the experimental error below a low threshold (e.g., always smaller than 10^{-6} , which is sufficiently small).

Additionally, as the power k increases, both the scale of $\text{Tr}(\rho^k)$ and the scale of the absolute error become very small, sometimes even dropping below the machine epsilon, which represents the smallest numerical difference a computer can accurately represent in floating-point arithmetic. To eliminate errors caused by floating-point precision limitations, we implemented our algorithm using integer fractions instead of floating-point types for iterative estimations. Specifically, we used Python's built-in `Fractions.Fraction` class, which represents rational numbers exactly as ratios of two integers. This allowed us to perform arithmetic operations with full precision, avoiding the accumulation of rounding errors that typically arise in floating-point computations. Thanks to this exact representation, our simulation was able to detect discrepancies as small as on the order of 10^{-200} .

We note that the use of exact rational arithmetic in our work is primarily for methodological purposes. In realistic near-term quantum experiments, errors from finite measurement statistics, decoherence, and other hardware imperfections are expected to far exceed floating-point precision limits. Nevertheless, we performed sampling-based simulations so that, even though such physical noise sources were not modeled, the results remain statistically meaningful. This setup aligns with the aim of our study, which is to evaluate and validate the intrinsic performance of the algorithm under idealized, noise-free conditions. Such extremely low error levels should not be expected in practice on current quantum devices.

Here, we uncover a new insight: in Eq. (A.58), the theoretical bound is derived using the scaling difference between factorial and exponential functions, such as $t! \geq 2^t$. However, this approach may not provide a sufficiently tight bound. Obtaining a closed-form lower bound for t analytically is extremely challenging, but considering Stirling's approximation,

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \dots\right), \quad (4.3)$$

we observe that the lower bound for t could be as low as

$$\mathcal{O}\left(\frac{\ln(k/\epsilon)}{\ln \ln(k/\epsilon)}\right), \quad (4.4)$$

suggesting a potentially looser bound than initially expected. And the simulation results based on this bound are included in Appendix B.

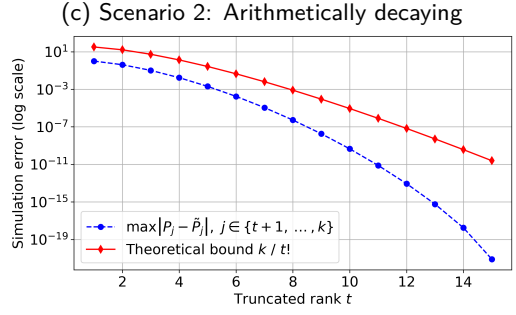
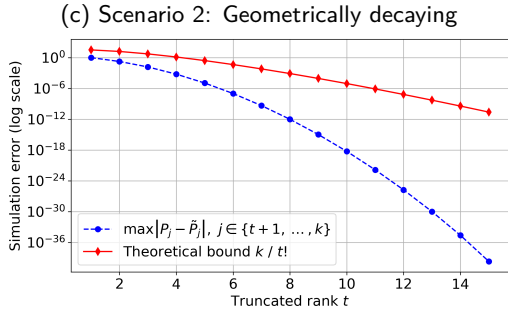
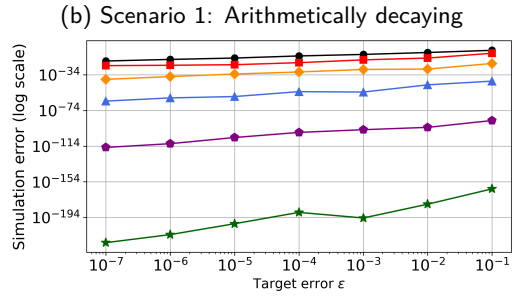
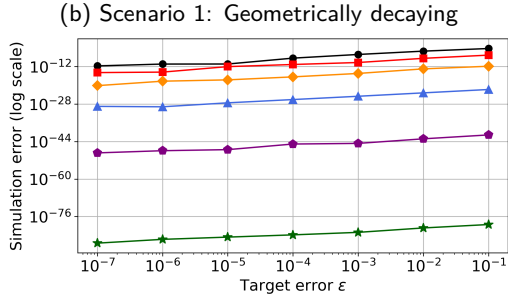
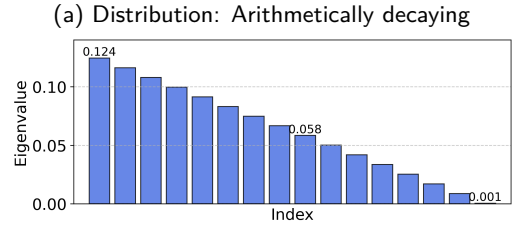
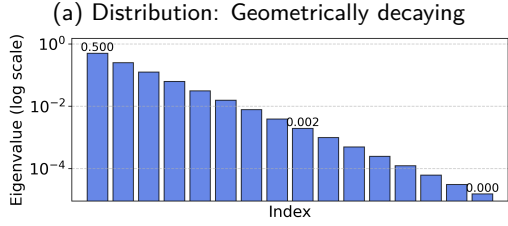


Figure 5: Simulation results for geometrically decaying eigenvalues.

Figure 6: Simulation results for arithmetically decaying eigenvalues.

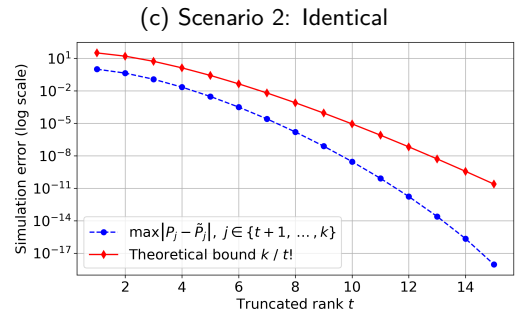
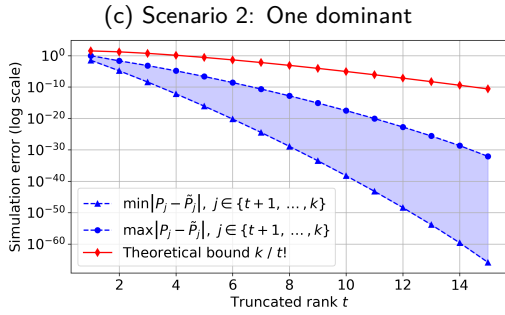
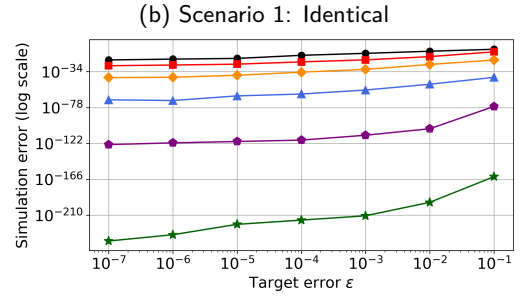
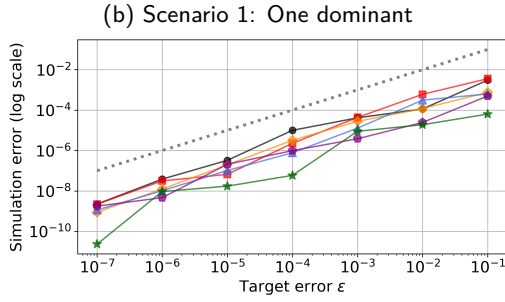
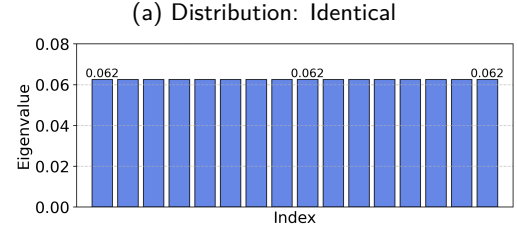
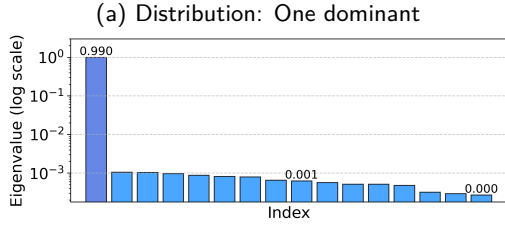


Figure 7: Simulation results for one dominant eigenvalue.

Figure 8: Simulation results for identical eigenvalues.

5 Applications in quantum information

Now, let's explore several use cases of how our *rank is all you need* and *effective rank is all you need* ideas can be efficiently applied to quantum information processing tasks.

5.1 Nonlinear function calculations for quantum states

Applying [Corollary 3.1](#) to [16, Theorem 5], we can enhance the theorem.

