Information recoverability of noisy quantum states

Xuanqiang Zhao\textsuperscript{1,2}, Benchi Zhao\textsuperscript{1}, Zihan Xia\textsuperscript{1}, and Xin Wang\textsuperscript{1}

\textsuperscript{1}Institute for Quantum Computing, Baidu Research, Beijing 100193, China
\textsuperscript{2}QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China

Extracting classical information from quantum systems is an essential step of many quantum algorithms. However, this information could be corrupted as the systems are prone to quantum noises, and its distortion under quantum dynamics has not been adequately investigated. In this work, we introduce a systematic framework to study how well we can retrieve information from noisy quantum states. Given a noisy quantum channel, we fully characterize the range of recoverable classical information. This condition allows a natural measure quantifying the information recoverability of a channel. Moreover, we resolve the minimum information retrieving cost, which, along with the corresponding optimal protocol, is efficiently computable by semidefinite programming. As applications, we establish the limits on the information retrieving cost for practical quantum noises and employ the corresponding protocols to mitigate errors in ground state energy estimation. Our work gives the first full characterization of information recoverability of noisy quantum states from the recoverable range to the recovering cost, revealing the ultimate limit of probabilistic error cancellation.

1 Introduction
1.1 Background

The dynamics of a closed quantum system is described as a unitary evolution [1]. However, quantum systems are rarely closed in practice as they unavoidably interact with the environment. Quantum channels, stemming from the unitary dynamics in a larger Hilbert space, are considered as the proper mathematical formalism depicting the evolution of general quantum systems [2]. Quantum channels represent the quantum information manipulation processes and are essential to quantum computation [2–8].

A central subroutine of quantum computation is to extract classical information from a quantum system. The expectation value of some chosen observable, also known as shadow information, characterizes physical properties of the quantum system, making estimating expectation values the main goal of many quantum algorithms [2, 9], such as variational quantum eigensolver (VQE) [10]. The importance of expectation values has motivated the study of shadow tomography and related problems [11–13]. However, inevitable noises modeled as quantum channels can corrupt the shadow information as they undesirably change the state of the quantum system, preventing us from estimating the expectation value accurately.

It is natural to ask how a quantum channel $\mathcal{N}$ affects the information stored in quantum states. The minimum fidelity [14–16] of some initial state $|\psi\rangle\langle\psi|$ and the final state $\mathcal{N}(|\psi\rangle\langle\psi|)$ is the one we often use to answer this question. But this method is not appropriate for quantifying how well a quantum channel preserves shadow information. Take the completely phase damping channel $\mathcal{N}_{PD}(\rho) = Z\rho Z$ as an example. While $F_{\text{min}}(\mathcal{N}_{PD}) = 0$, the expectation value of a diagonal observable (e.g., observable $Z$) is unaffected. Similarly, other fidelity-based measures such as average fidelity and entanglement fidelity [14, 17] are not suitable to depict the shadow information preservation either. Some other measures for information preservation are various channel capacities (e.g., classical capacity [18–20]), where each channel capacity quantifies a quantum channel’s

\textcopyright 2023-04-03, click title to verify. Published under CC-BY 4.0.
utility to transfer information for a certain purpose. However, these capacities are derived in asymptotic settings with multiple uses of the channel, and thus are not proper guides to practical tasks with finite quantum resources to some extent. Naturally, the following key questions then arise:

1. What is a proper way to quantify the level of shadow information preservation by a quantum channel?

2. How to quantify the cost of retrieving a certain piece of shadow information given it is preserved?

1.2 Contributions

To address these two questions in an operational way, we introduce a framework for retrieving shadow information from a noisy state. Both practically relevant and theoretically interesting, this framework combines an efficient way to manipulate quantum states and general quantum operations. Note that similar protocols are employed in predicting properties of quantum states [13, 21] and mitigating quantum errors [22–26]. In particular, we suppose that multiple copies of a noisy state $N(\rho)$ are available. For each copy we apply a quantum channel randomly sampled from an ensemble of channels $\{D_j\}$ and then measure the observable of interest, as shown in Fig. 1. We say that a channel $N$ preserves the classical information inquired by an observable $O$ if we can recover this information, that is: there exist a set of quantum channels $\{D_j\}$ and a set of real numbers $\{c_j\}$ such that

$$\sum_j c_j \text{Tr}[D_j \circ N(\rho)O] = \text{Tr}[\rho O]. \quad (1)$$

We establish a necessary and sufficient condition for the preservation of the desired shadow information in terms of a relation between $O$ and $N$ that should be satisfied. This condition motivates a measure called shadow destructivity characterizing a quantum channel’s level of destroying shadow information.

We define a measure called retrieving cost to quantify the minimum cost of recovering $\text{Tr}[\rho O]$, which also quantifies how well the channel preserves the shadow information queried by a specific observable. This measure, along with the concrete retrieving protocol, is efficiently computable with respect to the system dimension via semidefinite programming [27]. We analytically obtain the values of this measure and the retrieving protocols for generalized amplitude damping (GAD) channels and depolarizing channels. Our retrieving costs set ultimate limits on the cost required for shadow retrieving, and the corresponding protocols outperform existing probabilistic error cancellation (PEC) methods [22–26]. Specifically, we employ our method to estimate the ground state energies of several molecules with VQE. The results show that the gap of sampling overhead between our method and a conventional PEC method gets larger as the size of the quantum system grows, implying potential applications of our method in implementing near-term algorithms.
1.3 Preliminaries

Before introducing our results, we set the notations and define several quantities that will be used throughout this paper. We use symbols such as $H_A$ and $H_B$ to denote finite-dimensional Hilbert spaces associated with systems $A$ and $B$, respectively. We use $d_A$ to denote the dimension of the system $A$. The sets of linear and Hermitian operators acting on $A$ are denoted by $L_A$ and $L^H_A$, respectively.

In this work, we focus on linear maps having the same input and output dimension. A linear map $N_A \rightarrow A'$ transforms linear operators in $L_A$ to linear operators in $L_{A'}$, where $H_{A'}$ is isomorphic to $H_A$. We call a linear map $N_A \rightarrow A'$ a quantum channel if it is completely positive and trace-preserving (CPTP). By saying $N$ is completely positive (CP), we mean $id_R \otimes N$ is a positive map with a reference system $R$ of arbitrary dimension. By saying $N$ is trace-preserving (TP), we mean $\text{Tr}[N(X)] = \text{Tr}[X]$ for all $X \in L_A$. A linear map $N$ is called Hermitian-preserving (HP) if $N(X) \in L^H_{A'}$ for all $X \in L^H_A$. For any linear map $N_{A \rightarrow A'}$, its Choi-Jamiołkowski matrix is given by $J_N \equiv \sum_{i,j=0}^{d_A-1} |i\rangle \langle j| \otimes N_{A \rightarrow A'}(|i\rangle \langle j|)$, where $\{|i\rangle\}_{i=0}^{d_A-1}$ is an orthonormal basis in $H_A$.

2 Main Results

2.1 A Necessary and Sufficient Condition

We study the preservation of shadow information by considering an operational task: recovering the expectation value $\text{Tr}[\rho O]$ of an observable $O$ from multiple copies of a corrupted state $\rho$, where $N$ is the noisy channel of interest and $\rho$ is an unknown state. A usual way [25] is to simulate a Hermitian-preserving and trace-preserving (HPTP) map $D$, which we call a retriever in this work, such that $D \circ N = id$, where $id$ is the identity map. While this approach is sufficient for recovering the target information, its requirement on the retriever $D$ is generally not necessary as we are only concerned with the expectation value of a specific observable. Hence, the only necessary requirement for a retriever $D$ is

$$\text{Tr}[D \circ N(\rho)O] = \text{Tr}[\rho O]$$

for any state $\rho$ and a fixed observable $O$. Our framework extends the set of potential retrievers from HPTP maps to Hermitian-preserving and trace-scaling (HPTS) maps for that HPTS maps are physically simulatable in the way prescribed by Eq. (1), as shown in Lemma 6. We say a linear map $N$ is trace-scaling (TS) if $\text{Tr}_A[J_N] = pI_A$ for some real number $p$, and HPTP maps are special cases of HPTS maps when $p = 1$.

Here, given a quantum channel $N$ and an observable $O$, we give the only condition that a retriever $D$ needs to satisfy so that Eq. (2) holds for an arbitrary state:

$$N^{\dagger} \circ D^{\dagger}(O) = O,$$

where $N^{\dagger}$ and $D^{\dagger}$ are the adjoint maps of $N$ and $D$, respectively. This condition is a direct implication of Lemma 1.
Lemma 1 Given an observable \( O \), an HP map \( \mathcal{M} \) satisfies \( \text{Tr}[\mathcal{M}(\rho)O] = \text{Tr}[\rho O] \) for any state \( \rho \) if and only if it holds that \( \mathcal{M}^\dagger(O) = O \).

Proof First, we assume that \( \mathcal{M}^\dagger(O) = O \) is true. For \( \mathcal{M} \) being HP, its adjoint map \( \mathcal{M}^\dagger \) is also HP (see Remark 1). Hence, \( \text{Tr}[\mathcal{M}(\rho)O] = \text{Tr}[\rho(\mathcal{M}^\dagger(O))] = \text{Tr}[\rho \mathcal{M}^\dagger(O)] \) for both \( O \) and \( \mathcal{M}^\dagger(O) \) being Hermitian. Then, with our assumption, we have \( \text{Tr}[\mathcal{M}(\rho)O] = \text{Tr}[\mathcal{M}^\dagger(O)\rho] = \text{Tr}[\rho O] \).

For the other direction, given \( \text{Tr}[\mathcal{M}(\rho)O] = \text{Tr}[\rho O] \) for all states \( \rho \), we immediately have \( \text{Tr}[\rho \mathcal{M}^\dagger(O)] = \text{Tr}[\rho O] \). Suppose, for the sake of contradiction, that \( \mathcal{M}^\dagger(O) \neq O \). If we define \( \Delta O = \mathcal{M}^\dagger(O) - O \), then \( \text{Tr}[\rho \Delta O] = 0 \) for all \( \rho \). However, this implies that \( \Delta O = 0 \) since otherwise we can choose a \( \sigma \) from the nonzero eigenspace of \( \Delta O \) so that \( \text{Tr}[\sigma \Delta O] = 0 \). As \( \Delta O = 0 \) contradicts with the assumption that \( \mathcal{M}^\dagger(O) \neq O \), we conclude that \( \text{Tr}[\rho \mathcal{M}^\dagger(O)] = \text{Tr}[\rho O] \) holds for all \( \rho \) only if \( \mathcal{M}^\dagger(O) = O \).

