OUKA
The University of Osaka
Institutional Knowledge Archive

| | |
|---|---|
| Title | Optimization of quantum protocols and computing through variational algorithms and semidefinite programs for reducing quantum resources |
| Author(s) | 趙, 犇池 |
| Citation | 大阪大学, 2025, 博士論文 |
| Version Type | VoR |
| URL | https://doi.org/10.18910/101497 |
| rights | |
| Note | |

*The University of Osaka Institutional Knowledge Archive : OUKA*

https://ir.library.osaka-u.ac.jp/

The University of Osaka

# Optimization of quantum protocols and computing through variational algorithms and semidefinite programs for reducing quantum resources

Benchi Zhao

January, 2025

# Optimization of quantum protocols and computing through variational algorithms and semidefinite programs for reducing quantum resources

A dissertation submitted to
THE GRADUATE SCHOOL OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN SCIENCE

BY

Benchi Zhao

January, 2025

# Abstract

Quantum computers are believed to provide an exponential speedup compared to classical computers for specific problems, such as factoring and quantum simulations. After decades of research, we are in the noisy intermediate-scale quantum era, which means we can access the quantum computer with hundreds of physical qubits with noise. Due to the limited number of qubits on near-term quantum devices, every shot is valuable. It is natural to ask how to utilize near-term quantum devices more efficiently. Or we can ask how to reduce the requirement of quantum resources when we implement the quantum algorithm. In this thesis, we try to use optimization tools to reduce the requirement of quantum resources in the following three tasks: expectation value estimation, extracting non-linear features from noisy systems, and quantum channel simulation using coherence.

In the first part, we point out that the sampling times of expectation value estimation of Hamiltonian depends on the Pauli norm of Hamiltonian. Then we propose a variational quantum algorithm called variational quantum Hamiltonian engineering to minimize the Pauli norm of Hamiltonian, such that the sampling times for estimating expectation value can be reduced. We develop a theory to encode the Pauli norm optimization problem into the vector $l_1$-norm minimization problem. Then we devise an appropriate cost function and utilize the parameterized quantum circuits to minimize the cost function. We also conduct numerical experiments to show the effectiveness of our algorithm on the Ising Hamiltonian and molecules' Hamiltonian. This work makes a significant contribution to improving the efficiency of quantum computing.

In the second part, we focus on extracting accurate high-order moments from noisy quantum systems. To address the noise, we establish a method, called observable shift, for deriving protocols that using quantum operations and classical postprocessing only. Compared with the existing method, our method requires lower sampling overheads and easier implementations. We further construct the protocol for large quantum systems to retrieve the depolarizing channels, making the proposed method scalable. This work contributes to a deeper understanding of how quantum noise could affect high-order information extraction and provides a more efficient way to tackle it.

In the third part, we study channel simulation using coherence, which refers to realizing a target channel with coherent states and free operations. It was shown that if there are no constraints on the coherent states, we can simulate

arbitrary quantum channels with maximally incoherent operations (MIO) accurately and deterministically. However, in real cases, the coherent states are limited, which motivates us to consider the more general probabilistic framework. In this chapter, we develop the framework for probabilistic channel simulation using coherence with free operations. When the chosen set of free operations is the MIO, we provide an efficiently computable SDP to calculate the maximal success probability and derive the analytic expression of success probability for some special cases. When the chosen set of free operations is the dephasing-covariant incoherent operations (DIO), we show that if the target channel is not a resource nonactivating channel, then one cannot simulate it exactly both deterministically and probabilistically. The SDP for maximal success probability of simulating channel by DIO is also given correspondingly. This work fills an important gap in literature by establishing the probabilistic toolbox for the key resource of quantum coherence.

In the first and second parts of this thesis, we consider sampling as a quantum resource and find ways to reduce the requirement of such resources in the task of expectation value estimation and non-linear features estimation from noisy quantum states, respectively. In the third part, we consider coherence as a quantum resource and study the framework of probabilistic channel simulation using limited coherent states.

This thesis contributes to reducing the requirement of quantum resources in different quantum tasks, such as expectation value estimation, non-linear features estimation, and channel simulation using coherence, which makes the quantum devices more efficient. We are confident that this thesis pushes quantum computers a step toward the next milestone. One interesting research in the future is to apply our method on the quantum device to check the actual performance. Besides, another interesting research topic is proposing methods to solve the problems that quantum computers will really encounter.

# Contents

# Abbreviations

| | |
|---|---|
| VQA | Variational Quantum Algorithm |
| SDP | Semidefinite Program |
| NISQ | Noisy Intermediate-Scale Quantum |
| PQC | Parameterized Quantum Circuit |
| QWC | Qubit-Wise Commuting |
| GC | General Commuting |
| QEC | Quantum Error Correction |
| QEM | Quantum Error Mitigation |
| QPD | Quasi-Probability Decomposition |
| CP | Completely Positive |
| TP | Trace Preserving |
| HP | Hermitian Preserving |
| TS | Trace Scaling |
| MIO | Maximally Incoherent Operation |
| DIO | Dephasing-covariant Incoherent Operation |
| ROC | Robustness of Coherence |

# Chapter 1

# Introduction to Quantum Computing

## 1.1 Development of quantum computers

Quantum mechanics and computer science started merging in the 1980s, creating a new interdisciplinary field called *quantum computing*. Quantum computing is a cutting-edge computing model that manipulates units of quantum information according to the principles of quantum mechanics. Over the past forty years, quantum computing technology has advanced rapidly, leading to the development of quantum computers, which are different from classical computers. Quantum computers are considered to have the potential to speed up computing exponentially in particular problems, such as large number factoring [1], Hamiltonian simulation [2], and machine learning [3, 4]. Once we solve these vital problems with quantum computing, it will revolutionize the field of science. For example, it is well-known that the RSA algorithm [5] is one of the first public-key cryptosystems and is widely used for secure data transmission. The security of the RSA algorithm relies on the difficulty of factoring a large number into its prime components. If we can factorize large numbers efficiently with quantum computing, then it will impact the field of cryptography [6]. Besides, if we can simulate Hamiltonian with quantum computer efficiently, then it will impact quantum chemistry [7]. By integrating quantum chemistry into the design and optimization of sustainable technologies, we can significantly enhance the efficiency and scalability of processes crucial for reducing greenhouse gas emissions and transitioning to a low-carbon economy. This contributes to global issues such as carbon neutrality. If we apply quantum computing to machine learning [3], it is expected to lead to transformative advancements, especially for problems such as vast datasets [4] and high-dimensional spaces [8].

Since Feynman proposed the concept of quantum computing [2], research on quantum hardware has also advanced rapidly. As the hardware foundation for quantum computing, researchers have proposed many promising and

operationally feasible quantum hardware platforms. Representative platforms include: superconducting quantum computing [9]; trapped ion quantum computing [10]; rydberg atom quantum computing [11]; semiconductor quantum dot quantum computing [12]; photonic quantum computing [13]; nitrogen vacancy center quantum computing [14]; topological quantum computing [15].

It is widely believed that fault-tolerant quantum computers [16–18] are essential for solving complex problems such as factoring [1] and quantum chemistry [7] with theoretically proven quantum speedup. At this stage, we have achieved several milestones on our way to realizing fault-tolerant quantum computers, although there is still a long way to go before achieving our ultimate goal (solving significant science problems by quantum computers with advantages). In 2014, the high fidelity superconducting quantum computer was realized. The fidelities of single-qubit and two-qubit gates are 99.92% and 99.4%, respectively [19], and the measurement accuracy arrives at 99.8% [20]. Such high fidelity quantum gates and measurements exceed the error correction threshold [21, 22], making error correction possible on quantum devices. This was a great milestone and attracted a great of interest from commercial companies, such as Google and IBM. The development of quantum computers has gotten faster since then. In 2016, IBM developed a 5-qubit quantum computer [23]. Later on, the variational quantum eigensolver algorithm [8] was conducted on IBM's 6-qubit processor in 2017 [24]. In 2018, Google conducted experiments on a 49-qubit quantum computer [25]. One year later, Google posted a new device called Sycamore, which possesses 53 qubits [9].

Due to the property of quantum mechanics, simulating quantum systems requires exponential classical computing resources respect to the quantum system size. Thus, the increasing number of qubits makes it harder and harder for classical computers to simulate them. When the number of qubits exceeds some threshold, it is obvious that even the best classical supercomputer cannot simulate. We call such a phenomenon as quantum supremacy [9, 13, 26]. Quantum supremacy is considered the second milestone on the way to fault-tolerance quantum computing, for it breaks through the limits of our current computing power. In 2019, Google first claimed to have demonstrated the quantum supremacy [9] that the quantum processor performs the target computation in 200 seconds while a classical algorithm would take approximately 10,000 years in the world's fastest supercomputer to solve the same problem. The task that Google used to show supremacy is sampling the output of a pseudo-random quantum circuit [25, 27]. The reason they chose this task is that it does not possess structure and therefore allows for limited guarantees of computational hardness [25, 28]. However, researchers have developed better classical algorithms for the sampling problem used to claim quantum supremacy, vanishing the gap between Google's processor [9] and classical supercomputers [26]. Thus, the quantum supremacy claimed by Google is not the real supremacy. In 2020, Jianwei Pan's group claimed quantum supremacy by implementing Gaussian boson sampling [29] on 76 photons with their photonic quantum computer [13], which is faster than using the state-of-the-art simulation strategy and supercomputers by a factor of $10^{14}$. The efficiency of implementation and the po-

tential applications in graph-based problems [30] and quantum chemistry [31] make authors choose the task of Gaussian boson sampling to show quantum supremacy. In 2022, Xanadu reported another boson sampling experiment with a programmable photonic processor and claimed to have obtained a speedup 50 million times more than the classical supercomputer [32]. In 2024, Google [33] implemented an algorithm for random circuit sampling in the weak-noise phase with 67 qubits at 32 cycles and showed that such a task takes $10^{13}$ years for classical computers. Now, many evidences [13, 32, 33] has shown that quantum computers are reaching a milestone where classical computers are difficult to simulate for benchmarks such as random quantum circuits [25, 27].

## 1.2 Challenges of next milestone

The tasks used to demonstrate the quantum supremacy such as random quantum circuit [25, 27] and Gaussian boson sampling [13, 32] are not so important in reality, for it cannot solve practical problems. It makes no sense to build a machine that can only speed up something useless. The next milestone we should pursue is applying quantum computers for practically important problems and showing advantages.

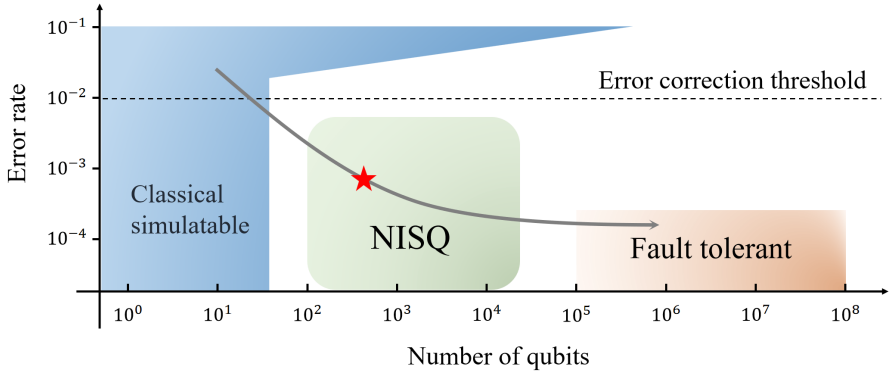



Figure 1.1: The roadmap of quantum computing. We are in the NISQ era with a few hundred noisy physical qubits.

There are many challenges on the way to the next milestone of quantum computing. One of the main obstacles is the limited access to the quantum computers. At this stage, we are in the noisy intermediate-scale quantum (NISQ) era [34], which means we can access to few hundred physical qubits with noise as shown in Figure 1.1. While solving practical problems such as large number factoring [1] and Hamiltonian simulation [2] requires more than millions error corrected qubits, in other words, fault-tolerant devices are necessary. The huge gap in the number of qubits between NISQ devices and fault-tolerant devices needs time to fill, and there is a long way to go. Instead of waiting for the

realization of fault-tolerant devices, we need to think about what we can do with current quantum devices. Since the limited access to quantum computers, it is very important to think about how to use the near-term quantum computers efficiently. In other words, *reducing the requirement of quantum resources when we implement quantum algorithms or quantum tasks* is a necessary problem we have to consider. In this thesis, we aim to propose methods to reduce the requirement of resources with optimization tools for three different tasks: expectation value estimation [35], non-linear features estimation [36], as well as channel simulation using coherent [37]. Specifically, we encode these problems into optimization problems (encode the requirement of quantum resources into the objective function) and choose suitable optimization tools such as variational quantum algorithms (VQA) or semidefinite programs (SDP) to minimize the objective functions (requirement of quantum resources). We believe that this thesis pushes quantum computers a step toward the next milestone.

## 1.3 Overview of this thesis

This thesis makes some attempts and explorations to reduce the requirement of quantum resources in different tasks with optimization tools. The layout of the thesis is shown as follows:

**Chapter 2: Preliminary**
In chapter 2, we first introduce the fundamental concepts in quantum computing, and then we introduce three problems we are going to solve in this thesis. Besides, we also introduce two optimization methods: *variational quantum algorithms (VQA)* and *semidefinite programs (SDP)*, which we utilized to address the three problems we introduced. In the following three chapters, we solve one proposed problem in each chapter with optimization tools.

**Chapter 3: Variational Quantum Hamiltonian Engineering**
In Chapter 3, we point out that the complexity of expectation value estimation of Hamiltonian depends on the Pauli norm of Hamiltonian. Then we propose a variational quantum algorithm called variational quantum Hamiltonian engineering (VQHE) to minimize the Pauli norm of Hamiltonian, such that the overhead for executing expectation value estimation can be reduced. We develop a theory to encode the Pauli norm optimization problem into the vector $l_1$-norm minimization problem. Then we devise an appropriate cost function and utilize the parameterized quantum circuits (PQC) to minimize the cost function. We also conduct numerical experiments to reduce the Pauli norm of the Ising Hamiltonian and molecules' Hamiltonian to show the efficiency of the proposed VQHE. This chapter is based on [Zhao and Fujii, arXiv:2406.08998 (2024)] with slight modifications to fit in the context.

**Chapter 4: Retrieving Non-Linear Features From Noisy Quantum States**

In Chapter 4, we focus on actually extracting high-order moments of quantum states. But in reality, inevitable quantum noise prevents us from accessing the desired value. We address this issue by systematically analyzing the feasibility and efficiency of extracting high-order moments from noisy states. We establish a method for deriving protocols using quantum operations and classical postprocessing only. Our protocols, in contrast to conventional ones, incur lower overheads and avoid sampling different quantum operations due to a novel technique called the *observable shift method*, making the protocols strong candidates for practical use on current quantum devices. The proposed method also indicates the power of entangled protocols in retrieving high-order information, whereas in the existing methods, entanglement does not help. We further construct the protocol for large quantum systems to retrieve the depolarizing channels, making the proposed method scalable. This chapter is based on [Zhao, Benchi, et al. PRX Quantum 5 (2), 020357 (2024)] with slight modifications to fit in the context.

**Chapter 5: Probabilistic Channel Simulation Using Coherence**
In Chapter 5, we study the channel simulation using coherence, which refers to realizing a target channel with coherent states and free operations, which is a fundamental problem in the quantum resource theory of coherence. The limitations of the accuracy of deterministic channel simulation motivate us to consider the more general probabilistic framework. In this chapter, we develop the framework for probabilistic channel simulation using coherence with free operations. When the chosen set of free operations is the maximally incoherent operations (MIO), we provide an efficiently computable semidefinite program (SDP) to calculate the maximal success probability and derive the analytic expression of success probability for some special cases. When the chosen set of free operations is the dephasing-covariant incoherent operations (DIO), we show that if the target channel is not a resource nonactivating channel, then one cannot simulate it exactly both deterministically and probabilistically. The SDP for maximal success probability of simulating channel by DIO is also given correspondingly. This chapter is based on [Zhao, Ito, and Fujii. Phys. Rev. Research 6, 043316 (2024)] with slight modifications to fit in the context.

**Chapter 6: Conclusion**
We summarize the thesis in chapter 6 and we also discuss some further possible research directions.

# Chapter 2

# Preliminary

This section introduces the fundamental concepts in quantum computing, including quantum states, quantum dynamics, and quantum measurements. Then we will introduce three obstacles in quantum computing and specify the problems we are going to solve in the following of this thesis. Next, we will introduce the optimization tools, such as variational quantum algorithms (VQA) and semidefinite programs (SDP), we used to solve the proposed problems.

## 2.1 Fundamental concepts

### 2.1.1 Quantum state

In classical computation, the *bit* is the fundamental unit, whose state is either 0 or 1. There is a counterpart in quantum computation, which is known as *quantum bit* or *qubit*. Two possible states for qubit are the state $|0\rangle$ and $|1\rangle$. The notation $|\cdot\rangle$ is known as the *Dirac notation*. The difference between bits and qubits is that a qubit can be in a state other than $|0\rangle$ and $|1\rangle$. It is also possible to form a linear combination of states, or called *superpositions:*

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,. \tag{2.1}$$

The number $\alpha$ and $\beta$ are complex numbers. According to quantum mechanics, when we measure a qubit we get either the result 0 with probability $|\alpha|^2$, or the result 1 with probability $|\beta|^2$. Naturally, $|\alpha|^2 + |\beta|^2 = 1$, for the probabilities must sum to one. The special states $|0\rangle$ and $|1\rangle$ are known as *computational basis states* and form an orthogonal basis for this vector space.

**State vector**

A state of a quantum system is described by a *state vector*

$$|\psi\rangle = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{d-1} \end{bmatrix} \tag{2.2}$$

on a complex $d$-dimensional system $\mathbb{C}^d$, where the symbol $|\cdot\rangle$ is called *ket* and indicates a complex column vector. Similarly, the notation $\langle\cdot|$ is called *bra* and indicates a complex row vector, and they are related complex conjugate,

$$\langle\psi| = |\psi\rangle^\dagger = \begin{bmatrix} c_0^* & c_1^* & \cdots & c_{d-1}^* \end{bmatrix} \tag{2.3}$$

A quantum state in the $d$-dimensional system can be described by

$$|\psi\rangle = \sum_{i=0}^{d-1} c_i \, |i\rangle \,, \tag{2.4}$$

where $|i\rangle$ are orthogonal basis, i.e., $\langle i|j\rangle = 0, \forall i \neq j$, and $\sum_{i=0}^{d-1} |c_i|^2 = 1$. Let's take the qubit case ($d = 2$) as an example

$$|\psi\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \tag{2.5}$$

with the basis

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.6}$$

**Density operator**

Besides state vector, the quantum state of a quantum system can be described in a more general way known as *density operator*. More precisely, suppose a quantum system is in one of a number of states $|\psi_i\rangle$, where $i$ is an index, with respective probability $p_i$. We call $\{p_i, |\psi_i\rangle\}$ a *ensemble of pure states*. The density operator for the system is defined by the equation

$$\rho := \sum_{i=1} p_i |\psi_i\rangle\langle\psi_i|, \quad \sum_i p_i = 1. \tag{2.7}$$

The density operator is often known as the *density matrix*. The density operator satisfies the following properties:

$$\rho \geq 0, \quad \mathrm{Tr}[\rho] = 1. \tag{2.8}$$

$\rho \geq 0$ means $\rho$ is a positive matrix, i.e., $\langle\psi| \, \rho \, |\psi\rangle \geq 0, \forall \, |\psi\rangle$.

A quantum system whose state $|\psi\rangle$ is known exactly is said to be in a *pure state*. In this case, the density operator is simply $\rho = |\psi\rangle\langle\psi|$. Otherwise, $\rho$ is a

*mixed state*, i.e., a mixture of different pure states. We say a quantum state $\rho$ is a pure state if and only if $\mathrm{Tr}[\rho^2] = 1$. Otherwise, $\rho$ is a mixed state.

The deepest application of the density operator is as a descriptive tool for *sub-systems* of composite quantum systems. Such a description is provided by the *reduced density operator*. Suppose we have a physical system $A$ and $B$, whose state is described by a density matrix operator $\rho^{AB}$. The reduced density operators are defined by

$$\rho^A = \mathrm{Tr}_B[\rho^{AB}], \quad \rho^B = \mathrm{Tr}_A[\rho^{AB}]. \tag{2.9}$$

### 2.1.2 Quantum dynamics

#### Unitary

In *closed* quantum system, the evolution of the system can be described by *unitary transformation*, mapping a quantum state from initial state $|\psi\rangle$ to the final state $|\psi'\rangle$,

$$|\psi'\rangle = U\,|\psi\rangle\,. \tag{2.10}$$

In quantum computing, the time evolution $U$ is sometimes called the quantum gate. The Pauli operators are the most fundamental quantum gates, which are

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.11}$$

$X$ gate is also called *NOT* gate or *bit flip* gate, while $Z$ gate is also known as *phase flip* gate. The set of Pauli operators forms an orthogonal basis set in operator space, which means that we can decompose any operator on the basis of Pauli. For example, a Hamiltonian $H$ can be decomposed into Pauli operators

$$H = \sum_i h_i P_i. \tag{2.12}$$

#### Quantum channel

In *open* quantum system, the system interacts with the environment, the quantum dynamics can be described by *quantum channel* $\mathcal{E}$ or called *super-operators*, mapping quantum system from initial state $\rho$ to final state

$$\rho' = \mathcal{E}(\rho). \tag{2.13}$$

We say a map $\mathcal{M}$ is a quantum channel if and only if it satisfies the completely positive (CP) and trace-preserving (TP) condition:

1. Completely positive (CP): $\mathcal{I}_k \otimes \mathcal{M}(X) \geq 0$, for all $k \geq 0$, for all positive operator $X$, where $\mathcal{I}_k$ is an identity map on reference system with dimension $k^2$.

2. Trace-preserving (TP): $\mathrm{Tr}[\mathcal{M}(Y)] = \mathrm{Tr}[Y]$, for all positive operator $Y$.

Trace-nonincreasing is known as a relaxation of trace-preserving condition, given by $\mathrm{Tr}[\mathcal{M}(Y)] \leq \mathrm{Tr}[Y]$. Here we emphasize that CPTP maps are the largest operation set on quantum hardware.

The quantum channel $\mathcal{E}$ can be well described in several equivalent ways. Here we are going to introduce two of them, namely *Kraus representation* and *Choi-Jamiołkowski representation.*

**Kraus representation**

A quantum channel can be described by a set of Kraus operators $\{E_i\}$ [22]

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger, \tag{2.14}$$

where $E_i$ are linear operators. Since the quantum channel must be TP, the Kraus operators must satisfy $\sum_i E_i^\dagger E_i = I$. Note that the Kraus representation is not unique.

**Choi-Jamiołkowski representation**

The Choi-Jamiołkowski representation [38–40] of a quantum channel provides a method to express a quantum channel as a bipartite operator, making it a fundamental concept in quantum information theory. For a quantum channel $\mathcal{E}_{A'\to B}$, its *Choi operator* is defined as

$$J_{\mathcal{E}_{A'\to B}} = \sum_{i,j} |i\rangle\langle j|_A \otimes \mathcal{E}_{A'\to B}(|i\rangle\langle j|_{A'}). \tag{2.15}$$

The *Choi state* of channel $\mathcal{E}_{A'\to B}$ is defined as

$$\Phi_{\mathcal{E}} = \frac{1}{d_A} J_{\mathcal{E}} = \frac{1}{d_A} \sum_{i,j} |i\rangle\langle j|_A \otimes \mathcal{E}_{A'\to B}(|i\rangle\langle j|_{A'}) \tag{2.16}$$

This representation is equivalent to applying the quantum channel $\mathcal{E}$ on the second system of the maximally entangled state $\Phi_{AA'}$, as shown in Figure. 2.1

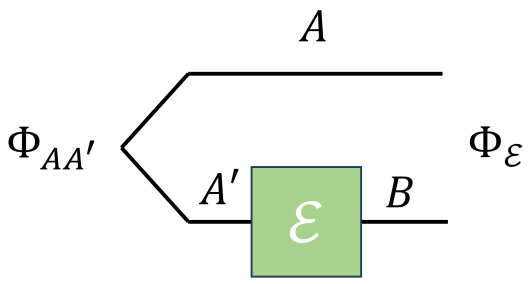



Figure 2.1: The Choi state $\Phi_{\mathcal{E}}$ of a quantum channel $\mathcal{E}$. $\Phi_{AA'}$ is the max entangled state.

The completely positive and trace-preserving (or trace-nonincreasing) conditions for the quantum channel have different expressions in Choi-Jamiołkowski representation as shown in the following [38]:

- $\mathcal{E}$ is completely positive $\iff J_{\mathcal{E}} \geq 0$;
- $\mathcal{E}$ is trace-preserving $\iff \mathrm{Tr}_B[J_{\mathcal{E}}] = \mathbb{I}_A$, where $\mathbb{I}$ refers to identity.
- $\mathcal{E}$ is trace non-increasing $\iff \mathrm{Tr}_B[J_{\mathcal{E}}] \leq \mathbb{I}_A$.

The quantum dynamics in Choi-Jamiołkowski representation is

$$\mathcal{E}_{A\to B}(\rho) = \mathrm{Tr}_A[(\rho_A^T \otimes \mathbb{I}_B)J_{\mathcal{E}}]. \tag{2.17}$$

Given two quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{B\to C}$, Choi representation of the composed channel $\mathcal{M}\circ\mathcal{N}_{A\to C}$ is given by

$$J_{\mathcal{M}\circ\mathcal{N}} = \mathrm{Tr}_B[(J_{\mathcal{N}_{A\to B}}^{T_B} \otimes \mathbb{I}_C)(\mathbb{I}_A \otimes J_{\mathcal{M}_{B\to C}})], \tag{2.18}$$

where the superscript $T_B$ means take partial transpose on system $B$.

### 2.1.3 Quantum measurement

Quantum measurements are described by a collection $M_m$ of *measurement operators*. The index $m$ refers to the measurement outcomes that may occur in the experiment. Given a quantum state $\rho$, and then make a measurement on it. The probability of getting result $m$ is

$$p(m) = \mathrm{Tr}[M_m \rho M_m^\dagger], \tag{2.19}$$

and the state after the measurement collapses to

$$\rho' = \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}[M_m \rho M_m^\dagger]}. \tag{2.20}$$

The probability of getting all results must sum to one, which implies the measurement operators are *complete*, i.e.,

$$\sum_m M_m^\dagger M_m = \mathbb{I}. \tag{2.21}$$

## 2.2 Quantum resources

Quantum resources [41] refer to specific features or properties of quantum systems that provide advantages over classical systems in performing various computational, informational, or physical tasks. These resources are central to quantum information theory and are crucial in applications. If we would like to achieve quantum advantage, then quantum resources are necessary. If a task does not use quantum resources at all, this implies that such a task is a classical task, which, of course, shows no quantum advantage. Now, we are in the NISQ era, which means that we can only access to few hundred noisy physical qubits. Due to the limited number of quantum computing resources, it is necessary to consider how to use them more efficiently. In this thesis, we attempt

to propose methods to reduce the requirement of quantum resources in different quantum tasks. Thus, implementing the same tasks with our methods needs fewer quantum resources than that of the conventional method. The quantum resources are saved, and the spared quantum resource could be used in other tasks, maximizing the use of quantum computers.

Quantum resource is not unique, the common quantum resources are entanglement [42–45], coherence [46–48] and magic [49–51].

- **Quantum entanglement** [42–45] is a phenomenon in quantum mechanics where two or more particles become linked in such a way that the state of one particle directly influences the state of the others, no matter how far apart they are. If there is no entanglement in quantum computing, then all the quantum systems are isolated and in the form of tenor products. In this case, it is hard to show the advantage of quantum computing.

- **Quantum coherence** [46–48] refers to the property of a quantum system where it exists in a superposition of different states, allowing the system to exhibit wave-like interference effects. Coherence is central to the concept of quantum superposition. If there is no coherence in quantum computing, then the quantum systems are just classical systems, and the computation is classical as well.

- **Quantum magic** [49–51] refers to a key concept in the field of quantum computing, particularly in the context of fault-tolerant quantum computation. It is known that a class of quantum computing, Clifford circuits, is classically efficiently simulatable even if it includes entanglement and coherence [52]. Thus, the non-Clifford (magic) resource state stands for the difficulty of a classical simulation. In other words, quantum magic determines the difficulty of classical simulation.

If we are considering the general quantum resources, then we need to take quantum sampling [53, 54] into account. Quantum sampling refers to the use of quantum systems to efficiently sample from complex probability distributions, which may be infeasible or computationally intensive for classical methods. Every shot executed on quantum device can be regarded as one sampling. Thus, if quantum sampling is forbidden in computing, then it will be meaningless to talk about quantum computing. In this thesis, we are going to focus on the quantum resources of sampling and coherence.

## 2.3 Quantum problems

In this section, we introduce three problems that we are going to solve in this thesis.

### 2.3.1 Expectation value of Hamiltonian

Estimating the expectation value of Hamiltonian is a critical step in many quantum algorithms such as variational quantum eigensolver (VQE) [24]. The expectation value of a Hamiltonian $H$ with respect to a quantum state $|\psi\rangle$ is given by $\langle H\rangle = \langle\psi| H |\psi\rangle$. This quantity represents the average energy of the quantum state $|\psi\rangle$ when measured on the basis of the eigenstates of the Hamiltonian $H$. For a given Hamiltonian $H$, we can write it as

$$H = \sum_{i=1}^{L} h_i P_i, \tag{2.22}$$

where $L$ is the number of terms, $h_i$ is the coefficient of Pauli terms $P_i$. To estimate the expectation value of Hamiltonian $H$, one can estimate the expectation value of each Pauli term and sum them up by post-processing $\langle H\rangle = \sum_i h_i \langle P_i\rangle$, with variance [55, 56]

$$\epsilon^2 = \sum_{i}^{L} \frac{|h_i|^2 \mathrm{Var}[P_i]}{S_i}, \tag{2.23}$$

where $S_i$ is the measurement time of each of the operators $P_i$, Var denotes the variance in the measurement of the operator for the given trial state.

Generally, the variance of Pauli $\mathrm{Var}[P_i]$ is unknown in advance. However, since the terms $P_i$ are Pauli, the variance is upper bounded by $\mathrm{Var}[P_i] \leq 1$ [55, 56]. One can confirm (e.g., via the use of Lagrange multipliers [57]) that the least number of measurements required to bound variance below $\epsilon^2$ can be found by choosing $S_i \propto |h_i|$ [55]. This implies the total measurement time $N$ is upper bounded by

$$N \leq \frac{\|H\|_P^2}{\epsilon^2}, \tag{2.24}$$

where $\|H\|_P = \sum_i |h_i|$ is called Pauli norm. Obviously, the total sampling time is determined by Pauli norm, and it is natural to ask a question that

*Can we find a method to engineer the Hamiltonian to minimize its Pauli norm, such that the sampling times reduce correspondingly?*

We propose an algorithm called variational quantum Hamiltonian engineering to solve this issue in Chapter 3.

### 2.3.2 Non-linear features estimation from noisy systems

Non-linear features of a quantum state, i.e., $\mathrm{Tr}[\rho^k]$ with integer $k$, is a fundamental and significant quantity in quantum information theory. However, quantum systems are rarely closed in practice as they unavoidably interact with the environment. Quantum channels $\mathcal{N}$, stemming from the unitary dynamics in a larger Hilbert space, are considered the proper mathematical formalism depicting the evolution of general quantum systems. A quantum state $\rho$, which

suffers from noise, is denoted as $\rho' = \mathcal{N}(\rho)$. The corresponding estimation of non-linear feature becomes $\mathrm{Tr}[\mathcal{N}(\rho)^k]$, which is not the desired value.

Quantum error correction (QEC) [58] is widely considered a potential solution to address quantum noise. The idea of QEC is encoding several physical qubits into logical qubits, such that the error can be detected and corrected. Many efforts have been devoted to studying the error correction codes such as color codes [59], stabilizer codes [21], and quantum low-density-parity-check (LDPC) codes [60]. However, achieving a high-fidelity logical qubit requires a huge amount of physical qubits (achieving a $10^{-6}$ error rate would require a distance-27 logical qubit using 1457 physical qubits [61]), making it impractical on near-term quantum devices. To get precise results from near-term quantum devices, quantum error mitigation (QEM) was proposed [62]. Unlike QEC, QEM aims to remove noise from classical information, while QEC aims to remove the noise from the quantum information. Specifically, given noisy states $\mathcal{N}(\rho)$, QEM outputs the accurate expectation value of some observables $O$, while QEC removes the noise and outputs perfect quantum state $\rho$, which is

$$\mathcal{N}(\rho) \xrightarrow{QEM} \mathrm{Tr}[O\rho], \quad \mathcal{N}(\rho) \xrightarrow{QEC} \rho. \tag{2.25}$$

The idea of QEM is to apply the inverse channel of noise channel $\mathcal{N}^{-1}$ to the quantum system, and then make measurements over observable $O$. In this case, the estimated expectation value $\zeta$ will be

$$\zeta = \mathrm{Tr}[O\mathcal{N}^{-1} \circ \mathcal{N}(\rho)] = \mathrm{Tr}[O\rho], \tag{2.26}$$

which is exactly the desired value. However, one critical problem is that the inverse channel $\mathcal{N}^{-1}$ is generally a Hermitian-preserving trace-preserving (HPTP) map, which means that we cannot apply such a channel on a quantum device directly. Fortunately, it was proved that any HPTP map can be decomposed into a linear combination of CPTP maps [54], i.e.,

$$\mathcal{N}^{-1} = c_1\mathcal{D}_1 + c_2\mathcal{D}_2, \tag{2.27}$$

where $c_1, c_2$ are real values, and $\mathcal{D}_1, \mathcal{D}_2$ are CPTP maps. The target value $\mathrm{Tr}[O\rho]$ can be obtained by the following steps:

1. For $t$-th sampling time, sample quantum channel $\mathcal{D}^t$ from $\{\mathcal{D}_1, \mathcal{D}_2\}$ with probability $\{\frac{|c_1|}{g}, \frac{|c_2|}{g}\}$, respectively, where $g = |c_1| + |c_2|$ is known as *sampling overhead*.
2. Apply the sampled channel onto the noisy state $\mathcal{D}^{(t)} \circ \mathcal{N}(\rho)$
3. Measure the observable of interest and record the measurement value $o^{(t)}$.
4. Repeat the above sampling and measuring process for $T$ times.
5. Post process the $N$ measurement results to get the desired value

$$\zeta = \frac{g}{T}\sum_{t=1}^{T} \mathrm{sgn}(c^{(t)})o^{(t)}, \tag{2.28}$$

where sgn stands for the sign of $c^{(t)}$.

By the Hoeffding's inequality, the total sampling time that obtains the estimation within an error $\epsilon$ with a probability no less than $1-\delta$ is upper bounded by $2g^2\frac{\log(2/\delta)}{\epsilon^2}$. Thus, if we set the total sampling time $T$ no less than the upper bound, i.e.,

$$T \geq 2g^2\frac{\log(2/\delta)}{\epsilon^2} \propto g^2, \tag{2.29}$$

then it is guaranteed that the error of estimated value within $\epsilon$ with a probability no less than $1-\delta$.

Note that the realization of Hermitian preserving map $\mathcal{N}^{-1}$ shown above requires sampling quantum channels with certain probability [54], which is complicated implementation. In Ref. [63], the authors proposed that a Hermitian preserving map $\Phi$ can be realized by injecting different quantum states, $\omega_i$ into a fixed quantum channel $\Lambda$, i.e., $\Phi(\cdot) = \sum_i c_i\Lambda(\cdot \otimes \omega_i)$. The new proposed method [63], no doubt, simplifies the realization of Hermitian preserving maps. However, we still need to sample different states with certain probabilities and inject them into quantum channel. Since both methods [54, 63] have to implement probabilistic sampling, it motivates us to consider the following problem:

*Can we design a more efficient way to mitigate noise when estimating non-linear features without probabilistic sampling?*

We propose a novel method called observable shift to solve this issue in Chapter 4.

### 2.3.3 Channel simulation using limited coherence

Quantum resource theory [41, 64] can be viewed as a framework for understanding the *interconversion* of various resources. This theory facilitates a wide range of quantum information processing tasks. There are three fundamental components of any resource theory: *resources* (such as entanglement), *free resources* (such as separable states), and a restricted set of *free operations* (such as local operations and classical communication). These components are interdependent; specifically, it is required that resources cannot be generated from free resources using free operations. Coherence is an important topic in quantum resource theories [41], which refers to the property of the superposition of states. It empowers various quantum tasks, such as cryptography [65], metrology [66–68], thermodynamics [69–71], and channel simulation [48, 72].

In previous works of resource theory of coherence, they focused on the quantification and interconversion of *static coherence*–the degree of superposition present in a state [73–75]. Later on, a more general framework was proposed, called *dynamic coherence*–the power to generate coherence itself [72]. How to convert static coherence into dynamic coherence (also known as *channel simulation using coherence*.) raises great interest, and many efforts have contributed to establishing the framework of deterministic channel simulation [48, 72, 76].

At the same time, the limitations of deterministic channel simulation are also shown. In Ref. [48, 72], it has been proved that any CPTP map can be simulated by maximally incoherent operations (MIO) with the help of maximally coherent states $\Psi_m$. However, in many cases, we cannot access the perfect maximally coherent state, and we can only access the limited resource state $\omega$ with few resources, then it is impossible to simulate the target channel deterministically and accurately. Here we are curious about the question that

*If we can only access the resource state $\omega$ with few resources, can we simulate the target channel accurately with probability?*

We develop a probabilistic framework for channel simulation using coherence to answer this question in Chapter 5.

## 2.4 Optimization tools

### 2.4.1 Variational quantum algorithms (VQA)

Variational Quantum Algorithms (VQAs) [8] have become the leading strategy for achieving quantum advantage on NISQ (Noisy Intermediate-Scale Quantum) devices. These algorithms are designed to handle the constraints of NISQ computers through optimization-based or learning-based approaches. They utilize classical optimization techniques by running parameterized quantum circuits on quantum computers and then outsourcing the parameter optimization to classical optimizers. This hybrid approach not only leverages classical computational strengths but also helps keep the quantum circuit depth shallow, thereby reducing noise compared to quantum algorithms developed for the fault-tolerant era.

One of the main advantages of Variational Quantum Algorithms (VQAs) is their flexibility, providing a general framework for solving a wide range of problems. VQAs have been successfully applied to various tasks such as searching ground state [24, 77] and excited state [78, 79], quantum classification [80, 81], quantum information analysis [82–84], and quantum data compression [85, 86].