Theorem 5.1. *Let ρ be a quantum state with rank r . Suppose there exist $\epsilon > 0$ and a slowly-growing function C (as a function of k) such that $g : \mathbb{R} \rightarrow \mathbb{R}$ is approximated by a degree k polynomial*

$$f(x) = \sum_{i=0}^k c_i x^i \quad (5.1)$$

on the interval $[0, 1]$, in the sense that

$$\sup_{x \in [0, 1]} |g(x) - f(x)| < \frac{\epsilon}{2r}, \quad (5.2)$$

and

$$\sum_{i=0}^k |c_i| < C. \quad (5.3)$$

Then estimating $\text{Tr}(g(\rho))$ within an ϵ additive error and with a success probability of at least $1 - \delta$, where $\delta \in (0, 1)$ requires

$$\mathcal{O}\left(\frac{C^2 k^2 \tilde{r}^4 \ln^2 \tilde{r} \ln(1/\delta)}{\epsilon^2}\right) = \tilde{\mathcal{O}}\left(\frac{C^2 k^2 \ln(1/\delta)}{\epsilon^2}\right) \quad (5.4)$$

copies of ρ and

$$\mathcal{O}\left(\frac{C^2 k^2 \tilde{r}^3 \ln^2 \tilde{r} \ln(1/\delta)}{\epsilon^2}\right) = \tilde{\mathcal{O}}\left(\frac{C^2 k^2 \ln(1/\delta)}{\epsilon^2}\right) \quad (5.5)$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(\tilde{r})$ qubits and $\mathcal{O}(\tilde{r})$ CSWAP operations. Here, the notation $\tilde{\mathcal{O}}(\cdot)$ hides polylogarithmic factors in k and $1/\epsilon$.

In the original theorem mentioned in [16],

$$\mathcal{O}\left(\frac{C^2 k^2 \ln(1/\delta)}{\epsilon^2}\right) \quad (5.6)$$

copies of ρ were required, and the circuit consisted of $\mathcal{O}(k)$ qubits and $\mathcal{O}(k)$ CSWAP operations.

According to [Theorem 3.1](#) and [3.2](#), it suffices to estimate $\{\text{Tr}(\rho^i)\}_{i=1}^{\tilde{r}}$ with additive error at most $\epsilon/(2Ck\tilde{r}\ln\tilde{r})$ in order to approximate $\{\text{Tr}(\rho^i)\}_{i=1}^k$ up to an additive error of ϵ/C . Following the same reasoning as in [Eq. \(3.10\)](#) and [Eq. \(3.11\)](#), we can determine both the circuit repetition count and the required number of copies of ρ for this estimation.

By applying the result of [Corollary 3.1](#), the total number of copies of ρ required to achieve an ϵ -additive estimation with failure probability at most δ is given by

$$\mathcal{O}\left(\frac{C^2 k^2 \tilde{r}^4 \ln^2 \tilde{r} \ln(1/\delta)}{\epsilon^2}\right). \quad (5.7)$$

The corresponding algorithm can be implemented on a constant-depth quantum circuit that acts on $\mathcal{O}(\tilde{r})$ qubits and uses $\mathcal{O}(\tilde{r})$ CSWAP gates.

Typically, k is much larger than $\tilde{r}$, so our enhanced theorem offers advantages for estimating $g(\rho)$. When $g(x) = e^{\beta x}$, C becomes $e^{|\beta|}$. We can efficiently estimate $\text{Tr}(e^{\beta \rho})$ using [Theorem 5.1](#), which has applications in thermodynamics and the density exponentiation algorithm [\[41, 42, 16\]](#).

5.2 Quantum Gibbs state preparation

We highlight that our method improves the efficiency of preparing the quantum Gibbs state. The truncated Taylor series:

$$S_q(\rho) = \sum_{i=1}^q \text{Tr}((\rho - I)^i \rho) \quad (5.8)$$

is used as the cost function for variational quantum Gibbs state preparation [\[10\]](#). It is shown that the fidelity $F(\rho(\theta_0), \rho_G)$ between the optimized state $\rho(\theta_0)$ and the Gibbs state ρ_G is bounded by

$$F(\rho(\theta_0), \rho_G) \geq 1 - \sqrt{2 \left(\beta \epsilon + \frac{2r}{q+1} (1 - \Delta)^{q+1} \right)}, \quad (5.9)$$

where β is the inverse temperature of the system, and Δ is a constant that satisfies

$$-\Delta \ln(\Delta) < \frac{1}{q+1} (1 - \Delta)^{q+1}. \quad (5.10)$$

By using the inequality

$$D(\rho(\theta_0), \rho_G) < \sqrt{1 - F(\rho(\theta_0), \rho_G)}, \quad (5.11)$$

to achieve $D(\rho(\theta_0), \rho_G) < \epsilon$, we need to set $q = \mathcal{O}(r/\epsilon^4)$, where D is the trace distance. Using previous methods, $q = \mathcal{O}(r/\epsilon^4)$ qubits and CSWAP operations are required, which are impractical for near-term quantum devices. Our work significantly reduces the number of qubits and CSWAP operations to $\mathcal{O}(\tilde{r})$, exponentially reducing the quantum resources. This demonstrates that our method makes the preparation of the quantum Gibbs state using the truncated Taylor series much more feasible.

5.3 Efficient estimation of the trace of products of quantum state powers

In this section, we discuss the efficient estimation of the set

$$\{\text{Tr}(\rho^i \sigma^j) : (i, j) \in [k] \times [l]\}. \quad (5.12)$$

The core routine relies on [\[Algorithm 2\]](#) from [Section 3.3](#). Extending the problem from the general estimation of the trace of quantum state powers to this broader setting is crucial, as it enables the estimation of various distance measures, making this generalization highly significant.

Theorem 5.2. *For the problem of estimating the trace of products of quantum state powers, given large integers k, l , it is possible to approximate $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(k,l)}$ within*

an additive error of ϵ using quantum resources only up to $\{\text{Tr}(\rho^i)\}_{i=1}^t$, $\{\text{Tr}(\sigma^i)\}_{i=1}^t$, and $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(t,t)}$, where t satisfies

$$t \geq \tilde{R} = \min \left\{ r, \left\lceil \ln \left(\frac{4k + 4l}{\epsilon} \right) \right\rceil \right\}. \quad (5.13)$$

We define $\tilde{R}$ as the effective rank of the quantum states ρ and σ , where

$$r = \max\{\text{rank}(\rho), \text{rank}(\sigma)\}. \quad (5.14)$$

To clarify the notation, $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(a,b)}$ refers to the set of values obtained by calculating $\text{Tr}(\rho^i \sigma^j)$ for all $(i, j) \in [a] \times [b]$ where $[n]$ denotes the set of natural numbers from 1 to n .

Proof. See the details in [Appendix A.8](#). $\square$

To aid in understanding the proof of [Theorem 5.2](#), we can summarize the idea sketch of the proof in the following manner, corresponding to [\[Algorithm 3\]](#).

[Algorithm 3] Estimation of $\text{Tr}(\rho^k \sigma^l)$

1. For a fixed index i , we begin by estimating the sets $\{\text{Tr}(\sigma^i)\}_{i=1}^t$ and $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(t,t)}$, which enables the application of [\[Algorithm 2\]](#) with $M = \rho^i$. This allows us to compute the estimated values for $\{\text{Tr}(\rho^i \sigma^j)\}_{j=1}^l$ for the fixed i .
2. Repeating Step 1 for $i = 1, 2, \dots, t$, we obtain the values for $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(t,l)}$.
3. For a fixed index j , we then estimate the sets $\{\text{Tr}(\rho^i)\}_{i=1}^t$ and $\{\text{Tr}(\rho^i \sigma^j)\}_{i=1}^t$, allowing us to apply [\[Algorithm 2\]](#) once more, this time setting $M = \sigma^j$. This step computes the estimated values for $\{\text{Tr}(\rho^i \sigma^j)\}_{i=1}^k$.
4. With the values for $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(t,l)}$ already obtained in Step 2, the final process yields the estimated values for $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(k,l)}$.

The estimation of the values of $\{\text{Tr}(\rho^i)\}_{i=1}^t$, $\{\text{Tr}(\sigma^i)\}_{i=1}^t$, and $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=(1,1)}^{(t,t)}$ requires the same procedure as Step 2 of [\[Algorithm 2\]](#). Consequently, the quantum resources required for this process can be estimated as follows: for the estimation of $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=1}^{(k,l)}$, at most $\mathcal{O}(t)$ qubits and $\mathcal{O}(t)$ CSWAP gates are necessary. This improves upon the previous result in [\[16\]](#), which required $\mathcal{O}(k + l)$ qubits and $\mathcal{O}(k + l)$ CSWAP gates. Furthermore, in terms of copy complexity, each quantum state ρ and σ is required $\tilde{\mathcal{O}}(k^2/\epsilon^2)$ and $\tilde{\mathcal{O}}(l^2/\epsilon^2)$ times, respectively (details in [Appendix A.8](#)). Here, the notation $\tilde{\mathcal{O}}(\cdot)$ hides polylogarithmic factors in k and l . This improves upon previous studies, where ρ and σ were required $\tilde{\mathcal{O}}(k^2 l/\epsilon^2)$ and $\tilde{\mathcal{O}}(k l^2/\epsilon^2)$ times, respectively. An efficient algorithm for estimating $\{\text{Tr}(\rho^i \sigma^j)\}_{(i,j)=1}^{(k,l)}$ can be widely applied to various quantum information tasks. For example, it can be used to compute the Schatten- p distance, defined as

$$\|\rho - \sigma\|_p = (\text{Tr}(|\rho - \sigma|^p))^{\frac{1}{p}}. \quad (5.15)$$

It also applies to the estimation of other distance measures, such as

$$K_\alpha(\rho, \sigma) = \text{Tr}((1 + \rho)^\alpha (1 + \sigma)^{1-\alpha}), \quad (5.16)$$

which satisfies faithfulness and the data processing inequality under unital quantum channels [\[16\]](#).