Remark 1 A linear map \( \mathcal{N} \) is Hermitian-preserving (HP) if and only if \( J_N \) is Hermitian. The adjoint of an HP map \( \mathcal{N} \) is also HP. This can be seen by checking the Hermiticity of its Choi matrix: \( (J_N^\dagger)^\dagger = J_N = J_N^\dagger \) for \( J_N \) being Hermitian.

Lemma 1 sets the criterion that we should refer to when searching for a retriever. We say a channel \( \mathcal{N} \) preserves the shadow information queried by an observable \( O \) if there exists a retriever \( \mathcal{D} \) satisfying Eq. (3). It is crucial to ask for what observables the channel preserves their corresponding shadow information. In Theorem 2, we establish a necessary and sufficient condition for the preservation of the shadow information by introducing the adjoint image of \( \mathcal{L}^\dagger \), under the channel \( \mathcal{N}_{A \to A'} \), which we define to be the image of \( \mathcal{L}^\dagger_{A'} \) under \( \mathcal{N}_{A \to A'}^\dagger \) and is denoted by \( \mathcal{N}^\dagger(\mathcal{L}^\dagger_{A'}) \).

Theorem 2 (Necessary and sufficient condition for information preservation) Given a quantum channel \( \mathcal{N}_{A \to A'} \) and an observable \( O \), there exists an HPTS map \( \mathcal{D} \) such that \( \mathcal{M} \equiv \mathcal{D} \circ \mathcal{N} \) satisfies \( \mathcal{M}^\dagger(O) = O \) if and only if \( O \in \mathcal{N}^\dagger(\mathcal{L}^\dagger_{A'}) \).

The proof for the necessity part is rather straightforward, while that for the sufficiency part is nontrivial. To prove the “if” part, we denote by \( Q \) a Hermitian operator that evolves into the observable \( O \) under the map \( \mathcal{N}^\dagger \), i.e., \( \mathcal{N}^\dagger(Q) = O \). Then, it is sufficient to construct an HPTS map \( \mathcal{D} \) such that \( \mathcal{D}^\dagger(O) = Q \). According to Lemma S1, we can instead construct an HP and unit-scaling map \( \mathcal{D}^\dagger \), which guarantees that the map \( \mathcal{D} \) is HPTS. By saying that \( \mathcal{D}^\dagger \) is unit-scaling, we mean that \( \mathcal{D}^\dagger \) scales the identity operator. The main difficulty of this construction is to ensure that \( \mathcal{D}^\dagger \) is unit-scaling. To overcome this difficulty, we split all the observables into three categories by their rank and trace.

We first consider observables that do not have full rank and are not traceless, i.e., \( \text{Tr}[O] \neq 0 \). For these observables, we exploit their kernel so that the constructed \( \mathcal{D}^\dagger \) is unit-scaling. With some manipulation, the other two cases can be reduced to this one. The detailed proof is given below.

Proof For the “only if” part, suppose such a map \( \mathcal{D} \) exists for the given channel \( \mathcal{N} \) and non-zero observable \( O \). Note that \( \mathcal{D}^\dagger \) is also a Hermitian-preserving map since its Choi operator is Hermitian, as we have noted in Remark 1. Then, we can let \( Q = \mathcal{D}^\dagger(O) \) so that \( \mathcal{N}^\dagger(Q) = \mathcal{N}^\dagger \circ \mathcal{D}^\dagger(O) = \mathcal{M}^\dagger(O) = O \). Hence, \( O \in \mathcal{N}^\dagger(\mathcal{L}^\dagger_{A'}) \) if there exists an HPTS map \( \mathcal{D} \) such that \( \mathcal{M} \equiv \mathcal{D} \circ \mathcal{N} \) satisfies \( \mathcal{M}^\dagger(O) = O \).

For the “if” part, suppose \( O \in \mathcal{N}^\dagger(\mathcal{L}^\dagger_{A'}) \), which means that there exists a Hermitian operator \( Q \) such that \( \mathcal{N}^\dagger(Q) = O \). It is sufficient to construct an HPTS map \( \mathcal{D} \) satisfying \( \mathcal{D}^\dagger(O) = Q \) so that \( \mathcal{N}^\dagger(Q) = \mathcal{N}^\dagger \circ \mathcal{D}^\dagger(O) = \mathcal{M}^\dagger(O) = O \). By Lemma S1, it is equivalent to construct a Hermitian-preserving and unit-scaling map \( \mathcal{D}^\dagger \) satisfying the above requirement. Denoting the rank of the observable \( O \) by \( k \), the trace of \( O \) by \( t \), and the dimension of the Hilbert space \( \mathcal{H}_A \) by \( d \), we complete the proof by constructing a Hermitian-preserving and unit-scaling map \( \mathcal{D}^\dagger \) satisfying \( \mathcal{D}^\dagger(O) = Q \) in each of the following three cases.

- Case 1: \( k < d \) and \( t \neq 0 \). Let \( P \) be the projection on the support of \( O \) and \( P^\perp \) be the projection on the kernel of \( O \). We can construct a map \( \mathcal{D}^\dagger \) such that

\[
\mathcal{D}^\dagger(\cdot) = \frac{Q}{t} \text{Tr}[P \cdot P] + \frac{k(I - Q)}{t(d - k)} \text{Tr}[P^\perp \cdot P^\perp].
\]
Clearly, $\mathcal{D}^\dagger$ is an HP map. Note that

\begin{equation}
\mathcal{D}^\dagger(O) = \frac{Q}{t} \text{Tr}[POP] + \frac{k(I - Q)}{t(d - k)} \text{Tr}[P^\perp OP^\perp] 
\end{equation}

\begin{equation}
= \frac{Q}{t} \text{Tr}[O] 
\end{equation}

\begin{equation}
= Q 
\end{equation}

and

\begin{equation}
\mathcal{D}^\dagger(I) = \frac{Q}{t} \text{Tr}[PIP] + \frac{k(I - Q)}{t(d - k)} \text{Tr}[P^\perp IP^\perp] 
\end{equation}

\begin{equation}
= \frac{Q}{t} \text{Tr}[P] + \frac{k(I - Q)}{t(d - k)} \text{Tr}[P^\perp] 
\end{equation}

\begin{equation}
= \frac{Q}{t} \cdot k + \frac{k(I - Q)}{t(d - k)} \cdot (d - k) 
\end{equation}

\begin{equation}
= \frac{k}{t} I, 
\end{equation}

which implies that $\mathcal{D}^\dagger$ is also unit-scaling and $\mathcal{D}^\dagger(O) = Q$.

- Case 2: $k < d$ and $t = 0$. Since $t = 0$, $O$ has at least two distinct eigenvalues. Let $O = \sum_j \lambda_j \langle \psi_j | \psi_j \rangle$ be the spectral decomposition of $O$, where each $\lambda_j \neq 0$. Define $\bar{O} \equiv O - \lambda_0 \langle \psi_0 | \psi_0 \rangle$ and $\bar{Q} \equiv Q + \Delta Q$, where $\Delta Q$ is a Hermitian operator that we will fix later. Let $k$ denote the rank of $\bar{O}$ and $\bar{t}$ the trace of $\bar{O}$. Following the definition of $\bar{O}$, we have $\bar{k} = k - 1$ and $\bar{t} = -\lambda_0$. As $\bar{k} < d$ and $\bar{t} \neq 0$, we can construct a Hermitian-preserving and unit-scaling map

\begin{equation}
\mathcal{D}^\dagger(\cdot) = \frac{\bar{Q}}{\bar{t}} \text{Tr}[\bar{P} \cdot \bar{P}] + \frac{\bar{k}(I - \bar{Q})}{\bar{t}(d - \bar{k})} \text{Tr}[\bar{P}^\perp \cdot \bar{P}^\perp] 
\end{equation}

\begin{equation}
= -\frac{\bar{Q}}{\lambda_0} \text{Tr}[\bar{P} \cdot \bar{P}] - \frac{(k - 1)(I - \bar{Q})}{\lambda_0(d - k + 1)} \text{Tr}[\bar{P}^\perp \cdot \bar{P}^\perp], 
\end{equation}

where $\bar{P}$ denotes the projection on the support of $\bar{O}$ and $\bar{P}^\perp$ be the projection on the kernel of $\bar{O}$. By the definition of $\bar{O}$, we have $\bar{P}OP = \sum_{j > 0} \lambda_j \langle \psi_j | \psi_j \rangle$ and $\bar{P}^\perp \bar{O} \bar{P}^\perp = \lambda_0 \langle \psi_0 | \psi_0 \rangle$, and thus $\text{Tr}[\bar{P}OP] = -\lambda_0$ and $\text{Tr}[\bar{P}^\perp \bar{P}^\perp] = \lambda_0$. Hence,

\begin{equation}
\mathcal{D}^\dagger(O) = -\frac{\bar{Q}}{\lambda_0} \cdot (-\lambda_0) - \frac{(k - 1)(I - \bar{Q})}{\lambda_0(d - k + 1)} \cdot \lambda_0 
\end{equation}

\begin{equation}
= \frac{d}{d - k + 1} Q + \frac{d}{d - k + 1} \Delta Q - \frac{k - 1}{d - k + 1} I. 
\end{equation}

Setting $\Delta Q = \frac{k - 1}{d} (I - Q)$, we will have $\mathcal{D}^\dagger(O) = Q$.

- Case 3: $k = d$. We first consider the case where $O = cI$ for some real coefficient $c$. As $Q$ can be any Hermitian operator satisfying $\mathcal{N}^\dagger(Q) = O$, and we know that $\mathcal{N}^\dagger$ is unital, we can let $Q = O$ so that $\mathcal{D}^\dagger$ being the identity map id will complete the proof.