To develop a VQA to solve a particular optimization problem, we need to follow the steps:

1. Encode the optimization problem into a cost function $C(\theta)$;
2. Prepare the parameterized quantum circuit $U(\theta)$;
3. Apply the circuit $U(\theta)$ and estimate the cost function on the quantum device;
4. Optimize parameter $\theta$ with classical optimizer to minimize the cost function $C(\theta)$;
5. Repeat 3-5 until the cost function converges to its minimum.

Here we take variational quantum eigensolver as an example, which aims to obtain the ground energy of a Hamiltonian. At the beginning, we are given a Hamiltonian, then we prepare the quantum circuit as shown in the yellow in Figure. 2.2. The parameter of rotation angles is chosen randomly at the beginning. Our aim is to find the minimal eigenvalue of Hamiltonian, so we cost function is set as

$$C(\theta) = \langle\psi(\theta)| H |\psi(\theta)\rangle = \langle 0| U(\theta)^\dagger H U(\theta) |0\rangle . \tag{2.30}$$

Next, we apply the unitary $U(\theta)$ to the trail state $|0\rangle$, and estimate the expectation value of Hamiltonian to get the cost function. Then we use classical optimization to minimize the cost function until the cost function converges. The converged value will be the desired ground engine value of Hamiltonian. The process of VQE is shown in Figure. 2.2.

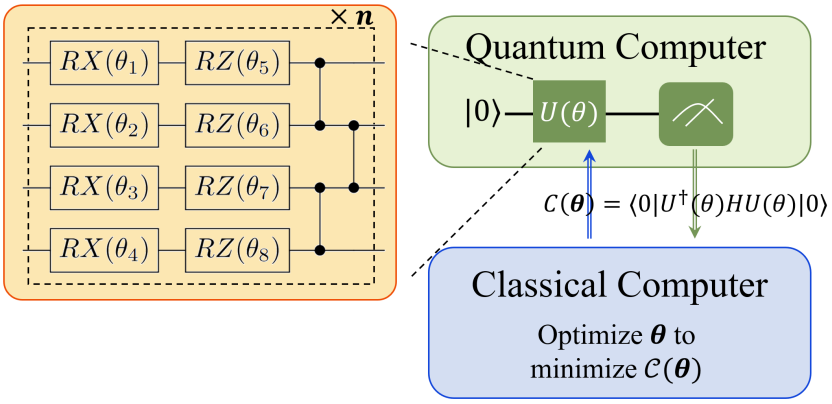



Figure 2.2: The diagram for variational quantum eigensolvers.

### 2.4.2 Semidefinite programs (SDP)

Semi-definite programs (SDPs) represent a crucial category of optimization problems frequently encountered in quantum information theory. An SDP is characterized by its optimization variable, which is a positive semi-definite operator $X$. The objective function in an SDP is linear with respect to $X$, and the constraints involve an operator inequality that features a linear function of $X$.

**Definition 2.1.** *Given a Hermiticity-preserving superoperator $\Phi$ and Hermitian operator $A$ and $B$, a semidefinite program (SDP) is defined by primal problem and dual problem*

$$\begin{array}{rlrl} \text{supremum} & \mathrm{Tr}[AX] & \text{infimum} & \mathrm{Tr}[BY] \\ \text{subject to} & \Phi(X) \leq B, & \text{subject to} & \Phi^\dagger(Y) \geq A, \\ & X \geq 0 & & Y \geq 0 \end{array}$$

where $\Phi^\dagger$ is the dual map to $\Phi$, which is $\mathrm{Tr}[Y\Phi(X)] = \mathrm{Tr}[X\Phi^\dagger(Y)]$.

A variable $X$ for the primal problem is called *feasible solution* if the two constraints in the primal problem are satisfied, i.e., $X \geq 0, \Phi(X) \leq B$; if a variable $X$ satisfies $X > 0, \Phi(X) < B$, then it is called *strictly feasible solution*. It is similar to the dual problem that a variable $Y$ is a feasible solution if $Y \geq 0, \Phi^\dagger(Y) \geq A$; a variable $Y$ is a strictly feasible solution if $Y > 0, \Phi^\dagger(Y) > A$.

We denote the optimal value obtained by the primal and dual problem as

$$\alpha = \sup_{X \geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\} \tag{2.31}$$

$$\beta = \inf_{Y \geq 0} \{\mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A\}. \tag{2.32}$$

If there is no primal feasible solution, then $\alpha = -\infty$, and if there is no dual feasible solution, then $\beta = +\infty$.

**Proposition 2.1.** *[38] For every SDP, the following weak duality inequality holds*

$$\alpha \leq \beta \tag{2.33}$$

This convenient relationship allows us to directly bound the optimal values of the primal problem by selecting a valid variable from the dual problem, and vice versa.

The condition where $\alpha = \beta$ is referred to as strong duality, which does not hold universally. However, it often holds for semidefinite programs encountered in practice. Several conditions can be used to verify strong duality. The following proposition outlines one set of conditions that ensure strong duality.

**Proposition 2.2.** *[38] Slater's theorem is a sufficient condition for strong duality to hold, and it is given as follows*

- *If there exists $X \geq 0$ such that $\Phi(X) \leq B$ and there exists $Y > 0$ such that $\Phi^\dagger(Y) > A$, then $\alpha = \beta$ and there exists a primal feasible operator $X$ for which $\mathrm{Tr}[AX] = \alpha$.*
- *If there exists $Y \geq 0$ such that $\Phi^\dagger(Y) \geq A$ and there exists $X > 0$ such that $\Phi(X) > B$, then $\alpha = \beta$ and there exists a primal feasible operator $Y$ for which $\mathrm{Tr}[BY] = \beta$.*

The primal and dual problems are connected by the max-min of a Lagrangian $\mathcal{L}(\Phi, A, B, X, Y)$,

$$\mathcal{L}(\Phi, A, B, X, Y) := \mathrm{Tr}[AX] + \mathrm{Tr}[BY] - \mathrm{Tr}[\Phi(X)Y] \tag{2.34}$$

$$= \mathrm{Tr}[AX] + \mathrm{Tr}[(B - \Phi(X))Y] \tag{2.35}$$

$$= \mathrm{Tr}[BY] + \mathrm{Tr}[(A - \Phi^\dagger(Y))]. \tag{2.36}$$

Taking an infimum over $Y \geq 0$ on $\mathcal{L}(\Phi, A, B, X, Y)$ and then a supremum over $X \geq 0$, such a max-min problem is equivalent to the primal problem, which is

$$\sup_{X\geq 0} \inf_{Y\geq 0} \mathcal{L}(\Phi, A, B, X, Y) = \sup_{X\geq 0} \{\mathrm{Tr}[AX] + \inf_{Y\geq 0} \mathrm{Tr}[(B - \Phi(X))Y]\} \tag{2.37}$$

$$= \alpha. \tag{2.38}$$

If $B$ does not satisfy the condition $B \geq \Phi(X)$, then the term $\inf_{Y\geq 0} \mathrm{Tr}[(B - \Phi(X))Y]$ will be $-\infty$, which has no feasible solution.

If we first take a supremum over $X \geq 0$ on $\mathcal{L}(\Phi, A, B, X, Y)$ and then take an infimum over $Y \geq 0$, such a min-max problem is equivalent to the dual problem

$$\inf_{Y\geq 0} \sup_{X\geq 0} \mathcal{L}(\Phi, A, B, X, Y) = \inf_{Y\geq 0} \{\mathrm{Tr}[BY] + \sup_{X\geq 0} \mathrm{Tr}[(A - \Phi^\dagger(Y))X]\} \tag{2.39}$$

$$= \beta. \tag{2.40}$$

Similarly, if $A$ does not satisfy the condition $A \leq \Phi^\dagger(Y)$, then the term $\sup_{X\geq 0} \mathrm{Tr}[(A - \Phi^\dagger(Y))X]$ will be $+\infty$, which has no feasible solution.

# Chapter 3

# Variational Quantum Hamiltonian Engineering

In this chapter, we point out that the complexity of expectation value estimation of Hamiltonian depends on the Pauli norm of Hamiltonian. Then we propose a variational quantum algorithm called variational quantum Hamiltonian engineering (VQHE) to minimize the Pauli norm of Hamiltonian, such that the overhead for executing expectation value estimation can be reduced. We develop a theory to encode the Pauli norm optimization problem into the vector $l_1$-norm minimization problem. Then we devise an appropriate cost function and utilize the parameterized quantum circuits (PQC) to minimize the cost function. We also conduct numerical experiments to reduce the Pauli norm of the Ising Hamiltonian and molecules' Hamiltonian to show the efficiency of the proposed VQHE. This chapter is based on [Zhao and Fujii, arXiv:2406.08998 (2024)] with slight modifications to fit in the context.

## 3.1 Introduction

Quantum computing, which harnesses the principles of quantum mechanics to revolutionize computation, is a promising solution to tackle problems that are currently intractable for classical computers. In recent decades, quantum technologies have emerged with a growing number of powerful applications across diverse fields such as optimization [87, 88], chemistry [89, 90], security [91, 92], and machine learning [3]. In quantum computing, the Hamiltonian of a quantum system, an operator that represents the total energy of a quantum system, plays a pivotal role in many critical tasks such as expectation value estimation of a Hamiltonian and Hamiltonian simulation.

Estimating the expectation value of a Hamiltonian is a fundamental task in quantum mechanics and quantum computing, especially when simulating quantum systems or solving quantum chemistry problems. The expectation value of a Hamiltonian $H$ with respect to a quantum state $|\psi\rangle$ is given by $\langle H\rangle =$

$\langle\psi| H |\psi\rangle$. This quantity represents the average energy of the quantum state $|\psi\rangle$ when measured on the basis of the eigenstates of the Hamiltonian $H$. The most straightforward method to estimate the expectation value is to decompose the Hamiltonian into Pauli terms, i.e., $H = \sum_{i=1}^{L} h_i P_i$, and estimate the expectation value of each term $\langle P_i \rangle$. Then calculate the expectation value by post-processing $\langle H \rangle = \sum_i h_i \langle P_i \rangle$. The total measurement time is $N \propto L|h_{\max}|^2$ [24, 56], where $L$ is the number of non-zero Pauli coefficients, and $h_{\max}$ is the largest Pauli coefficient. While the number of Pauli terms typically increases polynomially in the system size, it amounts to be large for complex quantum systems such as those huddled in quantum chemistry, which makes this method no longer efficient. To tackle this problem, one can sample the Pauli term $P_i$ of the target Hamiltonian $H$ randomly with a certain probability $p_i$, which is proportional to its absolute value of the coefficient, i.e., $p_i = \frac{|h_i|}{\sum_i |h_i|}$ [55, 57, 93]. In this case, the total measurement time is determined by the summation of the absolute value of Pauli coefficients, i.e., $N \propto (\sum_{i=1}^{L} |h_i|)^2$.

Hamiltonian simulation is another fundamental task in quantum computing, which simulates the dynamics of a quantum system. The standard methods for Hamiltonian simulation, such as Trotter-Suzuki decomposition [94, 95], are practical for sparse Hamiltonian. The gate count of such a method, which quantifies the complexity of Hamiltonian simulation, depends on the number of Pauli terms in the Hamiltonian. If the system's Hamiltonian is not sparse, then the consumption of the standard method will be too large to be acceptable. To address this problem, Campbell proposed an approach called quantum stochastic drift protocol (qDrift) [96]. This protocol weights the probability of gates by the corresponding interaction strength in the Hamiltonian, leading to a gate count independent of the number of terms in the Hamiltonian but the summation of the absolute value of Pauli coefficients, i.e., $\sum_{i=1}^{L} |h_i|$.

The complexity of both tasks, expectation value estimation and Hamiltonian simulation, are dependent on the summation of the absolute value of Pauli coefficients, which is called *Pauli norm* (or *stabilizer norm*) of Hamiltonian. For a given Hamiltonian $H = \sum_{i=1}^{4^n} h_i P_i$, the Pauli norm $\|H\|_P$ is defined by

$$\|H\|_P = \sum_{i=1}^{4^n} |h_i|. \tag{3.1}$$

It is natural to wonder if a method exists to reduce the Pauli norm of a given Hamiltonian $H$ such that the complexities of the two tasks can be mitigated further. Our answer to this question is positive. In this work, we proposed a variational quantum algorithm (VQA) [8, 97] called variational quantum Hamiltonian engineering (VQHE) to minimize the Pauli norm of Hamiltonian, and the complexity of the two tasks are reduced correspondingly. VQA is a popular paradigm for near-term quantum applications, which uses a classical optimizer to train parameterized quantum circuits (PQCs) [98] to achieve certain tasks. VQAs have been successfully applied to various tasks such as searching ground state [24, 77] and excited state [78, 79], quantum classification [80, 81], quantum

information analysis [82–84], and quantum data compression [85, 86]. In previous works [99, 100], it was proposed to diagonalize the Hamiltonian by VQA. Such a strategy can reduce the complexity of the expectation value of Hamiltonian because all the Pauli terms of the diagonalized Hamiltonian commute with each other, and they can be measured simultaneously. However, the VQAs to diagonalize the Hamiltonian are too costly. In the algorithm in Ref. [99], we have to estimate each eigenvalue of the Hamiltonian in every iteration. The non-zero eigenvalue of the Hamiltonian might increase exponentially with the system size, making the algorithm impossible. In Ref. [100], the first step of the algorithm to diagonalize the Hamiltonian is preparing the Hamiltonian thermal state, which is a QMA-hard [101] problem. Thus, the algorithm in Ref. [100] is not practical either.

In this work, we propose a pre-processing algorithm to engineer the Hamiltonian, called variational quantum Hamiltonian engineering (VQHE), to minimize the Pauli norm of the engineered Hamiltonian. With the engineered Hamiltonian, the processing of Hamiltonian such as expectation value estimation and Hamiltonian simulation becomes easier to implement. Specifically, we use a parameterized unitary $U(\boldsymbol{\theta})$ to reduce the Pauli norm of a given Hamiltonian $H$, i.e., $\min \|U(\boldsymbol{\theta})HU(\boldsymbol{\theta})^\dagger\|_P$. We first develop the theory to convert the Pauli norm optimization problem into the state $l_1$-norm minimization problem. Then design an appropriate cost function and minimize it variationally. The VQHE algorithm outputs the engineered Hamiltonian, whose Pauli norm is minimized. We then display how to apply the engineered Hamiltonian to the tasks of the expectation value estimation and the Hamiltonian simulation. Especially in the task of expectation value estimation, we emphasize that the engineered Hamiltonian is compatible with grouping strategy [24, 102–104], which is another method to reduce the measurement time. The numerical experiments are conducted by applying the VQHE algorithms to the Ising Hamiltonian and some molecules' Hamiltonian, which shows the effectiveness of our algorithm. This work proposes a variational quantum algorithm to reduce the measurement complexity of expectation value estimation and the gate count of Hamiltonian simulation, making a significant contribution to improving the efficiency of quantum computing.

## 3.2 Theoretical framework

In this section, we are going to encode the Hamiltonian into a quantum state and prove that applying a unitary channel on a Hamiltonian is equivalent to applying a unitary gate on the Hamiltonian state.

For a given $n$-qubit Hamiltonian $H$, we can always decompose it into Pauli basis as

$$H = \sum_{i=1}^{4^n} h_i P_i, \tag{3.2}$$

where $h_i$ is the real Pauli coefficient and the $P_i \in \{I, X, Y, Z\}^{\otimes n}$ is $n$-qubit

Pauli tensor product. The problem of our task is minimizing the Pauli norm of Hamiltonian $H$ by optimizing the unitary $U$, which can be written as

$$P_{\text{opt}}(H) = \min_{U} \|UHU^\dagger\|_P. \tag{3.3}$$

On quantum computers, it is not convenient to apply a unitary channel $\mathcal{U}$ on the Hamiltonian $H$ directly. For simplicity, our first step is to encode the Hamiltonian into a quantum state by the following Definition 3.1.

**Definition 3.1.** (Hamiltonian vectorization) *For a given $n$-qubit Hamiltonian $H = \sum_{i=1}^{4^n} h_i P_i$, it can be encoded into a $2n$-qubit state*

$$|H\rangle = \frac{1}{\lambda}(h_1, h_2, \cdots, h_{4^n})^T, \tag{3.4}$$

*where $|H\rangle$ is the Hamiltonian state, and Hamiltonian vector refers to the unnormalized Hamiltonian state, $T$ stands for transpose, and $\lambda$ refers to the normalization factor, i.e., $\lambda = \sqrt{\sum_{i=1}^{L=4^n} h_i^2}$.*

For example, if the given Hamiltonian is $H = I + 2X + 3Y - 4Z$, then the corresponding Hamiltonian state is $|H\rangle = \frac{1}{\sqrt{30}}(1, 2, 3, -4)^T$. The $l_1$-norm of the Hamiltonian state is $\| |H\rangle \|_1 = 10/\sqrt{30}$. The state initialization process can be efficiently executed on a quantum device. It was proved in Ref.[105] that arbitrary $n$-qubit, $d$-sparse quantum state can be deterministically prepared with a circuit depth $\Theta(\log(nd))$. Fortunately, the number of Pauli terms of the most meaningful Hamiltonian is polynomial, i.e., $\tilde{\mathcal{O}}(n^4)$. Thus, the corresponding Hamiltonian state is $\tilde{\mathcal{O}}(n^4)$-sparse, implying it can be prepared efficiently.

The engineered Hamiltonian state can be directly represented as $|H'\rangle = \left|UHU^\dagger\right\rangle$. Such a state can be equivalently expressed as an *encoded unitary* $V$ applied on the Hamiltonian state $|H\rangle$. The formal statement is shown in the following theorem.

**Theorem 3.1.** *For a given Hamiltonian $H$ and unitary $U$, the vectorized engineered Hamiltonian is $|H'\rangle = \left|UHU^\dagger\right\rangle$, which is equivalent to applying a unitary gate $V$ on the Hamiltonian state*

$$\left|UHU^\dagger\right\rangle = V|H\rangle, \tag{3.5}$$

*with $V = \begin{bmatrix} \langle c_1| \\ \vdots \\ \langle c_{4^n}| \end{bmatrix}$, where $|c_i\rangle = \left|U^\dagger P_i U\right\rangle$ is the engineered Hamiltonian state of Pauli $P_i$.*

*Proof.* For a given Hamiltonian $H$, we can decompose it into a linear combination of Pauli terms, i.e., $H = \sum_i h_i P_i$. The engineered Hamiltonian by unitary $U$ is $H' = UHU^\dagger = \sum_i h_i' P_i$. The engineered coefficients are obtained

by following

$$h'_i = \frac{1}{2^n}\mathrm{Tr}[UHU^\dagger P_i] \tag{3.6}$$
$$= \frac{1}{2^n}\mathrm{Tr}[\sum_j h_j P_j U^\dagger P_i U] \tag{3.7}$$
$$= \frac{1}{2^n}\mathrm{Tr}[\sum_j h_j P_j \sum_m c_{im} P_m] \tag{3.8}$$
$$= \frac{1}{2^n}\sum_{j,m} h_j c_{im}\mathrm{Tr}[P_j P_m] \tag{3.9}$$

Since $\mathrm{Tr}[P_j P_m] = 2^n\delta_{jm}$, where $P_j$ and $P_m$ are Pauli matrices, $n$ is number of qubits. Then

$$h'_i = \sum_j h_j c_{ij} = \langle c_i|H\rangle, \tag{3.10}$$

where $|c_i\rangle = \left|U^\dagger P_i U\right\rangle$ is the engineered Hamiltonian state of Pauli operator $P_i$. Here we can construct an operator $V$

$$V = \begin{bmatrix} \langle c_1| \\ \langle c_2| \\ \vdots \\ \langle c_{4^n}| \end{bmatrix} \tag{3.11}$$

such that

$$|H'\rangle = V|H\rangle\,. \tag{3.12}$$

Since unitary $U$ preserves the $l_2$-norm, so **(1)** $\langle c_i|c_i\rangle = 1$. Because

$$\mathrm{Tr}[UP_iU^\dagger UP_jU^\dagger] = \mathrm{Tr}[P_iP_j] = 0 \quad \text{for} \quad i \neq j,$$

then

$$\mathrm{Tr}[\sum_m c_{im}P_m \sum_t c_{jt}P_t] = \sum_{m,t} c_{im}c_{jt}\mathrm{Tr}[P_mP_t] = 0.$$

Since $\mathrm{Tr}[P_mP_t] = 2^n\delta_{mt}$, it becomes $\sum_m c_{im}c_{jm} = 0$, which implies **(2)** $\langle c_i|c_j\rangle = 0, \forall i \neq j$. The two properties imply that the matrix $V$ is unitary. □

This theorem informs us that engineering Hamiltonian is equivalent to applying a unitary $V$ on the Hamiltonian vector. Specifically, from definition 3.1, the Hamiltonian state requires normalizing, so it is straightforward to have the relation between Pauli norm of Hamiltonian $\|H\|_P$ and the $l_1$-norm of Hamiltonian state

$$\|UHU^\dagger\|_P = \lambda\|V|H\rangle\|_1. \tag{3.13}$$

Since unitary preserves $l_2$-norm, meaning $\lambda$ is a constant, the problem of minimizing Pauli norm $\|H\|_P$ is equivalent to minimizing $l_1$-norm of Hamiltonian state $\| |H\rangle \|_1$, which is

$$P_{\text{opt}}(H) = \lambda \min_V \|V |H\rangle\|_1. \tag{3.14}$$

Thus, when the optimal unitary $V$ is found to minimize the $l_1$-norm of $V|H\rangle$, it is equivalent to the optimal $U$ such that $\|UHU^\dagger\|_P$ is minimized. The specific relation between $U$ and $V$ is shown in Sec. 3.3.3.

Note $V$ is unitary (linear operation), so we can derive the following properties.

**Corollary 3.1.** *The properties of operation $V$*

$$(1) \quad \left|U_2U_1HU_1^\dagger U_2^\dagger\right\rangle = V_2V_1 |H\rangle \tag{3.15}$$

$$(2) \quad (\alpha V_1 + \beta V_2)|H\rangle = \alpha V_1 |H\rangle + \beta V_2 |H\rangle \tag{3.16}$$

$$(3) \quad \left|U_1 \otimes U_2HU_1^\dagger \otimes U_2^\dagger\right\rangle = V_1 \otimes V_2 |H\rangle \tag{3.17}$$

*Proof.* For property (1), we can denote the $\tilde{H} = U_1HU_1^\dagger$, then $\left|U_2U_1HU_1^\dagger U_2^\dagger\right\rangle = \left|U_2\tilde{H}U_2^\dagger\right\rangle = V_2\left|\tilde{H}\right\rangle$. Apply the theorem again, $V_2\left|\tilde{H}\right\rangle = V_2\left|U_1HU_1^\dagger\right\rangle = V_2V_1|H\rangle$. For property (2), it is obtained straightforwardly from the linearity of unitary $V$. For property (3), suppose $H = \sum_{ij} h_{ij}P_i \otimes P_j$, then we have

$$U_1 \otimes U_2HU_1^\dagger \otimes U_2^\dagger = \sum_{ij} h_{ij}U_1P_iU_1^\dagger \otimes U_2P_jU_2^\dagger \tag{3.18}$$

where $i, j \in \{1, 2, 3, 4\}$ such that $P_i, P_j$ corresponds to Pauli operators $\{I, X, Y, Z\}$ respectively. From Theorem 3.1, we have $UP_iU^\dagger = V\vec{a}_i$, where $\vec{a}_i$ refers to the basis vector, i.e., only position $i$ in $\vec{a}_i$ is 1 and other positions are all 0. In another word, $|P_i\rangle = \vec{a}_i$. Correspondingly, $|H\rangle = \frac{1}{\sqrt{\sum_{ij} h_{ij}^2}} \sum_{ij} h_{ij}\vec{a}_i \otimes \vec{a}_j$. Now, we have

$$U_1 \otimes U_2HU_1^\dagger \otimes U_2^\dagger = \sum_{ij} h_{ij}V_1\vec{a}_i \otimes V_2\vec{a}_j \tag{3.19}$$

$$= V_1 \otimes V_2 \sum_{ij} h_{ij}\vec{a}_i \otimes \vec{a}_j. \tag{3.20}$$

If we take Hamiltonian vectorization on both sides, it becomes

$$\left|U_1 \otimes U_2HU_1^\dagger \otimes U_2^\dagger\right\rangle = V_1 \otimes V_2 |H\rangle. \tag{3.21}$$

Note the fact that unitary is $l_2$- norm preserving operation, so the normalization factors are canceled out in this equation. □

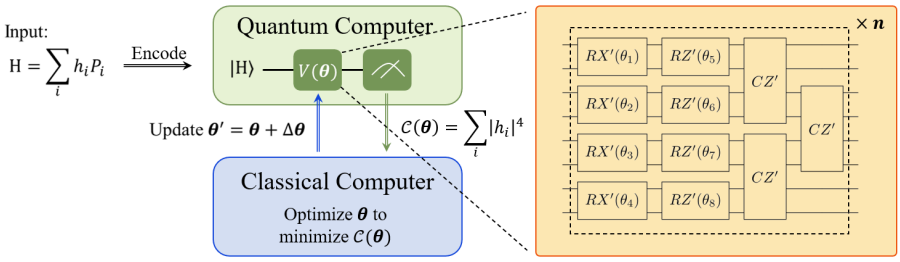



Figure 3.1: Diagram of the variational quantum algorithm for Hamiltonian engineering. Encode the given Hamiltonian $H$ into quantum state $|H\rangle$, then apply parameterized quantum circuit $V(\boldsymbol{\theta})$. Note that the rotation gates $RX'(\theta), RZ'(\theta)$, and control-Z gates $CZ'$ are transferred gates, the exact circuit representation is shown in Figure. 3.4. Measuring the circuit and calculating the cost function, which is the $l_1$-norm of the quantum state, the classical computer optimizes the cost function by updating the parameters in the PQC. After iterations, the cost function converges to the optimized value.

## 3.3 Variational quantum algorithms

It is well-known that due to the non-convexity of unitary, it is very hard to obtain the optimal unitary to satisfy our requirement as shown in Eq. (3.14). Here we propose to apply the variational quantum algorithm (VQA) to approach the optimal unitary. The workflow of our algorithm is shown in Figure. 3.1. At the beginning, we encode the Hamiltonian into a quantum state. Randomize the parameters in the PQC, and estimate the cost function by measurement. Then we use a classical computer to optimize the parameters in the PQC and update the parameters correspondingly. Repeat this process until the cost function is minimized.

In this section, we are going to present the details of the cost function and the choice of PQC in the algorithm design. The specific algorithm is also provided as shown in Algorithm 1.

**Algorithm 1** Variational quantum Hamiltonian engineering (VQHE)

**Input:** Classical description of $n$-qubit Hamiltonian $H = \sum_i h_i P_i$, parameterized quantum circuit (PQC) $V(\boldsymbol{\theta})$, number of iterations ITR,

**Output:** The optimized parameters $\boldsymbol{\theta}^*$, engineered Hamiltonian state $|H'\rangle$

1: Calculate the normalization factor of the Pauli coefficients $\lambda = \sum_i |h_i|$
2: Encode Hamiltonian $H$ into quantum state $|H\rangle$
3: Randomly initialize parameters $\theta$
4: **for** itr $= 1, \ldots,$ ITR **do**
5: Apply $V(\boldsymbol{\theta})$ to the Hamiltonian state$|H\rangle$
6: Apply the estimation algorithm (Sec 3.3.2) to estimate the cost function in Eq. (3.23)
7: Maximize the cost function $\mathcal{C}$ and update parameter $\boldsymbol{\theta}$
8: **end for**
9: Output the optimized parameters $\boldsymbol{\theta^*}$, and the engineered Hamiltonian state $|H'\rangle$

### 3.3.1 Cost function

We have shown that the complexities of many tasks depend on the Pauli norm of the Hamiltonian, such as expectation value estimation [102] and Hamiltonian simulation [96]. We then convert the problem of minimizing the Pauli norm (Eq. (3.3)) to the problem of minimizing the $l_1$-norm of Hamiltonian vector (Eq. (3.14)). It is natural to consider the $l_1$-norm of the corresponding Hamiltonian state as the cost function, i.e., $\| |H\rangle \|_1$. However, estimating the $l_1$-norm of a quantum state is generally hard on quantum devices. Fortunately, there is a quantum algorithm that can efficiently estimate the sum of the fourth power of elements in the Hamilton vector $|H\rangle$, i.e.,

$$Q = \sum_{i=0}^{d-1} |\langle i|H\rangle|^4 = \sum_{i=0}^{d-1} |h_i|^4 \tag{3.22}$$

where $d$ is the dimension of the system. The detail of the algorithm to estimate the quantity $Q$ is shown in the Sec 3.3.2. This algorithm is also used in inverse participation ratio estimation [106]. Both $l_1$-norm and $Q$ describe the uncertainty of a quantum state. From this point of view, minimizing $l_1$-norm and maximizing $Q$ have the same optimizing direction, which reduces the uncertainty of a quantum state. Thus, we set $Q$ as the cost function, i.e.,

$$\mathcal{C} = Q = \sum_{i=0}^{d-1} |h_i'|^4 \tag{3.23}$$

Although minimizing $l_1$-norm and maximizing $Q$ is not exactly the same problem, the effectiveness of setting $Q$ as the cost function emerges in the numerical experiment in Sec. 3.4.4.

### 3.3.2 Quantum algorithm for estimating cost function

We denote the summation of the fourth power of elements in an $n$-qubit pure state $|\psi\rangle = (h_0, \ldots, h_{d-1})^T$ as $Q$, which is

$$Q = \sum_{i=0}^{d-1} |\langle i|\psi\rangle|^4 = \sum_{i=0}^{d-1} |h_i|^4 \tag{3.24}$$

where $|i\rangle$ refers to the computational basis. There exists an efficient quantum algorithm to estimate the quantity $Q$, which is shown in Figure. 3.2.

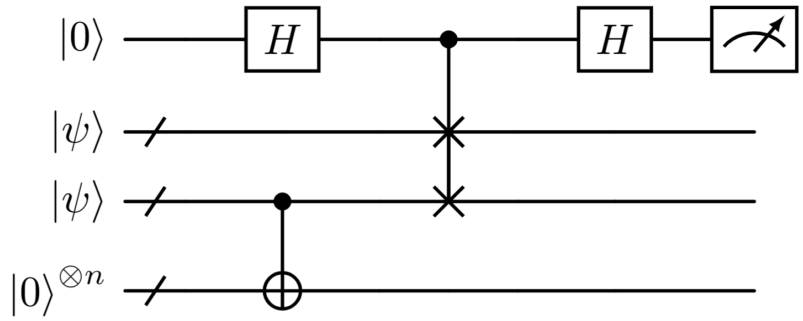



Figure 3.2: Quantum circuit for estimating $Q$. The quantum circuit consists of $n$ CNOT gates and $n$ controlled-SWAP gates. $Q$ is estimated through single-qubit measurements on the top qubit.

The first step of the algorithm is to initialize the quantum state. We need four registers in total. The first register is called the measurement register, which is initialized as $|0\rangle$. The second and third registers are prepared as the $n$-qubit quantum state $|\psi\rangle$. The last register is prepared as $n$-qubit zero state $|0\rangle^n$. At the beginning, we have

$$\begin{aligned} |\Psi\rangle &= |0\rangle \otimes |\psi\rangle \otimes |\psi\rangle \otimes |0\rangle^{\otimes n}, \\ &= \sum_i c_i |0\rangle \otimes |\psi\rangle \otimes |i\rangle \otimes |0\rangle^{\otimes n}. \end{aligned} \tag{3.25}$$

Apply CNOT gates on the third and fourth register, we then have

$$CNOT_{3,4}|\Psi_0\rangle = \sum_i c_i |0\rangle \otimes |\psi\rangle \otimes |i\rangle \otimes |i\rangle. \tag{3.26}$$

At this step, we can discard the fourth register and the quantum state on the third register becomes a mixed state $\rho = \sum_i |c_i|^2 |i\rangle\langle i|$. Take the SWAP test on the first three registers, the probability of getting the measurement as $+1$ is

$$P_{+1} = \frac{1}{2} + \frac{\mathrm{Tr}[\rho|\psi\rangle\langle\psi|]}{2} = \frac{1}{2} + \frac{Q}{2}, \tag{3.27}$$

where $Q = \sum_{i=0}^{d-1} |h_i|^4$ is the desired value. In Ref. [106], A similar algorithm is also used in inverse participation ratio estimation.

### 3.3.3 Parameterized quantum circuit

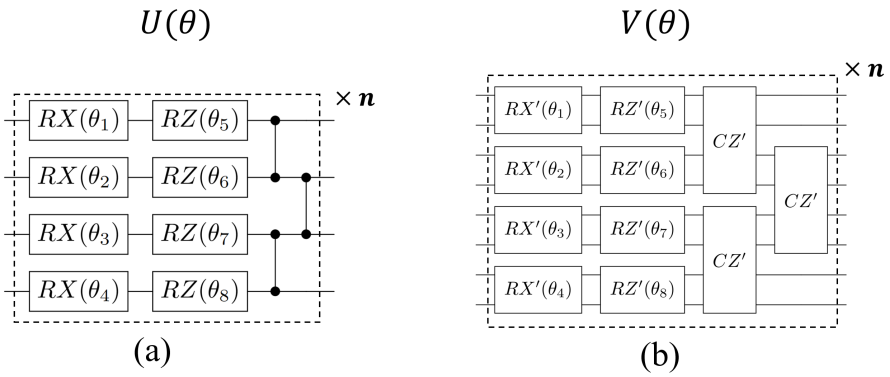



Figure 3.3: Structure of one entangled layer. (a) is the ansatz $U(\boldsymbol{\theta})$ for training Hamiltonian $H$. (b) is the ansatz for training Hamiltonian state $|H\rangle$. The circuits for RX', RY', and CZ' are displayed in Figure. 3.4.

From Theorem 3.1, the encoded unitary $V$ is constructed from the unitary $U$, so $U$ and $V$ have a one-to-one correspondence relation. We can consider unitary $V$ as a function of unitary $U$, i.e., $V = f(U)$. The function $f$ maps an $n$-qubit unitary $U$ into a $2n$-qubit unitary $V$. Here we take the hardware-efficient ansatz for $U$ as shown in Figure. 3.3 (a), and we automatically obtain the corresponding encoded ansatz $V$ as shown in Figure. 3.3 (b). The ansatz for $U$ contains single qubit rotation gates rotation-X (RX), rotation-Z (RZ), and entangled gates control-Z (CZ). The corresponding encoded unitary components $V$ are denoted as RX', RZ', and CZ', whose detailed information is shown in Figure 3.4. The encoded unitaries are all real, thus there is no need to worry about the imaginary element in the optimized Hamiltonian state. There is one thing we need to note that $U(\boldsymbol{\theta})$ and $V(\boldsymbol{\theta})$ share the same parameters $\boldsymbol{\theta}$. Once the parameter optimization in unitary $V(\boldsymbol{\theta})$ is done, we can obtain the unitary $U(\boldsymbol{\theta})$ directly by inserting the parameters $\boldsymbol{\theta}$ into $U$. The unitary $U$ is necessary when we apply the algorithm to expectation value estimation and Hamiltonian simulation, which will be discussed in Sec. 3.4.

## 3.4 Application

In this section, we are going to apply the proposed algorithm to reduce the measurement time in the task of expectation value estimation and the gate count in the task of Hamiltonian simulation, respectively. We especially emphasize

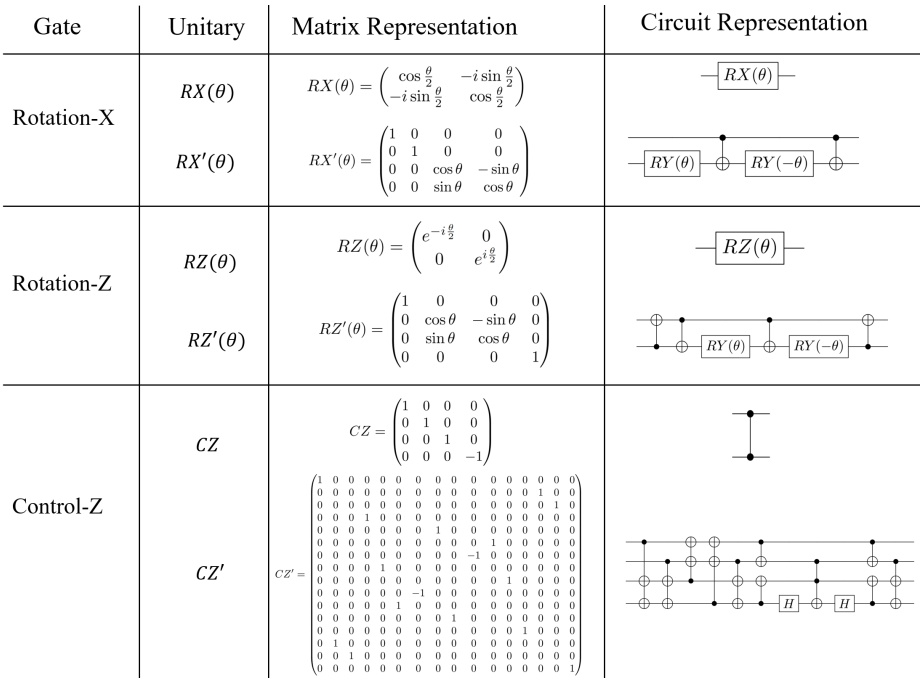



Figure 3.4: Unitary gates $U$ and the corresponding encoded unitary $V$.

that the proposed VQHE is compatible with grouping, which is another method to reduce the measurement complexity. The measurement complexity can be reduced for a further step. We also propose a partition trick to tackle the issue of scalability.

To apply the engineered Hamiltonian $H'$ to the tasks of expectation value estimation and Hamiltonian simulation, we must first obtain the engineered Pauli coefficients $h'_i$. Quantum state tomography [22] is a candidate to get such information from the engineered Hamiltonian state $|H'\rangle$. However, the cost of state tomography increases exponentially with respect to the system size. Fortunately, the engineered Hamiltonian state $|H'\rangle$ is optimized by reducing its uncertainty, hence the complex amplitude is expected to be concentrated on a certain computational basis. In other words, we expect the engineered Hamiltonian state $|H'\rangle$ to be a sparse pure state. In Ref. [107], a more efficient algorithm than quantum state tomography was proposed to reconstruct the sparse pure state. With the engineered Hamiltonian coefficients $h'_i$, the $l_1$-norm of the engineered Hamiltonian norm can be calculated $\|H'\|_P = \sum_i |h_i|$ straightforwardly.

### 3.4.1 Expectation value estimation

In quantum computing, estimating the expectation value of a Hamiltonian $H$ to a quantum state, i.e., $\langle H \rangle = \langle \psi | H | \psi \rangle$, is one of the most fundamental processes in many tasks. The total measurement time is determined by the Pauli norm, i.e., $N \sim \|H\|_P$. In order to reduce the measurement time, we can apply the proposed VQHE algorithm first as shown in Algorithm 1 as a pre-processing, which returns us the engineered Hamiltonian state $|H'\rangle$ and the optimized parameters $\boldsymbol{\theta}^*$. Then estimate the engineered Pauli coefficients $h'_i$ from the engineered Hamiltonian state $|H'\rangle$. The optimized Hamiltonian norm is obtained by summing up all the absolute values of the coefficients, i.e., $\|H'\|_P = \sum_i p_i$.