5.4 Entanglement detection

Determining whether a quantum state is separable or entangled is a fundamental problem in quantum information theory. It is well known that a separable quantum state ρ_{AB} always has a positive semi-definite (PSD) partial transpose (PT), denoted as $\rho_{AB}^{\Gamma_B}$. By contraposition, if $\rho_{AB}^{\Gamma_B}$ has a negative eigenvalue, then ρ_{AB} must be entangled. For brevity, we denote the partial transpose of ρ as ρ^Γ . The k -th PT moment is defined as

$$p_k^{\text{PT}} = \text{Tr}((\rho^\Gamma)^k). \quad (5.17)$$

PT moments are typically estimated using classical shadows [30, 43]. By leveraging PT moments and the Newton-Girard method, the presence of a negative eigenvalue can be detected. The PT moments required for entanglement detection are $p_1^{\text{PT}}, p_2^{\text{PT}}, \dots, p_r^{\text{PT}}$, where r is the rank of ρ^Γ . Let $\lambda_1, \dots, \lambda_r$ be the eigenvalues of ρ^Γ . The following lemma, restating Lemma 1 of [30], formalizes this criterion:

Lemma 5.1. *A quantum state ρ is entangled if*

$$e_i(\lambda_1, \dots, \lambda_r) < 0 \quad (5.18)$$

for some $i = 1, 2, \dots, r$, where p_i^{PT} are the PT moments of ρ^Γ , and $e_i(x_1, \dots, x_m)$ denotes the elementary symmetric polynomial in m variables, defined as

$$e_i(x_1, \dots, x_m) = \sum_{1 \leq j_1 < j_2 < \dots < j_i \leq m} x_{j_1} x_{j_2} \cdots x_{j_i}, \quad (5.19)$$

which satisfies the recursive formula

$$e_k = \frac{1}{k} \sum_{i=1}^k (-1)^{i-1} e_{k-i} p_i^{\text{PT}}. \quad (5.20)$$

To integrate our approach, suppose that the eigenvalues of ρ^Γ are all non-negative. In this case, ρ^Γ is a valid density matrix, allowing us to apply [Algorithm 2]. Computing $p_1^{\text{PT}}, p_2^{\text{PT}}, \dots, p_t^{\text{PT}}$ is sufficient to estimate higher-order PT moments, where $t = \mathcal{O}(\ln(r/\epsilon))$. Using these PT moments and the recursive formula Eq. (5.20), we compute $e_i(\lambda_1, \dots, \lambda_r)$ for $i = 1, 2, \dots, r$. If the inequality Eq. (5.18) holds for some i , then ρ is entangled. Combining Lemma 5.1 with our method establishes a new entanglement detection criterion that requires only $p_1^{\text{PT}}, p_2^{\text{PT}}, \dots, p_t^{\text{PT}}$.

We hypothesize that in practical scenarios, the required number of PT moments is significantly smaller. Numerical simulations in Section 4 suggest that $t = 8$ is sufficient in most cases. If experimental validation confirms that $t = 8$ is also adequate for entanglement detection, this could constitute a groundbreaking discovery. We leave quantitative analysis and experimental verification as future work.

6 Concluding remarks

6.1 Summary of findings

In this paper, we present an efficient algorithm for estimating the trace of quantum state powers. Our first key observation is the discovery of the rank dependence in this problem. Specifically, we find that for a large integer k , estimating $\text{Tr}(\rho^k)$ within an additive error ϵ requires only the computation of $\{\text{Tr}(\rho^i)\}_{i=1}^r$ using quantum resources. The remaining

values can then be efficiently estimated within the same additive error ϵ by employing a simple recurrence relation based on the Newton-Girard method.

Our second key observation reveals a condition even stronger than rank dependence. In practical experimental settings, estimating the rank of a given quantum state is often non-trivial and can introduce additional overhead. To address this issue, we introduce the concept of the effective rank $\tilde{r}$ and rigorously prove that, for a target power k , our approach requires only quantum resources proportional to $\ln k$. This result significantly reduces the resource requirements and enhances the feasibility of trace estimation in realistic quantum experiments.

By leveraging the concepts of rank dependence and effective rank, we successfully extended our efficient estimation algorithm to tackle not only the problem of estimating traces of quantum state powers with arbitrary observables, $\text{Tr}(M\rho^k)$, but also the more general problem of estimating traces of products of quantum state powers, $\text{Tr}(\rho^k\sigma^l)$. Our main ideas were rigorously validated through formal mathematical proofs and numerical simulations. Furthermore, we demonstrated several practical applications of our algorithm in quantum information processing. Specifically, we showed that it enables more resource-efficient quantum estimation of nonlinear functionals of quantum states, quantum Gibbs state preparation, and entanglement detection compared to previously known methods. Moreover, we illustrated its applicability to various distance measures, further highlighting its broad utility.

6.2 Future research directions

In our study, several important directions for future investigation remain.

- (1) A tighter upper bound on $|\tilde{P}_k - P_k|$ in [Lemma 3.2](#) needs to be established. While [Section 4](#) discusses several observations suggesting the possibility of a tighter bound, a more rigorous mathematical formulation and an explicit analytical expression would be valuable.
- (2) A more detailed quantitative analysis of entanglement detection is necessary to identify specific aspects where our algorithm offers concrete improvements.
- (3) Perhaps most critically, our current work presents an efficient quantum algorithm for estimating $\text{Tr}(\rho^k)$ with additive error under multi-copy joint measurements. The algorithm sequentially applies CSWAP gates across registers to construct a fully entangled state over all quantum samples, thereby enabling coherent global operations necessary for accurate estimation. However, the exponential decay of $\text{Tr}(\rho^k)$ with increasing k suggests that multiplicative-error estimators are necessary in many applications. Despite this importance, our understanding of both additive and multiplicative error dependence remains limited across various measurement models. For the incoherent measurement setting, [Liu et al. \[44\]](#) implicitly establish a lower bound for additive-error estimation, but a corresponding upper bound remains unknown. For multiplicative-error estimation, the first nontrivial upper bound in the fixed-basis incoherent setting was proposed only recently [\[28\]](#), and no meaningful lower bound is currently available.
 - (a) While the fixed-basis incoherent and coherent measurement settings are now partially understood, relatively little is known about other regimes, including randomized and adaptive measurements, with tight bounds under either additive or multiplicative error still largely missing.

- (b) Furthermore, in practically motivated constrained models such as the few-copy measurement setting (where only a small number of copies can be measured jointly) or the bounded quantum memory setting, our understanding is even more limited. While recent work [45] has characterized the memory-sample tradeoff for Pauli shadow tomography in these restricted settings, extending such analyses to nonlinear estimation tasks remains largely open and would be a particularly exciting direction. Establishing tight bounds in these models continues to pose significant analytical challenges, but offers a rich and important avenue for future work.
- (4) It is necessary to investigate how our proposed algorithm can enhance virtual distillation. In the virtual distillation process, the expectation value of an observable M with respect to the state $\rho^k/\text{Tr}(\rho^k)$ must be computed. Our algorithm is expected to accelerate this computation. However, whether virtual distillation can still yield error-free expectation values under certain conditions when the ideal state $|\psi\rangle$ cannot be prepared and only a faulty state ρ is available requires a more detailed analysis. For our algorithm to be useful in this context, it must be assumed that even if $|\psi\rangle$ cannot be directly prepared, the quantum computer can still prepare multiple copies of ρ simultaneously and perform joint operations on them. However, in a faulty quantum computer, the process of preparing multiple copies and performing joint operations may introduce additional errors, which could be larger than those arising in much simpler single-copy operations. Given these considerations, it would be interesting to analyze how our algorithm influences virtual distillation while accounting for these potential error sources.
- (5) A natural open problem is to explore how our approach can be extended to general real values of k , rather than just integer k . While the work of [32] addresses algorithms for non-integer k , it would be interesting to develop an iterative variant in the spirit of our approach. This leads to several open questions, including the challenge of obtaining tighter quantitative bounds.
- (6) Beyond the specific problem settings addressed in this paper, it would be highly interesting to explore rank-dependent quantum algorithms applicable to broader areas of quantum information processing, particularly those that are especially efficient for low-rank quantum states.

Data availability statement

The data and software that support the findings of this study can be found in the following repository: <https://github.com/tfoseel/trace-of-powers>

Acknowledgments

J.L. thanks Chirag Wadhwa for helpful discussions and feedback regarding the importance of considering the trace of powers problem under various measurement settings. The authors also thank the anonymous reviewers for their valuable comments that helped improve the manuscript. This work was supported by the National Research Foundation of Korea (NRF) through a grant funded by the Ministry of Science and ICT (Grant Nos. RS-2023-00211817; RS-2024-00404854; RS-2025-00515537). This work was also supported by the Institute for Information & Communications Technology Promotion (IITP) grants funded

by the Korean government (MSIP) (Grant Nos. RS-2025-02304540; 2019-II190003, Research and Development of Core Technologies for Programming, Running, Implementing, and Validating of Fault-Tolerant Quantum Computing Systems), the National Research Council of Science & Technology (NST) (Grant No. GTL25011-401), and Korea Institute of Science and Technology Information (KISTI: P25026).