Now, consider the case where $O \neq cI$ for any real coefficient $c$, which implies that $O$ must have at least two distinct eigenvalues. Let $O = \sum_j \lambda_j \langle \psi_j | \psi_j \rangle$ be the spectral decomposition of $O$, where each $\lambda_j \neq 0$ and $\lambda_0$ is the smallest eigenvalue. Then, we define $\bar{O} \equiv O - \lambda_0 I$ and $\bar{Q} \equiv Q + \Delta Q$, where $\Delta Q$ is a Hermitian operator that we will fix later. Let $\bar{k}$ denote the rank of $\bar{O}$ and $\bar{t}$ the trace of $\bar{O}$. Following the definition of $O$, we have $\bar{k} < d$ and $\bar{t} = t - d \cdot \lambda_0 > 0$. Hence, by Case 1, we can construct a Hermitian-preserving and unit-scaling map

\begin{equation}
\mathcal{D}^\dagger(\cdot) = \frac{\bar{Q}}{\bar{t}} \text{Tr}[\bar{P} \cdot \bar{P}] + \frac{\bar{k}(I - \bar{Q})}{\bar{t}(d - \bar{k})} \text{Tr}[\bar{P}^\perp \cdot \bar{P}^\perp] 
\end{equation}
\[ \frac{\hat{Q}}{t - d \cdot \lambda_0} \text{Tr}[\hat{P} \cdot \hat{P}] + \frac{\hat{k}(I - \hat{Q})}{(t - d \cdot \lambda_0)(d-k)} \text{Tr}[\hat{P}^\perp \cdot \hat{P}^\perp], \]  

(17)

where \( \hat{P} \) denotes the projection on the support of \( \hat{O} \) and \( \hat{P}^\perp \) be the projection on the kernel of \( \hat{O} \). By the definition of \( \hat{O} \), we have \( \hat{O} \hat{P} \hat{O}^\dagger = \sum_{\lambda \neq \lambda_0} \lambda_j \ket{\psi_j} \bra{\psi_j} \) and \( \hat{P}^\perp \hat{O} \hat{P}^\perp = \sum_{\lambda = \lambda_0} \lambda_j \ket{\psi_j} \bra{\psi_j} \), and thus \( \text{Tr}[\hat{P} \hat{O} \hat{P}^\dagger] = (t - (d - \hat{k}) \lambda_0) \) and \( \text{Tr}[\hat{P}^\perp \hat{O} \hat{P}^\perp] = (d - \hat{k}) \lambda_0 \). Hence,

\[ \mathcal{D}^\dagger(O) = \frac{\hat{Q}}{t - d \cdot \lambda_0} \cdot (t - (d - \hat{k}) \lambda_0) + \frac{\hat{k}(I - \hat{Q})}{(t - d \cdot \lambda_0)(d-k)} \cdot (d - \hat{k}) \lambda_0 \]

(18)

\[ = Q + \Delta Q + \frac{\hat{k}}{t - d \cdot \lambda_0} \lambda_0 I. \]

(19)

Setting \( \Delta Q = - \frac{\hat{k}}{t - d \cdot \lambda_0} \lambda_0 I \) results in \( \mathcal{D}^\dagger(O) = Q \).

Since any observable \( O \) can be categorized into one of the three cases above, we conclude that given a quantum channel \( \mathcal{N} \), if \( O \in \text{Image}(\mathcal{N}^\dagger) \), then there exists a Hermitian-preserving and unit-scaling map \( \mathcal{D}^\dagger \) such that \( \mathcal{D}^\dagger(O) = Q \) for some \( Q \) satisfying \( \mathcal{N}^\dagger(Q) = O \). As the adjoint of a Hermitian-preserving and unit-scaling map is HPPTS, we can always find an HPTS retriever \( \mathcal{D} \) so that \( \mathcal{N}^\dagger \circ \mathcal{D}^\dagger(O) = \mathcal{N}^\dagger(O) = O \) given \( O \in \mathcal{N}^\dagger(\mathcal{L}_A) \).

As illustrated in Fig. 2, Theorem 2 implies that the channel \( \mathcal{N} \) preserves the shadow information or, equivalently, the shadow information is retrievable, if and only if the corresponding observable \( O \) is in the adjoint image of \( \mathcal{L}_A^\perp \) under \( \mathcal{N} \). A physical interpretation of this theorem can be given within the Heisenberg picture, where quantum states are constant while observables evolve with time [3]. From this perspective, Theorem 2 is saying that \( \text{Tr}[\rho O] \) can be recovered if and only if there is an observable \( Q \) evolving into \( O \) under the backward dynamics prescribed by the given \( \mathcal{N} \). This theorem also manifest the ultimate limitation of a quantum channel on preserving the shadow information queried by some observables.

2.2 Quantifying the Level of Shadow Information Preservation

Following Theorem 2, for a quantum channel \( \mathcal{N}_0 \rightarrow A' \), we define the dimension of \( \mathcal{N}^\dagger(\mathcal{L}_A^\perp) \) as the channel \( \mathcal{N} \)'s effective shadow dimension \( d_s(\mathcal{N}) \), which quantifies how much shadow information is preserved by \( \mathcal{N} \). To compute the effective shadow dimension, we note that the set of all Hermitian operators \( \mathcal{L}_A^\perp \) can be viewed as a vector space over the field of real numbers, and the set of all linear operators \( \mathcal{L}_A^\perp \) is a vector space over the field of complex numbers. Both vector spaces are with dimension \( d_{A'} \), where \( d_{A'} = \text{dim}(\mathcal{H}_{A'}) \). Note that a basis for \( \mathcal{L}_A^\perp \) is also a basis for \( \mathcal{L}_A^\perp \), which implies that \( \text{dim}(\mathcal{N}^\dagger(\mathcal{L}_A^\perp)) = \text{dim}(\mathcal{N}^\dagger(\mathcal{L}_A^\perp)) \). Then, by the properties of the adjoint [28], the image of \( \mathcal{N}^\dagger \) is the space orthogonal to the null space of \( \mathcal{N} \). Hence, the dimension of \( \mathcal{N}^\dagger(\mathcal{L}_A^\perp) \) is \( d_{A'} - \text{dim}(\text{null}(\mathcal{N})) = \text{dim}(\text{Image}(\mathcal{N})) \), which equals the rank of a matrix of \( \mathcal{N} \) as a linear map.

A matrix of \( \mathcal{N} \) as a linear map can be obtained from the channel’s Kraus representation. Given \( \mathcal{N}(|i\rangle\langle j|) = \sum_k E_k |i\rangle\langle j| E_k^\dagger \), we have \( \text{vec}(\mathcal{N}(|i\rangle\langle j|)) = \sum_k \text{vec}(E_k^\dagger) \otimes E_k |i\rangle \) \( = (\sum_k \text{vec}(E_k \otimes E_k^\dagger)) |j\rangle \), where \( \text{vec}(\cdot) \) is the complex conjugate of \( \cdot \). Hence, \( M_N = \sum_k \text{vec}(E_k \otimes E_k^\dagger) \) is a matrix representation of \( \mathcal{N} \) as a linear map. The effective shadow dimension \( d_s(\mathcal{N}) \) is efficiently computable with respect to the system dimension as the matrix rank of \( M_N \).

Based on the effective shadow dimension, we define a measure called shadow destructivity:

\[ \zeta(\mathcal{N}) \equiv \log \frac{d^2}{d_s(\mathcal{N})}, \]

(20)

where \( d \) is the dimension of the Hilbert space that the input linear operators in \( \mathcal{L}_A^\perp \) act on. The shadow destructivity quantifies a quantum channel’s capability to destruct shadow information and has some meaningful properties:

1. (Faithfulness) \( \zeta(\mathcal{N}) = 0 \) if and only if all the shadow information is recoverable. In other words, \( \zeta \) is strictly larger than 0 if and only if some shadow information is irrevocably lost.

Proposition 3 (Faithfulness of shadow destructivity). The shadow destructivity is faithful in the sense that \( \zeta(\mathcal{N}) \geq 0 \) and it vanishes, i.e., \( \zeta(\mathcal{N}) = 0 \), if and only if for each observable \( O \), there exists a retriever \( \mathcal{D} \) such that \( \text{Tr}[\mathcal{D} \circ \mathcal{N}(\rho)O] = \text{Tr}[\rho O] \) for every quantum state \( \rho \).
Accepted in Quantum 2023-04-03, click title to verify. Published under CC-BY 4.0.
2.3 Quantifying the Cost of Retrieving Shadow Information

By saying that a quantum channel preserves some shadow information, we mean this information can be retrieved after a state is corrupted by this channel. Though the effective shadow dimension of a quantum channel captures its overall capability of preserving shadow information, a piece of shadow information being preserved does not mean that retrieving it is cost-free, and different pieces of information can have different retrieving cost. In the following part, we derive a measure called retrieving cost from the number of copies of the corrupted state required to estimate the desired shadow information within an acceptable accuracy.

Note that the retriever \(D\) is an HPTS map but not necessarily CPTP, i.e., not necessarily a quantum channel. While such a map \(D\) is not physically implementable, we can simulate its action by decomposing it as a linear combination of CPTP maps: \(D = \sum_j c_j D_j\), and utilizing Eq. (1). As Lemma 6 implies, a linear map is physically simulatable in this framework if and only if it is HPTS.

**Lemma 6** A linear map \(D\) is HPTS if and only if it can be written as a linear combination of quantum channels \(\{D_j\}\) with real numbers \(\{c_j\}\): \(D = \sum_j c_j D_j\). Furthermore, for any HPTS map \(D\), there exist two quantum channels \(D_1, D_2\) with real numbers \(c_1, c_2\) such that

\[
D = c_1 D_1 + c_2 D_2.
\]

**Proof** We first prove that a linear combination of quantum channels is HPTS. The Choi-Jamiolkowski matrix of such a linear combination \(D = \sum_j c_j D_j\) is \(J_D = \sum_j c_j J D_j\), where \(J D_j \geq 0\) and \(\text{Tr}_B[J D_j] = I_A\) for each \(j\). As a linear combination of Hermitian operators is also Hermitian, \(D\) is Hermitian-preserving. Also noting that \(\text{Tr}_B[J D] = \sum_j c_j \text{Tr}_B[J D_j] = \sum_j c_j I_A\), we have \(\text{Tr}_B[J D] = c I_A\), where \(c = \sum_j c_j\). Hence, \(D\) is HPTS.