When we estimate the expectation value of Hamiltonian $H$, we have

$$\langle \psi | H | \psi \rangle = \langle \psi | U^\dagger H' U | \psi \rangle , \tag{3.28}$$

where $H' = UHU^\dagger$ is the engineered Hamiltonian. Here we set the unitary $U$ as $U(\boldsymbol{\theta}^*)$, where $U(\boldsymbol{\theta}^*)$ is the Ansatz shown in Figure. 3.3 (a), and the parameters $\boldsymbol{\theta}^*$ are optimized by Algorithm 1. In this case, the corresponding measurement time of estimating Hamiltonian $H$ reduces to $N \sim \|U(\boldsymbol{\theta}^*)HU^\dagger(\boldsymbol{\theta}^*)\|_P$.

The measurement time can be reduced further by applying a grouping algorithm to the engineered Hamiltonian. Grouping [24, 102–104] is an efficient classical algorithm to reduce the measurement time in the task of expectation value estimation. The main idea of grouping is dividing the Pauli terms into collections, such that all the Pauli elements in a collection commute with each other, thus they can be measured simultaneously. Many grouping methods have been proposed, such as qubit-wise commuting (QWC) [24], general commuting (GC) [103, 104], and sorted insertion [102]. Here we focus on the sorted insertion strategy, which is described as follows, and the details of QWC and GC can be found in Appendix A.1 and Appendix A.2.

For a given Hamiltonian as shown in Eq. (3.2), the set $\{(h_i, P_i)\}_{i=1}^{4^n}$ is sorted by the absolute value of coefficients $h_j$ so that $|h_1| \geq |h_2| \geq \cdots \geq |h_{4^n}|$. Then, in the order $i = 1, \cdots 4^n$, it is checked whether $P_i$ commutes with all elements in an existing collection. If it does, it is added to that collection. If not, a new collection is created and $h_i P_i$ is inserted there. The collections are checked in order of their creation. After division, there are $N$ collections in total, and the $i$-th collection has $m_i$ Pauli terms. Here denote the grouped Pauli norm as $\|H\|_{gp} = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{m_i} |h_{ij}|^2}$, which is highly dependent on the group strategy. The total measurement time of the grouped collections is proportional to the square of the which is $N \sim \|H\|_{gp}$ [102]. The details and examples are shown in the Appendix A.3. If we apply the VQHE algorithm followed by the sorted grouping algorithm, the corresponding measurement time will be

$$N \sim \|U(\boldsymbol{\theta}^*)HU^\dagger(\boldsymbol{\theta}^*)\|_{gp}. \tag{3.29}$$

### 3.4.2 Hamiltonian simulation with qDrift

Consider a Hamiltonian $H = \sum_{j=1}^{L} h_j H_j$, where $H_j$ is hermitian and normalized, $h_j$ is the weight. For each $H_j$, the unitary $e^{-i\tau H_j}$ can be implemented on quantum computers for any $\tau$. In the Trotter formulae, one divides $U = e^{-itH}$ into segments so that $U = U_r^r$ with $U_r = e^{-itH/r}$ and uses $V_r = \prod_{i=1}^{L} e^{-ith_j H_j/r}$ to approach $U_r$ in the large $r$ limit. Repeat this process for $r$ times such that $V_r^r \to U$ in the large $r$ limit. It has been proven in Refs. [108, 109] that the total gate count is $G \sim L^3(\Lambda t)^2/2\epsilon$, where $\Lambda := \max_j h_j$ is the magnitude of the strongest term in the Hamiltonian, $\epsilon$ is the error tolerance. The gate count of this method depends on the terms of decomposition, which is not practical for electronic structure Hamiltonians due to its large number of terms.

The quantum stochastic drift protocol (qDrift) [96] was proposed to solve this problem. Each unitary in the sequence is chosen independently from an identical distribution. Denote $\gamma = \sum_i h_j$. The strength $\tau_j$ of each unitary is fixed $\tau_j = t\gamma/G$, which is independent of $h_j$, so we implement gates of the form $e^{-i\tau H_j}$. Then we randomly choose the unitary $e^{-i\tau H_j}$ with probability $p_j = h_j/\gamma$. The full circuit is described by an ordered list of $j$ values $\boldsymbol{j} = \{j_1, j_2, \cdots, j_N\}$ that corresponds to unitary

$$V_{\boldsymbol{j}} = \prod_{k=1}^{G} e^{-i\tau H_{j_k}}, \tag{3.30}$$

which is selected from the product distribution $P_{\boldsymbol{j}} = \gamma^{-G} \prod_{k=1}^{G} h_{j_k}$. The gate count of Hamiltonian simulation by qDrift method is $G \sim \mathcal{O}((\gamma t)^2/\epsilon)$, where $\gamma$ is the Pauli norm of Hamiltonian.

Before simulating a given Hamiltonian $H$, we can first apply the VQHE algorithm to engineer the Hamiltonian such that the Pauli norm of the engineered Hamiltonian $H' = U(\boldsymbol{\theta}^*)HU^\dagger(\boldsymbol{\theta}^*)$ is reduced. Instead of simulating the original Hamiltonian, we can simulate the engineered Hamiltonian. Because of Eq. (3.31)

$$e^{-iHt} = U^\dagger(\boldsymbol{\theta}^*)e^{-iH't}U(\boldsymbol{\theta}^*), \tag{3.31}$$

we need to sandwich the Hamiltonian evolution by unitaries. The corresponding gate count of simulating the engineered Hamiltonian is

$$G \sim \mathcal{O}((\gamma' t)^2/\epsilon) + \tilde{\mathcal{O}}(U(\boldsymbol{\theta}^*)), \tag{3.32}$$

where $\gamma' = \|U(\boldsymbol{\theta}^*)HU^\dagger(\boldsymbol{\theta}^*)\|_P$, and $\tilde{\mathcal{O}}(U(\boldsymbol{\theta}^*))$ is the gate count of the unitary, which can be ignored because it is a constant. The Eq. (3.31) can be proved directly as shown in the following

$$
\begin{aligned}
U^\dagger e^{-iH't}U &= U^\dagger e^{-iUHU^\dagger t}U \\
&= U^\dagger \sum_n \frac{(-it)^n}{n!}(UHU^\dagger)^n U \\
&= U^\dagger \sum_n \frac{(-it)^n}{n!}(UHU^\dagger UHU^\dagger \cdots UHU^\dagger)U \\
&= U^\dagger \sum_n \frac{(-it)^n}{n!}(UH^nU^\dagger)U \\
&= U^\dagger U \sum_n \frac{(-it)^n}{n!}(H^n)U^\dagger U \\
&= e^{-iHt}.
\end{aligned}
$$

The second and last equations come from the Taylor expansion.

### 3.4.3 Scalability

One obstacle to variational quantum algorithms is barren plateaus [110–112], which refers to the phenomenon that the gradient decreases exponentially with the increase of the quantum system. To solve this problem, we propose a straightforward trick called *partition trick*. Briefly speaking, we can divide the Hamiltonian into parts, where they have common factors. For each part, we extract the common factors and apply the VQHE to the rest Hamiltonian. We take the Hamiltonian $H = 3XIXYYZ + 2XIXZYI - XIXIIZ + IZXIYI + 2IZZIYI$ as an example. We can first divide the Hamiltonian into two parts, and extract the common factor, thus we have $H_1 = XIX \otimes (3YYZ + 2ZYI - IIZ)$ and $H_2 = (IZX + 2IZZ) \otimes IYI$. Then we apply VQHE to minimize $3YYZ + 2ZYI - IIZ$ and $IZX + 2IZZ$, respectively. The engineered Hamiltonian will be $H' = H_1' + H_2'$ with $H_1' = I \otimes U_1 \cdot H_1 \cdot I \otimes U_1^\dagger$ and similarly $H_2' = U_2 \otimes I \cdot H_2 \cdot U_2^\dagger \otimes I$.

The partition trick is compatible with estimating expectation value. When we estimate the expectation value, we have

$$
\langle\psi| H |\psi\rangle = \sum_i \langle\psi| H_i |\psi\rangle = \sum_i \langle\psi| U_i^\dagger H_i' U_i |\psi\rangle \tag{3.33}
$$

where $H_i$ is the partite sub-Hamiltonian, and $H_i' = U_i H_i U_i^\dagger$ is the engineered sub-Hamiltonian. We should estimate the expectation values of each sub-Hamiltonian first, then sum them up. In such a case, the measurement time is $N \sim \sum_i \|H_i'\|_P$. In the task of Hamiltonian simulation, Eq. (3.31) doesn't hold anymore, for the unitaries $U_i$ are not the same for different partition Hamiltonian $H_i$. The partition trick is not compatible with Hamiltonian simulation.

The partition trick gets rid of barren plateaus by dividing a big optimization problem into several small ones, which helps us find the effective unitary to

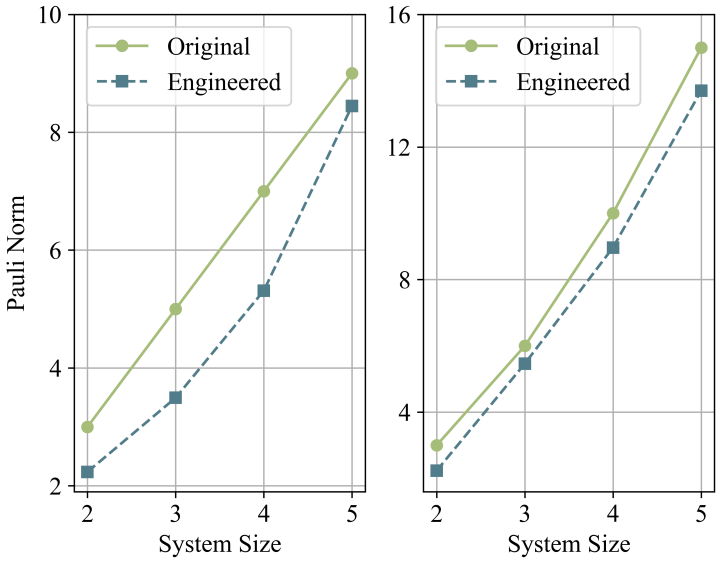



Figure 3.5: The Pauli norm comparison of different sizes of neighbor Ising Hamiltonian (left) and all-to-all Ising Hamiltonian (right). The solid line stands for the original Pauli norm, and the dashed line is the engineered Pauli norm.

reduce the Pauli norm. However, such a trick will also reduce the expressibility of the ansatz, making us unlikely to obtain the optimal unitary to engineer Hamiltonian. This trick can be understood as a trade-off between effectiveness and optimality.

### 3.4.4 Numerical experiments

In this part, we are going to show the effectiveness of the proposed VQHE, which can reduce the Pauli norm of Hamiltonian. Specifically, we apply VQHE to the Ising Hamiltonian and the molecules' Hamiltonian.

Ising model, a mathematical model of ferromagnetism in statistical mechanics, involves representing a material as a lattice of discrete points, where each point or "spin" can be in one of two possible states, typically denoted as "up" or "down." The spins interact with their nearest neighbors, and the interactions are characterized by a coupling constant. The Hamiltonian of the Ising model describes the energy of the system and is given by the sum of the interactions between neighboring spins. The basic form of the Hamiltonian for a one-dimensional Ising model is $H = -J_{ij} \sum_{\{i,j\}} \sigma_i \sigma_j - g_k \sum_k \sigma_k$ where $\{i, j, k\}$ refers to the sit.

In the first numerical experiment, we apply the VQHE algorithm to the Ising model with different system sizes and compare the original Pauli norm

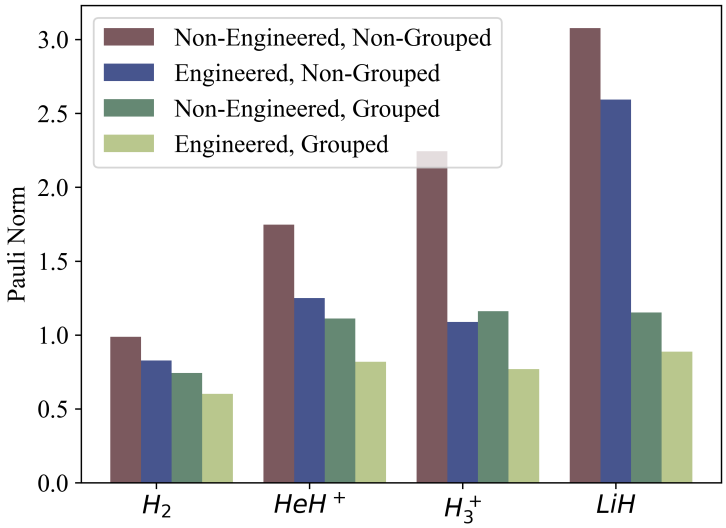



Figure 3.6: The Pauli norm of different Hamiltonian of molecules, including $H_2$ [24], $HeH^+$, $H_3^+$ [113] and $LiH$ [24]. We compare the Pauli norm of the original Pauli norm, engineered Pauli norm, grouped Pauli norm [102], and engineered grouped Pauli norm

and the engineered Pauli norm. For convenience, we set the coefficients in the Ising Hamiltonian to be the same, i.e., $J_{ij} = g_k = 1$. In this experiment, we consider two Hamiltonians which are *neighbor Ising Hamiltonian* $H_{\text{nei}}$ ($j = i+1$) and *all-to-all Ising Hamiltonian* $H_{\text{all}}$ ($i$ and $j$ are independent), respectively. Specifically, we have

$$H_{\text{nei}} = -\sum_{\{i,i+1\}} Z_i Z_{i+1} + \sum_k X_k \tag{3.34}$$

$$H_{\text{all}} = -\sum_{i,j} Z_i Z_j + \sum_k X_k, \tag{3.35}$$

where $X$ and $Z$ refer to Pauli-X and Pauli-Z operators, respectively. From the numerical results (Figure. 3.5), the engineered Hamiltonian shows advantages over the original Hamiltonian with smaller Pauli norms in both cases of neighbor Ising Hamiltonian and all-to-all Ising Hamiltonian.

We also conduct numerical experiments on more complicated systems, including $H_2$, $HeH^+$, $H_3^+$, and $LiH$ molecules. In this experiment, we apply VQHE to the tapered Hamiltonian [114, 115] of molecules for simplicity. The qubit tapering approach encodes the original quantum system into a smaller system, which preserves the ground state and ground energy. Thus, the tapered

Hamiltonian reduces the requirement of the number of qubits, making the variational quantum eigensolver (VQE) [24] more efficient. By applying VQHE to the tapered Hamiltonian of molecules, one observes the engineered Hamiltonian has a smaller Pauli norm compared with the original one. For a further step, we also take the sorted insertion grouping strategy into consideration. Specifically, we engineer the Hamiltonian with VQHE first, then apply the sorted insertion grouping method to the engineered Hamiltonian. The numerical results in Figure. 3.6 show that the engineered grouped Pauli norm of Hamiltonian $\|H'\|_{gp}$ can be reduced further compared with that of engineered Hamiltonian $\|H'\|_P$ and that of grouped Hamiltonian $\|H\|_{gp}$. These results emphasize that the proposed VQHE is compatible with the grouping strategies.

## 3.5 Conclusion

In this work, we propose the VQHE algorithm to reduce the Pauli norm of Hamiltonian such that the overheads for expectation value and Hamiltonian simulation can be reduced. We first develop the theory to convert the Pauli norm optimization problem into the vector $l_1$-norm minimization problem, and then design the cost function and parameterized quantum circuits to minimize Pauli norm variationally. We then display how to apply the proposed VQHE algorithm to expectation value estimation and Hamiltonian simulation. In the task of expectation value estimation, we also emphasize that the proposed algorithm is compatible with grouping, such that the measurement time can be reduced further. The numerical experiments are conducted by applying the VQHE algorithms to the Ising Hamiltonian and some molecules' Hamiltonian, which shows the effectiveness of VQHE algorithm.

For further research, it would also be fun to find the optimal Pauli norm of Hamiltonian we could obtain. It will also be interesting to apply the proposed algorithm VQHE to other applications in quantum computing. Besides, this framework may also be applied to tensor networks, empowering classical computers to solve quantum problems.

# Chapter 4

# Retrieving Non-Linear Features from Noisy Quantum States

In this chapter, we focus on actually extracting high-order moments of quantum states. But in reality, inevitable quantum noise prevents us from accessing the desired value. We address this issue by systematically analyzing the feasibility and efficiency of extracting high-order moments from noisy states. We establish a method for deriving protocols using quantum operations and classical post-processing only. Our protocols, in contrast to conventional ones, incur lower sampling overheads and have simple workflows, making the protocols strong candidates for practical use on current quantum devices. The proposed method also indicates the power of entangled protocols in retrieving high-order information, whereas in the existing methods, entanglement does not help. We further construct the protocol for large quantum systems to retrieve the depolarizing channels, making the proposed method scalable. This chapter is based on [Zhao, Benchi, et al. PRX Quantum 5 (2), 020357 (2024)] with slight modifications to fit in the context.

## 4.1 Introduction

Quantum computing has emerged as a rapidly evolving field with the potential to revolutionize the way we process and analyze information. Such an advanced computational paradigm stores and manipulates information in a quantum state, which forms an elaborate representation of a many-body quantum system [116]. One critical task for this purpose is to estimate the $k$-th *moment* of a quantum state's density matrix $\rho$, which is often denoted as $\mathrm{Tr}[\rho^k], k \in \mathbb{Z}^+$. For example, the second moment of $\rho$ is commonly known as the *purity* of $\rho$. Accurately computing $\mathrm{Tr}[\rho^k]$ provides an elementary precondition for ex-

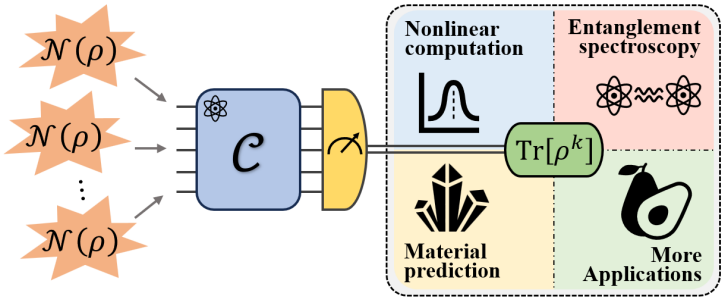



Figure 4.1: The general framework of recovering the high-order quantum information $\mathrm{Tr}[\rho^k]$ given copies of noisy resource $\mathcal{N}(\rho)$ based on our derived protocol, i.e., a quantum channel $\mathcal{C}$ and measurement-based post-processing. The information can be further employed in various applications in practical quantum computing.

tracting spectral information of the quantum state [22, 117], which is crucial in supporting the evaluation of non-linear functions in quantum algorithms [83, 118], applying to entanglement spectroscopy by determining measures of entanglement, e.g., *Rényi entropy* and *von Neumann entropy* [119, 120], and characterizing non-linear features of complex quantum systems in materials [121–124]. In particular, as a core-induced development, understanding and controlling quantum entanglement inspire various quantum information breakthroughs, including entanglement theories, quantum cryptography, teleportation and discrimination [42, 125–127].

Numerous methods have been proposed for efficiently estimating quantum state spectra on a quantum computer, including the deterministic quantum schemes processing intrinsic information of the state [128] and the variational quantum circuit learning for approximating non-linear quantum information functions [129, 130]. Meanwhile, a direct estimation method of $\mathrm{Tr}[\rho^k]$ through the Newton-Girard method and *generalized swap trick* [119] has been proposed in [123], and then it was further improved by [131].

However, quantum systems are inherently prone to the effects of noise, which can arise due to a variety of factors, such as imperfect state preparation, coupling to the environment, and imprecise control of quantum operations [132]. In definition, quantum noise can be described in a language of quantum operation denoted as $\mathcal{N}$. Such an operation can inevitably pose a significant challenge to the reliable estimation of $\mathrm{Tr}[\rho^k]$ from corrupted copies of quantum state $\mathcal{N}(\rho)$.

Previous works concentrated on the first-order situation by applying the inverse operation $\mathcal{N}^{-1}$ [54, 62, 133, 134] to each copy of the noisy state, such that $\mathcal{N}^{-1} \circ \mathcal{N} = \mathrm{id}$, where id means identity map. Such inverse operation might not be physically implementable, which requires the usage of the quasi-probability decomposition (QPD) and sampling techniques [135], decomposing $\mathcal{N}^{-1}$ into

a linear combination of quantum channels $\mathcal{N}^{-1} = \sum_i c_i \mathcal{C}_i$, where $c_i$ are real coefficients and $\mathcal{C}_i$ are quantum channels. Then, we have to sample and prepare the quantum channels with certain probabilities, i.e., $p_i \propto |c_i|$. Sampling and preparing different channels on quantum device is very complicated. To simplify the process, the authors proposed that a Hermitian preserving map $\mathcal{N}^{-1}$ can be realized by injecting different quantum states [63], $\omega_i$ into a fixed quantum channel $\Lambda$, i.e., $\mathcal{N}^{-1}(\cdot) = \sum_i c_i \Lambda(\cdot \otimes \omega_i)$. The new proposed method [63], no doubt, simplifies the realization of Hermitian preserving maps from sampling channels into sampling states. However, we still need to take the process of probabilistic sampling when realizing Hermitian preserving channels. The total required sampling times for both methods (sample channel and sample states) are the same, which are square proportional to *sampling overhead* $g = \sum_i |c_i|$ [136]. Nevertheless, the situations for estimating $\mathrm{Tr}[\rho^k]$ with $k > 1$ stay unambiguous apart from handling individual state noise. In this chapter, we are going to retrieve the $k$-th moment from noisy states, which is illustrated in Figure. 4.1. To systematically analyze the feasibility and efficiency of extracting high-order moment information from noisy states, as shown in Figure. 4.1, The following two questions are addressed:

1. *Under what conditions can we retrieve the high-order moments from noisy quantum states?*

2. *For such conditions, can we obtain a better sampling complexity than existing method?*

These two questions address the existence and efficiency of quantum protocols for retrieving high-order moment information and essential properties from noisy states, which help us to access accurate non-linear feature estimations.

In the present study, we aim to address both of these questions. For the first question, we establish a necessary and sufficient condition for the retrieval of high-order moments from noisy states, which states that a quantum protocol can achieve this goal if and only if the noisy channel is invertible. Regarding the second question, we propose a method called *observable shift* for deriving protocols using quantum operations and classical postprocessing only. In contrast to the conventional sampling techniques, our protocol only employs one quantum operation and avoid quasi-probability decomposition. Besides, our method requires smaller sampling complexity compared to quasi-probability decomposition. We further construct a protocol for large quantum systems to retrieve the depolarizing channels, making the observable shift method scalable.

We also demonstrate the advantages of our method over existing probabilistic sampling method [62, 63, 133, 135] with step-by-step protocols for some types of noise of common interest. Our protocols incur lower sampling overheads and have simple workflows, serving as strong candidates for practical usage on current quantum devices. The proposed method also indicates the power of entanglement in retrieving high-order information, whereas in the existing methods, entangled protocols do not help [54, 63]. In the end, numerical experiments are performed to demonstrate the effectiveness of our protocol with

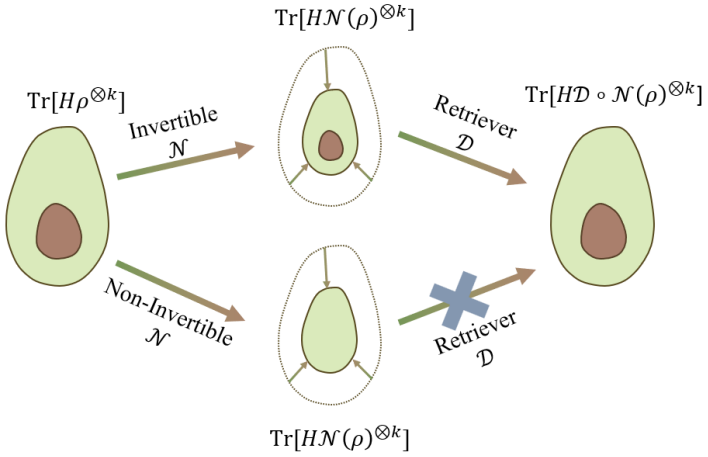



Figure 4.2: Illustration of Theorem 4.1. Suppose a state $\rho$ is corrupted by an invertible channel $\mathcal{N}$, and $H$ is the moment observable, such that $\mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]$. The state information is deformed but can be retrieved via applying $\mathcal{D}$ and post-processing (Top). However, for non-invertible $\mathcal{N}$, the high-order moment is completely destroyed and cannot be retrieved (Bottom).

depolarizing noise applied on the ground state of the Fermi-Hubbard model. Our sampling results illustrate a more accurate estimation on $\mathrm{Tr}[\rho^2]$ compared with no protocol applied.

## 4.2 Moment recoverability

In this section, we are going to address the first question proposed in the introduction. We discover a necessary and sufficient condition for the existence of a high-order moment extraction protocol as shown in Theorem 4.1.

**Theorem 4.1.** (Necessary and sufficient condition for existence of protocol) *Given a noisy channel $\mathcal{N}$, there exists a quantum protocol to extract the $k$-th moment $\mathrm{Tr}[\rho^k]$ for any state $\rho$ if and only if the noisy channel $\mathcal{N}$ is invertible.*

Intuitively, we can understand Theorem 4.1 from the following aspects. Estimating high-order moment demands complete information about quantum channels. If a noise channel $\mathcal{N}$ is invertible, it means information stored in quantum states is deformed, which can be carefully re-deformed back to the original information with extra resources of noisy states and sampling techniques. However, when the loss of information is unattainable, i.e., the noise is non-invertible. Part of the information stored in the quantum state is destroyed completely, leading to an infeasible estimation problem even with extra quantum resources. An illustration of the theorem is shown in Figure. 4.2.

In the following, we will present a sketch of proof for our main theorem. Starting with the definition of a *quantum protocol* which is usually described as a sequence of realizable quantum operations and post-processing steps used to perform a specific task in the domain of quantum information processing. Mathematically, we say there exists a quantum protocol to retrieve the $k$-th moment from copies of a noisy state $\mathcal{N}(\rho)$ if there exists an operation $\mathcal{D}$ such that

$$\mathrm{Tr}[H\mathcal{D}\circ\mathcal{N}^{\otimes k}(\rho^{\otimes k})] = \mathrm{Tr}[H\rho^{\otimes k}], \tag{4.1}$$

where $H$ is what we call the *moment observable*, as the usage of it is the core of extracting the high-order moment from quantum states, i.e., $\mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]$. For example, in estimating the purity of single-qubit states, the moment observable $H$ is just a SWAP operator correlating two qubits. It is proved in the Appendix B.1 that for any order $k$, there exists such a moment observable $H$ to extract the $k$-th moment information. The inspiration for our proof comes from the method used to simulate Hermitian-preserving maps on quantum devices, which has enjoyed great success in a variety of tasks, such as error mitigation [62, 133, 134], and entanglement detection [42, 137]. We extend our allowed operation $\mathcal{D}$ to the field covering the Hermitian-preserving maps.

If the noisy channel $\mathcal{N}$ is invertible, then there exists the inverse operation of the noisy channel $\mathcal{N}^{-1}$, which is generally a Hermitian-preserving map, and $(\mathcal{N}^{-1})^{\otimes k}$ stands for a feasible solution to the high-order moment retriever. On the other hand, by assuming a Hermitian-preserving map $\mathcal{D}$ satisfying Eq. (4.1) and non-invertible $\mathcal{N}$. In the view of the Heisenberg picture, the adjoint of the maps in Eq. (4.1) satisfies

$$\mathrm{Tr}[(\mathcal{N}^{\otimes k})^\dagger\circ\mathcal{D}^\dagger(H)\rho^{\otimes k}] = \mathrm{Tr}[H\rho^{\otimes k}]. \tag{4.2}$$

It has been proven in [138] that given an observable $O$, a Hermitian-preserving map $\mathcal{M}$ satisfies $\mathrm{Tr}[\mathcal{M}(\rho)O] = \mathrm{Tr}[\rho O]$ for any state $\rho$ if and only if it holds that $\mathcal{M}^\dagger(O) = O$. Thus, we can derive that as long as we find a Hermitian-preserving operation $\mathcal{D}$ such that the condition

$$(\mathcal{N}^{\otimes k})^\dagger\circ\mathcal{D}^\dagger(H) = H \tag{4.3}$$

is satisfied, the problem is solved. Since the effective rank of $H$ is full, whose definition and proof are shown in Appendix B.2, then from the fact that $\mathrm{rank}(B) \leq \min(\mathrm{rank}(A), \mathrm{rank}(B))$, we can deduce that $\mathcal{N}$ is invertible, contradicting to our assumption. This means there exists no quantum protocol for extracting high-order moments when the noise is non-invertible. The detailed proof is given in Appendix B.2.

## 4.3 Observable shift method

In the previous part, we mentioned that applying the inverse operation of a noisy channel $\mathcal{N}^{-1}$ to noisy states simultaneously to mitigate the error is one feasible solution to retrieve high-order moments. However, this channel inverse

method requires exponentially many resources with respect to $k$ to retrieve the $k$-th moment. Also, the implementation of inverse operation $\mathcal{N}^{-1}$ is not quantum device friendly because it has to sample and implement different quantum channels probabilistically.

In this section, we propose a new method called *observable shift* to retrieve high-order moment information from noisy states, which requires only one quantum operation with comparable sampling complexity.

**Lemma 4.1.** (Observable shift) *Given an invertible quantum channel $\mathcal{N}$ and an observable $O$, there exists a quantum channel $\mathcal{C}$, called retriever, and coefficients $t, f$ such that*

$$\mathcal{N}^\dagger \circ \mathcal{C}^\dagger(O) = \frac{1}{f}\left(O + tI\right), \tag{4.4}$$

*where $I$ is identity.*

*Proof.* Since the noise channel $\mathcal{N}$ is invertible, the image of the adjoint map of the noise channel $\mathcal{N}^\dagger(\mathcal{L})$ has full dimension and hence $O \in \mathcal{N}^\dagger(\mathcal{L})$. It was proved in [138] that there exists a HPTS map $\mathcal{D}$ such that $\mathcal{N}^\dagger \circ \mathcal{D}^\dagger(H) = H$. Then we take the trick of observable shift, which is

$$\mathcal{N}^\dagger\left(\mathcal{D}^\dagger(H)\right) = H + tI - tI. \tag{4.5}$$

Since $\mathcal{N}$ is CPTP, its adjoint map $\mathcal{N}^\dagger$ is unital-preserving, implying

$$\mathcal{N}^\dagger \circ \mathcal{D}^\dagger(H + tI) = H + tI. \tag{4.6}$$

Denote $\tilde{\mathcal{D}}^\dagger(H) = \mathcal{D}^\dagger(H) + tI$. Since $\mathcal{D}$ is an HPTS map, thus $\mathcal{D}^\dagger$ is a HPUS, i.e., $\mathcal{D}^\dagger(I) = aI$, where $a$ is a real coefficient. Correspondingly,

$$\tilde{\mathcal{D}}^\dagger(I) = \mathcal{D}^\dagger(I) + tI = (a + t)I \tag{4.7}$$

is also an HPUS map. Due to the fact that the adjoint of an HPUS map is HPTS, meaning $\tilde{\mathcal{D}}$ is an HPTS map. As long as the value $t$ no smaller than the absolute minimum eigenvalue of $\mathcal{D}^\dagger(H)$, i.e., $t \geq |\min\{eig(\mathcal{D}^\dagger(H))\}|$, the map $\tilde{\mathcal{D}}$ reduces to CPTS. Thus, we have

$$\mathcal{N}^\dagger \circ \tilde{\mathcal{D}}^\dagger(H) = H + tI \tag{4.8}$$

Note that any CPTS map and be written as a coefficient $f$ times a CPTP map $\mathcal{C}$, which is $\tilde{\mathcal{D}} = f\mathcal{C}$, meaning

$$f\mathcal{N}^\dagger \circ \mathcal{C}^\dagger(H) = H + tI \tag{4.9}$$

$$\Rightarrow \mathcal{N}^\dagger \circ \mathcal{C}^\dagger(H) = \frac{1}{f}(H + tI) \tag{4.10}$$

which complete the proof. □

We develop this observable shift technique since the expectation of $O + tI$ regarding any quantum states can be computed as $\mathrm{Tr}[O\rho] + t$ during the measurement procedures. Moreover, if one wants to maintain the retrievability of $\mathrm{Tr}[O\rho^k]$ from the noise channel $\mathcal{N}$ with respect to any possible quantum states, then the only change that could be made to the observable is to add constant identity since such a transformation could maintain the original information of $\rho$. Therefore, the trace value can be retrieved via measurements and post-processing. For instance, when we estimate $\mathrm{Tr}[O\rho]$, where $O = I + X + Z$, By skipping the identity, often called *shifting the observable*, the value of $\mathrm{Tr}[O\rho]$ can still be re-derived by post-adding a value of one to the expectation value of the *shifted observable* $O' = X + Z$, i.e., $\mathrm{Tr}[O\rho] = 1 + \mathrm{Tr}[O'\rho]$.

Besides, instead of mitigating noise states individually, our method utilizes entanglement to retrieve the information with respect to the moment observable $H$. Compared with the channel inverse method, the proposed observable shift method requires fewer quantum resources, and its implementation is easier.

In our cases, we aim to find a Hermitian-preserving map $\mathcal{D}$ such that Eq. (4.3) holds together with the allowance of observable shifting (4.4), i.e.,

$$(\mathcal{N}^{\otimes k})^\dagger \circ \mathcal{D}^\dagger(H) = H' - tI, \tag{4.11}$$

where $H' = H + tI$ is the shifted observable, and $t$ is a real coefficient. Note that the quantum channel $\mathcal{N}^{\otimes k}$ is a completely positive and trace preserving (CPTP) map, and the adjoint of a CPTP map is completely positive unital preserving [38], which refers $(\mathcal{N}^{\otimes k})^\dagger(I) = I$. Thus, we have

$$(\mathcal{N}^{\otimes k})^\dagger(\mathcal{D}^\dagger(H) + tI) = H', \tag{4.12}$$

where we can consider $\mathcal{D}^\dagger(H) + tI$ as a whole and denote it as $\tilde{\mathcal{D}}(H)$. With proper coefficient $t$, the map $\tilde{\mathcal{D}}$ could reduce to a completely positive map $\mathcal{C}$.

If we apply the quantum channel $\mathcal{C}$ to a noisy state and make measurements over moment observable $H$, the expectation value will be

$$\zeta = \mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}^{\otimes k}(\rho^{\otimes k})] = \mathrm{Tr}[(\mathcal{N}^{\otimes k})^\dagger \circ \mathcal{C}^\dagger(H)\rho^{\otimes k}] \tag{4.13}$$

$$= \frac{1}{f}\mathrm{Tr}[(H + tI)\rho^{\otimes k}] = \frac{1}{f}(\mathrm{Tr}[H\rho^{\otimes k}] + t). \tag{4.14}$$

Obviously, the desired high-order moment is given by $\mathrm{Tr}[H\rho^{\otimes k}] = f\zeta - t$. By the Hoeffding's inequality [136], obtaining the estimation within an error $\epsilon$ with a probability no less than $1 - p$ is upper bounded by $f^2 \frac{2}{\delta^2} \log\left(\frac{2}{p}\right)$. Thus, the number of total sampling times $T$ is should be

$$T \geq f^2 \frac{2}{\delta^2} \log\left(\frac{2}{p}\right). \tag{4.15}$$

Usually, the success probability $1 - p$ is fixed. Thus, we consider it as a constant in this chapter, and the corresponding sample complexity is $\mathcal{O}(f^2/\delta^2)$, which only depends on error tolerance $\delta$ and the *sampling overhead* $f$. Specifically, we have the following Definition 4.1.

**Definition 4.1.** *Given error tolerance $\delta$, the sufficient sampling complexity of retrieving $k$-th moment information by observable shift method is $\mathcal{O}(f^2(\mathcal{N},k)/\delta^2)$. The sampling overhead $f(\mathcal{N},k)$ is in the following set*

$$f(\mathcal{N},k) \in \left\{ f \,|(\mathcal{N}^{\otimes k})^\dagger \circ \mathcal{C}^\dagger(H) = \frac{1}{f}\left(H + tI\right), f \in \mathbb{R}^+, t \in \mathbb{R}, \mathcal{C} \in \mathrm{CPTP} \right\}, \tag{4.16}$$

*where $\mathcal{N}$ is the noisy channel, $\mathcal{C}$ is quantum channel, $t$ is the shifted distance, $H$ is the moment observable.*

Since sampling complexity depends on the overhead $f$, it is desirable to find a quantum retriever $\mathcal{C}$ and shift distance $t$ to make the sampling overhead $f$ as small as possible, and the minimum sapling overhead $f_{\min}(\mathcal{N},k)$ is defined as

$$f_{\min}(\mathcal{N},k) = \min \left\{ f \,|(\mathcal{N}^{\otimes k})^\dagger \circ \mathcal{C}^\dagger(H) = \frac{1}{f}\left(H + tI\right), f \in \mathbb{R}^+, t \in \mathbb{R}, \mathcal{C} \in \mathrm{CPTP} \right\}, \tag{4.17}$$

which can be calculated by the following SDP

$$\begin{aligned}
f_{\min}(\mathcal{N},k) = \min \quad & f && \text{(4.18a)}\\
\text{subject to} \quad & J_{\tilde{\mathcal{C}}} \geq 0 && \text{(4.18b)}\\
& \mathrm{Tr}_C[J_{\tilde{\mathcal{C}}_{BC}}] = f I_B && \text{(4.18c)}\\
& J_{\mathcal{F}_{AC}} \equiv \mathrm{Tr}_B[(J^{T_B}_{\mathcal{N}^{\otimes k}_{AB}} \otimes I_C)(I_A \otimes J_{\tilde{\mathcal{C}}_{BC}})] && \text{(4.18d)}\\
& \mathrm{Tr}_C[(I_A \otimes H_C^T) J^T_{\mathcal{F}_{AC}}] = H_A + tI_A. && \text{(4.18e)}
\end{aligned}$$

The $J_{\tilde{\mathcal{C}}}$ and $J_{\mathcal{N}^{\otimes k}}$ are the Choi-Jamiołkowski matrices for the completely positive trace-scaling map $\tilde{\mathcal{C}} = f\mathcal{C}$ and noise channel $\mathcal{N}^{\otimes k}$ respectively. Eq. (4.18b) corresponds to the condition that the map $\tilde{\mathcal{C}}$ is completely positive, and Eq. (4.18c) guarantees that $\tilde{\mathcal{C}}$ is a trace-scaling map. In Eq. (4.18d), $J_{\mathcal{F}}$ is the Choi matrix of the composed map $\tilde{\mathcal{C}} \circ \mathcal{N}^{\otimes k}$. Eq. (4.18e) corresponds to the constraint shown in Eq. (4.4). The dual SDP of the original problem is shown as follows

$$\begin{aligned}
f_{\min}(\mathcal{N},k) = \max \quad & -\mathrm{Tr}[KH] && \text{(4.19a)}\\
\text{subject to} \quad & \mathrm{Tr}[M] \leq 1 && \text{(4.19b)}\\
& \mathrm{Tr}[K] = 0 && \text{(4.19c)}\\
& M \otimes I + \mathrm{Tr}_A[(K_A^T \otimes I_B \otimes H_C)(J^{T_B}_{\mathcal{N}^{\otimes k}_{AB}} \otimes I_C)] \geq 0 && \text{(4.19d)}
\end{aligned}$$

Beyond retrieving particular non-linear features, i.e., $\mathrm{Tr}[\rho^k]$, we can also apply our method to estimate non-linear functions. For a toy example, if we wish to estimate the function $F(\rho) = \frac{1}{2}\mathrm{Tr}[\rho^2] + \frac{1}{3}\mathrm{Tr}[\rho^3]$, we should design the moment observable first, which is supposed to be $H = [\frac{1}{2}H_2 \otimes I + \frac{1}{3}H_3]$, where $H_2$ and $H_3$ are the moment observables for two and three qubits respectively.