Author Contributions

M.S. and J.L. contributed equally to this work, undertaking the primary responsibilities, including the development of the main ideas, mathematical proofs, initial drafting, and revisions of the paper. S.L. contributed to the numerical simulations and paper preparation. K.J. supervised the research. All authors discussed the results and contributed to the final paper.

References

- [1] Sonika Johri, Damian S. Steiger, and Matthias Troyer. “Entanglement spectroscopy on a quantum computer”. *Physical Review B* **96**, 195136 (2017).
- [2] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller. “Rényi entropies from random quenches in atomic hubbard and spin models”. *Physical Review Letters* **120**, 050406 (2018).
- [3] B. Vermersch, A. Elben, M. Dalmonte, J. I. Cirac, and P. Zoller. “Unitary n -designs via random quenches in atomic hubbard and spin models: Application to the measurement of rényi entropies”. *Physical Review A* **97**, 023604 (2018).
- [4] Artur K. Ekert, Carolina Moura Alves, Daniel K. L. Oi, Michał Horodecki, Paweł Horodecki, and L. C. Kwék. “Direct estimations of linear and nonlinear functionals of a quantum state”. *Physical Review Letters* **88**, 217901 (2002).
- [5] Todd A. Brun. “Measuring polynomial functions of states”. *Quantum Information and Computation* **4**, 401 (2004).
- [6] S. J. van Enk and C. W. J. Beenakker. “Measuring $\text{Tr}\rho^n$ on single copies of ρ using random measurements”. *Physical Review Letters* **108**, 110503 (2012).
- [7] You Zhou and Zhenhuan Liu. “A hybrid framework for estimating nonlinear functions of quantum states”. *npj Quantum Information* **10**, 62 (2024).
- [8] Fabio Antonio Bovino, Giuseppe Castagnoli, Artur Ekert, Paweł Horodecki, Carolina Moura Alves, and Alexander Vladimir Sergienko. “Direct measurement of nonlinear properties of bipartite quantum states”. *Physical Review Letters* **95**, 240407 (2005).
- [9] Justin Yirka and Yiğit Subaşı. “Qubit-efficient entanglement spectroscopy using qubit resets”. *Quantum* **5**, 535 (2021).
- [10] Youle Wang, Guangxi Li, and Xin Wang. “Variational quantum gibbs state preparation with a truncated taylor series”. *Physical Review Applied* **16**, 054035 (2021).
- [11] Mirko Consiglio, Jacopo Settino, Andrea Giordano, Carlo Mastroianni, Francesco Plastina, Salvatore Lorenzo, Sabrina Maniscalco, John Goold, and Tony J. G. Apollaro. “Variational gibbs state preparation on noisy intermediate-scale quantum devices”. *Physical Review A* **110**, 012445 (2024).

- [12] Barbara M. Terhal and David P. DiVincenzo. “Problem of equilibration and the computation of correlation functions on a quantum computer”. *Physical Review A* **61**, 022301 (2000).
- [13] Arnau Riera, Christian Gogolin, and Jens Eisert. “Thermalization in nature and on a quantum computer”. *Physical Review Letters* **108**, 080402 (2012).
- [14] Jingxiang Wu and Timothy H. Hsieh. “Variational thermal quantum simulation via thermofield double states”. *Physical Review Letters* **123**, 220502 (2019).
- [15] Yiğit Subaşı, Lukasz Cincio, and Patrick J Coles. “Entanglement spectroscopy with a depth-two quantum circuit”. *Journal of Physics A: Mathematical and Theoretical* **52**, 044001 (2019).
- [16] Yihui Quek, Eneet Kaur, and Mark M. Wilde. “Multivariate trace estimation in constant quantum depth”. *Quantum* **8**, 1220 (2024).
- [17] Hsin-Yuan Huang, Richard Kueng, and John Preskill. “Predicting many properties of a quantum system from very few measurements”. *Nature Physics* **16**, 1050–1057 (2020).
- [18] Aniket Rath, Cyril Branciard, Anna Minguzzi, and Benoît Vermersch. “Quantum fisher information from randomized measurements”. *Physical Review Letters* **127**, 260501 (2021).
- [19] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. “Quantum fingerprinting”. *Physical Review Letters* **87**, 167902 (2001).
- [20] Daniel Gottesman and Isaac Chuang. “Quantum digital signatures” (2001). [arXiv:quant-ph/0105032](https://arxiv.org/abs/quant-ph/0105032).
- [21] M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. “Beyond the swap test: Optimal estimation of quantum state overlap”. *Physical Review Letters* **124**, 060503 (2020).
- [22] Steph Foulds, Viv Kendon, and Tim Spiller. “The controlled swap test for determining quantum entanglement”. *Quantum Science and Technology* **6**, 035002 (2021).
- [23] Xavier Gitiaux, Ian Morris, Maria Emelianenko, and Mingzhen Tian. “Swap test for an arbitrary number of quantum states”. *Quantum Information Processing* **21**, 344 (2022).
- [24] Michał Oszmaniec, Daniel J. Brod, and Ernesto F. Galvão. “Measuring relational information between quantum states, and applications”. *New Journal of Physics* **26**, 013053 (2024).
- [25] Tuan-Yow Chien and Shayne Waldron. “A characterization of projective unitary equivalence of finite frames and applications”. *SIAM Journal on Discrete Mathematics* **30**, 976–994 (2016).
- [26] V. Bargmann. “Note on wigner’s theorem on symmetry operations”. *Journal of Mathematical Physics* **5**, 862–868 (1964).
- [27] P. W. Shor. “Fault-tolerant quantum computation”. In *Proceedings of 37th Conference on Foundations of Computer Science*. Pages 56–65. (1996).
- [28] Angelos Pelecanos, Xinyu Tan, Ewin Tang, and John Wright. “Beating full state tomography for unentangled spectrum estimation” (2025). url: <https://arxiv.org/abs/2504.02785>.

- [29] Zachary P. Bradshaw, Margarite L. LaBorde, and Mark M. Wilde. “Cycle index polynomials and generalized quantum separability tests”. *Proceedings of the Royal Society A* **479**, 20220733 (2023).
- [30] Antoine Neven, Jose Carrasco, Vittorio Vitale, Christian Kokail, Andreas Elben, Marcello Dalmonte, Pasquale Calabrese, Peter Zoller, Benoît Vermersch, Richard Kueng, et al. “Symmetry-resolved entanglement detection using partial transpose moments”. *npj Quantum Information* **7**, 152 (2021).
- [31] Rafael Wagner, Zohar Schwartzman-Nowik, Ismael L Paiva, Amit Te’eni, Antonio Ruiz-Molero, Rui Soares Barbosa, Eliahu Cohen, and Ernesto F Galvão. “Quantum circuits for measuring weak values, kirkwood–dirac quasiprobability distributions, and state spectra”. *Quantum Science and Technology* **9**, 015030 (2024).
- [32] Yupan Liu and Qisheng Wang. “On estimating the trace of quantum state powers”. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Pages 947–993. SIAM (2025).
- [33] H. Francis Song, Stephan Rachel, Christian Flindt, Israel Klich, Nicolas Laflorencie, and Karyn Le Hur. “Bipartite fluctuations as a probe of many-body entanglement”. *Physical Review B* **85**, 035409 (2012).
- [34] Armen Bagdasaryan, Serkan Araci, Mehmet Açikgöz, and H. M. Srivastava. “Analogues of newton–girard power-sum formulas for entire and meromorphic functions with applications to the riemann zeta function”. *Journal of Number Theory* **147**, 92–102 (2015).
- [35] José Luis Cereceda. “Sums of powers of integers and stirling numbers”. *Resonance* **27**, 769–784 (2022).
- [36] Ronald G Mosier. “Root neighborhoods of a polynomial”. *Mathematics of Computation* **47**, 265–273 (1986).
- [37] Antonio D. Córcoles, Abhinav Kandala, Ali Javadi-Abhari, Douglas T. McClure, Andrew W. Cross, Kristan Temme, Paul D. Nation, Matthias Steffen, and Jay M. Gambetta. “Challenges and opportunities of near-term quantum computing systems”. *Proceedings of the IEEE* **108**, 1338–1352 (2019).
- [38] Konstantinos Georgopoulos, Clive Emary, and Paolo Zuliani. “Modeling and simulating the noisy behavior of near-term quantum computers”. *Physical Review A* **104**, 062432 (2021).
- [39] Jin-Min Liang, Qiao-Qiao Lv, Zhi-Xi Wang, and Shao-Ming Fei. “Unified multivariate trace estimation and quantum error mitigation”. *Physical Review A* **107**, 012606 (2023).
- [40] William J. Huggins, Sam McArdle, Thomas E. O’Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. “Virtual distillation for quantum error mitigation”. *Physical Review X* **11**, 041036 (2021).
- [41] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. “Quantum principal component analysis”. *Nature Physics* **10**, 631–633 (2014).
- [42] Shelby Kimmel, Cedric Yen-Yu Lin, Guang Hao Low, Maris Ozols, and Theodore J. Yoder. “Hamiltonian simulation with optimal sample complexity”. *npj Quantum Information* **3**, 13 (2017).