For the other direction, if a map \(D\) is HPTS, its Choi-Jamiolkowski matrix \(J_D\) satisfies \(J_D\) being Hermitian and \(\text{Tr}_B[J D] = c I_A\) for some real coefficient \(c\). If \(c \neq 0\), \(D/c\) is an HPTP map and there exist two CPTP maps \(D_1, D_2\) and two real numbers \(c_1 \geq 0, c_2 \leq 0\) such that \(D/c = c_1 D_1 + c_2 D_2\) [25]. Then we have \(D = c \cdot c_1 D_1 + c \cdot c_2 D_2\). On the other hand, if \(c = 0\), \(D + \text{id}\) is an HPTP map and there still exist two CPTP maps \(D_1, D_2\) and two real numbers \(c_1 \geq 0, c_2 \leq 0\) such that \(D + \text{id} = c_1 D_1 + c_2 D_2\). Hence, \(D\) can be written as \(D = c_1 D_1 + (c_2 - 1) D_2\), where \(D_2 = c_2 D_2/(c_2 - 1) - \text{id}/(c_2 - 1)\) is a quantum channel.

Now let \(D\) be an HPTS map whose decomposition is \(\sum_j c_j D_j\). With this decomposition, we can simulate its action on the expectation value by the probabilistic sampling method prescribed by Eq. (1). Specifically, in the \(s\)-th round of total \(S\) times of sampling, we first sample a quantum channel \(D^{(s)}\) from \(\{D_j\}\) with probabilities \(\{|c_j|/\gamma\}\), where \(\gamma = \sum_j |c_j|\), and apply it to a copy of the corrupted state to obtain \(D^{(s)} \circ N(\rho)\). Then, we measure this state in the eigenbasis \(\{|\psi_j\rangle\langle\psi_j|\}\) of the given observable \(O = \sum_j o_j |\psi_j\rangle\langle\psi_j|\), where \(o_j \in [-1, 1]\), and obtain the measurement value \(o^{(s)}\). After \(S\) rounds of sampling, we attain an estimation for the expectation value \(\text{Tr}[\rho O]\) as

\[
\xi = \frac{\gamma}{S} \sum_{s=1}^{S} \text{sgn}(o^{(s)}) o^{(s)}.
\]

By the Hoeffding’s inequality [30], the number of rounds required to obtain the estimation within an error \(\varepsilon\) with a probability no less than \(1 - \delta\) is

\[
S = 2\gamma^2 \log(2/\delta)/\varepsilon^2.
\]

It can be seen that the required number of copies \(S\) of the corrupted state is directly related to \(\gamma\), the sum of absolute values of all \(c_j\). Hence, \(\gamma\) can be used to characterize the cost of simulating the HPTS map \(D\). It is desirable to find a decomposition of \(D\) making \(\gamma\) as small as possible, and we denote the minimum possible value of \(\gamma\) as \(\gamma_{\min}(D)\). In Ref. [25], the authors define the logarithm of \(\gamma_{\min}(D)\) as the physical implementability of an HPTP map \(D\).

Naturally, given a channel \(N\), we define the retrieving cost with respect to an observable \(O\) to be

\[
\gamma_O(N) = \min \{\gamma_{\min}(D)|D\text{ is HPTS, } N^\dagger \circ D^\dagger(O) = O\},
\]
which is the $\gamma_{\min}(D)$ minimized over $D$ that can retrieve the shadow information queried by $O$ from the noisy channel $N$.

Notably, Lemma 6 states that any HPTS map can be decomposed as a linear combination of two quantum channels (Eq. (26)), and we show in the Supplemental Material that the minimum retrieving cost can be achieved by such a decomposition. This is true because all channels with non-negative coefficients can be grouped into a single channel with a non-negative coefficient, and we can do the same thing for all channels with negative coefficients. The existence of such an optimal decomposition with two CPTP maps leads to an efficient way to compute this measure through a semidefinite program (SDP) in terms of linear maps’ Choi-Jamiolkowski matrices:

$$\gamma_{O}(N) = \min c_1 + c_2$$

subject to $J_D \equiv J_{D_1} - J_{D_2}$

$$\begin{aligned}
J_{D_1} &\geq 0, J_{D_2} \geq 0 \\
\text{Tr}_{A_i'}[J_{D_1A_i'A_i'}] &\equiv c_1 I_{A_i'}, \quad \text{Tr}_{A_i'}[J_{D_2A_i'A_i'}] = c_2 I_{A_i'} \\
J_{M_{A_i,A_i'}} &\equiv \text{Tr}_{A_i'A_i'}[(J_{N_{A_i,A_i'}}^T \otimes I_{A_i,A_i})(J_{d_{A_i,A_i}} \otimes J_{D_{A_i'A_i'}})] \\
\text{Tr}_{A_i'}[(I_{A_i} \otimes O_{M_{A_i,A_i'}}^T)J_{M_{A_i,A_i'}}^T] &\equiv O_{A_i'},
\end{aligned}$$

where $J_{d_{A_i,A_i}}$ is the Choi-Jamiolkowski matrix of an identity channel. Eq. (30c) corresponds to the condition that $D_1$ and $D_2$ are CP, and Eq. (30d) requires them to be TP. Eq. (30f) corresponds to the minimum requirement for a valid retriever as given in Eq. (3). Note that we can establish approximate versions of Eq. (30) by relaxing the requirement in Eq. (30f) such that $\text{Tr}_{A_i'}[(I_{A_i} \otimes O_{M_{A_i,A_i'}}^T)J_{M_{A_i,A_i'}}^T]$ is close to $O$ within a certain tolerance. This relaxation provides a remedy to cases where a perfect retriever does not exist as well as a trade-off between the retrieving cost and the estimation precision.

### 2.4 Case Study and Applications in Error Mitigation

To intuitively understand effective shadow dimension and shadow destructivity, we study two non-invertible channels as examples: $\tilde{N}_1(\cdot) = \frac{1}{2}I(\cdot)I + \frac{1}{2}X(\cdot)X$ and $\tilde{N}_2(\cdot) = \frac{1}{2}I(\cdot)I + \frac{1}{4}X(\cdot)X + \frac{1}{4}Y(\cdot)Y$.

For a quantum state $\rho = \begin{pmatrix} a & b + ci \\ b - ci & 1 - a \end{pmatrix}$, where $a, b, c$ are real numbers and $i \equiv \sqrt{-1}$, it can be easily verified that $\tilde{N}_1(\rho) = \begin{pmatrix} 1/2 & b \\ b & 1/2 \end{pmatrix}$ and $\tilde{N}_2(\rho) = \begin{pmatrix} 1/2 & (b + ci)/2 \\ (b - ci)/2 & 1/2 \end{pmatrix}$. While both channels are non-invertible, it is intuitive to see that channel $\tilde{N}_2$ preserves more information than channel $\tilde{N}_1$ since the imaginary part of off-diagonal elements is preserved. This intuition coincides with the effective shadow dimension of these two channels, $d_{\tilde{N}_2} = 3 > d_{\tilde{N}_1} = 2$, which implies that $\tilde{N}_2$ preserves one more shadow dimension than $\tilde{N}_1$. The shadow destructivity of these channels, $\zeta(\tilde{N}_1) = \log(4/2) = 1$ and $\zeta(\tilde{N}_2) = \log(4/3) \approx 0.415$, has similar implications.

Now recall the retrieving cost, which quantifies the resources required to recover a certain piece of shadow information. The corresponding optimal retrieving protocol can be considered as a method to mitigate errors, endowing the retrieving cost with a practical meaning in quantum error mitigation [22–26, 26, 31–44].

In error mitigation, PEC methods are promising to suppress noises in quantum circuits run on near-term quantum computers and return us unbiased estimation of expectation values [22–25]. Briefly speaking, the idea of PEC is to decompose the inverse (HPTP) of a noisy channel $N$ into a linear combination of physically implementable channels (CPTP), i.e., $N^{-1} = \sum_j a_j A_j$, with real coefficients $a_j$ and CPTP maps $A_j$, where $\sum_j a_j = 1$. The inverse map $N^{-1}$ is implemented by sampling the channels $A_j$ according to a probability distribution $p(j) = \frac{a_j}{\gamma_{\text{con}}}$, where $\gamma_{\text{con}} = \sum_j |a_j|^2$. In this way, the unbiased estimation of the expectation value of an observable $O$ with respect to a state $\rho$ is obtained by manipulating multiple copies of the noisy state $N(\rho)$:

$$\gamma_{\text{con}} \sum_j p(j) \text{Tr}[A_j \circ N(\rho)O] = \text{Tr}[N^{-1} \circ N(\rho)O] = \text{Tr}[\rho O].$$

(31)
The quantity $\gamma_{\text{con}}$ is known as the overhead (also called retrieving cost in the setting of our work) because it captures how many copies of the noisy state are needed to ensure the estimation within the acceptable precision. Therefore, $\gamma_{\text{con}}$ is usually used to evaluate the efficiency of an error mitigation protocol.

Our proposed method can be considered as an approach to mitigating errors, where the retrieving cost corresponds to the overhead of an error mitigation protocol. For a given noisy channel $\mathcal{N}$ and an observable $O$ of interest, the SDP in Eq. (30) can give us a retriever $D$ and its decomposition $D = c_1 D_1 + c_2 D_2$ with the optimized retrieving cost $\gamma_{\text{pro}} = |c_1| + |c_2|$. The expectation of the estimation made by our method is

\[
\gamma_{\text{pro}} \left( \frac{|c_1|}{\gamma_{\text{pro}}} \text{Tr}[D_1 \circ \mathcal{N}(\rho)O] + \frac{|c_2|}{\gamma_{\text{pro}}} \text{Tr}[D_2 \circ \mathcal{N}(\rho)O] \right) = \text{Tr}[D \circ \mathcal{N}(\rho)O] \quad (32)
\]

\[
= \text{Tr}[\rho^\dagger \circ D^\dagger(O)] \quad (33)
\]

\[
= \text{Tr}[\rho O]. \quad (34)
\]

Note that in our method, the retriever $D$ is not necessarily the inverse of the noisy channel, i.e., $D \neq \mathcal{N}^{-1}$.

In the following, we compare the retrieving cost between our proposed protocol $\gamma_{\text{pro}}$ and the conventional PEC method $\gamma_{\text{con}}$ (see, e.g., [22–26]), with the assumption that full knowledge of the noise is accessible and the noise is modeled as a channel just before the measurement.