Then, the retrieving protocol with sampling overhead $f_{\min}(\mathcal{N}, k)$ is given by SDP as shown in Eq. (4.18). With the power of estimating non-linear functions, our method applies to the entropy evaluation from noisy quantum states, showcasing the practical potential in quantum many-body correlation determination and entanglement detection. One can directly assist the estimation of *Rényi entropy* [139] provided a quantum state $\rho$, which is

$$H_\alpha(\rho) := \frac{1}{1-\alpha} \log \left( \mathrm{Tr}[\rho^\alpha] \right), \tag{4.20}$$

where $\alpha \in (0,1) \cup (1,\infty)$.

## 4.4 Protocols for particular noise channels

We have introduced the observable shift method in the previous part, next will provide the analytical protocol for retrieving the second-order moment information $\mathrm{Tr}[\rho^2]$ from noisy quantum states suffering from depolarizing channel and amplitude dimpling channel, respectively.

### 4.4.1 Mitigate depolarizing channel

Depolarizing channels have been extensively studied due to its simplicity and ability to represent a wide range of physical processes that can affect quantum states [22]. A quantum state that undergoes a depolarizing channel would be randomly replaced by a maximally mixed state with a certain error rate. The single-qubit depolarizing (DE) noise $\mathcal{N}^\epsilon_{\mathrm{DE}}$ has an exact form

$$\mathcal{N}^\epsilon_{\mathrm{DE}}(\rho) = (1-\epsilon)\rho + \epsilon \frac{I}{2}, \tag{4.21}$$

where $\epsilon$ is the noise level, and $I$ refers to the identity operator.

The retrieving protocol and retrieving overhead for such noise channel was well studied in [138]. Here we will focus on retrieving the second order information $\mathrm{Tr}[\rho^2]$ from two-copy of depolarizing noised state $\mathcal{N}(\rho \otimes \rho) = \mathcal{N}^\epsilon_{\mathrm{DE}} \otimes \mathcal{N}^\epsilon_{\mathrm{DE}}(\rho \otimes \rho)$, after applying such channel, the state becomes

$$\begin{aligned}
\mathcal{N}(\rho \otimes \rho) &= \mathcal{N}^\epsilon_{\mathrm{DE}}(\rho) \otimes \mathcal{N}^\epsilon_{\mathrm{DE}}(\rho) && (4.22)\\
&= [(1-\epsilon)\rho + \epsilon \frac{I}{2}] \otimes [(1-\epsilon)\rho + \epsilon \frac{I}{2}] && (4.23)\\
&= (1-\epsilon)^2 \rho \otimes \rho + \frac{\epsilon(1-\epsilon)}{2} I \otimes \rho + \frac{(1-\epsilon)\epsilon}{2} \rho \otimes I + \frac{\epsilon^2}{4} I \otimes I. && (4.24)
\end{aligned}$$

Given many copies of such noisy quantum states, our method derives a protocol for retrieving the second-order moment $\mathrm{Tr}[\rho^2]$ using only one quantum channel and post-processing. Specifically, we have Proposition 4.1.

**Proposition 4.1.** *Given two copies of noisy states, $\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes 2}$, and error tolerance $\delta$, the second order moment $\mathrm{Tr}[\rho^2]$ can be estimated by*

$$f_{\min}\mathrm{Tr}[H\mathcal{C}\left(\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes 2}\right)] - t,$$

*with sample complexity $\mathcal{O}(f_{\min}^2/\delta^2)$, where $f_{\min} = \frac{1}{(1-\epsilon)^2}$, $t = \frac{1-(1-\epsilon)^2}{2(1-\epsilon)^2}$. The term $\mathrm{Tr}[H\mathcal{C}\left(\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes 2}\right)]$ can be estimated by implementing a quantum retriever $\mathcal{C}$ on noisy states and making measurements over moment observable $H$. Moreover, there exists an ensemble of unitary operations $\{p_j, U_j\}_j$ such that the action of the retriever $\mathcal{C}$ can be interpreted as*

$$\mathcal{C}(\cdot) = \sum_{j=1} p_j U_j(\cdot)U_j^{\dagger} \tag{4.25}$$

*Proof.* To begin with, we give the explicit form of the retriever $\mathcal{C}$

$$\mathcal{C}(\cdot) = \sum_{j=1}^{12} \frac{1}{12} U_j \cdot U_j^{\dagger}, \tag{4.26}$$

where all the probabilities $p_j$ are the same, i.e., $p_j = \frac{1}{12}$, and the corresponding unitaries $U_j$ are

$$\begin{aligned}
&U_1 = I \otimes I; \quad U_2 = X \otimes X; \quad U_3 = Y \otimes Y; \quad U_4 = Z \otimes Z; \\
&U_5 = \frac{1}{2}\begin{pmatrix} -i & 1 & 1 & i \\ i & 1 & -1 & i \\ i & -1 & 1 & i \\ -i & -1 & -1 & i \end{pmatrix}; \; U_6 = \frac{1}{2}\begin{pmatrix} i & -1 & -1 & -i \\ i & 1 & -1 & i \\ i & -1 & 1 & i \\ i & 1 & 1 & -i \end{pmatrix}; \\
&U_7 = \frac{1}{2}\begin{pmatrix} i & i & i & i \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -i & i & i & -i \end{pmatrix}; U_8 = \frac{1}{2}\begin{pmatrix} -i & i & i & -i \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ i & i & i & i \end{pmatrix}; \\
&U_9 = \frac{1}{2}\begin{pmatrix} i & 1 & 1 & -i \\ -i & 1 & -1 & -i \\ -i & -1 & 1 & -i \\ i & -1 & -1 & -i \end{pmatrix}; U_{10} = \frac{1}{2}\begin{pmatrix} -i & -1 & -1 & i \\ -i & 1 & -1 & -i \\ -i & -1 & 1 & -i \\ -i & 1 & 1 & i \end{pmatrix}; \\
&U_{11} = \frac{1}{2}\begin{pmatrix} i & -i & -i & i \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -i & -i & -i & -i \end{pmatrix}; \; U_{12} = \frac{1}{2}\begin{pmatrix} -i & -i & -i & -i \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ i & -i & -i & i \end{pmatrix}.
\end{aligned} \tag{4.27}$$

To prove estimating the second order moment by implementing quantum channel $\mathcal{C}$ with sampling overhead $f_{\min} = \frac{1}{(1-\epsilon)^2}$, we divide the process into two steps. For the first step, we prove that quantum channel $\mathcal{C}$ is a feasible solution to the prime problem with sampling overhead $\frac{1}{(1-\epsilon)^2}$, meaning $f_{\min} \leq \frac{1}{(1-\epsilon)^2}$.

Then we are going to find one feasible solution to the dual problem with overhead $\frac{1}{(1-\epsilon)^2}$, meaning $f_{\min} \geq \frac{1}{(1-\epsilon)^2}$. Thus, we have overhead $f_{\min} = \frac{1}{(1-\epsilon)^2}$, which complete the proof.

Now, we are going to proof the first part. At first, we need to write the Choi matrix form of the retriever $\mathcal{C}$, which is

$$J_{\mathcal{C}} = \frac{1}{4}IIII + \frac{1}{12}(XX + YY + ZZ) \otimes (XX + YY + ZZ). \tag{4.28}$$

For an arbitrary state $\rho \otimes \rho$, after applying the noise $\mathcal{N}$ onto it, we denote the state as $\rho' = \mathcal{N}(\rho \otimes \rho)$, thus

$$\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho) = \mathcal{C}(\rho') = \mathrm{Tr}_A[(\rho'^T \otimes II)J_{\mathcal{C}}] \tag{4.29}$$

$$= \mathrm{Tr}_A\Big[(\rho'^T \otimes II)(\frac{1}{4}IIII + \frac{1}{12}(XX + YY + ZZ) \otimes (XX + YY + ZZ))\Big] \tag{4.30}$$

$$= \frac{1}{4}\mathrm{Tr}[\rho'^T]II + \frac{1}{12}\mathrm{Tr}[\rho'^T(XX + YY + ZZ)(XX + YY + ZZ)] \tag{4.31}$$

$$= \frac{1}{4}II + \frac{1}{12}\mathrm{Tr}[\rho'^T(XX + YY + ZZ)](XX + YY + ZZ))\Big] \tag{4.32}$$

Note that the above equations utilized the face that transpose operation is trace preserving, i.e., $\mathrm{Tr}[\rho^T] = \mathrm{Tr}[\rho] = 1$. Since the matrix $XX + YY + ZZ$ is symmetry, thus we have

$$\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho) = \frac{1}{4}II + \frac{1}{12}\mathrm{Tr}[\rho'^T(XX + YY + ZZ)^T](XX + YY + ZZ)) \tag{4.33}$$

$$= \frac{1}{4}II + \frac{1}{12}\mathrm{Tr}[\rho'(XX + YY + ZZ)](XX + YY + ZZ)) \tag{4.34}$$

The trace term can be calculated by substituting Eq. (4.24), we have

$$\begin{aligned}\mathrm{Tr}[\rho'(XX + YY + ZZ)] &= \mathrm{Tr}[((1-\epsilon)^2\rho \otimes \rho \\ &+ \frac{\epsilon(1-\epsilon)}{2}I \otimes \rho + \frac{(1-\epsilon)\epsilon}{2}\rho \otimes I + \epsilon^2(I \otimes I))(XX + YY + ZZ)]\end{aligned} \tag{4.35}$$

$$\begin{aligned}=&(1-\epsilon)^2\mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)] \\ &+ \frac{\epsilon(1-\epsilon)}{2}\mathrm{Tr}[X \otimes \rho X + Y \otimes \rho Y + Z \otimes \rho Z] \\ &+ \frac{(1-\epsilon)\epsilon}{2}\mathrm{Tr}[\rho X \otimes X + \rho Y \otimes Y + \rho Z \otimes Z] + \frac{\epsilon^2}{4}\mathrm{Tr}[XX + YY + ZZ]\end{aligned} \tag{4.36}$$

$$=(1-\epsilon)^2\mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)]. \tag{4.37}$$

In the second equation, since $X, Y, Z$ are all traceless Hermitian matrices, all terms are zeros except the first term. Replace the equation back to Eq. (4.34),

then

$$\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho) = \frac{1}{4} II + \frac{(1-\epsilon)^2}{12} \mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)](XX + YY + ZZ). \tag{4.38}$$

The information $\mathrm{Tr}[\rho^2]$ is estimated from $\mathrm{Tr}[H(\rho \otimes \rho)]$, where $H = \frac{1}{2}(II + XX + YY + ZZ)$ is cyclic permutation operator (in the 2-qubit case, $H$ is just SWAP operator). It is easy to check that

$$\begin{aligned} \mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] = \frac{1}{(1-\epsilon)^2} \Big[ & \frac{1}{4} \mathrm{Tr}[H * II] \\ & + \frac{(1-\epsilon)^2}{12} \mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)] \mathrm{Tr}[H * (XX + YY + ZZ)] \Big]. \end{aligned} \tag{4.39}$$

We can quickly get $\mathrm{Tr}[H] = 2$ and $\mathrm{Tr}[H * (XX + YY + ZZ)] = 6$. Then

$$\mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] = \frac{1}{2} + \frac{(1-\epsilon)^2}{2} \mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)] \tag{4.40}$$

$$= (1-\epsilon)^2 \Big[ \frac{1}{2(1-\epsilon)^2} + \frac{1}{2} \mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)] \Big] \tag{4.41}$$

$$= (1-\epsilon)^2 \Big[ \frac{1}{2(1-\epsilon)^2} - \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ)] \Big] \tag{4.42}$$

$$= (1-\epsilon)^2 \Big[ \frac{1}{2(1-\epsilon)^2} - \frac{(1-\epsilon)^2}{2(1-\epsilon)^2} + \frac{1}{2} \mathrm{Tr}[\rho \otimes \rho(XX + YY + ZZ + II)] \Big] \tag{4.43}$$

$$= (1-\epsilon)^2 \Big[ \frac{2\epsilon^2 - \epsilon^2}{2(1-\epsilon)^2} + \mathrm{Tr}[\rho \otimes \rho H] \Big] \tag{4.44}$$

$$= (1-\epsilon)^2 \Big[ \frac{2\epsilon^2 - \epsilon^2}{2(1-\epsilon)^2} + \mathrm{Tr}[\rho^2] \Big]. \tag{4.45}$$

The desired high-order moment $\mathrm{Tr}[\rho^2]$ value equals to

$$\mathrm{Tr}[\rho^2] = f \mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] - t \tag{4.46}$$

where $f = \frac{1}{(1-\epsilon)^2}$ is the sampling overhead and $t = \frac{1-(1-\epsilon)^2}{2(1-\epsilon)^2}$ is the shifted distance. In order to estimate the value with the error $\delta$, the sampling overhead should be $1/(1-\epsilon)^2$. Thus, we have $f_{\min} \leq 1/(1-\epsilon)^2$.

Next, we are going to use dual SDP to show that $f_{\min} \geq 1/(1-\epsilon)^2$. We set the dual variables as $M = \frac{1}{4} II - \frac{1}{12}(XX + YY + ZZ)$, and $K = q(XX + YY + ZZ)$ where $q = -\frac{1}{6(1-\epsilon)^2}$. We will show the variables $\{M, K\}$ is a feasible solution to the dual problem.

If we substitute the variables into the dual problem Eq. (4.19), we can easily

check that $\mathrm{Tr}[M] \leq 1, \mathrm{Tr}[K] = 0$. For the last condition, we have

$$\mathrm{Tr}_A[(K_A^T \otimes I_B \otimes H_C)(J_{\mathcal{N}_{AB}^{\otimes k}}^{T_B} \otimes I_C)] \tag{4.47}$$

$$= -\frac{1}{12}(XX + YY + ZZ)(II + XX + YY + ZZ) \tag{4.48}$$

Thus, we have

$$M \otimes II + \mathrm{Tr}_A[(K_A^T \otimes I_B \otimes H_C)(J_{\mathcal{N}_{AB}^{\otimes k}}^{T_B} \otimes I_C)] \tag{4.49}$$

$$= \frac{1}{4}IIII - \frac{1}{6}(XX + YY + ZZ)II - \frac{1}{12}(XX + YY + ZZ)(XX + YY + ZZ) \tag{4.50}$$

$$\geq 0, \tag{4.51}$$

which means $\{M, K\}$ is a feasible solution to the dual SDP. Therefore, we have $f_{\min} \geq -\mathrm{Tr}[KH] = \frac{1}{(1-\epsilon)^2}$. Combined with prime part, we have $f_{\min} = \frac{1}{(1-\epsilon)^2}$. □

As a result, the second order moment can be retrieved from depolarized states $\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes 2}$ by applying the unitaries $U_j$ randomly with equal probabilities and then performing measurements with respect to the moment observable $H$. After repeating these steps for $T$ rounds, where $T$ is given by Eq. (4.15), and averaging the measurement results, we can obtain the estimated expectation value $\zeta = \mathrm{Tr}[H\mathcal{C}(\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes 2})]$. Then, the desired second-order moment is given by

$$\mathrm{Tr}[\rho^2] = \frac{1}{(1-\epsilon)^2}\zeta - \frac{1-(1-\epsilon)^2}{2(1-\epsilon)^2}. \tag{4.52}$$

When estimating $\mathrm{Tr}[\rho^2]$ from copies of the noisy state $\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes 2}$, conventional methods incurs a sampling overhead $\frac{(1+\epsilon/2)^2}{(1-\epsilon)^2}$. On the other hand, our observable shift method offers a protocol with a lower sampling overhead $\frac{1}{(1-\epsilon)^2}$, which is much lower than that of the conventional method.

### 4.4.2 Mitigate amplitude damping channel

The quantum amplitude damping (AD) channel is another important model that we are interested in, which often appears in superconducting qubits or trapped ions. This type of noise is particularly relevant for the loss of energy or the dissipation of excited states [140], whose action results in the transition of a qubit's excited state to its ground state, offering a more realistic representation of energy relaxation processes in quantum systems. The AD channel $\mathcal{N}_{\mathrm{AD}}^{\varepsilon}$ is characterized by a single parameter $\varepsilon$, representing the damping rate, which has two Kraus operators: $A_0^{\varepsilon} \coloneqq |0\rangle\langle 0| + \sqrt{1-\varepsilon}|1\rangle\langle 1|$ and $A_1^{\varepsilon} \coloneqq \sqrt{\varepsilon}\,|0\rangle\langle 1|$, where $\varepsilon \in [0, 1]$. Similarly, given many copies of AD-produced quantum states, the second-order information $\mathrm{Tr}[\rho^2]$ can be retrieved by applying only one quantum channel and post-processing with our protocol in Proposition 4.2.

**Proposition 4.2.** *Given two copies of noisy states, $\mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho)^{\otimes 2}$, and error tolerance $\delta$, the second order moment $\mathrm{Tr}[\rho^2]$ can be estimated by*

$$f_{\min}\mathrm{Tr}[H\mathcal{C}\left(\mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho)^{\otimes 2}\right)] - t,$$

*with sample complexity $\mathcal{O}(f_{\min}/\delta^2)$, where $f_{\min} = \frac{1}{(1-\varepsilon)^2}$, $t = -\frac{\varepsilon^2}{(1-\varepsilon)^2}$. The term $\mathrm{Tr}[H\mathcal{C}\left(\mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho)^{\otimes 2}\right)]$ can be estimated by implementing a quantum retriever $\mathcal{C}$ on noisy states and making measurements. Moreover, the Choi matrix of such the retriever $\mathcal{C}$ is*

$$\begin{aligned} J_{\mathcal{C}} = &|00\rangle\langle 00| \otimes \frac{1}{6}((1+2\varepsilon)II + (1-4\varepsilon)H) \\ &+ |\Psi^+\rangle\langle\Psi^+| \otimes \frac{1}{6}((1+2\varepsilon)II + (1-4\varepsilon)H) \\ &+ |\Psi^-\rangle\langle\Psi^-| \otimes \frac{1}{2}(II - H) \\ &+ |11\rangle\langle 11| \otimes \frac{1}{6}(II + H), \end{aligned} \tag{4.53}$$

*where $|\Psi^{\pm}\rangle\langle\Psi^{\pm}| = \frac{1}{2}(|01\rangle \pm |10\rangle)(\langle 01| \pm \langle 10|)$ are Bell states.*

*Proof.* The amplitude damping (AD) channel is a physical channel that describes the energy leakage, dropping from a high energy state to a low energy state. The qubit amplitude damping channel $\mathcal{N}_{\mathrm{AD}}^{\varepsilon}$ has two Kraus operators: $A_0^{\varepsilon} \coloneqq |0\rangle\langle 0| + \sqrt{1-\varepsilon}|1\rangle\langle 1|$ and $A_1^{\varepsilon} \coloneqq \sqrt{\varepsilon}\,|0\rangle\langle 1|$. A single qubit state $\begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}$ after going through the AD channel is

$$\mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho) = \begin{pmatrix} \rho_{00} + \varepsilon\rho_{11} & \sqrt{1-\varepsilon}\rho_{01} \\ \sqrt{1-\varepsilon}\rho_{10} & (1-\varepsilon)\rho_{11} \end{pmatrix} \tag{4.54}$$

where $\varepsilon \in [0,1]$ is the damping factor. This part we will focus on retrieving the information $\mathrm{Tr}[\rho^2]$ from two-copy AD channel noised state $\mathcal{N}_{\mathrm{AD}}^{\varepsilon} \otimes \mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho \otimes \rho)$. The noised state is

$$\mathcal{N}(\rho\otimes\rho) = \mathcal{N}_{\mathrm{AD}}^{\varepsilon} \otimes \mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho\otimes\rho) \tag{4.55}$$

$$= \begin{pmatrix} (\rho_{00}+\varepsilon\rho_{11})^2 & (\rho_{00}+\varepsilon\rho_{11})(\sqrt{1-\varepsilon}\rho_{01}) & (\sqrt{1-\varepsilon}\rho_{01})(\rho_{00}+\varepsilon\rho_{11}) & (1-\varepsilon)\rho_{01}^2 \\ (\rho_{00}+\varepsilon\rho_{11})(\sqrt{1-\varepsilon}\rho_{10}) & (\rho_{00}+\varepsilon\rho_{11})(1-\varepsilon)\rho_{11} & (1-\varepsilon)\rho_{01}\rho_{10} & (1-\varepsilon)\sqrt{1-\varepsilon}\rho_{01}\rho_{11} \\ \sqrt{1-\varepsilon}\rho_{10}(\rho_{00}+\varepsilon\rho_{11}) & (1-\varepsilon)\rho_{10}\rho_{01} & (1-\varepsilon)\rho_{11}(\rho_{00}+\varepsilon\rho_{11}) & (1-\varepsilon)\sqrt{1-\varepsilon}\rho_{11}\rho_{01} \\ (1-\varepsilon)\rho_{10}^2 & (1-\varepsilon)\sqrt{1-\varepsilon}\rho_{10}\rho_{11} & (1-\varepsilon)\sqrt{1-\varepsilon}\rho_{11}\rho_{10} & (1-\varepsilon)^2\rho_{11}^2 \end{pmatrix} \tag{4.56}$$

To prove to estimate the second order moment by implementing quantum channel $\mathcal{C}$ with sampling overhead $f_{\min} = \frac{1}{(1-\varepsilon)^2}$, we divide the process into two steps. For the first step, we prove that quantum channel $\mathcal{C}$ is a feasible solution to the prime problem with sampling overhead $\frac{1}{(1-\varepsilon)^2}$, meaning $f_{\min} \leq \frac{1}{(1-\varepsilon)^2}$. Then we are going to find one feasible solution to the dual problem with overhead $\frac{1}{(1-\varepsilon)^2}$, meaning $f_{\min} \geq \frac{1}{(1-\varepsilon)^2}$. Thus, we have the sampling overhead $f_{\min} = \frac{1}{(1-\varepsilon)^2}$, which complete the proof.

Now, we are going to prove the first part. For an arbitrary state $\rho\otimes\rho$, after applying the noise $\mathcal{N}$ onto it, we denote the state as $\rho'=\mathcal{N}(\rho\otimes\rho)$, thus

$$\begin{aligned}
&\mathcal{C}\circ\mathcal{N}(\rho\otimes\rho)=\mathcal{C}(\rho')=\mathrm{Tr}_A[(\rho'^T\otimes II)J_{\mathcal{C}}]\\
&=\mathrm{Tr}_A\Big[(\rho'^T\otimes II)\Big(|00\rangle\langle 00|\otimes\frac{1}{6}((1+2\varepsilon)II+(1-4\varepsilon)H)\\
&\quad+\frac{1}{2}(|01\rangle+|10\rangle)(\langle 01|+\langle 10|)\otimes\frac{1}{6}((1+2\varepsilon)II+(1-4\varepsilon)H)\\
&\quad+\frac{1}{2}(|01\rangle-|10\rangle)(\langle 01|-\langle 10|)\otimes\frac{1}{2}(II-H)\\
&\quad+|11\rangle\langle 11|\otimes\frac{1}{6}(II+H)\Big)\Big] \qquad (4.57)\\
&=\mathrm{Tr}[\rho'^T|00\rangle\langle 00|]\Big(\frac{1}{6}((1+2\varepsilon)II+(1-4\varepsilon)H)\Big)\\
&\quad+\mathrm{Tr}[\rho'^T((|01\rangle+|10\rangle)(\langle 01|+\langle 10|)]\Big(\frac{1}{12}((1+2\varepsilon)II+(1-4\varepsilon)H)\Big)\\
&\quad+\mathrm{Tr}[\rho'^T((|01\rangle-|10\rangle)(\langle 01|-\langle 10|)]\Big(\frac{1}{4}(II-H)\Big)\\
&\quad+\mathrm{Tr}[\rho'^T|11\rangle\langle 11|]\Big(\frac{1}{6}(II+H)\Big) \qquad (4.58)
\end{aligned}$$

we can take transpose of the quantum noisy state $\rho'$, as shown in Eq. (4.56), then substitute it into our equation, we get

$$\begin{aligned}
\mathcal{C}\circ\mathcal{N}(\rho\otimes\rho)&=(\rho_{00}+\varepsilon\rho_{11})^2\Big(\frac{1}{6}((1+2\varepsilon)II+(1-4\varepsilon)H)\Big)\\
&\quad+[(1-\varepsilon)\rho_{11}(\rho_{00}+\varepsilon\rho_{11})+(1-\varepsilon)\rho_{01}\rho_{10}]\Big(\frac{1}{6}((1+2\varepsilon)II+(1-4\varepsilon)H)\Big)\\
&\quad+[(1-\varepsilon)\rho_{11}(\rho_{00}+\varepsilon\rho_{11})-(1-\varepsilon)\rho_{01}\rho_{10}]\Big(\frac{1}{2}(II-H)\Big)\\
&\quad+(1-\varepsilon)^2\rho_{11}^2\Big(\frac{1}{6}(II+H)\Big) \qquad (4.59)
\end{aligned}$$

The value $\mathrm{Tr}[\rho^2]$ is usually estimated by $\mathrm{Tr}[H(\rho\otimes\rho)]$, where $H=\frac{1}{2}(II+XX+YY+ZZ)$ is the two-qubit cyclic permutation operator. Then the estimated value from our protocol can be arrived at

$$\begin{aligned}
\mathcal{C}\circ\mathcal{N}(\rho\otimes\rho)&=(\rho_{00}+\varepsilon\rho_{11})^2\Big(\frac{1}{6}\mathrm{Tr}[H((1+2\varepsilon)II+(1-4\varepsilon)H)]\Big)\\
&\quad+[(1-\varepsilon)\rho_{11}(\rho_{00}+\varepsilon\rho_{11})+(1-\varepsilon)\rho_{01}\rho_{10}]\Big(\frac{1}{6}\mathrm{Tr}[H((1+2\varepsilon)II+(1-4\varepsilon)H)]\Big)\\
&\quad+[(1-\varepsilon)\rho_{11}(\rho_{00}+\varepsilon\rho_{11})-(1-\varepsilon)\rho_{01}\rho_{10}]\Big(\frac{1}{2}\mathrm{Tr}[H(II-H)]\Big)\\
&\quad+(1-\varepsilon)^2\rho_{11}^2\Big(\frac{1}{6}\mathrm{Tr}[H(II+H)]\Big) \qquad (4.60)
\end{aligned}$$

Note that the $\mathrm{Tr}[H] = 2$ and $\mathrm{Tr}[H \cdot H] = 4$, then we have

$$\begin{aligned}
\mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] &= (\rho_{00} + \varepsilon\rho_{11})^2(1-2\varepsilon) + [(1-\varepsilon)\rho_{11}(\rho_{00} + \varepsilon\rho_{11}) \\
&\quad + (1-\varepsilon)\rho_{01}\rho_{10}](1-2\varepsilon) - [(1-\varepsilon)\rho_{11}(\rho_{00} + \varepsilon\rho_{11}) \\
&\quad - (1-\varepsilon)\rho_{01}\rho_{10}] + (1-\varepsilon)^2\rho_{11}^2 && (4.61)\\
&= (1-2\varepsilon)\rho_{00}^2 + 2(1-\varepsilon)^2\rho_{01}\rho_{10} - 2\varepsilon^2\rho_{00}\rho_{11} + (1-2\varepsilon)\rho_{11}^2 && (4.62)\\
&= (1-\varepsilon)^2\rho_{00}^2 + 2(1-\varepsilon)^2\rho_{01}\rho_{10} + (1-\varepsilon)^2\rho_{11}^2 - \varepsilon^2\rho_{00}^2 + 2\varepsilon^2\rho_{00}\rho_{11} + \varepsilon^2\rho_{11}^2 && (4.63)\\
&= (1-\varepsilon)^2\Big[\rho_{00}^2 + \rho_{11}^2 + 2\rho_{10}\rho_{01} - \frac{\varepsilon^2}{(1-\varepsilon)^2}[\rho_{00}^2 + \rho_{11}^2 + \rho_{00}\rho_{11} + \rho_{11}\rho_{00}]\Big] && (4.64)\\
&= (1-\varepsilon)^2\Big[\mathrm{Tr}[H\rho \otimes \rho] - \frac{\varepsilon^2}{(1-\varepsilon)^2}\Big] && (4.65)\\
&= (1-\varepsilon)^2\Big[\mathrm{Tr}[\rho^2] - \frac{\varepsilon^2}{(1-\varepsilon)^2}\Big]. && (4.66)\\
& && (4.67)
\end{aligned}$$

The desired high-order moment $\mathrm{Tr}[\rho^2]$ value equals to

$$\mathrm{Tr}[\rho^2] = f\mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] - t \tag{4.68}$$

where $f = \frac{1}{(1-\varepsilon)^2}$ is the sampling overhead and $t = -\frac{\varepsilon^2}{(1-\varepsilon)^2}$ is the shifted distance. In order to estimate the value with the error $\delta$, the sampling overhead should be $1/(1-\varepsilon)^2$. Thus, we have $f_{\min} \leq 1/(1-\varepsilon)^2$

Next, we are going to use dual SDP to show that $f_{\min} \geq 1/(1-\varepsilon)^2$. We set the dual variables as

$$\begin{aligned}
M &= \frac{1}{4}(|01\rangle - |10\rangle)(\langle 01| - \langle 10|) + \frac{1}{2}|11\rangle\langle 11| && (4.69)\\
K &= \frac{1}{2(1-\varepsilon)^2}\Big[-\varepsilon|00\rangle\langle 00| - |11\rangle\langle 11| + \frac{1+\varepsilon}{2}(|01\rangle\langle 01| + |10\rangle\langle 10|) \\
&\quad + \frac{\varepsilon - 1}{2}(|01\rangle\langle 10| + |10\rangle\langle 01|)\Big]. && (4.70)
\end{aligned}$$

We will show the variables $\{M, K\}$ is a feasible solution to the dual problem. If we substitute the variables into the dual problem Eq. (4.19), we can easily check that $\mathrm{Tr}[M] \leq 1$ and $\mathrm{Tr}[K] = 1$. For the last condition, after simplifying, we have

$$\mathrm{Tr}_A[(K_A^T \otimes I_B \otimes H_C)(J_{\mathcal{N}_{AB}^{\otimes k}}^{T_B} \otimes I_C)] = (M - |11\rangle\langle 11|) \otimes H \tag{4.71}$$

Therefore,

$$\begin{aligned}
& M \otimes I + \mathrm{Tr}_A[(K_A^T \otimes I_B \otimes H_C)(J_{\mathcal{N}_{AB}^{\otimes k}}^{T_B} \otimes I_C)] && (4.72)\\
=& M \otimes I + (M - |11\rangle\langle 11|) \otimes H && (4.73)\\
\geq& 0 && (4.74)
\end{aligned}$$

which means $\{M, K\}$ is a feasible solution to the dual SDP, therefore, we have $f_{\min} \geq -\mathrm{Tr}[KH] = \frac{1}{(1-\varepsilon)^2}$. Combined with prime part, we have $f_{\min} = \frac{1}{(1-\varepsilon)^2}$. □

The above retriever $\mathcal{C}$ can be implemented based on the following measurements and post-processing. Given amplitude damping noisy states $\mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho)^{\otimes 2}$, we make measurements in the basis $\mathcal{B} = \{|00\rangle, |\Psi^+\rangle, |\Psi^-\rangle, |11\rangle\}$. From the Choi matrix of the retriever $\mathcal{C}$, which is shown in Eq. (4.53), we know that based on the obtained measurement results, the quantum system collapses to the states $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ correspondingly, where $\sigma_1 = \sigma_2 = \frac{1}{6}((1+2\varepsilon)II + (1-4\varepsilon)H)$, $\sigma_3 = \frac{1}{2}(II - H)$ and $\sigma_4 = \frac{1}{6}(II + H)$. Each state corresponds to a fixed expectation value $\mathrm{Tr}[H\sigma_i]$, which can be predetermined via direct matrix calculation with fixed $H$ and known $\sigma_i$. The next step is to run sufficient shots of basis-$\mathcal{B}$ measurements to determine the probability of measuring each basis state, denoted as $p_i$, respectively. The term $\mathrm{Tr}[H\mathcal{C}\left(\mathcal{N}_{\mathrm{AD}}^{\varepsilon}(\rho)^{\otimes 2}\right)]$ is then given by the estimated value $\zeta = \sum_{i=1}^{4} p_i \mathrm{Tr}[H\sigma_i]$. The desired second-order moment is obtained by

$$\mathrm{Tr}[\rho^2] = \frac{1}{(1-\varepsilon)^2}\zeta + \frac{\varepsilon^2}{(1-\varepsilon)^2}. \tag{4.75}$$

More details can be found in the Supplementary Material. The sampling overhead for conventional methods is $\frac{(1+\varepsilon)^2}{(1-\varepsilon)^2}$, while the overhead incurred by our method is still as low as $\frac{1}{(1-\varepsilon)^2}$, saying that our method requires fewer quantum resources.

## 4.5 Generalized observable shift method

The specific protocol of the observable shift method with minimal sampling overhead is given by SDP as shown in Eq. (4.18). However, when the size of the system increases, the computer memory for solving such SDP increases exponentially. In personal computers, a 5-qubit system is the largest problem size that SDP can solve. Achieving a specific error mitigation protocol on a large system is a crucial problem.

We present a scalable approach for creating an error mitigation protocol by exploiting the observable shift method recursively. More specifically, we demonstrate the capability to design a quantum protocol for mitigating depolarizing noises on arbitrary copies of noisy qudits, as evidenced in the subsequent Proposition 4.3.

**Proposition 4.3.** *Given arbitrary $k$ copies of noisy states $\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes k}$, the $k$-th order moment $\mathrm{Tr}[\rho^k]$ can be estimated by $f_k \mathrm{Tr}[H\mathcal{C}_k\left(\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes k}\right)] - t_k$, where the term $\mathrm{Tr}[H\mathcal{C}_k\left(\mathcal{N}_{\mathrm{DE}}^{\epsilon}(\rho)^{\otimes k}\right)]$ can be estimated by implementing a quantum retriever $\mathcal{C}_k$ on noisy states and making measurements. Moreover, such $\mathcal{C}_k$, $f_k$, $t_k$*

*can be recursively constructed as*

$$f_k = \frac{1}{(1-\varepsilon)^k}, \tag{4.76}$$

$$\mathcal{C}_k = \mathrm{id}_k + \sum_{l=2}^{k-1} \binom{k}{l} (1-\varepsilon)^l \varepsilon^{k-l} f_l \mathcal{R}_l^\dagger \circ (\mathcal{C}_l \otimes \mathrm{id}_{k-l}), \ \textit{and} \tag{4.77}$$

$$t_k = f_k \left[ \frac{\varepsilon^k}{d^k} + \frac{k(1-\varepsilon)\varepsilon^{k-1}}{d^{k-1}} - \sum_{l=2}^{k-1} \binom{k}{l} \frac{(1-\varepsilon)^l \varepsilon^{k-l}}{d^{k-l}} t_l \right], \tag{4.78}$$

*for $\mathcal{C}_2, t_2$ given in Proposition 4.1 and some CP maps $\mathcal{R}_l$.*

Note that the implementation of $\mathcal{C}_k$ requires post-selection of measurement outcomes, as $\mathcal{C}_k$ is completely positive but not necessarily trace-preserving. The detailed proof of Proposition 4.3 are deferred to Appendix B.3.2. We also numerically verify the feasibility of the number of state copies to be up to a hundred.

## 4.6 Comparison with existing protocols

To extract high-order moment information from noisy states, one straightforward method is to apply an inverse operation of noisy channel $\mathcal{N}^{-1}$ on quantum states to mitigate error, and then perform measurements over moment observable $H$, which is $\mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[H\left(\mathcal{N}^{-1}\right)^{\otimes k}\left(\mathcal{N}(\rho)^{\otimes k}\right)]$. However, the map $\mathcal{N}^{-1}$ might not be a physical quantum channel [54], thus we cannot implement it directly on a quantum system. Fortunately, we can simulate such channel by quasi-probability decomposition, which decomposes such non-physical map into a linear combination of physical quantum channels, i.e., $\mathcal{N}^{-1} = \sum_i c_i \mathcal{C}_i$, where $c_i$ are the real coefficients and $\mathcal{C}_i$ are physical quantum channels. We need to note that $c_i$ can be negative. From the aspect of physical implementation, in the $t$-th round of total $T$ times of sampling, we first sample a quantum channel $\mathcal{C}^{(t)}$ from $\{\mathcal{C}_i\}$ with probability $\{|c_i|/g\}$, where $g = \sum_i |c_i|$, and apply it to noisy state $\mathcal{C}^{(t)} \circ \mathcal{N}(\rho)$. Then we take measurements and get results $o^{(t)}$. After $T$ rounds of sampling, we attain an estimation for the expectation value $\zeta = \frac{g}{T} \sum_{t=1}^{T} \mathrm{sgn}(c_i^{(t)}) o^{(t)} = \mathrm{Tr}[O\rho]$.

In Ref. [63], the authors proposed a similar but simpler method to realize Hermitian preserving map, which replaces the probabilistic sampling of channels into probabilistic sampling of states. Specifically, a Hermitian preserving map $\mathcal{N}^{-1}$ can be realized by injecting different quantum states, $\omega_i$ with probability $\frac{|c_i|}{g}$ into a fixed quantum channel $\Lambda$, i.e., $\mathcal{N}^{-1}(\cdot) = \sum_i c_i \Lambda(\cdot \otimes \omega_i)$. The new proposed method [63], no doubt, simplifies the realization of Hermitian preserving maps to some extent. However, we still need to sample different states with certain probabilities and inject them into quantum channel. Besides, it was proved that the sampling overhead of probabilistic state sampling method should be the same as that of probabilistic channel sampling [63].