- [43] Andreas Elben, Richard Kueng, Hsin-Yuan Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, et al. “Mixed-state entanglement from local randomized measurements”. *Physical Review Letters* **125**, 200501 (2020).
- [44] Zhenhuan Liu, Weiyuan Gong, Zhenyu Du, and Zhenyu Cai. “Exponential separations between quantum learning with and without purification” (2024). [arXiv:2410.17718](#).
- [45] Sitan Chen, Weiyuan Gong, and Qi Ye. “Optimal tradeoffs for estimating pauli observables”. In 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS). Pages 1086–1105. (2024).

A Omitted proofs

A.1 Proof of Lemma 3.1

Proof. Let's try to find some properties of $|d_i|$.

- If $i = 1$,

$$d_1 = b_1 - a_1 = Q_1 - P_1 = \epsilon_1, \quad (\text{A.1})$$

this gives $|d_1| \leq |\epsilon_1|$.

- And if $i = 2$,

$$\begin{aligned} d_2 &= b_2 - a_2 \\ &= \frac{Q_1^2 - Q_2}{2} - \frac{P_1^2 - P_2}{2} = \frac{(Q_1 + P_1)(Q_1 - P_1) - (Q_2 - P_2)}{2} = \epsilon_1 - \frac{1}{2}\epsilon_2, \end{aligned} \quad (\text{A.2})$$

this gives $|d_2| \leq |\epsilon_1| + \frac{|\epsilon_2|}{2}$.

- And if $i = 3$,

$$d_3 = \frac{(b_2 - a_2)Q_1 - (b_1 - a_1)Q_2 + (Q_3 - P_3) - a_1(Q_2 - P_2) + a_2(Q_1 - P_1)}{3}. \quad (\text{A.3})$$

this gives,

$$|d_3| \leq \frac{|d_2|P_1 + |d_1|P_2 + |\epsilon_3| + a_1|\epsilon_2| + a_2|\epsilon_1|}{3} \leq |\epsilon_1| + \frac{|\epsilon_2|}{2} + \frac{|\epsilon_3|}{3}. \quad (\text{A.4})$$

Now, we suppose $|d_k| \leq \sum_{i=1}^k \frac{|\epsilon_i|}{i}$. Then,

$$\begin{aligned} d_{k+1} &= \frac{1}{k+1} \left\{ \sum_{i=1}^{k+1} (-1)^{i-1} b_{k+1-i} Q_i - \sum_{i=1}^{k+1} (-1)^{i-1} a_{k+1-i} P_i \right\} \\ &= \frac{1}{k+1} \left\{ \sum_{i=1}^{k+1} (-1)^{i-1} (b_{k+1-i} - a_{k+1-i}) Q_i + \sum_{i=1}^{k+1} (-1)^{i-1} a_{k+1-i} (Q_i - P_i) \right\}. \end{aligned} \quad (\text{A.5})$$

Taking the absolute value, and by using $|Q_i| \leq 1$ and $|a_{k+1-i}| \leq 1$,

$$\begin{aligned} |d_{k+1}| &\leq \frac{1}{k+1} \left(\sum_{i=1}^k |b_{k+1-i} - a_{k+1-i}| + \sum_{i=1}^{k+1} |Q_i - P_i| \right) \\ &\leq \frac{1}{k+1} \left(\sum_{i=1}^k \sum_{j=1}^{k+1-i} \frac{|\epsilon_j|}{j} + \sum_{i=1}^{k+1} |\epsilon_i| \right) \\ &= \frac{1}{k+1} \sum_{j=1}^{k+1} \left(\frac{k+1-j}{j} + 1 \right) |\epsilon_j| \\ &= \sum_{j=1}^{k+1} \frac{|\epsilon_j|}{j}. \end{aligned} \quad (\text{A.6})$$

So, we can conclude that

$$\forall k \in \mathbb{N}, |d_k| \leq \sum_{i=1}^k \frac{|\epsilon_i|}{i} \quad (\text{A.7})$$

by strong mathematical induction logic. $\square$

A.2 Proof of Theorem 3.1

Proof. We assume

$$|\epsilon_i| < \frac{\epsilon}{kt \ln t} \quad (\text{A.8})$$

holds for $i = 1, 2, \dots, t$. We first prove that

$$|Q_i - \tilde{P}_i| < \epsilon \quad (\text{A.9})$$

for $i = 1, 2, \dots, k$ always holds. Consider the recurrence relations defined by the Eq. (2.12), and Eq. (3.4). Then the difference between $\tilde{P}_{t+k}, Q_{t+k}$ becomes:

$$Q_{t+k} - \tilde{P}_{t+k} = \sum_{j=1}^t (-1)^{j-1} a_j (Q_{t+k-j} - \tilde{P}_{t+k-j}) + \sum_{j=1}^t (-1)^{j-1} (b_j - a_j) Q_{t+k-j}. \quad (\text{A.10})$$

Let $\tilde{\epsilon}_i = Q_i - \tilde{P}_i$ for all i . Then,

$$\tilde{\epsilon}_{t+k} = \sum_{j=1}^t \left\{ (-1)^{j-1} (a_j \tilde{\epsilon}_{t+k-j} + d_j Q_{t+k-j}) \right\}, \quad (\text{A.11})$$

$$\tilde{\epsilon}_{t+k-1} = \sum_{j=1}^t \left\{ (-1)^{j-1} (a_j \tilde{\epsilon}_{t+k-j-1} + d_j Q_{t+k-j-1}) \right\}. \quad (\text{A.12})$$

By exploiting $a_1 = 1$, we can sum up the above expressions in the form of

$$\begin{aligned} \tilde{\epsilon}_{t+k} &= \sum_{j=2}^t (-1)^{j-1} a_j \tilde{\epsilon}_{t+k-j} + \sum_{j=1}^t (-1)^{j-1} a_j \tilde{\epsilon}_{t+k-j-1} + \sum_{j=1}^t (-1)^{j-1} d_j (Q_{t+k-j} + Q_{t+k-j-1}) \\ &= \sum_{j=1}^{t-1} (-1)^{j-1} (a_j - a_{j+1}) \tilde{\epsilon}_{t+k-j-1} + (-1)^{t-1} a_t \tilde{\epsilon}_k + \sum_{j=1}^t (-1)^{j-1} d_j (Q_{t+k-j} + Q_{t+k-j-1}). \end{aligned} \quad (\text{A.13})$$

Since $a_j \geq a_{j+1}$ (trivial from Section 2.1),

$$|\tilde{\epsilon}_{t+k}| \leq \sum_{j=1}^{t-1} (a_j - a_{j+1}) |\tilde{\epsilon}_{t+k-j-1}| + a_t |\tilde{\epsilon}_k| + \sum_{j=1}^t |d_j| (Q_{t+k-j} + Q_{t+k-j-1}) \quad (\text{A.14})$$

$$\leq \sum_{j=1}^{t-1} (a_j - a_{j+1}) |\tilde{\epsilon}_{t+k-j-1}| + a_t |\tilde{\epsilon}_k| + 2 \sum_{j=1}^t |d_j|. \quad (\text{A.15})$$

Note that, from Eq. (3.3), $P_i = \tilde{P}_i$ for $i = 1, 2, \dots, t$. Therefore, $\epsilon_i = \tilde{\epsilon}_i$ for $i = 1, 2, \dots, t$. Let $\epsilon' := \max_{1 \leq j \leq t} |\epsilon_j|$. Suppose that

$$\tilde{\epsilon}_{t+m} \leq \epsilon' + m \sum_{j=1}^t |d_j| \quad (\text{A.16})$$

holds for $m = 1, 2, \dots, k-1$. Then,

$$\begin{aligned} |\tilde{\epsilon}_{t+k}| &\leq \sum_{j=1}^{t-1} (a_j - a_{j+1}) \left\{ \epsilon' + (k-2) \sum_{j=1}^t |d_j| \right\} + a_t \left\{ \epsilon' + (k-2) \sum_{j=1}^t |d_j| \right\} + 2 \sum_{j=1}^t |d_j| \\ &\leq a_1 \left\{ \epsilon' + (k-2) \sum_{j=1}^t |d_j| \right\} + 2 \sum_{j=1}^t |d_j| \\ &= \epsilon' + k \sum_{j=1}^t |d_j|. \end{aligned} \quad (\text{A.17})$$

Moreover, $m = k$ also holds. Since $m = 0$ trivially holds, by strong mathematical induction logic, for every m , [Eq. \(A.16\)](#) holds. By applying [Lemma 3.1](#), we get:

$$|d_j| \leq \sum_{i=1}^j \frac{\epsilon'}{i} \leq \epsilon' \ln j. \quad (\text{A.18})$$