As a concrete example, consider retrieving shadow information $\text{Tr}[\rho X]$ from a state $\rho$ corrupted by a GAD channel [45], where $X$ is the Pauli $X$ operator. The cost of our method is

\[
\gamma_{\text{pro}} = \frac{1}{\sqrt{1-\epsilon}},
\]

while the cost of the conventional method is $\gamma_{\text{con}} = \frac{1-2p\epsilon + 1}{1-\epsilon}$ [25], where $\epsilon$ is the damping factor and $p$ is the temperature indicator associated with the GAD channel. It is obvious to see that $\gamma_{\text{pro}} < \gamma_{\text{con}}$ for any $0 < \epsilon < 1$, as shown in Fig. 3. Detailed protocols are provided in the Supplemental Material.

Now we apply our protocol to ground state energy estimation with VQE. VQE [10] is a promising near-term algorithm for estimating the ground state energy of a Hamiltonian $H$. It aims to minimize the cost function $E = \text{Tr}[\psi(\theta)\psi(\theta)^\dagger H]$ by tuning parameters $\theta$ in a parameterized quantum circuit used to prepare the ansatz state $|\psi(\theta)\rangle$. In practice, $H$ is decomposed into a linear combination of tensor products of Pauli tensors $\{O_j\}$, i.e., $H = \sum_j h_j O_j$, where each $h_j$ is a real coefficient. The core subroutine of VQE is estimating the expectation value of each term $\text{Tr}[\rho O_j]$.

We use Eq. (28) to estimate the numbers of sampling rounds required by VQE to estimate the ground state energies of given molecules when there is depolarizing noise [2] on each qubit. Table 1 shows the Comparison between our protocol and the conventional protocol [25]. The retrieving cost of our proposed method is $\gamma_{\text{pro}} = \frac{1}{1-\epsilon}$ (see Supplemental Material), while the conventional
method is $\gamma_{\text{con}} = \frac{1+(1-2/d^2)\epsilon}{1-\epsilon}$, where $\epsilon$ is the noise level, and $d$ is the dimension of the system. It is clear that the number of sampling rounds of our method is smaller than that of the conventional method.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>H$_2$</th>
<th>HF</th>
<th>CO$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>$1.24 \times 10^5$</td>
<td>$7.84 \times 10^{10}$</td>
<td>$1.28 \times 10^{13}$</td>
</tr>
<tr>
<td>Conventional</td>
<td>$1.60 \times 10^5$</td>
<td>$1.79 \times 10^{11}$</td>
<td>$1.21 \times 10^{14}$</td>
</tr>
</tbody>
</table>

Table 1: Sampling times for estimating the ground state energies of three different molecules under the depolarizing noise with a noise level $\epsilon = 0.1$. The precision parameters in Eq. (28) are set to $\epsilon = \delta = 0.01$. For estimating the sampling times, each coefficient $h_j$ of a Hamiltonian $H = \sum_j h_j O_j$ is assumed to be $\max\{|h_j|\}$, and the term where $O_j$ is the identity operator is excluded.

Besides the advantage of having a lower cost, our method has a larger range of applicability. The conventional method implements the retriever as the inverse of the noisy channel. In contrast, it only works for the cases where the noise is invertible. However, our method, employing an observable-adaptive strategy, also works for some cases where the noise is non-invertible. As long as the observable of interest is in the image of the noise’s adjoint map, our method can provide an optimal protocol for recovering the corresponding shadow information.

3 Conclusions

In this work, we establish two measures from an operational perspective to quantify a quantum channel’s influence on shadow information encoded in quantum states, answering the questions posed at the beginning. Our work delivers a systematic framework of quantifying the information preservativity and destructivity of a quantum channel. Meanwhile it establishes an optimal way of extracting noiseless classical information from noisy states, which could be useful for near-term quantum information processing tasks.

While the shadow destructivity and the retrieving cost capture different aspects of shadow information recoverability, a single unified measure that captures both aspects may be worth further studies. For future work, it would be interesting to see how the techniques presented in this work can be combined with quantum error correction [46–48] or quantum error mitigation [37, 41–44]. We also expect that our ideas could be applied to near-term quantum tasks or applications on noisy quantum devices [49].

Acknowledgements

X. Z. and B. Z. contributed equally to this work. We would like to thank Chenghong Zhu for helpful discussions. We would also like to thank the referees for their helpful comments, which greatly improved the readability of the manuscript. Part of this work was done when X. Z., B. Z., and Z. X. were research interns at Baidu Research.

References


Accepted in *Quantum* 2023-04-03, click title to verify. Published under CC-BY 4.0.


1 Adjoint of a Trace-scaling Map

**Lemma S1** The adjoint of a trace-scaling map is unit-scaling, and vice versa. By calling a linear map $N$ unit-scaling, we mean it scales the identity operator, i.e., $N(I) = pI$ for some real number $p$.

**Proof** Suppose that $N$ is a trace-scaling map such that $\text{Tr}[N(\cdot)] = p \text{Tr}[\cdot]$ for some real number $p$. If $p \neq 0$, we can define a TP map $M \equiv N/p$. It is known that the adjoint of TP maps are unital [50], which preserves the identity operator, and vice versa. Thus, $M^\dagger$, the adjoint of $M$, is a unit map. Then, $N^\dagger = pM^\dagger$, the adjoint of the trace-scaling map $N$, is a unit-scaling map. On the other hand, if $p = 0$, then $N + \text{id}$ is TP and thus $N^\dagger + \text{id}$ is unital, making $N^\dagger$ a unit-scaling map. Hence, the adjoint of a trace-scaling map is unit-scaling.

Now suppose that $N$ is a unit-scaling map such that $N(I) = pI$ for some real number $p$. If $p \neq 0$, then $M \equiv N/p$ is a unital map. Thus, we know its adjoint $M^\dagger$ is TP, making $N^\dagger = pM^\dagger$ trace-scaling. On the other hand, if $p = 0$, then $N + \text{id}$ is unital and hence its adjoint $N^\dagger + \text{id}$ is TP, again making $N^\dagger$ trace-scaling. Hence, the adjoint of a unit-scaling map is trace-scaling. $\blacksquare$

2 Retrieving Cost

Recall that, given a quantum channel $N$, the shadow retrieving cost with respect to an observable $O$ is defined as

$$\gamma_O(N) = \min \{ \gamma_{\min}(D) | D \text{ is HPTS, } N^\dagger \circ D^\dagger(O) = O \}, \quad (S1)$$

where $\gamma_{\min}(D)$ is the minimum cost for simulating $D$. Here, we give a proof that the minimum cost of simulating an HPTS map can be achieved by a decomposition with two CPTP maps. The same thing has been proved for HPTP maps in Ref. [25], and the proof here is similar.

**Proposition S2** For any HPTS map $D$,

$$\gamma_{\min}(D) = \min \left\{ |c_1| + |c_2| |D = c_1D_1 + c_2D_2; \ D_1, D_2 \text{ are CPTP}; \ c_1, c_2 \in \mathbb{R} \right\}. \quad (S2)$$

**Proof** Lemma 6 implies that $\gamma_{\min}(D)$ is finite as $D$ can be decomposed into a linear combination of two CPTP maps. Suppose $\gamma_{\min}(D)$ is achieved by $D = \sum_{j=1}^T c_j^j D_j^j$, where $T \geq 3$ so that $\gamma_{\min}(D) = \sum_{j=1}^T |c_j^j|$. Then, it is always possible to construct two CPTP maps $D_1, D_2$ so that $D = c_1D_1 + c_2D_2$ and $|c_1| + |c_2| = \gamma_{\min}(D)$. In particular, let $c_1 = \sum_{j:j'<0} c_j^j$, $D_1 = \sum_{j:j'\geq 0} c_j/c_j^j \cdot D_j^j$, $c_2 = \sum_{j:j'<0} c_j^j$, and $D_2 = \sum_{j:j'\geq 0} c_j^j/c_2 \cdot D_j^j$. One can check that $D_1$ and $D_2$ are CPTP and $c_1D_1 + c_2D_2 = \sum_{j=1}^T c_j^j D_j^j$. Thus, this decomposition of $D$ is valid. Plus, $|c_1| + |c_2| = \sum_{j=1}^T |c_j^j|$, ensuring that this decomposition achieves $\gamma_{\min}(D)$. $\blacksquare$

2.1 Dual SDP for Retrieving Cost

For a quantum channel $N_{A\rightarrow A'}$, the primal SDP characterization of its retrieving cost with respect to an observable $O$ is

$$\gamma_O(N) = \min c_1 + c_2 \quad (S3a)$$

s.t. $J_D \equiv J_{D_1} - J_{D_2}$

$$J_{D_i} \geq 0, \quad J_{D_2} \geq 0 \quad (S3b)$$

$$\text{Tr}_{A'}[J_{D_1A'_{A''}}] = c_1 I_{A'}, \quad \text{Tr}_{A'}[J_{D_2A'_{A''}}] = c_2 I_{A'} \quad (S3c)$$

$$J_{iA''_{A'}} \equiv \text{Tr}_{A''_{A'}}[(J_{N_{A''_{A'}} A''_A''} \otimes I_{A'''})(J_{D_{A''''_{A''''}} A''''_A''''})] \quad (S3d)$$
The Lagrange function is

$$L(J_{D_1}, J_{D_2}, c_1, c_2, M, N, K) = c_1 + c_2 + \langle M, \text{Tr}_{A_c}[J_{D_2}] \rangle - c_1 I_{A_c} + \langle N, \text{Tr}_{A_c}[J_{D_2}] \rangle - c_2 I_{A_c} + \langle K, \text{Tr}_{A_c}[(I_{A_c} \otimes O_{A_c})J_{M_{A_c,A_c'}}] \rangle - O_{A_c} \text{.}$$

(S4)

Thus, the Lagrange function can be expressed as

$$
\langle K_{A_c}, \text{Tr}_{A_c}[I_{A_c} \otimes O_{A_c}^T]J_{M_{A_c,A_c'}} \rangle = \text{Tr}[K_{A_c} \text{Tr}_{A_c}[(I_{A_c} \otimes O_{A_c})J_{M_{A_c,A_c'}}] \rangle \text{.}
$$

(S6)