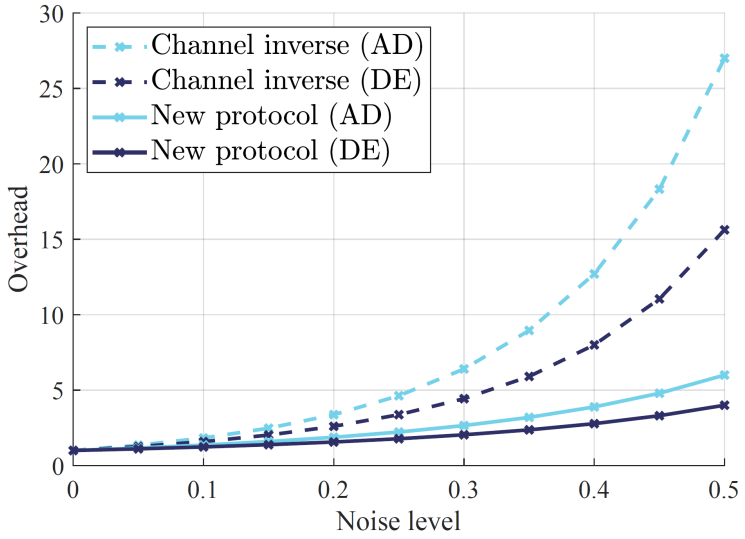



Figure 4.3: The sampling overhead respective to noise level for estimating $\mathrm{Tr}[\rho^3]$ from amplitude damping noise corrupted state $\mathcal{N}(\rho)$. The dashed curve refers to the overhead for the channel inverse, and the solid curve stands for the newly proposed method. The light blue and dark blue curves represent results from amplitude damping (AD) and depolarizing channel (DE), respectively.

The total sampling times $T$ is also given by Hoeffding's inequality as shown in Eq. (4.15), and the sampling overhead $g_{\min}(\mathcal{N})$ is given by [54, 63, 138]

$$g_{\min}(\mathcal{N}) = \min\Big\{ \sum_i |c_i| \; | \mathcal{N}^{-1} = \sum_i c_i \mathcal{C}_i, c_i \in \mathbb{R}, \mathcal{C}_i \in \mathrm{CPTP} \Big\}. \tag{4.79}$$

When we apply the channel inverse method to retrieve the $k$-th moment, we should apply the inverse operation simultaneously on $k$ quantum systems, the corresponding sampling overhead if given by $g_{\min}(\mathcal{N}, k)$, which is

$$g_{\min}(\mathcal{N}, k) = \min\Big\{ \sum_i |c_i| \; | (\mathcal{N}^{-1})^{\otimes k} = \sum_i c_i \mathcal{C}_i, c_i \in \mathbb{R}, \mathcal{C}_i \in \mathrm{CPTP} \Big\}, \tag{4.80}$$

With Eq.(4.16) and Eq. (4.80), we can make a comparison of the sampling overhead between our method and the conventional QPD channel inverse method, which leads to Lemma 4.2.

**Lemma 4.2.** *For arbitrary invertible quantum noisy channel $\mathcal{N}$, and moment order $k$, we have*

$$f_{\min}(\mathcal{N}, k) \leq g_{\min}(\mathcal{N}, k) \tag{4.81}$$

Lemma 4.2 implies that in the task of extracting the $k$-th moment $\mathrm{Tr}[\rho^k]$ from noisy states, the proposed method requires fewer sampling times, i.e.,

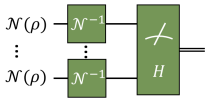
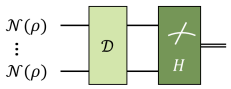
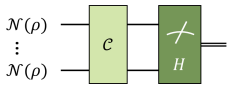
| Method | Channel Inverse | Info. recover | Obs. shift |
|---|---|---|---|
| |  |  |  |
| Constraints | $\mathcal{N}^{-1} \circ \mathcal{N} = \mathcal{I}$ | $\left(\mathcal{N}^{\otimes k}\right)^{\dagger} \mathcal{D}^{\dagger}(H) = H$ | $\left(\mathcal{N}^{\otimes k}\right)^{\dagger} \mathcal{C}^{\dagger}(H) = \frac{1}{f}(H + tI)$ |
| Implementation | Apply quantum channels with prob. | Apply quantum channels with prob. | Apply one quantum channel |
| Overhead | High | Medium | Low |

Figure 4.4: Comparison of the setting between three error mitigation methods.

consumes fewer quantum resources. The detailed proof is displayed in Supplementary Material. It has been proven by Regula et. al.in Ref. [63] that the sampling overhead for simulating a trace preserving linear map is equivalent to its diamond norm. Specifically, in the case of inverse operation $\mathcal{N}^{-1}$, we have $g_{\min}(\mathcal{N}) = \|\mathcal{N}^{-1}\|_{\diamond}$. Note that the diamond norm is multiplicative with respect to tensor product [54, 63], i.e., $\|(\mathcal{N}^{-1})^{\otimes k}\|_{\diamond} = \|\mathcal{N}^{-1}\|_{\diamond}^{k}$. Thus, we conclude that the sampling overhead for retrieving high-moment from noisy states increases exponential respect to the moment order $k$, which is

$$\begin{aligned} g_{\min}(\mathcal{N}, k) &= g_{\min}(\mathcal{N}^{\otimes k}) = \|(\mathcal{N}^{-1})^{\otimes k}\|_{\diamond} = \|\mathcal{N}^{-1}\|_{\diamond}^{k} \\ &= g_{\min}(\mathcal{N})^{k}. \end{aligned} \tag{4.82}$$

In order to illustrate the advantage of the proposed method over the channel inverse method in terms of sampling overhead, we conduct a numerical experiment to extract the third moment $\mathrm{Tr}[\rho^3]$ from amplitude damping noise channel with different noise levels. The results are shown in Figure. 4.3. The dashed and solid curves stand for the sampling overhead for the channel inverse and observable shift method, respectively. We also made a comparison with another method which was proposed in [138].

Here, we are going to make a comparison of the sampling overhead of three different methods, which are channel inverse, information recover, and observable shift, respectively. The straightforward comparison is shown in the following.

1. **Channel inverse.** This method was proposed in Ref. [54, 62], which cancels the noise by applying the inverse of the noise channel $\mathcal{N}^{-1}$ on noisy state $\mathcal{N}(\rho)$, such that the expectation value is unbiased, i.e., $\mathrm{Tr}[O\mathcal{N}^{-1} \circ \mathcal{N}(\rho)] = \mathrm{Tr}[O\rho]$. The minimal sampling overhead of this method can be

given by SDP:

$$\begin{aligned}
\min \quad & p_1 + p_2 && (4.83)\\
\text{subject to} \quad & J_{\mathcal{D}} \equiv J_1 - J_2 && (4.84)\\
& \mathrm{Tr}_B[J_1] = p_1 I_A, \quad \mathrm{Tr}_B[J_2] = p_2 I_A && (4.85)\\
& J_1 \geq 0, \quad J_2 \geq 0 && (4.86)\\
& J_{\mathcal{F}_{AC}} \equiv \mathrm{Tr}_B[(J_{\mathcal{N}_{AB}}^{T_B} \otimes I_C)(I_A \otimes J_{\mathcal{D}_{BC}})] && (4.87)\\
& J_{\mathcal{F}_{AC}} = J_{\mathrm{id}_{AC}} && (4.88)
\end{aligned}$$

$J_{\mathcal{D}}$ and $J_{\mathcal{N}}$ are the Choi matrices for retriever $\mathcal{D}$ and noisy $\mathcal{N}$ respectively. $J_{\mathrm{id}_{AB}}$ refers to the Choi matrix for the identity channel.

2. **Information recover.** This method was proposed in Ref. [138], which aims to find the hermitian preserving trace preserving map $\mathcal{D}$ such that the information can be recovered respective to observable $O$, i.e., $\mathrm{Tr}[O\mathcal{D} \circ \mathcal{N}(\rho)] = \mathrm{Tr}[O\rho]$. Note that this method differs from the channel inverse method. The channel inverse can be understood as recovering information for all observables, while this method only recovers information concerning a certain observable. The minimal sampling overhead of this method can be given by SDP:

$$\begin{aligned}
\min \quad & c_1 + c_2 && (4.89)\\
\text{subject to} \quad & J_{\mathcal{D}} \equiv J_1 - J_2 && (4.90)\\
& \mathrm{Tr}_B[J_1] = c_1 I_A, \quad \mathrm{Tr}_B[J_2] = c_2 I_A && (4.91)\\
& J_1 \geq 0, \quad J_2 \geq 0 && (4.92)\\
& J_{\mathcal{F}_{AC}} \equiv \mathrm{Tr}_B[(J_{\mathcal{N}_{AB}}^{T_B} \otimes I_C)(I_A \otimes J_{\mathcal{D}_{BC}})] && (4.93)\\
& \mathrm{Tr}_C[(I_A \otimes O_C^T) J_{\mathcal{F}_{AC}}^T] = O_A && (4.94)
\end{aligned}$$

$J_{\mathcal{D}}$ and $J_{\mathcal{N}}$ are the Choi matrices for retriever $\mathcal{D}$ and noisy $\mathcal{N}$ respectively. $T$, $T_B$ stand for transpose and partial transpose, respectively. $O$ refers to observable.

3. **Observable shift.** This method is proposed in this work. Details can be found in the OBSERVABLE SHIFT METHOD section in the main text. The minimal sampling overhead of this method is given by the following SDP:

$$\begin{aligned}
\min \quad & f && (4.95)\\
\text{subject to} \quad & J_{\tilde{\mathcal{C}}} \geq 0 && (4.96)\\
& \mathrm{Tr}_C[J_{\tilde{\mathcal{C}}_{BC}}] = f I_B && (4.97)\\
& J_{\mathcal{F}_{AC}} \equiv \mathrm{Tr}_B[(J_{\mathcal{N}_{AB}^{\otimes k}}^{T_B} \otimes I_C)(I_A \otimes J_{\tilde{\mathcal{C}}_{BC}})] && (4.98)\\
& \mathrm{Tr}_C[(I_A \otimes H_C^T) J_{\mathcal{F}_{AC}}^T] = H_A + t I_A. && (4.99)
\end{aligned}$$

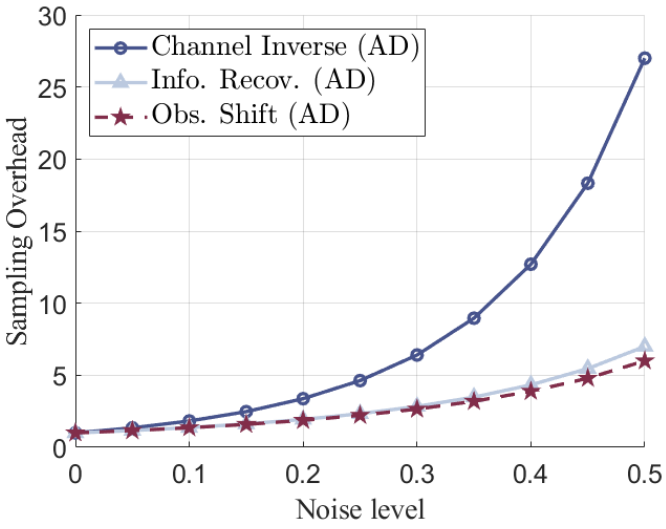



Figure 4.5: Comparison of sampling overhead between three error mitigation methods under amplitude damping channel.

$J_{\tilde{\mathcal{C}}}$ and $J_{\mathcal{N}}$ are the Choi matrices for retriever $\tilde{\mathcal{C}}$ and noisy $\mathcal{N}$ respectively. $I$ is identity. $T, T_B$ stand for transpose and partial transpose, respectively. $H$ refers to observable for estimating high order moment.

We also conduct a numerical simulation to compare the minimal sampling overhead of the three methods in the task of estimating $\mathrm{Tr}[\rho^3]$ from noisy states. The noise model is amplitude damping and the observable is $H = (S + S^\dagger)/2$, where $S$ is the cyclic permutation matrix. The results are shown in Figure. 4.5. It is straightforward to conclude that the new proposed observable shift method requires the lowest sampling overhead compared with the channel inverse method and observable shift method.

Compared with the existing method, the proposed observable shift method has at least two-fold advantages: lower sampling overhead and easier implementation. First, our method can achieve a lower sampling overhead, i.e., $f_{\min}(\mathcal{N}, k) \leq g_{\min}(\mathcal{N}, k)$. In previous, we have shown examples where $f_{\min}(\mathcal{N}, k)$ is strictly smaller, indicating the effectiveness of the newly proposed observable shift technique. The lower sampling overhead is also illustrated in numerical results (Figure. 4.3 and Figure. 4.5). Second, protocols given by our method are more hardware friendly. In previous method [54], to realize HPTP map $\Phi = \sum_i c_i \Lambda_i$, it is necessary to sample different quantum channels $\Lambda_i$ with certain probability $\{c_i / \sum_i |c_i|\}$ and prepare them on quantum devices, which is very complicated. Later on, such a process (realizing HPTP map $\Phi$) is simplified into feeding in the different states $\omega_i$ into the circuit which realizes the fixed operation $\Lambda$, i.e., $\Phi(\cdot) = \Lambda(\cdot \otimes \sum_i c_i \omega_i)$ [63], which simplifies the implementa-

tion greatly. However, the simplified method still needs to sample the quantum states with certain probability. While in our method, we only need to implement a fixed a quantum circuit without any probabilistic sampling process.

## 4.7 Application to Fermi-Hubbard model

The Fermi-Hubbard model is a key focus in condensed matter physics due to its relevance in metal-insulator transitions and high-temperature superconductivity [141, 142]. Recent studies have shown that entanglement spectroscopy can be utilized to extract critical exponents and phase transitions in the Fermi-Hubbard model [143–145]. As the model is characterized by a broad range of correlated electrons, it necessitates multi-determinant and highly accurate calculations [145, 146] which hence demand ingenious methods of quantum noise control.

In a physical system such as a metallic crystal with an $n_x \times n_y$ square lattice, each lattice point, known as a site, is assigned an index. The Hubbard model Hamiltonian takes on a fermionic form in second quantization

$$\begin{aligned} H_{\text{Hubbard}} = -J \sum_{\langle i,j\rangle,\sigma} (a^\dagger_{i\sigma} a_{j\sigma} + a^\dagger_{j\sigma} a_{i\sigma}) \\ + U \sum_i n_{i\uparrow} n_{i\downarrow} + H_{\text{local}}, \end{aligned} \tag{4.100}$$

where $a^\dagger_{i\sigma}, a_{i\sigma}$ are fermionic creation and annihilation operators; $n_{i\sigma} = a^\dagger_{i\sigma}, a_{i\sigma}$, are the number operators; the notation $\langle i, j\rangle$ associates adjacent sites in the $n_x \times n_y$ rectangular lattice; $\sigma \in \{\uparrow, \downarrow\}$ labels the spin orbital. The first term in Eq. (4.100) corresponds to the hopping term, where $J$ denotes the tunneling amplitude. The second term involves the on-site Coulomb repulsion, represented by $U$. The final term in the equation defines the local potential resulting from nuclear-electron interaction, which we have chosen to be the Gaussian form [147]

$$H_{\text{local}} = \sum_{j=1} \sum_{\nu=\uparrow,\downarrow} \epsilon_{j,\nu} n_{j,\nu}; \quad \epsilon_{j,\nu} = -\lambda_\nu e^{-\frac{1}{2}(j-m_\nu)^2/\sigma^2_\nu}. \tag{4.101}$$

In the following, we consider a specific 3-site (6-qubit) Fermi-Hubbard Hamiltonian with $J = 2, U = 3$ and $\lambda_{\uparrow,\downarrow} = 3, 0.1$, $m_{\uparrow,\downarrow} = 3, 3$. The standard deviation $\sigma_v$ for both spin-up and -down potentials are set to 1 guaranteeing a charge-spin symmetry around the center site ($i = 2$) of the chain system.

The ground state entanglement spectroscopy of the model identifies the topological-ordering signatures of the system which requires high-precision entropy estimations over each bipartite sector of the entire system. We show that the mitigation of the quantum noise can be achieved and therefore, enhance the determination of $\text{Tr}[\mathcal{N}(\rho_A)^2]$, via our proposed method, which is displayed in the Supplementary Material.

Figure. 4.6 displays the sampling distribution with and without error mitigation. The orange curve refers to the estimation distribution of second-order

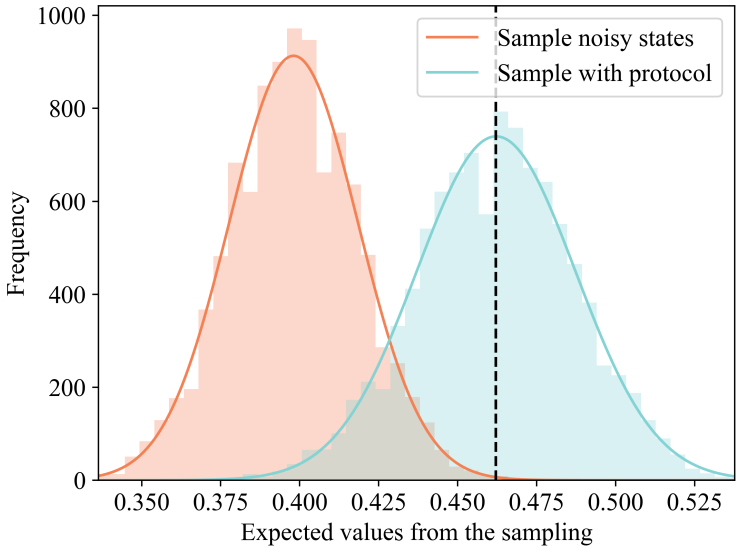



Figure 4.6: Simulation of high-order moment $\mathrm{Tr}[\rho^2]$ estimation. The curves are calculated from sampling. The orange curve represents the estimation from depolarizing noised state $\mathcal{N}(\rho)$ with noise level $\epsilon = 0.1$. The cyan curve is the estimation with the proposed error mitigation method. The black dashed line stands for the exact value of $\mathrm{Tr}[\rho^2]$.

information $\mathrm{Tr}[\rho^2]$ from noisy states, and the cyan curve shows the estimation distribution with error mitigation. And the black dash line is the exact value of $\mathrm{Tr}[\rho^2]$.

## 4.8 Conclusion and discussion

In this study, we establish that when quantum states are distorted by noises, the original moment information can still be retrieved through post-processing if and only if the noise is invertible. Furthermore, our proposed method, called observable shift, outperforms existing techniques in two aspects: (1) The proposed method requires lower quantum sampling complexity than the existing one. (2) The observable shift method is easier to implement than the existing method, which makes our method more friendly to quantum devices. We also propose the construction of a protocol to retrieve the depolarizing channel of large-size quantum systems. Our findings have implications for the dependable estimation of non-linear information in quantum systems and can influence various applications, including entanglement spectroscopy and ground-state property estimation.

For further work, one important task is to improve the scalability of the observable shift method, which makes this approach more practical. Also, in-

vestigating the approximate version of retrieving non-linear features is also interesting. We expect the observable shift technique can be incorporated into more algorithms and protocols to boost efficiency. It will be also interesting to explore other error mitigation methods [148–153] to extract non-linear features from noisy quantum states.

# Chapter 5

# Probabilistic Channel Simulation Using Coherence

In this chapter, we study the channel simulation using coherence, which refers to realizing a target channel with coherent states and free operations, which is a fundamental problem in the quantum resource theory of coherence. The limitations of the accuracy of deterministic channel simulation motivate us to consider the more general probabilistic framework. In this chapter, we develop the framework for probabilistic channel simulation using coherence with free operations. When the chosen set of free operations is the maximally incoherent operations (MIO), we provide an efficiently computable semidefinite program (SDP) to calculate the maximal success probability and derive the analytic expression of success probability for some special cases. When the chosen set of free operations is the dephasing-covariant incoherent operations (DIO), we show that if the target channel is not a resource nonactivating channel, then one cannot simulate it exactly both deterministically and probabilistically. The SDP for maximal success probability of simulating a channel by DIO is also given correspondingly. This chapter is based on [Zhao, Ito, and Fujii. Phys. Rev. Research 6, 043316 (2024)] with slight modifications to fit in the context.

## 5.1 Introduction

In recent years, many efforts have contributed to establishing the framework of quantum resource theories [41, 64] to understand the unique properties of quantum mechanical systems such as coherence [73, 74, 154–161], entanglement [42, 44, 45, 82, 162], and magic [49, 163–167]. In general, a resource theory is defined by specifying free states and free operations. Free states are states that do not possess the resource under consideration, while free operations are operations that preserve the set of free states. Taking the resource theory of entanglement as an example [42, 44, 45, 82, 162], the free states are separable states, which are not entangled, and one of the free operation sets is local operation and classical

communication (LOCC) [22, 168], which does not generate entanglement. Similar to entanglement, coherence is another important topic in quantum resource theories [41], which refers to the property of the superposition of states. It empowers various quantum tasks, such as cryptography [65], metrology [66–68], thermodynamics [69–71], and channel simulation [48, 72].

In the resource theory of coherence [154], the free states are defined as classical states, i.e., density operators that are diagonal in a given reference orthogonal basis $\{|i\rangle\}$. Such states are called *incoherent states* and denoted as $\mathcal{I}$. The corresponding *maximally coherent state* in dimension $m$ is the state $|\Psi_m\rangle = \frac{1}{\sqrt{m}} \sum_{j=0}^{m-1} |j\rangle$. In this work, we denote the density matrix of a maximally coherent state with rank $m$ as $\Psi_m = |\Psi_m\rangle\langle\Psi_m|$ for convenience. The resource theory of coherence does not have a gold-standard physics-motivated class of operations like LOCC in the entanglement resource theory. With this in mind, our task is to characterize the operational properties and applications of quantum coherence under several different sets of operations, such as dephasing-covariant incoherent operations (DIO) [169], and maximally incoherent operations (MIO) [73]. An operation $\mathcal{M}$ is in DIO if it commutes with $\Delta$, or equivalently $\mathcal{M}(|i\rangle\langle i|) \in \mathcal{I}$ and $\Delta(\mathcal{M}(|i\rangle\langle j|)) = 0,\ \forall i \neq j$, where $\Delta$ is the coherence destroying map [170] (completely dephasing channel), i.e., $\Delta(\cdot) = \sum_i |i\rangle\langle i| \cdot |i\rangle\langle i|$. MIO consists of all operations $\mathcal{M}$ such that $\mathcal{M}(\rho) \in \mathcal{I}$ for any free state $\rho \in \mathcal{I}$. From the definition, DIO is a subset of MIO, i.e., DIO $\subset$ MIO. Note that both MIO and DIO are defined in a mathematical way and don't have corresponding physical implementations. However, it doesn't mean MIO and DIO are useless. In the resource theory of coherence, physically incoherent operations (PIO) [169], a subset of MIO and DIO, are physically consistent operations. But the set of PIO is hard to characterize by mathematics. In order to investigate the property of PIO, relaxing the operation into a larger set of operations such as DIO or MIO is a common strategy. Although relaxing the operations cannot study the exact property of PIO, the bounds can be characterized efficiently.

Initial research primarily of coherence focused on the quantification and interconversion of *static coherence* [73–75], which refers to the degree of superposition within a state. Later on, a more general framework was proposed, called *dynamic coherence*–the power to generate coherence itself [72]. How to convert static coherence into dynamic coherence (also known as *channel simulation using coherence*, as shown in Fig. 5.1) raises great interest, and many efforts have contributed to establishing the framework of deterministic channel simulation [48, 72, 76]. At the same time, the limitations of deterministic channel simulation are also shown.

A given resource state does not necessarily allow for the exact and deterministic MIO simulation of a given target quantum channel, although for each channel there exists a state such that such a simulation is possible [48, 72]. In fact, if we can only access the resource state $\omega$ whose robustness of coherence is smaller than the requirement of the channel simulation, then it is impossible to simulate the target channel with no error. Also, in Ref. [76], the authors found that any coherent unitary channel cannot be simulated by DIO exactly (e.g. the

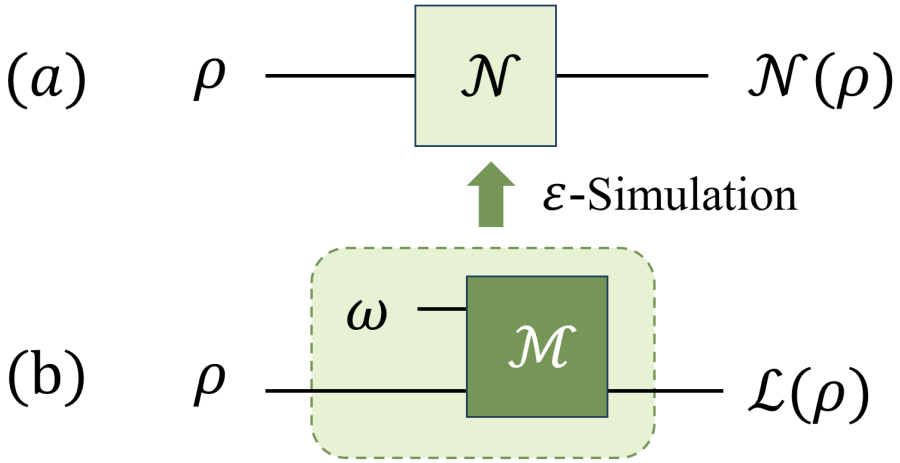



Figure 5.1: (a) The quantum channel $\mathcal{N}$. (b) Utilizing a free operation $\mathcal{M}$ and a given resource state $\omega$ to implement channel $\mathcal{L}$. The realized channel $\mathcal{L}$ is $\varepsilon$-close to the target channels $\mathcal{N}$.

Hadamard gate) no matter what resource state $\omega$ is provided. More generally, in Ref. [46], the authors showed that if the target channel is not a resource nonactivating channel [170], then it cannot be simulated by DIO exactly and deterministically.

The limitation of deterministic channel simulation motivates us to consider a more general probabilistic framework, in which the channel simulation will succeed only with some probability. The probabilistic framework has been applied to many quantum tasks and has shown advantages over deterministic ones. For example, in Ref. [171], the author observes that one can transfer the quantum state from $\rho$ to $\sigma$ probabilistically, while such transformation is forbidden in the deterministic scenario. Moreover, in the task of coherence distillation [155], with the same input state, the output state by probabilistic distillation is closer to the maximal coherent state than deterministic distillation.

In this chapter, we focus on the probabilistic channel simulation using coherence, characterizing the relation between the maximal success probability and the distance from simulated channel $\mathcal{L}$ and target channel $\mathcal{N}$. In the first part, we show three cases of channel simulation with MIO depending on the types of the resource state. (i) If the resource state is the rank-$m$ maximally coherent state $\omega = \Psi_m$, we provide the analytical expression of maximal success probability. (ii) If the resource state $\omega$ is a pure coherent state, we derive the non-zero success probability. (iii) If the resource state $\omega$ is a general coherent state, we provide an efficiently computable semidefinite program (SDP) to achieve the maximal success probability. In the second part, we concentrate on the channel simulation with DIO. It was proved in Ref. [46] that if the target channel $\mathcal{N}$ is

not a resource nonactivating channel, then it cannot be simulated exactly by DIO with any resource state deterministically. We further show that such an exact simulation is impossible even in probabilistic scenarios. We then provide the efficiently computable SDP for achieving the maximal success probability of channel simulation with DIO. Our work fills an important gap in the literature by establishing the probabilistic toolbox for the key resource of quantum coherence.

## 5.2 The problem of probabilistic channel simulation

To quantify the coherence of a quantum channel, the robustness of coherence (ROC) of a quantum channel was proposed

**Definition 5.1.** [72] *The robustness of coherence of a quantum channel $\mathcal{N}$, $C_R(\mathcal{N})$ is defined by*

$$1 + C_R(\mathcal{N}) := \min\{\lambda : \ \mathcal{N} \leq \lambda\mathcal{M}, \mathcal{M} \in \mathrm{MIO}\}. \tag{5.1}$$

*The inequality of $\mathcal{N} \leq \lambda\mathcal{M}$ means that the map $\lambda\mathcal{M} - \mathcal{N}$ is completely positive.*

**Definition 5.2.** [72] *The smoothed version of a robustness of channel is called $\varepsilon$-robustness of coherence of channel, which is defined by*

$$C_R^{\varepsilon}(\mathcal{N}) := \min\left\{C_R(\mathcal{L}) : \ \frac{1}{2}\|\mathcal{N} - \mathcal{L}\|_{\diamond} \leq \varepsilon\right\}. \tag{5.2}$$

If we consider replacement channel $\mathcal{N}_\sigma$, whose output is independent of the input state $\rho$, i.e., $\mathcal{N}_\sigma(\rho) = \mathrm{Tr}[\rho]\sigma$, then the ROC of a channel (Definition 5.1) reduces to the ROC of states [158, 160].

**Definition 5.3.** [158, 160] *Given an arbitrary quantum state $\rho$, the robustness of coherence of the state is defined as*

$$C_R(\rho) = \min_{\tau}\left\{s \geq 0 \middle| \frac{\rho + s\tau}{1+s} = \sigma \in \mathcal{I}\right\}, \tag{5.3}$$

*where $\mathcal{I}$ refers to the set of free states.*

Similar to the ROC of a state, the ROC of a channel quantifies the minimal mixing required to destroy all the coherence in a quantum channel $\mathcal{N}$ [72]. The ROC of quantum channel $C_R(\mathcal{N})$, as well as $\varepsilon$-ROC of a channel can be effectively calculated by SDP [72].

The task of probabilistic channel simulation can be defined as follows. For a given target channel $\mathcal{N}_{A\to B}$, which transforms linear operators in the system $A$ to the system $B$, resource state $\omega$ and error tolerance $\varepsilon$, we aim to find a free operation $\mathcal{M}$ such that $\mathcal{M}(\omega \otimes \cdot)$ probabilistically outputs a channel $\mathcal{L}$ which is close to $\mathcal{N}(\cdot)$ up to error $\varepsilon$. A single-bit classical flag register $F$ is

used to indicate if the simulation is successful. If the register $F$ is found to be $|0\rangle$, then it implies the output channel $\mathcal{L}(\cdot)$ simulates our target $\mathcal{N}(\cdot)$ up to an error $\varepsilon$. Otherwise, the simulation fails and we discard the "rubbish" output $\mathcal{K}(\cdot)$, where $\mathcal{K}$ can be arbitrary MIO. When the channel simulation fails, we can perform a replacement channel $\mathcal{N}_{\frac{\mathbb{I}}{d}}$, which replaces input state into identity, on $\mathcal{K}(\cdot)$ as post-operation, i.e., $\mathcal{N}_{\frac{\mathbb{I}}{d}} \circ \mathcal{K}(\cdot) = \mathrm{Tr}[\cdot]\mathbb{I}/d$, where $\mathbb{I}$ is identity and $d$ is the dimension of the system. Note that such post-operation won't change the success probability. The state version of this trick is shown in Ref. [155]. Our goal is to maximize the success probability of channel simulation. Here, we can define the problem as follows

**Problem 5.1.** *Given triplet $(\mathcal{N}, \omega, \varepsilon)$, what is the maximal success probability $P_{\mathcal{O}}(\mathcal{N}, \omega, \varepsilon)$ to simulate channel $\mathcal{N}$ up to error $\varepsilon$ with the given resource state $\omega$ and free operation class $\mathcal{O} \in \{\mathrm{MIO}, \mathrm{DIO}\}$? Mathematically, it is an optimization problem*

$$
\begin{aligned}
P_{\mathcal{O}}(\mathcal{N}, \omega, \varepsilon) = \max \quad & p; && \text{(5.4a)}\\
\text{s.t.} \quad & \mathcal{M}_{RA\to FB}(\omega \otimes \cdot) = p|0\rangle\langle 0|_F \otimes \mathcal{L}(\cdot) \\
& \qquad + (1-p)|1\rangle\langle 1|_F \otimes \mathrm{Tr}[\cdot]\frac{\mathbb{I}}{d}; && \text{(5.4b)}\\
& \frac{1}{2}||\mathcal{L} - \mathcal{N}||_\diamond \leq \varepsilon; && \text{(5.4c)}\\
& \mathcal{M} \in \mathcal{O}, && \text{(5.4d)}
\end{aligned}
$$

*where $\mathcal{L}$ are the probabilistically implemented channels, which approximates the target channel, and $\mathbb{I}$ refers to identity.*

System $R$ is the resource system containing the resource state, and system $F$ is the flag system, indicating if the channel simulation is successful. $||\cdot||_\diamond$ is known as the diamond norm [172], which has two operational meanings: First, it quantifies how well one physically discriminates between two quantum channels [173]. If we set $\varepsilon = 0$, it implies the exact implementation of the target channel. Second, it quantifies the cost for simulating a general hermitian preserving map with physical implementations [63, 174].

## 5.3 Probabilistic channel simulation with MIO

In this section, we are going to study the channel simulation with MIO under three cases. First, if the resource state is the rank-$m$ maximally coherent state, we provide the analytical expression of the maximal success probability. Second, if the resource state $\omega$ is a pure coherent state, we derive analytical lower bounds on the maximal success probability. Third, for general coherent states, we provide an efficiently computable SDP to determine the maximal success probability.

**Theorem 5.1.** *Given an arbitrary target channel $\mathcal{N}$ and a fixed maximally coherent state $\Psi_m$ of dimension $m \geq 2$, the maximal success probability of the exact channel simulation with MIO is*

$$P_{\text{MIO}}(\mathcal{N}, \Psi_m, \varepsilon = 0) = \min\left\{1, \frac{m-1}{C_R(\mathcal{N})}\right\}. \tag{5.5}$$

*Proof.* If a target channel $\mathcal{N}$ can be simulated by MIO and the given resource state $\Psi_m$, it means that there exists an operation $\mathcal{M} \in \text{MIO}$, such that

$$\mathcal{M}(\Psi_m \otimes \cdot) = \mathcal{L}(\cdot) = p|0\rangle\langle 0| \otimes \mathcal{N}(\cdot) + (1-p)|1\rangle\langle 1| \otimes \mathcal{K}(\cdot),$$

with $p > 0$. The output $\mathcal{K}(\cdot)$ is a failure branch, which should be discarded. And the channel $\mathcal{K}$ can be arbitrary MIO, i.e., $C_R(\mathcal{K}) = 0$. The main idea is to prove $C_R(\mathcal{L}) = pC_R(\mathcal{N})$, which will be obtained directly by proving $C_R(\mathcal{L}) \leq pC_R(\mathcal{N})$ and $C_R(\mathcal{L}) \geq pC_R(\mathcal{N})$.

Due to the convexity of the ROC of channels [72], we have

$$\begin{aligned} C_R(\mathcal{L}) &= C_R(p\,|0\rangle\langle 0| \otimes \mathcal{N} + (1-p)\,|1\rangle\langle 1| \otimes \mathcal{K}) \\ &\leq pC_R(|0\rangle\langle 0| \otimes \mathcal{N}) + (1-p)C_R(|1\rangle\langle 1| \otimes \mathcal{K}) \\ &= pC_R(\mathcal{N}) + (1-p)C_R(\mathcal{K}) \\ &= pC_R(\mathcal{N}). \end{aligned} \tag{5.6}$$

The third equation holds because removing and appending systems in incoherent states are in MIO.

On the other hand, from the definition of the ROC of a channel, we have

$$\begin{aligned} &1 + C_R(\mathcal{L}) \\ =&\min\{\lambda | p\mathcal{N}_{[0]} + (1-p)\mathcal{K}_{[1]} \leq \lambda\mathcal{M}, \mathcal{M} \in \text{MIO}\} \\ =&\min\left\{\lambda \middle| \mathcal{N}_{[0]} \leq \frac{\lambda\mathcal{M} - (1-p)\mathcal{K}_{[1]}}{p}, \mathcal{M} \in \text{MIO}\right\} \\ =&\min\left\{\lambda \middle| \mathcal{N}_{[0]} \leq \frac{\lambda+p-1}{p} \cdot \frac{\lambda\mathcal{M} - (1-p)\mathcal{K}_{[1]}}{\lambda+p-1}, \mathcal{M} \in \text{MIO}\right\} \qquad (5.7) \\ =&\lambda^*, \qquad (5.8) \end{aligned}$$

where we define $\mathcal{N}_{[0]} = |0\rangle\langle 0| \otimes \mathcal{N}$ and $\mathcal{M}_{[1]} = |1\rangle\langle 1| \otimes \mathcal{M}$. Here, the minimum $\lambda^*$ always exists, which can be straightforwardly derived from Theorem 2 in Ref. [72]. For convenience, we define $\mathcal{M}' := \frac{\lambda\mathcal{M} - (1-p)\mathcal{K}_{[1]}}{\lambda+p-1}$ for $\mathcal{M} \in MIO$. Because the minimum is attained by $\lambda \geq 1$, the inequality in Eq. (5.7) ensures that $\mathcal{M}'$ is completely positive, i.e., $0 \leq [p/(\lambda+p-1)]\mathcal{N}_{[0]} \leq \mathcal{M}'$. Since both $\mathcal{M}, \mathcal{K}_{[1]} \in \text{MIO}$, we have $\text{Tr}[\mathcal{M}(\rho)] = \text{Tr}[\rho] = \text{Tr}[\mathcal{K}_{[1]}(\rho)]$ and $\mathcal{M}(\rho), \mathcal{K}_{[1]}(\rho) \in \mathcal{I}$ for any free state $\rho \in \mathcal{I}$. Therefore, we have

$$\mathcal{M}'(\rho) = \frac{\lambda\mathcal{M}(\rho) - (1-p)\mathcal{K}_{[1]}(\rho)}{\lambda+p-1} \in \mathcal{I} \quad (\forall \rho \in \mathcal{I}),$$

and $\mathrm{Tr}[\mathcal{M}'(\rho)] = \mathrm{Tr}[\rho]$. Hence, we obtain $\mathcal{M}' \in \mathrm{MIO}$. Then, noting that $C_R(\mathcal{N}) = C_R(\mathcal{N}_{[0]})$, we have

$$\begin{aligned} 1 + C_R(\mathcal{N}) &= 1 + C_R(\mathcal{N}_{[0]}) \\ &= \min\{\tau | \mathcal{N}_{[0]} \leq \tau\mathcal{M}, \mathcal{M} \in \mathrm{MIO}\}, \\ &\leq \frac{\lambda^* + p - 1}{p}. \end{aligned}$$

Because $\lambda^* = 1 + C_R(\mathcal{L})$ from Eq. (5.8), we arrive at

$$pC_R(\mathcal{N}) \leq C_R(\mathcal{L}). \tag{5.9}$$

Combined with Eq. (5.6), we obtain

$$pC_R(\mathcal{N}) = C_R(\mathcal{L}). \tag{5.10}$$

From Theorem 4 in Ref. [72], there exists an MIO to simulate channel $\mathcal{L}$ if and only if $m - 1 \geq C_R(\mathcal{L})$. By considering Eq. (5.10), we can rephrase that there exists an MIO to simulate channel $\mathcal{N}$ probabilistically if and only if the success probability $p$ satisfies

$$\frac{m-1}{C_R(\mathcal{N})} \geq p. \tag{5.11}$$

Because the probability cannot exceed 1, the maximal success probability is

$$P_{\mathrm{MIO}}(\mathcal{N}, \Psi_m, \varepsilon = 0) = \min\left\{1, \frac{m-1}{C_R(\mathcal{N})}\right\}. \tag{5.12}$$

The proof is complete. □

Theorem 5.1 implies the success probability of simulating a channel by using MIO and maximally coherent states is always greater than 0, i.e., $P_{\mathrm{MIO}}(\mathcal{N}, \Psi_m, \varepsilon = 0) > 0$ with $m \geq 2$. In other words, any quantum channel can be simulated by MIO probabilistically with $\Psi_m$.