Finally,

$$|\tilde{\epsilon}_{t+k}| \leq \epsilon' + \epsilon' k \ln t! < \epsilon' k t \ln t < \epsilon. \quad (\text{A.19})$$

We proved $|Q_i - \tilde{P}_i| < \epsilon$ for $i = 1, 2, \dots, k$. To conclude the proof, we set $t = r$. We will show that if $t = r$, $P_i = \tilde{P}_i$ holds for all i . Consider the following polynomial,

$$x^r - a_1 x^{r-1} + a_2 x^{r-2} - \dots + (-1)^r a_r = (x - p_1)(x - p_2) \dots (x - p_r). \quad (\text{A.20})$$

Then,

$$p_i^{r+k} = \sum_{j=1}^r (-1)^{j-1} a_j p_i^{r+k-j}, \quad (\text{A.21})$$

And we have,

$$P_{r+k} = \sum_{j=1}^r (-1)^{j-1} a_j P_{r+k-j}, \quad (\text{A.22})$$

which is the same recurrence relation with [Eq. \(3.4\)](#), when $t = r$. Hence, $P_i = \tilde{P}_i$ and

$$|\epsilon_i| = |Q_i - P_i| = |Q_i - \tilde{P}_i| < \epsilon \quad (\text{A.23})$$

for $i = 1, 2, \dots, r$ which completes the proof. $\square$

A.3 Proof of [Corollary 3.1](#)

Proof. Using the multivariate trace estimation method [\[16\]](#), it is known that with

$$\mathcal{O}\left(\frac{\ln(1/\delta)}{\epsilon^2}\right) \quad (\text{A.24})$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(i)$ qubits and $\mathcal{O}(i)$ CSWAP operations, we can estimate each $\text{Tr}(\rho^i)$ within an ϵ additive error and with a success probability of no less than $1 - \delta$. Note that only the maximum error

$$\epsilon' := \max_{1 \leq j \leq t} |\epsilon_j| \quad (\text{A.25})$$

affects the error of our algorithm. Thus, with

$$\mathcal{O}\left(\frac{k^2 t^2 \ln^2 t \ln(1/\delta)}{\epsilon^2}\right) \quad (\text{A.26})$$

runs, we can satisfy the assumption in [Theorem 3.1](#). Hence, $\text{Tr}(\rho^i)$ ($\forall i \leq k$) can be estimated within an ϵ error and with a success probability of no less than $1 - \delta$. Finally, we set $t = r$ in [Theorem 3.1](#), which concludes the proof. $\square$

A.4 Proof of Lemma 3.2

Proof. Let $\epsilon_i = \tilde{P}_i - P_i$. By definition, we have $\epsilon_i = 0$ for $i = 1, 2, \dots, t$. For $i \geq t+1$, we have

$$\epsilon_{t+k} = \tilde{P}_{t+k} - P_{t+k}. \quad (\text{A.27})$$

From the definition of $\tilde{P}_{t+k}$, we have:

$$\epsilon_{t+k} = \left\{ \sum_{i=1}^t (-1)^{i-1} \tilde{P}_{t+k-i} a_i \right\} - P_{t+k}. \quad (\text{A.28})$$

Rewriting this, we split ϵ_{t+k} into two terms:

$$\epsilon_{t+k} = \left\{ \sum_{i=1}^t (-1)^{i-1} \epsilon_{t+k-i} a_i \right\} + \left\{ \sum_{i=1}^t (-1)^{i-1} P_{t+k-i} a_i - P_{t+k} \right\}. \quad (\text{A.29})$$

Define the second term as z_{t+k} for brevity:

$$z_{t+k} = \left\{ \sum_{i=1}^t (-1)^{i-1} P_{t+k-i} a_i - P_{t+k} \right\} \quad (\text{A.30})$$

$$= (-1)^{t-1} \sum_{\{\alpha_1, \dots, \alpha_{t+1}\} \subseteq [r]} \left(\sum_{i=1}^{t+1} p_{\alpha_i}^k \right) \prod_{i=1}^{t+1} p_{\alpha_i}. \quad (\text{A.31})$$

Thus,

$$\epsilon_{t+k} = \sum_{i=1}^t (-1)^{i-1} \epsilon_{t+k-i} a_i + z_{t+k}. \quad (\text{A.32})$$

We first bound z_{t+k} :

$$|z_{t+k}| \leq (t+1) \sum_{\{\alpha_1, \dots, \alpha_{t+1}\} \subseteq [r]} \prod_{i=1}^{t+1} p_{\alpha_i} = (t+1) a_{t+1}. \quad (\text{A.33})$$

Next, consider the recursive relation for ϵ_{t+k} :

$$\epsilon_{t+k} = \sum_{i=1}^t (-1)^{i-1} \epsilon_{t+k-i} a_i + z_{t+k}. \quad (\text{A.34})$$

Combining this with the relation for ϵ_{t+k-1} , we get:

$$\epsilon_{t+k} = \left\{ \sum_{i=1}^{t-1} (-1)^{i-1} \epsilon_{t+k-i-1} (a_i - a_{i+1}) \right\} + \epsilon_{k-1} a_t + z_{t+k} + z_{t+k-1}. \quad (\text{A.35})$$

Taking the absolute value, we bound $|\epsilon_{t+k}|$:

$$\begin{aligned} |\epsilon_{t+k}| &\leq \left\{ \sum_{i=1}^{t-1} |\epsilon_{t+k-i-1}| (a_i - a_{i+1}) \right\} + |\epsilon_{k-1}| a_t + 2(t+1) a_{t+1} \\ &\leq \epsilon_{\max} + 2(t+1) a_{t+1}, \end{aligned} \quad (\text{A.36})$$

where $\epsilon_{\max}$ defined as:

$$\epsilon_{\max} = \max_{i \leq t+k-2} |\epsilon_i|. \quad (\text{A.37})$$

By induction, we conclude:

$$|\epsilon_{t+k}| \leq k(t+1)a_{t+1}. \quad (\text{A.38})$$

Also, we can bound a_{t+1} as follows (see the details in [Appendix A.4.1](#)):

$$a_{t+1} \leq \binom{r}{t+1} \frac{1}{r^{t+1}} = \frac{r(r-1)\dots(r-t-1)}{(t+1)!r^{t+1}} \leq \frac{1}{(t+1)!} \left(1 - \frac{t}{r}\right). \quad (\text{A.39})$$

Combining the results, we have:

$$|\epsilon_k| = |P_k - \tilde{P}_k| \leq (k-t)(t+1)a_{t+1} \leq k(t+1)a_{t+1}. \quad (\text{A.40})$$

Substituting the bound for a_{t+1} , we get:

$$|\epsilon_k| \leq \frac{k}{t!} \left(1 - \frac{t}{r}\right). \quad (\text{A.41})$$

□

A.4.1 Bounding a_t

Define $A_{j,k}$ as:

$$A_{j,k} = \sum_{\{\alpha_1, \dots, \alpha_t\} \subseteq [k]} \left(\prod_{i=1}^j p_{\alpha_i} \right), \quad (\text{A.42})$$

where $p_i \geq 0$ and $\sum_{i=1}^k p_i = 1$. By definition, we have $a_t = A_{t,r}$, where r denotes the rank. Next, let $x = p_1$ and define

$$p'_i = \frac{p_{i+1}}{1-x}, \quad \text{for } i = 1, 2, \dots, k-1.$$

Note that $\sum_{i=1}^{k-1} p'_i = 1$. Define $A'_{j,k}$ as:

$$A'_{j,k} = \sum_{\{\alpha_1, \dots, \alpha_t\} \subseteq [k]} \left(\prod_{i=1}^j p'_{\alpha_i} \right). \quad (\text{A.43})$$

Importantly, x and $A'_{j,k}$ are independent. For $(j, k) < (t, r)$, suppose that $A_{j,k}$ is maximized when

$$p_1 = p_2 = \dots = p_k = \frac{1}{k}. \quad (\text{A.44})$$

We then obtain the recurrence relation:

$$A_{t,r} = x(1-x)^{t-1} A'_{t-1,r-1} + (1-x)^t A'_{t,r-1}. \quad (\text{A.45})$$

Since it is straightforward to verify that

$$\max_{p'_i} A'_{j,k} = \max_{p_i} A_{j,k}, \quad (\text{A.46})$$

it follows that (by assumption) $A'_{t-1,r-1}$ and $A'_{t,r-1}$ are maximized when

$$p'_1 = p'_2 = \cdots = p'_{r-1} = \frac{1}{r-1}. \quad (\text{A.47})$$

Thus, we obtain

$$\max A'_{t-1,r-1} = \binom{r-1}{t-1} \frac{1}{(r-1)^{t-1}}, \quad (\text{A.48})$$

$$\max A'_{t,r-1} = \binom{r-1}{t} \frac{1}{(r-1)^t}. \quad (\text{A.49})$$

Now, considering the maximization over x and p'_i , we derive:

$$\begin{aligned} \max_{p_i} A_{t,r} &= \max_{x, p'_i} A_{t,r} \\ &= \max_x \left\{ x(1-x)^{t-1} \max_{p'_i} A'_{t-1,r-1} + (1-x)^t \max_{p'_i} A'_{t,r-1} \right\} \\ &= \max_x \left\{ x(1-x)^{t-1} \binom{r-1}{t-1} \frac{1}{(r-1)^{t-1}} + (1-x)^t \binom{r-1}{t} \frac{1}{(r-1)^t} \right\} \\ &= \frac{1}{r(r-1)^t} \binom{r}{t} \times \max_x \left\{ (1-x)^{t-1} (r - rx - t + rtx) \right\}. \end{aligned}$$