Thus, the Lagrange function can be expressed as

$$L(J_{D_1}, J_{D_2}, c_1, c_2, M, N, K) = c_1(1 - \text{Tr}[M]) + c_2(1 - \text{Tr}[N]) - \text{Tr}[K_{O}]
$$

(S15)

For $J_{D_1} \geq 0$ and $J_{D_2} \geq 0$, it must hold that $\text{Tr}[M] \leq 1$, $\text{Tr}[N] \leq 1,$

$$
M_{A_c} \otimes I_{A_c} + \text{Tr}_{A_c}[:[(K_{A_c}^T \otimes J_{N_{A_c,A_c'}}^T \otimes O_{A_c})](J_{M_{A_c,A_c'}} \otimes I_{A_c,A_c'})] \rangle \geq 0, \text{ and}
$$

(S16)

$$
N_{A_c} \otimes I_{A_c} - \text{Tr}_{A_c}[:[(K_{A_c}^T \otimes J_{N_{A_c,A_c'}}^T \otimes O_{A_c})](J_{M_{A_c,A_c'}} \otimes I_{A_c,A_c'})] \rangle \geq 0.
$$

(S17)

Thus, we arrive at the following dual SDP:

$$
\gamma_{O}(N) = \max_{\text{Tr}[M] \leq 1, \text{Tr}[N] \leq 1} \text{Tr}[K_{O}]
$$

s.t. $M_{A_c} \otimes I_{A_c} + \text{Tr}_{A_c}[:[(K_{A_c}^T \otimes J_{N_{A_c,A_c'}}^T \otimes O_{A_c})](J_{M_{A_c,A_c'}} \otimes I_{A_c,A_c'})] \rangle \geq 0$

(S18a)

$$
N_{A_c} \otimes I_{A_c} - \text{Tr}_{A_c}[:[(K_{A_c}^T \otimes J_{N_{A_c,A_c'}}^T \otimes O_{A_c})](J_{M_{A_c,A_c'}} \otimes I_{A_c,A_c'})] \rangle \geq 0.
$$

(S18d)
3 Retrieving Cost for GAD channel

A GAD channel can be used to describe the energy dissipation to the environment at finite temperature [45]. It is one of the realistic sources of noise in superconducting quantum computing. For single-qubit cases, a GAD channel can be characterized by the following Kraus operators:

\[
E_0 = \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, E_1 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{\epsilon} \\ 0 & 0 \end{pmatrix}, \quad (S19) \\
E_2 = \sqrt{1-p} \begin{pmatrix} \sqrt{1-\epsilon} & 0 \\ 0 & 1 \end{pmatrix}, E_3 = \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{1-\epsilon} \end{pmatrix}, \quad (S20)
\]

where \( \epsilon \) is the damping factor and \( p \) is the indicator of the temperature of the environment. A quantum state \( \rho \) after going through the GAD channel is given by \( \mathcal{N}_{\text{GAD}}(\rho) = \sum_{i=0}^{3} E_i \rho E_i^\dagger \). Note that the amplitude damping channel is a special case of the GAD channel when \( p = 1 \). A single qubit state \( \rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \) after going through the GAD channel is

\[
\rho' = \mathcal{N}_{\text{GAD}}(\rho) = \begin{pmatrix} (1-\epsilon)\rho_{00} + p\epsilon(\rho_{00} + \rho_{11}) & \sqrt{1-\epsilon}\rho_{01} \\ \sqrt{1-\epsilon}\rho_{10} & \rho_{11} + \epsilon\rho_{00} - p\epsilon(\rho_{00} + \rho_{11}) \end{pmatrix}, \quad (S21)
\]

where \( \epsilon \) and \( p \) are noise parameters.

**Proposition S3** Given an observable \( O \in \{X,Y\} \) and a GAD channel \( \mathcal{N} \), the minimum cost \( \gamma \) to retrieve the information \( \text{Tr}[\rho O] \) is \( \gamma_0(\mathcal{N}) = \frac{1}{\sqrt{1-\epsilon}} \), and the corresponding retriever \( D \) can be written in the form of a Choi matrix

\[
J_D = c_1 J_{D_1} + c_2 J_{D_2},
\]

where \( c_1 = -c_2 = \frac{1}{2\sqrt{1-\epsilon}} \), \( J_{D_1} = \frac{1}{2} O^T \otimes O + \frac{1}{2} I \otimes I \), and \( J_{D_2} = -\frac{1}{2} O^T \otimes O + \frac{1}{2} I \otimes I \).

**Proof** First, we are going to prove \( \gamma_0(\mathcal{N}) \leq \frac{1}{\sqrt{1-\epsilon}} \) using SDP (S3). We show that the retriever \( D \) above is a feasible solution with a cost of \( \frac{1}{\sqrt{1-\epsilon}} \). So to be specific, we have

\[
D \circ \mathcal{N}(\rho) = D(\rho') = \text{Tr}[\mathcal{A}(\rho'T \otimes I) \cdot J_D] = \frac{1}{2\sqrt{1-\epsilon}} \text{Tr}[\mathcal{A}(\rho'T \otimes I)(O^T \otimes O)] = \frac{1}{2\sqrt{1-\epsilon}} \text{Tr}[\mathcal{A}\rho'T O^T \otimes IO] = \frac{1}{2\sqrt{1-\epsilon}} O \text{Tr}[\rho'O]. \quad (S26)
\]

This means that

\[
\text{Tr}[D \circ \mathcal{N}(\rho)O] = \frac{1}{2\sqrt{1-\epsilon}} \text{Tr}[O^2] \text{Tr}[\rho'O] = \frac{1}{\sqrt{1-\epsilon}} \text{Tr}[\rho'O] = \text{Tr}[O\rho], \quad (S29)
\]

where the second equality follows from the fact that \( O^2 = I \) for \( O \in \{X,Y\} \), and the third equality can be verified with direct calculation using Eq. (S21). Here, we have proven that the retriever \( D \) is a feasible solution to retrieve information with cost \( \frac{1}{\sqrt{1-\epsilon}} \), implying that \( \gamma_0(\mathcal{N}) \leq |c_1| + |c_2| = \frac{1}{\sqrt{1-\epsilon}} \).

Second, we use the dual SDP (S18) to show the cost \( \gamma_0(\mathcal{N}) \geq \frac{1}{\sqrt{1-\epsilon}} \). We show that \( \{M, N, K\} \) is a feasible solution to the dual problem, where \( M = N = \frac{1}{2} \) and \( K = -\frac{1}{2\sqrt{1-\epsilon}} O \). To be specific, we have

\[
\text{Tr}[A_i A_\alpha^\dagger (K_{A_i}^T \otimes J_{\alpha A_\alpha A_\alpha^\dagger}^O \otimes O_{A_\alpha})(J_{d A_i A_\alpha} \otimes I_{A_\alpha A_\alpha^\dagger})]
\]
Thus, we need a retriever to scale the expectation value back, and such a retriever corresponds to a retrieving cost of $1/\sqrt{1-\varepsilon}$, which aligns with the above proposition. The same phenomena are observed for the mixed Pauli noises with Pauli observables (see Proposition S5).

**Proposition S4** Given an observable $O = Z$ and a GAD channel $\mathcal{N}$, the minimum cost $\gamma_O(\mathcal{N})$ to retrieve the information $\text{Tr}[\rho O]$ is $\gamma_O(\mathcal{N}) = \frac{|1-2p|+1}{1-\varepsilon}$, and the corresponding retriever $D$ can be written in the form of a Choi matrix

$$J_D = c_1 J_{D_1} + c_2 J_{D_2},$$
where \( c_1 = -c_2 = \frac{|1-2p|^{e+1}}{2(1-\epsilon)} \), and

\[
J_{D_1} = \frac{1}{2(|1-2p|\epsilon + 1)} Z \otimes Z + \frac{\epsilon(1-2p)}{2(1-2p|\epsilon + 1)} I \otimes Z + \frac{1}{2} I \otimes I \quad (S41)
\]

\[
J_{D_2} = -\frac{1}{2(|1-2p|\epsilon + 1)} Z \otimes Z - \frac{\epsilon(1-2p)}{2(1-2p|\epsilon + 1)} I \otimes Z + \frac{1}{2} I \otimes I .
\quad (S42)
\]

**Proof** First, we are going to prove \( \gamma_{O}(\mathcal{N}) \leq \frac{|1-2p|^{e+1}}{1-\epsilon} \) by showing that the retriever \( \mathcal{D} \) is a feasible solution to the SDP \( (S3) \). To be specific, we have

\[
\mathcal{D} \circ \mathcal{N}(\rho) = \mathcal{D}(\rho')
\]

\[
= \text{Tr}_{A}[ (\rho'^{T} \otimes I) \cdot J_{D} ]
\]

\[
= \text{Tr}_{A} \left[ (\rho'^{T} \otimes I) \cdot \left\{ \frac{|1-2p|\epsilon + 1}{2(1-\epsilon)} Z \otimes Z + \frac{\epsilon(1-2p)}{1-2p|\epsilon + 1} I \otimes Z \right\} \right]
\]

\[
= \frac{1}{2(1-\epsilon)} Z \text{Tr}[\rho' Z] + \frac{\epsilon(1-2p)}{2(1-\epsilon)} Z \text{Tr}[\rho'] .
\quad (S46)
\]

Therefore, we further have

\[
\text{Tr}\left[Z \mathcal{D} \circ \mathcal{N}(\rho)\right] = \frac{1}{2(1-\epsilon)} \text{Tr}[Z^2] \text{Tr}[\rho' Z] + \frac{\epsilon(1-2p)}{2(1-\epsilon)} \text{Tr}[Z^2] \text{Tr}[\rho']
\]

\[
= \frac{1}{(1-\epsilon)} ((1-\epsilon) \rho_{00} + p\epsilon(\rho_{00} + \rho_{11}) - \rho_{11} - \epsilon \rho_{00} + \rho_{00} + \rho_{11}) + \epsilon(1-2p)(\rho_{00} + \rho_{11})
\]

\[
= \rho_{00} - \rho_{11}
\quad (S49)
\]

\[
= \text{Tr}[Z\rho] .
\quad (S50)
\]

where the second equality follows from Eq. \((S21)\) and the fact that \( Z^2 = I \). Hence, the retriever \( \mathcal{D} \) is a feasible solution to retrieve information, implying that \( \gamma_{O}(\mathcal{N}) \leq |c_1| + |c_2| = \frac{|1-2p|^{e+1}}{1-\epsilon} \).