According to Theorem 5.1, if given enough resource, i.e., $C_R(\Psi_m) = m - 1 \geq C_R(\mathcal{N})$, it implies the success probability equals to one, thus we can simulate channel deterministically. Otherwise, if we can only access a limited amount of coherence, i.e., $C_R(\Psi_m) = m - 1 < C_R(\mathcal{N})$, then the channel fails to be simulated deterministically. Even though, we can still succeed in such a simulation probabilistically, with the probability given by the ratio between the robustness of resource state $\Psi_m$ and the robustness of coherent channel $\mathcal{N}$, i.e.,

$$p = \frac{m-1}{C_R(\mathcal{N})}. \tag{5.13}$$

In this sense, Theorem 5.1 quantitatively gives in terms of the resource measures how the advantage of probabilistic channel simulation over the deterministic one

appears. Examples of coherent unitary channels are studied to demonstrate such an advantage at the end of this section.

If we are allowed to simulate target channel $\mathcal{N}$ up to error $\varepsilon$, it is equivalent to exactly simulate a channel $\mathcal{L}$ which satisfies $\frac{1}{2}\|\mathcal{L}-\mathcal{N}\|_\diamond \leq \varepsilon$. From the definition of $\varepsilon$-ROC of channel, we have $\min_{\mathcal{L}} C_R(\mathcal{L}) = C_R^\varepsilon(\mathcal{N})$, which leads to the following corollary directly.

**Corollary 5.1.** *Given an arbitrary target channel $\mathcal{N}$ and a fixed maximally coherent state $\Psi_m$ of dimension $m \geq 2$, the maximal success probability of channel simulation with MIO up to error $\varepsilon$ is*

$$\begin{aligned} P_{\mathrm{MIO}}(\mathcal{N}, \Psi_m, \varepsilon) &= \max_{\mathcal{L}} P_{\mathrm{MIO}}(\mathcal{L}, \Psi_m, 0) \\ &= \min\left\{1, \frac{m-1}{\min_{\mathcal{L}} C_R(\mathcal{L})}\right\} \\ &= \min\left\{1, \frac{m-1}{C_R^\varepsilon(\mathcal{N})}\right\}, \end{aligned} \tag{5.14}$$

*where $\mathcal{L}$ are $\varepsilon$-close to the target channels $\mathcal{N}$.*

From Corollary 5.1, we can approximately simulate the target channel. The more error we allow, the higher the success probability that we can simulate it.

For a further step, instead of using a maximally coherent state, we consider a general pure coherent state. We obtain the following corollary, which gives a lower bound of the maximal success probability for the general pure resource state.

**Corollary 5.2.** *Given target channel $\mathcal{N}$ and coherent pure state $\psi = |\psi\rangle\langle\psi|$, where $|\psi\rangle = \sum_{i=1}^n \psi_i |i\rangle$, $\psi_i \neq 0, n \geq 2$, the maximal success probability of channel simulation with MIO up to error $\varepsilon$ is lower bounded by*

$$P_{\mathrm{MIO}}(\mathcal{N}, \psi, \varepsilon) \geq P_{\mathrm{MIO}}(\mathcal{N}, \Psi_m, \varepsilon) \times P_{\mathrm{MIO}}^{\mathrm{distill}}(|\psi\rangle\langle\psi| \to \Psi_m), \tag{5.15}$$

$$\geq \min\left\{1, \frac{m-1}{C_R^\varepsilon(\mathcal{N})} \times \frac{n^2}{m(\sum_{i=1}^n |\psi_i|^{-2})}\right\}, \tag{5.16}$$

*where $P_{\mathrm{MIO}}^{\mathrm{distill}}(|\psi\rangle\langle\psi| \to \Psi_m)$ is the success probability of coherence distillation with MIO from input state $|\psi\rangle\langle\psi|$ to the rank-$m$ maximally coherent state for arbitrary integer $m \geq 2$.*

*Proof.* Before starting the proof, we need to note that arbitrary coherent pure state $|\psi\rangle$ can be distilled into maximal coherent state $\Psi_m$ by MIO with non-zero probability [155], which is

$$P_{\mathrm{MIO}}^{\mathrm{distill}}(\psi \to \Psi_m) \geq \frac{n^2}{m(\sum_{i=1}^n |\psi_i|^{-2})} > 0, \tag{5.17}$$

where $m \geq 2$ is an integer.

In order to simulate channel $\mathcal{N}$ using a given coherent pure state $\psi$, one feasible method is to probabilistically distill the maximally coherent state $\Psi_m$ from

$\psi$ first, then probabilistically implement the target channel using the distilled maximally coherent state. Therefore, the maximal success probability is lower bounded as

$$P_{\mathrm{MIO}}(\mathcal{N},\psi,\varepsilon) \geq P_{\mathrm{MIO}}(\mathcal{N},\Psi_m,\varepsilon) \times P_{\mathrm{MIO}}^{\mathrm{distill}}(|\psi\rangle\langle\psi| \to \Psi_m). \tag{5.18}$$

By considering Eq. (5.17) and Corollary 5.1, we arrive at the lower bound of the maximal success probability as

$$P_{\mathrm{MIO}}(\mathcal{N},\psi,\varepsilon) \geq \min\left\{1, \frac{m-1}{C_R^{\varepsilon}(\mathcal{N})}\right\} \times \frac{n^2}{m(\sum_{i=1}^{n}|\psi_i|^{-2})}. \tag{5.19}$$

The proof is completed. □

This corollary implies that if we take an arbitrary coherent pure state as a resource state, we can simulate an arbitrary channel by MIO with non-zero probability. In other words, any coherent pure state is "useful" in the probabilistic channel simulation with MIO. Even if only having a few resources, coherent pure states $\psi$ possess the potential to simulate the target channel, albeit with a small probability of success $P_{\mathrm{MIO}}(\mathcal{N},\psi,\varepsilon) > 0$.

So far, we have studied the probabilistic channel simulation with MIO by considering the resource states as maximally coherent states and coherent pure states, respectively. For a more general case, where the given resource state is a coherent mixed state, we provide an efficiently computable SDP to achieve the maximal success probability of channel simulation. Due to the non-linearity of Eq. (5.4b), we cannot formulate Problem 1 into an SDP directly. We thus consider a generalization of the set of free operations $O$, namely the set of free sub normalized quantum operations $O_{sub}$ which consist of completely positive and trace non-increasing maps that are free in the sense of $O$. By adopting this generalization, we convert the probability optimization into the following expression. A similar technique is also applied in Ref. [175, 176].

**Lemma 5.1.** *For any triplet $(\mathcal{N},\omega,\varepsilon)$ and operation class $\mathcal{O}$, the maximal success probability of coherence channel simulation $P_{\mathcal{O}}(\mathcal{N},\omega,\varepsilon) = \max\{p \in \mathbb{R}_+ | \mathcal{E}(\omega\otimes\cdot) = p\mathcal{L}(\cdot), \frac{1}{2}||\mathcal{L}-\mathcal{N}||_\diamond \leq \varepsilon, \mathcal{E}\in\mathcal{O}_{\mathrm{sub}}\}$, where $\mathcal{L}$ are the implemented channels.*

*Proof.* For any quantum operation $\Lambda(\omega\otimes\rho) = |0\rangle\langle 0|\otimes\mathcal{E}_0(\omega\otimes\rho)+|1\rangle\langle 1|\otimes\mathcal{E}_1(\omega\otimes\rho)$, where $\mathcal{E}_0$ and $\mathcal{E}_1$ are two subnormalized operations, we can check that $\Lambda\in\mathcal{O}$ if and only if $\mathcal{E}_0,\mathcal{E}_1\in\mathcal{O}_{\mathrm{sub}}$, and $\mathcal{E}_0+\mathcal{E}_1$ is trace preserving. Thus, the optimization in Eq. (5.4) is equivalent to finding the optimal subnormalized operations $\mathcal{E}_0$ and $\mathcal{E}_1$ such that $\mathcal{E}_0(\omega\otimes\rho) = p\mathcal{L}(\rho), \mathcal{E}_1(\omega\otimes\rho) = (1-p)\mathbb{I}/d, \frac{1}{2}||\mathcal{L}-\mathcal{N}||_\diamond \leq \varepsilon$, and $\mathcal{E}_0+\mathcal{E}_1$ is trace preserving. Since we can take $\mathcal{E}_1(\omega\otimes\rho) = (1-\mathrm{Tr}[\mathcal{E}_0(\omega\otimes\rho)])\frac{\mathbb{I}}{d}$, without compromising the success probability, the maximal success probability of coherence channel simulation is only dependent on $\mathcal{E}_0$, which complete the proof. □

By generalizing the free operation into the subnormalized version, Lemma 5.1 simplifies the optimization of the maximal success probability. For further steps, we formulate the success probability optimization as an efficiently computable SDP, which is shown as follows.

**Proposition 5.1.** *For a given triplet $(\mathcal{N}_{A\to B}, \omega, \varepsilon)$, the maximal success probability to simulate the target channel with MIO is given by $P_{\text{MIO}}(\mathcal{N}_{A\to B}, \omega, \varepsilon) = 1/t_{\min}$, where $t_{\min}$ is given by*

$$
\begin{aligned}
t_{\min} = & \min \ t; \\
& \text{s.t.} \operatorname{Tr}_R[J_{\tilde{\mathcal{E}}}(\omega^T \otimes \mathbb{I}_A \otimes \mathbb{I}_B)] = J_{\mathcal{L}_{AB}}; && \text{(5.20a)} \\
& J_{\tilde{\mathcal{E}}} \geq 0, \operatorname{Tr}_B[J_{\tilde{\mathcal{E}}}] \leq t\mathbb{I}_{RA}; && \text{(5.20b)} \\
& \operatorname{Tr}_{RA}[J_{\tilde{\mathcal{E}}}(|ij\rangle\langle ij|_{RA}^T \otimes \mathbb{I}_B)] \\
& = \Delta(\operatorname{Tr}_{RA}[J_{\tilde{\mathcal{E}}}(|ij\rangle\langle ij|_{RA}^T \otimes \mathbb{I}_B)]), \ \forall i, j; && \text{(5.20c)} \\
& \operatorname{Tr}_B[J_{\mathcal{L}}] = \mathbb{I}_A; && \text{(5.20d)} \\
& Z \geq 0, Z \geq J_{\mathcal{L}} - J_{\mathcal{N}}, \operatorname{Tr}_B[Z] \leq \varepsilon \mathbb{I}_A. && \text{(5.20e)}
\end{aligned}
$$

*Proof.* To get rid of the non-linearity, we consider the map $\tilde{\mathcal{E}} = t\mathcal{E}$, where $t = 1/p$, the inverse of success probability, and $\mathcal{E}$ is the subnormalized MIO. The notations of $J_{\tilde{\mathcal{E}}}$ and $J_{\mathcal{L}}$ are the Choi-Jamiołkowski matrices of maps $\tilde{\mathcal{E}}$ and $\mathcal{L}$, respectively. $J_{\mathcal{N}}$ is the Choi-Jamiołkowski matrix of the target channel $\mathcal{N}$. Eq. (5.20a) corresponds to the constraint $\tilde{\mathcal{E}}(\omega \otimes \cdot) = \mathcal{L}(\cdot)$. Eq. (5.20b) and Eq. (5.20c) implies that $\mathcal{E}$ is a subnormalized MIO. From the result in Ref. [172] (Sec.4), Eq. (5.20d) and Eq. (5.20e) guarantee that the simulated channel $\mathcal{L}$ includes the target channels $\mathcal{N}$ up to error $\varepsilon$, which is $\frac{1}{2}||\mathcal{L} - \mathcal{N}||_\diamond \leq \varepsilon$. □

A general qubit unitary possesses four real parameters. However, we can transform unitaries into each other without any additional cost by incorporating coherent unitaries before or after them. This observation implies the existence of an equivalence relation among qubit unitaries up to coherent unitaries. A unique representative of each equivalence class [48, 72] is given in Eq. (5.21)

$$
U_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{5.21}
$$

where $\theta \in [0, \pi/4]$.

Here we choose the unitary channel $\mathcal{U}_\theta^{(l)}(\cdot) = U_\theta^{\otimes l} \cdot U_\theta^{\dagger \otimes l}$, as the target channel to simulate. The success probability of unitary channel simulation $\mathcal{U}_\theta^{(l)}$ can be derived from the robustness of the channel. We take the default that $\mathcal{U}_\theta$ refers to $l = 1$. The specific statement is shown in the following.

**Proposition 5.2.** *Given the triplet $(\mathcal{U}_\theta^{(l)}, \Psi_m, \varepsilon = 0)$, the maximal success prob-*

*ability of channel simulation is*

$$P_{\mathrm{MIO}}(\mathcal{U}_\theta^{(l)}, \Psi_m, \varepsilon = 0) = \min\left\{1, \frac{m-1}{C_R(\mathcal{U}_\theta^{(l)})}\right\} \tag{5.22}$$

$$= \min\left\{1, \frac{m-1}{(1+\sin 2\theta)^l - 1}\right\}. \tag{5.23}$$

*Proof.* The cohering power [177, 178] of a channel $P_R(\mathcal{N})$ is defined as

$$P_R(\mathcal{N}) = \max_i \log(1 + C_R(\mathcal{N}(|i\rangle\langle i|))). \tag{5.24}$$

We denote $i^*$ to be the optimal solution, i.e., $P_R(\mathcal{N}) = \log(1 + C_R(\mathcal{N}(|i^*\rangle\langle i^*|)))$. In [72], it has been proved that the cohering power of a channel $P_R(\mathcal{N})$ is equivalent to the log-robustness of the channel, which is $P_R(\mathcal{N}) = \log(1 + C_R(\mathcal{N}))$. It is straightforward to have

$$C_R(\mathcal{N}) = C_R(\mathcal{N}(|i^*\rangle\langle i^*|)). \tag{5.25}$$

If we replace the unitary channel $\mathcal{U}_\theta$ into the equation, we can deduce $C_R(\mathcal{U}_\theta) = C_R(\mathcal{U}_\theta(|i^*\rangle\langle i^*|))$ directly. For a single-qubit channel, the robustness of a state is equal to its $l_1$-norm of coherence [158], which is the summation of the absolute value of all non-diagonal elements [74]. Then, we have

$$C_R(\mathcal{U}_\theta) = C_R(\mathcal{U}_\theta(|i^*\rangle\langle i^*|)) = C_{l_1}(\mathcal{U}_\theta(|i^*\rangle\langle i^*|)) = \sin 2\theta. \tag{5.26}$$

Note that the robustness is multiplicative under the tensor product of states [179], specifically

$$1 + C_R(\rho_1 \otimes \rho_2) = (1 + C_R(\rho_1))(1 + C_R(\rho_2)). \tag{5.27}$$

We correspondingly have

$$C_R(\mathcal{U}_\theta^{(l)}) = (1 + \sin 2\theta)^l - 1. \tag{5.28}$$

Recall Theorem 5.1, the maximal success probability is the ratio between the robustness of the resource state and the robustness of the target coherent channel, also the success probability $p$ cannot exceed 1. Then we have

$$P_{\mathrm{MIO}}(\mathcal{U}_\theta^{(l)}, \Psi_m, \varepsilon = 0) = \min\left\{1, \frac{m-1}{(1+\sin 2\theta)^{(l)} - 1}\right\}, \tag{5.29}$$

which completes the proof. □

Hence, we obtain the analytic expression of the maximal success probability of the exact simulation of unitary channel $\mathcal{U}_\theta^{(l)}$ with MIO. We also conduct numerical experiments to show the approximate probabilistic channel simulation. We consider 2-qubit unitary channels $\mathcal{U}_\theta^{l=2}$ $(0 \leq \theta \leq \pi/4)$ as the target channels. The resource state used is the rank-2 maximally coherent state $\Psi_2$. For different error tolerance $\varepsilon \in \{0, 0.05, 0.1, 0.15, 0.2\}$, the maximal success probabilities of channel simulation with MIO are calculated by the SDP given in Proposition 5.1. The results are shown in Figure. 5.2

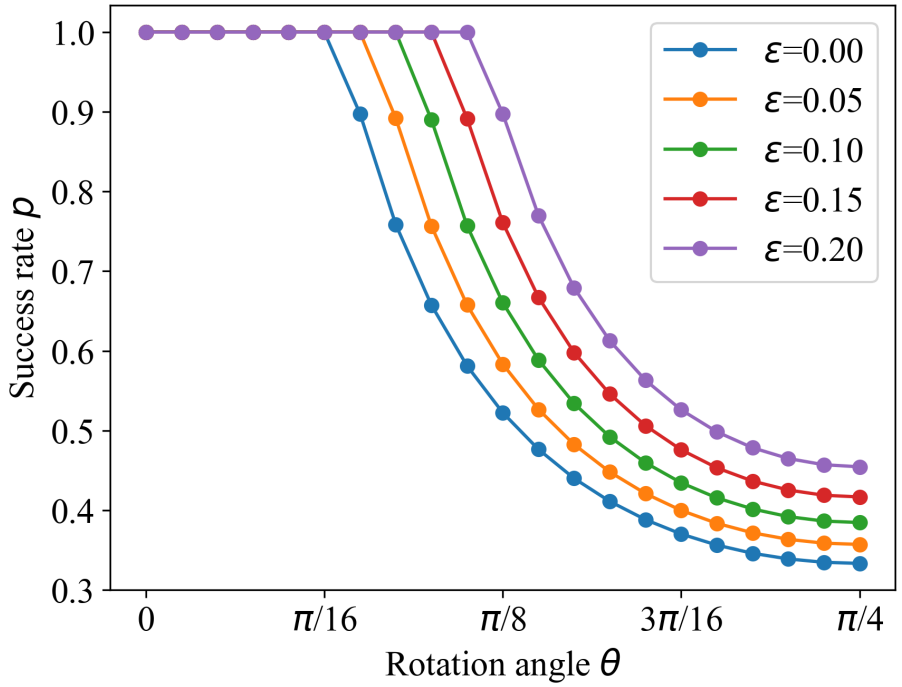



Figure 5.2: Success probability $p$ of unitary channel simulation with coherent state $\Psi_2$ and MIO. The rotation angle $\theta$ is the parameter of target unitary channel $\mathcal{U}_\theta^{(2)}$, which is chosen from 0 to $\pi/4$. The resource state is the rank-2 maximally coherent state $\Psi_2$. The five curves from bottom to top correspond to the error tolerance $\varepsilon$ equaling to $\{0, 0.05, 0.1, 0.15, 0.2\}$, respectively.

The robustness of the unitary channel $C_R(\mathcal{U}_\theta^{(2)})$ increases with respect to the increase of the rotation angle $\theta$. When the angle is small, the resource state $\Psi_2$ has more resources than the required resource of the task, so we can simulate such a channel deterministically. When the rotation angle exceeds some threshold, the coherence of the resource state $\Psi_2$ is not enough for its deterministic implementation. Thus, one can never implement the target channel exactly in the deterministic scenario, and can only realize it approximately. While in the probabilistic scenario, one can exactly implement the target channel at the expense of the reduced success probability.

## 5.4 Probabilistic channel simulation with DIO

It was proved in Ref. [46] that if the target channel $\mathcal{N}$ is not a resource nonactivating channel, then it cannot be simulated exactly by DIO with any resource state deterministically. We further show that such an exact simulation is not

possible even in probabilistic scenarios. Also, SDP for calculating the success probability of the channel simulation is provided.

In the previous part, we have shown that arbitrary channels can be simulated by using an appropriate resource state and MIO. But for DIO, it is not the same story. Explicitly, we have the following theorem.

**Theorem 5.2.** [46] (Appendix Eq.(109)) *If a quantum channel $\mathcal{N}$ can be implemented by a DIO channel $\mathcal{M}$ using resource state $\omega$, i.e. $\mathcal{M}(\omega\otimes\cdot)=\mathcal{N}(\cdot)$, then $\mathcal{N}$ satisfies $\Delta\circ\mathcal{N}=\Delta\circ\mathcal{N}\circ\Delta$.*

*Proof.* Since $\mathcal{N}(\rho)=\mathcal{M}(\omega\otimes\cdot)$ holds for any quantum state, we directly have $\Delta\circ\mathcal{N}(\cdot)=\Delta\circ\mathcal{M}(\omega\otimes\cdot)$. Considering $\mathcal{M}$ is a DIO channel, which means $\Delta\circ\mathcal{M}(\omega\otimes\cdot)=\mathcal{M}\circ\Delta(\omega\otimes\cdot)$, we arrive at

$$\Delta\circ\mathcal{N}(\cdot)=\mathcal{M}\circ\Delta(\omega\otimes\cdot). \tag{5.30}$$

If the input state is applied to the completely dephasing map, we directly have

$$\Delta\circ\mathcal{N}(\Delta(\cdot))=\mathcal{M}\circ\Delta(\omega\otimes\Delta(\cdot)) \tag{5.31}$$

$$\Rightarrow\Delta\circ\mathcal{N}\circ\Delta(\cdot)=\mathcal{M}\circ\Delta(\omega\otimes\cdot). \tag{5.32}$$

Combine Eq. (5.30) and Eq. (5.32), we have

$$\Delta\circ\mathcal{N}(\cdot)=\Delta\circ\mathcal{N}\circ\Delta(\cdot) \tag{5.33}$$

Eq. (5.33) holds for any quantum state, which implies $\Delta\circ\mathcal{N}=\Delta\circ\mathcal{N}\circ\Delta$, i.e., $\mathcal{N}$ is resource nonactivating channel. The proof is complete. □

In other words, this theorem implies that if a quantum channel $\mathcal{N}$ does not satisfy the condition $\Delta\circ\mathcal{N}=\Delta\circ\mathcal{N}\circ\Delta$ (also known as a resource nonactivating channel [170]), then it cannot be simulated by DIO. Also, it is straightforward to extend this theorem to the probabilistic scenario, which is shown in the following corollary 5.3.

**Corollary 5.3.** *If a quantum channel $\mathcal{N}$ can be simulated exactly by a DIO channel $\mathcal{M}$ using resource state $\omega$ with non-zero probability $p$, i.e., $\mathcal{M}(\omega\otimes\cdot)=p\,|0\rangle\langle 0|\otimes\mathcal{N}(\cdot)+(1-p)\,|1\rangle\langle 1|\otimes\mathrm{Tr}[\cdot]\frac{\mathbb{I}}{d}$, where $\mathbb{I}$ is the identity, $d$ is the dimension, then $\mathcal{N}$ satisfies $\Delta\circ\mathcal{N}=\Delta\circ\mathcal{N}\circ\Delta$.*

*Proof.* Denote $\mathcal{N}'(\cdot)=p\,|0\rangle\langle 0|\otimes\mathcal{N}(\cdot)+(1-p)\,|1\rangle\langle 1|\otimes\mathrm{Tr}[\cdot]\frac{\mathbb{I}}{d}$. From Theorem 5.2, $\mathcal{N}'$ is a nonactivating channel, then we directly have $\Delta\circ\mathcal{N}=\Delta\circ\mathcal{N}\circ\Delta$, which completes the proof. □

Theorem 5.2 and Corollary 5.3 tell us that not all quantum channels can be simulated exactly by DIO, even probabilistically. Take single-qubit unitary channel $\mathcal{U}_\theta$ and quantum state$|+\rangle\langle+|$ as an example. One can easily obtain

$$\Delta\circ\mathcal{U}_\theta(|+\rangle\langle+|)=\begin{pmatrix}\frac{1}{2}-\cos\theta\sin\theta & 0\\ 0 & \frac{1}{2}+\cos\theta\sin\theta\end{pmatrix}$$

$$\Delta\circ\mathcal{U}_\theta\circ\Delta(|+\rangle\langle+|)=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix}.$$

$\Delta \circ \mathcal{U}_\theta \neq \Delta \circ \mathcal{U}_\theta \circ \Delta$ represents that a single qubit unitary channel $\mathcal{U}_\theta$ is not a resource nonactivating channel. Thus, qubit unitary channel cannot be exactly simulated by DIO even probabilistically.

If a quantum channel $\mathcal{N}$ can be simulated by DIO, then the success probability can be efficiently computed by the SDP as shown in Proposition 5.3.

**Proposition 5.3.** *For a given triplet $(\mathcal{N}_{A\to B}, \omega, \varepsilon)$, the maximal success probability to simulate the target channel with DIO is given by $P_{\text{DIO}}(\mathcal{N}_{A\to B}, \omega, \varepsilon) = 1/t_{\min}$, where $t_{\min}$ is given by*

$$
\begin{aligned}
t_{\min} = & \min \ t; \\
& \text{s.t.} \operatorname{Tr}_R[J_{\tilde{\mathcal{E}}}(\omega^T \otimes \mathbb{I}_A \otimes \mathbb{I}_B)] = J_{\mathcal{L}_{AB}}; && \text{(5.34a)} \\
& J_{\tilde{\mathcal{E}}} \geq 0, \operatorname{Tr}_B[J_{\tilde{\mathcal{E}}}] \leq t\mathbb{I}_{RA}; && \text{(5.34b)} \\
& \operatorname{Tr}_{RA}[J_{\tilde{\mathcal{E}}}(|ij\rangle\langle ij|_{RA}^T \otimes \mathbb{I}_B)] \\
& = \Delta(\operatorname{Tr}_{RA}[J_{\tilde{\mathcal{E}}}(|ij\rangle\langle ij|_{RA}^T \otimes \mathbb{I}_B)]), \ \forall i, j; && \text{(5.34c)} \\
& \Delta(\operatorname{Tr}_{RA}[J_{\tilde{\mathcal{E}}}(|ij\rangle\langle mn|_{RA}^T \otimes \mathbb{I}_B)]) = 0, \ \forall ij \neq mn; && \text{(5.34d)} \\
& \operatorname{Tr}_B[J_{\mathcal{L}}] = \mathbb{I}_A; && \text{(5.34e)} \\
& Z \geq 0, Z \geq J_{\mathcal{L}} - J_{\mathcal{N}}, \operatorname{Tr}_B[Z] \leq \varepsilon \mathbb{I}_A; && \text{(5.34f)}
\end{aligned}
$$

Compared with the SDP for success probability of channel simulation with MIO in Eq. (5.20), the SDP in Eq. (5.34) has one more constraint as Eq. (5.34d), which corresponds to $\Delta(\mathcal{M}(|i\rangle\langle j|)) = 0, \ \forall i \neq j$.

We conduct a numerical experiment to simulate a target channel with DIO. We consider a 2-qubit random channel as the target channel, the Choi-Jamiołkowski matrix of which is shown in Appendix C. The resource state is a coherent pure state, $|\psi\rangle = \sqrt{\alpha}\,|0\rangle + \sqrt{1-\alpha}\,|1\rangle$ for $\alpha \in [0, 0.5]$. The state's coherence increases as parameter $\alpha$ increases. When $\alpha = 0.5$, the resource state is a rank-2 maximally coherent state $\Psi_2$. The maximal success probabilities of channel simulation by DIO with different error tolerances and coherent states are calculated by the SDP (Eq. (5.34)).

From the results shown in Figure. 5.3, we can see that the probabilistic channel simulation with smaller probability $p$ admits smaller error tolerance. In other words, probabilistic relaxation is also effective for DIO. As $\alpha$ increases (larger coherence), the success probability of simulation increases faster and gets closer to the case of the maximally coherent state ($\alpha = 0.5$).

## 5.5 Conclusion

We have focused on the probabilistic channel simulation with MIO and DIO, respectively. For the MIO part, we talked about three cases: If the resource state is the maximally coherent state, we provide an analytical expression for the maximal success probability. If we select any pure coherent state as the resource state, the maximal success probability is guaranteed to be greater than zero. If the resource state $\omega$ is a general coherent state, we offer an efficiently

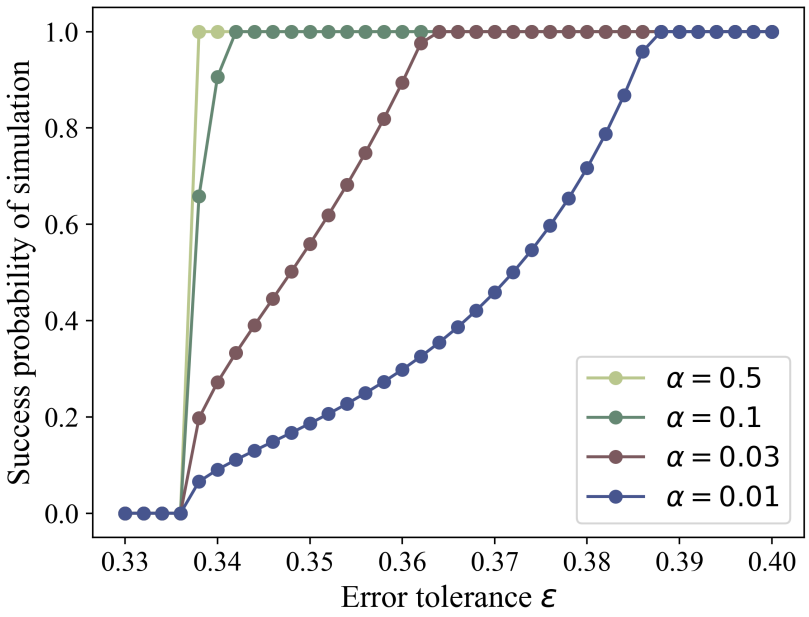



Figure 5.3: Success probability of simulating a random channel by DIO with different error tolerances. The resource state is pure coherent state $|\psi\rangle = \sqrt{\alpha}\,|0\rangle + \sqrt{1-\alpha}\,|0\rangle$, with $\alpha \in \{0.01, 0.03, 0.1, 0.5\}$.

computable SDP for achieving the maximal success probability. For the DIO part, we show that not all quantum channels can be exactly simulated by DIO, even probabilistically. Furthermore, we present an efficiently computable SDP for attaining the maximal success probability of channel simulation with DIO.

For further research, it would be interesting to apply the framework of probabilistic channel simulation to other important quantum resources such as entanglement and magic.

# Chapter 6

# Conclusion

In this thesis, we attempt to reduce the requirement of quantum resources in different quantum tasks with the help of optimization tools such as variational quantum algorithms (VQA) and semidefinite programs (SDP). This thesis focuses on three tasks including expectation value estimation, non-linear features estimation, as well as channel simulation using coherence. The details are as follows:

In Chapter 3, we propose the VQHE algorithm to reduce the Pauli norm of Hamiltonian such that the overheads for expectation value and Hamiltonian simulation can be reduced. We first develop the theory to convert the Pauli norm optimization problem into the vector $l_1$-norm minimization problem, and then design the cost function and parameterized quantum circuits to minimize Pauli norm variationally. We then display how to apply the proposed VQHE algorithm to expectation value estimation and Hamiltonian simulation. In the task of expectation value estimation, we also emphasize that the proposed algorithm is compatible with grouping, such that the measurement time can be reduced further. The numerical experiments are conducted by applying the VQHE algorithms to the Ising Hamiltonian and some molecules' Hamiltonian, which shows the effectiveness of VQHE algorithm.

In Chapter 4, we establish that when quantum states are distorted by noises, the original moment information can still be retrieved through post-processing if and only if the noise is invertible. Furthermore, our proposed method, called observable shift, outperforms existing methods in two aspects: (1) The proposed method requires lower quantum sampling complexity than the existing one, which implies the superiority of entangled protocols over product protocols. This contrasts with the multiplicativity of overhead observed in existing methods for quantum error mitigation. (2) The observable shift method is easier to implement than the existing method, making our method more friendly to quantum devices. We also propose the construction of a protocol to retrieve the depolarizing channel of large-size quantum systems. Our findings have implications for the dependable estimation of non-linear information in quantum systems and can influence various applications, including entanglement spec-

troscopy and ground-state property estimation.

In Chapter 5, we focus on the probabilistic channel simulation with MIO and DIO, respectively. For the MIO part, we talked about three cases: If the resource state is the maximally coherent state, we provide an analytical expression for the maximal success probability. If we select any pure coherent state as the resource state, the maximal success probability is guaranteed to be greater than zero. If the resource state $\omega$ is a general coherent state, we offer an efficiently computable SDP for achieving the maximal success probability. For the DIO part, we show that not all quantum channels can be exactly simulated by DIO, even probabilistically. Furthermore, we present an efficiently computable SDP for attaining the maximal success probability of channel simulation with DIO.

This thesis contributes to reducing the requirement of quantum resources in different quantum tasks, making the quantum devices more efficient in particular tasks such as expectation value estimation, non-linear features estimation, and channel simulation using coherence. We are confident that this thesis pushes quantum computers a step toward the next milestone. One interesting research in the future is to apply our method on the quantum device to check the actual performance. Besides, another interesting research topic is applying optimization methods such as SDP or VQA to solve the problems that quantum computers will really encounter.

# Appendix A

# Measurement times of expectation value estimation

Estimating the expectation value of Hamiltonian is a critical step in many quantum algorithms such as variational quantum eigensolver (VQE) [24]. The expectation value of a Hamiltonian $H$ with respect to a quantum state $|\psi\rangle$ is given by $\langle H\rangle = \langle\psi| H |\psi\rangle$. This quantity represents the average energy of the quantum state $|\psi\rangle$ when measured on the basis of the eigenstates of the Hamiltonian $H$. For a given Hamiltonian $H$, we can write it as

$$H = \sum_{i=1}^{L} h_i P_i, \tag{A.1}$$

where $L$ is the number of terms, $h_i$ is the coefficient of Pauli terms $P_i$. To estimate the expectation value of Hamiltonian $H$, one can estimate the expectation value of each Pauli term and sum them up by post-processing $\langle H\rangle = \sum_i h_i \langle P_i\rangle$, with variance [55, 56]

$$\epsilon^2 = \sum_{i}^{L} \frac{|h_i|^2 \mathrm{Var}[P_i]}{S_i}, \tag{A.2}$$

where $S_i$ is the measurement time of each of the operators $P_i$, Var denotes the variance in the measurement of the operator for the given trial state.

Generally, the variance of Pauli $\mathrm{Var}[P_i]$ is unknown in advance. However, since the terms $P_i$ are Pauli, the variance is upper bounded by $\mathrm{Var}[P_i] \leq 1$ [55, 56]. One can confirm (e.g., via the use of Lagrange multipliers [57]) that the least number of measurements required to bound variance below $\epsilon^2$ can be found by choosing $S_i \propto |h_i|$. This implies the total measurement time $N$ is upper bounded by

$$N \leq \frac{\|H\|_P^2}{\epsilon^2}, \tag{A.3}$$

where $\|H\|_P = \sum_i |h_i|$ is called Pauli norm.

To reduce the total sampling times for expectation value estimation, grouping strategy is proposed. The main idea is to partition the Pauli terms into groups, and in each group, all elements commute with each other. Since commuting Pauli terms can be measured simultaneously, all elements in a group can be measured at the same time. This kind of method is called *Clique Cover*. If the number of such groups is minimum, it is called *Minimum Clique Cover*.

## A.1 Qubit wise commuting (QWC)

Qubit-wise commuting (QWC)[103] is one of the simplest methods to partition the Hamiltonian terms, which only group the terms that commute in qubit wise. For example $\{I_1X_2, Z_1X_2\}$ is a QWC group because $I_1$ commutes with $X_1$ and $X_2$ commutes with $X_2$. While the case of $\{X_1X_2, Z_1Z_2\}$ is not a QWC group for $X_1$ and $Z_1$ do not commute. When doing the measurement, we only need to apply for local Clifford unitary $U$. This method can decrease 70% terms.

## A.2 General commuting (GC)

The general commuting method [103, 104] is a more powerful method compared with QWC. Note that QWC is simply the subset of GC corresponding to the case where the number of non-commuting indices is 0. When we make measurements, we need to apply the unitary $U$ transfer to the GC group into QWC group such that we can make measurements directly, i.e.,

$$\langle H\rangle = \sum_i \langle\phi| H_i |\phi\rangle \tag{A.4}$$

$$= \sum_i \langle\phi| U_i^\dagger U_i H_i U_i^\dagger U_i |\phi\rangle \tag{A.5}$$

$$= \sum_i \langle\phi| U_i^\dagger A_i U_i |\phi\rangle \tag{A.6}$$

$$= \sum_i \langle\psi| A_i |\psi\rangle = \sum_i \langle A_i\rangle, \tag{A.7}$$

where $A_i$ is the corresponding QWC group. The unitary $U$ can be decomposed into Clifford gates. This method can decrease two two-magnitude number of terms.

After partitioning into GC groups, we need to apply unitaries first, and then make simultaneous measurements. Since the unitaries are Clifford [180], the coefficient won't change after the unitary.

## A.3 The grouped Pauli norm

For a given operator $H$, we have

$$H = \sum_{i=1}^{N} H_i = \sum_{i=1}^{N} \sum_{j=1}^{m_i} a_{ij} P_{ij}, \tag{A.8}$$

where $N$ is the number of collections of mutually commuting operators, $m_i$ the number of operators in collection $i$, $P_{ij}$ is the $j$ Pauli operator in the $i$th collection and $a_{ij} \in \mathbb{R}$ is its coefficient.