Define

$$f(x) = (1-x)^{t-1} (r - rx - t + rtx). \quad (\text{A.50})$$

Differentiating $f(x)$ and solving for its maximum, we find that the optimal value occurs at $x = 1/r$. This implies that $A_{t,r}$ is maximized when

$$p_1 = x = \frac{1}{r}, \quad (\text{A.51})$$

$$p_2 = p_3 = \cdots = p_r = \frac{1}{r-1} (1-x) = \frac{1}{r}. \quad (\text{A.52})$$

By strong induction, we conclude that for all t, r ,

$$a_t \leq \max A_{t,r} = \binom{r}{t} \frac{1}{r^t}. \quad (\text{A.53})$$

A.5 Proof of Theorem 3.2

Proof. We adopt the same notation as in [Appendix A.4](#). Using the proof from [Appendix A.2](#), we conclude that if

$$\epsilon' = \frac{\epsilon}{2kt \ln t}, \quad (\text{A.54})$$

then the following condition holds:

$$|\tilde{P}_i - Q_i| < \frac{\epsilon}{2}. \quad (\text{A.55})$$

To estimate P_i using Q_i with an additive error ϵ , we must ensure that:

$$|\tilde{P}_i - P_i| < \frac{\epsilon}{2}. \quad (\text{A.56})$$

This ensures:

$$|P_i - Q_i| \leq |\tilde{P}_i - Q_i| + |\tilde{P}_i - P_i| < \epsilon, \quad (\text{A.57})$$

for all $i = 1, 2, \dots, k$. From [Lemma 3.2](#), we derive the following bound:

$$|\tilde{P}_i - P_i| = |\epsilon_i| \leq \frac{i}{t!} \left(1 - \frac{t}{r}\right) \leq \frac{k}{e^t}. \quad (\text{A.58})$$

To ensure $|\tilde{P}_i - P_i| < \epsilon/2$, it suffices to satisfy:

$$\frac{k}{e^t} < \frac{\epsilon}{2}. \quad (\text{A.59})$$

Taking logarithms and rearranging terms, we obtain:

$$t > \ln \left(\frac{2k}{\epsilon} \right) \geq \left\lceil \ln \left(\frac{2k}{\epsilon} \right) \right\rceil. \quad (\text{A.60})$$

□

A.6 Proof of [Theorem 3.3](#)

Proof. Let

$$\rho = \sum_{i=1}^r p_i |\psi_i\rangle \langle \psi_i|, \quad (\text{A.61})$$

and

$$m_i = \langle \psi_i | M | \psi_i \rangle. \quad (\text{A.62})$$

We introduce a new quantity, denoted as $\tilde{P}_{i,M}$, defined for $i \leq t$ as follows:

$$\tilde{P}_{i(\leq t),M} := \text{Tr}(M \rho^i) = \sum_{j=1}^r m_j p_j^i. \quad (\text{A.63})$$

For $i > t$, $\tilde{P}_{i,M}$ is recursively defined based on the Newton-Girard recurrence relations, where the elementary symmetric polynomials a_k are defined in [Eq. \(2.4\)](#).

$$\tilde{P}_{i(>t),M} := \sum_{k=1}^t (-1)^{k-1} a_k \tilde{P}_{i-k,M}. \quad (\text{A.64})$$

We assume that

$$|\epsilon_{i,M}| < \frac{\epsilon}{4}, \quad (\text{A.65})$$

and

$$|\epsilon_i| < \frac{\epsilon}{4 \|M\|_\infty k t \ln t}, \quad (\text{A.66})$$

for $i = 1, 2, \dots, t$. We will first prove that

$$|Q_{i,M} - \tilde{P}_{i,M}| < \frac{\epsilon}{2} \quad (\text{A.67})$$

holds for $i = 1, 2, \dots, k$. The difference between $\tilde{P}_{i,M}$ and $Q_{i,M}$ is given by:

$$Q_{t+k,M} - \tilde{P}_{t+k,M} = \sum_{i=1}^t (-1)^{j-1} a_j \left(Q_{t+k-j,M} - \tilde{P}_{t+k-j,M} \right) + \sum_{i=1}^t (-1)^{j-1} (b_j - a_j) Q_{t+k-j,M}. \quad (\text{A.68})$$

Let $\tilde{\epsilon}_{i,M} := Q_{i,M} - \tilde{P}_{i,M}$, so that we can write:

$$\tilde{\epsilon}_{t+k,M} = \sum_{j=1}^t \left\{ (-1)^{j-1} (a_j \tilde{\epsilon}_{t+k-j,M} + d_j Q_{t+k-j,M}) \right\}, \quad (\text{A.69})$$

$$\tilde{\epsilon}_{t+k-1,M} = \sum_{j=1}^t \left\{ (-1)^{j-1} (a_j \tilde{\epsilon}_{t+k-j-1,M} + d_j Q_{t+k-j-1,M}) \right\}. \quad (\text{A.70})$$

We define

$$\epsilon' = \max_{1 \leq j \leq t} |\epsilon_j|, \quad \epsilon'_M = \max_{1 \leq j \leq t} |\epsilon_{j,M}|. \quad (\text{A.71})$$

and by assumption, we have

$$\epsilon' < \frac{\epsilon}{4 \|M\|_\infty k t \ln t} \quad (\text{A.72})$$

and $\epsilon'_M < \epsilon/4$. Using the same logic as in the proof of [Theorem 3.1](#), and noting that $Q_{i,M} \leq \|M\|_\infty$, we conclude that for every k

$$|\tilde{\epsilon}_{t+k,M}| \leq \epsilon'_M + k \|M\|_\infty \sum_{j=1}^t |d_j| \quad (\text{A.73})$$

holds. By applying [Lemma 3.1](#), we can conclude that

$$|\tilde{\epsilon}_{t+k,M}| \leq \epsilon'_M + \epsilon' k t \ln t \|M\|_\infty < \frac{\epsilon}{2}. \quad (\text{A.74})$$

Therefore, $|Q_{i,M} - \tilde{P}_{i,M}| < \epsilon/2$ holds for $i = 1, 2, \dots, k$.

Next, we aim to prove that $|\tilde{P}_{i,M} - P_{i,M}| < \epsilon/2$. Note that

$$t = \ln \left(\frac{2k \|M\|_\infty}{\epsilon} \right). \quad (\text{A.75})$$

Let $\delta_i = \tilde{P}_{i,M} - P_{i,M}$. By definition, $\delta_i = 0$ for $i = 1, 2, \dots, t$. For $i \geq t+1$, we derive δ_{t+k} as follows:

$$\delta_{t+k} = \tilde{P}_{t+k,M} - P_{t+k,M}. \quad (\text{A.76})$$

Using the definition of $\tilde{P}_{t+k,M}$, we get

$$\delta_{t+k} = \left\{ \sum_{i=1}^t (-1)^{i-1} \tilde{P}_{t+k-i,M} a_i \right\} - P_{t+k,M}. \quad (\text{A.77})$$

Rewriting this expression, we split δ_{t+k} into two terms:

$$\delta_{t+k} = \left\{ \sum_{i=1}^t (-1)^{i-1} \delta_{t+k-i} a_i \right\} + \left\{ \sum_{i=1}^t (-1)^{i-1} P_{t+k-i,M} a_i - P_{t+k,M} \right\}. \quad (\text{A.78})$$

We define the second term as z_{t+k} for brevity:

$$z_{t+k} = \left\{ \sum_{i=1}^t (-1)^{i-1} P_{t+k-i,M} a_i - P_{t+k,M} \right\} \quad (\text{A.79})$$

$$= (-1)^{t-1} \sum_{\{\alpha_1, \dots, \alpha_{t+1}\} \subseteq [r]} \left(\sum_{i=1}^{t+1} m_{\alpha_i} p_{\alpha_i}^k \right) \prod_{i=1}^{t+1} p_{\alpha_i}. \quad (\text{A.80})$$

Thus, we have

$$\delta_{t+k} = \sum_{i=1}^t (-1)^{i-1} \delta_{t+k-i} a_i + z_{t+k}. \quad (\text{A.81})$$

We first bound z_{t+k} :

$$\begin{aligned} |z_{t+k}| &\leq (t+1) \|M\|_\infty \sum_{\{\alpha_1, \dots, \alpha_{t+1}\} \subseteq [r]} \prod_{i=1}^{t+1} p_{\alpha_i} \\ &= (t+1) \|M\|_\infty a_{t+1}. \end{aligned} \quad (\text{A.82})$$

Using the same induction logic as in the proof of [Lemma 3.2](#), we get:

$$|\delta_i| \leq i(t+1) \|M\|_\infty a_{t+1}. \quad (\text{A.83})$$

Since

$$a_{t+1} \leq \frac{1}{(t+1)!}, \quad t! \geq e^t, \quad t = \ln \left(\frac{2k \|M\|_\infty}{\epsilon} \right), \quad (\text{A.84})$$

we conclude that

$$|P_{i,M} - \tilde{P}_{i,M}| = |\delta_i| \leq \frac{i \|M\|_\infty}{2^t} < \frac{\epsilon}{2} \quad (\text{A.85})$$

holds for $i = 1, 2, \dots, k$.