Second, we use the dual SDP \( (S18) \) to show the cost \( \gamma_{O}(\mathcal{N}) \geq \frac{|1-2p|^{e+1}}{1-\epsilon} \). To prove this, we are going to split into several cases: \( 0 \leq p < \frac{1}{2} \), \( p = \frac{1}{2} \), and \( \frac{1}{2} < p \leq 1 \). First note that

\[
\text{Tr}_{A_{i}}[ (K^T_{A_{i}} \otimes J^T_{A_{i}} \otimes O_{A_{i}})(J_{d_{A_{i}}A_{o}} \otimes I_{A'_{i}A'_{o}}) ]
\]

\[
= \text{Tr}_{A_{i}}[ (K^T \otimes \sum_{i,j} \langle j | i \rangle \langle i | \otimes \mathcal{N}^{T}(\langle i | j \rangle) \otimes Z) \sum_{m,n} |m\rangle \langle n| \otimes |m\rangle \langle n| \otimes I \otimes I) ]
\]

\[
= \text{Tr}_{A_{i}}[ \sum_{i,j,m,n} K^T[m] \langle n| \otimes |j \rangle |i\rangle |m\rangle \langle n| \otimes \mathcal{N}^{T}(\langle i | j \rangle) \otimes Z]
\]

\[
= \sum_{i,j,n} \text{Tr}[K^T |i\rangle \langle n|] \text{Tr}[|j\rangle \langle n| \mathcal{N}^{T}(\langle i | j \rangle) \otimes Z]
\]

\[
= \sum_{i,j} \text{Tr}[K^T |i\rangle \langle j| \mathcal{N}^{T}(\langle i | j \rangle) \otimes Z]
\quad (S54)
\]

- When \( 0 \leq p < \frac{1}{2} \), we set the dual variables \( \{M, N, K\} \) as \( M = N = |0\rangle \langle 0|, K = \frac{1}{2}I + \frac{2p\epsilon - 1 - \epsilon}{2(1-\epsilon)} Z \) and prove that they form a feasible solution to the dual SDP.

\[
\text{Tr}_{A_{i}}[ (K^T_{A_{i}} \otimes J^T_{A_{i}} \otimes O_{A_{i}})(J_{d_{A_{i}}A_{o}} \otimes I_{A'_{i}A'_{o}}) ]
\]

\[
= \sum_{i,j} \left( -\frac{1}{2} \text{Tr}[|i\rangle \langle j|] + \frac{2p\epsilon - 1 - \epsilon}{2(1-\epsilon)} \text{Tr}[Z |i\rangle \langle j|] \right) \mathcal{N}^{T}(\langle i | j \rangle) \otimes Z
\]

\[
= \sum_{i} \left( -\frac{1}{2} + (-1)^{i} \frac{2p\epsilon - 1 - \epsilon}{2(1-\epsilon)} \right) \mathcal{N}^{T}(\langle i | i \rangle) \otimes Z
\quad (S55)
\]

\[
\text{Tr}_{A_{o}}[ (K^T_{A_{o}} \otimes J^T_{A_{o}} \otimes O_{A_{o}})(J_{d_{A_{i}}A_{o}} \otimes I_{A'_{i}A'_{o}}) ]
\]

\[
= \sum_{i} \left( -\frac{1}{2} + (-1)^{i} \frac{2p\epsilon - 1 - \epsilon}{2(1-\epsilon)} \right) \mathcal{N}^{T}(\langle i | i \rangle) \otimes Z
\quad (S56)
\]
\[
= \left(\frac{p - 1}{1 - \epsilon} N^T(\langle 0 | 0 \rangle) + \frac{\epsilon - p \epsilon}{1 - \epsilon} N^T(\langle 1 | 1 \rangle) \right) \otimes Z \tag{S57}
\]

\[
= -\langle 0 | 0 \rangle \otimes Z \tag{S58}
\]

\[
= -\langle 00 | 00 \rangle + \langle 01 | 01 \rangle. \tag{S59}
\]

Therefore,
\[
M'_{A_i} \otimes I_{A_o} + Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'_{A_iA_o}}^T \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = 2\langle 01 | 01 \rangle \geq 0 \tag{S60}
\]

\[
N_{A_i} \otimes I_{A_o} - Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'A_iA_o} \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = 2\langle 00 | 00 \rangle \geq 0. \tag{S61}
\]

- When \( p = \frac{1}{2} \), we set the dual variables \( \{ M, N, K \} \) as \( M = N = \frac{1}{2}I, K = -\frac{1}{2(1 - \epsilon)}Z \) and prove that they form a feasible solution to the dual SDP.

\[
Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'_{A_iA_o}}^T \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = \sum_{i,j} Tr \left[ \frac{1}{2(1 - \epsilon)}Z[i] \langle j | \right] N^T(\langle i | j \rangle) \otimes Z \tag{S62}
\]

\[
= -\frac{1}{2(1 - \epsilon)}(N^T(\langle 0 | 0 \rangle) - N^T(\langle 1 | 1 \rangle)) \otimes Z \tag{S63}
\]

\[
= -\frac{1}{2} Z \otimes Z. \tag{S64}
\]

Therefore,
\[
M'_{A_i} \otimes I_{A_o} + Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'_{A_iA_o}}^T \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = \frac{1}{2} I \otimes I - \frac{1}{2} Z \otimes Z \geq 0, \tag{S65}
\]

\[
N_{A_i} \otimes I_{A_o} - Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'A_iA_o} \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = \frac{1}{2} I \otimes I + \frac{1}{2} Z \otimes Z \geq 0. \tag{S66}
\]

- When \( \frac{1}{2} < p \leq 1 \), we set the dual variables \( \{ M, N, K \} \) as \( M = N = |1 \rangle \langle 1 |, K = \frac{1}{2}I + \frac{p - 2p\epsilon - 1}{2(1 - \epsilon)}Z \) and prove that they form a feasible solution to the dual SDP.

\[
Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'_{A_iA_o}}^T \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = \sum_{i,j} \left( \frac{1}{2} \right) Tr[|i \rangle \langle j |] + \left( \frac{\epsilon - 2p \epsilon - 1}{2(1 - \epsilon)} \right) Tr[Z[|i \rangle \langle j |]] N^T(\langle i | j \rangle) \otimes Z \tag{S67}
\]

\[
= \sum_{i} \left( \frac{1}{2} + (-1)^i \frac{\epsilon - 2p \epsilon - 1}{2(1 - \epsilon)} \right) N^T(\langle i | i \rangle) \otimes Z \tag{S68}
\]

\[
= \left( \frac{-p \epsilon}{1 - \epsilon} N^T(\langle 0 | 0 \rangle) + \frac{p + 1 - \epsilon}{1 - \epsilon} N^T(\langle 1 | 1 \rangle) \right) \otimes Z \tag{S69}
\]

\[
= |1 \rangle \langle 1 | \otimes Z \tag{S70}
\]

\[
= |10 \rangle \langle 10 | - |11 \rangle \langle 11 |. \tag{S71}
\]

Therefore,
\[
M'_{A_i} \otimes I_{A_o} + Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'_{A_iA_o}}^T \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = 2|10 \rangle \langle 10 | \geq 0. \tag{S72}
\]

\[
N_{A_i} \otimes I_{A_o} - Tr_{A_iA_o}[\{K_{A_i}^T \otimes J_{N'A_iA_o} \otimes O_{A_o}\}(J_{id_{A_iA_o}} \otimes I_{A_iA_o})] = 2|11 \rangle \langle 11 | \geq 0. \tag{S73}
\]

Hence, we conclude that under different \( p \) values, the corresponding solutions \( \{ M, N, K \} \) are all feasible. It can be easily verified that \( -\text{Tr}[KO] = \frac{|1 - 2p| + 1}{1 - \epsilon} \) for all these cases, which implies
that the inequality $\gamma_0(N) \geq \frac{1 - 2p + 1}{1 - e}$ always holds. Combining it with the primal part, we have $\gamma_0(N) = \frac{1 - 2p + 1}{1 - e}$.

From the proof, we also know that the above retriever $D$ is optimal. Moreover, since the Choi matrix has been given, the corresponding Kraus operators of the retriever can be easily derived.

- When $0 \leq p < \frac{1}{2}$, the Kraus operators for $D_1$ and $D_2$ are $E_{D_1} = \{\alpha|1\rangle\langle 0|, |0\rangle\langle 1|, |\beta\rangle\langle 0|\}$ and $E_{D_2} = \{|1\rangle\langle 0|, |0\rangle\langle 1|, |\beta\rangle\langle 0|\}$, where $\alpha = \sqrt{\frac{1}{1 + (1 - 2p)^2}}$ and $\beta = \sqrt{\frac{(1 - 2p)^2}{1 + (1 - 2p)^2}}$.

- When $\frac{1}{2} \leq p \leq 1$, the Kraus operators for $D_1$ and $D_2$ are $E_{D_1} = \{|1\rangle\langle 1|, |\alpha\rangle\langle 0|, |\beta\rangle\langle 0|\}$ and $E_{D_2} = \{|0\rangle\langle 1|, |\alpha\rangle\langle 1|, |\beta\rangle\langle 1|\}$, where $\alpha = \sqrt{\frac{1}{1 + (1 - 2p)^2}}$ and $\beta = \sqrt{\frac{(1 - 2p)^2}{1 + (1 - 2p)^2}}$.

## 4 Retrieving Cost for Mixed Pauli Channel

The mixed Pauli channel is a common noise model in quantum computers. For single-qubit cases, a quantum state corrupted by mixed Pauli becomes

$\mathcal{N}_{\text{Pauli}}(\rho) = p_1 \rho + p_2 X \rho X + p_3 Y \rho Y + p_4 Z \rho Z,$

where $p_i, p_2, p_3, p_4$ are the corresponding probabilities with $p_1 + p_2 + p_3 + p_4 = 1$ and $0 \leq p_i \leq 1$ for each $p_i \in \{p_1, p_2, p_3, p_4\}$. For general $n$-qubit cases, the noisy state is

$\mathcal{N}_{\text{Pauli}}(\rho) = \sum_{\sigma} p_\sigma \rho_\sigma,$

where the sum is over all the $n$-qubit Pauli operators $\{\sigma\}$, and $\{p_\sigma\}$ are the corresponding probabilities with $\sum_{p_\sigma} p_\sigma = 1$ and $0 \leq p_\sigma \leq 1$. Note that the depolarizing channel is a special case of the mixed Pauli channels, where $p_{X} = p_{Y} = p_{Z}$.