It has been shown [102] that the number of $m_i$ measurements of collection $i$ is

$$n_i = \frac{1}{\epsilon^2} \sqrt{\mathrm{Var}[H_i]} \sum_{j=1}^{N} \sqrt{\mathrm{Var}[H_j]} \tag{A.9}$$

Therefore, the total number of measurements is

$$M = \sum_{i}^{N} n_i = \left( \frac{1}{\epsilon} \sum_{i=1}^{N} \sqrt{\mathrm{Var}[H_i]} \right)^2, \tag{A.10}$$

where

$$\mathrm{Var}[H_i] = \mathrm{Cov}[H_i, H_i] = \langle H_i^2 \rangle - \langle H_i \rangle^2 \tag{A.11}$$

**Theorem A.1.** *[103] Given $M_1, M_2$, two commuting but non-identical Pauli strings, $\mathbb{E}(Cov[M_1, M_2]) = 0$ where the expectation is taken over a uniform distribution over all possible state vectors.*

For convenience, we assume we do not know the quantum state, and replace the variances and covariances by their expectation value over uniform spherical distribution. From Theorem A.1, it is straightforward to have

$$M = \sum_{i}^{N} n_i = \left( \frac{1}{\epsilon} \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{m_i} |h_{ij}|^2} \right)^2 \tag{A.12}$$

To make a comparison with the ungrouped Pauli norm of Hamiltonian, we denote it as the grouped Pauli norm

$$\|H\|_{gp} = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{m_i} |h_{ij}|^2}. \tag{A.13}$$

Here we give a simple example by setting the Hamiltonian as $H = 3XI - YY + 2ZZ$. The first step is to sort it by the absolute value of coefficients, which will be $\{3XI, 2ZZ, -1YY\}$. At this stage, no collection is created, so we need to create one for the first element, such that $\{3XI\}$. For the second element $2ZZ$, we have to create another collection for $ZZ$ anti-commutes to $XI$. Now

the collections will be $\{3XI\}, \{2ZZ\}$. For the third element $-1YY$, it does not commute with the element in the first collection, thus we need to check the commutativity of the second collection. Luckily, the third element commutes with all elements in the second collection, thus we can put it into this collection, which becomes $\{3XI\}, \{2ZZ, -1YY\}$. In this case, the corresponding grouped Pauli norm is $\|H\|_{gp} = 5.24$, while the Pauli norm without grouping is $\|H\|_P = 6$.

# Appendix B

# Proof for Retrieving Non-Linear Features from Noisy Quantum States

## B.1 Proof of the existence of observable for high-order moment

**Lemma B.1.** *Suppose $\rho$ is an $n$-qubit state. Then for a positive integer $k$, there exists an $nk$-qubit Hermitian matrix $H$ such that* $\mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]$.

*Proof.* Consider the computational basis $\{|x\rangle\}_x$ for an $n$-qubit system. The desired Hermitian matrix can be constructed from an $nk$-qubit cyclic permutation matrix

$$S_k = \sum_{\boldsymbol{x}:=x_1\cdots x_k=0}^{2^{nk}-1} \bigotimes_{j=1}^{k} |x_{\pi(j)}\rangle\langle x_j| \,, \tag{B.1}$$

where $\pi = (1, \ldots, k)$ is a permutation function in cyclic notation. Note that the conjugate transpose of this matrix permutes the subsystems in an inverse order i.e.

$$S_k^\dagger = \sum_{\boldsymbol{x}} \bigotimes_{j=1}^{k} |x_j\rangle\langle x_{\pi(j)}| = \sum_{\boldsymbol{x}} \bigotimes_{j=1}^{k} |x_{\pi^{-1}(j)}\rangle\langle x_j| \,. \tag{B.2}$$

In the rest of this proof, we will prove that

$$\mathrm{Tr}[S_k\rho^{\otimes k}] = \mathrm{Tr}[S_k^\dagger\rho^{\otimes k}] = \mathrm{Tr}[\rho^k] \tag{B.3}$$

for all $\rho$, and hence the result follows by setting $H = \frac{1}{2}\left(S_k + S_k^\dagger\right)$.

To start with, decomposing $\rho = \sum_{x,y} C_{xy} |x\rangle\langle y|$ gives

$$\rho^k = \sum_{\boldsymbol{x},\boldsymbol{y}} \prod_{j=1}^{k} C_{x_j y_j} |x_j\rangle\langle y_j| = \sum_{\boldsymbol{x},y_k} \left( \prod_{j=1}^{k-1} C_{x_j x_{\pi(j)}} \right) c_{x_k y_k} |x_1\rangle\langle y_k|$$

$$\implies \mathrm{Tr}[\rho^k] = \sum_{\boldsymbol{x}} \prod_{j=1}^{k} C_{x_j x_{\pi(j)}}, \tag{B.4}$$

$$\rho^{\otimes k} = \sum_{\boldsymbol{x},\boldsymbol{y}} \bigotimes_{j=1}^{k} C_{x_j y_j} |x_j\rangle\langle y_j| . \tag{B.5}$$

Apply $S_k$ on $\rho^{\otimes k}$ and we have

$$S_k \rho^{\otimes k} = \sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \bigotimes_{j=1}^{k} |z_{\pi(j)}\rangle\langle z_j| \cdot C_{x_j y_j} |x_j\rangle\langle y_j| = \sum_{\boldsymbol{x},\boldsymbol{y}} \bigotimes_{j=1}^{k} C_{x_j y_j} |x_{\pi(j)}\rangle\langle y_j| \tag{B.6}$$

$$\implies \mathrm{Tr}[S_k \rho^{\otimes k}] = \sum_{\boldsymbol{x}} \prod_{j=1}^{k} C_{x_j x_{\pi(j)}} = \mathrm{Tr}[\rho^k]. \tag{B.7}$$

Lastly, the statement $\mathrm{Tr}[S_k^\dagger \rho^{\otimes k}]$ holds from the fact that $\rho$ is Hermitian and thus

$$\begin{aligned} \mathrm{Tr}[S_k^\dagger \rho^{\otimes k}] &= \sum_{\boldsymbol{x}} \prod_{j=1}^{k} C_{x_j x_{\pi^{-1}(j)}} = \sum_{\boldsymbol{x}} \prod_{j=1}^{k} C^*_{x_{\pi^{-1}(j)} x_j} \\ &= \sum_{\boldsymbol{x}} \prod_{j=1}^{k} C^*_{x_j x_{\pi(j)}} = \mathrm{Tr}[(\rho^*)^k] = \mathrm{Tr}[\rho^k]. \end{aligned} \tag{B.8}$$

□

## B.2 Proof of necessary and sufficient condition for existence of mitigation protocol

The theorem is established through the vectorization of observables and their effective rank. To start with, we present the definitions of these two concepts.

**Definition B.1.** (Vectorization [38]) *The vectorization of a matrix* $M = \sum_{ij} M_{ij} |i\rangle\langle j|$ *is defined as*

$$|M\rangle\rangle := \sum_{i,j} M_{ij} |j\rangle \otimes |i\rangle . \tag{B.9}$$

**Definition B.2.** (Hermitian effective rank) *For a $2^n \times 2^n$ Hermitian matrix $O$ in a bipartite system $AB$, the effective rank $\mathcal{R}(O)$ is defined as*

$$\mathcal{R}(O) \coloneqq \mathrm{rank}(|O\rangle\langle O|_A), \tag{B.10}$$

*where $|O\rangle\langle O|_A = \mathrm{Tr}_B[|O\rangle\langle O|_{AB}]$.*

**Definition B.3.** (Matrix of channel [138]) *A quantum channel $\mathcal{N}$ can be represented in the matrix form $M_{\mathcal{N}}$ by its Kraus operators $\{E_k\}$. Given $\mathcal{N}(|i\rangle\langle j|) = \sum_k E_k |i\rangle\langle j| E_k^\dagger$, we can represent it in vector form, which is $vec(\mathcal{N}(|i\rangle\langle j|)) = \sum_k \bar{E}_k |j\rangle \otimes E_k |i\rangle = (\sum_k \bar{E}_j \otimes E_k) |j\rangle \otimes |i\rangle$, where $\bar{E}_k$ is the complex conjugate of $E_k$. Here we can define*

$$M_{\mathcal{N}} \coloneqq \sum_k \bar{E}_k \otimes E_k \tag{B.11}$$

*as the matrix representation of channel $\mathcal{N}$. The rank of such matrix $M_{\mathcal{N}}$ is called the channel rank of $\mathcal{N}$.*

In addition, here are a few results that will be used in the following proof of Theorem 4.1.

**Lemma B.2.** [138] *Suppose $H$ is an observable and $\mathcal{M}$ is a Hermitian-preserving is a map. Then $\mathrm{Tr}[(\mathcal{M}(\rho)O)] = \mathrm{Tr}[\rho O]$ for any state $\rho$ if and only if $\mathcal{M}^\dagger(O) = O$.*

**Lemma B.3.** *Suppose $H$ is an $nk$-qubit Hermitian matrix with $k$ subsystems, such that $\mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]$ for any $n$-qubit state $\rho$. Then the effective rank of $H$ for arbitrary subsystem $A$ is full.*

*Proof.* The property $\mathrm{Tr}[XY] = \langle X|Y\rangle$ implies

$$\left\langle \rho^{\otimes k} \middle| H \right\rangle = \mathrm{Tr}[\rho^k]. \tag{B.12}$$

Let $A$ be an $n$-qubit subsystem and $B$ be the rest of the subsystems. With respect to bipartite system $AB$, applying Schmidt decomposition on $|H\rangle$ gives

$$\left\langle \rho^{\otimes k} \middle| H \right\rangle = \sum_j c_j \langle \rho | u_j \rangle_A \cdot \left\langle \rho^{\otimes (k-1)} \middle| v_j \right\rangle_B, \text{ where } |H\rangle = \sum_j c_j |u_j\rangle_A |v_j\rangle_B . \tag{B.13}$$

Note that $|H\rangle\langle H|_A = \sum_j |c_j|^2 |u_j\rangle\langle u_j|$. Then the assumption $|\rho\rangle \in \ker(|H\rangle\langle H|_A)$ would imply

$$\sum_j |c_j|^2 \langle \rho | u_j \rangle \langle u_j | \rho \rangle = 0 \implies \langle \rho | u_j \rangle = 0 \ \forall j \implies \left\langle \rho^{\otimes k} \middle| H \right\rangle = 0 \implies \mathrm{Tr}[\rho^k] = 0, \tag{B.14}$$

which forms a contradiction as $\mathrm{Tr}[\rho^k] > 0$ for finite $k$. That is, the vectorization of all $n$-qubit quantum states are in the column space of $|H\rangle\langle H|_A$, and hence $|H\rangle\langle H|_A$ is of full rank. □

We are now ready to present the main theorem.

**Theorem 4.1** (Necessary and sufficient condition for existence of protocol) *Given a noisy channel $\mathcal{N}$, there exists a quantum protocol to extract the $k$-th moment $\mathrm{Tr}[\rho^k]$ for any state $\rho$ if and only if the noisy channel $\mathcal{N}$ is invertible.*

*Proof.* From Lemma B.1, there exists an observable $H$ such that $\mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]$. If the noisy channel $\mathcal{N}$ is invertible, then there exists the inverse operation of noisy channel $\mathcal{N}^{-1}$, such that we can apply the inverse operation on quantum system, and perform measurements with respect to moment observable $H$. Thus, the high-order moment can be retrieved, which is

$$\mathrm{Tr}[H \cdot (\mathcal{N}^{-1})^{\otimes k} \circ \mathcal{N}^{\otimes k}(\rho^{\otimes k})] = \mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]. \tag{B.15}$$

The proof for the "if" part is completed.

For the "only if" part, let's assume there exists a quantum protocol to extract $k$-th order information $\mathrm{Tr}[\rho^k]$ from noisy state $\mathcal{N}(\rho)$, with non-invertible noise channel $\mathcal{N}$. Then, from the definition of quantum protocol, the existence of quantum protocol to extract information $\mathrm{Tr}[\rho^k]$ refers to the existence of an HP map $\mathcal{D}$ such that the following equation holds

$$\mathrm{Tr}[H \cdot \mathcal{D} \circ \mathcal{N}^{\otimes k}(\rho^{\otimes k})] = \mathrm{Tr}[H\rho^{\otimes k}] = \mathrm{Tr}[\rho^k]. \tag{B.16}$$

From Lemma B.2, we convert the problem into that there exists an HP map $\mathcal{D}$ such that the following equation holds

$$(\mathcal{N}^{\otimes k})^\dagger \circ \mathcal{D}^\dagger(H) = H. \tag{B.17}$$

Since $\mathcal{D}$ is an HP map, then the adjoint map $\mathcal{D}^\dagger$ is an HP map as well, thus we have $D^\dagger(H) = Q$, where $Q$ is hermitian. Now we have

$$(\mathcal{N}^{\otimes k})^\dagger(Q) = H \tag{B.18}$$

Next we vectorize the Hermitian matrix $H$ and $Q$ into $|H\rangle$ and $|Q\rangle$, the corresponding channel $(\mathcal{N}^{\otimes k})^\dagger$ in vectoriztion representation is $M_{\mathcal{N}^\dagger}^{\otimes k}$, where $M_{\mathcal{N}^\dagger} = \sum_k \bar{E_k}^\dagger \otimes E_k^\dagger$, $\{E_k\}$ is the Kraus representation of channel $\mathcal{N}$ operator, and $\bar{E_k}$ is the complex conjugate of $E_k$. Thus the equation becomes

$$M_{\mathcal{N}^\dagger}^{\otimes k} |Q\rangle = |H\rangle\,, \tag{B.19}$$

which is equivalent to

$$M_{\mathcal{N}^\dagger}^{\otimes k} |Q\rangle\langle Q|\, M_{\mathcal{N}^\dagger}^{\otimes k\dagger} = |H\rangle\langle H|\,. \tag{B.20}$$

Then take partial trace on both side,

$$\mathrm{Tr}_B[M_{\mathcal{N}^\dagger}^{\otimes k} |Q\rangle\langle Q|\, M_{\mathcal{N}^\dagger}^{\otimes k\dagger}] = \mathrm{Tr}_B[|H\rangle\langle H|], \tag{B.21}$$

where $B$ is an arbitrary subsystem that contains $k-1$ copies of $\rho$. We can simplify it into

$$M_{\mathcal{N}^\dagger}\mathrm{Tr}_B[I\otimes M_{\mathcal{N}^\dagger}^{\otimes(k-1)}\,|Q\rangle\langle Q|\,I\otimes M_{\mathcal{N}^\dagger}^{\otimes(k-1)\dagger}]M_{\mathcal{N}^\dagger}^\dagger=\mathrm{Tr}_B[|H\rangle\langle H|]. \tag{B.22}$$

For further step, let's take $P=\mathrm{Tr}_B[I\otimes M_{\mathcal{N}^\dagger}^{\otimes(k-1)}\,|Q\rangle\langle Q|\,I\otimes M_{\mathcal{N}^\dagger}^{\otimes(k-1)\dagger}]$, and $T=\mathrm{Tr}_B[|H\rangle\langle H|]$, thus

$$M_{\mathcal{N}^\dagger}PM_{\mathcal{N}^\dagger}^\dagger=T. \tag{B.23}$$

From Lemma B.3, the matrix $T$ is full rank, thus the rank of $M_{\mathcal{N}^\dagger}PM_{\mathcal{N}^\dagger}^\dagger$ is full as well. Due to the fact that $\mathrm{rank}(XY)\leq\min(\mathrm{rank}(X),\mathrm{rank}(Y))$, the rank on the left hand side is

$$\mathrm{rank}(M_{\mathcal{N}^\dagger}PM_{\mathcal{N}^\dagger}^\dagger)\leq\min\{\mathrm{rank}(M_{\mathcal{N}^\dagger}),\mathrm{rank}(P),\mathrm{rank}(M_{\mathcal{N}^\dagger}^\dagger)\}, \tag{B.24}$$

implying that $M_{\mathcal{N}^\dagger}$, $P$ and $M_{\mathcal{N}^\dagger}^\dagger$ have full ranks. Then we have

$$M_{\mathcal{N}^\dagger}=\sum_k \bar{E_k}^\dagger\otimes E_k^\dagger=\sum_k(\bar{E_k}\otimes E_k)^\dagger=M_{\mathcal{N}}^\dagger, \tag{B.25}$$

since conjugate transpose won't affect the rank of a matrix, which means $M_{\mathcal{N}}$ has full rank. This concludes that $\mathcal{N}$ is invertible, and eventually, a contradiction is formed. Thus, no quantum protocol exists to extract high-order moment $\mathrm{Tr}[\rho^k]$. □

## B.3 Proof of Proposition 4.3

### B.3.1 Depolarizing noise retriever for the 2nd moment when $\rho$ is $n$-qubit state

**Proposition B.1.** *Given noised states $\mathcal{N}_{\mathrm{DE}}^\epsilon(\rho)^{\otimes 2}$, and error tolerance $\delta$ where $\rho$ is an $n$-qubit state, the second order moment $\mathrm{Tr}[\rho^2]$ can be estimated by $f\mathrm{Tr}[H\mathcal{C}\circ\mathcal{N}_{\mathrm{DE}}^\epsilon(\rho)^{\otimes 2}]-t$ where $f=\frac{1}{(1-\epsilon)^2}$ and $t=\frac{1-(1-\epsilon)^2}{2^n(1-\epsilon)^2}$. The term $\mathrm{Tr}[H\mathcal{C}\circ\mathcal{N}_{\mathrm{DE}}^\epsilon(\rho)^{\otimes 2}]$ can be estimated by implementing quantum retriever $\mathcal{C}$ with sample complexity $\mathcal{O}(1/\delta^2(1-\epsilon)^4)$. The Choi matrix of such a quantum retriever is,*

$$J_{\mathcal{C}}=\frac{1}{4^n}I^{\otimes n}I^{\otimes n}+\frac{1}{4^n(4^n-1)}(\sum_{i\neq 0}P_i\otimes P_i)\otimes(\sum_{j\neq 0}P_j\otimes P_j).$$

*Proof.* For an arbitrary state $\rho\otimes\rho$, after applying the noise $\mathcal{N}$ onto it, we denote

the state as $\rho' = \mathcal{N}(\rho \otimes \rho)$, thus

$$\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho) = \mathcal{D}(\rho') = \mathrm{Tr}_A[(\rho'^T \otimes I^{\otimes n}) J_{\mathcal{C}}] \tag{B.26}$$

$$= \mathrm{Tr}_A\Big[(\rho'^T \otimes I^{\otimes n})(\frac{1}{4^n} I^{\otimes n} I^{\otimes n} + \frac{1}{4^n(4^n-1)}(\sum_{i\neq 0} P_i \otimes P_i) \otimes (\sum_{j\neq 0} P_j \otimes P_j))\Big] \tag{B.27}$$

$$= \frac{1}{4^n}\mathrm{Tr}[\rho'^T] I^{\otimes n} + \frac{1}{4^n(4^n-1)}\mathrm{Tr}[\rho'^T(\sum_{i\neq 0} P_i \otimes P_i)](\sum_{j\neq 0} P_j \otimes P_j) \tag{B.28}$$

$$= \frac{1}{4^n} I^{\otimes n} + \frac{1}{4^n(4^n-1)}\mathrm{Tr}[\rho'^T(\sum_{i\neq 0} P_i \otimes P_i)](\sum_{j\neq 0} P_j \otimes P_j) \tag{B.29}$$

Note that the above equations utilized the face that transpose operation is trace-preserving, i.e., $\mathrm{Tr}[\rho^T] = \mathrm{Tr}[\rho] = 1$. Since the matrix $\sum_{i\neq 0} P_i$ is symmetric, thus we have

$$\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho) = \frac{1}{4^n} I^{\otimes n} + \frac{1}{4^n(4^n-1)}\mathrm{Tr}[\rho'^T(\sum_{i\neq 0} P_i \otimes P_i)^T](\sum_{j\neq 0} P_j \otimes P_j) \tag{B.30}$$

$$= \frac{1}{4^n} I^{\otimes n} + \frac{1}{4^n(4^n-1)}\mathrm{Tr}[\rho'(\sum_{i\neq 0} P_i \otimes P_i)](\sum_{j\neq 0} P_j \otimes P_j) \tag{B.31}$$

The trace term can be calculated by substituting Eq. (4.24), we have

$$\mathrm{Tr}[\rho'(\sum_{i\neq 0} P_i \otimes P_i)] = \mathrm{Tr}[((1-\epsilon)^2 \rho \otimes \rho + \frac{\epsilon(1-\epsilon)}{2^n} I^{\otimes n} \otimes \rho + \frac{(1-\epsilon)\epsilon}{2^n} \rho \otimes I^{\otimes n} + \frac{\epsilon^2}{4^n}(I^{\otimes n} \otimes I^{\otimes n}))(\sum_{i\neq 0} P_i \otimes P_i)] \tag{B.32}$$

$$= (1-\epsilon)^2 \mathrm{Tr}[\rho \otimes \rho(\sum_{i\neq 0} P_i \otimes P_i) \otimes (\sum_{j\neq 0} P_j \otimes P_j)]. \tag{B.33}$$

In the second equation, all other terms are traceless, thus, only the first term survives. Replace the equation back to Eq. (B.31), then

$$\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho) = \frac{1}{4^n} I^{\otimes n} + \frac{(1-\epsilon)^2}{4^n(4^n-1)}\mathrm{Tr}[\rho \otimes \rho(\sum_{i\neq 0} P_i \otimes P_i)](\sum_{j\neq 0} P_j \otimes P_j). \tag{B.34}$$

The information $\mathrm{Tr}[\rho^2]$ is estimated from $\mathrm{Tr}[H(\rho \otimes \rho)]$, where $H = \frac{1}{2^n}\sum_{i\neq 0} P_i \otimes P_i$ is cyclic permutation operator. It is easy to check that

$$\mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] = \frac{1}{4^n}\mathrm{Tr}[H * I^{\otimes n}] + \frac{(1-\epsilon)^2}{4^n(4^n-1)}\mathrm{Tr}[\rho \otimes \rho(\sum_{i\neq 0} P_i \otimes P_i)]\mathrm{Tr}[H * (\sum_{j\neq 0} P_j \otimes P_j)]. \tag{B.35}$$

We can quickly get $\mathrm{Tr}[H] = 2^n$ and $\mathrm{Tr}[H * (\sum_{j\neq 0} P_j \otimes P_j)] = 2^n(4^n - 1)$. Then

$$f\mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)] = \frac{1}{(1-\epsilon)^2}\Big[\frac{1}{2^n} + \frac{(1-\epsilon)^2}{2^n}\mathrm{Tr}[\rho \otimes \rho(\sum_{i\neq 0} P_i \otimes P_i)]\Big] \tag{B.36}$$

$$= \frac{1}{2^n(1-\epsilon)^2} + \frac{1}{2^n}\mathrm{Tr}[\rho \otimes \rho(\sum_{i\neq 0} P_i \otimes P_i)] \tag{B.37}$$

$$= \frac{1}{2^n(1-\epsilon)^2} - \frac{1}{2^n} + \frac{1}{2^n} + \frac{1}{2^n}\mathrm{Tr}[\rho \otimes \rho(\sum_{i\neq 0} P_i \otimes P_i)] \tag{B.38}$$

$$= \frac{1}{2^n(1-\epsilon)^2} - \frac{(1-\epsilon)^2}{2^n(1-\epsilon)^2} + \frac{1}{2^n}\mathrm{Tr}[\rho \otimes \rho(\sum_{i} P_i \otimes P_i)] \tag{B.39}$$

$$= \frac{1-(1-\epsilon)^2}{2^n(1-\epsilon)^2} + \mathrm{Tr}[\rho \otimes \rho H] \tag{B.40}$$

$$= t + \mathrm{Tr}[\rho^2]. \tag{B.41}$$

where $t = \frac{1-(1-\epsilon)^2}{2^n(1-\epsilon)^2}$ is the shift coefficient. The desired high-order moment $\mathrm{Tr}[\rho^2]$ is directly obtained by estimated value $\mathrm{Tr}[H\mathcal{C} \circ \mathcal{N}(\rho \otimes \rho)]$ substrate a constant $t$. □

### B.3.2 Depolarizing noise retriever for the $k$-th moment when $\rho$ is a qudit state

Appendix B.3.1 demonstrates the efficacy of the observable shift method in extracting the second moment of an $n$-qubit quantum state $\rho$ subjected to a depolarizing channel $D_\varepsilon$. This section extends the method's applicability, showing it can also determine the $k$-th moment of a $\rho$ given $k$ copies of $D_\varepsilon(\rho)$ for $100 \geq k > 2$. Our extension is based on the fact that the $k$-th moment of $\rho$ is intrinsically a linear combination of $\mathrm{Tr}[\rho^2], \ldots, \mathrm{Tr}[\rho^{k-1}]$ and $\mathrm{Tr}[D_\varepsilon(\rho)^k]$ with a constant, as evidenced by the following binomial expansion:

$$\mathrm{Tr}[D_\varepsilon(\rho)^k] = \mathrm{Tr}[((1-\varepsilon)\rho + \varepsilon I/d)^k] = \sum_{l=0}^{k}\binom{k}{l}\frac{(1-\varepsilon)^l\varepsilon^{k-l}}{d^{k-l}}\mathrm{Tr}[\rho^l], \tag{B.42}$$

where $I$ is the identity matrix of dimension $d = 2^n$. Then the recovery map can be recursively found by reducing $\mathrm{Tr}[\rho^l]$ to an expectation value of Hamiltonian $H_k$ under the depolarizing noise, and finally the observable shift method can be applied. Before presenting the details, we need an extra proposition to guarantee the existence of such a reduction, and the proof of which is deferred to Appendix B.3.3.

**Proposition B.2.** *Suppose* $100 \geq k > 2$. *Denote* $H_l = \frac{1}{2}(S_l + S_l^\dagger)$ *for the cyclic permutation operator* $S_l \in \mathbb{C}^{ld \times ld}$ *permuting* $l$ *subsystems. Then there exists CP maps* $\mathcal{T}_k$ *and* $\widetilde{\mathcal{T}}_k$ *such that*

$$\mathcal{T}_k(H_k) = H_{k-1} \otimes I/d, \quad \widetilde{\mathcal{T}}_k(H_k) = -H_{k-1} \otimes I/d. \tag{B.43}$$

To see how this proposition works, for any $k > l \geq 2$, one can construct a CP map

$$\mathcal{R}_l = (\mathcal{T}_{l+1} \otimes \mathrm{id}_{k-l-1}) \circ \ldots \circ (\mathcal{T}_{k-2} \otimes \mathrm{id}_2) \circ (\mathcal{T}_{k-1} \otimes \mathrm{id}) \circ \widetilde{\mathcal{T}}_k. \tag{B.44}$$

By Proposition B.2,

$$\begin{aligned}
\mathcal{R}_l(H_k) &= (\mathcal{T}_{l+1} \otimes \mathrm{id}_{k-l-1}) \circ \ldots \circ (\mathcal{T}_{k-2} \otimes \mathrm{id}_2) \circ (\mathcal{T}_{k-1} \otimes \mathrm{id}) \circ \widetilde{\mathcal{T}}_k(H_k) && \text{(B.45)}\\
&= -(\mathcal{T}_{l+1} \otimes \mathrm{id}_{k-l-1}) \circ \ldots \circ (\mathcal{T}_{k-2} \otimes \mathrm{id}_2)\left(\mathcal{T}_{k-1}\left(H_{k-1}\right) \otimes I/d\right) && \text{(B.46)}\\
&= -\mathcal{T}_{l+1}\left(H_{k-l+1}\right) \otimes I_{k-l-1}/d^{k-l-1} = -H_l \otimes I_{k-l}/d^{k-l}, && \text{(B.47)}
\end{aligned}$$

where the negative sign here is to preserve completely-positiveness of $\mathcal{C}_k$ in the later proposition. Then for any state $\sigma$, the computation of $\mathrm{Tr}[\sigma^l]$ is understood as an expectation value of $H_k$:

$$\mathrm{Tr}[\sigma^l] = \mathrm{Tr}[H_l \sigma^{\otimes l}] \propto -\mathrm{Tr}[\mathcal{R}_l(H_k)\sigma^{\otimes k}] = -\mathrm{Tr}[H_k \cdot \mathcal{R}_l^\dagger\left(\sigma^{\otimes k}\right)]. \tag{B.48}$$

We are ready to present the main conjecture in this section.

**Proof of Proposition 4.3** The statement is proved by induction.
(**Base case**: $k = 2$) The base case follows by Proposition B.1.
(**Inductive hypothesis**) For all $k > l \geq 2$, there exists a CP map $C_l$, constants $f_l$ and $t_l$ such that

$$f_l \mathrm{Tr}[H_l \cdot \mathcal{C}_l \circ D_\varepsilon^{\otimes l}\left(\rho^{\otimes l}\right)] - t_l = \mathrm{Tr}[\rho^l]. \tag{B.49}$$

(**Inductive step**) Proposition B.1 and Equation (B.42) implies

$$\begin{aligned}
\mathrm{Tr}[H_k D_\varepsilon(\rho)^{\otimes k}] = \mathrm{Tr}[D_\varepsilon(\rho)^k] =& (1-\varepsilon)^k \mathrm{Tr}[\rho^k] + \frac{\varepsilon^k}{d^k} + \frac{k(1-\varepsilon)\varepsilon^{k-1}}{d^{k-1}} \\
&+ \sum_{l=2}^{k-1} \binom{k}{l} \frac{(1-\varepsilon)^l \varepsilon^{k-l}}{d^{k-l}} \mathrm{Tr}[\rho^l]. \qquad \text{(B.50)}
\end{aligned}$$

From the induction hypothesis, we have

$$\begin{aligned}
&\mathrm{Tr}[H_k \cdot \mathrm{id}_k\left(D_\varepsilon(\rho)^{\otimes k}\right)] = \\
&(1-\varepsilon)^k\left(\mathrm{Tr}[\rho^k] + t_k\right) + \sum_{l=2}^{k-1} \binom{k}{l} (1-\varepsilon)^l \varepsilon^{k-l} f_l \mathrm{Tr}[H_l/d^{k-l} \cdot \mathcal{C}_l\left(D_\varepsilon(\rho)^{\otimes l}\right)],
\end{aligned} \tag{B.51}$$

where the constant $t_k$ is constructed as

$$t_k = \frac{1}{(1-\varepsilon)^k}\left[\frac{\varepsilon^k}{d^k} + \frac{k(1-\varepsilon)\varepsilon^{k-1}}{d^{k-1}} - \sum_{l=2}^{k-1}\binom{k}{l}\frac{(1-\varepsilon)^l\varepsilon^{k-l}}{d^{k-l}}t_l\right]. \tag{B.52}$$

Further, from the construction of $\mathcal{R}_l$ in Equation (B.44), the trace quantity on the last equality can be expanded as

$$\begin{aligned}
\mathrm{Tr}[H_l/d^{k-l}\cdot\mathcal{C}_l\left(D_\varepsilon(\rho)^{\otimes l}\right)] &= \mathrm{Tr}[\left(H_l\otimes I_{k-l}/d^{k-l}\right)\cdot(\mathcal{C}_l\otimes\mathrm{id}_{k-l})\left(D_\varepsilon(\rho)^{\otimes k}\right)] && \text{(B.53)}\\
&= -\mathrm{Tr}[\mathcal{R}_l(H_k)\cdot(\mathcal{C}_l\otimes\mathrm{id}_{k-l})\left(D_\varepsilon(\rho)^{\otimes k}\right)] && \text{(B.54)}\\
&= -\mathrm{Tr}[H_k\cdot\mathcal{R}_l^\dagger\circ(\mathcal{C}_l\otimes\mathrm{id}_{k-l})\left(D_\varepsilon(\rho)^{\otimes k}\right)]. && \text{(B.55)}
\end{aligned}$$

By moving the term around, we eventually get

$$\begin{aligned}
\mathrm{Tr}[\rho^k] &= \frac{1}{(1-\varepsilon)^k}\{\mathrm{Tr}[H_k\cdot\mathrm{id}_k\left(D_\varepsilon(\rho)^{\otimes k}\right)] && \text{(B.56)}\\
&\quad+\sum_{l=2}^{k-1}\binom{k}{l}(1-\varepsilon)^l\varepsilon^{k-l}f_l\mathrm{Tr}[H_k\cdot\mathcal{R}_l^\dagger\circ(\mathcal{C}_l\otimes\mathrm{id}_{k-l})\left(D_\varepsilon(\rho)^{\otimes k}\right)]\} - t_k && \text{(B.57)}\\
&= \frac{1}{(1-\varepsilon)^k}\mathrm{Tr}[H_k\cdot\left(\mathrm{id}_k+\sum_{l=2}^{k-1}\binom{k}{l}(1-\varepsilon)^l\varepsilon^{k-l}f_l\mathcal{R}_l^\dagger\circ(\mathcal{C}_l\otimes\mathrm{id}_{k-l})\right)\left(D_\varepsilon(\rho)^{\otimes k}\right)]\\
&\quad - t_k && \text{(B.58)}\\
&= f_k\mathrm{Tr}[H_k\cdot\mathcal{C}_k\left(D_\varepsilon(\rho)^{\otimes k}\right)] - t_k, && \text{(B.59)}
\end{aligned}$$

where $f_k = 1/(1-\varepsilon)^k$ and

$$\mathcal{C}_k = \mathrm{id}_k + \sum_{l=2}^{k-1}\binom{k}{l}(1-\varepsilon)^l\varepsilon^{k-l}f_l\mathcal{R}_l^\dagger\circ(\mathcal{C}_l\otimes\mathrm{id}_{k-l}). \tag{B.60}$$

The CP property of $\mathcal{C}_k$ follows by the positiveness of $f_l$ and the CP properties of $\mathcal{C}_l, \mathcal{R}_l$. ■

### B.3.3 Proof for Proposition B.2

This section presents the necessary lemma for proving the existence of physical protocol to retrieve generalized depolarizing noises. Particularly, we show that there exists CP maps that convert the permutation unitary $S_k$ to $\pm S_{k-1}$ for $k > 2$, as stated in Proposition B.2. Denote $\omega_k = \exp(2\pi i/k)$ as the $k$-th unity root. Let $\mathrm{Cr}_d^k$ be the minimal set of length-$k$ combinations that generates $\{0,\ldots,d-1\}^{\times k}$ under cyclic permutations, i.e.,

$$\{0,\ldots,d-1\}^{\times k} = \{x_{\pi^l(1)}\cdots x_{\pi^l(k)} \mid x_1\cdots x_k \in \mathrm{Cr}_d^k, l\in\mathbb{N}\}. \tag{B.61}$$

For example, $\mathrm{Cr}_2^3$ can be non-uniquely constructed as $\{000, 001, 011, 111\}$. Note that $|\mathrm{Cr}_d^k| = \frac{d^k-d}{k} + d$. One can then have the following result:

**Lemma B.4** (Spectral decomposition of $S_k$)**.** *The cyclic permutation matrix can be decomposed as*

$$S_k = \sum_{m=0}^{k-1} \omega_k^{-m} \sum_{\boldsymbol{x}\in \mathrm{Cr}_d^k} |\psi_{m,\boldsymbol{x}}\rangle\langle\psi_{m,\boldsymbol{x}}|\,, \tag{B.62}$$

*for* $\boldsymbol{x} = x_1 \cdots x_k$ *and eigenstates defined as*

$$|\psi_{m,\boldsymbol{x}}\rangle = \frac{1}{\sqrt{k}} \sum_{l=0}^{k-1} \omega_k^{ml} \bigotimes_{j=1}^{k} \left|x_{\pi^l(j)}\right\rangle. \tag{B.63}$$

*Proof.* We first prove that $|\psi_{m,\boldsymbol{x}}\rangle$ is an eigenstate of $S_k$ with eigenvalue $\omega_k^m$, then show that the state set forms a basis of $\mathcal{H}_{d^k}$. Applying $S_k$ on $|\psi_{m,\boldsymbol{x}}\rangle$ gives

$$S_k\,|\psi_{m,\boldsymbol{x}}\rangle = \left[\sum_{\boldsymbol{y}} \bigotimes_j |y_{\pi(j)}\rangle\langle y_j|\right] \cdot \left[\sum_l \omega_k^{ml} \bigotimes_j \left|x_{\pi^l(j)}\right\rangle\right] \tag{B.64}$$

$$= \sum_{\boldsymbol{y}} \sum_l \omega_k^{ml} \bigotimes_j \delta_{y_j x_{\pi^l(j)}} \left|y_{\pi(j)}\right\rangle \tag{B.65}$$

$$= \sum_{l=0}^{k-1} \omega_k^{ml} \bigotimes_j \left|x_{\pi^{l+1}(j)}\right\rangle = \sum_{l=1}^{k} \omega_k^{m(l-1)} \bigotimes_j \left|x_{\pi^l(j)}\right\rangle \tag{B.66}$$

$$= \omega_k^{-m}\,|\psi_{m,\boldsymbol{x}}\rangle\,. \tag{B.67}$$

Note that for fixed point $\boldsymbol{x}$ such that $x_j = x_{\pi(j)}$ for all $j$, $|\psi_{m,\boldsymbol{x}}\rangle \neq 0$ if and only if $m = 0$. Then the total number of such eigenstates are

$$d + k \cdot \Big(|\mathrm{Cr}_d^k| - d\Big) = d^k. \tag{B.68}$$

As for the orthogonality, for $0 \le m, m' < k$, $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{Cr}_d^k$,

$$\langle\psi_{m,\boldsymbol{x}}|\psi_{m',\boldsymbol{x}'}\rangle = \frac{1}{k} \sum_{l,l'} \omega_k^{-ml} \omega_k^{m'l'} \prod_{j=1}^{k} \left\langle x_{\pi^l(j)} \middle| x'_{\pi^{l'}(j)} \right\rangle \tag{B.69}$$

$$= \frac{1}{k} \sum_{l,l'} \omega_k^{-ml} \omega_k^{m'l'} \left(\prod_{j=1}^{k} \delta_{x_j x'_{\pi^{(l'-l)}(j)}}\right) \tag{B.70}$$

$$= \left(\frac{1}{k} \sum_l \omega_k^{l(m'-m)}\right) \delta_{\boldsymbol{x}\boldsymbol{x}'} = \delta_{mm'}\delta_{\boldsymbol{x}\boldsymbol{x}'}, \tag{B.71}$$

where $\boldsymbol{x} \neq \boldsymbol{x}'$ if and only if there does not exists $l$ such that $x_j \neq x'_{\pi^l(j)}$ for all $j$. □

**Proof of Proposition B.2** One can observe that for $k > 2$, there exists non-negative matrices $Q, \widetilde{Q} \in \mathbb{R}_+^{(k-1)\times k}$ such that for all $0 \le l < k-2$,

$$\sum_{m=0}^{k-1} Q_{lm}\omega_k^m = \omega_{k-1}^l, \quad \sum_{m=0}^{k-1} \widetilde{Q}_{lm}\omega_k^m = -\omega_{k-1}^l. \tag{B.72}$$

Such $Q, \widetilde{Q}$ can construct the desired transformations in Proposition B.2. Using the package CVX [181, 182], Equation (B.72) can be numerically verified for $100 \ge k$ [1]. For a general $k > 2$, $Q$ and $\widetilde{Q}$ can be explicitly constructed as

$$Q_{\text{odd } k} = \begin{pmatrix} 1 & & & & & & & & \\ & x_1 & y_1 & & & & & & \\ & & \ddots & \ddots & & & & & \\ & & & x_m & y_m & & & & \\ & & & & y_m & x_m & & & \\ & & & & & \ddots & \ddots & & \\ & & & & & & y_1 & x_1 \end{pmatrix}, \tag{B.73}$$

$$Q_{\text{even } k} = \begin{pmatrix} 1 & & & & & & & & \\ & x_1 & y_1 & & & & & & \\ & & \ddots & \ddots & & & & & \\ & & & x_{m-1} & y_{m-1} & & & & \\ & & & & x_m & x_m & & & \\ & & & & & y_{m-1} & x_{m-1} & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & y_1 & x_1 \end{pmatrix}, \tag{B.74}$$

$$\widetilde{Q}_{\text{odd } k} = (X \otimes I_{(k-1)/2}) \cdot Q_{\text{odd } k}, \quad \widetilde{Q}_{\text{even } k} = Q_{\text{even } k} \cdot (X \otimes I_{k/2}) \tag{B.75}$$

dependent on the parity of $k$, where $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is the Pauli-$X$ gate, $m = \lfloor k/2 \rfloor$ and constants $x_l, y_l$ are given as

$$x_l = \csc\frac{2\pi}{k} \sin\frac{2(k-1-l)\pi}{k(k-1)}, \quad y_l = \csc\frac{2\pi}{k} \sin\frac{2l\pi}{k(k-1)}. \tag{B.76}$$

The fact that $k > 2$ ensures the angles exist in csc are no greater than $\frac{2\pi}{3}$, and those exist in sin are no greater than $\frac{\pi}{2}$, and hence these constants are always non-negative.