Thus, $|Q_{i,M} - P_{i,M}| < \epsilon$ holds for $i = 1, 2, \dots, k$. $\square$

A.7 Proof of [Corollary 3.3](#)

Proof. Using the multivariate trace estimation method [\[16\]](#), it is known that with

$$\mathcal{O} \left(\frac{\ln(1/\delta)}{\epsilon^2} \right) \quad (\text{A.86})$$

runs on a constant-depth quantum circuit consisting of $\mathcal{O}(i)$ qubits and $\mathcal{O}(i)$ CSWAP operations, we can estimate each $\text{Tr}(\rho^i)$ within ϵ additive error and with success probability not smaller than $1 - \delta$.

Note that only the maximum error

$$\epsilon' := \max_{1 \leq j \leq t} |\epsilon_j|, \quad (\text{A.87})$$

and

$$\epsilon'_M := \max_{1 \leq j \leq t} |\epsilon_{j,M}|, \quad (\text{A.88})$$

affects the error of our algorithm. So with

$$\mathcal{O} \left(\frac{k^2 \|M\|_\infty t^2 \ln^2 t \ln(1/\delta)}{\epsilon^2} \right) \quad (\text{A.89})$$

runs for estimating $\text{Tr}(\rho^{j'})$ ($j' \leq r$), and

$$\mathcal{O} \left(\frac{c^2 N_M \ln(1/\delta)}{\epsilon^2} \right) \quad (\text{A.90})$$

runs for estimating $\text{Tr}(M\rho^j)$ ($j \leq r$), we can satisfy the assumption in [Theorem 3.3](#) with success probability not smaller than $1 - \delta$. Hence, $\text{Tr}(M\rho^i)$ ($\forall i \leq k$) can be estimated within ϵ error and with success probability not smaller than $1 - \delta$. Finally, we set $t = \tilde{r}_M$ in [Theorem 3.3](#), which concludes the proof by ignoring the logarithmic terms. $\square$

A.8 Proof of Theorem 5.2

Proof. For a fixed index $i \leq t$, we begin by estimating the sets $\{\text{Tr}(\sigma^j)\}_{j=1}^t$ and $\{\text{Tr}(\rho^i \sigma^j)\}_{j=1}^t$, which enables the application of [Algorithm 2] with $M = \rho^i$. By Theorem 3.3, we obtain:

$$\{\text{Tr}(\sigma^j)\}_{j=1}^t \text{ within additive error } \frac{\epsilon}{8lt \ln t}, \quad (\text{A.91})$$

and

$$\{\text{Tr}(\rho^i \sigma^j)\}_{j=1}^t \text{ within additive error } \frac{\epsilon}{8}. \quad (\text{A.92})$$

We then set

$$t \geq \tilde{R} = \min \left\{ r, \left\lceil \ln \left(\frac{4k + 4l}{\epsilon} \right) \right\rceil \right\} \quad (\text{A.93})$$

which allows us to compute the estimated values for

$$\{\text{Tr}(\rho^i \sigma^j)\}_{j=1}^l \text{ within additive error } \frac{\epsilon}{2}, \quad (\text{A.94})$$

for $i = 1, 2, \dots, t$. For a fixed index j , obtaining

$$\{\text{Tr}(\rho^i)\}_{i=1}^t \text{ within additive error } \frac{\epsilon}{4kt \ln t}, \quad (\text{A.95})$$

and using Eq. (A.94) enables the application of [Algorithm 2] with $M = \sigma^j$. This, in turn, allows the computation of the estimated values for

$$\{\text{Tr}(\rho^i \sigma^j)\}_{i=1}^k \text{ within additive error } \epsilon, \quad (\text{A.96})$$

for $j = 1, 2, \dots, l$. To satisfy Eq. (A.91), we need $\tilde{O}(l^2/\epsilon^2)$ copies of σ , and to satisfy Eq. (A.95), we need $\tilde{O}(k^2/\epsilon^2)$ copies of ρ . The contributions to the copy complexity from other conditions, such as Eq. (A.92) and Eq. (A.94), are analogous. Therefore, to follow the algorithm, we can conclude that the required number of copies of ρ is $\tilde{O}(k^2/\epsilon^2)$ and the required number of copies of σ is $\tilde{O}(l^2/\epsilon^2)$. Here, the notation $\tilde{O}(\cdot)$ hides polylogarithmic factors in k and l . So, setting $t \geq \tilde{R}$ is sufficient. $\square$

B Additional numerical simulations

In Section 4.2, we anticipated that a lower bound on t could be expressed as

$$\mathcal{O} \left(\frac{\ln(k/\epsilon)}{\ln \ln(k/\epsilon)} \right). \quad (\text{B.1})$$

While a rigorous mathematical proof remains an open problem for future research, we conducted experiments under Scenario 1, as described in Section 4.1, by setting

$$t = \min \left\{ r, \left\lceil \frac{\ln(k/\epsilon)}{\ln \ln(k/\epsilon)} \right\rceil \right\} \quad (\text{B.2})$$

and evaluating four different distributions. (The values of t for different (k, ϵ) are listed in Table 4.) The results, presented in Fig. 9, show that although the error is larger compared to when $\tilde{r}$ was used, the estimation still successfully remains below the target additive error in all cases.

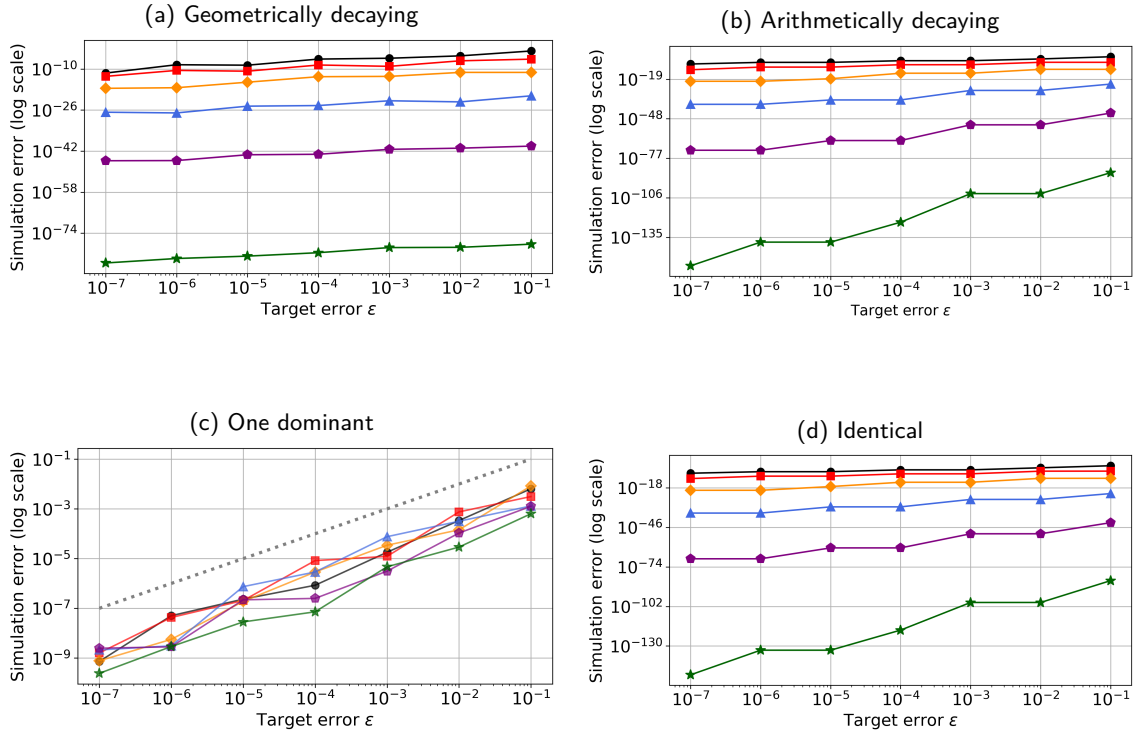


Figure 9: Simulation results obtained by modifying t according to Eq. (B.2) in the Scenario 1 described in Section 4.1.

(k, ϵ)	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
8	3	4	5	5	6	6	7
16	4	4	5	5	6	6	7
32	4	4	5	5	6	7	7
64	4	5	5	6	6	7	7
128	4	5	5	6	6	7	7
256	4	5	5	6	7	7	8

Table 4: **The value of t as a function of (k, ϵ) in Appendix B.** The value of t used in additional numerical simulation is $\min \{r, \lceil \ln(k/\epsilon) / \ln \ln(k/\epsilon) \rceil\}$. Note that $r = 16$.