**Proposition S5** Given an $n$-qubit observable $O = \bigotimes_{i=1}^{n} \sigma_i$, where $\sigma_i \in \{X, Y, Z, I\}$, and a mixed Pauli channel $\mathcal{N}$. The minimum cost $\gamma_0(N)$ to retrieve the information $\text{Tr}[\rho O]$ is $\gamma_0(N) = \sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^-$, and the corresponding retriever $D$ can be written in the form of a Choi matrix $J_D = c_1 J_{D_1} + c_2 J_{D_2}$, where $J_{D_1} = \frac{1}{2n} \rho^{1/2} \rho^{1/2}$, $J_{D_2} = \frac{1}{2} (\mathbb{1} \otimes O) \rho$, and $c_1 = -c_2 = \frac{2}{\sqrt{\sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^-}}$. The $\{\sigma^+\}$ are all the Pauli operators that commute with the observable, i.e., $\sigma^+ \in P_n, \{\sigma^+, O\} = 0$, and correspondingly $\sigma^- \in P_n, \{\sigma^-, O\} = 0$.

**Proof** First, we are going to prove $\gamma_0(N) \leq \frac{1}{\sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^-}$. We prove this by utilizing SDP (S3) and showing that the retriever $D$ is a feasible solution.

For an arbitrary state $\rho$, after the mixed Pauli channel, it becomes $\rho' = \mathcal{N}(\rho) = \sum_{\sigma^+} p_\sigma^+ \sigma^+ \rho \sigma^+ + \sum_{\sigma^-} p_\sigma^- \sigma^- \rho \sigma^-$. Then, we have

$D \circ \mathcal{N}(\rho) = D(\rho')$

$$= \text{Tr}_A[\rho'^T \otimes I] \cdot J_D$$

$$= \frac{1}{2^n} \cdot \left( \sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^- \right) \text{Tr}_A[\rho'^T \otimes I] \cdot (\mathbb{1} \otimes O)$$

$$= \frac{1}{2^n} \cdot \left( \sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^- \right) \text{Tr}_A[\rho'^T \otimes I]$$(S78)

$$= \frac{1}{2^n} \cdot \left( \sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^- \right) \text{Tr}[\rho'^T O]$$

(S80)

Since the transpose is linear, we have $\text{Tr}[\rho'^T O] = \text{Tr}[(\rho' O)^T] = \text{Tr}[\rho' O]$. Therefore,

$D \circ \mathcal{N}(\rho) = \frac{1}{2^n} \cdot \left( \sum_{\sigma^+} p_\sigma^+ - \sum_{\sigma^-} p_\sigma^- \right) \text{Tr}[\rho' O]$
\begin{align}
\frac{1}{2^n} \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) O \text{Tr} \left[ \left( \sum_{\sigma_+} p_{\sigma_+} \sigma^+ + \sum_{\sigma_-} p_{\sigma_-} \sigma^- \right) O \right] &= \frac{1}{2^n} \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) O \text{Tr} \left[ \left( \sum_{\sigma_+} p_{\sigma_+} \sigma^+ - \sum_{\sigma_-} p_{\sigma_-} \rho O \sigma^+ \right) \right] (S83) \\
\frac{1}{2^n} \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) O \text{Tr} \left[ \left( \sum_{\sigma_+} p_{\sigma_+} \rho O \sigma^+ + \sum_{\sigma_-} p_{\sigma_-} \rho O \sigma^- \right) \right] &= \frac{1}{2^n} \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) O \text{Tr} \left[ \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) \rho O \right] (S84) \\
\frac{1}{2^n} \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) O \text{Tr} \left[ \left( \sum_{\sigma_+} p_{\sigma_+} \sigma^+ \right) \right] &= \frac{1}{2^n} \left( \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \right) O \text{Tr} [\rho O] (S87)
\end{align}

where the third equality follows from the commutative properties of \( \{ \sigma_+ \} \) and \( \{ \sigma_- \} \) with respect to the observable, the fourth equality follows from the cyclic property of trace, and the fifth equality follows from the identities that \( \sigma^+ \sigma^+ = \sigma^- \sigma^- = I^\otimes n \). Then, it is easy to check that

\[
\text{Tr}[OD \circ N(\rho)] = \frac{1}{2^n} \text{Tr}[O^2] \text{Tr}[\rho O] \geq \frac{1}{2^n} \text{Tr}[I^\otimes n] \text{Tr}[\rho O] = \text{Tr}[\rho O] (S88)
\]

which means that the retriever \( D \) is a feasible protocol, and thus we have \( \gamma_O(N) \leq |c_1| + |c_2| = \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \).

Second, we use the dual SDP (S18) to show that \( \gamma_O(N) \geq \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-} \). We set the dual variables as \( M = N = \frac{1}{2^n} I, K = qO, \) where \( n \) is the number of qubits and \( q \equiv -\frac{1}{2^n} (\sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma_-} p_{\sigma_-}) \). We will show the variables \( \{ M, N, K \} \) is a feasible solution to the dual problem.

\[
\text{Tr}_{A_i A_n}(K^T_{A_i} \otimes J^T_{A_n A_n'} \otimes O_{A_n})(J_{d A_i A_n} \otimes I_{A_n A_n'})
\]

\[
= \sum_{i,j,m,n} \text{Tr}_{A_i A_n} \left[ q_{ij} \sigma^T \otimes \frac{1}{2^n} \left( \sum_{\sigma} p_{\sigma} \sigma^+ \right) O \right] (S91)
\]

\[
= \sum_{i,j,m,n} \text{Tr}_{A_i A_n} \left[ \sum_{\sigma} q_{ij} \sigma^T \otimes \frac{1}{2^n} \left( \sum_{\sigma} p_{\sigma} \sigma^+ + \sum_{\sigma_-} p_{\sigma_-} \sigma^- \right) O \right] (S92)
\]

\[
= \sum_{i,j,m,n} \text{Tr}_{A_i A_n} \left[ \sum_{\sigma} q_{ij} \sigma^T \otimes O \right] (S93)
\]

\[
= \sum_{i,j,m,n} \text{Tr}_{A_i A_n} \left[ \sum_{\sigma} q_{ij} \sigma^T \otimes O \right] (S94)
\]

\[
= \sum_{i,j,m,n} \text{Tr}_{A_i A_n} \left[ \sum_{\sigma} q_{ij} \sigma^T \otimes O \right] (S95)
\]

\[
= \sum_{\sigma} q_{ij} \sigma^T \otimes O (S96)
\]

\[
= q \left( \sum_{\sigma^+} p_{\sigma^+} \sigma^+ \otimes O + \sum_{\sigma^-} p_{\sigma^-} \sigma^- \otimes O \right) (S97)
\]

\[
= -\frac{1}{2^n} O \otimes O, (S98)
\]
where $O_T^j \equiv \langle j|O^T|i \rangle$. This implies that

$$M_{A'_i} \otimes I_{A'_o} + \text{Tr}_{A'_i A'_o} [(K_{A'_i}^T \otimes J_{A'_i}^T \otimes O_{A'_i})(J_{id_{A'_i A'_o}} \otimes I_{A'_i A'_o})] = \frac{1}{2} I \otimes I - \frac{1}{2} O^T \otimes O \geq 0,$$

(S99)

$$N_{A'_i} \otimes I_{A'_o} - \text{Tr}_{A'_i A'_o} [(K_{A'_i}^T \otimes J_{A'_i}^T \otimes O_{A'_i})(J_{id_{A'_i A'_o}} \otimes I_{A'_i A'_o})] = \frac{1}{2} I \otimes I + \frac{1}{2} O^T \otimes O \geq 0,$$

(S100)

which means that $\{M,N,K\}$ is a feasible solution to the dual SDP (S18). Therefore, we have

$$\gamma_O (N) \geq - \text{Tr}[KO] = \frac{1}{2} \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma-} p_{\sigma_-}.$$

Combining this with the primal part, we conclude that

$$\gamma_O (N) = \frac{1}{2} \sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma-} p_{\sigma_-}.$$

From the proof, we also know that the above retriever $D$ is optimal. Moreover, from the given Choi matrix, one could derive that the corresponding Kraus operators of $D_1$ and $D_2$ are

$$E_{D_1} = \left\{ \sqrt{\frac{1}{2n-1}} E_i \Big| E_i \in \{I,X,Y,Z\}^\otimes_n, E_i O = O E_i \right\}$$

and

$$E_{D_2} = \left\{ \sqrt{\frac{1}{2n-1}} E_i \Big| E_i \in \{I,X,Y,Z\}^\otimes_n, E_i O = - O E_i \right\}.$$

Note that the depolarizing channel is a special case of the mixed Pauli channels, where all noise coefficients are the same. For the single-qubit case, the depolarizing channel is $N_{\text{depo}}(\rho) = (1-\epsilon)\rho + \epsilon I$, which is equivalent to Eq. (S74) by setting $p_x = p_y = p_z = \frac{1}{4}$ and $p_I = 1 - p_x - p_y - p_z = 1 - \frac{3}{4}$. The $n$-qubit depolarizing channel is $N_{n-\text{depo}}(\rho) = (1-\epsilon)\rho + \epsilon I^n$, by setting the parameters in Eq. (S75) as $p_\sigma = \frac{1}{4}$ and $p_I = 1 - \sum_\sigma p_\sigma = 1 - \epsilon n - 1$, where $\{\sigma\}$ are the $n$-qubit Pauli operators excluding $I^n$, and $\{p_\sigma\}$ are the corresponding probabilities. It is obvious that $I^n$ commutes with any observable, i.e., $I \in \sigma^+$. Since the rest probabilities are identical, we have $\sum_{\sigma_+} p_{\sigma_+} - \sum_{\sigma-} p_{\sigma_-} = 1 - \epsilon$, which means that the retrieving cost for depolarizing channel is $\gamma_O (N_{n-\text{depo}}) = 1 - \epsilon^{-1}$ for $O$ being an $n$-qubit Pauli operator.

Accepted in Quantum 2023-04-03, click title to verify. Published under CC-BY 4.0.