Now we use $Q, \widetilde{Q}$ to construct the maps. To be specific, denote the eigenspace with respect to $\omega_k^m$ as $\Pi_k^{(m)}$, i.e., $\Pi_k^{(m)} = \sum_{\boldsymbol{x}\in \mathrm{Cr}_d^k} |\psi_{-m,\boldsymbol{x}}\rangle\langle\psi_{-m,\boldsymbol{x}}|_k$. Consider the map $\mathcal{T}_k, \widetilde{\mathcal{T}}_k$ satisfying

$$\mathcal{T}_k\left(\Pi_k^{(m)}\right) = \sum_l Q_{lm}\Pi_{k-1}^{(l)} \otimes I/d, \quad \widetilde{\mathcal{T}}_k\left(\Pi_k^{(m)}\right) = \sum_l \widetilde{Q}_{lm}\Pi_{k-1}^{(l)} \otimes I/d \tag{B.77}$$

[1] For the data and verification files, see the GitHub repository https://github.com/Dragon-John/high-moment-info

Then it follows that

$$\mathcal{T}_k\left(S_k\right) = \sum_{m=0}^{k-1} \omega_k^m \mathcal{T}_k(\Pi_k^{(m)}) = \sum_{m=0}^{k-1} \omega_k^m \sum_{l=0}^{k-2} Q_{lm} \Pi_{k-1}^{(l)} \otimes I/d \tag{B.78}$$

$$= \sum_l \left( \sum_m Q_{lm} \omega_k^m \right) \Pi_{k-1}^{(l)} \otimes I/d = \sum_l \omega_{k-1}^l \Pi_{k-1}^{(l)} = S_{k-1} \otimes I/d, \tag{B.79}$$

and analogously

$$\widetilde{\mathcal{T}}_k\left(S_k\right) = -S_{k-1} \otimes I/d. \tag{B.80}$$

Since $\{\Pi_k^{(m)}\}_{m=1}^k$ is an orthogonal set and $Q$ is non-negative, $\mathcal{T}_k$ can be completely positive and hence $\mathcal{T}_k\left(S_k^\dagger\right) = \mathcal{T}_k\left(S_k\right)^\dagger$. We conclude that

$$\mathcal{T}_k\left(H_k\right) = \frac{1}{2}\left[\mathcal{T}_k\left(S_k\right) + \mathcal{T}_k\left(S_k^\dagger\right)\right] = \frac{1}{2}\left[S_{k-1} + S_{k-1}^\dagger\right] \otimes I/d = H_{k-1} \otimes I/d. \tag{B.81}$$

Similar statement holds for $\widetilde{\mathcal{T}}_k$. ■

# Appendix C

# Random channel simulation with DIO

The target channel $\mathcal{N}$ used in the numerical experiment of channel simulation with DIO (Figure. 5.3) is a random 2-qubit rank-4 real channel, whose Choi-Jamiołkowski matrix $J_{\mathcal{N}}$ is shown below.

We can easily check that $\Delta \circ \mathcal{N} \neq \Delta \circ \mathcal{N} \circ \Delta$, implying the target channel $\mathcal{N}$ is not a resource nonactivating channel.

$$J_{\mathcal{N}} = \begin{pmatrix}
0.1425 & -0.0717 & 0.0663 & -0.0582 & 0.1640 & 0.0037 & -0.0010 & 0.0504 & 0.0095 & 0.0560 & 0.0765 & -0.1369 & -0.0963 & -0.1184 & 0.0484 & -0.0148 \\
-0.0717 & 0.2731 & -0.0653 & 0.0888 & -0.1178 & -0.0300 & 0.0830 & -0.2387 & 0.2253 & -0.1143 & 0.0338 & -0.0622 & 0.1673 & 0.2035 & -0.3014 & -0.1055 \\
0.0663 & -0.0653 & 0.2067 & 0.0887 & 0.1565 & -0.0341 & -0.0497 & 0.0280 & 0.1282 & -0.1035 & 0.0352 & 0.0825 & -0.1642 & -0.1078 & 0.0018 & 0.0301 \\
-0.0582 & 0.0888 & 0.0887 & 0.3776 & -0.0611 & -0.2899 & -0.2548 & -0.0843 & 0.1231 & -0.1234 & 0.1396 & 0.0695 & -0.1340 & -0.0933 & -0.1351 & -0.1089 \\
0.1640 & -0.1178 & 0.1565 & -0.0611 & 0.2334 & 0.0313 & 0.0094 & 0.0779 & 0.0515 & 0.0115 & 0.0580 & -0.0741 & -0.1535 & -0.1478 & 0.0720 & 0.0257 \\
0.0037 & -0.0300 & -0.0341 & -0.2899 & 0.0313 & 0.2686 & 0.2506 & 0.0248 & -0.0198 & 0.0210 & -0.1566 & 0.0291 & 0.1374 & 0.1432 & 0.0514 & 0.1116 \\
-0.0010 & 0.0830 & -0.0497 & -0.2548 & 0.0094 & 0.2506 & 0.2845 & -0.0765 & 0.0879 & -0.0219 & -0.1227 & -0.0273 & 0.1929 & 0.2113 & -0.0801 & 0.0578 \\
0.0504 & -0.2387 & 0.0280 & -0.0843 & 0.0779 & 0.0248 & -0.0765 & 0.2135 & -0.2269 & 0.1179 & -0.0351 & 0.0502 & -0.1302 & -0.1712 & 0.2746 & 0.0904 \\
0.0095 & 0.2253 & 0.1282 & 0.1231 & 0.0515 & -0.0198 & 0.0879 & -0.2269 & 0.3708 & -0.2111 & 0.0589 & -0.0127 & 0.0375 & 0.1278 & -0.3174 & -0.0736 \\
0.0560 & -0.1143 & -0.1035 & -0.1234 & 0.0115 & 0.0210 & -0.0219 & 0.1179 & -0.2111 & 0.1715 & 0.0095 & -0.1161 & -0.0078 & -0.0855 & 0.1671 & 0.0079 \\
0.0765 & 0.0338 & 0.0352 & 0.1396 & 0.0580 & -0.1566 & -0.1227 & -0.0351 & 0.0589 & 0.0095 & 0.1489 & -0.1334 & -0.0962 & -0.1113 & -0.0639 & -0.1010 \\
-0.1369 & -0.0622 & 0.0825 & 0.0695 & -0.0741 & 0.0291 & -0.0273 & 0.0502 & -0.0127 & -0.1161 & -0.1334 & 0.3088 & -0.0268 & 0.0376 & 0.0640 & 0.1119 \\
-0.0963 & 0.1673 & -0.1642 & -0.1340 & -0.1535 & 0.1374 & 0.1929 & -0.1302 & 0.0375 & -0.0078 & -0.0962 & -0.0268 & 0.2532 & 0.2514 & -0.1346 & -0.0071 \\
-0.1184 & 0.2035 & -0.1078 & -0.0933 & -0.1478 & 0.1432 & 0.2113 & -0.1712 & 0.1278 & -0.0855 & -0.1113 & 0.0376 & 0.2514 & 0.2868 & -0.1947 & 0.0025 \\
0.0484 & -0.3014 & 0.0018 & -0.1351 & 0.0720 & 0.0514 & -0.0801 & 0.2746 & -0.3174 & 0.1671 & -0.0639 & 0.0640 & -0.1346 & -0.1947 & 0.3599 & 0.1212 \\
-0.0148 & -0.1055 & 0.0301 & -0.1089 & 0.0257 & 0.1116 & 0.0578 & 0.0904 & -0.0736 & 0.0079 & -0.1010 & 0.1119 & -0.0071 & 0.0025 & 0.1212 & 0.1001
\end{pmatrix}$$

# Acknowledgments

First and foremost, I would like to thank my supervisor Keisuke Fujii for his patient guidance and supervision during my research. He is so kind to offer any help as long as I need. As a professor, he is busy all the time, but I can always get his feedback in a short time. I am so lucky to be supervised by Professor Fujii, and I am so honored to be his student.

A special thanks goes to Xin Wang who leads me to the charming quantum world. If I hadn't met him, I wouldn't have gone into academia. During my PhD period, he was so kind to accept my visiting request, and I really enjoyed the time in his group. Under his guidance, my research skills and badminton skills have significantly improved.

Most of my work has been collaborative, and I sincerely thank all my co-authors: Keisuke Fujii, Xin Wang, Xuanqiang Zhao, Zhan Yu, Zhixin Song, Youle Wang, Ranyiliu Chen, Mingrui Jing, Lei Zhang, Yu-Ao Chen, Kun Wang, Zihan Xia, and Kosuke Ito. In particular, I want to thank Xin Wang, and Giulio Chiribella for their hospitable host during my visits. I would also like to thank Kosuke Mitarai, Chenghong Zhu, Chengkai Zhu, Zhiping Liu, Yin Mo, Yingjian Li and Hideaki Hakoshima, Zanqiu Shen for their insightful discussions. I especially thank Xuanqiang Zhao for helping in many aspects. To me, he is a peer, a teacher, a buddy, a champion.

In the end, I would like to show my great gratitude to my beloved parents and my wife, who are always on my side and support me without any hesitation.

# List of Activities

## Publications included in thesis

1. **Zhao, B.**, & Fujii, K. (2024). *Variational quantum Hamiltonian engineering*. arXiv preprint arXiv:2406.08998.

2. **Zhao, B.**, Jing, M., Zhang, L., Zhao, X., Chen, Y. A., Wang, K., & Wang, X. (2024). *Retrieving Nonlinear Features from Noisy Quantum States*. PRX Quantum, 5(2), 020357.

3. **Zhao, B.**, Ito, K., & Fujii, K. (2024). *Probabilistic channel simulation using coherence*. arXiv preprint arXiv:2404.06775. (Accepted by PR Research)

## Publications not included in thesis

1. Zhao, X., **Zhao, B.**, Xia, Z., & Wang, X. (2023). *Information recoverability of noisy quantum states*. Quantum, 7, 978.

2. Zhao, X., Zhang, L., **Zhao, B.**, & Wang, X. (2023). *Power of quantum measurement in simulating unphysical operations*. arXiv preprint arXiv:2309.09963.

3. Yu, Z., Zhao, X., **Zhao, B.**, & Wang, X. (2023). *Optimal quantum dataset for learning a unitary transformation*. Physical Review Applied, 19(3), 034017.

## International conferences

1. "Retrieving non-linear features from noisy quantum states", Poster presentation in Quantum Information Processing (QIP 2024)

2. "Retrieving non-linear features from noisy quantum states", Short talk in Quantum Error, Sensing and Control: Principles, Applications and Engineering (Q-Escape 2024)

3. "Probabilistic channel simulation using coherence", Poster presentation in Asian Quantum Information Science Conference (AQIS 2024)

4. "Retrieving non-linear features from noisy quantum states", Poster presentation in Asian Quantum Information Science Conference (AQIS 2024)

# Bibliography


[1]P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", SIAM review **41**, 303–332 (1999).

[2]R. P. Feynman, "Simulating physics with computers", in *Feynman and computation* (cRc Press, 2018), pp. 133–153.

[3]J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning", Nature **549**, 195–202 (2017).

[4]S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis", Nature physics **10**, 631–633 (2014).

[5]R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems", Communications of the ACM **21**, 120–126 (1978).

[6]J. Buchmann, *Introduction to cryptography*, Vol. 335 (Springer, 2004).

[7]I. N. Levine, D. H. Busch, and H. Shull, *Quantum chemistry*, Vol. 6 (Pearson Prentice Hall Upper Saddle River, NJ, 2009).

[8]M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, et al., "Variational quantum algorithms", Nature Reviews Physics **3**, 625–644 (2021).

[9]F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., "Quantum supremacy using a programmable superconducting processor", Nature **574**, 505–510 (2019).

[10]C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, "Trapped-ion quantum computing: progress and challenges", Applied Physics Reviews **6** (2019).

[11]I. Cong, H. Levine, A. Keesling, D. Bluvstein, S.-T. Wang, and M. D. Lukin, "Hardware-efficient, fault-tolerant quantum computation with rydberg atoms", Physical Review X **12**, 021049 (2022).

[12]X. Zhang, H.-O. Li, G. Cao, M. Xiao, G.-C. Guo, and G.-P. Guo, "Semiconductor quantum computation", National Science Review **6**, 32–54 (2019).

[13]H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, et al., "Quantum computational advantage using photons", Science **370**, 1460–1463 (2020).

[14]V. Acosta and P. Hemmer, "Nitrogen-vacancy centers: physics and applications", MRS bulletin **38**, 127–130 (2013).

[15]A. Stern and N. H. Lindner, "Topological quantum computation—from basic concepts to first experiments", Science **339**, 1179–1184 (2013).

[16]P. W. Shor, "Fault-tolerant quantum computation", in Proceedings of 37th conference on foundations of computer science (IEEE, 1996), pp. 56–65.

[17]J. Preskill, "Fault-tolerant quantum computation", in *Introduction to quantum computation and information* (World Scientific, 1998), pp. 213–269.

[18]D. Gottesman, "Theory of fault-tolerant quantum computation", Physical Review A **57**, 127 (1998).

[19]R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, et al., "Superconducting quantum circuits at the surface code threshold for fault tolerance", Nature **508**, 500–503 (2014).

[20]E. Jeffrey, D. Sank, J. Mutus, T. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, et al., "Fast accurate state measurement with superconducting qubits", Physical review letters **112**, 190504 (2014).

[21]D. Gottesman, *Stabilizer codes and quantum error correction* (California Institute of Technology, 1997).

[22]M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).

[23]D. García-Martín and G. Sierra, "Five experimental tests on the 5-qubit ibm quantum computer", arXiv preprint arXiv:1712.05642 (2017).

[24]A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets", Nature **549**, 242–246 (2017).

[25]S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices", Nature Physics **14**, 595–600 (2018).

[26]F. Pan, K. Chen, and P. Zhang, "Solving the sampling problem of the sycamore quantum circuits", Physical Review Letters **129**, 090502 (2022).

[27]S. Aaronson and L. Chen, "Complexity-theoretic foundations of quantum supremacy experiments", arXiv preprint arXiv:1612.05903 (2016).

[28]M. J. Bremner, A. Montanaro, and D. J. Shepherd, "Average-case complexity versus approximate simulation of commuting quantum computations", Physical review letters **117**, 080501 (2016).

[29]C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, "Gaussian boson sampling", Physical review letters **119**, 170501 (2017).

[30]J. M. Arrazola and T. R. Bromley, "Using gaussian boson sampling to find dense subgraphs", Physical review letters **121**, 030503 (2018).

[31]J. Huh, G. G. Guerreschi, B. Peropadre, J. R. McClean, and A. Aspuru-Guzik, "Boson sampling for molecular vibronic spectra", Nature Photonics **9**, 615–620 (2015).

[32]L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, et al., "Quantum computational advantage with a programmable photonic processor", Nature **606**, 75–81 (2022).

[33]A. Morvan, B. Villalonga, X. Mi, S. Mandra, A. Bengtsson, P. Klimov, Z. Chen, S. Hong, C. Erickson, I. Drozdov, et al., "Phase transition in random circuit sampling", Nature **634**, 328–333 (2024).

[34]J. Preskill, "Quantum computing in the nisq era and beyond", Quantum **2**, 79 (2018).

[35]B. Zhao and K. Fujii, "Variational quantum hamiltonian engineering", arXiv preprint arXiv:2406.08998 (2024).

[36]B. Zhao, M. Jing, L. Zhang, X. Zhao, Y.-A. Chen, K. Wang, and X. Wang, "Retrieving nonlinear features from noisy quantum states", PRX Quantum **5**, 020357 (2024).

[37]B. Zhao, K. Ito, and K. Fujii, "Probabilistic channel simulation using coherence", arXiv preprint arXiv:2404.06775 (2024).

[38]S. Khatri and M. M. Wilde, "Principles of quantum communication theory: a modern approach", arXiv:2011.04672 (2020).

[39]M.-D. Choi, "Completely positive linear maps on complex matrices", Linear algebra and its applications **10**, 285–290 (1975).

[40]A. Jamiołkowski, "Linear transformations which preserve trace and positive semidefiniteness of operators", Reports on mathematical physics **3**, 275–278 (1972).

[41]E. Chitambar and G. Gour, "Quantum resource theories", Reviews of modern physics **91**, 025001 (2019).

[42]R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement", Reviews of modern physics **81**, 865 (2009).

[43]K. Audenaert, M. B. Plenio, and J. Eisert, "Entanglement cost under positive-partial-transpose-preserving operations", Physical review letters **90**, 027901 (2003).

[44]V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, "Quantifying entanglement", Physical Review Letters **78**, 2275 (1997).

[45]X. Wang and M. M. Wilde, "Cost of quantum entanglement simplified", Physical Review Letters **125**, 040502 (2020).

[46]T. Theurer, D. Egloff, L. Zhang, and M. B. Plenio, "Quantifying operations with an application to coherence", Physical review letters **122**, 190405 (2019).

[47]E. Chitambar and M.-H. Hsieh, "Relating the resource theories of entanglement and quantum coherence", Physical review letters **117**, 020402 (2016).

[48]K. B. Dana, M. G. Díaz, M. Mejatty, and A. Winter, "Resource theory of coherence: beyond states", Physical Review A **95**, 062327 (2017).

[49]M. Howard and E. Campbell, "Application of a resource theory for magic states to fault-tolerant quantum computing", Physical review letters **118**, 090501 (2017).

[50]J. R. Seddon and E. T. Campbell, "Quantifying magic for multi-qubit operations", Proceedings of the Royal Society A **475**, 20190251 (2019).

[51]J. R. Seddon, B. Regula, H. Pashayan, Y. Ouyang, and E. T. Campbell, "Quantifying quantum speedups: improved classical simulation from tighter magic monotones", PRX Quantum **2**, 010345 (2021).

[52]S. Bravyi and A. Kitaev, "Universal quantum computation with ideal clifford gates and noisy ancillas", Physical Review A **71**, 10.1103/physreva.71.022316 (2005).

[53]A. P. Lund, M. J. Bremner, and T. C. Ralph, "Quantum sampling problems, bosonsampling and quantum supremacy", npj Quantum Information **3**, 15 (2017).

[54]J. Jiang, K. Wang, and X. Wang, "Physical implementability of linear maps and its application in error mitigation", Quantum **5**, 600 (2021).

[55]D. Wecker, M. B. Hastings, and M. Troyer, "Progress towards practical quantum variational algorithms", Physical Review A **92**, 042303 (2015).

[56]J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms", New Journal of Physics **18**, 023023 (2016).

[57]N. C. Rubin, R. Babbush, and J. McClean, "Application of fermionic marginal constraints to hybrid quantum algorithms", New Journal of Physics **20**, 053020 (2018).

[58]D. A. Lidar and T. A. Brun, *Quantum error correction* (Cambridge university press, 2013).

[59]A. J. Landahl, J. T. Anderson, and P. R. Rice, "Fault-tolerant quantum computing with color codes", arXiv preprint arXiv:1108.5738 (2011).

[60]P. Panteleev and G. Kalachev, "Quantum ldpc codes with almost linear minimum distance", IEEE Transactions on Information Theory **68**, 213–229 (2021).

[61]R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, J. Atalaya, et al., "Quantum error correction below the surface code threshold", Nature (2024).

[62]K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits", Physical review letters **119**, 180509 (2017).

[63]B. Regula, R. Takagi, and M. Gu, "Operational applications of the diamond norm and related measures in quantifying the non-physicality of quantum maps", Quantum **5**, 522 (2021).

[64]M. Horodecki and J. Oppenheim, "(quantumness in the context of) resource theories", International Journal of Modern Physics B **27**, 1345019 (2013).

[65]P. J. Coles, E. M. Metodiev, and N. Lütkenhaus, "Numerical approach for unstructured quantum key distribution", Nature communications **7**, 11712 (2016).

[66]V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology", Nature photonics **5**, 222–229 (2011).

[67]F. Fröwis and W. Dür, "Stable macroscopic quantum superpositions", Physical review letters **106**, 110402 (2011).

[68]G. Tóth and I. Apellaniz, "Quantum metrology from a quantum information science perspective", Journal of Physics A: Mathematical and Theoretical **47**, 424006 (2014).

[69]M. Lostaglio, D. Jennings, and T. Rudolph, "Description of quantum coherence in thermodynamic processes requires constraints beyond free energy", Nature communications **6**, 6383 (2015).

[70]F. G. Brandao, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, "Resource theory of quantum states out of thermal equilibrium", Physical review letters **111**, 250404 (2013).

[71]G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Y. Halpern, "The resource theory of informational nonequilibrium in thermodynamics", Physics Reports **583**, 1–58 (2015).

[72]M. G. Díaz, K. Fang, X. Wang, M. Rosati, M. Skotiniotis, J. Calsamiglia, and A. Winter, "Using and reusing coherence to realize quantum processes", Quantum **2**, 100 (2018).

[73]J. Aberg, "Quantifying superposition", arXiv preprint quant-ph/0612146 (2006).

[74]T. Baumgratz, M. Cramer, and M. B. Plenio, "Quantifying coherence", Physical review letters **113**, 140401 (2014).

[75]A. Winter and D. Yang, "Operational resource theory of coherence", Physical review letters **116**, 120404 (2016).

[76]B. D. Jones, P. Skrzypczyk, and N. Linden, "The hadamard gate cannot be replaced by a resource state in universal quantum computation", arXiv preprint arXiv:2312.03515 (2023).

[77]A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, "A variational eigenvalue solver on a photonic quantum processor", Nature communications **5**, 4213 (2014).

[78]J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. De Jong, "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states", Physical Review A **95**, 042308 (2017).

[79]K. M. Nakanishi, K. Mitarai, and K. Fujii, "Subspace-search variational quantum eigensolver for excited states", Physical Review Research **1**, 033062 (2019).

[80]M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers", Physical Review A **101**, 032308 (2020).

[81]E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors", arXiv preprint arXiv:1802.06002 (2018).

[82]R. Chen, B. Zhao, and X. Wang, "Near-term efficient quantum algorithms for entanglement analysis", Physical Review Applied **20**, 024071 (2023).

[83]R. Chen, Z. Song, X. Zhao, and X. Wang, "Variational quantum algorithms for trace distance and fidelity estimation", Quantum Science and Technology **7**, 015019 (2021).

[84]X. Zhao, B. Zhao, Z. Wang, Z. Song, and X. Wang, "Practical distributed quantum information processing with loccnet", npj Quantum Information **7**, 159 (2021).

[85]J. Romero, J. P. Olson, and A. Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data", Quantum Science and Technology **2**, 045001 (2017).

[86]C. Cao and X. Wang, "Noise-assisted quantum autoencoder", Physical Review Applied **15**, 054012 (2021).

[87]E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm", arXiv preprint arXiv:1411.4028 (2014).

[88]M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo, et al., "Quantum approximate optimization of non-planar graph problems on a planar superconducting processor", Nature Physics **17**, 332–336 (2021).

[89]S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry", Reviews of Modern Physics **92**, 015003 (2020).

[90]G. A. Quantum, Collaborators*†, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, et al., "Hartree-fock on a superconducting qubit quantum computer", Science **369**, 1084–1089 (2020).

[91]C. H. Bennett and G. Brassard, "Quantum cryptography: public key distribution and coin tossing", Theoretical computer science **560**, 7–11 (2014).

[92]A. K. Ekert, "Quantum cryptography based on bell's theorem", Physical review letters **67**, 661 (1991).

[93]W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers", npj Quantum Information **7**, 1–9 (2021).

[94]H. F. Trotter, "On the product of semi-groups of operators", Proceedings of the American Mathematical Society **10**, 545–551 (1959).

[95]M. Suzuki, "General theory of fractal path integrals with applications to many-body theories and statistical physics", Journal of Mathematical Physics **32**, 400–407 (1991).

[96]E. Campbell, "Random compiler for fast hamiltonian simulation", Physical review letters **123**, 070503 (2019).

[97]K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning", Physical Review A **98**, 032309 (2018).

[98]M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, "Parameterized quantum circuits as machine learning models", Quantum Science and Technology **4**, 043001 (2019).

[99]X. Wang, Z. Song, and Y. Wang, "Variational quantum singular value decomposition", Quantum **5**, 483 (2021).

[100]J. Zeng, C. Cao, C. Zhang, P. Xu, and B. Zeng, "A variational quantum algorithm for hamiltonian diagonalization", Quantum Science and Technology **6**, 045009 (2021).

[101]Y. Wang, G. Li, and X. Wang, "Variational quantum gibbs state preparation with a truncated taylor series", Physical Review Applied **16**, 054035 (2021).

[102]O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, "Efficient quantum measurement of pauli operators in the presence of finite sampling error", Quantum **5**, 385 (2021).

[103]P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, "Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families", arXiv preprint arXiv:1907.13623 (2019).

[104]V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, "Measurement optimization in the variational quantum eigensolver using a minimum clique cover", The Journal of chemical physics **152**, 124114 (2020).

[105]X.-M. Zhang, T. Li, and X. Yuan, "Quantum state preparation with optimal circuit depth: implementations and applications", Physical Review Letters **129**, 230504 (2022).

[106]Y. Liu, P. Sierant, P. Stornati, M. Lewenstein, and M. Płodzień, "Quantum algorithms for inverse participation ratio estimation in multi-qubit and multi-qudit systems", arXiv preprint arXiv:2405.03338 (2024).

[107]B. Gulbahar, "K-sparse pure state tomography with phase estimation", arXiv preprint arXiv:2111.04359 (2021).

[108]A. M. Childs, A. Ostrander, and Y. Su, "Faster quantum simulation by randomization", Quantum **3**, 182 (2019).

[109]A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, "Toward the first quantum simulation with quantum speedup", Proceedings of the National Academy of Sciences **115**, 9456–9461 (2018).

[110]J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes", Nature communications **9**, 4812 (2018).

[111]C. O. Marrero, M. Kieferová, and N. Wiebe, "Entanglement-induced barren plateaus", PRX Quantum **2**, 040316 (2021).

[112]H.-k. Zhang, C. Zhu, M. Jing, and X. Wang, *Statistical analysis of quantum state learning process in quantum neural networks*, 2023, arXiv:2309.14980 [quant-ph].

[113]A. Aguado, O. Roncero, and C. Sanz-Sanz, "Three states global fittings with improved long range: singlet and triplet states of h+ 3", Physical Chemistry Chemical Physics **23**, 7735–7747 (2021).

[114]S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, "Tapering off qubits to simulate fermionic hamiltonians", arXiv preprint arXiv:1701.08213 (2017).

[115]K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, "Reducing qubit requirements for quantum simulations using molecular point group symmetries", Journal of Chemical Theory and Computation **16**, 6091–6097 (2020).

[116]C. H. Bennett and D. P. DiVincenzo, "Quantum information and computation", en, Nature **404**, Number: 6775 Publisher: Nature Publishing Group, 247–255 (2000).

[117]S. Chakraborty, A. Nema, and F. Buscemi, *One-shot purity distillation with local noisy operations and one-way classical communication*, 2022, arXiv:2208.05628 [quant-ph].

[118]A. M. Childs, R. Kothari, and R. D. Somma, "Quantum algorithm for systems of linear equations with exponentially improved dependence on precision", SIAM Journal on Computing **46**, 1920–1950 (2017).

[119]S. Johri, D. S. Steiger, and M. Troyer, "Entanglement spectroscopy on a quantum computer", Physical review B **96**, 195136 (2017).

[120]C.-M. Chung, L. Bonnes, P. Chen, and A. M. Läuchli, "Entanglement spectroscopy using quantum Monte Carlo", en, Physical review B **89**, 195147 (2014).

[121]F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, "Entanglement spectrum of a topological phase in one dimension", Physical review B **81**, 064439 (2010).

[122]G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, "Entanglement in quantum critical phenomena", Physical review letters **90**, 227902 (2003).

[123]Y. Subasi, L. Cincio, and P. J. Coles, "Entanglement spectroscopy with a depth-two quantum circuit", Journal of Physics A: Mathematical and Theoretical **52**, 044001 (2019).

[124]H. Li and F. D. M. Haldane, "Entanglement spectrum as a generalization of entanglement entropy: identification of topological order in non-abelian fractional quantum hall effect states", Physical review letters **101**, 010504 (2008).

[125]J. Yin, Y.-H. Li, S.-K. Liao, M. Yang, Y. Cao, L. Zhang, J.-G. Ren, W.-Q. Cai, W.-Y. Liu, S.-L. Li, et al., "Entanglement-based secure quantum cryptography over 1,120 kilometres", Nature **582**, 501–505 (2020).

[126]S. Pirandola, J. Eisert, C. Weedbrook, A. Furusawa, and S. L. Braunstein, "Advances in quantum teleportation", Nature photonics **9**, 641–652 (2015).

[127]M. Hayashi, D. Markham, M. Murao, M. Owari, and S. Virmani, "Bounds on multipartite entangled orthogonal state discrimination using local operations and classical communication", Physical review letters **96**, 040501 (2006).

[128]S. Subramanian and M.-H. Hsieh, "Quantum algorithm for estimating $\alpha$-renyi entropies of quantum states", Physical review A **104**, 022428 (2021).

[129]K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning", Physical review A **98**, 032309 (2018).

[130]K. C. Tan and T. Volkoff, "Variational quantum algorithms to estimate rank, quantum entropies, fidelity, and fisher information via purity minimization", Physical Review Research **3**, 033251 (2021).

[131]J. Yirka and Y. Subasi, "Qubit-efficient entanglement spectroscopy using qubit resets", en-GB, Quantum **5**, 535 (2021).

[132]A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, "Introduction to quantum noise, measurement, and amplification", Reviews of Modern Physics **82**, 1155 (2010).

[133]S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications", Physical review X **8**, 031027 (2018).

[134]R. Takagi, "Optimal resource cost for error mitigation", Physical Review Research **3**, 033178 (2021).

[135]C. Piveteau, D. Sutter, and S. Woerner, "Quasiprobability decompositions with reduced sampling overhead", npj Quantum Information **8**, 12 (2022).

[136]W. Hoeffding, "Probability inequalities for sums of bounded random variables", The collected works of Wassily Hoeffding, 409–426 (1994).

[137]A. Peres, "Separability criterion for density matrices", Physical Review Letters **77**, 1413 (1996).

[138]X. Zhao, B. Zhao, Z. Xia, and X. Wang, "Information recoverability of noisy quantum states", Quantum **7**, 978 (2023).

[139]M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties", Journal of Mathematical Physics **54**, 122203 (2013).

[140]H.-P. Breuer and F. Petruccione, *The theory of open quantum systems* (Oxford University Press, USA, 2002).

[141]C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, "Strategies for solving the fermi-hubbard model on near-term quantum computers", Physical review B **102**, 10.1103/physrevb.102.235122 (2020).

[142]E. Dagotto, "Correlated electrons in high-temperature superconductors", Reviwes of modern physics **66**, 763–840 (1994).

[143]C. Kokail, B. Sundar, T. V. Zache, A. Elben, B. Vermersch, M. Dalmonte, R. van Bijnen, and P. Zoller, "Quantum variational learning of the entanglement hamiltonian", Physical review letters **127**, 170501 (2021).

[144]N. M. Linke, S. Johri, C. Figgatt, K. A. Landsman, A. Y. Matsuura, and C. Monroe, "Measuring the rényi entropy of a two-site fermi-hubbard model on a trapped ion quantum computer", Physical review A **98**, 052334 (2018).

[145]A. Szasz, J. Motruk, M. P. Zaletel, and J. E. Moore, "Chiral spin liquid phase of the triangular lattice hubbard model: a density matrix renormalization group study", Physical review X **10**, 021042 (2020).

[146]D. L. B. Ferreira, T. O. Maciel, R. O. Vianna, and F. Iemini, "Quantum correlations, entanglement spectrum, and coherence of the two-particle reduced density matrix in the extended hubbard model", Physical review B **105**, 115145 (2022).

[147]D. Wecker, M. B. Hastings, and M. Troyer, "Progress towards practical quantum variational algorithms", Physical Review A **92**, 10.1103/physreva.92.042303 (2015).

[148]Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, "Quantum error mitigation", Reviews of Modern Physics **95**, 045005 (2023), arXiv:2210.00921.

[149]S. Endo, S. C. Benjamin, and Y. Li, "Practical Quantum Error Mitigation for Near-Future Applications", Physical review X **8**, 031027 (2018), arXiv:1712.09271.

[150]Z. Cai, X. Xu, and S. C. Benjamin, "Mitigating coherent noise using Pauli conjugation", npj Quantum Information **6**, 17 (2020), arXiv:1906.06270.

[151]B. Koczor, "Exponential Error Suppression for Near-Term Quantum Devices", Physical review X **11**, 031057 (2021), arXiv:2011.05942.

[152]Z. Cai, "Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications", npj Quantum Information **7**, 10.1038/s41534-021-00404-3 (2021), arXiv:2007.01265.

[153]W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, "Virtual distillation for quantum error mitigation", Physical review X **11**, 041036 (2021).

[154]A. Streltsov, G. Adesso, and M. B. Plenio, "Colloquium: quantum coherence as a resource", Reviews of Modern Physics **89**, 041003 (2017).

[155]K. Fang, X. Wang, L. Lami, B. Regula, and G. Adesso, "Probabilistic distillation of quantum coherence", Physical review letters **121**, 070404 (2018).

[156]B. Regula, K. Fang, X. Wang, and G. Adesso, "One-shot coherence distillation", Physical review letters **121**, 010401 (2018).

[157]Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and X. Ma, "One-shot coherence dilution", Physical review letters **120**, 070403 (2018).

[158]C. Napoli, T. R. Bromley, M. Cianciaruso, M. Piani, N. Johnston, and G. Adesso, "Robustness of coherence: an operational and observable measure of quantum coherence", Physical review letters **116**, 150502 (2016).

[159]E. Chitambar and G. Gour, "Comparison of incoherent operations and measures of coherence", Physical Review A **94**, 052336 (2016).

[160]M. Piani, M. Cianciaruso, T. R. Bromley, C. Napoli, N. Johnston, and G. Adesso, "Robustness of asymmetry and coherence of quantum states", Physical Review A **93**, 042107 (2016).

[161]H. Tajima and R. Takagi, "Gibbs-preserving operations requiring infinite amount of quantum coherence", arXiv preprint arXiv:2404.03479 (2024).

[162]C. Zhu, C. Zhu, and X. Wang, "Estimate distillable entanglement and quantum capacity by squeezing useless entanglement", arXiv preprint arXiv:2303.07228 (2023).

[163]S. Bravyi, G. Smith, and J. A. Smolin, "Trading classical and quantum computational resources", Physical Review X **6**, 021043 (2016).

[164]X. Wang, M. M. Wilde, and Y. Su, "Quantifying the magic of quantum channels", New Journal of Physics **21**, 103002 (2019).

[165]X. Wang, M. M. Wilde, and Y. Su, "Efficiently computable bounds for magic state distillation", Physical review letters **124**, 090505 (2020).

[166]S. Bravyi and J. Haah, "Magic-state distillation with low overhead", Physical Review A **86**, 052329 (2012).

[167]C. Zhu, Z. Liu, C. Zhu, and X. Wang, "Limitations of classically-simulable measurements for quantum state discrimination", arXiv preprint arXiv:2310.11323 (2023).

[168]E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, "Everything you always wanted to know about locc (but were afraid to ask)", Communications in Mathematical Physics **328**, 303–326 (2014).

[169]E. Chitambar and G. Gour, "Critical examination of incoherent operations and a physically consistent resource theory of quantum coherence", Physical review letters **117**, 030401 (2016).

[170]Z.-W. Liu, X. Hu, and S. Lloyd, "Resource destroying maps", Physical review letters **118**, 060502 (2017).

[171]B. Regula, "Probabilistic transformations of quantum resources", Physical Review Letters **128**, 110505 (2022).

[172]J. Watrous, "Semidefinite programs for completely bounded norms", arXiv preprint arXiv:0901.4709 (2009).

[173]M. F. Sacchi, "Optimal discrimination of quantum operations", Physical Review A **71**, 062340 (2005).

[174]X. Zhao, L. Zhang, B. Zhao, and X. Wang, "Power of quantum measurement in simulating unphysical operations", arXiv preprint arXiv:2309.09963 (2023).

[175]F. Buscemi and G. Gour, "Quantum relative lorenz curves", Physical Review A **95**, 012110 (2017).

[176]S. Ishizaka and M. B. Plenio, "Multiparticle entanglement manipulation under positive partial transpose preserving operations", Physical Review A **71**, 052303 (2005).

[177]A. Mani and V. Karimipour, "Cohering and decohering power of quantum channels", Physical Review A **92**, 032331 (2015).

[178]K. Bu, A. Kumar, L. Zhang, and J. Wu, "Cohering power of quantum operations", Physics Letters A **381**, 1670–1676 (2017).

[179]H. Zhu, M. Hayashi, and L. Chen, "Coherence and entanglement measures based on rényi relative entropies", Journal of Physics A: Mathematical and Theoretical **50**, 475303 (2017).

[180]T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, "Measuring all compatible operators in one series of single-qubit measurements using unitary transformations", Journal of chemical theory and computation **16**, 2400–2409 (2020).

[181]M. Grant and S. Boyd, *CVX: matlab software for disciplined convex programming, version 2.1*, `http://cvxr.com/cvx`, Mar. 2014.

[182]M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs", Lecture Notes in Control and Information Sciences, edited by V. Blondel, S. Boyd, and H. Kimura, `http://stanford.edu/~boyd/graph_dcp.html`, 95–110 (2008).